Treatment options and emotional well-being in patients with rosacea: An unsupervised machine learning analysis of over 200,000 posts

Karan Rajalingam, BS,[a] Nicole Levin, BS,[a] Oge Marques, PhD,[b] James Grichnik, MD, PhD,[c] Ann Lin, DO,[c] and Wei-Shen Chen, MD, PhD[c]

Background: Many patients with rosacea join online support groups to gather and disseminate information about disease management and provide emotional support for others.

Objective: To better understand rosacea patient's primary concerns for the disease as well as their disease search patterns online.

Methods: Overall, 207,038 posts by 41,400 users were collected from June 1, 2017, to June 1, 2022, in a popular online forum. We applied Latent Dirichlet Allocation (LDA), an unsupervised machine learning model, to organize the posts into topics. Keywords for each topic supplied by LDA were used to manually assign topic and category labels.

Results: Twenty-three significant topics of conversation were identified and organized into 4 major categories, including *Management* (50.33%), *Clinical Presentation* (24.14%), *Emotion* (21.97%), and *Information Appraisal* (3.57%).

Limitations: Although we analyzed the largest forum on the internet for rosacea, generalizability is limited given the presence of other smaller forums and the skewed demographics of forum users.

Conclusion: Social media forums play an important role for disease discussion and emotional venting. Although rosacea management was the most frequently discussed topic, emotional posting was a significantly prevalent occurrence. (JAAD Int 2023;13:172-8.)

Key words: machine learning; medical dermatology; rosacea; social media.

INTRODUCTION

Rosacea is a chronic inflammatory disease affecting the face with an estimated prevalence of 5% of the world population.[1] The management of this disease consists of avoidance of triggers, antibiotics, and symptomatic treatment of flushing, lesions, and ocular manifestations.[2] Although scarring and ocular involvement can lead to significant morbidity, the most common complication is the negative psychosocial impact suffered because of the publicly visible manifestation of the disease.[3] A survey conducted by the National Rosacea Society revealed that 90% of rosacea patients have feelings of low self-esteem and a 4.81 times increased likelihood from the general population to develop depression.[4] Social media platforms have become an online safe space for patients with chronic disease to exchange information and gather emotional support.[5-7] How

Correspondence to: Wei-Shen Chen, MD, PhD, Department of Dermatology and Cutaneous Surgery, University of South Florida Morsani College of Medicine, 13330 USF Laurel Dr., 6th Floor, Tampa, FL 33612. E-mail: weishenchen@usf.edu.

rosacea patients use these forums and the extent to which it influences their medical decision-making has yet to be fully characterized.

Previous studies of social media platforms for rosacea have focused the analysis on discussions pertaining to disease management.[8] They concluded that patients are primarily concerned about the adverse effects of treatments and engage more with anecdotal treatment advice supplied by nonhealth care sources than evidence-based information from health care sources. Currently, a deficit in literature exists regarding patient discussions outside of disease management and treatment modalities preferences.

Reddit is the sixth most popular website in the US and is frequented by approximately 1.78 billion visits monthly.[9] Previous studies have demonstrated extensive information exchange regarding dermatological conditions.[10,11] Reddit.com hosts the largest rosacea forum with 41,400 members. We hypothesize that by categorizing all the posts generated in this forum into various topics, we can understand patient concerns on a more comprehensive level and uncover trends in rosacea treatment. The insights gathered will also provide treating physicians with a greater awareness of rosacea patients' needs and perspectives.

METHODS

Reddit forums are composed of chronologically sorted initial posts referred to as submissions and response posts referred to as comments. We scraped all submissions and comments over the 5-year period of Jun 1, 2017, to June 1, 2022, from the rosacea-specific Reddit forum accessed through the link "https://www.reddit.com/r/rosacea." All the subsequent analyses were performed using python libraries.[12]

A total of 19,064 submissions and 187,992 comments for a combined 207,038 posts were identified for analysis (Supplemental Figure 1, available via Mendeley at https://data.mendeley.com/datasets/xzhjy7h94b/1). The web scraping and post aggregation was performed with python libraries json and pandas.[13] We removed 7,593 duplicate posts, defined as those with the same content associated with the same username, in order to filter out posts automatically generated by internet bots. We then searched the post content of the remaining 199,445 posts to extract posting dates, hyperlinks, and mentions of treatments.

Next, the posts were preprocessed for natural language processing. Posts <20 characters in length (n = 12,174) were discarded to exclude terse posts from analysis. The remaining 187,271 posts were stripped of punctuation and the individual words were lemmatized into their root form. This preprocessing step was performed with the python ntlk library.[14,15]

Finally, Latent Dirichlet Allocation (LDA), an unsupervised machine learning algorithm, was applied to classify documents into groups. LDA uses the frequency of word occurrence within posts to associate them with similar posts to constitute a group.[16] The number of groups for LDA to split posts into was empirically determined. We adjusted the number of groups in a stepwise fashion, starting from 5 groups, adding 5 in each step until we reached an interpretable consensus that did not incur significant overlap. This analysis was done with pythons nltk and scikit-learn libraries.[15,17]

We extracted 2 lists of representative word sets from each of the LDA's post groupings in order to assign them labels. The first set consisted of the most frequently occurring words belonging to each group. The second set was the most relevant words occurring in each group. These most relevant words were generated through a Term Frequency-Inverse Document Frequency algorithm. In Term Frequency-Inverse Document Frequency, the relevance of words is weighted directly by their frequency and inversely by their coexistence in other groups.[18] Each group was subsequently assigned with a topic label based on 10 words within each word set. Topics that were labeled the same were combined. These topics were then organized into larger categories and the frequencies of posts within each topic and category were tabulated.

RESULTS
Temporal trends

Posts were made to the forum at fluctuating rates throughout the study period. A date histogram revealed several peaks and troughs (Fig 1). All the peaks occurred during springtime, indicating heightened discussion in the forum. This is consistent with

CAPSULE SUMMARY

- Rosacea is a prevalent skin condition that can significantly impact quality of life.

- Uncovering the most frequently discussed topics in a popular rosacea forum, including emotional venting, gives dermatologists the insight to provide more comprehensive care, addressing both the patients' external and internal well-being.

Abbreviation used:

IPL: intense pulsed light
LDA: Latent Dirichlet Allocation

the known rosacea trigger of changing weather.[19] Interestingly, a similar trend is also independently noted using Google trends, showing an elevated interests during springtime in Google Trends using "rosacea" as the search key word.[20]

The posts were then searched for mentions of a predefined list of widely used treatments and associated generic names for rosacea. The 10 most frequently mentioned treatments were ivermectin ($n = 13,232$), azelaic acid ($n = 12,541$), metronidazole ($n = 6289$), doxycycline ($n = 5674$), intense pulsed light (IPL) ($n = 4368$), retinoids ($n = 3921$), sulfacetamide sulfur ($n = 3681$), pulsed dye laser ($n = 3127$), vitamin b3 ($n = 3063$), and isotretinoin ($n = 2712$) and their temporal trends are illustrated (Fig 2). Between years 2017 and 2022, most of these treatment options remain stable in their appearing frequencies in the forum. However, few treatments were increasingly discussed, including isotretinoin, retinoid, intense pulsed light, and pulsed dye laser. The frequency of mentioning Vitamin B3, however, appears to be trending down lately.

Hyperlinks

Forum participants often included links to external resources to supplement their posts. A total of 10,241 links were shared during the study period. The most commonly shared links pointed to other Reddit posts ($n = 1698$) (Fig 3). The next most frequently shared links directed to product recommendations ($n = 1038$), image hosting platforms where people shared personal images of their rosacea-affected skin ($n = 914$), evidence-based research articles ($n = 547$), rosacea patient advocacy organizations ($n = 312$), YouTube videos ($n = 241$), and tools to divulge skincare ingredients ($n = 125$).

Topic modeling

LDA was applied to model the topics and the associated vocabularies. During this process, we empirically assigned the number of topics to generate the models and found that too few resulted in insufficient discrimination of topics, whereas too many generated redundant ones. It was determined that executing the LDA algorithm to split posts into 40 topics yielded the most interpretable word sets. After detailed inspection and combining redundant topics, we identified 23 unique topics, which were then labeled into broader categories. Topics and
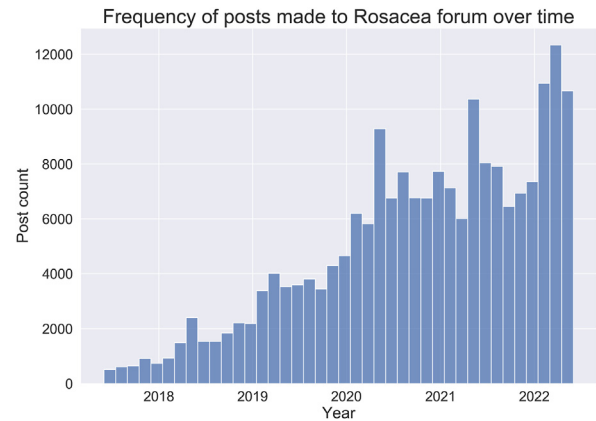
Fig 1. Temporal trends of rosacea post frequency over time.
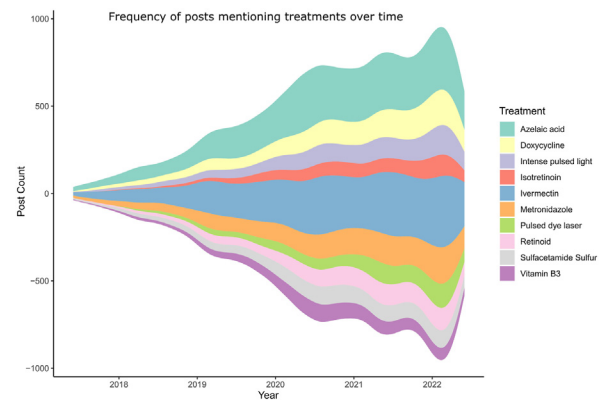

Fig 2. Streamgraph representing the frequency of posts mentioning different rosacea treatments over time.

categories are illustrated (Table I). An abridged word set is listed, and consists of the highest frequency words occurring in each topic or the most relevant words, which were determined by Term Frequency-Inverse Document Frequency analysis.

All the posts could be broadly categorized into *Management* (50.33%), *Clinical Presentation* (24.14%), *Emotion* (21.97%), and *Information Appraisal* (3.57%). Within the *Management* category, the most discussed topics were *Moisturizers* (10.31%), *Ivermectin* (6.87%), and *Diet* (4.07%). The *Clinical Presentation* category included discussion of *Flares* (10.56%), *Diagnosis* (6.80%), and *Keratosis Pilaris* (3.30%). The *Emotional* category consisted of *Negative Sentiment* (11.59%) and *Positive Sentiment* (10.38%). Finally, the *Information Appraisal* category comprised discussion about *Misinformation* (2.22%), and *Demodex* (1.35%).

Treatment analysis

The topics labeled positive sentiment and negative sentiment were further investigated for attitudes
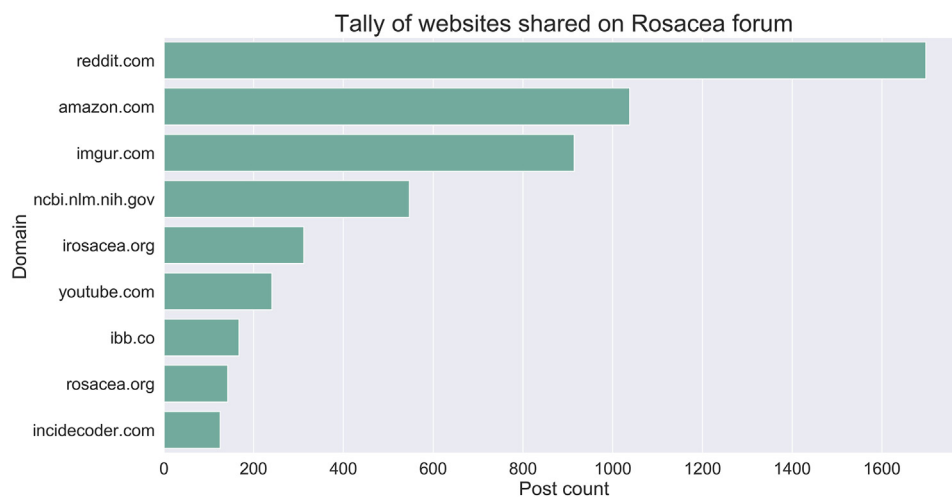
Fig 3. Tally of websites shared on rosacea forum.

toward specific topical, oral and laser treatments used in rosacea. Data from treatments mentioned in more than 50 posts in the positive and negative sentiments topic were tabulated (Fig 4). The overall sentiment toward various treatments was estimated by calculating the proportion of posts in the positive sentiment category mentioning the treatment divided by the total number of positive or negative posts mentioning the treatment. The most positively discussed treatments were metronidazole (52.9%), vitamin C (52.3%), and spirinolactone (52.1%). The least positively discussed treatments discussed were isotretinoin (29.9%), brimonidine (30.5%), and ivermectin (31.3%).

DISCUSSION

People living with chronic diseases are increasingly seeking help online for advice on disease management.[13,14] In the case of rosacea, the diagnosis is associated with high prevalence, complex management, and psychosocial burden. In addition, the peer-based counseling free of costs makes social media forums an even more appealing platform for patients with rosacea to seek help. Analyzing all the posts made to a large online forum for rosacea shed light on major topics of concern and indicated preferences in treatment. In addition, we discovered previously unexplored narratives of positive or negative sentiments, including venting and messages of support.

Management preferences

A study from 2013 analyzing rosacea discussion boards for medical therapies showed that antibiotic treatment, metronidazole, IPL, isotretinoin and clonidine were among the most highly discussed treatments.[17] In comparison, our study demonstrated that the most discussed treatments were ivermectin, azelaic acid, metronidazole, doxycycline, IPL, and retinoids (Fig 2). The contrast in discussion frequency between treatment modalities from 2013 and 2022 reflects the introduction of new treatment options and a shift in treatment preferences over time among rosacea patients.

Additionally, topic modeling revealed that discussion about specific medical therapies formed only a part of management strategies for rosacea. A comparable amount of discussion focused on adjunctive forms of management including use of moisturizers, skin care routines, cleansers, diet, sunscreen, and makeup received. Forum users inquire others about recommendations for these adjunctive strategies to experiment and personalize strategies to manage their rosacea.

Interestingly, treatments were often cited within emotional posts (Supplementary Table I, available via Mendeley at https://data.mendeley.com/datasets/xzhjy7h94b/1). Treatments with the most positive feedback from patients were metronidazole, vitamin C, spironolactone, and clonidine. Whereas, isotretinoin, brimonidine, ivermectin, and oxymetazoline exhibited the least positive associations. This discernable contrast in patient perception may be partially influenced by cost prohibitions. On average, metronidazole and vitamin c are less expensive options than isotretinoin and oxymetazoline. The emotional preferences toward treatments may also be influenced by factors besides treatment effectiveness determined by clinical trials. For example, pulsed dye laser and IPL achieve the comparable level of approval (Fig 4), despite recent studies demonstrating that pulsed dye laser is

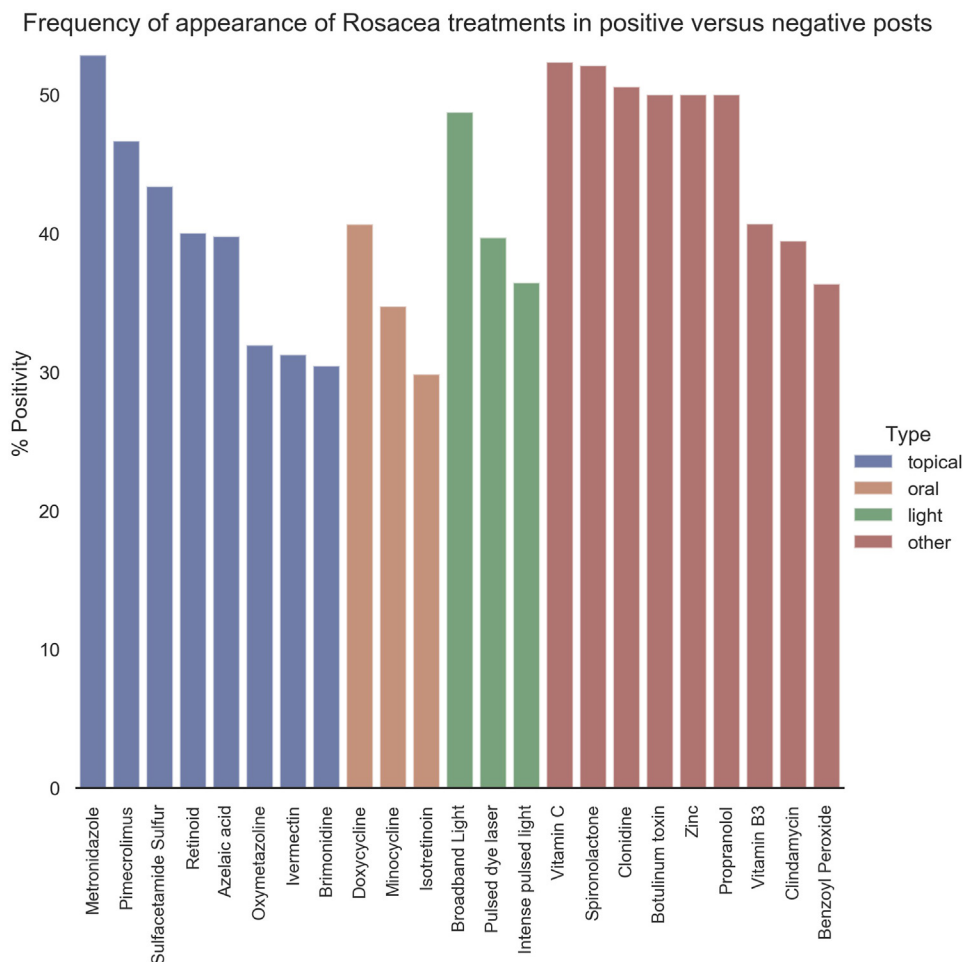**Table I.** Frequency of topics discussed in the Rosacea Reddit Forum

Category	Topic	Posts	%	Word set
Management	Moisturizers	19,314	10.31%	skin, dry, barrier, use, moisture, sensitive, oily, feel, really, moisturizer
50.33%	Ivermectin	12,858	6.87%	ivermectin, doctor, prescription, horse, paste, use, soolantra, gel, cheap
	Skincare routines	8431	4.50%	product, use, start, skincare, ago, routine, skin, new, pore, rosacea, old
	Cleansers	8402	4.49%	cleanser, gentle, skin, use, moisturizer, cerave, sensitive, hydrating, lrp
	Diet	7622	4.07%	carb, meat, elimination, cheese, keto, mast, bread, egg, stomach, fodmap
	Antibiotics	7542	4.03%	resistance, minocycline, erythromycin, tetracycline, clindamycin, systemic
	Lasers	5813	3.10%	bruise, bbl, pdl, pulse, swelling, yag, bruising, genesis, ktp, maintenance, nm
	Ingredients	5614	3.00%	try, ingredient, product, free, fragrance, list, tolerate, irritating, skin, order
	Sunscreen	5351	2.86%	sunscreen, spf, mineral, use, skin, chemical, white, zinc, wear, like, sun, 50
	Azelaic acid	3433	1.83%	acid, azelaic, finacea, use, 15, help, try, gel, ordinary, routine, prescription
	Makeup	3315	1.77%	cream, makeup, use, wear, skin, dr, foundation, cover, routine, day, layer
	Vitamins	2674	1.43%	serum, niacinamide, use, vitamin, ordinary, choice, acid, try, product, zinc
	Sun avoidance	2038	1.09%	trigger, avoid, sun, react, thing, personally, different, skin, common, heat
	Alcohol avoidance	1839	0.98%	vodka, beer, cetyl, cetearyl, drinking, liquor, carbomer, stearyl, gin, alcoholic
Clinical presentation	Flares	19,772	10.56%	face, red, burn, flare, feel, hot, itchy, day, make, just, wash, rosacea
24.14%	Diagnosis	12,740	6.80%	rosacea, acne, diagnose, like, dermatologist, look, symptom, advice
	Keratosis pilaris	6185	3.30%	concealer, kprf, sulfacetamide, keratosis, grisi, pilaris, congrats, rubra
	SIBO	3560	1.90%	immune, sibo, evidence, chronic, disorder, discuss, mast, gastric, blocker
	Ocular involvement	2946	1.57%	compress, ophthalmologist, tear, preservative, systane, eyelash, blepharitis
Emotion	Negative sentiment	21,701	11.59%	sad, ugly, conscious, esteem, rude, embarrassed, depression, date, stare
21.97%	Positive sentiment	19,436	10.38%	yay, encouraging, insight, congratulation, message, informative, reassurance
Information appraisal	Misinformation	4155	2.22%	misinformation, internet, bot, moderator, discussion, community, delete
3.57%	Demodex	2530	1.35%	egg, release, hatch, demodex, dust, population, eyelash, itch, decompose

effective for vascular and erythema symptoms and has been shown to be less painful than IPL.[21]

Emotional posting

Rosacea affects the face, which can lead to negative perceptions of self-image. Many posts share the negative impact of the disease on self-esteem, relationships, and anxiety levels, mainly because of concerns over how others will perceive their skin. Additional negative sentiments uncovered were feelings of isolation, expressions of frustration over unsuccessful treatments, and unexpected adverse effects to treatment (Supplementary Table I, available via Mendeley at https://data.mendeley.com/

Frequency of appearance of Rosacea treatments in positive versus negative posts

Fig 4. Frequency of appearance of rosacea treatments in positive vs negative posts.

datasets/xzhjy7h94b/1). On the other hand, the positive sentiments from the forum included expressions of support, encouragement, and excitement over efficacy of treatment (Supplementary Table II, available via Mendeley at https://data.mendeley.com/datasets/xzhjy7h94b/1).

Emotional posts comprised approximately 22.0% of all posts on the rosacea forum. The high prevalence of these posts implies that a great number of these patients face this psychologic toll. This widespread impact on mental health suggests that clinicians should continue to thoroughly investigate for and offer guidance through anxiety or depression related to the disease.

Analysis of other chronic disease forums has raised concern over the exposure to negative commentary that can psychologically harm patients and further isolate them.[19] However, while examining posts, we found that many negative submissions were replied to with positive posts, and there were a comparable number of positive and negative posts. This finding suggests that online forums can serve as a safe space to vent, and that online support plays an important role in patients' lives and care.

Limitations

We analyzed the largest forum on the internet for rosacea; however, other smaller online forums exist. Reddit users tend to be young males, whereas rosacea patients are usually middle-aged females. Therefore, these data may not be generalizable to all rosacea patients. Additionally, discussion forums are moderated by volunteers and posts deemed too negative or off-topic may be removed. Another limitation to consider is not all users posting have rosacea, whereby a portion of users post to ask whether or not they have rosacea based on symptomatology, which may skew the distribution of the results. Additionally, the machine learning model used to analyze the posts is unsupervised, which means the model outputs have no standard of comparison. Our findings warrant further investigation with prospective trials.

CONCLUSION

Users of rosacea-based social media forums support each other by extensively posting on treatment options, and their messages have far-reaching impact outside of medical clinics. Those who post represent a minor fraction of patients who browse for information. Thus, it is imperative that accurate health information is relayed to patients where they increasingly receive consultations outside of medical settings. Health care professionals and internet forum hosts might need to establish a presence on these forums in order to arbitrate on potential misinformation and encourage office evaluation by dermatologists when appropriate. Additionally, discrepancies in clinical trial-based efficacy of treatment and popularity of discussions in forums may be a result of cost or adverse effects, which should be investigated further. Lastly, chronic diseases like rosacea are associated with psychologic comorbidities, such as anxiety and depression, which can profoundly impact patients' quality of life. Our study demonstrated that this online forum could serve as a safe space for anonymous venting, social support, and positive encouragement among rosacea patients. Having knowledge of what patients are concerned about outside of medical office could better guide physicians caring for patients with rosacea, as well as guide assessment of comorbid psychologic conditions for improved comprehensive care.

Conflicts of interest

Dr Grichnik serves as a consultant to Galileo Group and Canfield Scientific and serves on Skin Advisory Board for Regeneron and Dermatology Advisory Council for Melanoma Research Foundation. Clinical trial support from Novartis, Eli Lilly, Dermira, Elorac, Boehringer, and Amgen. Authors Ramalingam, Levin, and Drs Marques, Lin, and Chen have no conflicts of interest to declare.

REFERENCES

1. Farshchian M, Daveluy S. Rosacea. In: StatPearls. StatPearls Publishing; 2022. Accessed March 10, 2022. http://www.ncbi.nlm.nih.gov/books/NBK557574/
2. Del Rosso JQ, Baum EW. Comprehensive medical management of rosacea: an interim stud report and literature review. *J Clin Aesthet Dermatol.* 2008;1(1):20-25.
3. Rainer BM, Kang S, Chien AL. Rosacea: epidemiology, pathogenesis, and treatment. *Dermatoendocrinol.* 2017;9(1):e1361574. https://doi.org/10.1080/19381980.2017.1361574
4. Coping With Rosacea | Rosacea.org. Accessed June 25, 2022. https://www.rosacea.org/patients/materials/coping-with-rosacea/managing-psychological-and-social-aspects-of-rosacea#Managing
5. Gupta AK, Ivanova IA, Renaud HJ. How good is artificial intelligence (AI) at solving hairy problems? A review of AI applications in hair restoration and hair disorders. *Dermatol Ther.* 2021;34(2):e14811. https://doi.org/10.1111/dth.14811
6. Merolli M, Gray K, Martin-Sanchez F. Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances. *J Biomed Inform.* 2013;46(6):957-969. https://doi.org/10.1016/j.jbi.2013.04.010
7. Patel R, Chang T, Greysen SR, Chopra V. Social media use in chronic disease: A systematic review and novel taxonomy. *Am J Med.* 2015;128(12):1335-1350. https://doi.org/10.1016/j.amjmed.2015.06.015
8. Alinia H, Moradi Tuchayi S, Farhangian ME, et al. Rosacea patients seeking advice: qualitative analysis of patients' posts on a rosacea support forum. *J Dermatolog Treat.* 2016;27(2):99-102. https://doi.org/10.3109/09546634.2015.1133881
9. Global top websites by monthly visits 2020 | Statista. Accessed June 26, 2022. https://www.statista.com/statistics/1201880/most-visited-websites-worldwide/
10. Okon E, Rachakonda V, Hong HJ, Callison-Burch C, Lipoff JB. Natural language processing of Reddit data to evaluate dermatology patient experiences and therapeutics. *J Am Acad Dermatol.* 2020;83(3):803-808. https://doi.org/10.1016/j.jaad.2019.07.014
11. Chu B, Fathy R, Nobles AL, Lipoff JB. Patient crowdsourcing of dermatologic consults on a Reddit social media community. *J Am Acad Dermatol.* 2021;85(1):226-227. https://doi.org/10.1016/j.jaad.2020.08.009
12. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? *Milbank Q.* 2012;90(2):219-249. https://doi.org/10.1111/j.1468-0009.2012.00662.x
13. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks — a publishing format for reproducible computational workflows. In: Loizides F, Scmidt B, eds. IOS Press; 2016:87-90. https://doi.org/10.3233/978-1-61499-649-1-87
14. McKinney W. Data structures for statistical computing in python. *Proc 9th Python Sci Conf.* Published online. 2010:56-61. https://doi.org/10.25080/Majora-92bf1922-00a
15. Bird S, Klein E, Loper E. Natural language processing with python. 1st ed. O'Reilly Media, Inc; 2009.
16. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res.* 2003;3:993-1022.
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12(85):2825-2830.
18. Ramos J. Using TF-IDF to determine word relevance in document queries. In: *Proceedings of the First Instructional Conference on Machine Learning.* New Jersey: IEEE; 2003:29-48; Vol. 242.
19. Changing Seasons Can Trigger Flare-Ups, New Survey Shows. Accessed June 26, 2022. Rosacea.org https://www.rosacea.org/rosacea-review/2014/summer/changing-seasons-can-trigger-flare-ups-new-survey-shows
20. Marchitto MC, Chien AL. Rosacea and associated comorbidities: a google search trends analysis. *J Clin Aesthet Dermatol.* 2020;13(7):36-40.
21. Husein-ElAhmed H, Steinhoff M. Light-based therapies in the management of rosacea: a systematic review with meta-analysis. *Int J Dermatol.* 2022;61(2):216-225. https://doi.org/10.1111/ijd.15680