ORIGINAL RESEARCH

OUTCOMES AND QUALITY

Sex, Age, and Patient Experience in Cardiologist Reviews

A Large-Scale Artificial Intelligence-Enabled Analysis

Andrew Yang, BA, BE, Fatima Rodriguez, MD, MPH, Jennifer P. Woo, MD

ABSTRACT

BACKGROUND Patients are increasingly using online reviews to evaluate cardiologists. Online reviews can provide insights into factors driving patient satisfaction. Little is known about the effects of age and sex on the patient experience with cardiologists.

OBJECTIVES The purpose of this study was to apply natural language processing techniques on online reviews to determine the factors underlying positive and negative patient experiences and the effects of age and sex on the patient experience with cardiologists.

METHODS Mixed effects logistic regression and sentiment analysis were applied to online cardiologist reviews from Healthgrades between 1998 and 2023. The results were then analyzed by sex and age to show trends with respect to rating statistics, sentiment analysis, and frequency of 2-word phrases.

RESULTS There were 100,334 online reviews of 9,461 cardiologists. Female cardiologists received lower average ratings compared to male cardiologists and were 34.5% less likely to receive a positive review (OR: 0.655; 95% CI: 0.481-0.893; $P = 0.015$). Older cardiologists received lower average ratings compared to younger cardiologists (4.145 ± 0.908 vs 4.348 ± 0.795; $P < 0.01$). Positive reviews were associated with time spent with patients (OR: 1.383; 95% CI: 1.251-1.528; $P < 0.01$), answering questions (OR: 2.622; 95% CI: 2.324-2.959; $P < 0.01$), and patients feeling they could trust their providers' decisions (OR: 2.285; 95% CI: 2.053-2.543; $P < 0.01$).

CONCLUSIONS Positive reviews were associated with cardiologists being comprehensive and patients feeling a sense of trust in the relationship. There was a difference in ratings based on age and sex with female and older cardiologists receiving lower ratings. (JACC Adv 2024;3:101046) © 2024 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

T he patient experience is at the core of patient-centered medical care. Online physician review websites are increasingly being used to measure patient experience and publicly evaluate individual physicians.[1] Information available on review websites is being used by over 60% of patients. These websites provide information on physician performance, clinic metrics, and patient satisfaction. A national survey showed that 59% of patients found physician rating websites to be

important when choosing a physician, with 37% avoiding certain providers due to perceived negative feedback.[2] Despite its wide use, the association between patient experience and physician and clinical encounter characteristics remains unclear. Few studies have studied online reviews of cardiologists let alone the effects of age and sex on the patient experience with cardiologists.

The results from prior studies on physician demographics and their associations with aspects of the patient-physician relationship have yielded mixed results. Some studies suggest improved patient experiences with patient-physician racial/ethnic or sex concordance while others do not.[3-6] The patient experience is subject to inherent biases and these biases tend to negatively affect female physicians.[3] Sex disparity is more evident when looking at sex distribution in medical subspecialties, especially in cardiology. Almost 42.6% of internal medicine residents are female; yet only 12.6% of practicing cardiologists are female.[7,8] In light of the existing sex disparity within cardiology, a better understanding of cardiologist demographics such as age and sex in the patient experience is crucial in our collective goal of achieving health equity.

Existing studies of online reviews of cardiologists tend to have small sample sizes and aim to understand the factors that affect the quantitative ratings.[6] Natural language processing (NLP) techniques from rule-based lexicographical approaches to sentiment analysis have allowed us to expand on this work and to harvest valuable qualitative rather than quantitative insights into textual accounts of factors that are important to patients.[9-13] Fine-tuned large language models (LLMs) result in significant improvements in prediction and thematic analyses over previously known models.[14,15] These NLP techniques have the potential to generate deeper insights beyond ratings by identifying nuances in language and by detecting emotional intensity. We developed and fine-tuned a LLM analysis of online cardiologist reviews to provide a more nuanced understanding of patient experience beyond the quantitative 1 to 5 star reviews and to examine the effects of age and sex on the patient experience with cardiologists.

METHODS

DATA COLLECTION AND DATA SET. This is a retrospective data analysis using online reviews from the patient review website, Healthgrades, from 1998 to 2023. We examined physician reviews under the Healthgrades practicing specialty "cardiology." Reviews are self-submitted feedback from internet community members on Healthgrades.com. Healthgrades reviews must "pass all verification steps and rules that govern content."[16] Each unique review consisted of: a written comment, provider rating (from 1-5, with 5 being best), and a combination of subratings. Reviews without written feedback were discarded as we wanted to preserve consistency with written reviews for natural language analysis. The subratings (scored between 1 and 5, with 5 being best) examined were office scheduling, office environment, provider staff, provider time allocation, provider answers, provider explanation, and provider trust. Wait time was an additional subrating scored between 1 and 5, with 1 indicating short wait time, and 5 indicating long wait time. Data were gathered on a state-by-state basis, covering all 50 states, Washington, DC, and Puerto Rico. States were grouped into geographic regions; a breakdown of region compositions can be found in Supplemental Table 1. The data consisted of cardiologists with 3 or more reviews. Physician characteristics, including sex, were obtained from Healthgrades. Physician ages were sorted into those 55 years and older and under age 55 years. Healthgrades did not collect information on race/ethnicity. Institutional Review Board approval was not required due to the online, publicly accessible nature of the data.

STATISTICAL ANALYSIS. For demographics, reviews were grouped by unique providers then ratings were averaged to generate a physician-based data set. A unique provider was determined by each unique combination of name, age, state, degree, and Healthgrades specialty. By performing mean and standard deviation statistical analysis at a provider level, we aimed to avoid confounders at the review level. T-tests were utilized to perform significance testing for demographics using the physician-based data set. To better understand the factors that drive patients to rate their physicians highly, we looked at the subset of subratings when the overall cardiologist rating was high at 5 stars. We then looked at the distribution of subratings when the overall cardiologist rating was low at 1 star.

SENTIMENT ANALYSIS. The NLP tools Valence Aware Dictionary and sEntiment Reasoner (VADER) and *bert-base-multilingual-uncased-sentiment* (BERT) were used for sentiment analysis by quantifying the positivity and negativity in word choice and language used, in addition to each user's rating.[17,18] The lexicographically based sentiment VADER analysis is reported on a scale of −1 to +1. A sentiment score

TABLE 1 Physician Demographics by Age (N = 9,461)	
Age, y	57.68 ± 10.58
Age, y	
≥ 55	5,666 (60%)
<55	3,795 (40%)
Sex	
Male	8,424 (89%)
Female	1,037 (11%)
Region	
Northeast	1,943 (20.5%)
Midwest	1,541 (16.3%)
South	4,156 (43.9%)
Pacific	1,810 (19.1%)
Puerto Rico	11 (0.1%)
No. of surveys per physician	10.7

Values are mean ± SD or n (%). For physician characteristics, values were reported as n and % of all physicians.

of −1 suggests reviews with negative language, 0 suggests reviews with neutral language, and +1 suggests reviews with positive language. The BERT-based sentiment analysis is reported on a scale of 1 to 5 (1 being negative language and 5 being positive language). A t-test was utilized to perform significance testing for sentiment analysis.

DATA PREPARATION FOR LOGISTIC REGRESSIONS.
To prepare reviews for logistic regressions, reviews with missing subratings (incomplete reviews) were eliminated. Eliminating incomplete reviews was preferred to imputation, as the distribution of subratings was likely non-normal and non-random. Demographics were one-hot encoded with respect to those under 55 years of age and male. Reviews were classified as high (4 or 5 stars) or low (1 or 2 stars). Reviews with 3 stars (n = 346) were excluded from the logistic regressions due to the bimodal distribution of ratings.

LOGISTIC REGRESSION. A binary logistic regression was fit with office environment rating, office scheduling rating, provider staff rating, provider time allocation rating, provider answers rating, provider explanation rating, provider trust rating, wait time, sex, and age to predict the odds of receiving a high rating (4 or 5 stars). The goal of the initial logistic regression was to determine subrating effects and adjust for the mixed effects logistic regression (MELR). Due to complexity from subratings, only a binary logistic regression model achieved convergence. In the initial binary logistic regression, a significant amount of collinearity was detected (Supplemental Table 2).

The high multicollinearity was corrected by compounding the subratings into an average subrating (Supplemental Table 3). To capture same-physician random effects, a MELR was used for average subrating, wait time, sex, and age to predict the odds of receiving a high rating (4 or 5 stars).
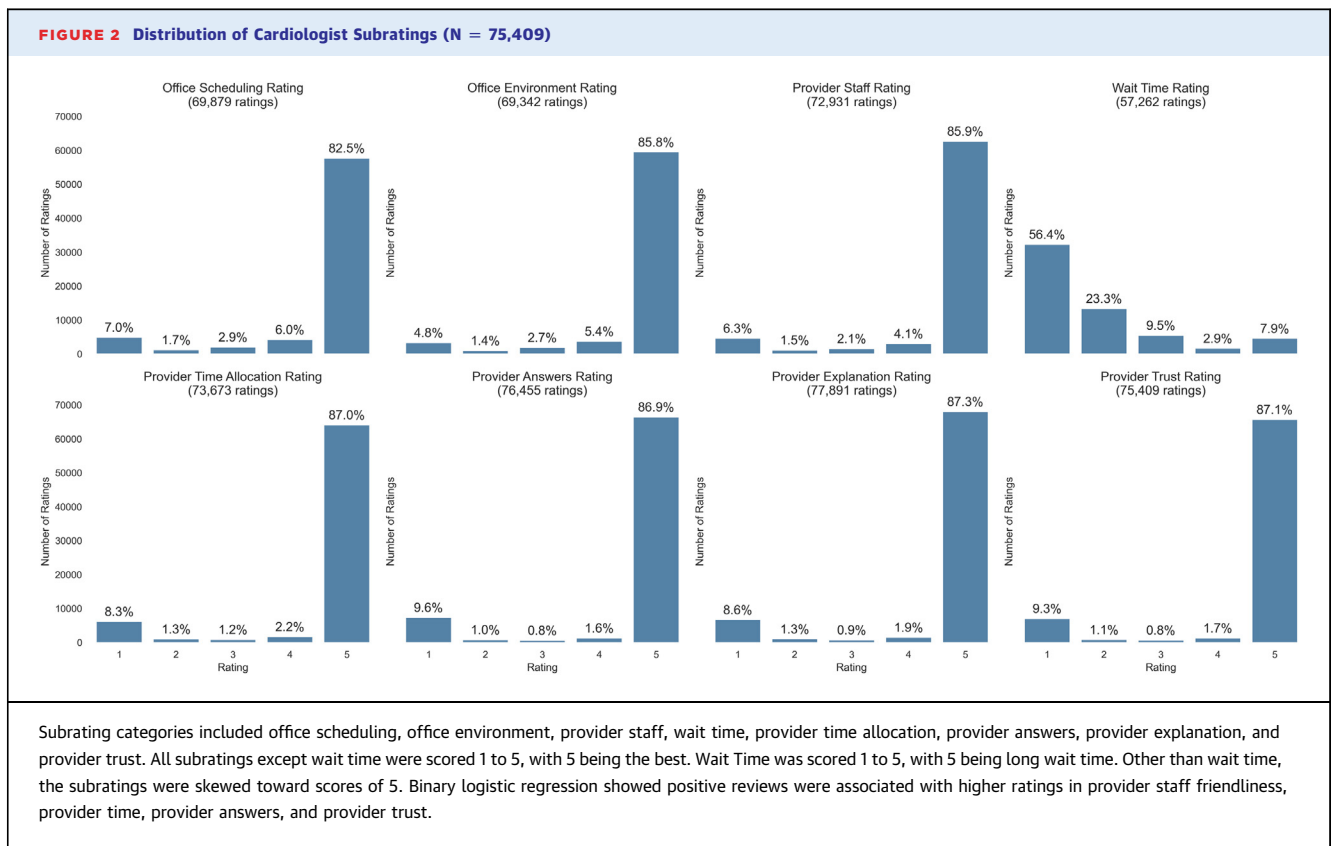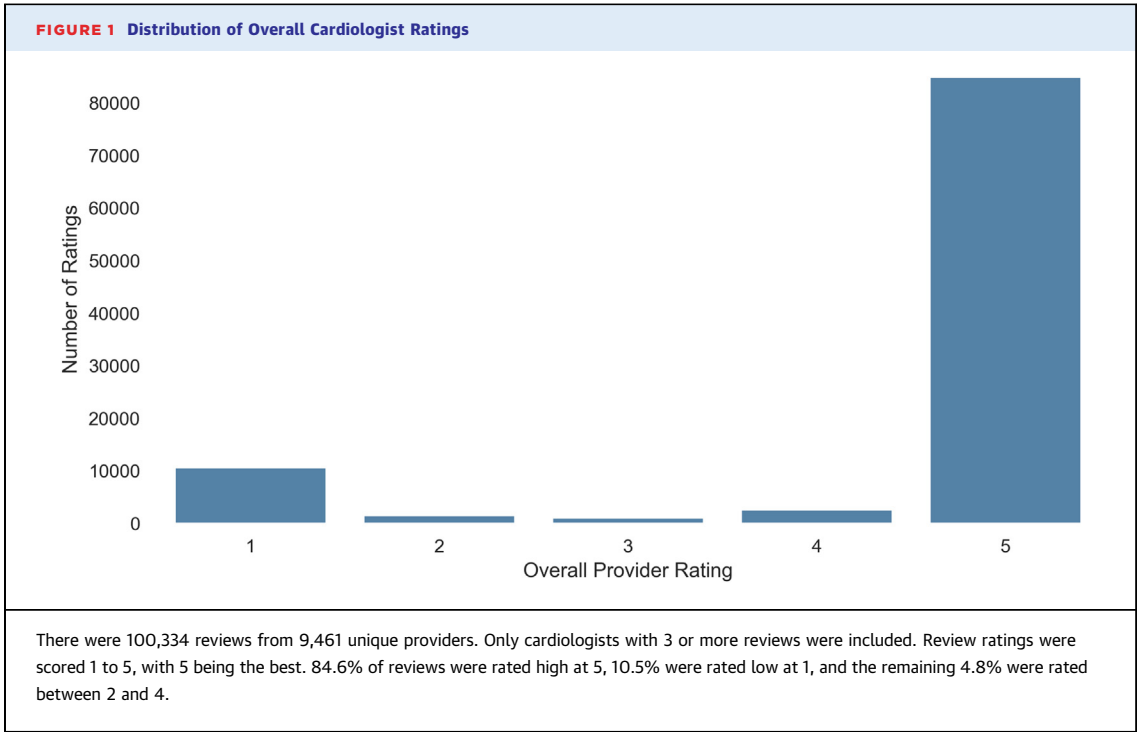
BIGRAM ANALYSIS. We then assessed the frequency of 2-word phrases in written reviews and determined if there were any associations among specific 2-word phrases and positive vs negative reviews, age, and gender.

RESULTS

OVERALL CARDIOLOGIST RATINGS AND SUBRATINGS.
There were a total of 100,334 reviews from 9,461 unique cardiologists. Cardiologist characteristics are summarized in **Table 1**. The mean age was 57.68 ± 10.58 years. Sixty percent were 55 years or older and 40% were under the age of 55 years. Eighty-nine percent of cardiologists were male and 11% were female. By geographic regions, 43.9% of the cardiologists practiced in the South, 20.5% in the Northeast, 19.1% in the Pacific, 16.3% in the Midwest, and only 0.1% in Puerto Rico. The mean number of reviews per cardiologist was 10.7 reviews. The cardiologist with the most reviews had 536 reviews. 2,237 cardiologists had exactly 3 reviews. The distribution of overall cardiologist ratings is shown in **Figure 1**. Each review included a cardiologist rating from 1 to 5, with 5 being best. The cardiologist ratings were skewed toward positive reviews. 84.6% of reviews were rated high at 5, 10.5% were rated low at 1, and the remaining 4.8% were rated between 2 to 4.

The distribution of provider subratings (in office scheduling, office environment, provider staff, wait time, provider time allocation, provider answers, provider explanation, and provider trust) is shown in **Figure 2**. Other than wait time, the subratings were all skewed toward the highest rating of 5. Wait time received a rating of 5, corresponding to a long wait time, only 7.9% of the time. The initial binary logistic regression showed positive reviews were associated with higher ratings in provider staff friendliness (OR: 1.591; 95% CI: 1.467-1.727; $P < 0.01$), provider time (OR: 1.383; 95% CI: 1.251-1.528; $P < 0.01$), provider answers (OR: 2.622; 95% CI: 2.324-2.959; $P < 0.01$), and provider trust (OR: 2.285; 95% CI: 2.053-2.543; $P < 0.01$) (**Central Illustration**, **Table 2**). A long wait time was 70.7% less likely to be associated with a positive review. To better understand what patient experience factors influenced positive and negative

FIGURE 1 Distribution of Overall Cardiologist Ratings

There were 100,334 reviews from 9,461 unique providers. Only cardiologists with 3 or more reviews were included. Review ratings were scored 1 to 5, with 5 being the best. 84.6% of reviews were rated high at 5, 10.5% were rated low at 1, and the remaining 4.8% were rated between 2 and 4.

FIGURE 2 Distribution of Cardiologist Subratings (N = 75,409)

Subrating categories included office scheduling, office environment, provider staff, wait time, provider time allocation, provider answers, provider explanation, and provider trust. All subratings except wait time were scored 1 to 5, with 5 being the best. Wait Time was scored 1 to 5, with 5 being long wait time. Other than wait time, the subratings were skewed toward scores of 5. Binary logistic regression showed positive reviews were associated with higher ratings in provider staff friendliness, provider time, provider answers, and provider trust.

**CENTRAL ILLUSTRATION** Patient Reviews of Cardiologists

Analysis of online cardiologist reviews from Healthgrades between 1998 to 2023.
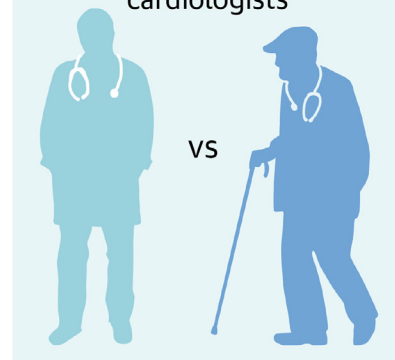Reviews analyzed by gender, age, and frequency of 2-word phrases

4- or 5-star reviews associated with *trust* in provider, *time* spent with patients, and providing *answers*

Women cardiologists were *34.5% less* likely to receive a positive review
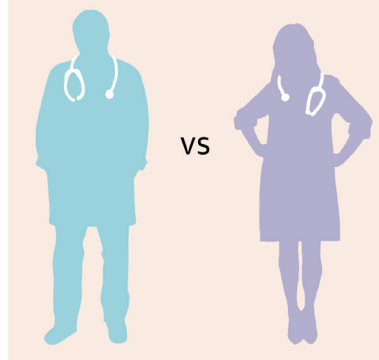
VS

Older cardiologists received *lower* average ratings compared to younger cardiologists

VS

Yang A, et al. JACC Adv. 2024;3(7):101046.

Our investigation of 100,334 online reviews of cardiologists found that positive reviews were associated with cardiologists being comprehensive and patients feeling a sense of trust in the relationship. There was a difference in ratings based on age and sex with female and older cardiologists receiving lower ratings. When controlling for age and subratings, female cardiologists were 34.5% less likely to receive a high rating.

reviews, we examined the distribution of subratings when the overall cardiologist rating was high at 5 (**Figure 3**). Not surprisingly, when the overall cardiologist rating was 5, there was an even more pronounced trend toward scores of 5's in the

subratings especially in the categories of provider time allocation, provider answers, and explanation and provider trust. Patients gave 5-star ratings for provider time allocation 98.0%, provider answers 98.8%, provider explanation 98.5%, and provider trust 98.9% of the time. We then examined the distribution of subratings when the overall cardiologist rating was low at 1 (**Figure 4**). When a cardiologist received an overall score of 1, we see a trend toward lower scores in provider time allocation, provider answers, provider explanation, and provider trust. Patients gave 1-star ratings for provider time allocation 77.7% of the time, provider answers 84.8% of the time, provider explanation 80.6% of the time, and provider trust 83.7% of the time.

The most frequent 2-word phrases found in positive and negative reviews are shown in **Table 3**. Two-word phrases associated with positive reviews included "takes time," "saved life," "highly recommend," "recommend doctor," and "answers question," Two-word phrases associated with negative reviews included "office staff," "stress test,"

TABLE 2 Likelihood of a Positive Review by Subratings

|	OR (95% CI)	P Value
Office environment rating	0.381695 (0.348-0.419)	<0.01
Office scheduling rating	1.039310 (0.963-1.121)	0.315
Provider staff rating	1.591331 (1.467-1.727)	<0.01
Wait time rating	0.292752 (0.278-0.309)	<0.01
Provider time allocation rating	1.383037 (1.251-1.528)	<0.01
Provider answers rating	2.622440 (2.324-2.959)	<0.01
Provider explanation rating	1.149205 (1.009-1.308)	0.07
Provider trust rating	2.284880 (2.053-2.543)	<0.01

Binary logistic regression model determined which subratings had the largest odds of yielding a positive review (defined as 4 or 5 stars) from a negative review (defined as 1 or 2 stars). Coefficients (β) are reported as the OR ($e^β$) and CI. *P* values were derived from the Holm-Bonferroni method to determine the statistical significance of the coefficients.

FIGURE 3   Distribution of Subratings When the Overall Cardiologist Rating Was 5 (N = 64,790)
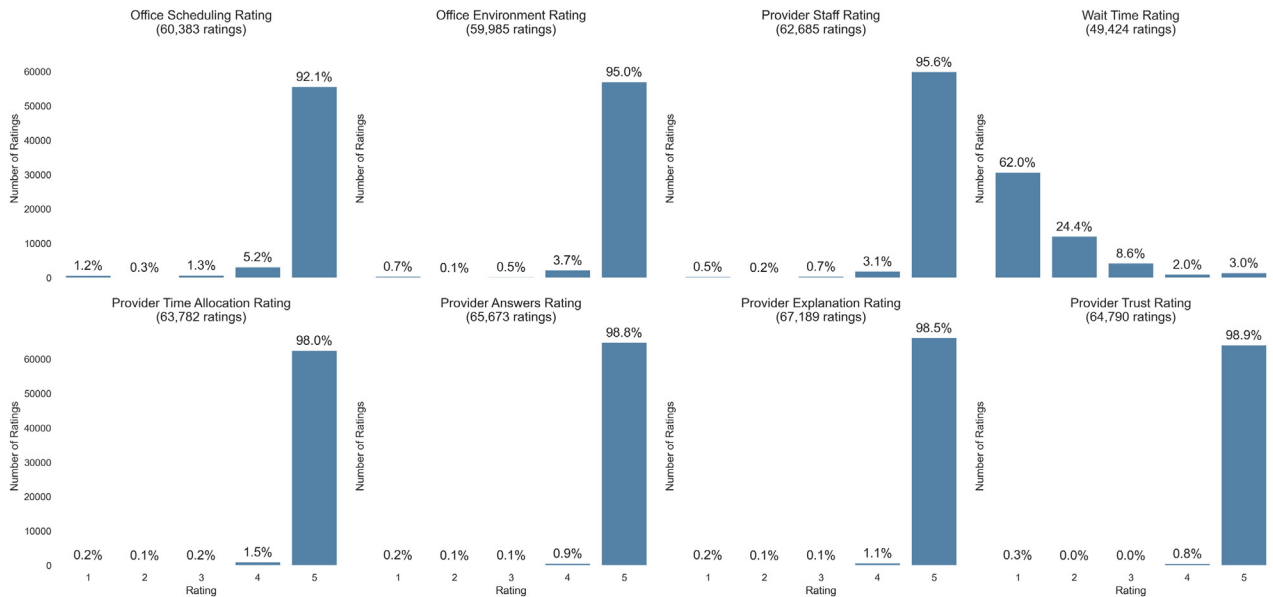
Review ratings were scored 1 to 5, with 5 being the best. When a cardiologist received an overall score of 5, other than wait time, all subratings were skewed toward scores of 5.

FIGURE 4   Distribution of Subratings When the Overall Cardiologist Rating Was 1
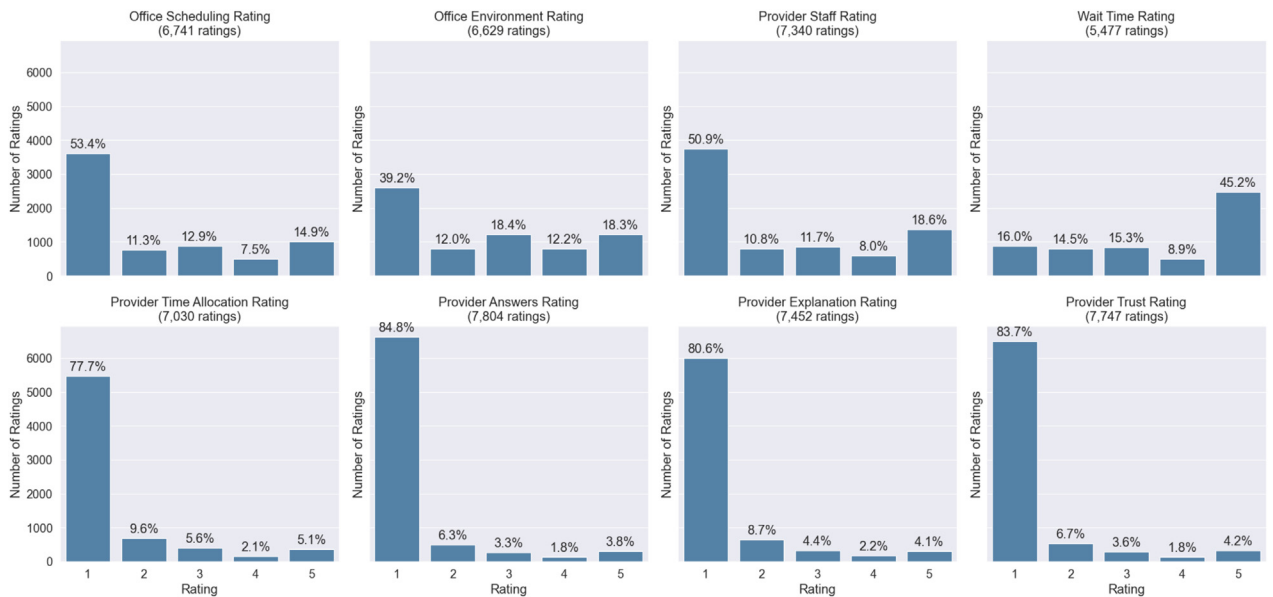
Review ratings were scored 1 to 5, with 5 being the best. When a cardiologist received an overall score of 1, we see a trend toward lower scores in provider time allocation (77.7%), provider answers (84.8%), provider explanation (80.6%), and provider trust (83.7%).

TABLE 3 Top 2-Word Phrases Associated With Negative and Positive Reviews

Most frequent 2-word phrases found in positive reviews (4, 5 stars)	
("highly," "recommend")	7,418
("takes," "time")	6,056
("answers," "questions")	5,866
("answered," "questions")	
("answer," "questions")	
("recommend," "dr")	4,020
("would," "recommend")	3,209
("saved," "life")	3,008
Most frequent 2-word phrases found in negative reviews (1, 2 stars)	
("office," "staff")	687
("stress," "test")	586
("bedside," "manner")	552
("would," "recommend")	532
("blood," "pressure")	446

Overall provider reviews with 4 or 5 stars were considered positive reviews and those with 1 or 2 were considered negative reviews.

"bedside manner," "would recommend," and "blood pressure."

EFFECT OF AGE AND SEX ON PROVIDER RATINGS. The mean overall provider ratings by age and sex are summarized in Table 4. Cardiologists 55 years and older had statistically significant lower overall provider ratings compared to cardiologists under age 55 years (4.145 ± 0.908 vs 4.348 ± 0.795; $P < 0.01$). Female cardiologists received statistically significant lower provider ratings compared to male cardiologists (4.161 ± 0.911 vs 4.235 ± 0.865; $P < 0.01$). Our MELR, controlling for all subratings and age, showed that female cardiologists were 34.5% less likely to receive a positive review compared to male cardiologists (Central Illustration, Table 5). Age did not appear to be a significant predictor for a positive review under the MELR (Table 5).

TABLE 4 Cardiologist Overall Ratings by Age and Sex

	Mean ± SD	P Value
Age		
<55 y	4.348 ± 0.795	<0.01
≥55 y	4.145 ± 0.908	
Sex		
Female	4.161 ± 0.911	<0.01
Male	4.235 ± 0.865	

Review ratings were scored 1 to 5, with 5 being the best. A *t*-test was utilized to perform significance testing. *P* values are associated with testing across demographic groups.

Sentiment analysis of the word choice and language in the reviews was performed (Table 6). Sentiment did not appear to be affected by sex (BERT scores 3.924 ± 0.668 for female vs 3.960 ± 0.642 for male, $P = 0.085$; VADER scores 0.546 ± 0.301 for female vs 0.552 ± 0.279 for male, $P = 0.526$). Older cardiologists 55 years and older had slightly lower sentiment scores and this finding was statistically significant (BERT scores 4.041 ± 0.596 for age <55 years vs 3.899 ± 0.670 for age ≥55 years, $P < 0.01$; VADER scores 0.588 ± 0.262 for age <55 years vs 0.526 ± 0.292 for age ≥55 years, $P < 0.01$).

The most frequent 2-word phrases found in positive and negative reviews by sex are shown in Tables 7 and 8. Positive reviews of male and female cardiologists shared many of the same 2-word phrases such as "highly recommend," "takes time," "recommend doctor," "saved life," and "bedside manner." There were 2 exceptions. "Cares patients" was more frequently used to describe female cardiologists and "great doctor" was more frequently used to describe male cardiologists. There was no significant difference in the 2-word phrases that were frequently found in negative reviews for female and male cardiologists.

DISCUSSION

In this comprehensive study of 100,334 patient reviews of 9,461 cardiologists across the United States, we demonstrated that cardiologists received higher ratings when they spent more time with patients, answered questions thoroughly, and instilled trust in the patient-physician relationship. This is similar to previously reported studies on patient reviews in cardiology and other subspecialties.[6,19,20] Our results reinforce that physician bedside manners, amount of time spent, and being comprehensive remain very important to patients and significantly influence patient satisfaction with their cardiologists.

We also found a statistically significant effect of sex on cardiologist reviews. Female cardiologists were 34.5% less likely to receive positive reviews, when controlling for age and subratings. A recent study of 563 Yelp reviews also revealed a negative bias of patients toward female cardiologists.[6] The sentiment in word choice and language in the reviews; however, appeared to be not statistically significant. The only difference we detected was in a 2-word phrase used to describe female cardiologists. The 2-word phrase "cares patients" was unique to female cardiologists and came up 204 times in positive reviews of female cardiologists. In primary care, female physicians have been found to have better

TABLE 5 Likelihood of a Positive Review by Age and Sex

	OR (95% CI)	P Value
Sex (female vs male)	0.655 (0.481-0.893)	0.015
Age (≥55 vs <55 y)	0.877 (0.713-1.078)	0.212
Average subrating	45.53 (38.01-54.54)	<0.01
Wait time	0.864 (0.802-0.930)	<0.01

A mixed effects logistic regression model (corrected for multicollinearity and controlling for all other characteristics/subratings) of age and sex and the odds of yielding a positive review (defined as 4 or 5 stars). In the model, a lower wait time value indicates a shorter wait time. Coefficients (β) are reported as the OR (e^β) and CI. P values were derived from the Holm-Bonferroni method to determine the statistical significance of the coefficients.

TABLE 7 Top 2-Word Phrases Associated With Positive Reviews by Sex

Female (n = 4,260)		Male (n = 41,619)	
2-Word Phrases	Count	2-Word Phrases	Count
(highly, recommend)	728	(highly, recommend)	6,690
(takes, time)	631	(takes, time)	5,425
(recommend, dr)	318	(recommend, dr)	3,702
(would, recommend)	293	(would, recommend)	2,916
(took, time)	286	(saved, life)	2,798
(bedside, manner)	251	(bedside, manner)	2,488
(would, highly)	248	(took, time)	2,392
(recommend, anyone)	228	(would, highly)	2,252
(time, explain)	221	(office, staff)	2,055
(saved, life)	210	(great, doctor)	2,024
(cares, patients)	204	(time, explain)	1,982

Overall provider reviews with 4 or 5 stars were considered positive reviews and those with 1 or 2 were considered negative reviews.

bedside manners than male colleagues spending 15% more time with patients and more time counseling patients.[4,21] Despite female cardiologists described as being more caring in our study, they continued to be rated lower than male cardiologists when controlling for all other subratings. Analyzing the frequency of 2-word phrases associated with negative reviews did not provide additional insight into why female cardiologists received lower ratings. We confirm that there is a negative bias against female cardiologists; however, it is not clear what drives this bias. The reason for the lack of differences in review sentiment between female and male cardiologists, despite a difference in rating, remains unclear. We suspect that the frequency of 2-word phrases in reviews may not have sufficiently captured the theme of the review and further research could use LLMs to analyze major themes across multiple written reviews. Moreover, future longitudinal studies examining changes over time would be valuable in understanding trends of ratings and sentiments.

Overall cardiologist rating was higher among younger cardiologists and sentiment analysis detected more positive word choice and language in

TABLE 6 Sentiment Analysis of Reviews by Age and Sex

	BERT		VADER	
	Mean ± SD	P Value	Mean ± SD	P Value
Age				
<55 y	4.041 ± 0.596	<0.01	0.588 ± 0.262	<0.01
≥55 y	3.899 ± 0.670		0.526 ± 0.292	
Sex				
Female	3.924 ± 0.668	0.085	0.546 ± 0.301	0.526
Male	3.960 ± 0.642		0.552 ± 0.279	

A t-test was utilized to perform significance testing. P values are associated with testing sentiment across demographic groups.

BERT = bert-base-multilingual-uncased-sentiment; VADER = Valence Aware Dictionary and sEntiment Reasoner.

reviews of younger cardiologists. Our results for average ratings are similar to prior studies observed in other subspecialties demonstrating higher ratings in younger physicians compared to older physicians.[20-22] However, our MELR showed that age is not predictive of receiving a high rating when controlling for sex and all subratings. This would suggest that age or 1 or more subrating is associated with older cardiologists receiving lower average ratings. Perhaps older cardiologists are being rated lower on 1 of the subratings, or patients of older cardiologists have higher wait times, or there is a sex difference associated with age. Our MELR may not have the resolution to examine age as a predictor, as age was classified as younger/older than 55 years. Nonetheless, our study finds a statistically significant difference in average ratings between younger and older cardiologists. The retrospective nature of the study limited our ability to determine a cause for the differences found in perceived patient experiences by age and sex. Another limitation is whether the small differences in ratings, although statistically significant, translates to being practically different (ie, BERT scores 4.041 vs 3.899 in younger vs older cardiologists). More studies are needed to determine if these differences in ratings by age are reproducible on other physician rating website platforms and to better understand why younger cardiologists are rated higher than older cardiologists.

To our knowledge, this study is one of the most comprehensive analyses of cardiologist reviews incorporating written reviews, subratings, and sentiment analyses of positivity and negativity in word choice and language. The findings from our study

TABLE 8 Top 2-Word Phrases Associated With Negative Reviews by Sex

Female (n = 598)		Male (n = 4,149)	
2-Word Phrases	Count	2-Word Phrases	Count
(office, staff)	84	(office, staff)	603
(bedside, manner)	76	(stress, test)	529
(would, recommend)	71	(bedside, manner)	476
(stress, test)	57	(would, recommend)	461
(test, results)	50	(blood, pressure)	404

Overall provider reviews with 4 or 5 stars were considered positive reviews and those with 1 or 2 were considered negative reviews.

could be used by cardiology training programs to address areas of improvement highlighted by patient reviews as well as develop policies to support the needs of female and older cardiologists.

STUDY LIMITATIONS. This study has several limitations including the inability to assess reviewer bias (such as sex, age and race/ethnicity) and the assumption that Healthgrades reviews are accurate and provided by real patients. The use of a single source of patient reviews from Healthgrades is a limitation; other physician review websites may prompt users differently or market to other demographics. Future studies examining reviews from multiple patient review websites would minimize biases based on the platform used. Patients who are willing to post on a public review may be self-selecting. For instance, those motivated to leave a review may do so only if they had an overwhelmingly positive or negative experience. This may have been reflected in our reviews as the ratings in our study were significantly skewed toward 5 stars.

Provider demographics, including sex, are provided by Healthgrades. It was not possible to verify the accuracy of these metrics; however, Healthgrades states that these data directly reflects the National Professional Identifier Registry from the Centers for Medicare and Medicaid Services. The physician demographic information also does not indicate nonmale and nonfemale sexes. There are limitations to the demographic analysis for nonbinary sexes.

Our binary logistic regressions and NLP techniques also possess limitations. In our logistic regressions, effects of neutral reviews are excluded from our study. Due to convergence, subrating effects (time spent with the patient, trust, and explanation) were based on the initial binary logistic regression and may be dependent on physician-level effects. However,

average subrating remained a strong determinant of a positive review in the mixed effects model (Table 5). VADER primarily utilizes syntaxial rules, and as such is unable to understand linguistic context. For instance, a sarcastic review may be interpreted incorrectly. The BERT model and linguistic analysis may have inherent biases, giving a false depiction of neutral sentiment across sex and age. Additionally, not every review may adhere strictly to correct English grammar.

CONCLUSIONS

With the growth of online reviews, it is important to understand the effects of age and sex on patient experience. Positive reviews were associated with providers being comprehensive and patients feeling like they can trust their cardiologists. Female and older cardiologists were rated lower than male and younger cardiologists; however, the reasons for these negative ratings are not clear. There is a need for further research to determine the factors underlying negative reviews in female and older cardiologists.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

ADDRESS FOR CORRESPONDENCE: Dr Jennifer P. Woo, Stanford University School of Medicine, 300 Pasteur Drive, 3rd Floor A32, Palo Alto, California 94304, USA. E-mail: jpwoo@stanford.edu.

PERSPECTIVES

COMPETENCY IN PATIENT CARE/INTERPERSONAL COMMUNICATION SKILLS: Positive patient experiences were associated with cardiologists being comprehensive—spending time explaining and answering questions, and feeling like they can trust their cardiologists. Such insights may be useful in developing future training programs.

TRANSLATIONAL OUTLOOK: There was an effect of age and sex on the patient-physician relationship with female and older cardiologists consistently rated lower than male and younger cardiologists. However, age did not affect likelihood of receiving a higher review. A better understanding of the negative biases that exist in the patient-physician experience is crucial in our collective goal of achieving health equity.

REFERENCES

1. Murphy GP, Radadia KD, Breyer BN. Online physician reviews: is there a place for them? *Risk Manag Healthc Policy*. 2019;12:85–89.

2. Hanauer D, Zheng K, Singer D. Public awareness, perception, and use of online physician rating sites. *JAMA*. 2014;311:734–735.

3. Rogo-Gupta LJ, Altamirano J, Homewood LN, et al. Women physicians receive lower Press Ganey patient satisfaction scores in a multicenter study of outpatient gynecology care. *Am J Obstet Gynecol*. 2023;229:304.e1–304.e9.

4. Flocke SA, Gilchrist V. Physician and patient gender concordance and the delivery of comprehensive clinical preventive services. *Med Care*. 2005;43:486–492.

5. Takeshita J, Wang S, Loren AW, et al. Association of racial/ethnic and gender concordance between patients and physicians with patient experience ratings. *JAMA Netw Open*. 2020;3:e2024583.

6. Mark E, Oswald M, Kundar P, Gulati M. Patient-centered insights and biases Regarding cardiologists via online review platform analysis. *J Am Heart Assoc*. 2023;12:e027405.

7. Ribeiras R. Women in cardiology: between the "glass ceiling" and the "sticky floor". *Rev Port Cardiol (Engl Ed)*. 2021;40:505–508.

8. Mehta LS, Fisher K, Rzeszut AK, et al. Current demographic status of cardiologists in the United States. *JAMA Cardiol*. 2019;4:1029–1033.

9. Paul MJ, Wallace BC, Dredze M. *What Affects Patient (Dis)Satisfaction? Analyzing Online Doctor Ratings With a Joint Topic-Sentiment Model*. AAAI Press Technical Reports; 2013.

10. Vasan V, Cheng CP, Lerner DK, Vujovic D, van Gerwen M, Iloreta AM. A natural language processing approach to uncover patterns among online ratings of otolaryngologists. *J Laryngol Otol*. 2023;137:1–5.

11. Jo JJ, Cheng CP, Ying S, Chelnis JG. Physician review websites: understanding patient satisfaction with Ophthalmologists using Natural Language processing. *J Ophthalmol*. 2023;2023:4762460.

12. Davagdorj K, Wang L, Li M, Pham VH, Ryu KH, Theera-Umpon N. Discovering Thematically coherent biomedical documents using contextualized bidirectional encoder representations from transformers-based clustering. *Int J Environ Res Public Health*. 2022;19:5893.

13. Gupta A, Gupta R, White MD, et al. Patient satisfaction reviews for 967 spine neurosurgeons on Healthgrades. *J Neurosurg Spine*. 2022;36:869–875.

14. Sohail SS. A promising start and not a panacea: ChatGPT's Early Impact and potential in medical science and biomedical engineering research. *Ann Biomed Eng*. 2023;52:1131–1135.

15. Dunivin Z, Zadunayski L, Baskota U, Siek K, Mankoff J. Gender, soft skills, and patient experience in online physician reviews: a large-scale Text analysis. *J Med Internet Res*. 2020;22:e14455.

16. Healthgrades community review Guidelines. Healthgrades. 2020. Accessed January 29, 2024. https://www.healthgrades.com/content/community-review-guidelines

17. Loureiro D, Barbieri F, Neves L, Espinosa Anke L, Camacho-collados J. TimeLMs: Diachronic Language Models from Twitter. In: Basile V, Kozareva Z, Stajner S, eds. *Proc Annu Meet Assoc Comput linguist Syst Demonstrations*. Assoc Comput Linguist; 2022:251–260.

18. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the International AAAI Conference on Web and Social Media*. 8. 2014:216–225. PKP Publishing Services Network.

19. Queen D, Trager MH, Fan W, Samie FH. Patient satisfaction of general Dermatology providers: a quantitative and qualitative analysis of 38,008 online reviews. *JID Innov*. 2021;1:100049.

20. Gao GG, McCullough JS, Agarwal R, Jha AK. A changing Landscape of physician quality reporting: analysis of patients' online ratings of their physicians over A 5-Year Period. *J Med Internet Res*. 2012;14:e38.

21. Ganguli S, Sheridan B, Gray J, Chernew M, Rosenthal MB, Neprash H. Physician work Hours and the gender Pay Gap - Evidence from primary care. *N Engl J Med*. 2020;383:1349–1357.

22. Jack RA, Burn MB, M PC, Liberman SR, Varner KE, Harris JD. Does experience matter? A meta-analysis of physician rating websites of Orthopaedic Surgeons. *Musculoskelet Surg*. 2018;102:63–71.

KEY WORDS age, gender bias, health equity, natural language processing, online reviews

---

APPENDIX For supplemental tables, please see the online version of this paper.