Genetics inMedicine

Check for updates

SYSTEMATIC REVIEW

Opportunities and pitfalls of social media research in rare genetic diseases: a systematic review

Emily G. Miller [ID][1], Amanda L. Woodward[2], Grace Flinchum[3], Jennifer L. Young[1], Holly K. Tabor[1,4] and Meghan C. Halley[1 ✉]

PURPOSE: Social media may be particularly valuable in research in rare genetic diseases because of the low numbers of patients and the rare disease community's robust online presence. The goal of this systematic review was to understand how social media is currently used in rare disease research and the characteristics of the participants in these studies.
METHODS: We conducted a systematic review of six databases to identify studies published in English between January 2004 and November 2020, of which 120 met inclusion criteria.
RESULTS: Most studies were observational ($n = 114$, 95.0%) and cross-sectional ($n = 107$, 89.2%), and more than half ($n = 69$, 57.5%) utilized only surveys. Only 101 rare diseases were included across all studies. Participant demographics, when reported, were predominantly female (70.1% ± 22.5%) and white (85.0% ± 11.0%) adult patients and caregivers.
CONCLUSION: Despite its potential benefits in rare disease research, the use of social media is still methodologically limited and the participants reached may not be representative of the rare disease population by gender, race, age, or rare disease type. As scholars explore using social media for rare disease research, careful attention should be paid to representativeness when studying this diverse patient community.

INTRODUCTION

Rare diseases affect an estimated 300 million people worldwide [1]. Though definitions vary across countries, the National Institutes of Health defines a rare disease as one affecting fewer than 200,000 people in the United States [2]. Approximately 80% of rare diseases are genetic in etiology [2]. Patients with rare diseases experience extended diagnostic odysseys lasting an average of six years, while some individuals remain undiagnosed indefinitely [1]. Though specific symptoms vary widely, many rare disease patients suffer from complex, poorly understood medical conditions, and the vast majority of rare diseases lack a Federal Drug Administration–approved treatment [3]. Patients and families struggle to access health-care providers with sufficient knowledge of their conditions and must often coordinate health care across multiple specialty providers [4].

Conducting health research on patients with rare diseases is particularly challenging. The low prevalence of each condition means that those who share a given diagnosis are likely to be geographically dispersed, making it difficult to recruit sufficient numbers of patients for research [4]. However, the rapid expansion of social media over the last 15 years has provided a new opportunity for rare disease patients to find each other. There is evidence that rare disease patients use social media frequently for a range of purposes, including social and informational support, research, and advocacy [5]. Due in part to rare disease patients' high utilization of social media, there has been growing interest in using social media to facilitate rare disease research across the spectrum of clinical and translational research [6].

Social media methods have several advantages over traditional methods (e.g., clinic-based recruitment) in health research, including increased access to patients, larger sample sizes, and more efficient recruitment [7]. Additionally, the content users generate on social media may provide valuable patient-reported data on disease course, health outcomes, and quality of life, as well as a new forum for delivery and evaluation of targeted health interventions [8]. For rare disease research, social media has been used to collect patient histories [9], examine patient needs via content analysis of support group posts [10], recruit rare disease patients for studies [11], and allow for data mining of information on symptoms and health outcomes [12]. Additionally, research has examined social media itself to understand how and why rare disease patients and their families use these online platforms [12]. Social media presents an opportunity to conduct new types of research in rare disease and to address longstanding challenges in research with these patients and families.

There are also potential drawbacks to using social media for studying rare diseases, including issues of representation and generalizability in study samples. Research in other conditions suggests that samples drawn from social media may be subject to biases in terms of gender, race/ethnicity, and age [13]. As there are over 7,000 rare diseases, findings or methods applicable to one disease may not apply to others [2]. This is particularly a concern in studies that make claims about the experiences of "rare disease patients" overall. Systematic over- or underrepresentation of certain diseases, or of patients from certain sociodemographic subgroups, could have a cumulative effect of biasing our knowledge of rare diseases. While social media represents a potentially powerful tool for rare disease research, it is necessary to understand these possible drawbacks.

The goal of this study is to systematically review the peer-reviewed academic literature on the use of social media in rare disease research. In this review, we examine how social media has

---

[1]Stanford Center for Biomedical Ethics, Stanford University, Stanford, CA, USA. [2]Lane Medical Library, Stanford University, Stanford, CA, USA. [3]Department of Political Science, Davidson College, Davidson, NC, USA. [4]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ✉email: mhalley@stanford.edu
Co-senior authors: Holly K. Tabor, Meghan C. Halley

been used in rare disease research, the types of research questions examined, the methods used, and the characteristics of participants included in these studies, with a focus on identifying gaps and opportunities in rare disease research using social media.

MATERIALS AND METHODS

We conducted a systematic review of the peer-reviewed academic literature. Our preregistered protocol containing our detailed methods is available at Open Science Framework (https://osf.io/, protocol ID 97fd6).

Eligibility criteria

We included studies that met the following criteria: (1) focused on the topic of rare or undiagnosed genetic diseases, (2) used social media to conduct the research, and (3) were published in English in a peer-reviewed journal between 1 January 2004 and 10 November 2020. We chose 2004 as our start date because MySpace, the first widely used social media site, was launched in this year [14]. We included rare diseases that meet the US definition of rare disease (a disease with a prevalence of fewer than 200,000 US citizens) [2]. We defined social media as any online site with user-generated content that also allowed for direct communication between user-specific profiles and groups. Examples included Facebook, Twitter, Reddit, and YouTube. We included studies that focused on rare diseases in general, a single rare disease, a group of diseases including at least one rare disease, or caregivers for those with rare diseases. We excluded studies that focused on rare infectious diseases or rare diseases with a known nongenetic etiology, as research on these acquired diseases may focus on issues such as prevention, and therefore may not be applicable to most rare diseases.

Search strategy

We developed a sensitive search strategy in collaboration with an academic reference librarian (A.L.W.). Our rare disease search terms included general terms for rare disease as well as keywords based on the rare disease categories defined by the Genetic and Rare Diseases Information Center (GARD) and Orphanet [15]. Our social media search terms included synonyms for social media (e.g., "online forum") as well as the names of the ten social media sites with the largest number of global users during our search period [14]. We adapted our search string for six databases. The databases used were PubMed, Embase, and CINAHL (via Ebsco) to find biomedical literature; PsycINFO (via ProQuest) to find psychology literature; Communications & Mass Media Complete (via ProQuest) to find social media and communications literature; and Web of Science to find interdisciplinary literature. Search strings for each database can be found in Appendix A.

To supplement our broad search, we also conducted targeted searches within the following journals that focus on rare or genetic diseases and/or online research: *Rare Diseases, Intractable & Rare Diseases Research, Orphanet Journal of Rare Diseases, Rare Diseases of the Immune System, Rare Tumors, Journal of Genetic Counseling, Genetics in Medicine,* and *The Journal of Medical Internet Research*. Finally, we manually reviewed the reference lists of all included articles, as well as any relevant reviews identified through our search, and screened all articles with keywords such as "rare disease" and "social media" based on our inclusion criteria.

Article selection

The results of our database and manual searches were exported into Zotero [16]. Duplicates and retracted papers were removed, and the remaining articles were uploaded into Covidence [17]. One reviewer (E.G.M.) screened each article for eligibility by title and abstract. Articles that clearly failed the eligibility criteria were

excluded. Two reviewers (E.G.M. and M.C.H.) then independently reviewed the full text of the remaining articles to assess eligibility. Disagreements between reviewers were resolved through an iterative consensus process involving multiple rounds of deliberative discussion.

Data extraction

We extracted detailed study characteristics and recorded them in Microsoft Excel, version 16.45 [18]. One author (E.G.M.) extracted verbatim text from each article relevant to publication details, study aims, methods, participants, results, strengths, and limitations from all eligible studies. The verbatim text excerpts were then uploaded into Dedoose, where categorical variables were created and assigned by two independent reviewers (E.G.M. and G.F.) [19]. Variables were created to categorize study aims, methods used, study design, disease categories, specific disease (s) studied, role of social media, social media site(s) used, justification for social media sampling frame, and countries represented in the study. We used disease categories taken from the GARD website to categorize rare diseases into groups [2]. A full codebook with definitions can be found in Appendix B.

Data analysis

We conducted descriptive data analysis using Excel to generate summary statistics for the entire sample. We also summarized participant demographics across each study, as well as patient demographics when a given study participant was a caregiver.

RESULTS

Our initial search yielded 3,437 articles, and 12 more were found in our manual search. After removal of duplicates and screening, 120 articles were included (Fig. 1) [6, 8–12, 20–133]. Supplementary Table 1 includes detailed data for each study.

Study characteristics

The studies reflected four broad goals, with some studies addressing multiple goals. The most common goal was the evaluation of the psychosocial challenges and needs of patients or caregivers, including topics related to mental health, social support, coping behaviors, and health-care access ($n = 48$, 40.0%), e.g., [29], [65], [88]. Many studies sought to evaluate patient physical health status or outcomes ($n = 40$, 33.3%), examining topics like survival comparisons between countries [21], the effectiveness of various treatments, e.g., [24], [58], [113], and phenotypic characterizations of diseases, e.g., [44], [53], [74]. Some studies aimed to gather information on patients' use of social media itself ($n = 34$, 28.3%), e.g., [28], [38], [56], and 14 studies (11.7%) aimed to develop social media research methods, typically through feasibility studies e.g., [27], [49], [70] (Table 1).

The majority of studies were observational ($n = 114$, 95.0%), e.g., [6], [40], [73], and cross-sectional ($n = 107$, 89.2%), e.g., [35], [67], [126]. More than half ($n = 69$, 57.5%) used surveys to collect data from rare disease patients or caregivers, e.g., [32], [66], [105], and a third ($n = 37$, 30.8%) conducted secondary data analyses of existing social media content (e.g., posts, videos, tweets), e.g., [42], [63], [102]. Other methodologies included telephone, video, or in-person interviews ($n = 10$, 8.3%), e.g., [27], [35], [96], and/or clinical research ($n = 11$, 9.2%) such as physical evaluations or natural history studies, e.g., [40], [59], [74]. Five studies (4.2%) involved an online intervention in which a social media group was created for the purposes of the study [8, 20, 67, 81, 94] (Table 1).

The number of studies published annually increased over time. No studies were published before 2007. The year with the most studies published ($n = 24$) was 2020, even though our review only
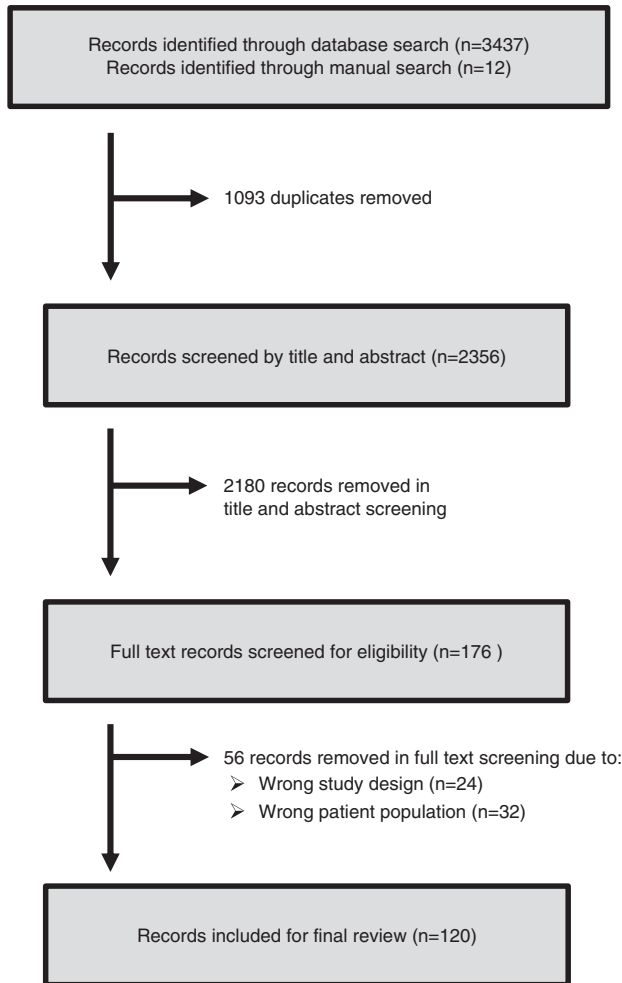
Fig. 1 PRISMA flow diagram detailing screening and selection of articles. Final sample size = 120 articles.

included studies published before 10 November 2020, and so did not capture the entire calendar year. The use of surveys increased the most from 2004 to 2020 compared to other methods (Supplementary Fig. 1).

Types and uses of social media

Eleven different social media sites were used across the 120 studies, and 22 studies (18.3%) used more than one platform, e.g., [73], [112], [120]. The most commonly used platforms were Facebook ($n = 59$, 49.2%), e.g., [77], [123], [131], and Twitter ($n = 28$, 23.3%), e.g., [9], [85], [100] (Table 2).

Across all studies, 79 (65.8%) used social media for recruitment of study participants, including for online surveys, in-person studies, clinical trials, or phone, video, and in-person interviews, e.g., [33], [92], [96]. Of these 79 studies, 64 also used social media for primary data collection (including through a link to an online survey hosted by a different platform), e.g., [88], [106], [118]. Secondary data analysis of the existing content of social media sites was the second most common method used ($n = 38$, 31.7%), e.g., [10], [104], [120]. Only four studies used social media for intervention delivery. Examples of intervention delivery included the creation of a social media site to determine its impact on the psychological well-being of rare disease patients or caregivers [20, 22, 94] and the use of social media to provide postoperative care training for rare disease patients [8] (Table 2).

Table 1. Study characteristics.

	Number of studies, n (%)[a]
Study goal(s)[a]	
Evaluation of psychosocial needs of patient or caregiver	48 (40.0%)
Evaluation of patient health status or outcomes	40 (33.3%)
Evaluation of social media use	34 (28.3%)
Development of research methods	14 (11.7%)
Other	6 (5.0%)
Study type[a]	
Observational	114 (95.0%)
Cross-sectional	107 (89.2%)
Longitudinal	11 (9.2%)
Experimental	5 (4.2%)
Other	1 (0.8%)
Methodology[a]	
Survey	69 (57.5%)
Secondary data analysis	37 (30.8%)
Interview	10 (8.3%)
Treatment study	11 (9.2%)
Online intervention	5 (4.2%)
Other	5 (4.2%)
Year of publication	
2004 to 2011	9 (7.5%)
2012 to 2014	12 (10.0%)
2015 to 2017	40 (33.3%)
2017 to 2020	59 (49.2%)

The number and percent of the $n = 120$ papers by study goals, study type, methods, and year of publication.
[a]Categories are not mutually exclusive, percentages may sum to >100.

Diseases studied

The 120 studies included 101 different diseases. The majority of studies ($n = 89$, 74.2%) focused on a single rare disease, e.g., [21], [87], [128], while others focused on a subset of multiple rare diseases ($n = 18$, 15.0%) e.g., [50], [116], [117], or a mixture of rare and nonrare diseases ($n = 7$, 5.8%) [9, 45, 47, 57, 121, 126, 127]. A subset ($n = 5$, 4.2%) stated a focus on "rare diseases" but did not specify which diseases were included [11, 12, 27, 38, 60], and one of the studies (0.8%) focused on undiagnosed rare diseases (Fig. 2) [81].

Across all studies, cystic fibrosis (CF; $n = 14$, 11.6%) was the most frequently studied disease, e.g., [35], [75], [88], followed by amyotrophic lateral sclerosis (ALS; $n = 7$, 5.8%), e.g., [56], [57], [126], blastic plasmacytoid dendritic cell neoplasm (BPDCN) [100–103], and Huntington disease [10, 26, 41, 49] were each included in 4 (3.3%) studies, and Hirschsprung disease [8, 128, 129], myeloproliferative neoplasm [99, 103, 132], neurofibromatosis type 1 [20, 70, 110], and sickle cell disease [28, 82, 120] were each included in 3 (2.5%) studies (Fig. 2). All other diseases were represented in two or fewer studies. Eighteen (15.0%) of the studies focused on rare cancers, e.g., [64], [94], [100]. Supplementary Tables 2 and 3 contain a full list of diseases and disease categories included across studies.

Study participant characteristics

Of the 84 studies that included data collection from human subjects, 40.5% ($n = 34$) included only rare disease patient participants, e.g., [43], [83], [88]; 19.0% ($n = 16$) included only caregivers of rare disease patients, e.g., [20], [94], [96], [114]; and 39.3% ($n = 33$) included both caregivers and patients, e.g., [50], [81], [97]. Demographic reporting included patient self-report, caregiver self-report, and caregivers reporting the demographics of the patients they cared for. In addition, some studies included both patients and caregivers, but did not disaggregate patient demographic data by reporter.

Across all studies, race/ethnicity of both patients and caregivers was underreported compared to age and sex. Demographics of caregivers were reported less frequently than patient demographics (Fig. 3).

Race/ethnicity

Only 20 studies reported information on patient race and/or ethnicity e.g., [70, 27, 117, 133] (Fig. 3). ethnicity, /ethnicity, we summarized demographics across studies using sociopolitical categories (i.e., Black, White), though studies used a range of both social and ancestry-based terms [134]. Across all studies reporting patient race/ethnicity ($n = 20$ papers) (including patient self-reported, caregiver reporting patients, and patients for whom the reporter is not specified) a mean of 85.0% ($\pm11.0\%$) of patients identified as White. For the subset of these papers in which rare disease patients self-reported their race/ethnicity ($n = 12$ papers), a mean of 88.6% ($\pm12.3\%$) of patients identified as White. No studies reported the race/ethnicity of the rare disease patient when only the caregiver was reporting. When caregivers reported their own race/ethnicity ($n = 6$ papers), a mean of 78.6% ($\pm22.0\%$) of caregivers identified as White (Table 3).

Reporting of non-White racial and ethnic categories was widely variable. Of the 26 studies that reported either patient or caregiver race/ethnicity, 10 reported only percentage White, e.g., [40], [58], [114], 5 reported only percentage White and percentage Hispanic [22, 27, 29, 48, 117], and 11 reported percentage White and at least one other race, e.g., [26], [70], [133]. Among the 16 studies reporting race/ethnic categories other than White, an average of 4.6% of respondents identified as Black ($\pm3.2\%$), 5.4% as Asian American race ($\pm3.7\%$), 15.4% as Hispanic ($\pm27.7\%$), and 1.03% as Native American ($\pm1.2\%$).

Sex

In studies with rare disease patient participants ($n = 35$ papers), an average of 80.0% were female ($\pm21.0\%$) (Table 3). However, when we excluded studies that focused on diseases that disproportionately affected one sex or that used gender-biased recruitment methods (e.g., recruiting from a women's group) ($n = 17$ papers), e.g., [35], [67], [125], the mean dropped to 63.8% ($\pm15.6\%$) female.

In studies with only caregivers reporting on patient demographics ($n = 17$ papers), 52.9% reported the sex of the cared-for patient ($n = 9$ papers). For those studies that did report patient sex, a mean of half (51.2% $\pm$ 6.7%) of patients cared for were female. Caregivers reported their own sex in only 33.3% of studies ($n = 16$ papers out of 45 studies with at least some caregiver reporting), with an average of 77% female ($\pm23.5\%$) across studies. One study had skewed recruitment because it recruited from a fathers' group [94]. When this study was excluded, an average of 83% of caregiver participants were female across studies ($\pm11.8\%$) (Table 3 and Fig. 3).

Table 2. Social media use characteristics.

	Number of studies n (%)[a]
Social media platform(s) used	
Facebook	59 (49.2%)
Twitter	28 (23.3%)
Specific rare disease website or forum	19 (15.8%)
Unspecified "online support group"	13 (10.8%)
Unspecified "social media"	13 (10.8%)
PatientsLikeMe	6 (5.0%)
YouTube	4 (3.3%)
Instagram	2 (1.7%)
Reddit	2 (1.7%)
Other[b]	7 (5.8%)
Role(s) of social media	
Recruitment	79 (65.8%)
Primary data collection	64 (53.3%)
Secondary data analysis	38 (31.7%)
Intervention delivery	4 (3.3%)
Other	1 (0.8%)

Frequency of social media platforms and role among $n = 120$ studies.
[a]Categories are not mutually exclusive, percentages may sum to >100.
[b]Social media sites included in "Other" were each used by one study. These included Pinterest, WeChat, Tumblr, LinkedIn, RareConnect, Yahoo, and "Internet based blogs."

Most Frequently Studied Diseases	Number of Studies (n, %)
Cystic Fibrosis	14 (11.6%)
ALS	7 (5.8%)
BPDCN	4 (3.3%)
Huntington's disease	4 (3.3%)
Myeloproliferative neoplasm	3 (2.5%)
Neurofibromatosis type 1	3 (2.5%)
Sickle cell disease	3 (2.5%)

Pie chart segments: 89, 74% (Single rare disease); 18, 15% (Multiple rare diseases); 7, 6% (Mixed rare and non-rare diseases); 5, 4% (Unspecified "rare diseases"); 1, 1%.
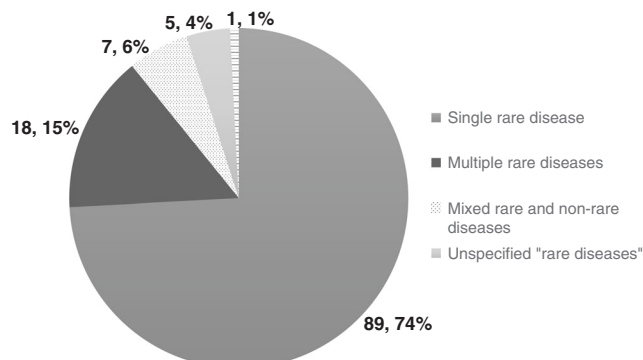
Fig. 2 Diseases studied. Pie chart of diseases studied by rare disease focus. Accompanying table of the most commonly studied rare diseases across all studies. *ALS:* amyotrophic lateral sclerosis, *BPDCN:* blastic plasmacytoid dendritic cell neoplasm.

Across all studies (including patient-reported, caregiver-reported, and those in which the reporter was not specified) ($n = 62$ papers), approximately two-thirds (70.1% ± 22.5%) of all patients were female, suggesting that rare disease research on social media may underreport accounts from male rare disease patients.

Age

In studies with rare disease patient participants ($n = 36$ papers), 91.7% ($n = 33$) reported patient age. Of these, all studies included adult patients, but only 12.1% ($n = 4$) also included pediatric patients (<18 years of age) [47, 65, 75, 130]. In studies with only caregiver participants ($n = 17$ papers), 88.2% ($n = 15$) reported the patients' ages, e.g., [24], [37], [53]. Of these, all included caregivers of pediatric patients, and 66.7% of studies ($n = 10$) also included caregivers of adult patients, e.g., [11], [84], [92]. The majority of pediatric patients were therefore studied through their caregivers. The methods used to collect and report pediatric patients' age varied widely; among those papers that reported a mean patient age ($n = 7$ papers), the average age of pediatric patient participants across studies was 7.6 years old (±5.7 years).
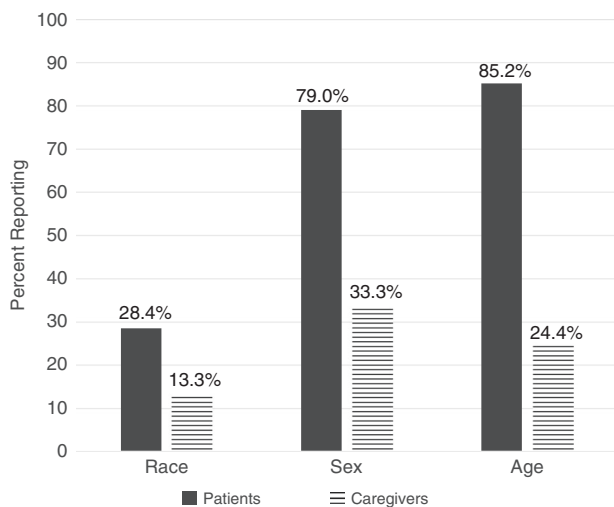

Fig. 3 Percentage of studies reporting age, race, and sex of patients and caregivers. Patient group includes self-reported patients, caregiver-reported patients, and patients for whom the reporter is unclear.

Across all studies that reported patient age, regardless of reporter ($n = 67$ papers), more studies included adult rare disease patients than pediatric rare disease patients (88.4% vs. 50.7%). Caregiver age was reported in 24.4% ($n = 11$) of papers that included at least some caregivers ($n = 45$ papers), and all caregivers were adults (Fig. 3).

Location

Fifty-four (45.2%) of the studies reported at least one country or region for study participants, e.g., [20]–[22]. The mean number of countries reported was 6.5 (±8.6), though 22.5% ($n = 27$) of studies included participants from only one country, e.g., [10], [40], [41]. The highest number of countries represented by one study was 40, though individual countries were not specified [70]. The majority of countries were Western and English-speaking, and the most frequently reported countries were the United States ($n = 36$), e.g., [126], [130], [131]; the United Kingdom ($n = 20$), e.g., [32], [35], [44]; Canada ($n = 18$), e.g., [21], [30], [36]; Australia ($n = 16$), e.g., [67], [68], [74]; and New Zealand ($n = 9$), e.g., [6], [30], [44]. This is not surprising given that our review was limited to studies published in English. See Supplementary Table 4 for the frequencies of all countries reported.

Participant sampling and sample size

All samples were convenience samples; no studies reported using a representative sampling frame. Over half ($n = 72$, 60.0%) of the studies did not provide justifications for their choices of social media platforms or sampling frames for participants, e.g., [26], [29], [85]. When reported, the most common justification for choice of social media was the size of a particular online community (e.g., "this platform has the most users") ($n = 25$, 20.8%), e.g., [42], [70], [132], followed by an existing collaboration with a group or organization ($n = 19$, 15.8%), e.g., [23], [55], [76]. The remaining studies that provided justifications for their sampling indicated that they chose a certain platform because it included a format of information (e.g., videos) or type of data (e.g., group was unmoderated, information was public, participants were verified) that was specific to study goals ($n = 14$, 11.7%), e.g., [9], [57], [104].

Study sample sizes ranged from 2 to 4,860 people. Six studies claimed to have reached the largest cohort of their specific rare disease ever recruited for a single study, though we did not independently verify these claims [6], [42], [46], [55], [90], [117].

Table 3. Study participant demographics by reporter.

	Patient demographics		Caregiver demographics
	Self-reported (36 studies)	Caregiver-reported (17 studies)	Self-reported (45 studies)
Sex	35[a] reported sex	9 reported sex	16 reported sex
Percent female M (± SD) across studies	80.0% (±21.0%)	51.2% (±6.7%)	77.0% (±23.5%)
Race	12 reported race	0 reported race	6 reported race
Percent White M (± SD) across studies	88.6% (±12.3%)	0	78.6% (±22.0%)
Age	33 reported age	15 reported age	11 reported age
Includes pediatric participants, number of studies (%)	4 (12.1%)	15 (100%)	0
Includes adult participants, number of studies (%)	33 (100.0%)	10 (66.7%)	11 (100.0%)

Demographics of rare disease patients and caregivers. Patient demographics are broken down by self-reported and caregiver-reported (i.e., studies in which a caregiver participant reported information about the patient).
[a]Number of studies.

DISCUSSION

The results of this systematic review indicate that there has been a rapid increase in the use of social media for rare disease research over the last 13 years. However, this research is still limited in terms of goals, methods, and study designs, as well as its representativeness of the broader rare disease community, both in terms of disease type and patient demographics.

Our results indicate that, in rare disease research, social media has primarily been used for recruitment in observational, cross-sectional studies. This is in contrast to social media research in other fields, where researchers have developed methods for employing interventional and longitudinal designs in social media research, as well as strategies to help reach a more representative sample using social media [135, 136]. In our sample, a small number of studies used social media in unique ways, for example, by mining social media data to identify adverse reactions to medications in order to guide drug development for rare diseases [137], or as a component of communication for postdischarge follow-up with patients and caregivers [8]. Social media has the potential to be further utilized by those in the rare disease community—as it has in other fields, such as cancer prevention and adolescent health—to increase mental and physical well-being of patients and to share health information and emerging research [13, 138].

In rare disease research, the cohorts of patients recruited through social media may not represent the broader rare disease community in terms of gender and race/ethnicity. Despite social media use being nearly equal across racial groups and only slightly higher in women than men, White and female participants were overrepresented in our included studies [139]. This is consistent with other reviews that have found an overrepresentation of White and female participants in studies using social media [140, 141]. Race/ethnicity was also highly underreported and, when reported, used a range of sociopolitical and ancestry-based categories in direct contrast to recent guidelines [134]. This lack of diversity is a problem in genomics studies in general, as individuals identifying as Black, Native American, or Hispanic/Latino are rarely included in genome-wide association studies [142]. Our findings indicate that, in addition to being excluded from genome sequencing studies, racial and ethnic minorities also are excluded from studies of rare diseases using social media. This is concerning, as our review demonstrates that a key focus of these studies is the psychosocial and health needs and challenges of rare disease patients and their families. By excluding non-White populations, who may face additional challenges associated with racial bias in our health-care system and society more broadly, we are likely overlooking the needs of many rare disease patients and families.

Additionally, only 4 of the 120 studies in our sample included a pediatric perspective that came directly from the pediatric patient. This may be due to the challenges of obtaining parental consent for research using social media and suggests that different approaches may be needed to understand the pediatric perspective on rare disease. It is difficult to assess the extent to which participants recruited through social media research are representative of the broader rare disease community, as there is currently no unified system for tracking epidemiologic data on rare diseases in the United States [143]. Regardless, it is clear that rare diseases affect all races, genders, and ages, and therefore information drawn from cohorts that are mostly White, female, and adult might not be generalizable to the rare disease community at large.

Studies were not representative in terms of disease type, with only 101 (1.4%) of the estimated 7,000 rare diseases represented. Some diseases were also over- or underrepresented relative to their prevalence. For example, cystic fibrosis was by far the most studied disease ($n = 14$, 11.6%), while sickle cell disease, which has a prevalence in the United States of over three times that of cystic fibrosis (100,000 vs. 30,000) [2, 144] was only included in three studies. Further, many rare disease communities may not currently be reachable through social media, as suggested by a recent study showing that only one in five rare pediatric diseases has a disease-specific group on Facebook [145]. It is possible that patients without dedicated groups for their specific diseases are present in groups for the broader rare disease community, but more research is needed to understand the distribution of rare diseases across the landscape of social media. Over- or underrepresentation of certain rare diseases may bias the general body of rare disease research, making it more difficult to draw conclusions about the needs and characteristics of rare disease patients as a whole.

Based on the findings of this review, we suggest a few key steps researchers can take to improve the quality and utility of future rare disease research using social media:

1. Researchers can focus on increasing representativeness in study samples. They should explore new methods to reach underrepresented demographics, such as reaching out to social media groups that specifically include male or non-White rare disease patients and their families, and/or diversify their recruitment methods beyond social media. Researchers could also use social media to contact community groups (many of which have a presence on social media), and these groups could help recruit patients and families outside of the social media platform. If that is not possible, researchers should, at a minimum, discuss the potential implications of bias in their results.

2. Researchers can more thoroughly report participant demographics and study methods. This should include clear reporting of race, gender, age, and nationality for both patients and caregivers. While there may be some occasions —for example, when studying ultrarare diseases—when concerns about confidentiality may be an obstacle to reporting detailed demographic data for a given study sample, researchers should still consider how gender, race/ethnicity, and age bias may influence their findings. Furthermore, new recommendations for publishing research in genetics and genomics have emphasized the need for authors to explicitly define race in genetics research, and for journals to provide clear guidelines for reporting race and other sociopolitical characteristics [134].

3. Researchers can clearly report methodological details including the social media platform used. They should justify both their choice of platform and choice of sampling frame for recruitment. Social media is not a monolith, and different groups may attract different types of rare disease patients. While some groups are solely for rare disease patients, others cater specifically to family members or health-care providers; others are open to all these categories [5]. Researchers should carefully consider their study goals when choosing the type of social media group to target. For further guidance on rigor in research using social media, rare disease researchers can turn to standardized reporting checklists developed specifically for online research [146].

Our study has several limitations. First, although we developed a detailed search string and augmented our search with manual review (see Appendix A), there is no way to ensure we captured every study on every rare disease. This limitation is inherent in the study of "rare diseases," which are, by their very nature, a large group of heterogeneous conditions that are methodologically challenging to study. However, we feel our carefully designed search string substantially mitigated against this limitation. Second, our review only includes studies published in the peer-reviewed research literature, and therefore does not include studies that may have been conducted by advocacy or other nonprofit organizations. However, the peer-reviewed literature is

widely considered to represent the most rigorous scientific work available and should be held to the high standards outlined in our conclusion. Third, our review was limited to studies published in English. Given the global burden of rare disease, we likely missed studies published in other languages, and therefore our review may lack a global perspective on how social media is used in rare disease research.

Conclusion
Social media is increasingly used to study hard-to-reach populations, including rare disease patients and their caregivers, in innovative and important ways. While social media is a potentially powerful tool, its current application in rare disease research is limited to primarily observational, cross-sectional studies using surveys to examine patient experiences and patient-reported outcomes. In addition, rare disease patients and caregivers reached by social media studies may not be representative of the rare disease population by gender or race/ethnicity, and represent only a small percentage of the over 7,000 identified rare diseases. As scholars explore new approaches to using social media for rare disease research, careful attention should be paid to representation within this large and diverse patient community.

DATA AVAILABILITY
The data set supporting the current study is included as a supplemental table.

REFERENCES
1. Shire. Rare disease impact report: insights from patients and the medical community. 2013. https://globalgenes.org/wp-content/uploads/2013/04/ShireReport-1.pdf.
2. Diseases. GARD: Genetic and Rare Diseases Information Center. National Institutes of Health, https://rarediseases.info.nih.gov/diseases.
3. Field MJ, Boat TF, Institute of Medicine (US), National Academies Press (US), eds. Rare Diseases and Orphan Products: Accelerating Research and Development. Washington (DC): National Academies Press; 2010.
4. Stoller JK. The challenge of rare diseases. Chest. 2018;153:1309–1314.
5. Iyer AA, Barzilay JR, Tabor HK. Patient and family social media use surrounding a novel treatment for a rare genetic disease: a qualitative interview study. Genet Med. 2020;22:1830–1837. https://doi.org/10.1038/s41436-020-0890-6.
6. Schumacher KR, et al. Social media methods for studying rare diseases. Pediatrics. 2014;133:e1345–e1353.
7. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res. 2016;18: e286.
8. Shen Y, Fang Y, Wu D, Bai J, Lin Y. Application of WeChat-assisted peri-operative care in children with congenital megacolon. J Paediatr Child Health. 2020 ;56:1551–6. https://doi.org/10.1111/jpc.14990.
9. Yamaguchi A, Queralt-Rosinach N. A proof-of-concept study of extracting patient histories for rare/intractable diseases from social media. Genomics Inform. 2020;18:e17.
10. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns. 2007;68:173–178.
11. Marshall DA, et al. The value of diagnostic testing for parents of children with rare genetic diseases. Genet Med. 2019;21:2798–2806.
12. Subirats L, Reguera N, Bañón A, Gómez-Zúñiga B, Minguillón J, Armayones M. Mining Facebook data of people with rare diseases: a content-based and temporal analysis. Int J Environ Res Public Health. 2018;15:1877.
13. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health. 2018;4:205520761877175.
14. Ortiz-Ospina E. The rise of social media. Our World in Data. 2019. https://ourworldindata.org/rise-of-social-media. Accessed 30 October 2020.
15. Orphanet. Linearization of disorders. http://www.orphadata.org/cgi-bin/rare_free.html. Accessed 4 November 2020.
16. Zotero. 2020. Center for History and New Media at George Mason University, Fairfax County, Virginia. www.zotero.org.
17. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org.
18. Microsoft Excel for Mac. 2021. Microsoft Corporation, Redmond, Washington, U.S.
19. Dedoose Version 8.3.45, web application for managing, analyzing, and presenting qualitative and mixed method research data (2020). Los Angeles, CA: SocioCultural Research Consultants, LLC. www.dedoose.com.
20. Akre C, Polvinen J, Ullrich NJ, Rich M. Children's at home: pilot study assessing dedicated social media for parents of adolescents with neurofibromatosis type 1. J Genet Couns. 2018;27:505–517.
21. Abhyankar P, Anand A, Stanbrook MB. On survival comparisons between adult cystic fibrosis patients in Canada versus the United States: Twitter discussions from @respandsleepjc (#rsjc). Can J Respir Crit Care Sleep Med. 2018;2:180–181.
22. Acher CW, Ostlie DJ, Leys CM, Struckmeyer S, Parker M, Nichol PF. Long-term outcomes of patients with tracheoesophageal fistula/esophageal atresia: survey results from tracheoesophageal fistula/esophageal atresia online communities. Eur J Pediatr Surg. 2016;26:476–480.
23. Albright K, et al. Seeking and sharing: why the pulmonary fibrosis community engages the web 2.0 environment. BMC Pulm Med. 2016;16:1–7.
24. Ali R, et al. Use of social media to assess the effectiveness of vagal nerve stimulation in Dravet syndrome: a caregiver's perspective. J Neurol Sci. 2017;375:146–149.
25. Al-Shaqsi SZ, Rai A, Forrest C, Phillips J. Public perception of a normal head shape in children with sagittal craniosynostosis. J Craniofac Surg. 2020;31:940–944.
26. Anderson KE, et al. Quality of care for Huntington's disease in the United States: findings from a national survey of patients and caregivers. J Huntingt Dis. 2019;8:509–519.
27. Applequist J, et al. A novel approach to conducting clinical trials in the community setting: utilizing patient-driven platforms and social media to drive web-based patient recruitment. BMC Med Res Methodol. 2020;20:58.
28. Barriteau CM, Thompson AL, Meier ER, Pecker LH. Sickle cell disease related internet activity is three times less frequent than cystic fibrosis related internet activity. Pediatr Blood Cancer. 2016;63:2061–2062.
29. Baxter M, Erby L, Roter D, Bernhardt BA, Terry P, Guttmacher A. Health screening behaviors among adults with hereditary hemorrhagic telangiectasia in North America. Genet Med. 2017;19:659–666.
30. Bedgood R, Sadurski R, Schade RR. The use of the internet in data assimilation in rare diseases. Dig Dis Sci. 2007;52:307–312.
31. Bi Q, et al. Determining the topic evolution and sentiment polarity for albinism in a Chinese online health community: machine learning and social network analysis. JMIR Med Inform. 2020;8:e17813.
32. Bowmer G, et al. Adherence to infection prevention and control guidelines: a vignette-based study of decision-making and risk-taking in young adults with cystic fibrosis. J Cyst Fibros. 2017;16:146–150.
33. Bruckner AL, et al. The challenges of living with and managing epidermolysis bullosa: insights from patients and caregivers. Orphanet J Rare Dis. 2020;15. https://doi.org/10.1186/s13023-019-1279-y.
34. Calthorpe RJ, et al. What effective ways of motivation, support and technologies help people with cystic fibrosis improve and sustain adherence to treatment? BMJ Open Respir Res. 2020;7. https://doi.org/10.1136/bmjresp-2020-000601.
35. Cammidge SL, Duff AJA, Latchford GJ, Etherington C. When women with cystic fibrosis become mothers: psychosocial impact and adjustments. Pulm Med. 2016;2016:9458980.
36. Canalichio KL, et al. Patient-reported outcomes in adult females with bladder exstrophy: a study of long-term sexual, reproductive and urinary outcomes using social media. J Pediatr Urol. 2020. https://doi.org/10.1016/j.jpurol.2020.06.020.
37. St Carter MT, Pierre SA, Zackai EH, Emanuel BS, Boycott KM. Phenotypic delineation of Emanuel syndrome (supernumerary derivative 22 syndrome): clinical features of 63 individuals. Am J Med Genet A. 2009;149A:1712–1721.
38. Castillo Esparcia A, López Villafranca P, Carretón Ballester MC. Online communication of patients with rare diseases in Spain. Rev Lat Comun Soc. 2015;70:673–688.
39. Collins R, et al. Cholesteatoma and family history: an international survey. Clin Otolaryngol. 2020;45:500–505.
40. Comerford M, Fogel R, Bailey JR, Chilukuri P, Chalasani N, Lammert CS. Leveraging social networking sites for an autoimmune hepatitis genetic repository: pilot study to evaluate feasibility. J Med Internet Res. 2018;20:e14.
41. Cotter K, Siskind CE, Sha SJ, Hanson-Kahn AK. Positive attitudes and therapeutic misconception around hypothetical clinical trial participation in the Huntington's disease community. J Huntingtons Dis. 2019;8:421–430.

42. Cotton AC, Bell RB, Jinnah HA. Expert opinion vs patient perspective in treatment of rare disorders: tooth removal in Lesch-Nyhan disease as an example. JIMD Rep. 2018;41:25–27.

43. Cragg S, Lafreniere K. Effects of Turner syndrome on women's self-esteem and body image. J Dev Phys Disabil. 2010;22:433–445.

44. Czajka CM, DiCaprio MR. What is the proportion of patients with multiple hereditary exostoses who undergo malignant degeneration? Clin Orthop. 2015;473:2355–2361.

45. DasMahapatra P, Raja P, Gilbert J, Wicks P. Clinical trials from the patient perspective: survey in an online patient community. BMC Health Serv Res. 2017;17:166.

46. David MP, et al. Perspectives of patients with dermatofibrosarcoma protuberans on diagnostic delays, surgical outcomes, and nonprotuberance. JAMA Netw Open. 2019;2:e1910413.

47. De Nardi L, Trombetta A, Ghirardo S, Genovese MRL, Barbi E, Taucar V. Adolescents with chronic disease and social media: a cross-sectional study. Arch Dis Child. 2020;105:744–748.

48. DiBenedetti DB, Coles TM, Sharma T, Pericleous L, Kulkarni R. Assessing patients' and caregivers' perspectives on stability of factor VIII products for haemophilia A: a web-based study in the United States and Canada. Haemophilia. 2014;20:e296–303.

49. Ekins S, Clark AM, Williams AJ. Open drug discovery teams: a chemistry mobile app for collaboration. Mol Inform. 2012;31:585–597.

50. El Abiad JM, et al. Natural history of Ollier disease and Maffucci syndrome: patient survey and review of clinical literature. Am J Med Genet A. 2020;182:1093–1103.

51. Eley SEA, McKechanie AG, Campbell S, Stanfield AC. Facilitating individuals and families affected by fragile X syndrome to participate in medication trials. J Intellect Disabil Res. 2020;64:864–874.

52. Elfassy C, Cachecho S, Snider L, Dahan-Oliel N. Participation among children with arthrogryposis multiplex congenita: a scoping review. Phys Occup Ther Pediatr. 2020;40:610–636.

53. Engwerda A, et al. The phenotypic spectrum of proximal 6q deletions based on a large cohort derived from social media and literature reports. Eur J Hum Genet. 2018;26:1478–1489.

54. Ferrara L, Morando V, Tozzi V. The transition of patients with rare diseases between providers: the patient journey from the patient perspective. Int J Integr Care. 2017;17:1–2.

55. Fogel R, Comerford M, Chilukuri P, Orman E, Chalasani N, Lammert C. Extrahepatic autoimmune diseases are prevalent in autoimmune hepatitis patients and their first-degree relatives: survey study. Interact J Med Res. 2018;7:e18.

56. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res. 2008;10:e15.

57. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res. 2011;13:e6.

58. Gelbard A, et al. Comparative treatment outcomes for patients with idiopathic subglottic stenosis. JAMA Otolaryngol Head Neck Surg. 2020;146:20–29.

59. Goel K, Tweet M, Olson TM, Maleszewski JJ, Gulati R, Hayes SN. Familial spontaneous coronary artery dissection: evidence for genetic susceptibility. JAMA Intern Med. 2015;175:821–826.

60. Gow J, Moffatt C, Blackport J. Participation in patient support forums may put rare disease patient data at risk of re-identification. Orphanet J Rare Dis. 2020;15:226.

61. Greenberg BM, Krishnan C, Harder L. New onset transverse myelitis diagnostic accuracy and patient experiences. Mult Scler Relat Disord. 2019;30:42–44.

62. Guthrie G, Davies RM, Fleming CK, Browning AC. YouTube as a source of information about retinitis pigmentosa. Eye (Lond). 2014;28:499–500.

63. Haik D, et al. The online support group as a community: a thematic content analysis of an online support group for idiopathic subglottic stenosis. Ann Otol Rhinol Laryngol. 2019;128:293–299.

64. Haller J, David MP, Lee NE, Shalin SC, Gardner JM. Impact of pathologist involvement in sarcoma and rare tumor patient support groups on Facebook: a survey of 542 patients and family members. Arch Pathol Lab Med. 2018;142:1113–1119.

65. Hamilton KV, et al. Exploring the medical and psychosocial concerns of adolescents and young adults with craniofacial microsomia: a qualitative study. Cleft Palate Craniofacial J. 2018;55:1430–1439.

66. Hawkins DM, Eide MJ, Lim HW. Perception of xeroderma pigmentosum support group members of xeroderma pigmentosum lifestyle impact. Photodermatol Photoimmunol Photomed. 2012;28:338–339.

67. Holton S, Fisher J, Button B, Williams E, Wilson J. Childbearing concerns, information needs and preferences of women with cystic fibrosis: an online discussion group. Sex Reprod Healthc. 2019;19:31–35.

68. Jacobs R, Boyd L, Brennan K, Sinha CK, Giuliani S. The importance of social media for patients and families affected by congenital anomalies: a Facebook cross-sectional analysis and user survey. J Pediatr Surg. 2016;51:1766–1771.

69. Janvier A, Farlow B, Wilfond BS. The experience of families with children with trisomy 13 and 18 in social networks. Pediatrics. 2012;130:293–298.

70. Johnson KJ, Mueller NL, Williams K, Gutmann DH. Evaluation of participant recruitment methods to a rare disease online registry. Am J Med Genet A. 2014;164A:1686–1694.

71. Judd-Glossy L, Ariefdjohan M, Curry S, Ketzer J, Pena A, Bischoff A. A survey of adults with anorectal malformations: perspectives on educational, vocational, and psychosocial experiences. Pediatr Surg Int. 2019;35:953–961.

72. Kauw D, Repping-Wuts H, Noordzij A, Stikkelbroeck N, Hermus A, Faber M. The contribution of online peer-to-peer communication among patients with adrenal disease to patient-centered care. J Med Internet Res. 2015;17:e54–e54.

73. Kelleher E, Giampietro PF, Moreno MA. Marfan syndrome patient experiences as ascertained through postings on social media sites. Am J Med Genet A. 2015;167A:2629–2634.

74. Kennedy J, et al. KAT6A Syndrome: genotype-phenotype correlation in 76 patients with pathogenic KAT6A variants. Genet Med. 2019;21:850–860.

75. Keogh RH, Bilton D, Cosgriff R, Kavanagh D, Rayner O, Sedgwick PM. Results from an online survey of adults with cystic fibrosis: accessing and using life expectancy information. PLoS One. 2019;14:e0213639.

76. Kirk S, Milnes L. An exploration of how young people and parents use online support in the context of living with cystic fibrosis. Health Expect. 2016;19:309–321.

77. Kleinendorst L, van den Heuvel LM, Henneman L, van Haelst MM. Who ever heard of 16p11.2 deletion syndrome? Parents' perspectives on a susceptibility copy number variation syndrome. Eur J Hum Genet. 2020;28:1196–1204.

78. Knepper KNK, Arrington MI. Social support and coping in an online PHPV forum. Tex Speech Commun J. 2020;44:19–31.

79. Krahe AM, Adams RD, Nicholson LL. Features that exacerbate fatigue severity in joint hypermobility syndrome/Ehlers-Danlos syndrome—hypermobility type. Disabil Rehabil. 2018;40:1989–1996.

80. Krischer J, et al. Experience with direct-to-patient recruitment for enrollment into a clinical trial in a rare disease: a web-based study. J Med Internet Res. 2017;19:e50.

81. Kühnle L, Mücke U, Lechner WM, Klawonn F, Grigull L. Development of a social network for people without a diagnosis (RarePairs): evaluation study. J Med Internet Res. 2020;22:e21849.

82. Liao BT, Busse J, Ender KL, Schechter WS. Exploring social media for patient perspectives of sickle cell disease. Pediatr Hematol Oncol. 2016;33:134–135.

83. Lowery EM, Afshar M, West N, Kovacs EJ, Smith B, Joyce C. Self-reported alcohol use in the cystic fibrosis community. J Cyst Fibros. 2020;19:84–90.

84. Macdonald M, Hudson A, Bladon A, Ratcliffe E, Blake K. Experiences in feeding and gastrointestinal dysfunction in children with CHARGE syndrome. Am J Med Genet A. 2017;173:2947–2953.

85. Mallett AJ, et al. Precision medicine diagnostics for rare kidney disease: Twitter as a tool in clinical genomic translation. Kidney Med. 2019;1:315–318.

86. Marques-da-Silva D, et al. An electronic questionnaire for liver assessment in congenital disorders of glycosylation (LeQCDG): a patient-centered study. JIMD Rep. 2019;44:55–64.

87. Mastboom MJ, Planje R, van de Sande MA. The patient perspective on the impact of tenosynovial giant cell tumors on daily living: crowdsourcing study on physical function and quality of life. Interact J Med Res. 2018;7:e4.

88. Mc Hugh R, Mc Feeters D, Boyda D, O'Neill S. Coping styles in adults with cystic fibrosis: implications for emotional and social quality of life. Psychol Health Med. 2016;21:102–112.

89. Meade O, Buchanan H, Coulson N. The use of an online support group for neuromuscular disorders: a thematic analysis of message postings. Disabil Rehabil. 2018;40:2300–2310.

90. Mehner LC, Jung JL, Maloney JA, McCourt EA. Ocular findings in pontine tegmental cap dysplasia. Cornea. 2020;39:1516–1519.

91. Might M, Might CC. What happens when N = 1 and you want plus 1? Prenat Diagn. 2017;37:70–72.

92. Myers KA, et al. The epileptology of Koolen-de Vries syndrome: electro-clinico-radiologic findings in 31 patients. Epilepsia. 2017;58:1085–1094.

93. Nakamura C, Bromberg M, Bhargava S, Wicks P, Zeng-Treitler Q. Mining online social network data for biomedical research: A comparison of clinicians' and patients' perceptions about amyotrophic lateral sclerosis treatments. J Med Internet Res. 2012;14:268–274.

94. Nicholas DB, Chahauver A, Brownstone D, Hetherington R, McNeill T, Bouffet E. Evaluation of an online peer support network for fathers of a child with a brain tumor. Soc Work Health Care. 2012;51:232–245.

95. Nieroba E, Larysz D. The internet as an additional source of information on rare illness. A Facebook support group for parents of children with craniosynostosis as a case study. Pol Ann Med. 2020;27:53–58.

96. Pacione M, Siskind CE, Day JW, Tabor HK. Perspectives on Spinraza (nusinersen) treatment study: views of individuals and parents of children diagnosed with spinal muscular atrophy. J Neuromuscul Dis. 2019;6:119–131.

97. Palser SC, Rayner OC, Leighton PA, Smyth AR. Perception of first respiratory infection with Pseudomonas aeruginosa by people with cystic fibrosis and those close to them: an online qualitative study. BMJ Open. 2016;6:e012303.

98. Peipert BJ, Goswami S, Helenowski I, Yount SE, Sturgeon C. Financial burden is associated with worse health-related quality of life in adults with multiple endocrine neoplasia type 1. Surgery. 2017;162:1278–1285.

99. Pemmaraju N, Gupta V, Mesa R, Thompson MA. Social media and myeloproliferative neoplasms (MPN)—focus on Twitter and the development of a disease-specific community: #MPNSM. Curr Hematol Malig Rep. 2015;10:413–420.

100. Pemmaraju N, Gupta V, Thompson MA, Lane AA. Social media and Internet resources for patients with blastic plasmacytoid dendritic cell neoplasm (BPDCN). Curr Hematol Malig Rep. 2016;11:462–467.

101. Pemmaraju N, Utengen A, Gupta V, Thompson MA, Lane AA. Blastic plasmacytoid dendritic cell neoplasm (BPDCN) on social media: #BPDCN-increasing exposure over two years since inception of a disease-specific Twitter community. Curr Hematol Malig Rep. 2018;13:581–587.

102. Pemmaraju N, Utengen A, Gupta V, Thompson MA, Lane AA. Analysis of first-year Twitter metrics of a rare disease community for blastic plasmacytoid dendritic cell neoplasm (BPDCN) on social media: #BPDCN. Curr Hematol Malig Rep. 2017;12:592–597.

103. Pemmaraju N, Thompson MA, Qazilbash M. Disease-specific hashtags and the creation of Twitter medical communities in hematology and oncology. Semin Hematol. 2017;54:189–192.

104. Perrone ME, Carmody D, Philipson LH, Greeley SAW. An online monogenic diabetes discussion group: supporting families and fueling new research. Transl Res J Lab Clin Med. 2015;166:425–431.

105. Peterson AM, Kallogjeri D, Spitznagel E, Chakinala MM, Schneider JS, Piccirillo JF. Development and validation of the nasal outcome score for epistaxis in hereditary hemorrhagic telangiectasia (NOSE HHT). JAMA Otolaryngol Head Neck Surg. 2020. https://doi.org/10.1001/jamaoto.2020.3040.

106. Pohlig F, et al. IT-based psychosocial distress screening in patients with sarcoma and parental caregivers via disease-specific online social media communities. Vivo Athens Greece. 2017;31:443–450.

107. Rowbotham NJ, et al. Adapting the James Lind Alliance priority setting process to better support patient participation: an example from cystic fibrosis. Res Involv Engagem. 2019;5:24.

108. Rowe CK, Shnorhavorian M, Block P, Ahn J, Merguerian PA. Using social media for patient-reported outcomes: a study of genital appearance and sexual function in adult bladder exstrophy patients. J Pediatr Urol. 2018;14:322.e1–322.e6.

109. Schumacher KR, et al. Fontan-associated protein-losing enteropathy and plastic bronchitis. J Pediatr. 2015;166:970–977.

110. Seidlin M, et al. Characterization and utilization of an international neurofibromatosis web-based, patient-entered registry: an observational study. PloS One. 2017;12:e0178639.

111. Sepehripour S, McDermott AL, Lloyd MS. Microtia and social media: patient versus physician perspective of quality of information. J Craniofac Surg. 2017;28:643–645.

112. Shalhub S, et al. Assessment of the information sources and interest in research collaboration among individuals with vascular Ehlers-Danlos syndrome. Ann Vasc Surg. 2020;62:326–334.

113. Smith S, et al. How can we relieve gastrointestinal symptoms in people with cystic fibrosis? An international qualitative survey. BMJ Open Respir Res. 2020;7. https://doi.org/10.1136/bmjresp-2020-000614.

114. Tsai J-H, Scheimann AO, McCandless SE, Strong TV, Bridges JFP. Caregiver priorities for endpoints to evaluate treatments for Prader-Willi syndrome: a best-worst scaling. J Med Econ. 2018;21:1230–1237.

115. Tweet MS, Gulati R, Aase LA, Hayes SN. Spontaneous coronary artery dissection: a disease-specific, social networking community-initiated study. Mayo Clin Proc. 2011;86:845–850.

116. Umbaugh HM, et al. Microtia and craniofacial microsomia: content analysis of facebook groups. Int J Pediatr Otorhinolaryngol. 2020;138. https://doi.org/10.1016/j.ijporl.2020.110301.

117. van der Heijden L, Piner SR, van de Sande MAJ. Pigmented villonodular synovitis: a crowdsourcing study of two hundred and seventy two patients. Int Orthop. 2016;40:2459–2468.

118. Van Tongel A, Piepers I, De Wilde L. The significance of the clavicle on shoulder girdle function. J Shoulder Elbow Surg. 2015;24:e255–259.

119. Victoria Diniz M, et al. Experience with the use of an online community on Facebook for Brazilian patients with gestational trophoblastic disease: netnography study. J Med Internet Res. 2018;20:e10897.

120. Walker AL, Gaydos LM, Farzan R, De Castro L, Jonassaint C. Social media discussions provide new insight about perceptions of hydroxyurea in the sickle cell community. Am J Hematol. 2019;94:E134–E136.

121. Walker KK. A content analysis of cognitive and affective uses of patient support groups for rare and uncommon vascular diseases: comparisons of may thurner, thoracic outlet, and superior mesenteric artery syndrome. Health Commun. 2015;30:859–871.

122. Wallace SE, Gilvary S, Smith MJ, Dolan SM. Parent perspectives of support received from physicians and/or genetic counselors following a decision to continue a pregnancy with a prenatal diagnosis of trisomy 13/18. J Genet Couns. 2018;27:656–664.

123. Wang T, Lund B. Categories of information need expressed by parents of individuals with rare genetic disorders in a Facebook community group: a case study with implications for information professionals. J Consum Health Internet. 2020;24:20–34.

124. Weidema ME, et al. Health-related quality of life and symptom burden of epithelioid hemangioendothelioma patients: a global patient-driven Facebook study in a very rare malignancy. Acta Oncol Stockh Swed. 2020;59:975–982.

125. Weijenborg PTM, Kluivers KB, Dessens AB, Kate-Booij MJ, Both S. Sexual functioning, sexual esteem, genital self-image and psychological and relational functioning in women with Mayer-Rokitansky-Küster-Hauser syndrome: a case-control study. Hum Reprod. 2019;34:1661–1673.

126. Wicks P, MacPhee GJA. Pathological gambling amongst Parkinson's disease and ALS patients in an online community (PatientsLikeMe.com): gambling in PD and ALS. Mov Disord. 2009;24:1085–1088.

127. Wicks P, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res. 2010;12:e19.

128. Wittmeier KD, et al. Identifying information needs for Hirschsprung disease through caregiver involvement via social media: a prioritization study and literature review. J Med Internet Res. 2018;20:e297.

129. Wittmeier K, et al. Analysis of a parent-initiated social media campaign for Hirschsprung's disease. J Med Internet Res. 2014;16:e288.

130. Yi JJ, et al. Impact of psychiatric comorbidity and cognitive deficit on function in 22q11.2 deletion syndrome. J Clin Psychiatry. 2015;76:1262–1270.

131. Zaid T, et al. Use of social media to conduct a cross-sectional epidemiologic and quality of life survey of patients with neuroendocrine carcinoma of the cervix: a feasibility study. Gynecol Oncol. 2014;132:149–153.

132. Pemmaraju N, Utengen A, Gupta V, Kiladjian J-J, Mesa R, Thompson MA. Rare cancers and social media: analysis of Twitter metrics in the first 2 years of a rare-disease community for myeloproliferative neoplasms on social media—#MPNSM. Curr Hematol Malig Rep. 2017;12:598–604.

133. Lotan I, Bacon T, Kister I, Levy M. Paroxysmal symptoms in neuromyelitis optica spectrum disorder: results from an online patient survey. Mult Scler Relat Disord. 2020;46. https://doi.org/10.1016/j.msard.2020.102578.

134. Brothers KB, Bennett RL, Cho MK. Taking an antiracist posture in scientific publications in human genetics and genomics. Genet Med. 2021;23:1004–1007.

135. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr. 2014;14:439–447.e4.

136. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res. 2017;19:e290.

137. Price J. What can big data offer the pharmacovigilance of orphan drugs? Clin Ther. 2016;38:2533–2545.

138. Han CJ, Lee YJ, Demiris G. Interventions using social media for cancer prevention and management: a systematic review. Cancer Nurs. 2018;41:E19–E31.

139. Pew Research Center. Demographics of social media users and adoption in the United States. https://www.pewresearch.org/internet/fact-sheet/social-media/. Accessed 14 May 2021.

140. Alshaikh F, Ramzan F, Rawaf S, Majeed A. Social network sites as a mode to collect health data: a systematic review. J Med Internet Res. 2014;16:e171.

141. Valdez RS, et al. Beyond traditional advertisements: leveraging Facebook's social structures for research recruitment. J Med Internet Res. 2014;16:e243.

142. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538:161–164.

143. de la Paz MP, Villaverde-Hueso A, Alonso V, János S, Zurriaga O, Pollán M, Abaitua-Borda I. Rare diseases epidemiology research. Adv Exp Med Biol. 2010;686:17-39. https://doi.org/10.1007/978-90-481-9485-8_2.

144. Centers for Disease Control and Prevention. Data & statistics on sickle cell disease. https://www.cdc.gov/ncbddd/sicklecell/data.html. Accessed 4 February 2021.

145. Titgemeyer SC, Schaaf CP. Facebook support groups for rare pediatric diseases: quantitative analysis. JMIR Pediatr Parent. 2020;3:e21694.
146. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-surveys (CHERRIES). J Med Internet Res. 2004;6:e34.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Conceptualization: M.C.H, H.K.T, E.G.M. Data curation: E.G.M., G.F., A.L.W. Data Analysis: E.G.M., J.L.Y. Investigation: A.L.W. Supervision: M.C.H., H.K.T. Writing—original draft: E.G.M., M.C.H. Writing—review & editing: E.G.M., M.C.H., H.K.T., J.L.J.

COMPETING INTERESTS

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41436-021-01273-z.

Correspondence and requests for materials should be addressed to M.C.H.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.