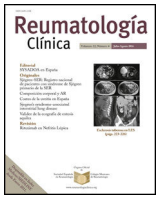
Reumatol Clin. 22 (2026) 502048

Sociedad Española de Reumatología - Colegio Mexicano de Reumatología

Reumatología Clínica

www.reumatologiaclinica.org



Original Article

Rheumatology on Reddit: A descriptive and sentiment analysis

Alfredo Madrid-García,[a,*] Luis Rodríguez-Rodríguez,[a] Beatriz Merino-Barbancho[b]

[a] *Grupo de Patología Musculoesquelética, Hospital Clínico San Carlos, Instituto de Investigación Sanitaria San Carlos (IdISSC), Madrid, Spain*
[b] *Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain*

ARTICLE INFO

Article history:
Received 6 August 2025
Accepted 27 November 2025

Keywords:
Reddit
Rheumatic and musculoskeletal diseases
Social media
Descriptive analysis
Sentiment analysis
Natural language processing

ABSTRACT

Background: Social-media data are increasingly used in medicine for diverse applications. Although several studies have examined social media in rheumatology, most have focused on individual diseases. To date, no work has systematically explored the full spectrum of rheumatology-specific communities on Reddit, and sentiment analysis has not been broadly applied across these communities. This study addresses that gap by identifying rheumatology-related subreddits, describing their characteristics, and analysing the sentiment of their discussions.
Methods: The Reddit search engine was used to identify candidate subreddits. For each subreddit we collected its name, creation date, subscriber count, public status, and activity since May 2023. We included only active subreddits with >1000 subscribers, a clear rheumatology focus, and data retrievable through Pushshift.io. Descriptive metrics were calculated to characterise the selected communities, and a pre-trained, fine-tuned sentiment-analysis model was applied to classify posts.
Results: Twenty subreddits met the inclusion criteria, with subscriber counts ranging from 2000 (r/Behcets) to 70,000 (r/Fibromyalgia). All communities exhibited near-exponential growth from 2016 to 2017 onward. Analysis of the ten most-commented threads in each subreddit yielded 32 thematic categories; the most frequent were "*Patients Like Me*", "*Asking for Emotional Support*", "*Asking for Demographic Information*", and "*COVID-19*". Negative sentiment predominated in subreddits devoted to musculoskeletal disorders of mechanical origin (e.g., costochondritis, back pain) and fibromyalgia. Themes such as "*Expressing Hopelessness*" and "*Asking for Help: Symptom Management*" were also associated with higher levels of negativity.
Conclusions: Rheumatology communities use Reddit to discuss health-related issues, suggesting opportunities to enhance patient support and engagement. Study limitations include the demographic skew of Reddit users, reliance on a model trained on Twitter data, and the exclusion of subreddits with fewer than 1000 subscribers, potentially omitting smaller emerging communities.

© 2025 The Author(s). Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Reumatología en Reddit: un análisis descriptivo y de sentimiento

Palabras clave:
Reddit
Enfermedades reumáticas y musculoesqueléticas
Redes sociales
Análisis descriptivo
Análisis de sentimiento
Procesamiento del lenguaje natural

RESUMEN

Antecedentes: Los datos de redes sociales se utilizan cada vez más en medicina para múltiples aplicaciones. Aunque varios estudios han analizado las redes sociales en reumatología, la mayoría se ha centrado en enfermedades individuales. Hasta la fecha, ningún trabajo ha explorado de forma sistemática todo el espectro de comunidades específicas de reumatología en Reddit, y el análisis de sentimiento apenas se ha aplicado de manera amplia en estas comunidades. Este estudio cubre esa laguna identificando las comunidades relacionadas con la reumatología, describiendo sus características y analizando el sentimiento de sus discusiones.

Abbreviations: AI, artificial intelligence; AMA, Ask Me Anything; COVID-19, coronavirus disease 2019; RA, rheumatoid arthritis; RMDs, rheumatic and musculoskeletal diseases; RoBERTa, robustly optimized BERT pretraining approach; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; URL, uniform resource locator.

* Corresponding author.
E-mail address: fredymad@msn.com (A. Madrid-García).

2173-5743/© 2025 The Author(s). Published by Elsevier España, S.L.U. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Métodos: Se empleó el motor de búsqueda de Reddit para identificar comunidades candidatas. De cada comunidad se recopilaron el nombre, la fecha de creación, el número de suscriptores, el estado de visibilidad, el tamaño y la actividad desde mayo de 2023. Se incluyeron únicamente comunidades activas con más de 1.000 suscriptores, con un enfoque claramente reumatológico, y cuyos datos fueran recuperables a través de Pushshift.io. Se calcularon métricas descriptivas para caracterizar las comunidades seleccionadas, y se aplicó un modelo de análisis de sentimiento, previamente entrenado y ajustado, para clasificar las publicaciones.
Resultados: Veinte comunidades cumplieron los criterios de inclusión, con recuentos de suscriptores que oscilaron entre 2,000 (r/Behcets) y 70,000 (r/Fibromyalgia). Todas las comunidades mostraron un crecimiento casi exponencial a partir de 2016-2017. El análisis de los 10 hilos con más comentarios en cada comunidad, arrojó 32 categorías temáticas; las más frecuentes fueron «Pacientes como yo», «Solicitar apoyo emocional», «Solicitar información demográfica» y «COVID-19». El sentimiento negativo predominó en las comunidades dedicadas a trastornos musculoesqueléticos de origen mecánico (p. ej., costocondritis, dolor de espalda) y fibromialgia. Temas como «Expresar desesperanza» y «Solicitar ayuda: manejo de síntomas» también se asociaron con mayores niveles de negatividad.
Conclusiones: Las comunidades de reumatología utilizan Reddit para debatir cuestiones relacionadas con la salud, lo que sugiere oportunidades para mejorar el apoyo y la implicación de los pacientes. Entre las limitaciones del estudio se incluyen sesgos demográficos de los usuarios, la dependencia de un modelo entrenado con datos de Twitter® y la exclusión de comunidades con menos de 1.000 suscriptores, lo que podría haber omitido comunidades emergentes más pequeñas.
© 2025 Los Autores. Publicado por Elsevier España, S.L.U. Se reservan todos los derechos, incluidos los de minería de texto y datos, entrenamiento de IA y tecnologías similares.

Background

Medical social-media data (i.e., web-based narrative text containing medical content generated by patients, physicians, and other healthcare professionals[1]) have been increasingly recognised as a valuable resource in healthcare research.[2,3] Applications include surveillance systems, pharmacovigilance, patient-healthcare, professional relationships, patient recruitment, regulatory decision-making and drug development.

In rheumatic and musculoskeletal diseases (RMDs), social media has emerged as an important platform for people to share experiences, express concerns, and seek medical advice.[4–6] Studies indicate that RMD patients actively use social media for health-related discussions,[7] and their integration into clinical practice has been examined.[8] The utility of social-media has been described across several RMDs, such as axial spondyloarthritis,[9] back pain,[10] lupus,[11] and ankylosing spondylitis.[12,13]. For instance, recent work highlights the growing role of platforms like TikTok,[14,15] where paediatric rheumatology content has gained significant attention, with millions of views and interactions.

Among major social platforms such as Facebook, X (formerly Twitter), LinkedIn, PatientsLikeMe or Reddit, the latter stands out as a widely used forum where individuals openly discuss medical conditions, making it a valuable source for understanding public perceptions, concerns, and disease experiences.

Prior mapping efforts have catalogued Reddit communities in other specialties such as otolaryngology[16] and dermatology.[17] However, to our knowledge, no study has systematically characterised Reddit communities related to RMDs. Accordingly, our aims were to:

- To identify rheumatology-related subreddits
- To conduct a comprehensive descriptive analysis of their characteristics, including the identification of themes or topics present in their most commented threads
- To analyse the predominant sentiments within these communities and themes

Therefore, this research aims to provide insight into how people discuss rheumatology on Reddit, contributing to a better understanding of the online community's role in supporting patients and professionals in the field of rheumatology.

Materials and methods

Methodology for identifying RMD-related subreddits

The methodology followed to identify the RMDs-related subreddits is similar to previously described methods.[18] Briefly, five main steps were taken:

1. *Search strategy*: The Reddit search engine, with the '*Community*' filter enabled, was used. To ensure thoroughness, a pre-compiled list of terms commonly associated with rheumatology was used to guide the search: *arthritis*, *autoimmune*, *back pain*, *behçet*, *chondropathies*, *connective tissue*, *fibromyalgia*, *gout*, *lupus*, *myositis*, *osteopathy*, *psoriasis*, *Raynaud*, *rheumatology*, *scleroderma*, *Sjogren*, *spondylitis*, *tendinitis*, *thritis*, *uveitis*, *vasculitis*. Initially, all subreddits that could be related, to some extent, to RMDs were manually annotated.
2. *Metadata extraction*: For each potential subreddit identified, we collected the following information: disease; community name; creation date; public, private, or restricted status; community description, number of subscribers (as of 14 May 2024) and last year's activity (i.e., at least one new post).
3. *Relevance classification*: After that, a rheumatologist manually reviewed the subreddits and classified them according to their relevance in the RMDs field into three categories: no affinity, low affinity, or high affinity.
4. *Eligibility screening*: Only subreddits with over 1000 subscribers that were active between May 2023 and May 2024 (i.e., hosting at least one new post), with a strong affinity, and retrievable from Pushshift.io were initially considered for further analysis. If multiple subreddits associated with the same pathology remained, we either kept both or selected just one for inclusion (i.e., the largest community).
5. *Data retrieval verification*: Finally, to ensure a diverse sample of autoimmune diseases, we relaxed the requirement that data be sourced exclusively from the Pushshift.io torrent. In cases where a subreddit was not retrievable from Pushshift.io, a manual request was made to the Pushshift.io Reddit administrator to obtain such data.

The data used in this study were obtained exclusively via a publicly available torrent. More details about Pushshift.io can be found in Supplementary Material "*Pushshift.io*" section.

Table 1
Potential subreddits found through Reddit search engine. Words used: arthritis, autoimmune, back pain, behcet, chondropathies, connective tissue, fibromyalgia, gout, lupus, myositis, osteopathy, psoriasis, raynaud, rheumatology, scleroderma, sjogren, spondylitis, tendinitis, thritis, uveitis, vasculitis.

Identified subreddits		
r/Allopurinol	r/FibroSupport4Adults	r/PsoriasisRemedies
r/ankylosingspondylitis	r/foreverbackpain	r/PsoriaticArthritis
r/arthritis	r/gout	r/Raynauds
r/Autoimmune	r/gout_and_diet	r/RaynaudsSupport
r/AutoimmuneMicrobiome	r/GoutCrew	r/rheumatoid
r/AutoimmuneNeurology	r/GranulPolyangiitis	r/rheumatoidarthritis
r/autoimmuneneutropenia	r/im30andmybackhurts	r/Rheumatology
r/autoimmunity	r/ItsNeverLupus	r/Sciatica
r/Autoinflammatory	r/Keto4Psoriasis	r/Scleroderma
r/back_pain	r/KneeInjuries	r/scoliosis
r/backpain	r/LivingWithLupus	r/ShoulderInjuries
r/Backpaintip	r/lowerbackpain	r/Sjogrens
r/Behcets	r/lupus	r/SjogrensSyndrome
r/cfs	r/LupusAwareness	r/spinalcordinjuries
r/ChronicPain	r/LupusMicrobiome	r/SpineFusion
r/costochondritis	r/LupusResearch	r/Spondylitis
r/CrohnsDisease	r/lupussupport	r/SynovialChondro
r/CutaneousLupus	r/LupusWarriorsUnite	r/Tendinitis
r/disability	r/mctd	r/TensionMyositisSyndrm
r/discoidlupus	r/menhavelupus	r/thoracicbackpain
r/EGPAsupport	r/Myositis	r/Thritis
r/Exercises4BackPain	r/neuropathy	r/TMJ
r/fibro	r/Osteoarthritis	r/UCTD
r/FibroArtsAndCrafts	r/Osteopathic	r/UlcerativeColitis
r/Fibromyalgia	r/Osteopathy	r/Uveitis
r/FibromyalgiaIsReal	r/PiriformisChronicPain	r/Vasculitis
r/FibromyalgiaResearch	r/Psoriasis	r/WegenersGPA
r/fibrosupport	r/PsoriasisDiet	

Methodology for characterising and describing the identified subreddits

Different metrics and statistics (e.g., number of users, messages, comments; thread length, word count per submission) were computed to characterise the different subreddits. To ensure the accuracy of distinct metrics while preserving the integrity of total activity data, we applied a stratified preprocessing approach:

1. *Platform-level volume metrics*: For aggregate statistics (e.g., total submissions, activity over time), no filtration was applied. Entries marked as *[deleted]* or *[removed]* in the body field, those containing one character or fewer, and automated messages (e.g., starting with "*Welcome to r/*") were included. This ensures that the metadata representing the total traffic load on the community was preserved.
2. *User-level metrics*: For statistics describing user behaviour (e.g., messages per user), records where the author field appeared as [deleted] were excluded. This filter was applied to prevent the aggregation of multiple distinct users who had deleted their accounts into a single "deleted" entity, which would otherwise introduce significant bias into user participation distribution.
3. *Engagement ratios*: When calculating engagement metrics (e.g., number of comments on the top ten most commented posts), comments missing a linkage to a parent submission were excluded to prevent the skewing of ratio calculations.

Continuous variables were summarised using the median and the first and third quartiles [Q1–Q3]. The full list of metrics and statistics is shown in the Supplementary Material "*Previous use of Reddit in rheumatology*" section. Finally, correlations between subreddits characteristics were analysed using Spearman correlation, adjusting *p*-values using the false discovery rate.

Methodology for characterisation the ten most-commented posts in each subreddit

To categorise discussion posts effectively, we manually identified recurring themes across posts and established a set of relevant categories based on content, user intent, and emotional tone. Each post was then classified into one or more of these predefined categories.

Methodology for conducting the sentiment analysis

To assess the emotional tone across the selected communities, a sentiment analysis was performed on all submissions from the 20 selected subreddits. A pre-trained RoBERTa-based language model, *cardiffnlp/twitter-roberta-base-sentiment-latest*, fine-tuned with social media data from Twitter was used to classify each submission as negative, neutral, or positive.[19] Briefly, this RoBERTa-based model evaluates sentiment by utilising a deep neural network architecture that has been pre-trained on a vast collection of texts to learn the contextual relationships between words. By fine-tuning the model with labelled Twitter sentiment data, it becomes adept at analysing text inputs, capturing both the semantic meaning and subtle contextual nuances, thereby facilitating accurate classification of sentiment.

The submissions were processed as follows: URLs, and title tags were removed, and submissions without a title or body were excluded (i.e., *[deleted]* or *[removed]*). Finally, the title and the body of each submission were combined into a single record and analysed. Comments on those submissions from other users were not considered. We did not apply topic filters beyond subreddit membership; therefore, posts may occasionally include community-building or off-topic content. Sentiment was computed at the post level, independent of topic.

R version 4.3.2 was used for the descriptive analyses, while Python 3.11 was employed for the sentiment and topic modelling analyses.

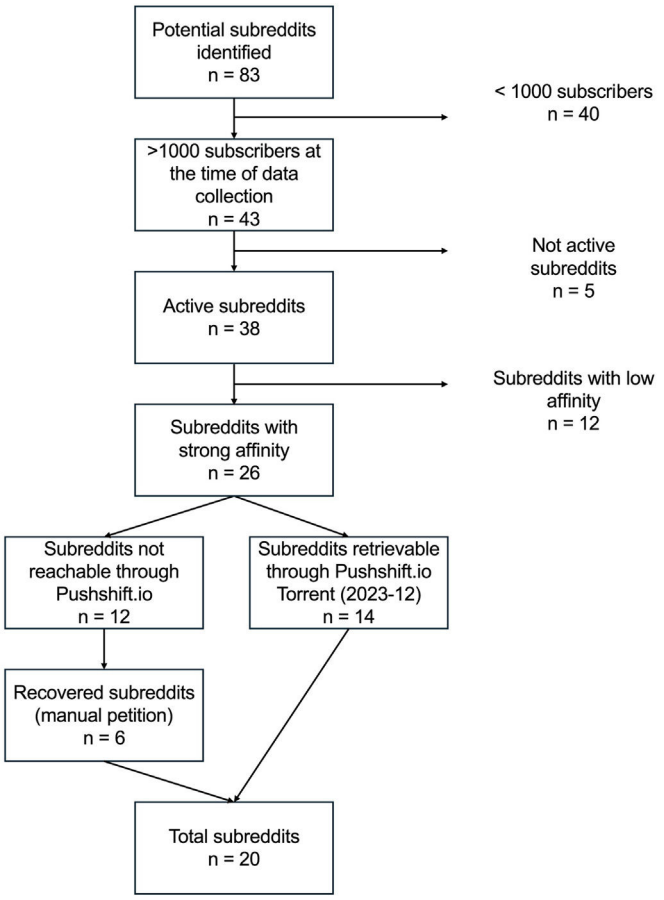


Fig. 1. Inclusion criteria diagram.

The reporting of this study conforms to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.[20]

Results

Identification of the RMD-related subreddits

The pre-compiled list of terms allowed us to identify 83 subreddits or communities potentially related to RMDs (Table 1). Notably, over 70 of them were classified as having some degree of relevance in the RMDs field. Supplementary Excel File Subreddits shows the general information of the 83 subreddits identified, including their fulfilment of the eligibility criteria. After applying these criteria, 20 communities were selected for further analysis (Fig. 1 presents the eligibility criteria diagram).

Characterisation of the RMD-related subreddits

Table 2 presents the main characteristics of the 20 selected subreddits. The oldest identified community is r/Fibromyalgia, which dates back to 2009, whereas the newest is r/rheumatoidarthritis, created in 2018. Regarding the number of subscribers, the community with the highest number was r/Fibromyalgia, and the lowest was r/Behcets. We observed a non-significant correlation between the lifetime of these communities and the number of subscribers (rho = 0.69, $p = 0.095$). Furthermore, when studying annual. User participation in each subreddit (Supplementary Figs. 1 and 2), a common characteristic was observed: from 2016 to 2017 onward, growth typically followed an exponential pattern.

Regarding the subreddit content, we observed substantial heterogeneity across communities in the cumulative number of submissions and comments, the median daily activity rate (i.e., number of messages, including submissions and comments, posted per day), and the level of engagement. On the one hand, we observed that both the cumulative number of submissions and comments were positively correlated with the number of subscribers (rho = 0.97, $p < 0.0001$; rho = 0.92, $p < 0.0001$, respectively). Furthermore, the median number of messages per day was also correlated with the number of subscribers (rho = 0.74, $p = 0.028$), as well as with the cumulative number of both submissions and comments (rho = 0.81, $p = 0.002$; rho = 0.86, $p = 0.0002$, respectively). On the other hand, regarding engagement, the comment-to-submission ratio showed only a significant correlation with the median number of messages per user (rho = 0.86, $p = 0.0002$).

Peak user interaction typically occurred between 16:00 and 23:00, as shown in Supplementary Figs. 3 and 4. The days of the week exhibiting the highest levels of activity were Thursdays, Wednesdays, Fridays, and Tuesdays, with 9, 7, 2, and 2 communities, respectively (Supplementary Figs. 5 and 6). For all but one of the communities analysed, r/autoimmunity, the majority of activity was concentrated in the second half of the year, with December (8 communities), October (4 communities), and August (3 communities) experiencing the highest levels of engagement (Supplementary Figs. 7 and 8). We observed no significant correlations between these times of peak activity (hour, weekday or month) and the other studied characteristics nor among themselves.

Thematic characterisation of most -commented posts

Finally, the titles and bodies of the ten most-commented threads per subreddit were collected (Supplementary Excel File Most Commented Threads, "posts" sheet). In Table 3, the titles of the threads with the highest number of comments per subreddit are shown. After manual review of titles and bodies, 32 different themes or categories were identified (Supplementary Excel File Most Commented Threads_R1, "categories" sheet). A brief definition of each category is also included. We observed that 107, 62, 26, and 5 threads were characterised using 1, 2, 3 or 4 themes, respectively.

When analysing the themes' characteristics (Supplementary Excel File Most Commented Threads "categories_analysis" sheet), we observed that "Patients Like Me", "Asking for Emotional Support", and "Asking for Demographic Information" were the three most frequently identified. In addition, they appeared across the highest number of communities. Furthermore, a significant number of the identified themes (8 out of 32, 25%) appeared across ten or more communities. Finally, themes present in posts with the highest number of comments were those related to "Asking for Impact of Environment/Life Events on Disease", "Ask Me Anything (AMA)" and "Sharing Working Experience".

Sentiments found across the different communities

The top five most positive and negative comments from each community were extracted and can be seen in Supplementary Excel File Sentiment Analysis. Moreover, Fig. 2 depicts the prevalence of negative, neutral, and positive classes across all analysed submissions in each subreddit. The highest percentage of negative comments pertained to musculoskeletal disorders of mechanical origin, such as costochondritis or back pain, as well as to fibromyalgia. r/Fibromyalgia is the most polarised community with the fewest neutral comments. The sentiment analysis conducted revealed the emotional experiences of individuals with RMDs. These posts frequently described intense physical pain, emotional

Table 2
Main characteristics of the selected subreddits. Data as of May, 15th 2024.

Disease	Subreddit	Creation date	Top 40k subreddits Pushshift.io	Subscribers (15 May 2024)	Submissions (S)	Comments (C)	Ratio C/S
Ankylosing spondylitis	r/ankylosingspondylitis	2012-03-01	1	21.5k	16820	207859	12.36
Arthritis	r/Thritis	2011-02-11	1	17.4k	8464	63961	7.56
Back pain	r/backpain	2009-12-06	1	39.5k	29519	149881	5.08
Back pain	r/Sciatica	2012-10-17	1	32.8k	23529	225745	9.59
Behcet's disease	r/Behcets	2015-06-07	0	2.0k	1177	9361	7.95
Costochondritis	r/costochondritis	2015-07-16	1	16.1k	13555	103390	7.63
Fibromyalgia	r/Fibromyalgia	2009-04-18	1	70k	62516	739395	11.83
Gout	r/gout	2011-10-07	1	24.5k	15395	176077	11.44
Mix	r/Autoimmune	2016-03-03	1	14.9k	6356	41182	6.48
Mix	r/autoimmunity	Private	1	Private	5938	37877	6.38
Mixed connective tissue disease	r/mctd	2014-05-31	0	2.2k	714	5816	8.15
Osteoarthritis	r/Osteoarthritis	2015-12-10	0	5.6k	1852	7125	3.85
Psoriatic arthritis	r/PsoriaticArthritis	2014-01-14	1	13.3k	8821	108968	12.35
Raynaud syndrome	r/Raynauds	2015-06-06	0	9.4k	4248	27833	6.55
Rheumatoid arthritis	r/rheumatoid	2012-03-18	1	25.8k	18135	199863	11.02
Rheumatoid arthritis	r/rheumatoidarthritis	2018-05-29	1	13.2k	6169	59495	9.64
Scleroderma	r/Scleroderma	2013-07-08	0	3.4k	2043	11868	5.81
Sjogren's syndrome	r/Sjogrens	2012-10-16	1	10.5k	6377	74571	11.69
Systemic lupus erythematosus	r/lupus	2010-06-09	1	32.2k	30553	279714	9.16
Uveitis	r/Uveitis	2015-05-21	0	2.3k	1471	12199	8.29
Total					263652	2542180	9.64

Table 3
Titles of submissions with most comments per subreddit.

Subreddit	Title
r/ankylosingspondylitis	*What was your age at diagnosis?*
r/Autoimmune	*Multiple AutoImmune Disorders?*
r/autoimmunity	*What's an interesting autoimmune disease to cover in my presentation?*
r/backpain	*The backpain is getting worse.*
r/Behcets	*Ethnicity and Behcets? Would everyone be willing to list theirs? And the big question . . .*
r/costochondritis	*Covid vaccine & Costochondritis*
r/Fibromyalgia	*How old are you all?*
r/gout	*How Old Are We At The Moment.*
	I Want To See Age Variation With Gout?
r/lupus	*Undiagnosed Seeking Diagnosis: Is this Lupus?*
r/mctd	*Vaccine?*
r/Osteoarthritis	*Looking for advice on cortisone injections and whether I should have them done by my knee surgeon?*
r/PsoriaticArthritis	*Who here is under the age of 40?*
	Sometimes having PA makes me feel so alone . . .
r/Raynauds	*Age Poll - First Signs*
r/rheumatoid	*How old is everyone on this sub?*
r/rheumatoidarthritis	*When were you diagnosed?*
r/Sciatica	*Your Sciatica and Back Pain Experiences Megathread*
r/scleroderma	*I need help please*
r/Sjogrens	*COVID-19 vaccination x Sjogren's experience thread*
r/Thritis	*How old is everyone here?*
r/Uveitis	*Covid Vaccine effect on Uveitis*

distress, and frustration related to the progression of chronic conditions. Phrases such as "*cripples me under a load of severe pain*" and "*I feel so worthless*" illustrate the profound impact that chronic illness can have on quality of life. Users also expressed dissatisfaction with medical treatments, highlighting the perceived inefficacy of certain interventions, delayed diagnoses, or mismanagement by healthcare professionals (e.g., "*I hate doctors. My rheumatologist gaslit me and doesn't think my pain is real. Said she thinks it's just the area overly sensitised. Waited 6 months to speak with her and this bullshit happened*"). Moreover, a recurring sentiment across multiple subreddits was frustration with the healthcare system (e.g., "*USA have the worst healthcare in the world*"). Many users expressed dissatisfaction with long wait times, misdiagnoses, financial burdens (e.g., "*Cannot afford the endless drugs and tests and don't know if I'll be able to work when I finally find work*"), and lack of access to effective treatments. Finally, chronic pain and its impact on daily life and mental health were recurring topics, with users expressing how debilitating these conditions can be.

Despite the overall dominance of negative sentiment, positive sentiments were present, particularly in posts where users reported successful treatments, improvements in symptoms, or expressed gratitude for community support. These posts were often marked by expressions of relief and optimism. For example, posts like "*Cosentyx is working!!!*" and "*My doctor signed off for permanent disability placard!!!:) it's a good day everybody*".

Users frequently expressed appreciation for the support and advice shared within these online spaces. Positive posts often referenced the sense of belonging and relief that came from discussing their experiences with others facing similar challenges (e.g., "*One week to go until I run 50 miles. This community has been amazing, and thank you all for the support!*").

Finally, sentiment analysis was combined with the thematic characterisation of the ten most-commented posts in each community. The aggregation of positive/negative/neutral probabilities by category allowed us to identify dominant emotional trends within each theme. The categories associated with the highest negative

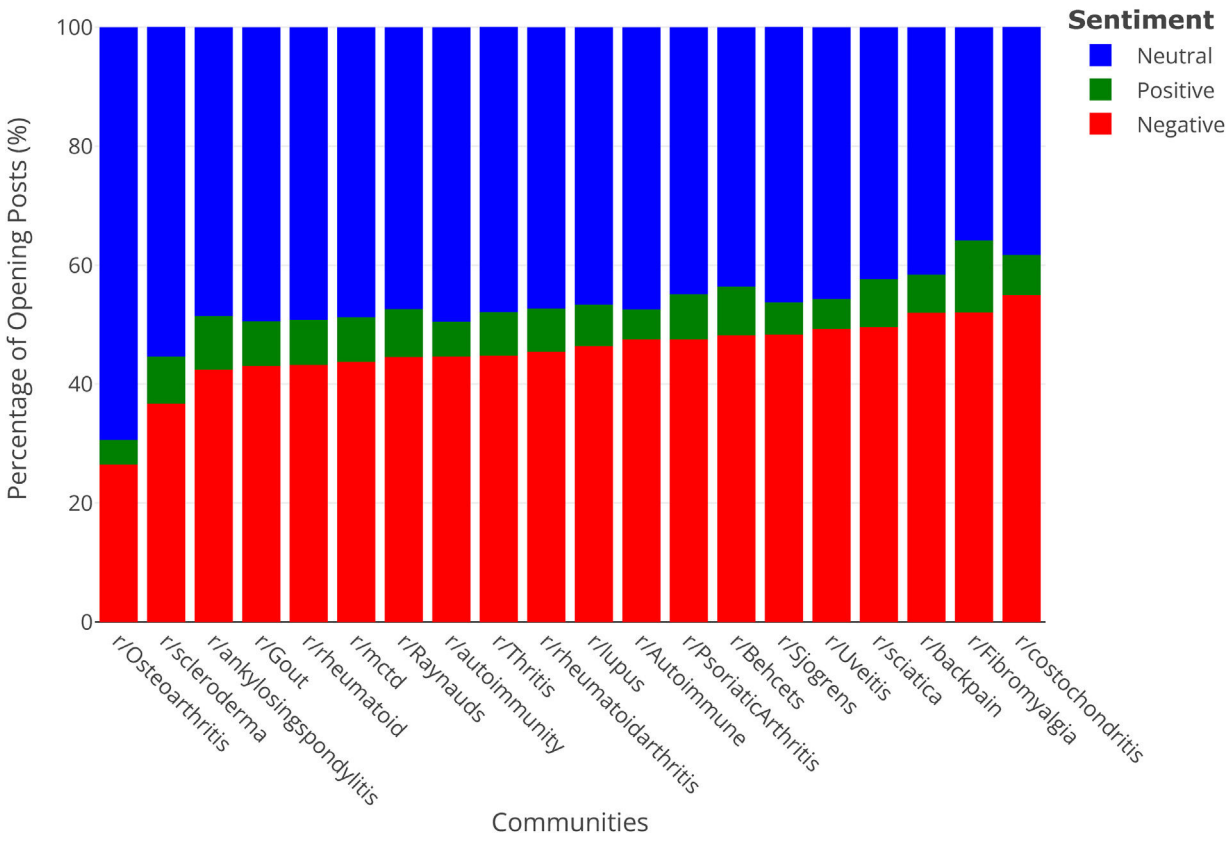


Fig. 2. Sentiment in RMDs Reddit communities sorted by negative percentage.

sentiments were those related to "Expressing Hopelessness", "Asking for Help: Symptom Management", and "Lack of Understanding from Healthy Peers". Conversely, those themes exhibiting greater neutral sentiments were "Asking for Demographic Information", "Asking for Experiences with Medical Procedures", and "Asking for Causes of Disease". No theme showed predominant positive sentiments.

Discussion

We have identified more than 70 rheumatology-related subreddits and selected 20 for further analysis. A thorough description of each community was presented, including measures of engagement. Subsequently, we identified several themes contained in the title and body of the ten most commented threads of each subreddit, and finally, applied sentiment analysis to characterise both the communities and the identified themes.

Regarding community characteristics, the number of users who participate in each RMD-related subreddit has increased in recent years. This yearly growth (Supplementary Figs. 1 and 2) may indicate a trend towards a more informed and community-oriented disease management. The ability to share experiences and advice instantaneously has fostered a global community of patients who support each other, challenging the isolation that often accompanies these chronic conditions. It is fundamental to consider that, as the lack of adequate social support contributes to disability in the context of RMDs,[21] the role of social media becomes crucial for patients with these diseases. In fact, a recent study showed that people are eager to share their personal experiences with chronic diseases on social media platforms despite possible privacy and security issues.[22] Additionally, a study conducted in 2023 showed that the number of Reddit users grew during the COVID-19 pandemic.[23] According to our results, this theme was the 4th most ubiquitous, being present in 18 posts from 11 communities. Furthermore, it was present in the most commented posts from r/mctd, r/costochondritis, r/Sjogrens, and r/Uveitis communities. This suggests how RMD patients experienced the COVID-19 pandemic and how they used social media seeking support or advice.

In terms of community history, r/Fibromyalgia is the oldest community among those we studied. Its considerable size and level of activity are understandable, given the characteristics and challenges faced by patients with this condition. The association between fibromyalgia and social media has also been described elsewhere.[24] When comparing the number of subscribers of r/Fibromyalgia to that of the largest dermatology-related community in 2017, r/SkincareAddiction, the latter was three times larger (i.e., 210k).[17] However, when compared with the largest otolaryngology community in 2021, r/earrumblersassemble, their subscriber counts were similar (i.e., 66k vs 70k). Regarding the themes identified, nine were associated with the most commented threads, with "Asking for Impact of Environment/life events on Disease", "Asking for Emotional Support", and "Asking for Causes of Disease", the most frequent, being present in three, two, and two posts. Notably, although the first theme was present in other communities, half of the posts belonged to r/Fibromyalgia, suggesting this community's particular interest in the role of environmental and life events in fibromyalgia.

Regarding engagement, the r/ankylosingspondylitis community has the highest comment-to-submission ratio, and the messages per day, per user, and number of words are among the top three. This suggests that patients with ankylosing spondylitis receive greater community support. In addition, this community is among those with a high number of themes identified in their most commented thread (14 categories), which could reflect the high activity

of the members of this community. Those themes present in the highest number of posts were "Asking for Symptom Experiences", "Asking for Disease Impact in Life", and "Patients Like Me". Considering these themes, we hypothesise that this community is particularly interested in how the disease affects daily life and how it is experienced by other members.

Regarding the identified themes, the "Ask Me Anything (AMA)" category was observed almost exclusively on r/gout. In these AMA sessions, professionals respond to questions within a specific time frame (e.g., "*I'm Dr. Rick Johnson, a medical professor who has studied uric acid for 20 years. Ask me anything on March 14!*"). This highlights gout patients' interest in exchanging information with healthcare professionals via social media. In addition, apart from this theme, "Asking for Demographic Information" (mostly related to age), and "Sharing Experiences with Disease", were the other most frequent themes in this community.

When analysing the number of subscribers, those subreddits related to low-prevalence diseases (e.g., r/Behcets, r/mctd, r/Uveitis) were much smaller than the rest. In fact, we hypothesise that the activity and size of a Reddit community, in the context of RMDs (both autoimmune and non-autoimmune), depend on different factors: average age at diagnosis, the prevalence of the disease, the level of digitalisation, the patients' proficiency in English, the lack of social and family support available to patients, and patient's social and economic resources for accessing the Internet.

The sentiment analysis highlights the dual emotional landscape of chronic illness, where users grapple with both the challenges of their conditions and the comfort found in community engagement. Moreover, the predominance of negative sentiment in fibromyalgia and mechanical-pain communities points to unmet needs that can affect treatment adherence, illness perceptions, and clinician-patient communication. In conditions with limited objective biomarkers and high diagnostic/therapeutic uncertainty, persistent pain, perceived invalidation, and lengthy care pathways may erode treatment expectancy and trust, potentially reducing persistence and fostering treatment switching.

Finally, data collected from Reddit could be used to perform artificial intelligence and natural language processing analyses, such as topic modelling for characterising the latent topics or graph analysis to identify hubs and authoritative Reddit users in each community. Moreover, these data can be used to feed large-language models, enhancing the capabilities of AI-driven chatbots.[18,25]

Limitations

- It should be noted that the information extracted from Reddit is not representative of the general population suffering from musculoskeletal diseases. Patients with RMDs are usually elderly people and women. Conversely, Reddit users tend to be male between 18 and 34 years old, as highlighted by Proferes et al.[26] Furthermore, given Reddit's predominance of English-language content, our analysis inherently excludes patients with no knowledge of English. Other existing challenges in the use of Reddit include privacy concerns, self-reported inaccuracies, misinformation, low data quality, limited content reliability, and ethical aspects. Fiesler et al.[27] have recently described ethical considerations in Reddit research.
- The use of Pushshift.io to obtain Reddit data is not without limitations. First, not all subreddits are accessible through Pushshift. Second, because the data are not collected in real time, some posts and comments may be missing or deleted. Third, uncompressing and processing the large data files can be technically challenging. Beyond these practical issues, there are also ethical and legal considerations. Pushshift may retain metadata from posts and comments that have subsequently been deleted on Reddit, which can restrict users' ability to exercise their "right to erasure" under the GDPR. In this study, we did not attempt to recover deleted content and did not analyze personal information. To minimize re-identification risks, no usernames were collected or reported, and all verbatim excerpts were carefully screened by two authors for potential re-identifying information (including usernames, names, specific locations or institutions, exact dates, or highly idiosyncratic personal details). As a result, no paragraph was modified. While our approach is consistent with prior peer-reviewed work using Pushshift, we acknowledge that recent disputes over its compliance with Reddit's API policies raise concerns about the long-term sustainability and legitimacy of this data source.
- Our study focused on subreddits with at least 1000 subscribers. While this approach helped us characterise well-established and active communities, it may also exclude smaller, potentially emerging subreddits that represent rarer or niche aspects of RMDs. This threshold was a pragmatic choice to ensure study feasibility and the robustness of our analyses, as communities with very low activity can yield unreliable conclusions and present data acquisition challenges. Consequently, several subreddits for less common RMDs, such as r/WegenersGPA, r/Myositis, and r/Vasculitis, were not included in the final analysis. We did not conduct a sensitivity analysis including subreddits with fewer than 1000 subscribers, so the impact of excluding these smaller but potentially important communities could not be quantified and should be explored in future work. However, by identifying these smaller communities in the Supplementary Material, we provide a foundation for future targeted studies that could explore the unique dynamics of these valuable, niche patient populations.
- The classification of subreddits regarding their relevance to RMDs was performed by a single rheumatologist without inter-rater reliability assessment. While we consider the explicit nature of community descriptions to minimise ambiguity, we acknowledge the potential for selection bias due to the lack of a second reviewer.
- A key limitation is the external validity of our sentiment analysis, as the RoBERTa-based model was fine-tuned on Twitter data. While both are social media platforms, their linguistic characteristics, user context, and community cultures differ significantly. Twitter's character limits historically encouraged brevity, abbreviations, and a faster pace of communication. In contrast, Reddit posts are often long-form, narrative, and highly contextualised within the specific culture and norms of a subreddit. This mismatch means that nuances specific to Reddit's conversational style such as sarcasm, complex narratives of patient experiences, or community-specific jargon might not be captured as effectively by a model trained on Twitter's distinct discourse patterns. Consequently, the accuracy of sentiment classification may be affected, and this should be considered when interpreting the sentiment results. To reduce the impact of this limitation we examined the generalisability of the sentiment model by evaluating it against a labelled Reddit-specific dataset as described in the Supplementary Material.
- We did not perform a formal, in-domain evaluation of the sentiment classifier using manually labelled Reddit data. Specifically, we did not obtain a stratified sample of posts from rheumatology subreddits annotated by multiple human coders to quantify inter-annotator agreement and model performance. Such an evaluation would provide a more precise estimate of how well a Twitter-trained model generalises to medical Reddit.
- Poorly structured texts and nuanced language, such as sarcasm, idiomatic expressions, or culturally specific references commonly found on social media, can lead to misclassification of sentiment. Moreover, while positive, negative, and neutral classes are com-

monly used in sentiment analysis, they often fail to capture the full spectrum of emotions and the varying intensities with which individuals may express them.
- When analysing periods of heightened activity, it is important to consider that Reddit is an international platform, with users distributed across various time zones.
- There are other metrics that could have been considered for the descriptive analysis, such as the number of *upvotes* per message, or the presence of multimedia content embedded in the messages.
- We did not account for bots and automated messages, which could have artificially inflated the message count.

Conclusion

Reddit provides a valuable platform for studying patient concerns, needs, and perceptions about their RMDs as evidenced by the number of subscribers, the diversity of communities, and the extent of their activity and size.

This paper advances our understanding of the communities dealing with rheumatic and musculoskeletal conditions on Reddit. By examining their characteristics, the study provides valuable information that could inform the development of more effective patient support and engagement strategies in digital environments. Moreover, by identifying the various communities related to RMDs on this platform, we establish a foundation for researchers to conduct more focused studies using data from this source.

CRediT authorship contribution statement

Alfredo Madrid-García: Conceptualization of this study, methodology, data curation, coding, review, writing (original draft preparation). Luis Rodríguez-Rodríguez: Conceptualization, writing (review & editing). Beatriz Merino-Barbancho: Formal analysis, data curation, writing (original draft preparation).

Ethical approval

Comité de Ética de la Investigación con Medicamentos Hospital Clínico San Carlos Ethics Review Board approval (24/417-E) was obtained on June 10, 2024 with no objections raised.

Consent for publication

Not applicable.

Clinical trial number

Not applicable.

Declaration of generative AI and AI-assisted technologies in the writing process

Large language models were employed to perform spelling and grammar checks.

Funding

This study did not receive funding.

Declaration of competing interest

None declared.

Data availability

The authors would like to thank the *Pushshift.io* team and r/pushshift/, specially to Watchful1 Reddit user.

Much of the data in this study is publicly accessible through Academic Torrents. However, the subreddits listed with a '0' value in the *Top 40k Subreddits* column of Table 1 are not available. The code used for pre-processing the subreddits was obtained from Madrid et al.[18]

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version available at https://doi.org/10.1016/j.reumae.2026.502048.

References

1. Denecke K. Health web science: social media data for healthcare. Springer; 2015.
2. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res. 2021;23, e17917.
3. Kanchan S, Gaidhane A. Social media role and its impact on public health: a narrative review. Cureus. 2023;15, http://dx.doi.org/10.7759/cureus.33737.
4. Berenbaum F. The social (media) side to rheumatology. Nat Rev Rheumatol. 2014;10:314–8.
5. Wilson N, Liu J, Adamjee Q, Di Giorgio S, Steer S, Hutton J, et al. Exploring the emotional impact of axial spondyloarthritis: a systematic review and thematic synthesis of qualitative studies and a review of social media. BMC Rheumatol. 2023;7:26.
6. Abbasi-Perez A, Alvarez-Mon MA, Donat-Vargas C, Ortega MA, Monserrat J, Perez-Gomez A, et al. Using twitter data analysis to understand the perceptions beliefs, and attitudes about pharmacotherapy used in rheumatology: an observational study. Healthcare. 2023;11:1526.
7. Studenic P, Alunno A, Stones SR, Ritschl V, Nikiphorou E. Social media use for health-related purposes by people with rheumatic and musculoskeletal diseases – results of a global survey. Arthritis Rheum. 2018;70:2613–5.
8. Erdogan M, Aydin O, Seyahi E. Patients with rheumatic diseases are ready to use social media in clinical practice; what about rheumatologists? A cross-sectional survey. Rheumatol Int. 2022;42:717–23.
9. Reuter K, Danve A, Deodhar A. Harnessing the power of social media: how can it help in axial spondyloarthritis research? Curr Opin Rheumatol. 2019;31:321–8.
10. Taik FZ, Bensaid R, Adnine A, El Mansouri N, Aharrane FZ, Amar A, et al. Use of social media as a source of health information among patients with chronic low back pain. Musculoskeletal Care. 2024;22:e1846, http://dx.doi.org/10.1002/msc.1846.
11. Blackie CA, Gualtieri L, Kasturi S. Listening to patients with lupus: why not proactively integrate the internet as a resource to drive improved care? J Med Internet Res. 2023;25, e44660.
12. Dzubur E, Khalil C, Almario CV, Noah B, Minhas D, Ishimori M, et al. Patient concerns and perceptions regarding biologic therapies in ankylosing spondylitis: insights from a Large-Scale survey of social media platforms. Arthritis Care Res (Hoboken). 2019;71:323–30.
13. Reilly E, Sengupta R. Back pain, ankylosing spondylitis and social media usage; a descriptive analysis of current activity. Rheumatol Int. 2020;40:1493–9.
14. La Bella S, Di Ludovico A, Mainieri F, Lauriola F, Silvestrini L, Ciarelli F, et al. Quality and characteristics of pediatric rheumatology content on social media: toward a new era of education for patients and caregivers? J Rheumatol. 2024;52:640–3, http://dx.doi.org/10.3899/jrheum.2024-0039, jrheum.2024-0039.
15. La Bella S, Breda L, Ravelli A. "*Gallia est omnis divisa in partes tres*": social media platforms as a new educational channel for pediatric rheumatology. J Rheumatol. 2024;51:741–3, http://dx.doi.org/10.3899/jrheum.2024-0408.
16. Xie DX, Boss EF, Stewart CM. An exploration of otolaryngology in the Reddit community. Laryngoscope. 2022;132:284–6.
17. Buntinx-Krieg T, Caravaglio J, Domozych R, Dellavalle RP. Dermatology on Reddit: elucidating trends in dermatologic communications on the world wide web. Dermatol Online J. 2017;23.
18. Madrid-García A, Merino-Barbancho B, Freites-Núñez D, Rodríguez-Rodríguez L, Menasalvas-Ruíz E, Rodríguez-González A, et al. From Web to RheumaLpack: creating a linguistic corpus for exploitation and knowledge discovery in rheumatology. Comput Biol Med. 2024;179:108920, http://dx.doi.org/10.1016/j.compbiomed.2024.108920.
19. Loureiro D, Barbieri F, Neves L, Anke LE, Camacho-Collados J. TimeLMs: diachronic language models from Twitter; 2022.
20. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg. 2014;12:1500–24, http://dx.doi.org/10.1016/j.ijsu.2014.07.014.
21. Gwinnutt JM, Norton S, Hyrich KL, Lunt M, Combe B, Rincheval N, et al. Influence of social support financial status, and lifestyle on the disparity between inflam-

mation and disability in rheumatoid arthritis. Arthritis Care Res (Hoboken). 2023;75:1026–35, http://dx.doi.org/10.1002/acr.24996.
22. Foufi V, Timakum T, Gaudet-Blavignac C, Lovis C, Song M. Mining of textual health information from Reddit: analysis of chronic diseases with extracted entities and their relations. J Med Internet Res. 2019;21, e12876.
23. Veselovsky V, Anderson A. Reddit in the time of COVID. Proceedings of the international AAAI conference on web and social media. 2023;Vol. 17:878–89, http://dx.doi.org/10.1609/icwsm.v17i1.22196.
24. Külekçioğlu S, Çetin A. Social media use in patients with fibromyalgia and its effect on symptom severity and sleep quality. Adv Rheumatol. 2021;61:51.
25. Benavent D, Madrid-García A. Large language models and rheumatology: are we there yet? Rheumatol Adv Pract. 2024;9:rkae119, http://dx.doi.org/10.1093/rap/rkae119.
26. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying reddit: a systematic overview of disciplines, approaches, methods, and ethics. Soc Media Soc. 2021;7:20563051211019004, http://dx.doi.org/10.1177/20563051211019004.
27. Fiesler C, Zimmer M, Proferes N, Gilbert S, Jones N. Remember the human: a systematic review of ethical considerations in Reddit research. Proc ACM Hum Comput Interact. 2024;8:1–33, http://dx.doi.org/10.1145/3633070.