Natural language processing to identify reasons for sex disparity in statin prescriptions

Celeste Witting [a,1], Zahra Azizi [a,b,1], Sofia Elena Gomez [a], Alban Zammit [c], Ashish Sarraju [d], Summer Ngo [a], Tina Hernandez-Boussard [e,f,g,2], Fatima Rodriguez [a,2,*]

[a] Stanford University Division of Cardiovascular Medicine and Cardiovascular Institute, Department of Medicine, Stanford University, Center for Academic Medicine, Mail Code 5687, 453 Quarry Road, Palo Alto, Stanford, CA, USA
[b] Center for Digital Health, Stanford University, Stanford, CA, USA
[c] Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA
[d] Department of Cardiovascular Medicine, Heart, Vascular & Thoracic Institute, Cleveland Clinic, Cleveland, OH, USA
[e] Department of Medicine, Biomedical Informatics, Stanford University, Stanford, CA, USA
[f] Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
[g] Department of Surgery, Stanford University School of Medicine, Stanford, CA, USA

GRAPHICAL ABSTRACT
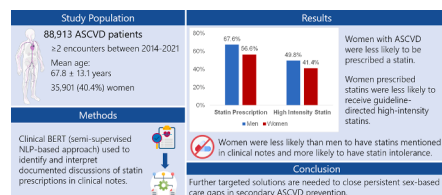
ARTICLE INFO

ABSTRACT

Background: Statins are the cornerstone of treatment of patients with atherosclerotic cardiovascular disease (ASCVD). Despite this, multiple studies have shown that women with ASCVD are less likely to be prescribed statins than men. The objective of this study was to use Natural Language Processing (NLP) to elucidate factors contributing to this disparity.
Methods: Our cohort included adult patients with two or more encounters between 2014 and 2021 with an ASCVD diagnosis within a multisite electronic health record (EHR) in Northern California. After reviewing structured EHR prescription data, we used a benchmark deep learning NLP approach, Clinical Bidirectional Encoder Representations from Transformers (BERT), to identify and interpret discussions of statin prescriptions documented in clinical notes. Clinical BERT was evaluated against expert clinician review in 20% test sets.
Results: There were 88,913 patients with ASCVD (mean age 67.8 ± 13.1 years) and 35,901 (40.4%) were women. Women with ASCVD were less likely to be prescribed statins compared with men (56.6% vs 67.6%, $p < 0.001$), and, when prescribed, less likely to be prescribed guideline-directed high-intensity dosing (41.4% vs 49.8%, $p < 0.001$). These disparities were more pronounced among younger patients, patients with private insurance, and those for whom English is their preferred language. Among those not prescribed statins, women were less likely

* Corresponding author.
  E-mail address: frodrigu@stanford.edu (F. Rodriguez).
[1] These authors contributed equally to this work.
[2] Joint senior authors.

than men to have statins mentioned in their clinical notes (16.9% vs 19.1%, *p* <0.001). Women were less likely than men to have statin use reported in clinical notes despite absence of recorded prescription (32.8% vs 42.6%, *p* <0.001). Women were slightly more likely than men to have statin intolerance documented in structured data or clinical notes (6.0% vs 5.3%, *p*=0.003).

Conclusions: Women with ASCVD were less likely to be prescribed guideline-directed statins compared with men. NLP identified additional sex-based statin disparities and reasons for statin non-prescription in clinical notes of patients with ASCVD.

Non-standard Abbreviations and Acronyms

ASCVD	atherosclerotic cardiovascular disease
BERT	Bidirectional Encoder Representations from Transformers
EHR	electronic health record
ICD	International Classification of Diseases
NLP	natural language processing

1. Introduction

Cardiovascular disease is the leading cause of death for men and women worldwide [1]. Contemporary practice guidelines recommend statins for all patients with atherosclerotic cardiovascular disease (ASCVD) [2]. A large body of evidence supports the efficacy and safety of statins for secondary prevention of major adverse cardiovascular events, with comparable efficacy among both women and men [3–5]. However, women with ASCVD are significantly less likely to be treated with statins as compared with men [1,6–9]. The reasons for this disparity are multifactorial. Women are less likely than men to be informed on their ASCVD risk and offered a statin by their clinician [9, 10], but are also less likely to fill their prescription and more likely to discontinue their statin [11], in part because women may experience greater side effects from statins [9,10].

Prior studies seeking to identify reasons for sex-based disparities in statin use have often relied on patient and clinician surveys from specialty practice sites [9,10,12], which may limit generalizability to real-world practice settings. Free-text notes in the electronic health record (EHR) provide a more direct view into clinical practice and are increasingly accessible on a large scale with the expansion of artificial intelligence technologies including natural language processing (NLP) [13]. The aim of this study was to use NLP to analyze clinical notes in the EHR of patients with ASCVD, with a goal of identifying drivers behind sex-based disparities in statin use.

2. Methods

This is a multi-site retrospective cohort study at Stanford Health Care Alliance (SHA) from 2014 to 2021. SHA is an integrated health system which comprises three sites (academic, community hospital and community practice network). The study was approved by the Stanford University Institutional Review Board (Protocol 47644). Informed consent was waived under exemption 4: research on existing data.

All patients between the ages of 18 and 89 years diagnosed with ASCVD between 2014 and 2021, and with at least 2 encounters in SHA databases were included. The diagnosis of ASCVD was defined using International Classification of Diseases (ICD) codes 9 and 10. These codes encompass coronary artery disease, cerebrovascular disease, and peripheral artery disease (Table S1). Polyvascular disease was defined as the presence of atherosclerosis in two or more vascular beds. The index diagnosis was considered at the time of first diagnosis. Socio-demographic (age, sex, race and ethnicity, insurance type, preferred language), clinical (comorbidities) and laboratory (cholesterol, hemoglobin A1C) data were obtained within 6 months prior to the initial diagnosis.

2.1. Outcomes

The primary endpoint of this study was the presence of a statin prescription and prescribed dosage (intensity). We first identified statin prescriptions and dosage using EHR structured data fields and RxNorm codes (Table S2). Statin prescriptions were included at the time of index ASCVD diagnosis, or within six months prior or one month after index diagnosis. Statin dosages were classified to high versus low or moderate dose using well-established definitions [14].

The secondary endpoint of the study was the reason for statin non-prescription. For patients without a statin prescription or a statin allergy documented in structured data, we analyzed EHR clinical notes for mentions of statins, using an NLP model (Fig. 1). Clinical notes were included from inpatient and outpatient encounters. Statin mentions were extracted from clinical notes within one month after the index diagnosis. The reasons for statin non-prescription were classified into adherence reported (e.g., patient taking a statin that is not captured in structured data), patient side effects (muscle and non-muscle), guideline-discordant practice (e.g., clinician not thinking a statin is indicated in patient with history of ASCVD), patient preference (e.g., patient does not want to take a statin), and non-specific. We also analyzed statin intolerance among non-prescribed patients, defined as having a statin allergy in structured data or side effects mentioned in clinical notes.

2.2. Natural language processing

We used Clinical Bidirectional Encoder Representations from Transformers (Clinical BERT), an NLP deep learning model pretrained on clinical notes from the MIMIC III dataset [15], as previously described [13]. We finetuned this model to predict the reason for statin non-prescription (downstream task). To develop a gold standard (ground truth) for NLP training, we manually annotated a random sample of clinical notes (N=1742). Four co-authors manually annotated the notes to determine statin prescription, or, in case of statin non-prescription, extracted reasons for non-prescription according to the aforementioned categories. To assess agreement between reviewers, overlapping review was performed in a set of 50 patients and a set of 100 patients. In case of discrepancy between reviewers, the notes were reviewed by a third reviewer (clinician expert). After manual annotation, the dataset was randomly split to 80% training set and 20% test set. We used 10-fold cross-validation to validate the model and tune the hyperparameters (learning rate, number of epochs, strength of weight decay, and Adam's epsilon value). For assessing statin prescription and non-prescription, a first fine-tuned model consisted of a binary classification for mention in clinical notes versus no mention in clinical notes. Extracting reasons for statin non-prescription involved a second model, namely a multi-class classification according to the aforementioned categories. To evaluate model performance, we assessed precision, recall, Area Under Curve score and F1 score. We used bootstrapping to calculate 95% confidence intervals for model performances. After training the datasets and evaluation, the highest performing model was applied across all clinical notes to determine the primary and secondary outcomes.
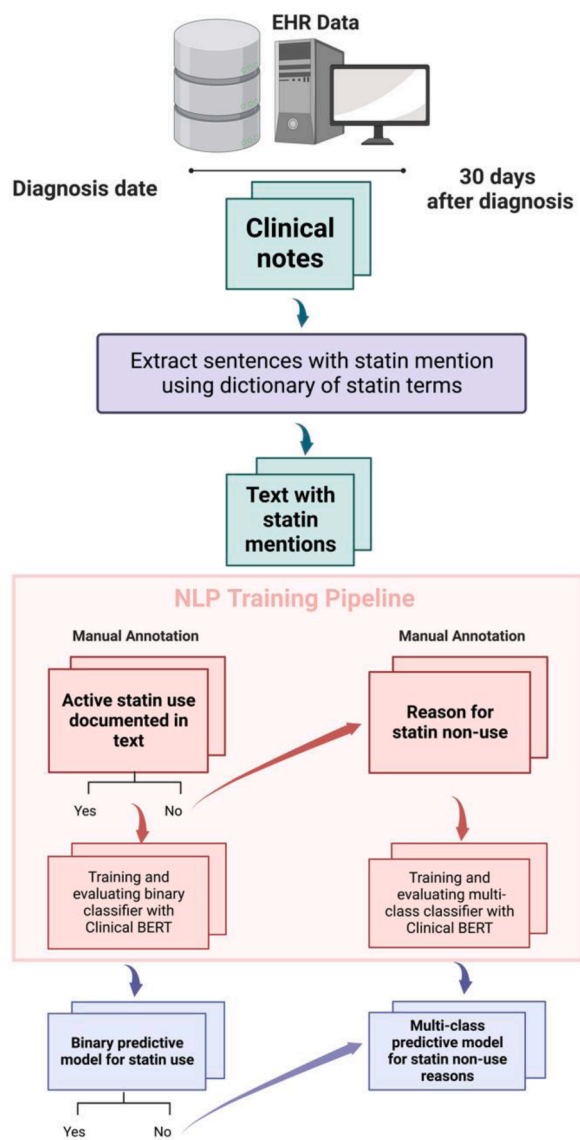
Fig. 1. Schematic of training, internal validation, and application of a deep learning model (Clinical BERT) for identifying prevalence and reasons of statin non-use from unstructured clinical notes of patients with ASCVD.
Legend: Abbreviations: BERT: Bidirectional Encoder Representations from Transformers, NLP: natural language processing.

2.3. Statistical analysis

Descriptive results were reported as means and standard deviations for continuous variables and percentages for categorical variables. Chi-squared and t-test were used for categorical variables and continuous variables, respectively, to assess the differences in baseline characteristics between women and men in normally distributed data. Wilcoxon Rank test was used for non-normally distributed continuous variables. Multivariable Logistic Regression models were used to identify socio-demographic factors associated with statin prescription and non-prescription. The models were adjusted for clinical factors associated with statin use. Pre-defined subgroup analyses were conducted to assess the consistency of the associations between sex and statin prescription and intensity across each of the demographic, social and clinical factors. Within each subgroup, the model was adjusted for other covariates including socio-demographic (age, sex, race and ethnicity, insurance type, preferred language, and provider location),and clinical factors (ASCVD type, comorbidities, and other prescribed lipid lowering

agents). Analyses were performed using Python (v.3.7) for NLP models and R software (v.4.2) for other models.

3. Results

Of 88,913 patients with ASCVD (mean age 67.8 ± 13.1 years), 40.4% were women. Compared with men, women were less likely to have private insurance and more likely to have a non-English preferred language (Table 1). Moreover, women were less likely to smoke, had lower rates of comorbidities or polyvascular disease, and had higher cholesterol levels.

Overall, 56,147 (63.1%) of patients with ASCVD were prescribed a statin (Figs. 2 and 3). Women were less likely to be prescribed statins as compared with men (56.6% vs 67.6%, $p < 0.001$, OR: 0.63 with 95% CI 0.61–0.64; Figs. 2 and 3). The results were consistent across subgroups, although the magnitude of effect differed by age, race, ethnicity, ASCVD type, insurance status, preferred language, and comorbidities (Fig. 4). The sex disparity was more prominent in younger patients, non-Hispanic white patients, English speakers, those with a diagnosis of coronary artery disease, and those without diabetes. In additional subgroup analyses, Black and Hispanic patients and patients on Medicaid were less likely to be prescribed statins than non-Hispanic White and privately insured patients, respectively (Fig. S1); however, the sex disparity in statin use was less prominent in these subgroups (Fig. 4).

Among those who were prescribed statins, less than half (46.8%)

Table 1
Baseline patient characteristics.

	Overall N=88,913	Women N=35,901	Men N=53,012
Socio-demographics			
Sex			
Female	35,901 (40.4)		
Age (years)	67.8 (13.1)	68.3 (14.0)	67.4 (12.5)
Race/Ethnicity			
Hispanic	9009 (10.1)	4080 (11.4)	4929 (9.3)
Non-Hispanic Asian	13,811 (15.5)	5713 (15.9)	8098 (15.3)
Non-Hispanic Black	4871 (5.5)	2478 (6.9)	2393 (4.5)
Non-Hispanic White	49,189 (55.3)	19,063 (53.1)	30,126 (56.8)
Other	9149 (10.3)	3517 (9.8)	5632 (10.6)
Unknown	2884 (3.2)	1050 (2.9)	1834 (3.5)
Insurance Type			
Private	18,548 (20.9)	6602 (18.4)	11,946 (22.5)
Medicaid	6207 (7.0)	2759 (7.7)	3448 (6.5)
Medicare	48,567 (54.6)	20,486 (57.1)	28,081 (53.0)
Other	9073 (10.2)	3379 (9.4)	5694 (10.7)
Unknown	6518 (7.3)	2675 (7.5)	3843 (7.2)
Preferred Language			
Non-English	11,223 (12.6)	5341 (14.9)	5882 (11.1)
Comorbidities			
Smoking	4695 (5.3)	1617 (4.5)	3078 (5.8)
BMI (kg/m^2)	28.2 (31.8)	27.8 (22.9)	28.5 (36.7)
ASCVD type			
Cerebrovascular	19,493 (21.9)	10,708 (29.8)	8785 (16.6)
Coronary Artery	46,619 (52.4)	15,372 (42.8)	31,247 (58.9)
Peripheral Arterial	8282 (9.3)	4223 (11.8)	4059 (7.7)
Polyvascular	14,519 (16.3)	5598 (15.6)	8921 (16.8)
CKD	14,732 (16.6)	5268 (14.7)	9464 (17.9)
Liver Disease	6521 (7.3)	2737 (7.6)	3784 (7.1)
Diabetes	21,940 (24.7)	8455 (23.6)	13,485 (25.4)
Hypertension	22,326 (25.1)	8605 (24.0)	13,721 (25.9)
2-year Charlson Index	2.5 (2.7)	2.6 (2.6)	2.5 (2.7)
Lab data			
Total Cholesterol (mg/dL)	169.5 (46.2)	182.2 (46.3)	160.4 (44.0)
LDL (mg/dL)	95.8 (38.1)	101.3 (38.9)	91.9 (37.0)
HDL (mg/dL)	53.7 (18.6)	61.0 (20.0)	48.5 (15.5)
HbA1C (%)	6.3 (1.4)	6.3 (1.4)	6.4 (1.4)

Categorical: Number (Percentage), Numerical: Mean (Standard Deviation).
Abbreviations: BMI: body mass index, ASCVD: atherosclerotic cardiovascular disease, CKD: Chronic Kidney Disease, LDL: Low Density Lipoprotein, HDL: High Density Lipoprotein, HbA1C: Hemoglobin A1c.
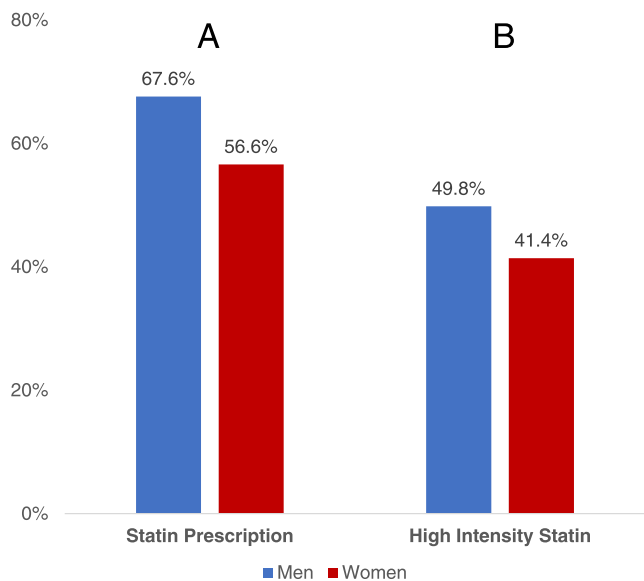
Fig. 2. Statin prescriptions and high-intensity statin prescription in men and women with atherosclerotic cardiovascular disease.

Legend: A: Statin prescription rates, in all patients with atherosclerotic cardiovascular disease (ASCVD), stratified by sex. **B**: High-intensity statin use, in patients with ASCVD who were prescribed statins with known dose, stratified by sex.

were on guideline-recommended therapy with a high-intensity statin (compared with low- or moderate-intensity). Women were less likely to be prescribed a high-intensity statin than men (41.4% vs 49.8%, *p* <0.001, OR: 0.71 with 95% CI 0.68–0.73; Figs. 2 and 3). Similar to rates of statin prescription overall, sex differences in the prescription of high-

intensity statins were present in all subgroups (Fig. 5).

Women were more likely to have a statin allergy documented in structured data fields (3.7% women vs 2.8% men, *p* < 0.001). Among patients without statin prescriptions in structured data fields or documented statin allergies, NLP analysis of clinical notes showed that nearly 20% of patients had mentions of statins in their clinical notes (Fig. 3). Women were less likely to have a mention of statin in their clinical notes as compared with men (16.9% women vs 19.1% men, *p* < 0.001). Women were also less likely than men to have notes indicating that they do in fact take a statin, despite lack of prescription in structured data (32.8% women vs 42.6% of men, *p* < 0.001). Women were more likely than men to have statin intolerance indicated in their structured data or clinical notes (6.0% women vs 5.3% men, *p* = 0.003). Other reasons for non-prescription were not significantly different by sex.

4. Discussion

In this study of over 80,000 patients with ASCVD receiving care within a large health care system, we found persistent sex disparities in guideline-directed statin prescriptions and documented reasons for statin non-prescription. Review of clinical notes uncovered greater sex-based disparities in statin prescription than seen with structured data alone, and that women were more likely than men to have statin intolerance as their reason for statin non-prescription.

We found that women were less likely to be prescribed a statin than men, and, among patients prescribed statins, women were less likely than men to be prescribed a guideline-recommended high-intensity statin. This is in alignment with previous research on sex disparities in statin prescription in patients with ASCVD, with women using statins approximately 7–11% less than men [6–9]. While women are less likely to be prescribed statins in all age groups, we found a more pronounced disparity in statin prescription by sex among younger patients (age 20–50 years old). This finding may be explained due to the pervasive
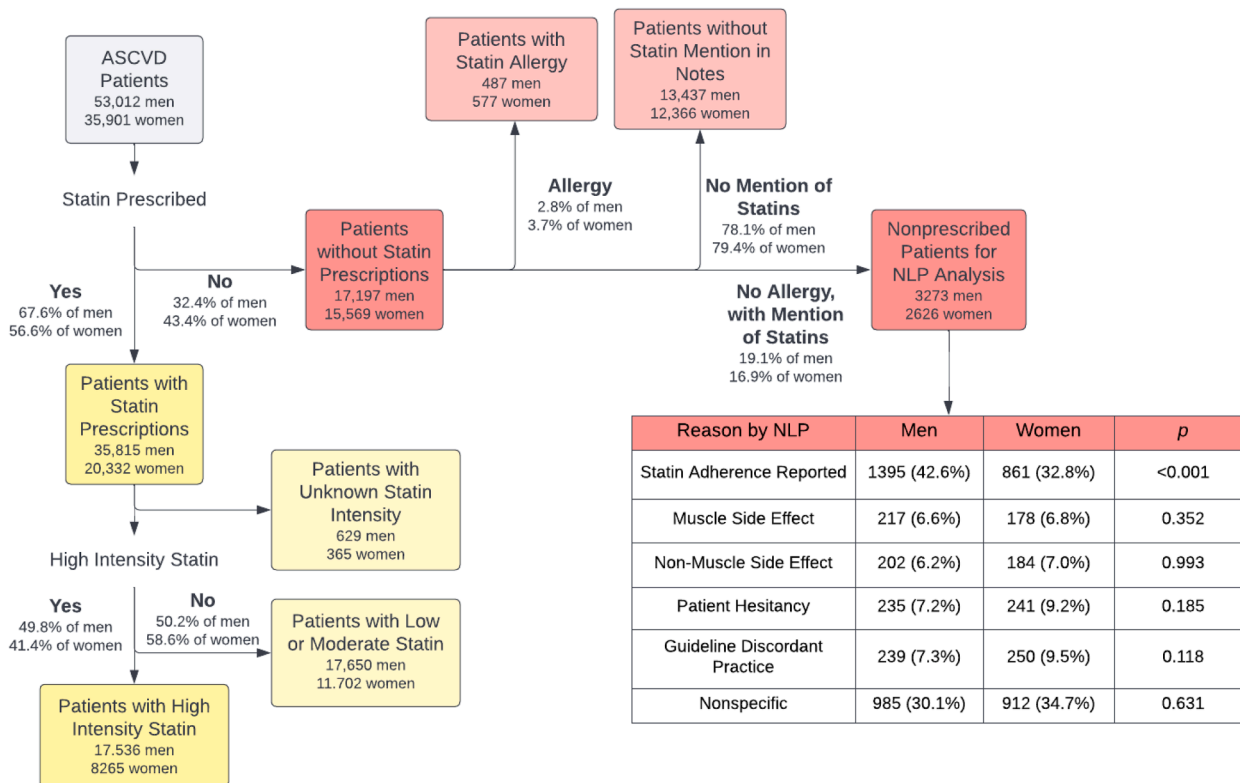

Fig. 3. Statin Prescriptions by Intensity and NLP analysis of clinical notes.
Legend: ASCVD: atherosclerotic cardiovascular disease, NLP: natural language processing.

Subgroup	Female vs Male	Male (%)	Female (%)
Age			
20-50	0.48(0.44-0.53)	51.1%	31.0%
50-75	0.67(0.64-0.69)	68.6%	57.5%
≥75	0.78(0.75-0.83)	70.4%	63.0%
Race and Ethnicity			
Non-Hispanic White	0.65(0.62-0.68)	68.0%	55.6%
Non-Hispanic Black	0.86(0.76-0.97)	61.7%	57.5%
Non-Hispanic Asian	0.69(0.64-0.74)	70.7%	60.3%
Hispanic	0.69(0.63-0.76)	64.7%	54.7%
Other	0.73(0.66-0.80)	66.4%	57.5%
Insurance Status			
Private	0.53(0.49-0.56)	65.2%	45.5%
Medicare	0.73(0.70-0.76)	69.7%	60.9%
Medicaid	0.71(0.64-0.79)	61.1%	52.0%
Other	0.74(0.68-0.82)	64.9%	55.3%
Preferred Language			
English	0.66(0.64-0.68)	67.4%	55.4%
Not English	0.81(0.75-0.88)	68.4%	63.3%
Smoking			
No	0.68(0.66-0.70)	67.3%	56.5%
Yes	0.63(0.55-0.72)	71.2%	59.3%
ASCVD type			
Polyvascular	0.72(0.67-0.78)	73.7%	67.1%
Coronary Artery	0.63(0.60-0.65)	70.4%	60.6%
Cerebrovascular	0.71(0.67-0.75)	59.1%	50.3%
Peripheral Arterial	0.82(0.75-0.90)	49.5%	44.0%
Hypertension			
No	0.67(0.64-0.69)	65.2%	53.4%
Yes	0.73(0.69-0.78)	74.2%	66.8%
Diabetes			
No	0.65(0.63-0.67)	63.8%	51.4%
Yes	0.81(0.76-0.87)	78.5%	73.5%
Chronic Kidney Disease			
No	0.67(0.65-0.69)	93.8%	38.4%
Yes	0.74(0.69-0.80)	74.7%	68.3%
Liver Disease			
No	0.67(0.65-0.69)	68.1%	57.0%
Yes	0.8(0.72-0.89)	59.4%	51.9%
2-year Charlson Index			
<3	0.66(0.63-0.68)	64.6%	52.3%
≥3	0.73(0.69-0.76)	72.5%	64.1%

0.0   0.5   1.0   1.5
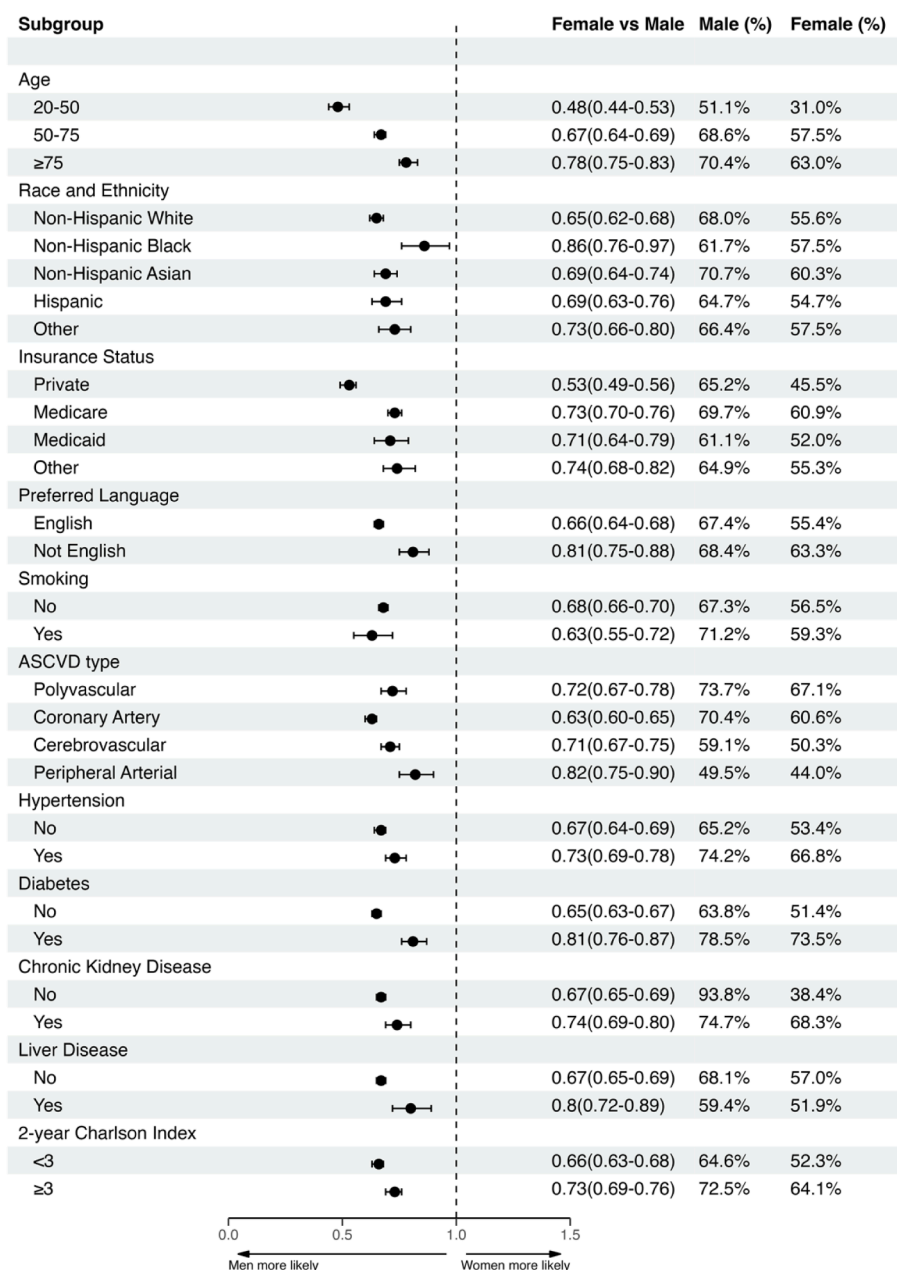
←— Men more likely   Women more likely —→

Fig. 4. Subgroup analysis of association between sex and statin prescription.

Legend: Subgroup analyses were conducted to assess the consistency of the associations between sex and statin use across each of the demographic, social and clinical factors. Within each subgroup the model is adjusted for other covariates including socio-demographic (age, sex, race and ethnicity, insurance type, preferred language, and provider location), and clinical factors (ASCVD type, comorbidities, and other prescribed lipid lowering agents). ASCVD: atherosclerotic cardiovascular disease.

under-recognition and undertreatment of heart disease in younger women and concerns of around statin use in women of reproductive age, both from patients and providers [16]. Our subgroup analyses also revealed that there is greater sex disparity in statin prescription among non-Hispanic White patients (compared with non-Hispanic Black patients), among patients with private insurance (compared with Medicare or Medicaid), and among patients for whom English is their preferred language (compared with those with another preferred language). These differences highlight the importance of considering intersectionality in unpacking reasons behind disparities. It is possible that groups which are under- prescribed statins overall (in our patients, Black, Hispanic, and patients on Medicaid) have a less prominent sex disparity because both sexes are under-prescribed statins [17,18].

The use of NLP in this study expands on prior research by allowing large-scale analysis of unstructured data in clinical notes. Prior research pointing to patient and clinician factors for sex disparities in statin use have relied on targeted survey data [9,10], which is prone to bias and fails to capture the magnitude of each factor on statin non-prescription.

In analyzing reasons for statin non-prescription within clinical notes, we aimed to assess the relative weight of patient and clinician factors by how often different factors are mentioned. Although patients and clinicians may report reasons for statin non-prescription when given a survey, these discussions may not be routinely happening in real-world practice settings or may point to a lack of documentations. Further investigation is needed to better understand the reasons for statin non-prescription in whom there is no documentation of a statin discussion, and why there is a sex disparity in documentation.

Our NLP analysis of clinical notes showed that, among patients without statin prescriptions in structured data and with mentions of statins in their notes, women are less likely than men to have statin adherence reported in their notes. In these patients, although structured EHR fields do not capture statin use through prescription data, the clinician documents that the patient is taking statin. This may represent an outside prescription but highlights the need to consider data from clinical notes to understand practice patterns and opportunities for targeted interventions to reduce care gaps. This shows that the inclusion
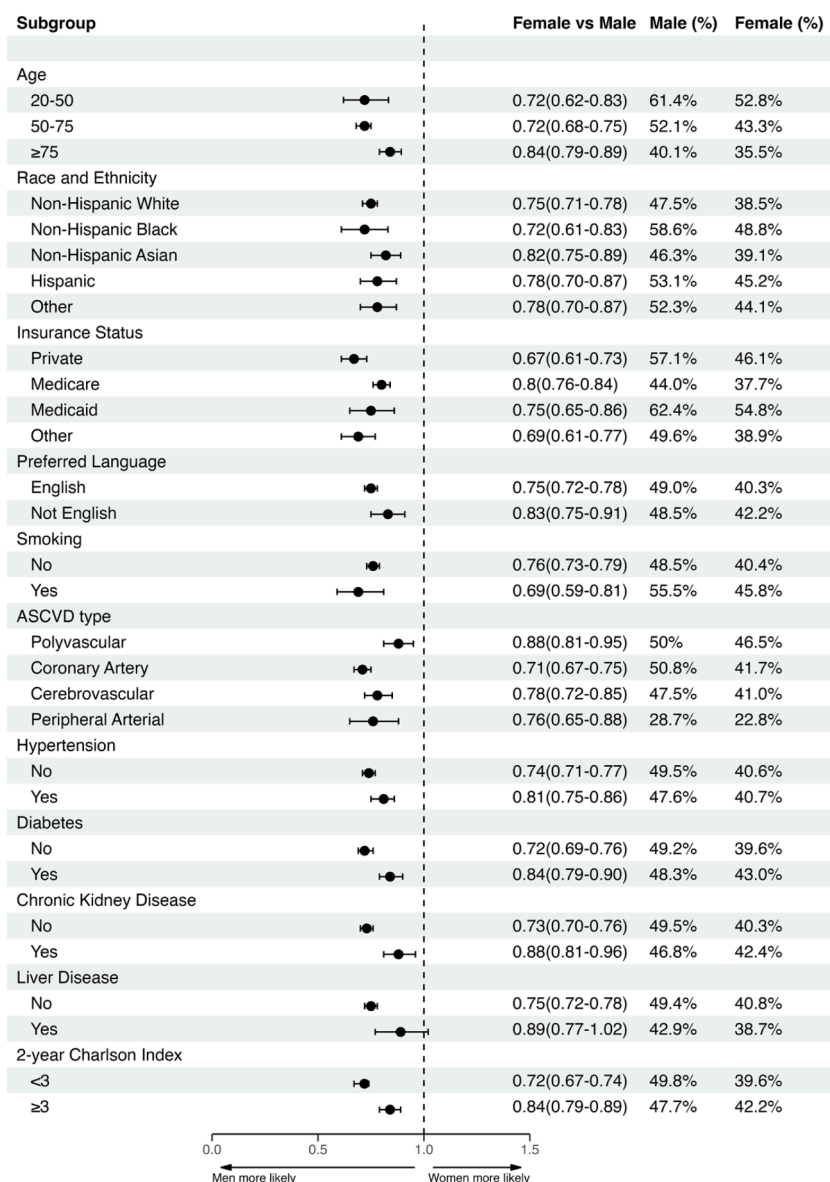
Subgroup		Female vs Male	Male (%)	Female (%)
Age				
20-50		0.72(0.62-0.83)	61.4%	52.8%
50-75		0.72(0.68-0.75)	52.1%	43.3%
≥75		0.84(0.79-0.89)	40.1%	35.5%
Race and Ethnicity				
Non-Hispanic White		0.75(0.71-0.78)	47.5%	38.5%
Non-Hispanic Black		0.72(0.61-0.83)	58.6%	48.8%
Non-Hispanic Asian		0.82(0.75-0.89)	46.3%	39.1%
Hispanic		0.78(0.70-0.87)	53.1%	45.2%
Other		0.78(0.70-0.87)	52.3%	44.1%
Insurance Status				
Private		0.67(0.61-0.73)	57.1%	46.1%
Medicare		0.8(0.76-0.84)	44.0%	37.7%
Medicaid		0.75(0.65-0.86)	62.4%	54.8%
Other		0.69(0.61-0.77)	49.6%	38.9%
Preferred Language				
English		0.75(0.72-0.78)	49.0%	40.3%
Not English		0.83(0.75-0.91)	48.5%	42.2%
Smoking				
No		0.76(0.73-0.79)	48.5%	40.4%
Yes		0.69(0.59-0.81)	55.5%	45.8%
ASCVD type				
Polyvascular		0.88(0.81-0.95)	50%	46.5%
Coronary Artery		0.71(0.67-0.75)	50.8%	41.7%
Cerebrovascular		0.78(0.72-0.85)	47.5%	41.0%
Peripheral Arterial		0.76(0.65-0.88)	28.7%	22.8%
Hypertension				
No		0.74(0.71-0.77)	49.5%	40.6%
Yes		0.81(0.75-0.86)	47.6%	40.7%
Diabetes				
No		0.72(0.69-0.76)	49.2%	39.6%
Yes		0.84(0.79-0.90)	48.3%	43.0%
Chronic Kidney Disease				
No		0.73(0.70-0.76)	49.5%	40.3%
Yes		0.88(0.81-0.96)	46.8%	42.4%
Liver Disease				
No		0.75(0.72-0.78)	49.4%	40.8%
Yes		0.89(0.77-1.02)	42.9%	38.7%
2-year Charlson Index				
<3		0.72(0.67-0.74)	49.8%	39.6%
≥3		0.84(0.79-0.89)	47.7%	42.2%

0.0   0.5   1.0   1.5
← Men more likely   Women more likely →

Fig. 5. Subgroup analysis of association between sex and rate of high-intensity statin prescription.

Legend: Subgroup analyses were conducted to assess the consistency of the associations between sex and statin intensity (high vs low or moderate), amongst patients with statin prescription with known dose, across each of the demographic, social and clinical factors. Within each subgroup the model is adjusted for other covariates including socio-demographic (age, sex, race and ethnicity, insurance type, preferred language, and provider location), and clinical factors (ASCVD type, comorbidities, and other prescribed lipid lowering agents).

of unstructured data from clinical notes may allow for more accurate quantification of disparities in medication use. We found that addition of statin use captured in clinical notes magnified the sex-based disparities in statin prescriptions. Our analysis also showed that women were more likely than men to be intolerant of statins, when combining structured allergy data with mentions of side effects in clinical notes. This finding has been previously documented [19,20], but it is noteworthy that statin intolerance was only documented in 6.0% of women without statin prescriptions, and that the difference between men and women, although statistically significant, was only 0.7% of unprescribed patients, which is of questionable clinical significance. Prior survey-based studies have reported higher rates of statin side effects in women, ranging from 8 to 31% [9,10], compared with 6.0% in our study. This is likely due to differences in methodology, where patients self-report statin side effects in surveys at a higher rate than their providers are aware of and documenting in clinical notes.

Possible interventions to address disparities in statin therapy include decision support tools and prompts when clinicians are seeing patients with an ASCVD diagnosis [21–23]. If statins are not prescribed, clear documentation of the reasons why can then be used to develop targeted interventions or alternative non-statin therapies. Further use of NLP

approaches in unstructured clinical documentation may identify other barriers to statin use. Effective interventions will need to target the patient, clinician, and system reasons for statin nonadherence, while considering that certain patient groups such as women and racial and ethnic minorities face additional treatment gaps [24].

This study should be interpreted within the context of several limitations. First, this was a single health system study in Northern California which limits generalizability to other populations across the United States, although our system comprises an academic hospital, a community hospital, and a community practice network. Second, although we restricted our cohort to those with at least two ASCVD encounters within our system, some patients also received care elsewhere leading to data fragmentation, as evidenced by our patients who did not have a structured statin prescription but had notes indicating that they were taking a statin. Third, we were unable to disaggregate racial and ethnic groups further or include more robust socioeconomic information due to data limitations. Finally, although BERT performed well with internal validity assessment, further study of NLP models' external generalizability and cost-effectiveness is needed prior to widespread use in clinical practice.

In conclusion, we found that women with ASCVD were under-

prescribed statins as compared with men, and this disparity was more prominent when including analysis of clinical notes using NLP. Documentation of reasons for statin non-prescription was less common in clinical notes for women, and women were slightly more likely to have statin intolerance than men. Further work is needed to identify targeted solutions to close this persistent sex-based care gap in secondary ASCVD prevention.

Sources of funding

CRediT authorship contribution statement

Celeste Witting: Conceptualization, Methodology, Investigation. **Zahra Azizi:** Methodology, Formal analysis, Investigation. **Sofia Elena Gomez:** Investigation. **Alban Zammit:** . **Ashish Sarraju:** Conceptualization, Methodology, Investigation. **Summer Ngo:** Investigation. **Tina Hernandez-Boussard:** Conceptualization, Methodology, Investigation, Supervision. **Fatima Rodriguez:** Conceptualization, Methodology, Investigation, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests.

Disclosures

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajpc.2023.100496.

References

[1] Virani SS, Alonso A, Aparicio HJ, et al. Heart disease and stroke statistics-2021 update: a report from the American Heart Association. Circulation 2021;143(8): e254–743. https://doi.org/10.1161/CIR.0000000000000950.

[2] Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: Executive summary: a report of the American College of Cardiology/ American Heart Association task force on clinical practice guidelines. J Am Coll Cardiol 2019;73(24):3168–209. https://doi.org/10.1016/j.jacc.2018.11.002.

[3] LaRosa JC, Grundy SM, Waters DD, et al. Intensive lipid lowering with atorvastatin in patients with stable coronary disease. N Engl J Med 2005;352(14):1425–35. https://doi.org/10.1056/NEJMoa050461.

[4] Truong QA, Murphy SA, McCabe CH, Armani A, Cannon CP, T.I.M.I. Study Group. Benefit of intensive statin therapy in women: results from PROVE IT-TIMI 22. Circ Cardiovasc Qual Outcomes 2011;4(3):328–36. https://doi.org/10.1161/ CIRCOUTCOMES.110.957720.

[5] Cholesterol Treatment Trialists Collaboration, Fulcher J, O'Connell R, et al. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174,000 participants in 27 randomised trials. Lancet 2015; 385(9976):1397–405. https://doi.org/10.1016/S0140-6736(14)61368-4.

[6] Peters SAE, Colantonio LD, Zhao H, et al. Sex differences in high-intensity statin use following myocardial infarction in the United States. J Am Coll Cardiol 2018; 71(16):1729–37. https://doi.org/10.1016/j.jacc.2018.02.032.

[7] Virani SS, Woodard LD, Ramsey DJ, et al. Gender disparities in evidence-based statin therapy in patients with cardiovascular disease. Am J Cardiol 2015;115(1): 21–6. https://doi.org/10.1016/j.amjcard.2014.09.041.

[8] Eindhoven DC, Hilt AD, Zwaan TC, Schalij MJ, Borleffs CJW. Age and gender differences in medical adherence after myocardial infarction: women do not receive optimal treatment - the Netherlands claims database. Eur J Prev Cardiol 2018;25(2):181–9. https://doi.org/10.1177/2047487317744363.

[9] Nanna MG, Wang TY, Xiang Q, et al. Sex differences in the use of statins in community practice. Circ Cardiovasc Qual Outcomes 2019;12(8):e005562. https:// doi.org/10.1161/CIRCOUTCOMES.118.005562.

[10] Karalis DG, Wild RA, Maki KC, et al. Gender differences in side effects and attitudes regarding statin use in the understanding statin use in America and gaps in patient education (USAGE) study. J Clin Lipidol 2016;10(4):833–41. https://doi.org/ 10.1016/j.jacl.2016.02.016.

[11] Olmastroni E, Boccalari MT, Tragni E, et al. Sex-differences in factors and outcomes associated with adherence to statin therapy in primary care: need for customisation strategies. Pharmacol Res 2020;155:104514. https://doi.org/10.1016/j. phrs.2019.104514.

[12] Cohen JD, Brinton EA, Ito MK, Jacobson TA. Understanding statin use in America and Gaps in Patient Education (USAGE): an internet-based survey of 10,138 current and former statin users. J Clin Lipidol 2012;6(3):208–15. https://doi.org/ 10.1016/j.jacl.2012.03.003.

[13] Sarraju A, Coquet J, Zammit A, et al. Using deep learning-based natural language processing to identify reasons for statin nonuse in patients with atherosclerotic cardiovascular disease. Commun Med 2022;15:88. https://doi.org/10.1038/ s43856-022-00157-w.

[14] Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. Circulation 2019;140(11):e563–95. https://doi.org/10.1161/ CIR.0000000000000677.

[15] Johnson AEW, Pollard TJ, Shen L, et al. Mimic-III, A freely accessible critical care database. Sci Data 2016;3:160035. https://doi.org/10.1038/sdata.2016.35.

[16] U.S. Food & Drug Administration. Statins: drug safety communication - FDA requests removal of strongest warning against using cholesterol-lowering statins during pregnancy. FDA 2022. Published online July 20, 2021. Accessed December 2, 2022. https://www.fda.gov/safety/medical-product-safety-information/statins -drug-safety-communication-fda-requests-removal-strongest-warning-against-usin g-cholesterol. Published online July 20, 2021. Accessed December 2, 2022.

[17] Lewey J, Shrank WH, Bowry ADK, Kilabuk E, Brennan TA, Choudhry NK. Gender and racial disparities in adherence to statin therapy: a meta-analysis. Am Heart J 2013;165(5):665–78. https://doi.org/10.1016/j.ahj.2013.02.011. 678.e1.

[18] Nanna MG, Navar AM, Zakroysky P, et al. Association of patient perceptions of cardiovascular risk and beliefs on statin drugs with racial differences in statin use: insights from the patient and provider assessment of lipid management registry. JAMA Cardiol 2018;3(8):739–48. https://doi.org/10.1001/ jamacardio.2018.1511.

[19] Petretta M, Costanzo P, Perrone-Filardi P, Chiariello M. Impact of gender in primary prevention of coronary heart disease with statin therapy: a meta-analysis. Int J Cardiol 2010;138(1):25–31. https://doi.org/10.1016/j.ijcard.2008.08.001.

[20] Bukkapatnam RN, Gabler NB, Lewis WR. Statins for primary prevention of cardiovascular mortality in women: a systematic review and meta-analysis. Prev Cardiol 2010;13(2):84–90. https://doi.org/10.1111/j.1751-7141.2009.00059.x.

[21] Vani A, Kan K, Iturrate E, et al. Leveraging clinical decision support tools to improve guideline-directed medical therapy in patients with atherosclerotic cardiovascular disease at hospital discharge. Cardiol J 2022;29(5):791–7. https:// doi.org/10.5603/CJ.a2020.0126.

[22] Adusumalli S, Westover JE, Jacoby DS, et al. Effect of passive choice and active choice interventions in the electronic health record to cardiologists on statin prescribing: a cluster randomized clinical trial. JAMA Cardiol 2021;6(1):40–8. https://doi.org/10.1001/jamacardio.2020.4730.

[23] Shah NN, Ghazi L, Yamamoto Y, et al. Rationale and design of a pragmatic trial aimed at improving treatment of hyperlipidemia in outpatients with very high risk atherosclerotic cardiovascular disease: a pragmatic trial of messaging to providers about treatment of hyperlipidemia (PROMPT-LIPID). Am Heart J 2022;253:76–85. https://doi.org/10.1016/j.ahj.2022.07.002.

[24] Ferdinand KC, Senatore FF, Clayton-Jeter H, et al. Improving medication adherence in cardiometabolic disease: practical and regulatory implications. J Am Coll Cardiol 2017;69(4):437–51. https://doi.org/10.1016/j.jacc.2016.11.034.