Using Social Media Listening to Characterize the Flare Lexicon in Patients with Sjögren's Disease

Chiara Baldini [1], Maurice Flurie [2], Zachary Cline [2], Colton Flowers [2], Coralie Peter Bouillot [3], Linda J. Stone [4], Lauren Dougherty [2], Christopher DeFelice [2] and Maria Picone [2,*]

[1] Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Pisa, 56126 Pisa, Italy
[2] TREND Community, Philadelphia, PA 19102, USA; mflurie@trend.community (M.F.); zcline@trend.community (Z.C.); cflowers@trend.community (C.F.); ldougherty@trend.community (L.D.); chris@trend.community (C.D.)
[3] Sjögren Europe, 1196 Gland, Switzerland; coralie@sjogreneurope.org
[4] Sjögren's UK, Birmingham B31 3TE, UK; lj.stone@me.com
[*] Correspondence: maria@trend.community; Tel.: +1-215-370-4605

Abstract

Background/Objectives: Sjögren's disease (SjD) flares are incompletely understood. The patient perspective is critical to closing this gap. This retrospective social media listening (SML) study characterized the flare lexicon within the online Reddit SjD community using novel machine learning and natural language processing. Methods: Documents (posts/comments) were analyzed from the subreddit group "r/Sjogrens" (October 2012 to August 2023). Outcomes were as follows: (1) Frequency of documents mentioning flare, and contexts in which flare was mentioned; (2) clinical concepts associated with flare (analyzed using co-occurrence and pointwise mutual information [PMI]); (3) proportion of flare vs. non-flare documents relevant to SYMPTOMS or TESTING (compared using a two-proportion z-test); and (4) primary emotions mentioned in flare documents. Results: Of 59,266 documents with 5025 authors, flare was mentioned 3330 times (4.4% of documents from 19.1% of authors). Flare was discussed as a symptom (1423 instances), disease (13), or with no clinical category (1890). Flare-associated clinical concepts (co-occurrence > 100 and $PMI^2 > 3$) included SYMPTOMS (pain, fatigue, dryness of eye, xerostomia, arthralgia, stress) and BODY PARTS (eye, mouth, joints, whole body). More flare vs. non-flare documents mentioned a SYMPTOM, whereas fewer mentioned a TEST ($p < 0.001$ for both). Within flare documents, 36.5% expressed emotions, primarily fear (40.5% of primary emotions), happiness (17.8%), sadness (15.7%), and anger (15.5%). Conclusions: The SjD community discusses flare frequently and in context with symptoms, specifically pain, eye and mouth dryness, and fatigue. Flare conversations frequently involve negative emotions. Additional research is required to clarify the patient experience of flare, its clinical parameters, and implications.

Keywords: autoimmune disease; dry eye syndromes; symptom exacerbation; fatigue; pain; xerostomia

1. Introduction

Sjögren's disease (SjD) is a chronic autoimmune disorder characterized by lymphocytic infiltration of the exocrine glands, primarily lachrymal and salivary [1]. SjD onset peaks in the fifth decade of life, and is nine times more common in women than men [2]. As glandular secretory function progressively worsens over time, patients experience debilitating

xeropthalmia (dry eyes) and xerostomia (dry mouth) [1]. For women, vaginal dryness is also common [3]. Patients may eventually experience extraglandular, systemic manifestations involving the joints and other organs (i.e., skin, kidneys, lungs), as well as the gastrointestinal, reproductive, and central and peripheral nervous systems [1,3,4]. Prevalence estimates vary, but a 2015 international meta-analysis found 43 cases per 100,000 individuals [2], and recent data suggest increased SjD incidence subsequent to the SARS-CoV-2 pandemic [5]. The hallmark symptoms of SjD are dry eyes and mouth, fatigue, arthralgia (joint pain), and myalgia (tendon and muscle pain) [1,6,7]. Anxiety and depression are also common [8]. SjD is notable for its heterogeneous profile [9,10], with symptoms varying substantially from patient to patient, day to day, and within each day [7,9–11]; however, compared to the population as a whole, patients with SjD consistently experience decreased health-related quality of life (HRQoL) and increased disability [6,7,12].

Importantly, many aspects of the patient experience of SjD remain incompletely understood [7,11,13]. Although medical treatment guidelines suggest that SjD progresses from dryness or sicca-related symptoms to systemic manifestations that include increased joint pain and fatigue and abnormal hematologic parameters [14–16], real-world research shows discordance between subjective patient symptoms and objective laboratory results [13]. This creates the potential for suboptimal care, and also suggests the existence of distinct endotypic profiles and unexplained variations in symptom sensitivity [10,13]. Relevant to this, Sjögren's flares, or temporary, acute exacerbations of disease activity, are a poorly understood aspect of patients' experience of SjD [7,17]. Flares are common in relapsing–remitting autoimmune diseases, including rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) [17], and have been well-studied in these conditions, for example, by the Outcome Measures in Rheumatology Clinical Trials (OMERACT) RA Flare Definition Working Group [18–20]. In contrast, SjD is a slow-progressing disease [21], and its flares are quite unusual and not well-studied or defined [11,17]. However, survey evidence, social media-listening research, patient education websites, and reports from clinical experts all indicate that some patients with SjD report flare activity [7,11,17,22–25]. Nonetheless, epidemiologic data quantifying the frequency of flare in SjD are extremely limited, with one single-center case series (1997–2001, N = 59) reporting a crude prevalence of 15% [17].

Although there is no consensus definition of flare in SjD, existing patient descriptions of SjD flares describe an unpredictable and sudden onset of one or more symptoms that are disruptive and worsen over several days before resolving [7,11]. These symptoms include anxiety and depression, eye irritation/vision issues, fatigue/exhaustion, forgetfulness and brain fog, global pain, and tingling and numbness [7,11]. Based on limited evidence, no direct relationship has been identified between flare as defined by patients and changes in biomarker tests (e.g., erythrocyte sedimentation rate, C-reactive protein, immunoglobulin, or complement C3 or C4 levels) [17]. Additionally, flares are rarely discussed in the SjD clinical literature [7,11,17], no official definition of SjD flare exists [11], and the most commonly used assessment indices may not capture all flare symptoms reported by patients [11,26,27]. Specifically, research evaluating key disease indices, including the clinician-focused European Alliance of Associations for Rheumatology (EULAR) Sjögren's Syndrome Disease Activity Index [ESSDAI] and the patient-focused EULAR Sjögren's Syndrome Patient Reported Index [ESSPRI], has found limited responsiveness with existing patient scores (i.e., the ability to detect small, clinically important changes) and a low correlation between the ESSDAI and ESSPRI instruments [28]. Such discrepancies pose a dilemma for clinicians [11,13,29]. Lacking a shared understanding of flare between patients and physicians (as well as caregivers), there is the potential for under- or over-treatment, and/or misinterpretation of treatment response [29].

Social media listening (SML), or the extraction and analysis of information from social media channels, is an emerging approach used to research real-world patient perspectives on topics like disease-related symptomology, diagnosis, medical treatment, and HRQoL [7,30,31]. SML can also capture clinically meaningful patient experience data that would otherwise remain overlooked [31]. Because patients on social media sites openly exchange health-related information, as well as provide and receive social support, these platforms have emerged as a natural setting to evaluate patient-reported data [31]. Considerable advancements in analytic approaches, including the utilization of natural language processing (NLP) and machine learning, have substantially enhanced the precision of this research [31]. SML research is particularly useful in conditions like SjD, where patient–clinician discordance exists with respect to symptom reporting [13]. The importance of flare in SjD was identified as a salient patient issue in a prior TREND communications analysis evaluating sleep and mental health across rheumatic diseases, including SjD [32]. The objective of this retrospective, non-interventional study was to use SML to develop a flare lexicon for SjD. Specifically, this analysis sought to gauge the extent to which the SjD community talks about flares; examine the contexts in which flares are mentioned; identify overarching clinical concepts associated with flares; and describe the emotional impact of flares on the SjD community. As such, this lexicon would capture patients' experience of flare based on their stated terms regarding SjD-related flare and its emotional and mental impact.

2. Materials and Methods

2.1. Study Design and Data Source

The study was conducted by TREND Community, a Certified B Corporation that uses digital health analytics to conduct research using available information on social media in order to benefit rare disease communities and assist in clinical research related to patient-defined efficacy and improved patient quality of life. Using the Reddit application programming interface, TREND Community downloaded documents (posts or comments) posted between 16 October 2012, and 7 August 2023, from the subreddit group, "r/Sjogrens." This group self-describes as "a science-positive, evidence-based community for people to discuss Sjögren's syndrome."

The study design was evaluated by an independent, commercial institutional review board (IRB) (Western Institutional Review Board), which determined that this research was exempt from IRB approval under Common Rule parts 45 CFR x 46.104(d)(2) and 45 CFR x 46.104(d)(4). Part 45 CFR x 46.104(d)(2) exempts research that includes only interactions involving educational tests, survey or interview procedures, or observations of public behavior, as long as this information is recorded in such a manner that the identity of study subjects cannot readily be ascertained. In this case, the authors of r/Sjogrens are voluntary, anonymous participants in a publicly visible discussion board. Part 45 CFR x 46.104(d)(4) exempts research involving the use of secondary, identifiable private information and/or biospecimens in situations where this information is recorded by the investigator in such a manner that the identity of the subjects cannot be readily ascertained, the investigator does not contact the subjects, and the investigator will not re-identify the subjects. Additionally, because Reddit does not collect or publicly release demographic information, such data were not available for this analysis (except in cases where authors voluntarily disclosed their demographics, in which case precautions were taken to protect individual data).

2.2. Outcomes

This study aimed to construct a flare lexicon by identifying and analyzing the following within the dataset: (1) Flare recognition, including frequency of flare documents and how the term "flare" was discussed; (2) the overarching clinical concepts associated with flare (i.e., the clinical concepts mentioned in the same context as flare); (3) the proportion of flare vs. non-flare documents mentioning terms relevant to the clinical categories of SYMPTOM and TEST; and (4) the most common emotions mentioned in flare documents.

2.3. Data Processing

The subreddit document dataset was analyzed using TREND Community's proprietary NLP engine, the Clinical Span Recognition and Linking (CSRL) system. CSRL uses machine learning and pattern-matching techniques to extract clinical terms of interest related to SjD (e.g., "flare", "fatigue", "dry mouth", "doctor") from social media documents and links these terms to overarching concepts. To do this, the engine recognizes and extracts clinical terms related to each of 24 distinct, broad clinical categories, including SYMPTOM, DISEASE, BODY PART, DRUG, TEST, MEDICAL SPECIALTY, and EMOTION. Note that CSRL captures all iterations of clinical terms (e.g., "flare", "flare-up", "flaring") and categorizes these terms not only based on the term itself, but also the context in which it is used (Figure A1). Next, the extracted terms are linked to specific clinical concepts from an existing lexical knowledge base, called TRENDnet. TRENDnet was built using the SNOMED [33] and RxNorm [34] vocabularies within the Unified Medical Language System (UMLS) [35], a National Institutes of Health-sponsored collection of vocabularies used to organize and understand biomedical terminology. RxNorm and SNOMED provide coverage of a vast number of internationally recognized medical terms. To find the closest (i.e., synonymous) concept for each term in TRENDnet, terms were linked to concepts utilizing cosine similarity (a measure of distance between data points in a shared vector space) between term frequency–inverse document frequency-encodings based on character 3-g.

To explore the potential emotional and mental impacts of flare experiences, the distribution of emotions mentioned in flare documents was examined. As noted above, EMOTION was 1 of the 24 categories used for term extraction. Extracted emotions were classified as related to 1 of the 6 primary emotions (anger, disgust, fear, happiness, sadness, and surprise); these primary emotions were initially identified by Ekman [36], and subsequently supported as semantically distinct by Bann and Bryson [37]. Before being categorized as primary emotions (e.g., *fear*), EMOTION terms were binned into specific secondary emotions (e.g., "worry" → *worry*; "nervous" → *nervousness*). Importantly, although this analysis was conducted to classify language related to mental health, it was not intended to diagnose any mental health conditions.

2.4. Statistical Analysis

To put the prevalence of flare into perspective, the proportion of documents mentioning flare was compared to that of each SYMPTOM concept. To understand how "flare" was discussed in conversation, the instances in which CSRL recognized ≥2 "flare" mentions under a single clinical category were tabulated. To measure the strength of association between flare and other clinical concepts, co-occurrence and pointwise mutual information (PMI) were used. Co-occurrence refers to the number of documents in which two concepts occurred together (e.g., "flare" and "dryness of eye"). PMI is a measure of association utilized in information theory and NLP to determine the relative probability of 2 concepts occurring together compared to the probability of the 2 concepts occurring independently [38]. Because PMI tends to give high scores to rare events [39], a variant of PMI, namely PMI^2, was used for this study. With conventional PMI, if a concept occurred only once across all

documents, but it occurred with "flare," it would receive a high PMI score. PMI^2 aims to correct for this bias by boosting scores for high-frequency concept pairs.

The frequencies of broad clinical categories (SYMPTOM, TESTING) mentioned in flare vs. non-flare documents were compared using a two-proportion z-test. To ensure consistency across these comparisons, document length was controlled for. This was done because longer documents are naturally more likely to mention any clinical category, leaving open the possibility of document length confounding observed differences between flare and non-flare documents. Document length was controlled for by stratifying the dataset by document length (i.e., number of characters) into the following bins: (0–100), (101–200), (201–400), (401–800), and ($\geq$801). Each stratum was analyzed separately. The null hypothesis was that the true proportions of flare documents and non-flare documents with at least one mention of that clinical category type were equivalent. For statistical testing, the two-sided alternative hypothesis was that they were different, and the alpha level, indicating statistical significance, was set at 0.05. For statistically significant results, Cohen's h was used to determine effect size. All analyses were conducted using Python (Version 3.10), specifically the NumPy (version 1.23.5) and statsmodels (version 0.13.5) packages.

2.5. Patient and Public Involvement

This work builds upon prior patient-led work to define Sjögren's flare, which culminated in a poster presentation at the EULAR 2023 Annual Meeting [11]. This presentation sparked a workshop and panel discussion conducted in collaboration with Sjögren Europe (a federation of national associations representing patients with SjD), which ultimately led to the genesis of this paper. These discussions included three authors of this paper (CB, CPB, LJS), two of whom are also SjD patients. During these discussions, patient participants voiced their experiences with flare, which helped to inform the patient-focused research objective and outcomes of this work. Sjögren Europe will also help to disseminate these study results to relevant, wider patient communities."

3. Results

On the date of document download (7 August 2023), the subreddit group had 4193 posts and 55,073 comments from 5025 authors, for a total of 59,266 documents. Authors were primarily from the US, but also Argentina, Canada, England, France, Germany, India, Ireland, the Netherlands, Norway, Singapore, and Taiwan.

"Flare" was mentioned 3330 times in 2598/59,266 (4.4%) documents from 962/5025 (19.1%) authors in the r/Sjogrens subreddit community. Flare was discussed as a SYMPTOM in 1423 instances or a DISEASE in 13 instances, although most flare mentions did not adhere to a specific clinical category (1890 instances) (Figure A2). The only symptoms more common than flare were pain (7.4% of all documents), dryness of eye (6.5%), and xerostomia (5.8%) (Figure 1). Fatigue (4.3%) was mentioned at a similar rate as flare (4.4%). Symptoms less common than flare, in descending order, were arthralgia, swelling, stress, neuropathy, headache, feeling tired, pruritus, and dry skin (all <3% of documents).

Figure 2 shows the clinical concepts that co-occurred most frequently with flare within flare documents (i.e., seen in the same context as flare), as well as the uniqueness of these in relation to flare documents using PMI^2. The concepts associated with flare (co-occurrence > 100 and PMI^2 > 3) included SYMPTOMS (seven concepts: symptoms, pain, fatigue, dryness of eye, xerostomia, arthralgia, and stress); BODY PARTS (four concepts: eye, mouth, joints, and whole body); MEDICATIONS (three concepts: pharmaceutical preparations, prednisone (the same class of medication as prednisolone) and Plaquenil (the brand name for hydroxychloroquine)); and MEDICAL PROFESSIONALS (two concepts: rheumatologist and physicians).
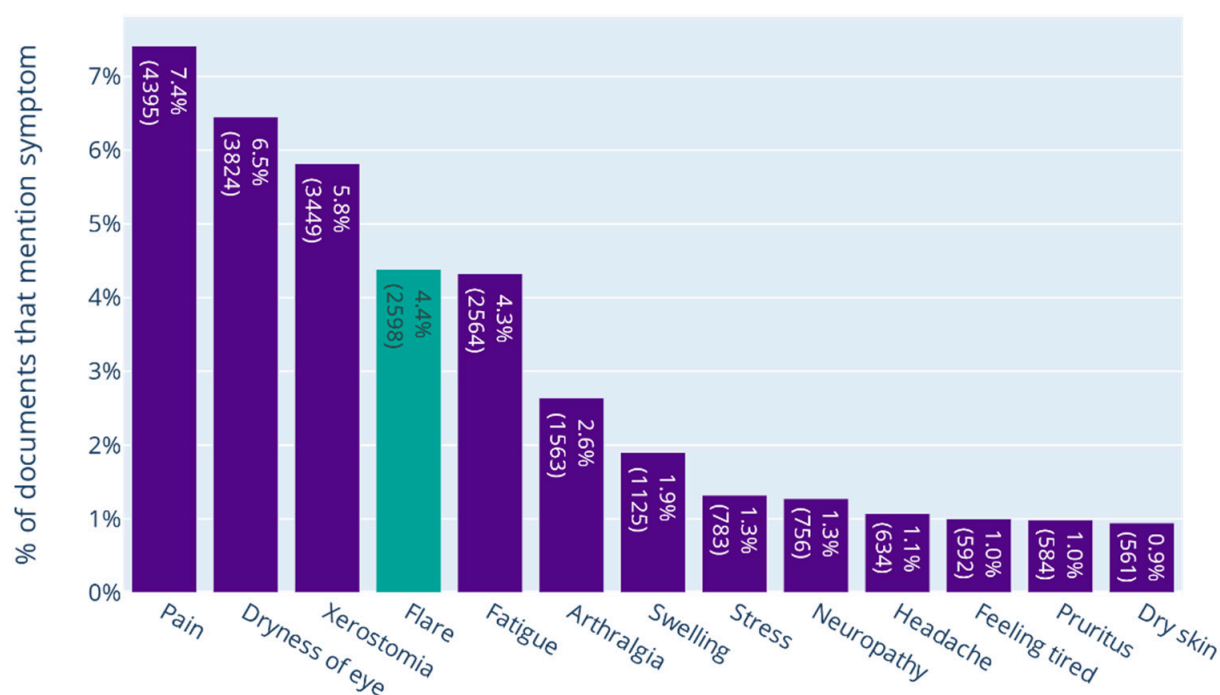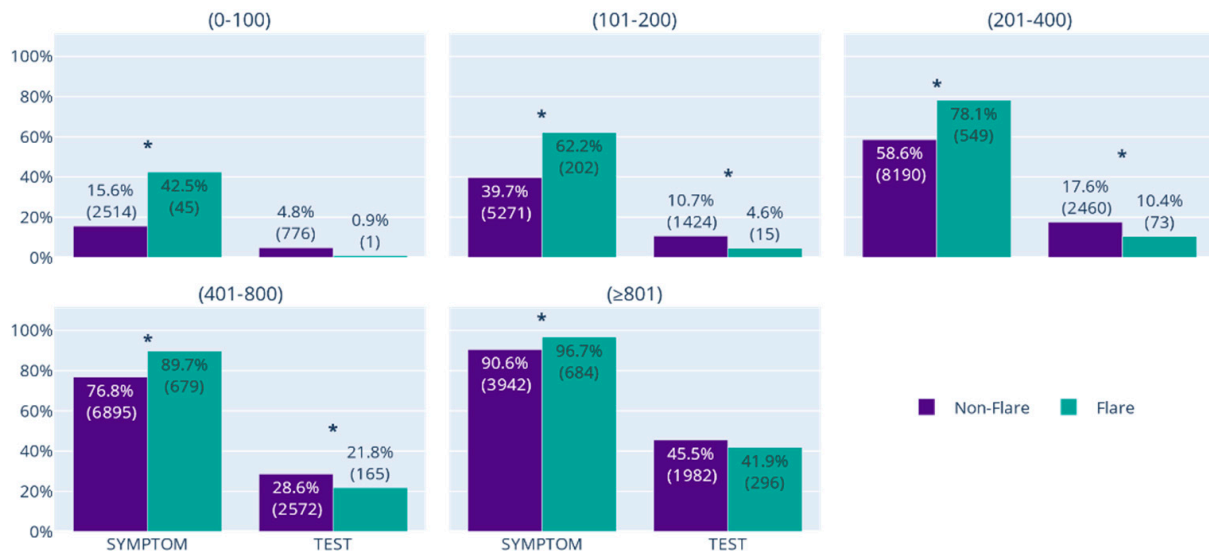
Figure 1. The Proportion of Documents Mentioning Symptoms in the Dataset. This bar chart represents the distribution of the 12 most common SjD symptoms recognized by CSRL in the r/Sjogrens subreddit community.
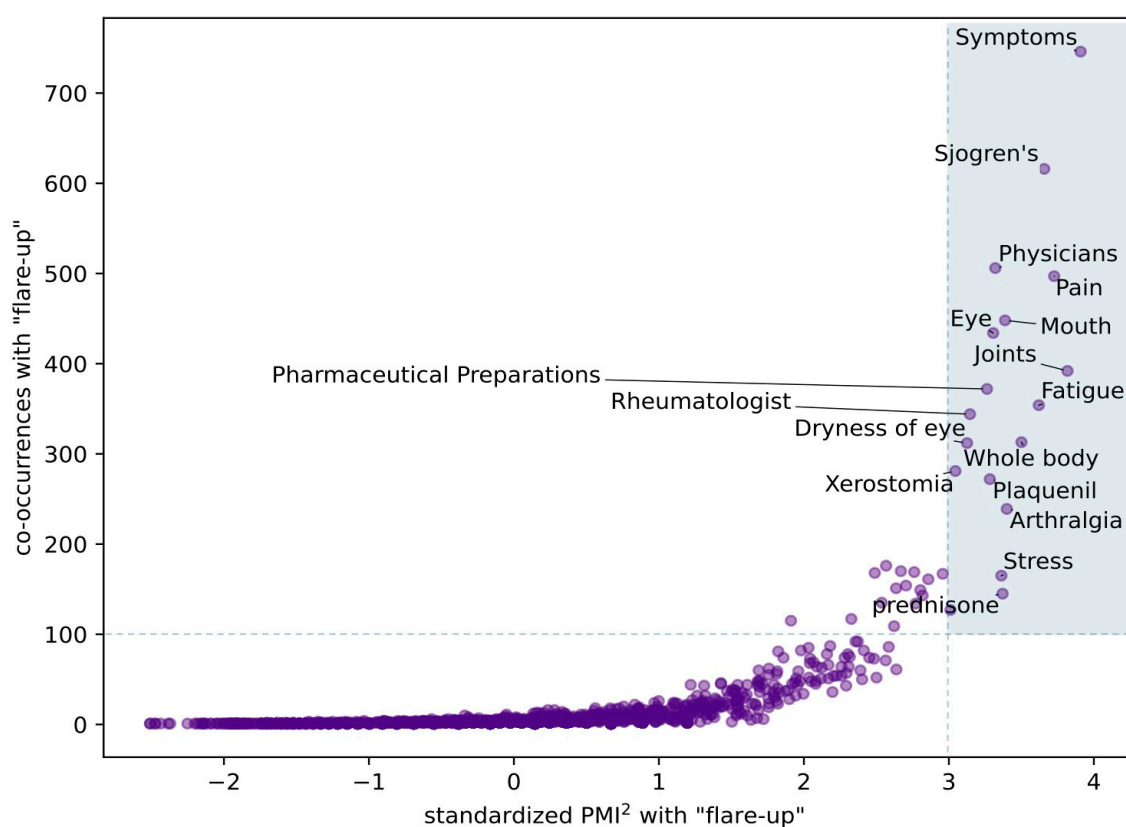
Figure 2. Flare Clinical Associations. The shaded area represents clinical concepts that co-occurred most frequently within flare documents with the highest degree of uniqueness (co-occurrence > 100 and PMI^2 > 3).

Figure 3 shows the proportion of flare and non-flare documents that mention any terms categorized as SYMPTOM or TEST, stratified by document length. After stratifying

for document length, a higher proportion of flare vs. non-flare documents mentioned a SYMPTOM, whereas a lower proportion mentioned a TEST. This was statistically significant for SYMPTOM in all document lengths ($p < 0.001$ for all); and for TEST in each stratum ($p < 0.001$) except the shortest ($\leq$100 characters; $p = 0.06$) and the longest ($\geq$801 characters; $p = 0.07$) documents (Table 1). The effect size for the difference in SYMPTOM mentions for flare vs. non-flare ranged from small to medium, while the effect size for TEST mentions was small.


Figure 3. Proportion of Flare and Non-Flare Documents That Mention SYMPTOM or TEST, Stratified by Document Length. After stratifying for document length, a higher proportion of flare vs. non-flare documents mentioned a SYMPTOM, whereas a lower proportion mentioned a TEST. This was statistically significant for SYMPTOM in all document lengths ($p < 0.000001$ for all); and for TEST in each stratum ($p < 0.001$) except the shortest ($\leq$100 characters; $p = 0.06$) and the longest ($\geq$801 characters; $p = 0.07$) documents. The asterisks indicate which differences in proportions between the Flare and Non-Flare groups are statistically significant.

Table 1. Results of Z-Tests and Cohen's h for Comparison of Flare vs. Non-Flare Documents Mentioning SYMPTOM or TEST.

	SYMPTOM			**TEST**		
Stratum	**Z**	***p***	**Cohen's h**	**Z**	***p***	**Cohen's h**
(0–100)	7.56	<0.000001	0.607	−1.86	0.062689	NA
(101–200)	8.14	<0.000001	0.452	−3.54	0.000397	−0.235
(201–400)	10.25	<0.000001	0.423	−4.95	0.000001	−0.210
(401–800)	8.20	<0.000001	0.352	−4.03	0.000057	−0.158
(≥801)	5.44	<0.000001	0.261	−1.82	0.068489	NA

Figure 4 displays the distribution of primary emotions and their relative secondary emotions in flare documents. A total of 351/962 (36.5%) flare authors in the r/Sjogrens subreddit community expressed an emotion across 546/2598 (21.0%) flare documents. Most emotions expressed were negative, such as fear (40.5% of primary emotions expressed), sadness (15.7%), anger (15.5%), and disgust (7.3%). Positive emotions were also indicated, such as happiness (17.8%).
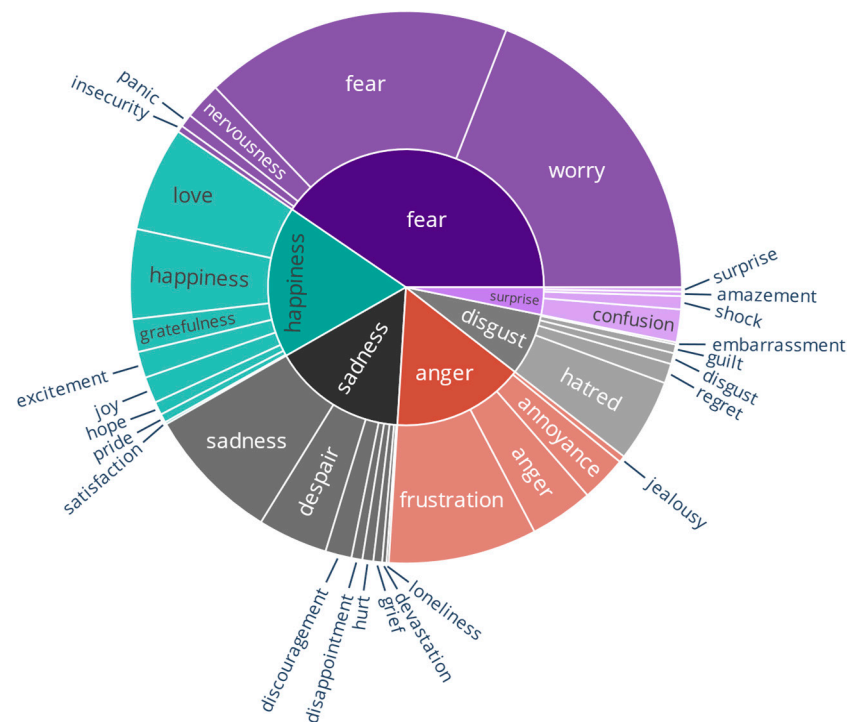
Figure 4. Distribution of Emotions in Documents That Mention Flare. Emotion terms were first binned into specific secondary emotions, represented by the outer ring of the circle, before being categorized into 1 of 6 primary emotions, shown in the inner ring.

4. Discussion

This SML study shows that the SjD community is discussing flare frequently. In the r/Sjogrens subreddit, flare was among the top five symptoms mentioned, with more than one in four conversations containing mention of flare. The heterogeneous, systemic nature of SjD was also visible in these results, with some patients reporting swelling, stress, neuropathy, headache, fatigue, pruritus, and dry skin in addition to flare and other hallmark symptoms of the disease. This study also shows that, when the SjD community discusses flare, they do so in the context of symptoms, rather than testing. Most commonly, when discussing flare symptoms, patients mentioned pain, dryness of the eyes and mouth, and fatigue. Other associated concepts included stress, as well as the clinical concept of the whole body. Flare also has an emotional impact on patients, wherein conversations about flare frequently mention fear. Collectively, these clinical concepts and emotional impacts may substantially affect a patient's HRQoL, including daily functioning and social and psychological well-being [7].

Beyond SjD, the methodological framework used in this study—combining SML with NLP and machine learning—may have broader applicability. Autoimmune conditions such as SLE, RA, and fibromyalgia also involve unpredictable symptom fluctuations [20,40–42] and patient–clinician discordance in the perception of disease activity [43–45]. Applying similar approaches in these contexts could facilitate the development of condition-specific lexicons that capture the lived experience of flare in patients' own words. In the longer term, such insights may also contribute to personalized treatment strategies by providing clinicians and researchers with complementary, patient-reported data that can inform decisions about disease monitoring and therapeutic intervention.

This SML study confirms prior research regarding the frequency and characteristics of SjD flare [7,11], providing valuable corroboration on topics that have been historically underexamined [7,11,17]. Over one decade of follow-up, approximately 20% of patients discussed SjD flare, a proportion that aligns with the 15% prevalence observed in a 2005

multi-year case series [17]. This study also confirms the varied and symptomatic patient experience of flare, as identified previously in only two studies [7,11]. In another SML study, conducted from April 2019 to March 2021, of relevant SjD content from Twitter, Reddit, and other publicly accessible patient forums, the symptom domains most strongly associated with flare were (in order of frequency) neuropathy; cognitive fatigue and brain fog; eye irritation; other vision issues; fatigue and exhaustion; nausea and digestive issues; emotional struggle; joint, bone, and muscle pain; and dry mouth. Patients also expressed that flare symptoms severely impacted their ability to think and move [7]. More recently, an effort to define patient flares was initiated by seven members of the Patient Advisory Group (PAG) associated with the European NECESSITY project. The symptoms that PAG members discussed as being most prominent with flare included dryness, fatigue, morning stiffness, global pain, forgetfulness and brain fog, depression and anxiety, and sleep disorders [11]. Captured in all of this research, including the present study, is the finding that patients with SjD experience a wide range of symptoms during Sjögren's flare episodes.

The failure to recognize disease activity in SjD, including flare, may cause failure to treat and therefore the potential for SjD-related damage to accrue [46]. This situation has been observed in RA, where no valid mechanistic biomarkers exist to identify flares [20], although flare frequency and duration are associated with radiographic deterioration, increased cardiovascular disease risk, and reduced physical activity [20,47–49]. The potential to overlook disease activity is especially likely in SjD because subjective manifestations of SjD (pain, fatigue, etc) are difficult to measure objectively, and patient interoception and perception may modify their experience of symptom severity. This was observed in the UK Primary Sjögren's Syndrome Registry (UKPSSR; N = 688), wherein Bezzina et al. 2017 found that subjective symptoms of pain and fatigue were strong predictors of patient sensitivity to SjD-related ocular and oral dryness [13]. Using the same registry, Tarn et al. 2019 identified four distinct endotypic profiles of SjD based on patient-reported symptom severity, with corresponding clinical and biological profiles, as well as therapeutic responses [10].

Understanding patient-reported, subjective measures of flare may be critical to preventing damage accrual. Therefore, there is a need to further develop the definition of "disease activity" to encompass both objective physician evaluation (e.g., objective tests, biomarkers, evaluation of organ damage) and the patient's subjective experience, including flare. While the ESSPRI instrument seeks to incorporate the patient perspective, it may fall short of identifying the true complexity of disease burden once patient experience is accounted for. Following on this, and with findings from the current study showing that flares are a salient feature of the patient experience, additional research is warranted to comprehensively define and clarify the clinical expressions of SjD flare, including a responsive patient-reported outcomes measure. A clear definition of flare would also help patients, caregivers, and physicians determine when clinical intervention is warranted. This is especially important for patients who have difficulty accessing care and those who may not recognize flare until after the fact (for example, if symptom[s] come on gradually and/or resolve quickly). Finally, defining flare is also highly relevant in the context of new drug development, as clinical trial outcomes that do not consider flare may neglect to detect elements of drug effectiveness (or lack thereof) [7,29].

To accomplish this, the SjD community may look to a similar effort undertaken in the RA disease space. Lacking a standard definition of flare to measure its intensity and impact, the OMERACT RA Flare Definition Working Group set out in 2009 to develop such a tool [18,20]. This effort culminated in the RA Flare Questionnaire (RA-FQ), a five-item patient-reported outcome with potential use in both daily practice and clinical trials [19,50]. Patients with RA who completed the RA-FQ have reported that the questionnaire has

potential applicability in enhancing patient-provider communication, as well as monitoring of day-to-day status and self-management [50].

Strengths and Limitations

This research is timely as it corresponds with US Food and Drug Administration guidance encouraging the consideration of patient perspectives in clinical research [51], recognizing the role of patient-generated data in the development of real-world evidence [52], and recommending social media analysis as a specific strategy for qualitative and quantitative data collection [53]. Furthermore, user information is private, which may allow authors to post more candidly about sensitive topics, and with fewer privacy barriers to participation (e.g., no risk of being recognized).

With respect to limitations, baseline demographic information, including patient diagnosis, was not captured or controlled for because such information was not consistently available. While some authors explicitly self-identified as having SjD, the diagnosis could not be verified for each author because the dataset was derived from a public online forum. Additionally, this study was not able to evaluate the duration of flares, nor did it capture or control for documents that mentioned patient comorbidities. The former warrants additional research because existing patient research indicates that temporal patterns are an intrinsic quality of flare [7,11], while the latter creates the potential for confounding, especially since some patients with SjD also have diseases with recognized flare patterns (e.g., SLE, RA, and fibromyalgia) [1,20,40–42]. Concepts linked to flare were limited to those that exist in the lexical knowledge base, in this case, TrendNet. Although the emotional state of patients in flare documents was captured, these are not indicative of mental health diagnoses, and caution is warranted in interpreting these results. Finally, these data are primarily from US authors, and selection bias may exist [54] because it is not known whether this Subreddit patient population is representative of the general population (research suggests demographic disparities in social media data) [55].

5. Conclusions

This SML analysis confirms that flares are a salient and consistent feature of the SjD patient experience. The SjD community is discussing flares frequently and in context with their symptoms, specifically pain, dryness of the eyes and mouth, and fatigue. The emotional impact of flares is also evident, wherein flare conversations frequently mention fear and other negative emotions. Additional research is required to more clearly understand the patient experience of Sjögren's flare, define the clinical parameters of flare, and clarify the clinical implications of such disease activity.

Author Contributions: Conceptualization, C.B., M.F., Z.C., C.P.B., L.J.S., C.D. and M.P.; methodology, C.B., M.F., Z.C., C.F., C.D. and M.P.; software, Z.C., C.F. and C.D.; validation, M.F. and Z.C.; formal analysis, M.F., Z.C., C.P.B. and L.J.S.; investigation, M.F., Z.C. and M.P.; resources, L.D., C.F. and M.P.; data curation, M.F., Z.C., C.P.B., L.J.S., C.D. and M.P.; writing—original draft preparation, C.B., M.F., Z.C. and M.P.; writing—review and editing, C.B., M.F., Z.C., C.P.B., L.J.S., L.D., C.D. and M.P.; visualization, M.F. and Z.C.; supervision, C.B., M.F., L.D., C.D. and M.P.; project administration, L.D. and M.P.; funding acquisition, L.D. and M.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study design was evaluated by an independent, commercial institutional review board (IRB) (Western Institutional Review Board), which determined that this research was exempt from IRB approval under Common Rule parts 45 CFR x 46.104(d)(2) and 45 CFR x 46.104(d)(4). Part 45 CFR x 46.104(d)(2) exempts research that includes only interactions involving educational tests, survey or interview procedures, or observations of public behavior, as

long as this information is recorded in such a manner that the identity of study subjects cannot readily be ascertained. In this case, the authors of r/Sjogrens are voluntary, anonymous participants in a publicly visible discussion board. Part 45 CFR x 46.104(d)(4) exempts research involving the use of secondary, identifiable private information and/or biospecimens in situations where this information is recorded by the investigator in such a manner that the identity of the subjects cannot be readily ascertained, the investigator does not contact the subjects, and the investigator will not re-identify the subjects. Additionally, because Reddit does not collect or publicly release demographic information, such data were not available for this analysis (except in cases where authors voluntarily disclosed their demographics, in which case precautions were taken to protect individual data).

Abbreviations

The following abbreviations are used in this manuscript:

CSRL	Clinical Span Recognition and Linking
ESSDAI	EULAR Sjögren's Syndrome Disease Activity Index
ESSPRI	EULAR Sjögren's Syndrome Patient Reported Index
EULAR	European Alliance of Associations for Rheumatology
HRQoL	Health-related quality of life
IRB	Institutional review board
NLP	Natural language processing
OMERACT	Outcome Measures in Rheumatology Clinical Trials
PAG	Patient Advisory Group
PMI	Pointwise mutual information
RA	Rheumatoid arthritis
RA-FQ	RA Flare Questionnaire
SjD	Sjögren's disease
SLE	Systemic lupus erythematosus
SML	Social media listening
UKPSSR	UK Primary Sjögren's Syndrome Registry
UMLS	Unified Medical Language System

Appendix A
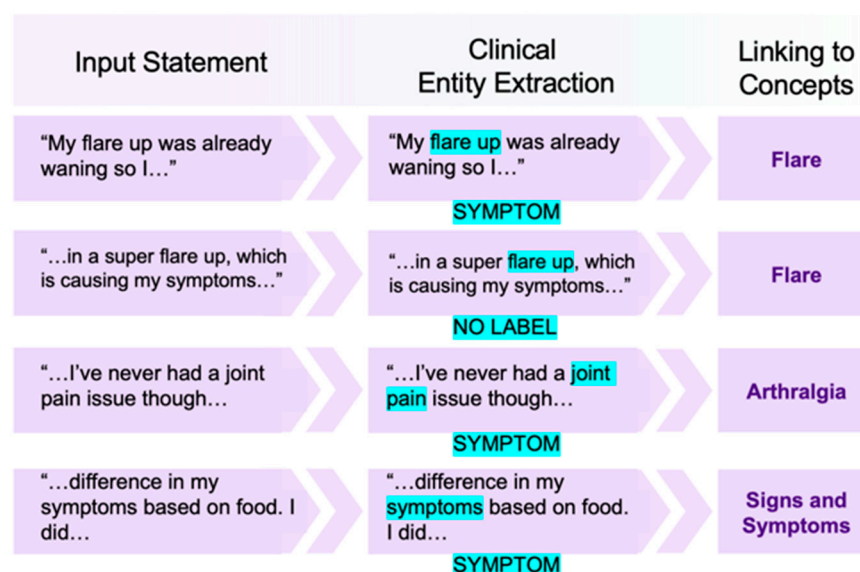
Figure A1. Example of Clinical Entity Recognition and Linking (CSRL). Note that CSRL categorizes clinical terms not only based on the term itself, but also the context in which it is used. Words in upper case represent broad clinical categories under which clinical terms were recognized and extracted.
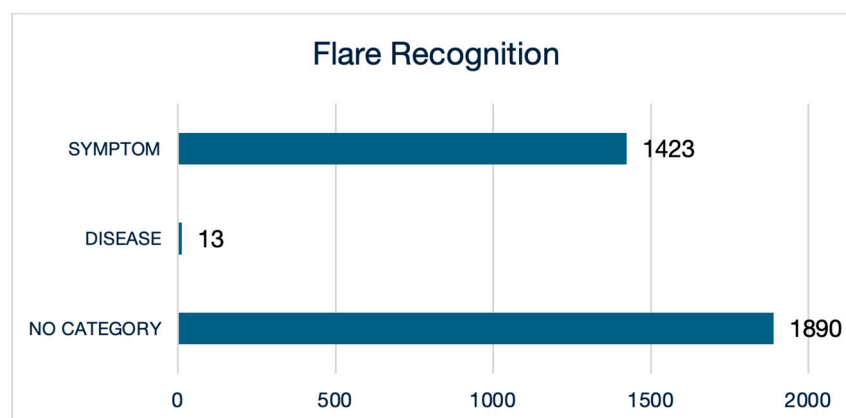
Figure A2. Clinical Categories in Which CSRL Recognized ≥ 2 Instances of "Flare" Mentions. CSRL recognized the term flare used as a SYMPTOM in 1423 mentions or a DISEASE in 13 mentions, although most 1890 flare mentions did not adhere to a specific clinical category.

References

1. Vitali, C.; Minniti, A.; Pignataro, F.; Maglione, W.; Del Papa, N. Management of Sjogren's syndrome: Present issues and future perspectives. *Front. Med.* **2021**, *8*, 676885. [CrossRef]
2. Qin, B.; Wang, J.; Yang, Z.; Yang, M.; Ma, N.; Huang, F.; Zhong, R. Epidemiology of primary Sjogren's syndrome: A systematic review and meta-analysis. *Ann. Rheum. Dis.* **2015**, *74*, 1983–1989. [CrossRef]
3. van Nimwegen, J.F.; van der Tuuk, K.; Liefers, S.C.; Verstappen, G.M.; Visser, A.; Wijnsma, R.F.; Vissink, A.; Hollema, H.; Mourits, M.J.E.; Bootsma, H.; et al. Vaginal dryness in primary Sjogren's syndrome: A histopathological case-control study. *Rheumatology* **2020**, *59*, 2806–2815. [CrossRef]
4. Kim-Lee, C.; Suresh, L.; Ambrus, J.L., Jr. Gastrointestinal disease in Sjogren's syndrome: Related to food hypersensitivities. *Springerplus* **2015**, *4*, 766. [CrossRef]
5. Martelli Junior, H.; Gueiros, L.A.; de Lucena, E.G.; Coletta, R.D. Increase in the number of Sjogren's syndrome cases in Brazil in the COVID-19 era. *Oral Dis.* **2022**, *28* (Suppl. S2), 2588–2590. [CrossRef]
6. Meijer, J.M.; Meiners, P.M.; Huddleston Slater, J.J.; Spijkervet, F.K.; Kallenberg, C.G.; Vissink, A.; Bootsma, H. Health-related quality of life, employment and disability in patients with Sjogren's syndrome. *Rheumatology* **2009**, *48*, 1077–1082. [CrossRef] [PubMed]

7. Perella, C.; Steenackers, M.; Robbins, B.; Stone, L.; Gervais, R.; Schmidt, T.; Goswami, P. Patient experience of Sjogren's disease and its multifaceted impact on patients' lives. *Rheumatol. Ther.* **2023**, *10*, 601–614. [CrossRef]

8. Valtysdottir, S.T.; Gudbjornsson, B.; Lindqvist, U.; Hallgren, R.; Hetta, J. Anxiety and depression in patients with primary Sjogren's syndrome. *J. Rheumatol.* **2000**, *27*, 165–169. [PubMed]

9. Gairy, K.; Ruark, K.; Sinclair, S.M.; Brandwood, H.; Nelsen, L. An innovative online qualitative study to explore the symptom experience of patients with primary Sjogren's syndrome. *Rheumatol. Ther.* **2020**, *7*, 601–615. [CrossRef] [PubMed]

10. Tarn, J.R.; Howard-Tripp, N.; Lendrem, D.W.; Mariette, X.; Saraux, A.; Devauchelle-Pensec, V.; Seror, R.; Skelton, A.J.; James, K.; McMeekin, P.; et al. Symptom-based stratification of patients with primary Sjögren's syndrome: Multi-dimensional character-isation of international observational cohorts and reanalyses of randomised clinical trials. *Lancet Rheumatol.* **2019**, *1*, e85–e94. [CrossRef] [PubMed]

11. Bouillot, C.; Hammitt, K.M.; Lindland, A.H.; Oosterbaan, M.; Pincemin, M.; Stone, L. AB1732-PARE "Flare, did you say flare?" Flares in Sjögren's disease: The patient perspective. *Ann. Rheum. Dis.* **2023**, *82*, 2103. [CrossRef]

12. Cornec, D.; Devauchelle-Pensec, V.; Mariette, X.; Jousse-Joulin, S.; Berthelot, J.M.; Perdriger, A.; Puechal, X.; Le Guern, V.; Sibilia, J.; Gottenberg, J.E.; et al. Severe health-related quality of life impairment in active primary Sjogren's syndrome and patient-reported outcomes: Data from a large therapeutic trial. *Arthritis Care Res.* **2017**, *69*, 528–535. [CrossRef]

13. Bezzina, O.M.; Gallagher, P.; Mitchell, S.; Bowman, S.J.; Griffiths, B.; Hindmarsh, V.; Hargreaves, B.; Price, E.J.; Pease, C.T.; Emery, P.; et al. Subjective and objective measures of dryness symptoms in primary Sjögren's syndrome: Capturing the discrepancy. *Arthritis Care Res.* **2017**, *69*, 1714–1723. [CrossRef] [PubMed]

14. Baer, A.N.; Vivino, F.B. Overview of the Management and Prognosis of SjöGren's Disease. 2023. Available online: https://www.uptodate.com/contents/overview-of-the-management-and-prognosis-of-sjogrens-disease?search=overview%20of%20the%20management%20and%20prognosis%20of%20sjogrens%20disease&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1 (accessed on 19 September 2025).

15. Price, E.J.; Rauz, S.; Tappuni, A.R.; Sutcliffe, N.; Hackett, K.L.; Barone, F.; Granata, G.; Ng, W.F.; Fisher, B.A.; Bombardieri, M.; et al. The British Society for Rheumatology guideline for the management of adults with primary Sjogren's Syndrome. *Rheumatology* **2017**, *56*, e24–e48. [CrossRef]

16. Ramos-Casals, M.; Brito-Zeron, P.; Bombardieri, S.; Bootsma, H.; De Vita, S.; Dorner, T.; Fisher, B.A.; Gottenberg, J.E.; Hernandez-Molina, G.; Kocher, A.; et al. EULAR recommendations for the management of Sjogren's syndrome with topical and systemic therapies. *Ann. Rheum. Dis.* **2020**, *79*, 3–18. [CrossRef]

17. Stevens, R.J.; Hamburger, J.; Ainsworth, J.R.; Holmes, G.; Bowman, S.J. Flares of systemic disease in primary Sjogren's syndrome. *Rheumatology* **2005**, *44*, 402–403. [CrossRef]

18. Alten, R.; Pohl, C.; Choy, E.H.; Christensen, R.; Furst, D.E.; Hewlett, S.E.; Leong, A.; May, J.E.; Sanderson, T.C.; Strand, V.; et al. Developing a construct to evaluate flares in rheumatoid arthritis: A conceptual report of the OMERACT RA Flare Definition Working Group. *J. Rheumatol.* **2011**, *38*, 1745–1750. [CrossRef]

19. Bartlett, S.J.; Barbic, S.P.; Bykerk, V.P.; Choy, E.H.; Alten, R.; Christensen, R.; den Broeder, A.; Fautrel, B.; Furst, D.E.; Guillemin, F.; et al. Content and construct validity, reliability, and responsiveness of the Rheumatoid Arthritis Flare Questionnaire: OMERACT 2016 Workshop Report. *J. Rheumatol.* **2017**, *44*, 1536–1543. [CrossRef]

20. Bozzalla-Cassione, E.; Grignaschi, S.; Xoxi, B.; Luvaro, T.; Greco, M.I.; Mazzucchelli, I.; Bugatti, S.; Montecucco, C.; Manzo, A. Insights into the concept of rheumatoid arthritis flare. *Front. Med.* **2022**, *9*, 852220. [CrossRef] [PubMed]

21. Gannot, G.; Lancaster, H.E.; Fox, P.C. Clinical course of primary Sjogren's syndrome: Salivary, oral, and serologic aspects. *J. Rheumatol.* **2000**, *27*, 1905–1909. [PubMed]

22. CareClinic. Sjogrens Flare: How Long They Last, Causes, Management Tips. Available online: https://careclinic.io/sjogrens-flare/ (accessed on 30 October 2023).

23. Jung, J.; Zazay, I.; Kalia, P.; Shaghaghi, N.; Ross, N.; Homan, B. Semaglutide as a potential therapeutic adjunct for reducing flare-ups in Sjogren's syndrome: A case report. *SAGE Open Med. Case Rep.* **2025**, *13*, 2050313X251372813. [CrossRef]

24. de Fremont, G.M.; Costedoat-Chalumeau, N.; Lazaro, E.; Belkhir, R.; Guettrot-Imbert, G.; Morel, N.; Nocturne, G.; Molto, A.; Goulenok, T.; Diot, E.; et al. Pregnancy outcomes in women with primary Sjogren's syndrome: An analysis of data from the multicentre, prospective, GR2 study. *Lancet Rheumatol.* **2023**, *5*, e330–e340. [CrossRef]

25. Park, H.S.; Martinez-Martinez, L.; Magallares Lopez, B.; Castellvi, I.; Moya, P.; Codes-Mendez, H.; Hernandez Sosa, N.; Diaz-Torne, C.; Laiz, A.; Sainz, L.; et al. Prognostic significance of lymphocytic foci composition in minor salivary gland biopsies for severe disease flare and severity in Sjogren's syndrome: A 3-year follow-up cohort study. *Front. Immunol.* **2024**, *15*, 1332924. [CrossRef] [PubMed]

26. Seror, R.; Bowman, S.J.; Brito-Zeron, P.; Theander, E.; Bootsma, H.; Tzioufas, A.; Gottenberg, J.E.; Ramos-Casals, M.; Dorner, T.; Ravaud, P.; et al. EULAR Sjogren's Syndrome Disease Activity Index (ESSDAI): A user guide. *RMD Open* **2015**, *1*, e000022. [CrossRef] [PubMed]

27. Seror, R.; Ravaud, P.; Mariette, X.; Bootsma, H.; Theander, E.; Hansen, A.; Ramos-Casals, M.; Dorner, T.; Bombardieri, S.; Hachulla, E.; et al. EULAR Sjogren's Syndrome Patient Reported Index (ESSPRI): Development of a consensus patient index for primary Sjogren's syndrome. *Ann. Rheum. Dis.* **2011**, *70*, 968–972. [CrossRef] [PubMed]

28. Seror, R.; Theander, E.; Brun, J.G.; Ramos-Casals, M.; Valim, V.; Dorner, T.; Bootsma, H.; Tzioufas, A.; Solans-Laque, R.; Mandl, T.; et al. Validation of EULAR primary Sjogren's syndrome disease activity (ESSDAI) and patient indexes (ESSPRI). *Ann. Rheum. Dis.* **2015**, *74*, 859–866. [CrossRef]

29. Dougados, M.; Huizinga, T.W.; Choy, E.H.; Bingham, C.O., 3rd; Aassi, M.; Bernasconi, C. Evaluation of the disease activity score in twenty-eight joints-based flare definitions in rheumatoid arthritis: Data from a three-year clinical trial. *Arthritis Care Res.* **2015**, *67*, 1762–1766. [CrossRef]

30. Cook, N.; Mullins, A.; Gautam, R.; Medi, S.; Prince, C.; Tyagi, N.; Kommineni, J. Evaluating patient experiences in dry eye disease through social media listening research. *Ophthalmol. Ther.* **2019**, *8*, 407–420. [CrossRef]

31. McDonald, L.; Malcolm, B.; Ramagopalan, S.; Syrad, H. Real-world data and the patient perspective: The PROmise of social media? *BMC Med.* **2019**, *17*, 11. [CrossRef]

32. Flurie, M.; Converse, M.; Davidson, K. The whole person: Exploring sleep and mental health across rheumatic diseases. In Proceedings of the EULAR 2023 Congress, Milan, Italy, 31 May–3 June 2023.

33. SNOMED International. About Us. Available online: https://www.snomed.org/about-us (accessed on 6 November 2023).

34. National Institutes of Health. RxNorm. Available online: https://www.nlm.nih.gov/research/umls/rxnorm/index.html (accessed on 6 November 2023).

35. Bodenreider, O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res.* **2004**, *32*, D267–D270. [CrossRef]

36. Ekman, P. Facial expressions of emotion: New findings, new questions. *Psychol. Sci.* **1992**, *3*, 34–38. [CrossRef]

37. Bann, E.Y.; Bryson, J.J. The conceptualisation of emotion qualia: Semantic clustering of emotional tweets. In *Computational Models of Cognitive Processes*; Mayor, J., Gomez, P., Eds.; World Scientific: Singapore, 2013; pp. 249–263.

38. Bestaver, S.; Shah, S. Analyzing Text for Distinctive Terms Using Pointwise Mutual Information. *Pew Research Center—Decoded*. 13 July 2022. Available online: https://www.pewresearch.org/decoded/2022/07/13/analyzing-text-for-distinctive-terms-using-pointwise-mutual-information/ (accessed on 6 November 2023).

39. Role, F.; Nadif, M. Handling the impact of low frequency events on co-occurrence based measures of word similarity—A case study of pointwise mutual information. In Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR 2011), Paris, France, 26–29 October 2011.

40. Adamichou, C.; Bertsias, G. Flares in systemic lupus erythematosus: Diagnosis, risk factors and preventive strategies. *Mediterr. J. Rheumatol.* **2017**, *28*, 4–12. [CrossRef] [PubMed]

41. Choi, B.Y.; Oh, H.J.; Lee, Y.J.; Song, Y.W. Prevalence and clinical impact of fibromyalgia in patients with primary Sjogren's syndrome. *Clin. Exp. Rheumatol.* **2016**, *34*, S9–S13. [PubMed]

42. Gomez-Arguelles, J.M.; Moreno-Zazo, M.; Maestu, C. Characterizing fibromyalgia flares: A prospective observational study. *Reumatologia* **2022**, *60*, 242–246. [CrossRef]

43. Elefante, E.; Cornet, A.; Andersen, J.; Somers, A.; Mosca, M. The communication GAP between patients and clinicians and the importance of patient reported outcomes in systemic lupus erythematosus. *Best Pract. Res. Clin. Rheumatol.* **2023**, *37*, 101939. [CrossRef]

44. Curtis, J.R.; McLean, R.R.; Lee, I.H.; Mackey, R.H.; Moore, P.C.; Haubrich, R.; Greenberg, J.D.; Wu, A. Prevalence and factors associated with patient-clinician discordance among patients with rheumatoid arthritis initiating advanced therapy. *ACR Open Rheumatol.* **2024**, *6*, 253–264. [CrossRef]

45. Rogers, J.; Eudy, A.M.; Criscione-Schreiber, L.; Pisetsky, D.; Sun, K.; Doss, J.; Clowse, M.E.B. The use of patient-reported outcome measures to assess clinical features of fibromyalgia in SLE. *Arthritis Rheumatol.* **2018**, *70* (Suppl. S9). Available online: https://acrabstracts.org/abstract/the-use-of-patient-reported-outcome-measures-to-assess-clinical-features-of-fibromyalgia-in-sle/ (accessed on 19 September 2025).

46. Baldini, C.; Ferro, F.; Pepe, P.; Luciano, N.; Sernissi, F.; Cacciatore, C.; Martini, D.; Tavoni, A.; Mosca, M.; Bombardieri, S. Damage accrual in a single centre cohort of patients with primary Sjögren's syndrome followed up for over 10 years. In Proceedings of the ACR/ARHP Annual Meeting, San Diego, CA, USA, 25–30 October 2013. Abstract Number 505.

47. Jacquemin, C.; Molto, A.; Servy, H.; Sellam, J.; Foltz, V.; Gandjbakhch, F.; Hudry, C.; Mitrovic, S.; Granger, B.; Fautrel, B.; et al. Flares assessed weekly in patients with rheumatoid arthritis or axial spondyloarthritis and relationship with physical activity measured using a connected activity tracker: A 3-month study. *RMD Open* **2017**, *3*, e000434. [CrossRef]

48. Kuettel, D.; Glinatsi, D.; Ostergaard, M.; Terslev, L.; Primdahl, J.; Moller, S.; Pedersen, A.; Petersen, R.; Weber, U.; Horslev-Petersen, K. Serial magnetic resonance imaging and ultrasound examinations demonstrate differential inflammatory lesion patterns in soft tissue and bone upon patient-reported flares in rheumatoid arthritis. *Arthritis Res. Ther.* **2020**, *22*, 19. [CrossRef]

49. Myasoedova, E.; Chandran, A.; Ilhan, B.; Major, B.T.; Michet, C.J.; Matteson, E.L.; Crowson, C.S. The role of rheumatoid arthritis (RA) flare and cumulative burden of RA severity in the risk of cardiovascular disease. *Ann. Rheum. Dis.* **2016**, *75*, 560–565. [CrossRef]

50. Bartlett, S.; Barbic, S.; Bykerk, V.; Fautrel, B.; Guillemin, F.; Den Broeder, A.; Alten, R.; Christensen, R.; Choy, E.; Furst, D.; et al. SAT0132 The rheumatoid arthritis flare questionnaire (RA-FQ): Results of Rasch analysis and feedback on real-world applications from international RA patients and clinicians. *Ann. Rheum. Dis.* **2017**, *76*, 818–819. [CrossRef]

51. Food and Drug Administration (FDA). Patient-Focused Drug Development: Collecting Comprehensive and Representative Input. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. Available online: https://www.fda.gov/media/139088/download (accessed on 8 November 2023).

52. Food and Drug Administration (FDA). Framework for FDA's Real-World Evidence Program. Available online: https://www.fda.gov/media/120060/download (accessed on 8 November 2023).

53. Food and Drug Administration (FDA). Patient-Focused Drug Development: Methods to Identify What Is Important to Patients. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. Available online: https://www.fda.gov/media/131230/download (accessed on 8 November 2023).

54. Giorgi, S.; Lynn, V.E.; Gupta, K.; Ahmed, F.; Matz, S.; Ungar, L.H.; Schwartz, H.A. Correcting sociodemographic selection biases for population prediction from social media. *Proc. Int. AAAI Conf. Weblogs Soc. Media* **2022**, *16*, 228–240. [CrossRef] [PubMed]

55. Cesare, N.; Grant, C.; Hawkins, J.B.; Brownstein, J.S.; Nsoesie, E.O. Demographics in Social Media Data for Public Health Research: Does It Matter? 24 September 2017. Available online: https://arxiv.org/ftp/arxiv/papers/1710/1710.11048.pdf (accessed on 8 November 2023).