

Original Paper

# Large Language Model–Based Analysis of Statin Therapy Discussions and Sentiment on Social Media: Cross-Sectional Observational Study

Siru Liu<sup>1,2</sup>, PhD; Jialin Liu<sup>3,4</sup>, MD

<sup>1</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

<sup>2</sup>Department of Computer Science, Vanderbilt University, Nashville, TN, United States

<sup>3</sup>Department of Medical Informatics, West China Medical School, Chengdu, China

<sup>4</sup>Department of Otolaryngology-Head and Neck Surgery, West China Hospital, Sichuan University, Chengdu, China

**Corresponding Author:**

Siru Liu, PhD

Department of Biomedical Informatics

Vanderbilt University Medical Center

2525 West End Ave #1475

Nashville, TN 610041

United States

Phone: 1 615-875-5216

Email: [siru.liu@vumc.org](mailto:siru.liu@vumc.org)

## Abstract

**Background:** Statin therapy, despite proven cardiovascular benefits, remains underused. Social media platforms may capture patient perspectives that are less visible in clinical encounters.

**Objective:** This study aimed to characterize themes, sentiment, and decision-making factors related to statin therapy through large language model (LLM)–based analysis of Reddit discussions.

**Methods:** This cross-sectional observational study analyzed English-language Reddit posts and comments mentioning statins from January 2022 to May 2025, identified via keyword-based Reddit application programming interface searches ( $\leq 1000$  posts per keyword). A total of 5328 retrieved discussions ( $n=1661$ , 31.2% posts and  $n=3667$ , 68.8% keyword-containing comments) from public subreddits were included. Themes, sentiments (positive, neutral, or negative), guideline-informed clinical relevance, information-seeking behavior, adverse effect mentions, decision factors, and adherence-related content were extracted using an LLM-based pipeline.

**Results:** Among 5328 discussions, prominent topics included adverse effects ( $n=1697$ , 31.9%), decision-making references related to laboratory results and physician advice ( $n=2767$ , 51.9% and  $n=2034$ , 38.2%, respectively), and alternative approaches ( $n=2485$ , 46.6%). Overall sentiment was neutral in 34% ( $n=1812$ ) of discussions, negative in 30.9% ( $n=1646$ ), and positive in 16.9% ( $n=900$ ); the remainder were mixed or unclear. Statin-directed sentiment was neutral in 44.1% ( $n=2350$ ) of discussions, negative in 25.2% ( $n=1343$ ), and positive in 12.5% ( $n=666$ ); the remainder did not express statin-directed sentiment. High clinical relevance was identified in 12.6% ( $n=672$ ) of discussions. Adherence-related issues were mentioned in 29.8% ( $n=1587$ ) of discussions. Among adverse effect mentions, muscle pain ( $n=129$ , 7.6%) and fatigue ( $n=110$ , 6.5%) were common.

**Conclusions:** LLM-enabled analysis of Reddit discourse highlights substantial negative sentiment, adherence-related concerns, and adverse effect narratives surrounding statin therapy. These findings suggest opportunities for patient-centered communication and shared decision-making strategies that address symptom attribution, uncertainty, and information needs in digital information environments.

*J Med Internet Res* 2026;28:e85057; doi: [10.2196/85057](https://doi.org/10.2196/85057)

**Keywords:** statin therapy; social media; large language model; sentiment analysis; Reddit; cardiovascular disease; patient perspectives; adherence; adverse effects; cohort study

## Introduction

Statins are among the most widely prescribed medications globally, with more than 200 million individuals using these agents to reduce cardiovascular morbidity and mortality [1]. These 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors remain the cornerstone of atherosclerotic cardiovascular disease (ASCVD) prevention, demonstrating well-established efficacy in both primary and secondary prevention settings [2,3]. Despite their proven benefits, real-world statin therapy is frequently limited by concerns regarding safety and tolerability. In routine care, statin-related events are commonly documented, with myalgia and myopathy representing the most frequent category, and these events often precipitate temporary discontinuation and subsequent rechallenge [4]. Consequently, long-term use remains suboptimal; a recent systematic review and meta-analysis found that only approximately 62% of patients achieve good adherence ( $\geq 80\%$  use) over a median follow-up of 24 months [5], while population-based primary care data show frequent discontinuation with substantial restarting, consistent with an intermittent use pattern in practice [6].

Traditional clinical research methods, including randomized controlled trials and observational studies, may not fully capture the breadth of patient experiences and concerns in routine care [7]. Effective communication between patients and health care providers is essential for shared decision-making and treatment adherence; however, barriers such as limited consultation time, health literacy challenges, and patient reluctance to disclose adverse experiences may constrain insight into the real-world challenges of medication use [8,9]. Patients may hesitate to report symptoms or concerns to clinicians due to fear of judgment, desire to avoid conflict with prescribers, or uncertainty about whether their experiences warrant clinical attention [9]. Social media platforms, especially Reddit (Reddit, Inc), provide access to unfiltered patient narratives that may not be expressed during clinical encounters [10-13]. Reddit hosts a wide range of health-related communities and has more than 50 million daily active users [14]. Within these forums, individuals share personal medical experiences, seek peer advice, and discuss treatment decisions. These discussions may provide valuable insights into patient beliefs, motivations, and barriers that influence medication adherence and cardiovascular risk management [15-17].

Recent advances in large language models (LLMs) offer scalable methods to analyze such unstructured, high-volume textual data, enabling systematic characterization of patient perspectives at a granularity that would be infeasible using manual qualitative approaches alone [18]. In the context of cardiovascular prevention, applying LLMs to social media discourse may provide novel insights into how patients interpret statin-related information, evaluate perceived risks and benefits, and make adherence decisions outside the clinical setting. Accordingly, this study aimed to systematically analyze Reddit discussions related to statin use using an LLM to characterize patient-reported experiences, identify recurrent concerns and misconceptions, and explore factors

influencing decision-making and adherence. By integrating patient-generated narratives with computational text analysis, this work seeks to complement traditional evidence sources and inform more patient-centered approaches to cardiovascular risk management.

## Methods

### Study Design and Data Source

This cross-sectional study analyzed publicly available Reddit discussions about statin therapy posted between January 1, 2022, and May 1, 2025. Reddit was selected because it hosts large, topic-specific health communities in which users openly share medication experiences, treatment decisions, and interactions with health care providers in naturalistic settings. Data were accessed via the official Reddit application programming interface (API) using Python Reddit API Wrapper, limited to public subreddits, and no user contact or reidentification attempts were made.

### Data Collection

#### Search Strategy

We retrieved content using “statin,” generic statin names (eg, atorvastatin, rosuvastatin, simvastatin, pravastatin, lovastatin, fluvastatin, and pitavastatin), and US brand names (eg, Lipitor, Crestor, Zocor, Pravachol, Livalo, Lescol, and Mevacor); full queries are provided in [Multimedia Appendix 1](#).

#### API Procedure

Using Python Reddit API Wrapper (version 7.7.1) in Python (version 3.9; Python Software Foundation; May 15-May 20, 2025), we queried Reddit’s search end point (sorted by “new” where available). Due to indexing and API constraints, each keyword returned up to 1000 submissions. For each submission, we extracted metadata (submission ID, created\_utc, subreddit, title, selftext, author, score, comment count, and permalink) and downloaded its comment tree (comment ID, parent ID, created\_utc, author, body, and score). We retained submissions within the study window; pagination, time stamps, and rate-limit handling are described in [Multimedia Appendix 1](#).

#### Eligibility

We included English-language posts and comments from public subreddits that matched 1 or more keywords (case-insensitive) within the study window and excluded removed and deleted placeholders, duplicates, non-English text (langdetect v1.0.9), and promotional and spam content ([Multimedia Appendix 1](#)).

#### Unit of Analysis

The unit of analysis was a Reddit discussion thread (eligible submission plus associated comments), enabling contextual interpretation of patient narratives and peer responses and supporting downstream LLM-based thematic extraction [19].

## Data Cleaning and Preprocessing

We removed URLs and normalized whitespace while preserving the original wording (no stemming or lemmatization). Privacy protection included removing @mentions and replacing detected email addresses and phone numbers with placeholders (eg, “[EMAIL],” “[PHONE]”); subreddit names were retained as public metadata, with reidentification risk minimized via aggregate reporting and avoidance of traceable quotations [20]. English language posts were identified using fastText (lid.176.bin) [21] (retain  $\geq 0.80$ , adjudicate 0.60-0.80 by 2 investigators, and exclude  $< 0.60$ ). For analysis, comments were retained if they contained 1 or more keywords; misspellings and variants were captured via regex-based partial matching with guardrails. Exact duplicates were removed via SHA-256 hashing, and near duplicates were excluded using MinHash with Jaccard similarity  $\geq 0.85$  [22]. All steps were deterministic and version controlled; a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)-style flow diagram is provided in [Multimedia Appendix 1](#).

## LLM-Based Content Extraction

Structured extraction used GPT-4.1 [23] via the OpenAI API (version 1.52.0) between May and June 2025 with standardized parameters (temperature=0.1, top\_p=1.0, and max\_tokens=4096). A single prompt template and predefined JSON schema were applied to all discussions. Two investigators iteratively developed the schema and decision rules through pilot testing on 50 purposively sampled discussions and consensus refinement of edge cases, consistent with clinical reasoning standards [18,24]. For 11 analytic domains ([Table 1](#)), outputs included structured variables, brief summaries, and verbatim evidence quotes supporting key classifications. Dietary changes and exercise were coded as alternative approaches only when explicitly discussed as substitutes for statin therapy (eg, attempting to avoid, delay, or discontinue statins); when described as adjuncts alongside statins, they were not classified as alternatives. Outputs failing JSON validation or missing mandatory evidence were automatically requeryed up to 3 times; only schema-valid outputs were retained ([Multimedia Appendix 2](#)).

**Table 1.** GPT-4.1 content analysis framework with examples.

Domain	Description	Example extractions
Primary themes	1 to 3 main discussion topics	<ul style="list-style-type: none"> <li>Concerns about statin adverse effects, cholesterol management through lifestyle changes, and physician-patient communication</li> </ul>
Statin medications	Specific statin names mentioned	<ul style="list-style-type: none"> <li>Generic: atorvastatin and rosuvastatin</li> <li>Brand: Lipitor and Crestor</li> </ul>
Experience type	Postclassification	<ul style="list-style-type: none"> <li>Personal experience, information request, and medical advice</li> </ul>
Sentiment analysis	Multitarget emotional assessment	<ul style="list-style-type: none"> <li>Neutral, negative, or positive</li> <li>Overall, toward statins and toward physicians</li> </ul>
Clinical relevance	Actionable insight assessment based on AHA <sup>a</sup> and ACC <sup>b</sup> guidelines	<ul style="list-style-type: none"> <li>High: ASCVD<sup>c</sup> history (MI<sup>d</sup>, stroke, and PAD<sup>e</sup>), LDL-C<sup>f</sup> <math>\geq 190</math> mg/dL, diabetes (in individuals aged 40-75 years), and 10-year ASCVD risk <math>\geq 7.5\%</math></li> <li>Medium: family history of premature ASCVD, chronic kidney disease, inflammatory conditions, CAC<sup>g</sup> score <math>\geq 100</math>, and smoking</li> <li>Low: general cardiovascular health discussions</li> </ul>
Adverse effects	Adverse reactions mentioned	<ul style="list-style-type: none"> <li>Muscle pain, fatigue, brain fog, dizziness, nausea, and memory issues</li> </ul>
Decision factors	Treatment choice influences	<ul style="list-style-type: none"> <li>Laboratory results, physician recommendation, and adverse effects</li> </ul>
Information seeking	Behavioral patterns	<ul style="list-style-type: none"> <li>Asking for advice, sharing experience, and seeking alternatives</li> </ul>
Adherence issues	Medication compliance	<ul style="list-style-type: none"> <li>Mentions discontinuation, compliance issues, and dose changes</li> </ul>
Alternative treatments	Nonstatin options	<ul style="list-style-type: none"> <li>Diet changes, exercise, ezetimibe, and coenzyme Q10</li> </ul>
Emotional indicators	Expressed emotions	<ul style="list-style-type: none"> <li>Frustration, confusion, hope, and anxiety</li> </ul>

<sup>a</sup>AHA: American Heart Association.

<sup>b</sup>ACC: American College of Cardiology.

<sup>c</sup>ASCVD: atherosclerotic cardiovascular disease.

<sup>d</sup>MI: myocardial infarction.

<sup>e</sup>PAD: peripheral artery disease.

<sup>f</sup>LDL-C: low-density lipoprotein cholesterol.

<sup>g</sup>CAC: coronary artery calcium.

All content was analyzed as user-generated narratives. References to clinician recommendations within discussions (eg, “my doctor recommended starting a statin”) were coded as patient-reported physician communication rather than independently verified clinical input. Posts containing general medical guidance without a described patient-clinician interaction were not coded under the “decision factors” physician recommendation category. Clinician perspectives

were not independently ascertained; physician-related content was captured only through users’ own accounts of clinical interactions.

## Clinical Relevance Tiering

A deterministic rule-based algorithm aligned with American Heart Association, American College of Cardiology, and US Preventive Services Task Force indicators [25,26]

assigned high, medium, or low clinical relevance ([Multimedia Appendix 2](#)). High relevance required explicit mention of ASCVD, low-density lipoprotein cholesterol  $\geq 190$  mg/dL, diabetes (in individuals aged 40-75 years), or elevated 10-year ASCVD risk, and medium relevance reflected risk-enhancing factors (eg, family history of premature ASCVD, chronic kidney disease, inflammatory conditions, and coronary artery calcium  $\geq 100$  Agatston units) [25,26]; otherwise, discussions were low relevance. Given self-reported data, absence of criteria was coded as “unreported,” and tier assignment defaulted downward unless explicit evidence supported a higher tier.

### Detection of LLM-Related Content

We searched for mentions of LLM tools (eg, ChatGPT [OpenAI], Claude [Anthropic], and Google Gemini) and categorized them as (1) seeking medical advice, (2) seeking general health information, (3) sharing LLM-generated content, or (4) incidental mentions. Categories 1 to 3 were flagged for sensitivity analyses ([Multimedia Appendix 2](#)).

### Validation

LLM outputs were validated via expert review ([Multimedia Appendix 3](#)). A stratified random sample of 50 discussions (0.9% of 5328) spanning key domains was independently assessed by 2 reviewers (a clinical informaticist and a physician) using a 5-point accuracy scale (1=very poor to 5=excellent). Interrater agreement used Cohen  $\kappa$ . Reviewers also documented recurrent errors. Verbatim evidence requirements enabled direct verification of extracted labels.

### Statistical Analysis

Multivariable logistic regression identified factors associated with negative sentiment toward statins. Cluster-robust

SEs were estimated at the subreddit level. Covariates were prespecified (clinical relevance tier, adverse effect mentions, adherence-related content, information-seeking behavior, specific statins, and subreddit category). Absence of a mention was coded as “not reported.” Diagnostics included variance inflation factors and influential observation checks. Results are reported as odds ratios with 95% CIs; 2-sided  $P < .05$  indicated significance.

### Ethical Considerations

This study analyzed publicly accessible Reddit posts and comments as secondary data and involved no direct contact, interaction, intervention, or attempt to reidentify users. Informed consent was not obtained because the study used publicly available online content, the investigators did not interact with users, and obtaining consent from all posters was not feasible. No compensation was provided because no participants were recruited or contacted. The study was exempt from institutional review board approval and aligned with the Association of Internet Researchers' Internet Research: Ethical Guidelines 3.0 [27].

## Results

### Dataset Characteristics and Distribution

After systematic filtering, the final dataset comprised 5328 discussions contributed by 4832 unique users across multiple Reddit communities. The most frequently represented forum was r/Cholesterol (n=2722, 51.1%). Nearly half of the discussions (n=2552, 47.9%) referenced a specific statin. The most commonly mentioned agents were rosuvastatin (n=1276, 23.9%) and atorvastatin (n=1013, 19%). Dataset characteristics and statin mentions are summarized in [Table 2](#).

**Table 2.** Characteristics of Reddit discussions and statin mentions (N=5328).

Characteristic and category	Mentions, n (%)
Document type	
Submissions (posts)	1661 (31.2)
Comments	3667 (68.8)
Source community	
r/Cholesterol	2722 (51.1)
r/PeterAttia	576 (10.8)
r/AskDocs	300 (5.6)
r/keto	92 (1.7)
r/HeartAttack	87 (1.6)
r/diabetes_t2	82 (1.5)
r/stroke	46 (0.9)
Other subreddits	1423 (26.7)
Primary discussion type	
Personal experience sharing	3210 (60.2)
Information request	882 (16.6)
Medical advice provision	734 (13.8)
General discussion	502 (9.4)
Mention of a specific statin	

Characteristic and category	Mentions, n (%)
Yes	2552 (47.9)
No	2776 (52.1)
Most frequent statins <sup>a</sup>	
Rosuvastatin (Crestor)	1276 (23.9)
Atorvastatin (Lipitor)	1013 (19)
Pravastatin (Pravachol)	279 (5.2)
Simvastatin (Zocor)	211 (4)

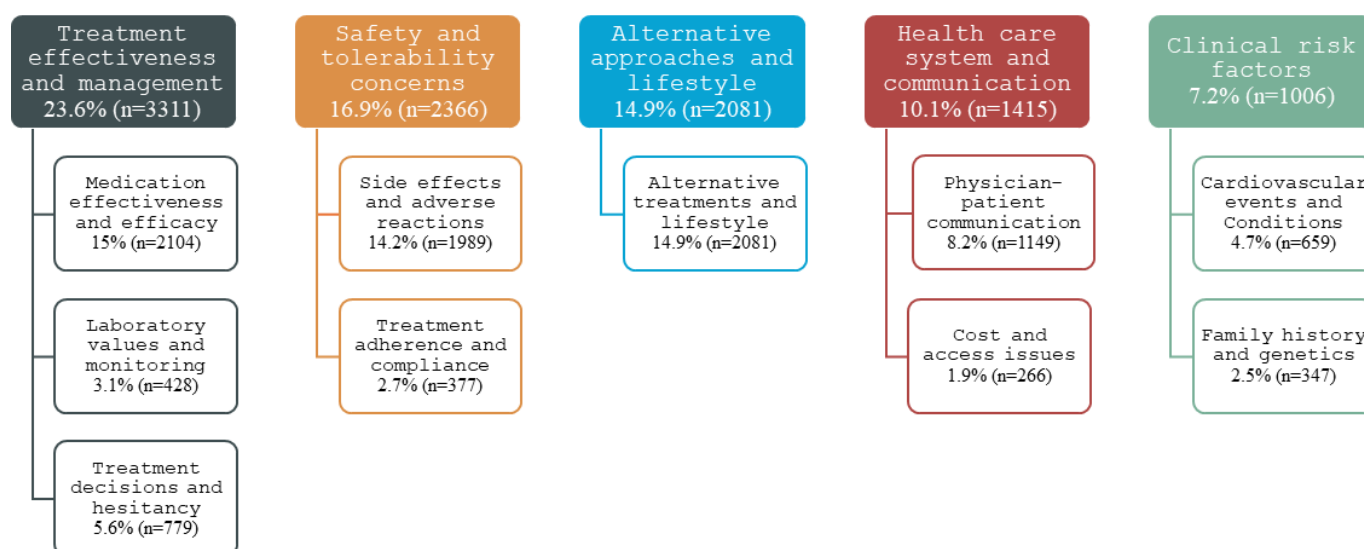
<sup>a</sup>Generic and brand name mentions were combined.

## Thematic Analysis and Primary Discussion Topics

Thematic analysis showed that discussions most commonly focused on treatment effectiveness (3311/14,007, 23.6% of thematic mentions), followed by safety and tolerability

concerns (n=2366, 16.9%) and alternative lifestyle interventions (n=2081, 14.9%; [Figure 1](#)). Privacy-preserving, paraphrased exemplar posts (1-2 per major theme) and the corresponding LLM theme classifications are provided in [Multimedia Appendix 4](#).

**Figure 1.** Quantitative distribution of themes in Reddit discussions related to statins (January 2022 to May 2025). The hierarchical tree map illustrates the primary discussion topics identified by the large language model pipeline across 5328 discussions, organized into 5 overarching thematic groups and 10 topic categories. Each discussion could receive up to 3 thematic codes; percentages represent the proportion of total thematic mentions (N=14,007). The 10 most frequent topic categories are shown; remaining thematic mentions (n=3828, 27.3%) comprised less frequent topics and are not displayed.



## Sentiment Analysis

Overall sentiment was neutral in 34% of discussions, negative in 30.9%, and positive in 16.9%, with the remainder classified as mixed or unclear. Sentiment distributions varied significantly by community type ( $\chi^2=335.7$ ;  $P<.001$ ). The proportion of discussions with negative sentiment was substantially higher in medical advice-seeking forums (eg, r/AskDocs: 77.3%) than in lifestyle-oriented forums (eg, r/Biohackers: 9.1%).

## Clinical Relevance Assessment

Among 5328 discussions, using guideline-informed criteria, 12.6% (n=672) of discussions were classified as high clinical relevance, 22.4% (n=1193) as medium relevance, and 65% (n=3463) as low relevance. High-relevance content was most frequent in r/stroke (40/46, 87.0%), r/HeartAttack (41/87, 47.1%), r/diabetes\_t2 (29/82, 35.4%), and r/AskDocs (82/300,

27.3%). Clinical information elements were reported as follows: laboratory values (n=2048, 38.4%), cardiovascular events (n=362, 6.8%), family history (n=908, 17%), and lifestyle factors (n=2616, 49.1%).

## Adverse Effect Reports and Safety Concerns

Among 5328 discussions, adverse effects were reported in 31.9% (n=1697; 95% CI 30.7%-33.2%) of discussions. Among discussions that referenced adverse effects, the most frequent included muscle pain (n=129, 7.6%) and fatigue (n=110, 6.5%), followed by cognitive symptoms such as “brain fog” (n=61, 3.6%). A detailed breakdown of adverse effects, including neuropsychiatric symptoms, is provided in [Multimedia Appendix 5](#). In 0.3% (n=17) of discussions, users explicitly stated that no adverse effects occurred.

## Decision-Making Factors and Treatment Influences

Of 5328 discussions, the most frequently coded factors associated with statin-related decisions were laboratory results (n=2767, 51.9%), physician recommendations (n=2034, 38.2%), adverse effects (n=1593, 29.9%), and family history of cardiovascular disease (n=822, 15.4%). Additional factors included lifestyle modifications (n=1092, 20.5%), online research (n=573, 10.8%), cost (n=316, 5.9%), and insurance coverage (n=72, 1.4%). Genetic predisposition and clinical guideline recommendations were each coded in 0.5% (n=26 for both) of discussions.

## Alternative Treatment Discussions

Among 5328 discussions, alternative treatments were coded in 46.6% (n=2485) of discussions. The most frequent nonpharmacologic alternatives were dietary changes (n=449, 8.4%) and exercise (n=406, 7.6%). Pharmacologic alternatives included ezetimibe (n=168, 3.2%), Repatha (n=84, 1.6%), and Zetia (n=70, 1.3%). Supplements included fish oil (n=60, 1.1%), coenzyme Q10 (n=57, 1.1%), and red yeast rice (n=53, 1%). Weight loss was coded in 1.4% (n=73) of discussions.

## Information-Seeking Behaviors and Community Engagement

Among 5328 discussions, experience sharing was coded in 78.2% (n=4167) of discussions and advice seeking was coded in 46.1% (n=2456). Discussions focused on alternative treatments occurred in 7.2% (n=386) of cases, while 5.3% (n=284) included questioning medication necessity. Advice seeking was most frequent in r/AskDocs (n=292, 97.3%).

## Emotional Indicators and Adherence Issues

Of the 5328 discussions, emotional expressions were identified in 85.2% (n=4537) of discussions. The most frequently coded emotions were frustration (n=1705, 32%) and confusion (n=911, 17.1%). Medication adherence issues were present in 29.8% (n=1587) of the discussions, including discontinuation (n=522, 9.8%) and dose modification

(n=437, 8.2%). Multiple codes per discussion were permitted; therefore, percentages may sum to more than 100%.

## Mentions of LLM Tools in Reddit Discussions

Explicit mentions of LLM tools occurred in 23/5,328 (0.4%) discussions, totaling 38 mentions. Of the 38 mentions, ChatGPT accounted for 32 (84.2%), followed by other GPT variants (n=3, 7.9%) and Claude, Bard/Gemini, or DeepSeek (n=3, 7.9% combined). Mentions were coded as statin-related medical decision-making (n=2, 5.3%), information seeking (n=6, 15.8%), and general references (n=30, 78.9%). Discussions containing LLM mentions were most commonly observed in r/PeterAttia (6/23, 26.1%) and r/Cholesterol (5/23, 21.7%). Mentions increased from 6 in 2023 to 32 in January 2025 to May 2025.

## Expert Validation Results

Expert validation of a stratified sample of 50 discussions showed a mean validation score of 4.67 (SD 0.74). Interrater agreement was substantial (Cohen  $\kappa$ =0.85; 95% CI 0.78-0.92). Expert feedback noted recurring coding challenges in distinguishing genetic risk factors (eg, lipoprotein[a]) from family history, as well as in identifying cardiovascular event-related content (eg, coronary artery calcium scores).

## Statistical Associations

Sentiment distributions differed across communities ( $\chi^2=335.7$ ;  $P<.001$ ). Negative sentiment was more frequent in medical advice-seeking communities than in lifestyle-focused communities (eg, r/AskDocs: 77.3% vs r/Biohackers: 9.1%;  $P<.001$ ). Adverse effect reporting also differed by community type ( $\chi^2=124.6$ ;  $P<.001$ ). Adverse effect mentions and adherence issues were associated with negative sentiment ( $P<.001$  for both). In adjusted multivariable analysis, factors associated with negative sentiment included adverse effect mentions (adjusted odds ratio [aOR] 3.42, 95% CI 2.89-4.05;  $P<.001$ ), posts in medical advice communities (aOR 2.76, 95% CI 2.31-3.30;  $P<.001$ ), adherence issues (aOR 2.18, 95% CI 1.84-2.58;  $P<.001$ ), and high clinical relevance content (aOR 1.47, 95% CI 1.18-1.84;  $P=.001$ ; Table 3).

**Table 3.** Multivariable logistic regression of factors associated with negative sentiment.

Variable	Adjusted odds ratio (95% CI)	P value
Adverse effect mentions	3.42 (2.89-4.05)	<.001
Posts in medical advice communities	2.76 (2.31-3.30)	<.001
Adherence issues	2.18 (1.84-2.58)	<.001
High clinical relevance content	1.47 (1.18-1.84)	.001

## Discussion

### Main Findings

In this large-scale analysis of statin-related discussions on Reddit, patient discourse was shaped primarily by perceived adverse effects, uncertainty about benefits, and

peer validation, with relatively limited reference to formal cardiovascular risk stratification. Although statins are strongly endorsed by clinical guidelines for ASCVD prevention [2,26], real-world patient narratives were frequently framed around experiential and emotional factors, including fear of long-term harm, symptom attribution, and ambivalence toward medical authority. This

divergence between guideline-based evidence and patient-centered concerns helps explain persistently suboptimal statin adherence despite decades of robust trial data [28].

Notably, fewer than 1 in 8 discussions met criteria for high clinical relevance based on American Heart Association, American College of Cardiology, and US Preventive Services Task Force risk thresholds [2,26]. Instead, most conversations focused on nonspecific symptoms, laboratory fluctuations, or lifestyle considerations, suggesting that engagement with statin therapy often occurs outside a formal risk-benefit calculus. Therefore, effective statin counseling may require attention to patients' beliefs and concerns, beyond clinical risk factors alone.

### ***The Adverse Effect Paradox: Perception vs Trial Evidence***

Adverse effect concerns were prominent in our dataset ( $n=1697$ , 31.9% of discussions), which is not directly comparable to trial-based incidence estimates but may reflect a perception-evidence gap that shapes beliefs and adherence decisions. Evidence from blinded randomized trials and individual-participant meta-analyses suggests that the excess risk of muscle symptoms attributable to statins is small and that most muscle symptom reports under blinded conditions are not attributable to statin therapy [29]. This divergence is consistent with nocebo-related expectation and attribution mechanisms within the broader information environment. In the Anglo-Scandinavian Cardiac Outcomes Trial–Lipid-Lowering Arm, muscle-related adverse events were reported more frequently during the unblinded extension than during blinded treatment despite comparable exposure, supporting expectation-driven symptom attribution [30].

Cognitive concerns (eg, “brain fog”) appeared in 3.6% ( $n=61$ ) of discussions, although randomized trials have not demonstrated statin-attributable cognitive impairment [31]. Regulatory agencies have noted rare, generally reversible postmarketing reports, and the prominence of such concerns online may reflect expectation-driven symptom attribution [32]. Mechanistically, true statin myopathy occurs in a subset of patients, including those with *SLCO1B1* variants or drug-drug interactions, but the high salience of adverse effects in online communities likely reflects both heterogeneity in susceptibility and selection effects (symptomatic users preferentially posting) [33,34]. Clinicians should anticipate that patients initiate statins within an information ecosystem that can magnify harm narratives. Proactive counseling, clear differentiation of evidence-based risks, and structured symptom assessment (eg, Statin-Associated Muscle Symptom Clinical Index) may reduce nocebo-driven discontinuation [35,36].

### ***Emotional Burden as a Hidden Driver of Nonadherence***

Beyond physical symptoms, statin discussions carried substantial emotional content ( $n=4537$ , 85.2%), with frustration ( $n=1705$ , 32%) and confusion ( $n=911$ , 17.1%) predominating. These patterns suggest that statin decisions may be shaped by emotional responses (eg, frustration

and anxiety) in addition to objective risk appraisal. Negative emotion is consistently associated with poorer adherence across chronic disease contexts. A meta-analysis by DiMatteo et al [37] found markedly higher nonadherence among patients with depression. Similarly, medication-specific emotional distress has been associated with treatment discontinuation, even after accounting for clinical depression [38,39]. In our dataset, adherence problems ( $n=1587$ , 29.8%) frequently co-occurred with adverse effect narratives and negative sentiment, suggesting a pathway in which perceived harms generate distress that undermines persistence. These emotional dimensions may be underelicited in routine care, where time-constrained visits often prioritize biomarker review and dose adjustment. Incorporating a brief assessment of medication-related distress into cardiovascular prevention workflows, paired with motivational interviewing, could identify patients at risk for emotionally driven discontinuation earlier in the treatment course.

### ***Community as an Information Filter: The Ecology of Online Health Discourse***

Sentiment and content varied substantially across Reddit communities. Negative sentiment was more prevalent in medical advice-seeking subreddits (eg, r/AskDocs: 77.3%) than in lifestyle-focused subreddits (eg, r/Biohackers: 9.1%). Condition-specific forums (eg, r/stroke and r/HeartAttack) contained the highest proportion of clinically relevant content, while dietary communities (eg, r/keto) more often reflected cholesterol skepticism that diverged from guideline framing [2].

These patterns are consistent with community selection as an “information filter,” in which users preferentially encounter narratives aligned with prevailing community norms and their own concerns; similar dynamics have been described in other online health communities [40,41]. Such community-specific patterns may reinforce users' existing perspectives. For example, dietary forums showed more skepticism toward statins, while condition-specific communities (eg, r/HeartAttack) contained more secondary prevention content. Therefore, community-specific outreach may be more effective than generic education. Strategies could include engaging trusted voices within skeptical forums, tailoring evidence presentation to community values, and directly addressing prevalent misconceptions (eg, “cholesterol myth” narratives), thereby complementing traditional patient counseling.

### ***The Emerging Role of Artificial Intelligence in Patient Decision-Making***

Explicit mentions of LLMs were uncommon ( $n=23$ , 0.4%) but increased from 6 in 2023 to 32 in early 2025, suggesting growing uptake. Users described using tools such as ChatGPT to interpret laboratory results, weigh treatment options, and prepare questions for clinical encounters. To the best of our knowledge, few studies have quantified artificial intelligence (AI) tool use within real-world cardiovascular treatment discussions. Prior research has evaluated LLM performance on medical queries [42,43], but implications for

patient expectations and decision-making remain incompletely characterized. These findings raise questions about how AI-mediated information may shape patient perspectives before clinical visits, warranting further investigation and practical guidance for clinicians.

### Limitations

This study has several limitations. First, Reddit API retrieval imposes sampling constraints. Search results are not uniformly sampled. Each keyword query is capped ( $\leq 1000$  submissions), and ranking algorithms can overrepresent newer posts. This may yield a recency- and visibility-biased corpus. Accordingly, theme and sentiment frequencies should be interpreted as visibility-weighted patterns within the retrieved dataset rather than population-level prevalence for all statin-related content on Reddit; this may inflate high-salience topics (eg, adverse effects, discontinuation, and emotionally charged narratives) and complicate temporal comparisons. Second, Reddit's pseudonymous structure prevents verification of demographics, diagnoses, lipid values, comorbidities, and outcomes. Users are unlikely to be representative of the broader statin-using population, limiting generalizability—particularly to older adults and those with lower digital access or health literacy. Third, self-reported social media content is subject to recall bias and selective posting, potentially overrepresenting unusual or negative experiences. Fourth, although we used expert-guided prompt development and validation, LLM-based extraction may misclassify nuanced language (eg, sarcasm or irony) or clinical attribution, which could affect domain-specific estimates. Evidence-quote requirements and human validation mitigate but do not eliminate this risk. Fifth, findings are limited to 1 platform and English-language content;

discourse may differ across platforms and languages. Finally, this cross-sectional design cannot establish causality, assess within-person changes, or link discourse to clinical outcomes. Because nondisclosure in social media reflects “unreported” rather than missing at random, our conservative handling of unreported variables may underestimate clinical relevance and attenuate associations. We mitigate these limitations through transparent reporting, stratified expert validation, and emphasis on associations rather than causal claims. Future work should consider time-stratified sampling, multiplatform triangulation, and linkage to external clinical data (eg, registries, claims, or electronic health record-linked cohorts) where feasible.

### Conclusions

This study highlights a substantial misalignment between guideline-based statin risk stratification and the lived experiences expressed by patients online. Across 5328 Reddit discussions, discourse centered on perceived adverse effects ( $n=1697$ , 31.9%), emotional distress ( $n=4537$ , 85.2%), and mentions of “natural” alternatives ( $n=2485$ , 46.6%), patterns that may help explain persistent challenges in long-term adherence. These findings suggest that effective cardiovascular prevention must go beyond information provision to address nocebo effects and treatment-related frustration that shape patients' responses to therapy. Methodologically, our validated LLM-enabled pipeline captures nuanced dimensions of treatment experience (eg, ambivalence and peer influence) that are poorly measured by conventional surveillance approaches. As patients increasingly navigate AI-mediated information environments, evidence-informed engagement with digital spaces may support shared decision-making and improve long-term outcomes.

### Acknowledgments

During manuscript preparation, the authors used ChatGPT (OpenAI) for language editing to improve clarity and readability. The tool was used only to assist with wording and expression. All artificial intelligence-generated suggestions were critically reviewed and revised by the authors, who take full responsibility for the accuracy, integrity, and originality of the final manuscript.

### Funding

This work was supported by the National Institutes of Health (grant R00LM014097).

### Data Availability

The datasets generated or analyzed during this study are not publicly available because sharing Reddit-derived data could increase the risk of user traceability and would be inconsistent with the study's privacy protections.

### Authors' Contributions

Conceptualization: SL, JL

Data curation: SL

Formal analysis: SL

Investigation: SL

Methodology: SL, JL

Software: SL

Writing—original draft: SL

Writing—review and editing: SL, JL

### Conflicts of Interest

None declared.

**Multimedia Appendix 1**

Data cleaning and filtering.

[\[DOCX File \(Microsoft Word File\), 19 KB-Multimedia Appendix 1\]](#)

**Multimedia Appendix 2**

Large language model-based extraction protocol and prompt.

[\[DOCX File \(Microsoft Word File\), 27 KB-Multimedia Appendix 2\]](#)

**Multimedia Appendix 3**

Expert validation protocol.

[\[DOCX File \(Microsoft Word File\), 33 KB-Multimedia Appendix 3\]](#)

**Multimedia Appendix 4**

Example posts with large language model-derived theme and sentiment classifications.

[\[DOCX File \(Microsoft Word File\), 18 KB-Multimedia Appendix 4\]](#)

**Multimedia Appendix 5**

Frequency of specific adverse effects mentioned in Reddit discussions on statin therapy.

[\[DOCX File \(Microsoft Word File\), 20 KB-Multimedia Appendix 5\]](#)

**References**

1. Collins R, Reith C, Emberson J, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet*. Nov 19, 2016;388(10059):2532-2561. [doi: [10.1016/S0140-6736\(16\)31357-5](https://doi.org/10.1016/S0140-6736(16)31357-5)] [Medline: [27616593](https://pubmed.ncbi.nlm.nih.gov/27616593/)]
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*. Jun 25, 2019;73(24):3168-3209. [doi: [10.1016/j.jacc.2018.11.002](https://doi.org/10.1016/j.jacc.2018.11.002)] [Medline: [30423391](https://pubmed.ncbi.nlm.nih.gov/30423391/)]
3. Visseren FLJ, Mach F, Smulders YM, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur J Prev Cardiol*. Feb 19, 2022;29(1):5-115. [doi: [10.1093/eurjpc/zwab154](https://doi.org/10.1093/eurjpc/zwab154)] [Medline: [34558602](https://pubmed.ncbi.nlm.nih.gov/34558602/)]
4. Zhang H, Plutzky J, Skentzos S, et al. Discontinuation of statins in routine care settings: a cohort study. *Ann Intern Med*. Apr 2, 2013;158(7):526-534. [doi: [10.7326/0003-4819-158-7-201304020-00004](https://doi.org/10.7326/0003-4819-158-7-201304020-00004)] [Medline: [23546564](https://pubmed.ncbi.nlm.nih.gov/23546564/)]
5. Basios A, Markozannes G, Ntzani EE, et al. Prevalence and determinants of adherence to statin therapy: a systematic review and meta-analysis. *Eur J Prev Cardiol*. Dec 16, 2025:zwaf769. [doi: [10.1093/eurjpc/zwaf769](https://doi.org/10.1093/eurjpc/zwaf769)] [Medline: [41401202](https://pubmed.ncbi.nlm.nih.gov/41401202/)]
6. Vinogradova Y, Coupland C, Brindle P, Hippisley-Cox J. Discontinuation and restarting in patients on statin treatment: prospective open cohort study using a primary care database. *BMJ*. Jun 28, 2016;353:i3305. [doi: [10.1136/bmj.i3305](https://doi.org/10.1136/bmj.i3305)] [Medline: [27353261](https://pubmed.ncbi.nlm.nih.gov/27353261/)]
7. Anker SD, Agewall S, Borggrefe M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. *Eur Heart J*. Aug 7, 2014;35(30):2001-2009. [doi: [10.1093/eurheartj/ehu205](https://doi.org/10.1093/eurheartj/ehu205)] [Medline: [24904027](https://pubmed.ncbi.nlm.nih.gov/24904027/)]
8. Magnani JW, Mujahid MS, Aronow HD, et al. Health literacy and cardiovascular disease: fundamental relevance to primary and secondary prevention: a scientific statement from the American Heart Association. *Circulation*. Jul 10, 2018;138(2):e48-e74. [doi: [10.1161/CIR.0000000000000579](https://doi.org/10.1161/CIR.0000000000000579)] [Medline: [29866648](https://pubmed.ncbi.nlm.nih.gov/29866648/)]
9. Brinton EA. Understanding patient adherence and concerns with STatins and MedicatION Discussions With Physicians (ACTION): a survey on the patient perspective of dialogue with healthcare providers regarding statin therapy. *Clin Cardiol*. Jun 2018;41(6):710-720. [doi: [10.1002/clc.22975](https://doi.org/10.1002/clc.22975)] [Medline: [29749101](https://pubmed.ncbi.nlm.nih.gov/29749101/)]
10. Lardon J, Abdellaoui R, Bellet F, et al. Adverse drug reaction identification and extraction in social media: a scoping review. *J Med Internet Res*. Jul 10, 2015;17(7):e171. [doi: [10.2196/jmir.4304](https://doi.org/10.2196/jmir.4304)] [Medline: [26163365](https://pubmed.ncbi.nlm.nih.gov/26163365/)]
11. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc*. May 2015;22(3):671-681. [doi: [10.1093/jamia/ocu041](https://doi.org/10.1093/jamia/ocu041)] [Medline: [25755127](https://pubmed.ncbi.nlm.nih.gov/25755127/)]
12. Golder S, O'Connor K, Wang Y, Klein A, Gonzalez Hernandez G. The value of social media analysis for adverse events detection and pharmacovigilance: scoping review. *JMIR Public Health Surveill*. Sep 6, 2024;10:e59167. [doi: [10.2196/59167](https://doi.org/10.2196/59167)] [Medline: [39240684](https://pubmed.ncbi.nlm.nih.gov/39240684/)]

13. Somani S, van Buchem MM, Sarraju A, Hernandez-Boussard T, Rodriguez F. Artificial intelligence-enabled analysis of statin-related topics and sentiments on social media. *JAMA Netw Open*. Apr 3, 2023;6(4):e239747. [doi: [10.1001/jamanetworkopen.2023.9747](https://doi.org/10.1001/jamanetworkopen.2023.9747)] [Medline: [37093597](https://pubmed.ncbi.nlm.nih.gov/37093597/)]
14. Press - Reddit. Reddit. URL: <https://www.redditinc.com/press> [Accessed 2025-05-20]
15. Rupert DJ, Gard Read J, Amoozegar JB, et al. Peer-generated health information: the role of online communities in patient and caregiver health decisions. *J Health Commun*. Nov 2016;21(11):1187-1197. [doi: [10.1080/10810730.2016.1237592](https://doi.org/10.1080/10810730.2016.1237592)] [Medline: [27805496](https://pubmed.ncbi.nlm.nih.gov/27805496/)]
16. Golder S, O'Connor K, Hennessy S, Gross R, Gonzalez-Hernandez G. Assessment of beliefs and attitudes about statins posted on Twitter: a qualitative study. *JAMA Netw Open*. Jun 1, 2020;3(6):e208953. [doi: [10.1001/jamanetworkopen.2020.8953](https://doi.org/10.1001/jamanetworkopen.2020.8953)] [Medline: [32584408](https://pubmed.ncbi.nlm.nih.gov/32584408/)]
17. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*. Apr 23, 2013;15(4):e85. [doi: [10.2196/jmir.1933](https://doi.org/10.2196/jmir.1933)] [Medline: [23615206](https://pubmed.ncbi.nlm.nih.gov/23615206/)]
18. Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. *JAMA*. Jan 28, 2025;333(4):319-328. [doi: [10.1001/jama.2024.21700](https://doi.org/10.1001/jama.2024.21700)] [Medline: [39405325](https://pubmed.ncbi.nlm.nih.gov/39405325/)]
19. Goglia D, Vega D. Structure and dynamics of growing networks of Reddit threads. *Appl Netw Sci*. 2024;9:48. [doi: [10.1007/s41109-024-00654-y](https://doi.org/10.1007/s41109-024-00654-y)]
20. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. *J Med Internet Res*. Jun 6, 2017;19(6):e195. [doi: [10.2196/jmir.7082](https://doi.org/10.2196/jmir.7082)] [Medline: [28588006](https://pubmed.ncbi.nlm.nih.gov/28588006/)]
21. Language identification. FastText. URL: <https://fasttext.cc/docs/en/language-identification.html> [Accessed 2025-05-20]
22. Broder AZ. On the resemblance and containment of documents. Presented at: Compression and Complexity of SEQUENCES 1997; Jun 13, 1997; Salerno, Italy. [doi: [10.1109/SEQUEN.1997.666900](https://doi.org/10.1109/SEQUEN.1997.666900)]
23. Introducing GPT-4.1 in the API. OpenAI. 2025. URL: <https://openai.com/index/gpt-4-1/> [Accessed 2025-04-14]
24. Croskerry P. A universal model of diagnostic reasoning. *Acad Med*. Aug 2009;84(8):1022-1028. [doi: [10.1097/ACM.0b013e3181ace703](https://doi.org/10.1097/ACM.0b013e3181ace703)] [Medline: [19638766](https://pubmed.ncbi.nlm.nih.gov/19638766/)]
25. 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: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. Jun 18, 2019;139(25):e1082-e1143. [doi: [10.1161/CIR.0000000000000625](https://doi.org/10.1161/CIR.0000000000000625)] [Medline: [30586774](https://pubmed.ncbi.nlm.nih.gov/30586774/)]
26. US Preventive Services Task Force, Mangione CM, Barry MJ, et al. Statin use for the primary prevention of cardiovascular disease in adults: US Preventive Services Task Force recommendation statement. *JAMA*. Aug 23, 2022;328(8):746-753. [doi: [10.1001/jama.2022.13044](https://doi.org/10.1001/jama.2022.13044)] [Medline: [35997723](https://pubmed.ncbi.nlm.nih.gov/35997723/)]
27. Franzke AS, Bechmann A, Zimmer M, Ess C. Internet research: ethical guidelines 3.0. Association of Internet Researchers; 2020. URL: <https://aoir.org/reports/ethics3.pdf> [Accessed 2026-04-06]
28. Ingersgaard MV, Helms Andersen T, Norgaard O, Grabowski D, Olesen K. Reasons for nonadherence to statins - a systematic review of reviews. *Patient Prefer Adherence*. 2020;14:675-691. [doi: [10.2147/PPA.S245365](https://doi.org/10.2147/PPA.S245365)] [Medline: [32308373](https://pubmed.ncbi.nlm.nih.gov/32308373/)]
29. Cholesterol Treatment Trialists' Collaboration. Effect of statin therapy on muscle symptoms: an individual participant data meta-analysis of large-scale, randomised, double-blind trials. *Lancet*. Sep 10, 2022;400(10355):832-845. [doi: [10.1016/S0140-6736\(22\)01545-8](https://doi.org/10.1016/S0140-6736(22)01545-8)] [Medline: [36049498](https://pubmed.ncbi.nlm.nih.gov/36049498/)]
30. Gupta A, Thompson D, Whitehouse A, et al. Adverse events associated with unblinded, but not with blinded, statin therapy in the Anglo-Scandinavian Cardiac Outcomes Trial-Lipid-Lowering Arm (ASCOT-LLA): a randomised double-blind placebo-controlled trial and its non-randomised non-blind extension phase. *Lancet*. Jun 24, 2017;389(10088):2473-2481. [doi: [10.1016/S0140-6736\(17\)31075-9](https://doi.org/10.1016/S0140-6736(17)31075-9)] [Medline: [28476288](https://pubmed.ncbi.nlm.nih.gov/28476288/)]
31. Zhou Z, Ryan J, Ernst ME, et al. Effect of statin therapy on cognitive decline and incident dementia in older adults. *J Am Coll Cardiol*. Jun 29, 2021;77(25):3145-3156. [doi: [10.1016/j.jacc.2021.04.075](https://doi.org/10.1016/j.jacc.2021.04.075)] [Medline: [34167639](https://pubmed.ncbi.nlm.nih.gov/34167639/)]
32. Cholesterol-lowering drugs get labeling changes. U.S. Food and Drug Administration. 2015. URL: <https://www.fda.gov/drugs/special-features/cholesterol-lowering-drugs-get-labeling-changes> [Accessed 2025-11-10]
33. Cooper-DeHoff RM, Niemi M, Ramsey LB, et al. The Clinical Pharmacogenetics Implementation Consortium Guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and statin-associated musculoskeletal symptoms. *Clin Pharmacol Ther*. May 2022;111(5):1007-1021. [doi: [10.1002/cpt.2557](https://doi.org/10.1002/cpt.2557)] [Medline: [35152405](https://pubmed.ncbi.nlm.nih.gov/35152405/)]
34. SEARCH Collaborative Group, Link E, Parish S, et al. SLCO1B1 variants and statin-induced myopathy--a genomewide study. *N Engl J Med*. Aug 21, 2008;359(8):789-799. [doi: [10.1056/NEJMoa0801936](https://doi.org/10.1056/NEJMoa0801936)] [Medline: [18650507](https://pubmed.ncbi.nlm.nih.gov/18650507/)]
35. Howard JP, Wood FA, Finegold JA, et al. Side effect patterns in a crossover trial of statin, placebo, and no treatment. *J Am Coll Cardiol*. Sep 21, 2021;78(12):1210-1222. [doi: [10.1016/j.jacc.2021.07.022](https://doi.org/10.1016/j.jacc.2021.07.022)] [Medline: [34531021](https://pubmed.ncbi.nlm.nih.gov/34531021/)]

36. Rosenson RS, Miller K, Bayliss M, et al. The Statin-Associated Muscle Symptom Clinical Index (SAMS-CI): revision for clinical use, content validation, and inter-rater reliability. *Cardiovasc Drugs Ther.* Apr 2017;31(2):179-186. [doi: [10.1007/s10557-017-6723-4](https://doi.org/10.1007/s10557-017-6723-4)] [Medline: [28421332](https://pubmed.ncbi.nlm.nih.gov/28421332/)]
37. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. *Arch Intern Med.* Jul 24, 2000;160(14):2101-2107. [doi: [10.1001/archinte.160.14.2101](https://doi.org/10.1001/archinte.160.14.2101)] [Medline: [10904452](https://pubmed.ncbi.nlm.nih.gov/10904452/)]
38. Hoogendoorn CJ, Krause-Steinrauf H, Uschner D, et al. Emotional distress predicts reduced type 2 diabetes treatment adherence in the Glycemia Reduction Approaches in Diabetes: a comparative effectiveness study (GRADE). *Diabetes Care.* Apr 1, 2024;47(4):629-637. [doi: [10.2337/dc23-1401](https://doi.org/10.2337/dc23-1401)] [Medline: [38227900](https://pubmed.ncbi.nlm.nih.gov/38227900/)]
39. Gonzalez JS, Shreck E, Psaros C, Safren SA. Distress and type 2 diabetes-treatment adherence: a mediating role for perceived control. *Health Psychol.* May 2015;34(5):505-513. [doi: [10.1037/hea0000131](https://doi.org/10.1037/hea0000131)] [Medline: [25110840](https://pubmed.ncbi.nlm.nih.gov/25110840/)]
40. Chou WYS, Oh A, Klein WMP. Addressing health-related misinformation on social media. *JAMA.* Dec 18, 2018;320(23):2417-2418. [doi: [10.1001/jama.2018.16865](https://doi.org/10.1001/jama.2018.16865)] [Medline: [30428002](https://pubmed.ncbi.nlm.nih.gov/30428002/)]
41. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. *Health Commun.* Sep 2018;33(9):1131-1140. [doi: [10.1080/10410236.2017.1331312](https://doi.org/10.1080/10410236.2017.1331312)] [Medline: [28622038](https://pubmed.ncbi.nlm.nih.gov/28622038/)]
42. Wei Q, Yao Z, Cui Y, Wei B, Jin Z, Xu X. Evaluation of ChatGPT-generated medical responses: a systematic review and meta-analysis. *J Biomed Inform.* Mar 2024;151:104620. [doi: [10.1016/j.jbi.2024.104620](https://doi.org/10.1016/j.jbi.2024.104620)] [Medline: [38462064](https://pubmed.ncbi.nlm.nih.gov/38462064/)]
43. Liu M, Okuhara T, Chang X, et al. Performance of ChatGPT across different versions in medical licensing examinations worldwide: systematic review and meta-analysis. *J Med Internet Res.* Jul 25, 2024;26:e60807. [doi: [10.2196/60807](https://doi.org/10.2196/60807)] [Medline: [39052324](https://pubmed.ncbi.nlm.nih.gov/39052324/)]

## Abbreviations

**AI** : artificial intelligence

**aOR**: adjusted odds ratio

**API**: application programming interface

**ASCVD**: atherosclerotic cardiovascular disease

**LLM**: large language model

**PRISMA**: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

*Edited by Amaryllis Mavragani; peer-reviewed by Leon Wreyford, Yenan Zhu; submitted 30.Sep.2025; final revised version received 02.Feb.2026; accepted 24.Mar.2026; published 10.Apr.2026*

*Please cite as:*

*Liu S, Liu J*

*Large Language Model–Based Analysis of Statin Therapy Discussions and Sentiment on Social Media: Cross-Sectional Observational Study*

*J Med Internet Res* 2026;28:e85057

URL: <https://www.jmir.org/2026/1/e85057>

doi: [10.2196/85057](https://doi.org/10.2196/85057)

© Siru Liu, Jialin Liu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 10.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.