

# Perceptions and Responses to Diseases among Patients with Inflammatory Bowel Disease: Text Mining Analysis of Posts on a Japanese Patient Community Website

Eujin Lee<sup>a</sup> Hiroaki Tsuchiya<sup>a</sup> Hajime Iida<sup>a</sup> Katsumasa Nagano<sup>a</sup>  
Yoko Murata<sup>a</sup> Atsuo Maemoto<sup>b</sup>

<sup>a</sup>Medical Affairs Division, Janssen Pharmaceutical K.K., Tokyo, Japan; <sup>b</sup>Inflammatory Bowel Disease Center, Sapporo Higashi Tokushukai Hospital, Sapporo, Japan

## Keywords

Inflammatory bowel disease · Crohn's disease · Ulcerative colitis · Text mining · Community forum

## Abstract

**Introduction:** Patients with inflammatory bowel disease (IBD) are increasingly using online platforms to communicate with other patients and healthcare professionals seeking disease-related information and support. Free-text posts on these platforms could provide insights into patients' everyday lives, which could help improve patient care. In this proof-of-concept (POC) study, we applied text mining to extract patient needs from free-text posts on a community forum in Japan, holistically visualized the patients' perceptions and their connections, and explored the patient characteristic-dependent trends in the use of words. **Methods:** Free-text posts written between May 11, 2020 and May 31, 2022 on the community forum were retrieved and subjected to text mining analysis. Trends in the use of words were extracted from the posts for correspondence and co-occurrence network analyses using KH Coder open-source text mining software. **Results:** Seventy-four posts were analyzed. Using text mining methods, we

successfully extracted and visualized a variety of patient concerns and their connections. The correspondence and co-occurrence analyses revealed patient segment-dependent trends in the use of words. For example, patients with a disease duration of  $\leq 5$  years were more likely to use words related to emotions or their desire to change or quit their job, such as "anxiety" and "resignation." Patients with a disease duration of  $>10$  years were more likely to use words showing that they are finding ways to live with or accept their disease, and are getting used to the lifestyle, but some patients continued to experience worsening disease. **Conclusions:** We found that free-text posts on an IBD community forum can be a useful source of information to capture the wide variety of thoughts of patients. Text mining procedures can help visualize the relative importance of the topics identified from free-text posts. Our findings of this POC study will be useful for generating new hypotheses to better understand and address the needs of patients with IBD.

© 2024 The Author(s).  
Published by S. Karger AG, Basel

Results were presented as an abstract and oral presentation at the 11th Annual Meeting of the Asian Organization for Crohn's and Colitis 2023 (April 13–15; Busan, Korea).

## Introduction

Inflammatory bowel diseases (IBDs), such as Crohn's disease (CD) and ulcerative colitis (UC), are chronic diseases that have a significant detrimental impact on patient-centered outcomes, such as quality of life (QOL) [1–3]. In the last 10–20 years, social media and online community websites/forums have become a leading source of information for patients, who may post comments/information in a candid, unsolicited manner. The free-text comments posted on such platforms could provide important qualitative insight into the QOL and true unmet needs of patients, and unveil possible gaps between patients and their healthcare team [4]. Because patients with IBD are likely to be relatively young and familiar with community platforms, the value of community platforms is evident among the patients with IBD.

Text mining approaches have proven useful to extract patient needs, perceptions, and attitudes from free-text posts on social media [4–7], but few studies have applied such methods to patients with IBD. Prior studies that applied text mining methods to extract IBD patient insights from social media or forums primarily focused on themes across the sample population rather than themes in specific segments of the population because they lacked details on patient demographics. Therefore, trends across finely segmented patient groups are yet to be elucidated.

TOMONOWA IBD (<https://www.tomonowa.jp/ibd>) is an online community forum for supporting patients with IBD (UC or CD) in Japan. This forum allows patients to write free-text posts sharing their experiences with peers and to post questions for specialists. Unlike other online community forums, this platform collects some patient background information, including some mandatory and optional characteristics, and therefore allows patients to be segmented into groups based on their background characteristics for in-depth research.

Here, we performed a proof-of-concept (POC) study to investigate whether a text mining approach could be used to extract patient needs from the free-text posts on a Japanese IBD community forum. We sought to holistically capture a wide variety of patient concerns and observe links between them, and attempted to visualize the patient segment-dependent trends in the use of words.

## Materials and Methods

### *TOMONOWA IBD Forum*

TOMONOWA IBD is a website and Japanese community forum hosted by Janssen Pharma K.K. that was launched in 2019. People aged  $\geq 18$  years old

throughout Japan can register if they have CD/UC and are receiving treatment at a Japanese medical institution. Patients can find the website through organic searches, referral (by healthcare practitioners, family members, or friends), or via advertisements. Upon registration, members must provide a username, age range, disease type, and an email address. Members can (optionally) specify their gender, disease duration, and the name of their medical institution. All members must agree to the terms and conditions (<https://www.tomonowa.jp/s/term-of-use>) at registration; this includes consent for the reuse of their posts together with the anonymized membership registration information. The website does not provide medical advice; it only allows patients to share their experiences, ask general questions, or ask questions to professionals (limited to financial planners, psychological counselors, and career advisors). The website and forum are managed by a third-party vendor, which provided Janssen Pharma (the sponsor) with anonymized data.

We used posts created after May 11, 2020, when the community forum was opened to all patients with IBD in Japan, through to May 31, 2022, when there were approximately 196 active members. The posts were anonymized and could not be linked to specific members, but we could identify which posts were written by the same user according to the post identifiers that were not linked to the individual identifiable information.

The background characteristics of the members who created the eligible posts were retrieved and analyzed anonymously to compile information on their gender, age group, disease duration, and disease type (Table 1). These characteristics were used to define segments of the study population. The age groups 60–69, 70–79, and  $\geq 80$  years were combined due to the low number of responses. Data were not analyzed among segments by disease type due to the small number of patients with CD.

### *Text Mining and Data Analysis*

All posts were written in Japanese. The text mining procedures and analyses were performed on the Japanese text, and the results were translated into English for publication purposes. The text mining procedure and data analysis are described in detail in online supplementary File 1 (for all online suppl. material, see <https://doi.org/10.1159/000541837>) and summarized in Figure 1. Text mining comprised four primary elements: data cleansing, morphological analysis, QOL item creation, and data coding; the specific procedures are presented in online supplementary File 1 (Section 1).

**Table 1.** Demographic characteristics

Patient segment	Number of patients in segment <sup>a</sup>	Disease duration <sup>b</sup>				
		<1 year	2–5 years	6–9 years	≥10 years	Unknown
Overall	74 (100.0%)	15 (20.3%)	15 (20.3%)	4 (5.4%)	31 (41.9%)	9 (12.2%)
Disease type						
CD	19 (25.7%)	1 (5.3%)	2 (10.5%)	1 (5.3%)	10 (52.6%)	5 (26.3%)
UC	55 (74.3%)	14 (25.5%)	13 (23.6%)	3 (5.5%)	21 (38.2%)	4 (7.3%)
Gender						
Women	36 (48.6%)	10 (27.8%)	9 (25.0%)	2 (5.6%)	13 (36.1%)	2 (5.6%)
Men	31 (41.9%)	4 (12.9%)	5 (16.1%)	1 (3.2%)	17 (54.8%)	4 (12.9%)
Unknown	7 (9.5%)	1 (14.3%)	1 (14.3%)	1 (14.3%)	1 (14.3%)	3 (42.9%)
Age						
18–29 years	16 (21.6%)	3 (18.8%)	6 (37.5%)	2 (12.5%)	2 (12.5%)	3 (18.8%)
30–39 years	13 (17.6%)	3 (23.1%)	2 (15.4%)	1 (7.7%)	6 (46.2%)	1 (7.7%)
40–49 years	20 (27.0%)	3 (15.0%)	4 (20.0%)	0	11 (55.0%)	2 (10.0%)
50–59 years	22 (29.7%)	6 (27.3%)	2 (9.1%)	1 (4.5%)	10 (45.5%)	3 (13.6%)
≥60 years	3 (4.1%)	0	1 (33.3%)	0	2 (66.7%)	0
Women						
18–29 years	11 (14.9%)	2 (18.2%)	4 (36.4%)	2 (18.2%)	2 (18.2%)	1 (9.1%)
30–39 years	6 (8.1%)	2 (33.3%)	2 (33.3%)	0	2 (33.3%)	0
40–49 years	6 (8.1%)	1 (16.7%)	2 (33.3%)	0	2 (33.3%)	1 (16.7%)
50–59 years	13 (17.6%)	5 (38.5%)	1 (7.7%)	0	7 (53.8%)	0
≥60 years	0	0	0	0	0	0
Men						
18–29 years	4 (5.4%)	1 (25.0%)	1 (25.0%)	0	0	2 (50.0%)
30–39 years	7 (9.5%)	1 (14.3%)	0	1 (14.3%)	4 (57.1%)	1 (14.3%)
40–49 years	12 (16.2%)	1 (8.3%)	2 (16.7%)	0	9 (75.0%)	0
50–59 years	5 (6.8%)	1 (20.0%)	1 (20.0%)	0	2 (40.0%)	1 (20.0%)
≥60 years	3 (4.1%)	0	1 (33.3%)	0	2 (66.7%)	0

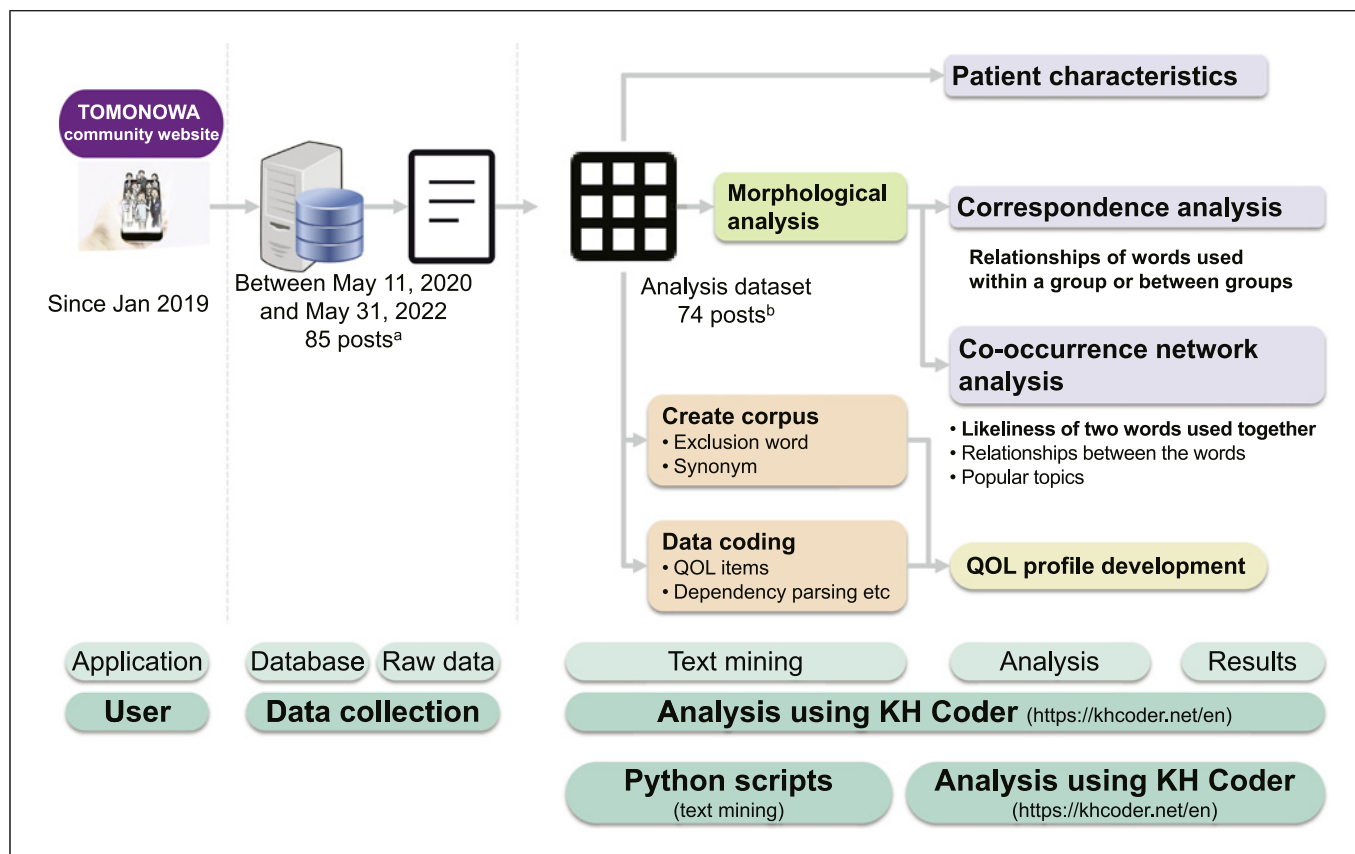
Values are *n* (%). CD, Crohn's disease; UC, ulcerative colitis. <sup>a</sup>Percentages were calculated using the total number (74) as the denominator. <sup>b</sup>Percentages were calculated using the number of patients in each segment as the denominator.

Following text mining, we performed correspondence analysis and co-occurrence network analysis, as described in online supplementary File 1 (Section 2). Correspondence analysis depicts the relationships between words and segments in two-dimensional plots, to visualize the trends in data without assessing the statistical significance. The relationships between the category (patient segment) and the words used are assessed based on the distance and the direction each word extends from the center and whether they extend in a similar direction to a specific patient segment; items further away from the center in the same direction show stronger associations than items closer to the center [8]. Items located close to the center are relevant to all segments. The top 60 words that appeared at least five times across the whole sample population and showed highly distinctive

patterns with large differences in their occurrence are included. The eigenvalues on the *x* and *y* dimensions represent the relative contribution of each dimension to the observed data; the sum of both values indicates the cumulative contribution.

Co-occurrence network analysis was applied to each patient segment to identify the trending topics within each segment and visualize which words/topics may be related. Networks are constructed to assess the likelihood of two words used together in the same post from the Jaccard coefficient, which is widely used to assess relationship units (e.g., species, diseases). Terms that occurred more than three times and included in the top 60 strongest co-occurrences are displayed.

Where possible, we checked the context in which each term was used to help us understand the meaning of the



**Fig. 1.** Overview of procedures. <sup>a</sup>Of 85 posts, 11 were excluded from the data analyses for the following reasons: 8 posts were created by family members/friends, 2 posts were related to website operation, and 1 was posted by a patient living outside Japan. <sup>b</sup>Comprising 66 initial posts and 8 responses; the responses were handled as unique posts.

terms included in the correspondence analysis and co-occurrence networks. The interpretations of all analyses were reviewed by a physician (author Atsuo Maemoto).

Finally, we developed QOL profiles consisting of QOL categories and sentiment scores to identify which QOL category had the most negative effect on the patients and to visualize the unmet needs in a radar chart form. To thoroughly cover the topics mentioned and to extract the true unmet needs of the patients, all posts were reviewed to create a list of QOL items. The items were selected based on those used in IBD Disk [9] and included derived items. Any words that were considered likely to appear in the posts were added to the list of QOL items. IBD Disk was used as the main source of terminology because of its value in assessing the daily life burden or disability associated with IBD and for assessing patient acceptability [10–12]. IBD Disk was developed from the IBD Disability Index, which showed good correlations with Short-Form 36 [13–15]. All QOL items were reviewed by a physician (author Atsuo Maemoto). All

posts were checked for the presence of QOL items. Further details are described in online supplementary File 1 (Section 3).

## Results

### *Characteristics of the Analyzed Posts and Patient Segmentation*

We retrieved 85 posts written between May 11, 2020 and May 31, 2022, of which 74 were included in the analyses (Fig. 1). These posts were written by 63 patients according to the post-identification labels that were not linked to the individual identifiable information. The total number of extracted sentences was 492, and the mean length of the 74 posts (in Japanese) was 185.4 characters. Most of the posts were written in paragraphs; few of the posts comprised a single sentence. Overall, 38 posts were questions for professionals (9 to financial advisors, 15 to career advisors, and 14 to

psychologists), 34 were comments or experiences shared by the patients, and 2 were general questions.

Table 1 shows the characteristics of the members who wrote the analyzed posts. These categories were used for patient segmentation. The age and gender distributions broadly reflect the general demographic trends of patients with IBD in Japan [16, 17].

### *Trends by Disease Duration*

#### Correspondence Analysis

Figure 2 shows the correspondence analysis for trends by disease duration (i.e., <1, 2–5, 6–9, and ≥10 years). The number of posts for the segment 6–9 years was too small for meaningful analysis of this segment.

#### Disease Duration <1 Year

The results of the correspondence analysis showed that the words extended in the direction of the <1 year segment and further away from the center, meaning that they are strongly associated with this segment, included “pain” and “diarrhea,” which referred to the presence of these symptoms. “Postoperative” found in this direction was linked to postoperative symptoms.

#### Disease Duration 2–5 Years and 6–9 Years

Terms that extended in the direction of the segments with relatively shorter disease duration (<1 year and 2–5 years) on the top right hand included “anxiety/anxious,” “hard,” and “job change.” In context, the patients were “anxious” about the unknown future, the disease, family-related concerns, or for non-specified reasons. The patients found their situation to be “hard” and they wished to change or had previously changed their “job.”

Terms that extended in the direction of the 2–5 years segment were “resignation,” “necessary,” “mind,” and “live/life.” “Resignation” was used in a similar way to “job change.” Patients mentioned that they started to care (“mind”) less or more about their disease. In addition, although “live/life” was located slightly closer to the center, this word was showing the start of the changes in the patients’ way of thinking or their lifestyles. For example, patients mentioned that they were starting to “live” normally or were trying to “live” a regular life. “Necessary” was used in varying contexts, such as, “*is it necessary to gather information?*” and “*hospitalization was necessary.*”

The main term extending in the direction of the 2–5 years and 6–9 years segments was “self.” This term was used in various ways, including “*I find myself (self) not being able to...*,” “*I am (self) not the only one,*” “*why do*

*I (self) have to go through this?*”. This term indicated that patients were more self-aware/conscious and were reflecting on themselves after some years of experience with the disease.

The only term strongly associated with the 6–9 years segment was “follow-up,” which was repeatedly used by 1 patient. When we excluded this term from the analysis, the other relationships between words and segments remained. The patient used this term to express regret about not taking time to learn about what follow-up entails and resisting the treatment or follow-up processes. As a post hoc analysis, we combined the 2–5 and 6–9 years segments into a single category, and the results were essentially comparable to those obtained for the separate categories (data not shown).

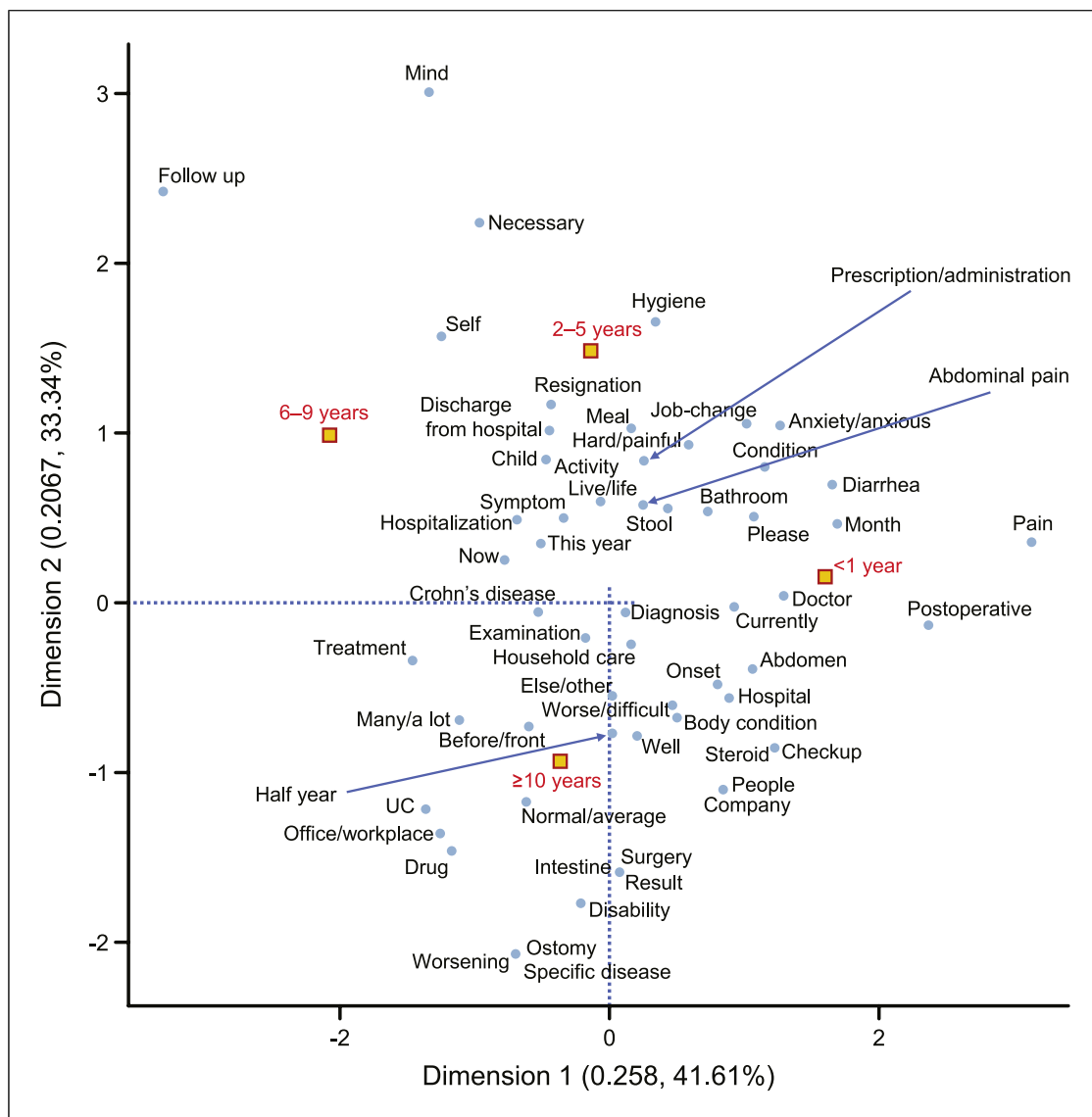
#### Disease Duration ≥10 Years

Terms associated with a disease duration of ≥10 years were located in the lower half of the plot. Some terms used in this segment were “normal/average,” “office/workplace,” “disability,” “surgery,” “specific disease,” and “ostomy.” Patients used “normal/average” together with salary, life, and person to share that they were living a normal life with a normal salary, like people who are not ill. However, “worsening” also appeared and was used in relation to symptoms. Although some patients were starting to live a normal life, others were still experiencing worsening of symptoms.

The term “office/workplace” was mostly used in relation to the understanding of workplace colleagues “*who do (or do not) understand the situation.*” “Disability” was used as in the context of a disability basic pension, disability certificate, and disability grade. Patients were asking about eligibility for disability status, which would allow them to qualify for financial support provided by the National Subsidy System. Some patients wanted to know about the system because they had developed complications or were concerned about the side effects that may require additional treatment and therefore additional cost. “Specific disease” was also used in the same context, to inquire about eligibility for a specific disease certificate to receive support. Patients in this group had experienced “ostomy” or removed “ostomy.”

#### Co-occurrence Network Analysis

The co-occurrence networks with subgraphs showing clusters of closely connected words in disease duration segments (<1 year, 2–5 years, and ≥10 years) are presented in online supplementary Figure 1 (online suppl. File 2). Co-occurrence networks were not obtained for



**Fig. 2.** Correspondence analysis showing the use of words among patients segmented by disease duration ( $n = 65$ ). Words that appeared at least five times are presented. The co-occurrence networks for the trends in words that appeared at least three times for patients with a disease duration of <1 year, 2–5 years, and  $\geq 10$  years are shown in online supplementary Figure 1 (online suppl. File 2). Dimensions 1 and 2 accounted for 41.61% and 33.34%, respectively, of the model (total contribution: 74.95%). UC, ulcerative colitis.

disease duration 6–9 years due to the small sample size. The co-occurrence network for <1 year showed that the words used by the patients in this segment were related to the presence of symptoms (“pain,” “diarrhea,” “symptoms”) and resulting struggles with continuation of work, such as needing time off work (“job”) because of their poor “body condition.” Anxiety was also mentioned by the patients.

The co-occurrence network for the 2–5 years segment included terms associated with multiple hospitalizations

(“hospitalization,” “discharge from hospital,” “necessary”) that forced them to quit their job (“job,” “part-time job”), and the presence of symptoms making the environment inconvenient, including bathrooms located far from them (“gas,” “stool,” “bathroom”). “Mind” was used the same way as seen in correspondence analysis. “Dental” and “hygiene” were used repeatedly by 1 patient, who was diagnosed soon after qualifying as a dental hygienist. Excluding these words from the analysis did not affect the relationships

between words, although the list of the most frequent words changed slightly.

The co-occurrence network for the  $\geq 10$  years segment indicated that there were broadly two types of situations for patients in this group, as seen in the correspondence analysis. The first group was still experiencing worsening symptoms and undergoing surgical operations (“worsening,” “ostomy,” “resection,” “bloody stool,” “many”). The second group was adapting to daily life and engaging with work, although still facing some challenges related to their colleagues’ understanding (“normal/average,” “live/life,” “job,” “office/workplace,” “understanding,” “body condition,” “worse/difficult”). “Applicable” and “certificate” were related to eligibility for government support including the disability basic pension.

#### Summary of Disease Duration-Dependent Trends

These analyses suggest that patients with disease duration  $< 1$  year were struggling with symptoms and feeling overwhelmed with negative emotions because of the unknown future. After 2–5 years, they were considering a job change or resignation, possibly because they found it difficult to live with the disease in the setting. Patients also started to either care less or care more about the disease and find ways to live normally. They also reflected on themselves. Patients with a disease duration of  $\geq 10$  years seemed to experience one of the following: they had started to live a normal life or were still experiencing worsening symptoms. Those who had managed to live normally had possibly shifted their attention to chronic conditions affecting day-to-day life at their workplace, and they felt that their colleagues’ understanding was important. Patients in this segment were likely to be older than those in the other disease duration segments and more likely to be susceptible to complications and side effects. Accordingly, they were thinking about the additional costs of treating complications/side effects, had financial concerns, and were seeking a disability certificate to qualify for the National Subsidy System.

#### Trends by Age

We next analyzed the age-related trends among four segments (18–29, 30–39, 40–49, and 50–59 years). In the correspondence analysis (Fig. 3), dimension 1 differentiates the patients according to their age, with younger patients (18–29 years) located on the left hand side of the graph and older patients on the right hand side of the graph. The co-occurrence networks for the four segments are depicted in online supplementary Figure 2a–d (online suppl. File 2). In summary, patients aged 18–20 years had

concerns related to age-specific life events including school, finding a job, and having a child, and they were emotionally affected. The older segments were asking about the type of governmental support and discussing their symptoms and work.

#### Trends by Gender

The correspondence analysis of the posts written by men and women is shown in Figure 4. The co-occurrence networks for men and women are shown in online supplementary Figure 3a, b (online suppl. File 2). Men were likely to use terms related to symptoms and factors that made their life practically inconvenient or events happening in their life. In co-occurrence networks, the subgraphs related to work included different topics for men (subgraphs 3 and 7 in online suppl. Fig. 3a) and women (subgraphs 1 and 2 in online suppl. Fig. 3b). Women were more likely to mention their struggles with work/housework and acknowledge the importance of coworkers/families’ understandings. Women were also concerned about medications and reflecting on themselves.

#### QOL Profile Development

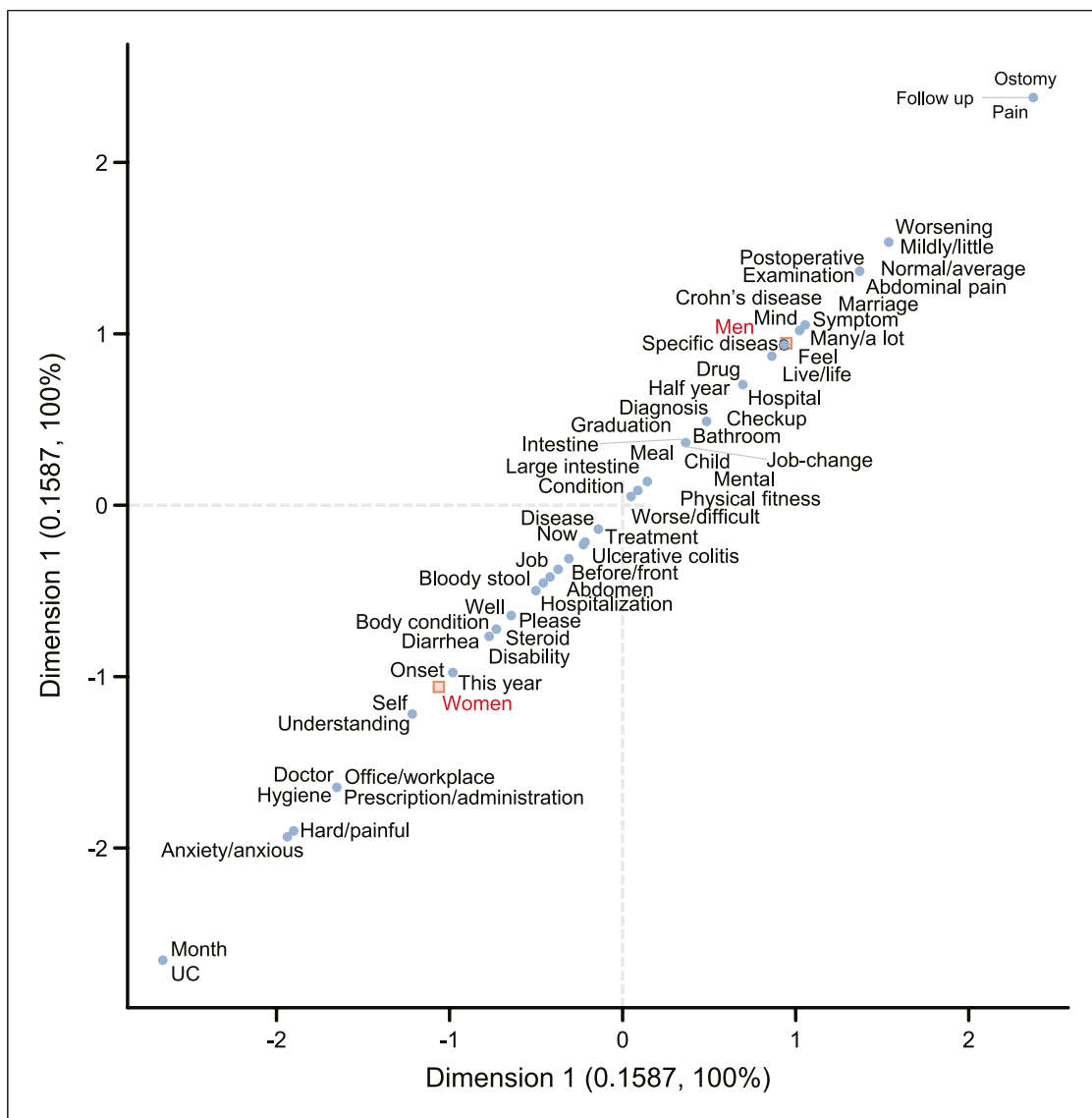
We also developed a list of QOL categories/items (online suppl. Table 1 [File 3]), which comprised 49 items in two classifications and six categories. The first classification, QOL, comprised “abdominal symptoms,” “systemic symptoms,” “emotions,” and “social activities.” The second classification, “other,” comprised “workplace” and “treatment.” Individual QOL profiles were subsequently prepared for the 74 posts, and the average profile is shown in online supplementary Figure 4 (online suppl. File 2) to visualize the degree of the negative sentiment in each QOL category. Larger scores correspond to greater negative sentiments. The greatest negative influences were systemic symptoms and social activity, followed by emotions and workplace, while abdominal symptoms and treatment had a weaker negative influence on QOL.

## Discussion

In this POC study, we performed text mining of posts on an online patient community forum to obtain insight into the true unmet needs of patients with IBD. While text mining methods have been used in other settings, they have rarely been applied to IBD. Furthermore, previous studies generally investigated themes across the patient cohort, and trends among patient segments remain to be elucidated.





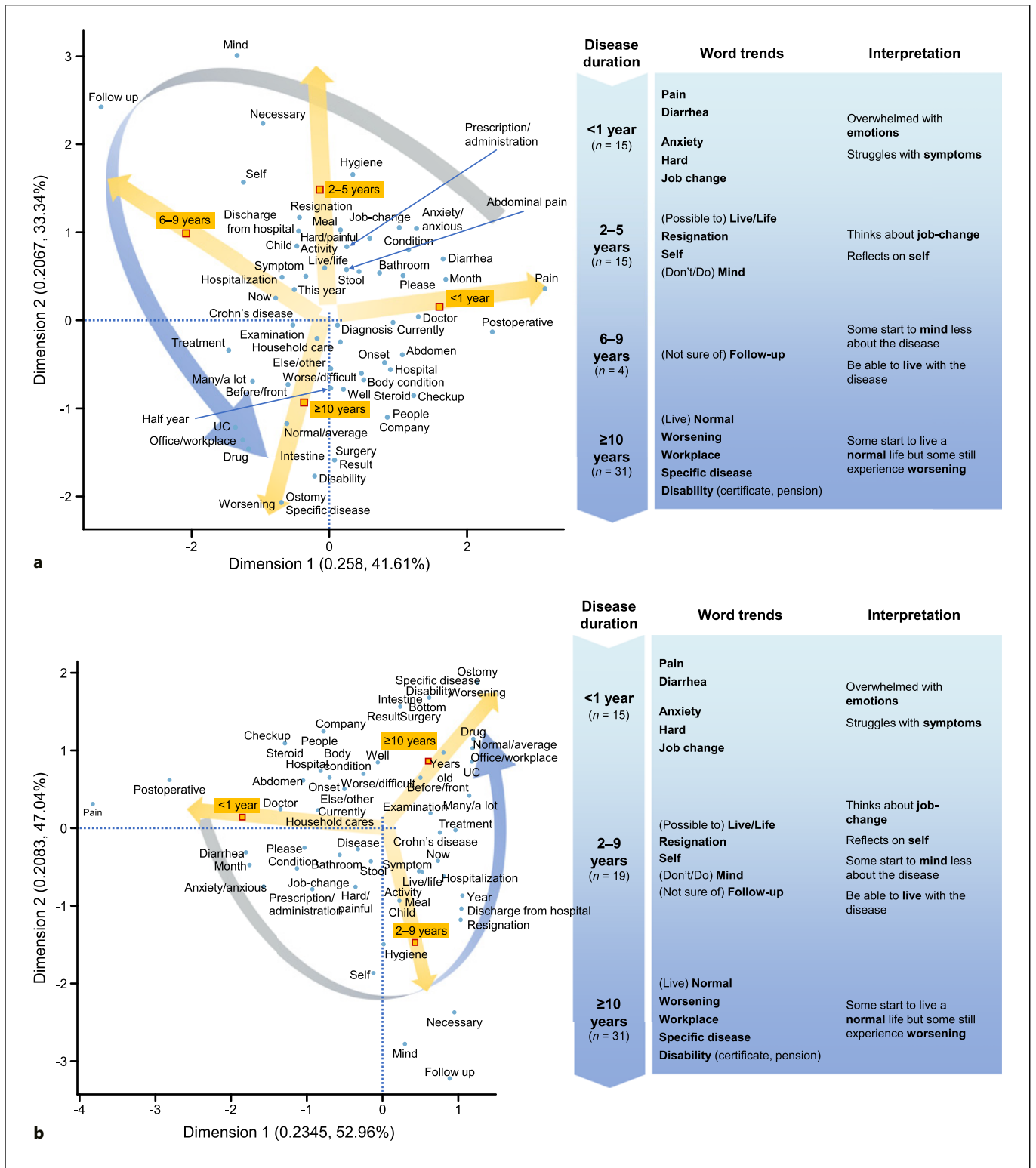


**Fig. 4.** Correspondence analysis showing the use of words among patients segmented by gender ( $n = 67$ ). Words that appeared at least five times are presented. The co-occurrence networks for the trends in words that appeared at least three times in men and women are shown in online supplementary Figure 3 (online suppl. File 2). Dimension 1 (on both axes) accounted for 100.00% of the model. UC, ulcerative colitis.

community forum posts can be useful to capture the psychological impact of IBD on patients. They reported a variety of emotions, including gratitude, anxiety/fear, empathy, sadness/depression, shame/guilt, and loneliness. Meanwhile, Pérez-Pérez et al. [23] determined the sentiment of Tweets on X (formerly Twitter) written by patients with IBD and found that they were frequently related to negative emotions, possibly because patients tend to use social media to express their emotions when there is no cure for the disease. Additionally, Keller et al.

[25] used social media posts to understand how patients in their reproductive period make decisions regarding medication use during pregnancy and perceived safety of IBD therapies. However, because these studies utilized social media posts that lacked detailed demographic information, their insight into the needs of patient segments was limited.

In this POC study, we obtained some preliminary insights and knowledge about the differences in perceptions among patients divided into segments by disease



**Fig. 5.** Evolution of terms expressed by patients over the course of IBD for four segments of disease duration (**a**; <1 year, 2–5 years, 6–9 years, ≥10 years) or in a post hoc analysis using three segments of disease duration (**b**; <1 year, 2–9 years, ≥10 years). UC, ulcerative colitis.

duration, age, and gender. In particular, for disease duration, we found that the terms evolved with increasing disease duration (Fig. 5a), with some overlap in the analysis by age. The findings were essentially unchanged when patients were divided into three disease duration segments (1 year, 2–6 years, and  $\geq 10$  years; Fig. 5b). Patients with a disease duration of  $< 1$  year mostly focused on symptom-related words, while other terms related to emotions and job change extended in the direction of the relatively short ( $< 1$  year) or intermediate (2–5 years or 2–9 years) disease duration segments. The intermediate segment also used terms such as “resignation” concerning changing their employer. These findings suggest that, in earlier years, patients may struggle with their symptoms, become overwhelmed with emotions, especially anxiety, and find it difficult to live with the disease. With longer disease duration, the patients were contemplating changing their job or resigning from work for a while due to their symptoms. Some patients with a disease duration of  $\geq 10$  years appeared to have learned to live with the disease and were paying attention to their work environment, but some were still experiencing worsening symptoms. Patients with this disease duration have also experienced surgery and were seeking information on eligibility for disease certification/disability benefits.

We also performed correspondence analysis after dividing the posts according to the purpose of the visit, whether they were questions for professionals or described patient experiences, and observed duration-dependent trends. Most of the trends overlapped with those obtained in the analyses described in the results, although words related to living a normal life and getting used to the disease or other people’s understandings were scattered elsewhere in posts related to patient experiences (data not shown).

Overall, these findings indicate that the perspectives and needs of patients evolve and differ among patient segments. Therefore, patient communication and support may need to be tailored for individual patients, considering their background characteristics, putting weight on different topics.

As another aspect of this study, we sought to identify and visualize which QOL categories were most negatively affected. For this, all eligible posts were reviewed to thoroughly cover the topics mentioned, and we identified 49 QOL items that could be split into two major classifications (QOL and “other”) and six categories (online suppl. Table 1 [File 3]). Many of the items grouped as “other” were related to environmental/external factors that are not usually covered in QOL questionnaires but were mentioned in the patients’ posts. From the list of identified items, we may

infer that patient posts could be useful to capture a wide variety of patient concerns in a real-world setting that may not necessarily be captured by conventional questionnaires.

## Limitations

There are some limitations that we should mention. In particular, the use of posts on an online patient community forum carries a high risk of selection bias because the members who write the posts are actively seeking information and support from their peers and professionals, especially regarding careers, finance, and psychological issues due to the nature of this website. Additionally, the sample size was relatively small (74 posts), although we consider that the sample size was sufficient for this POC of applying text mining to community forum posts. Therefore, the study cohort may not reflect the wider population of patients with IBD in Japan. This also limited finer segmentation of the patients, and some segments contained few patients, limiting the possible insight that could be gained. In the future, we hope to perform larger studies using this approach in order to gain more detailed insight into the varying needs across segments of patients with IBD.

## Conclusions

Here, we applied a text mining method to analyze posts published on an IBD community forum. While the results of this POC study may be considered preliminary, we have shown that it is possible to capture the broad potential needs of patients with IBD and to visualize the links between topics. This information could help fill the existing gaps between patients and physicians. We have also demonstrated that the needs vary among patient segments. This implies that the clinical and psychological support may need to be tailored for individual patients, taking into account their background characteristics or other clinically relevant factors. Although the insights uncovered in this POC study should be considered preliminary due to the nature of the study, our findings will be useful for generating new hypotheses to better understand and address the needs of patients with IBD.

## Acknowledgments

The authors thank Yoshiji Yamoto (Janssen Pharmaceutical K.K.) and Tomoyoshi Ishikawa (Deloitte Tohmatsu LHit Data Visionary Co., Ltd.) for their contributions to the analysis and

interpretation of the data. The authors thank Deloitte Tohmatsu LHit Data Visionary Co., Ltd. for supporting the data analysis and Nicholas D. Smith (EMC K.K.) for medical writing support.

### Statement of Ethics

Under the terms and conditions for use of TOMONOWA IBD, all members agree that their posts may be subject to reproduction, quotation, disclosure, provision, editing, translation, publication, or distribution, for example, without notice or consent from the member. As this study involved secondary analysis of anonymized free-text posts, without involving human subjects, patient consent was not required in accordance with local or national guidelines. The study was approved by an independent ethics committee, the Non-Profit Organization MINS Research Ethics Committee (MINS-REC-220219; <http://www.npo-mins.com/>). We followed the Standards for Reporting Qualitative Research in the preparation of the manuscript.

### Conflict of Interest Statement

Eujin Lee, Hiroaki Tsuchiya, Hajime Iida, Katsumasa Nagano, and Yoko Murata are employees of Janssen Pharmaceutical K.K./ Johnson & Johnson K.K. and own stock in Johnson & Johnson K.K. Atsuo Maemoto received support for this article from Janssen Pharmaceutical K.K.

### References

- 1 Al Khoury A, Balram B, Bessissow T, Afif W, Gonczi L, Abreu M, et al. Patient perspectives and expectations in inflammatory bowel disease: a systematic review. *Dig Dis Sci.* 2022;67(6):1956–74. <https://doi.org/10.1007/s10620-021-07025-y>
- 2 Armuzzi A, Liguori G. Quality of life in patients with moderate to severe ulcerative colitis and the impact of treatment: a narrative review. *Dig Liver Dis.* 2021;53(7):803–8. <https://doi.org/10.1016/j.dld.2021.03.002>
- 3 Le Berre C, Ananthakrishnan AN, Danese S, Singh S, Peyrin-Biroulet L. Ulcerative colitis and Crohn's disease have similar burden and goals for treatment. *Clin Gastroenterol Hepatol.* 2020;18(1):14–23. <https://doi.org/10.1016/j.cgh.2019.07.005>
- 4 Correia RB, Wood IB, Bollen J, Rocha LM. Mining social media data for biomedical signals and health-related behavior. *Annu Rev Biomed Data Sci.* 2020;3:433–58. <https://doi.org/10.1146/annurev-biodatasci-030320-040844>
- 5 Cammel SA, De Vos MS, van Soest D, Hettne KM, Boer F, Steyerberg EW, et al. How to automatically turn patient experience free-text responses into actionable insights: a natural language programming (NLP) approach. *BMC Med Inform Decis Mak.* 2020;20(1):97. <https://doi.org/10.1186/s12911-020-1104-5>
- 6 Smink W, Sools AM, van der Zwaan JM, Wiegersma S, Veldkamp BP, Westerhof GJ. Towards text mining therapeutic change: a systematic review of text-based methods for therapeutic change process research. *PLoS One.* 2019;14(12):e0225703. <https://doi.org/10.1371/journal.pone.0225703>
- 7 Wagland R, Recio-Saucedo A, Simon M, Bracher M, Hunt K, Foster C, et al. Development and testing of a text-mining approach to analyse patients' comments on their experiences of colorectal cancer care. *BMJ Qual Saf.* 2016;25(8):604–14. <https://doi.org/10.1136/bmjqs-2015-004063>
- 8 Sourial N, Wolfson C, Zhu B, Quail J, Fletcher J, Karunanathan S, et al. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. *J Clin Epidemiol.* 2010;63(6):638–46. <https://doi.org/10.1016/j.jclinepi.2009.08.008>
- 9 Ghosh S, Louis E, Beaugerie L, Bossuyt P, Bouguen G, Bourreille A, et al. Development of the IBD disk: a visual self-administered tool for assessing disability in inflammatory bowel diseases. *Inflamm Bowel Dis.* 2017;23(3):333–40. <https://doi.org/10.1097/MIB.0000000000001033>
- 10 Naftali T, Richter V, Mari A, Khoury T, Shirin H, Broide E. The inflammatory bowel disease disk application: a platform to assess patients' priorities and expectations from treatment. *J Dig Dis.* 2021;22(10):582–9. <https://doi.org/10.1111/1751-2980.13045>
- 11 Singh A, Gupta YK, Dhaliwal AS, Kahlon BK, Bansal V, Mahajan R, et al. Inflammatory bowel disease (IBD)-disk accurately predicts the daily life burden and parallels disease activity in patients with IBD. *Intest Res.* 2023;21(3):375–84. <https://doi.org/10.5217/ir.2022.00037>
- 12 Tadbiri S, Nachury M, Bouhnik Y, Serrero M, Hébuterne X, Roblin X, et al. The IBD-disk is a reliable tool to assess the daily-life burden of patients with inflammatory bowel disease. *J Crohns Colitis.* 2021;15(5):766–73. <https://doi.org/10.1093/ecco-jcc/jjaa244>
- 13 Allen PB, Kamm MA, Peyrin-Biroulet L, Studd C, McDowell C, Allen BC, et al. Development and validation of a patient-reported disability measurement tool for patients with inflammatory bowel disease. *Aliment Pharmacol Ther.* 2013;37(4):438–44. <https://doi.org/10.1111/apt.12187>
- 14 Gower-Rousseau C, Sarter H, Savoye G, Tavernier N, Fumery M, Sandborn WJ, et al. Validation of the inflammatory bowel disease disability index in a population-based cohort. *Gut.* 2017;66(4):588–96. <https://doi.org/10.1136/gutjnl-2015-310151>

### Funding Sources

This study and medical writing support were funded by Janssen Pharmaceutical K.K.

### Author Contributions

Conceptualization and writing – review and editing: all authors. Data curation: Eujin Lee, Hiroaki Tsuchiya, Hajime Iida, and Katsumasa Nagano. Formal analysis, software, and validation: Hiroaki Tsuchiya. Funding acquisition: Yoko Murata. Investigation and methodology: Eujin Lee, Hiroaki Tsuchiya, Hajime Iida, Katsumasa Nagano, and Atsuo Maemoto. Project administration and writing – original draft: Eujin Lee. Supervision: Atsuo Maemoto. Visualization: Eujin Lee and Hiroaki Tsuchiya. The sponsor was involved in study design, the collection, analysis, and interpretation of the data, and reviewed the manuscript.

### Data Availability Statement

Although these data are not currently available for public sharing to protect patient privacy, requests for sharing data can be sent to the corresponding author (Hiroaki Tsuchiya; email: [htsuchi7@its.jnj.com](mailto:htsuchi7@its.jnj.com)) and reasonable requests will be evaluated on an individual basis.

- 15 Soares JB, Pereira R, Costa JM, Arroja B, Gonçalves R. The inflammatory bowel disease-disability index: validation of the Portuguese version according to the COSMIN checklist. *Eur J Gastroenterol Hepatol*. 2016;28(10):1151–60. <https://doi.org/10.1097/MEG.0000000000000701>
- 16 Okabayashi S, Kobayashi T, Hibi T. Inflammatory bowel disease in Japan-is it similar to or different from Westerns? *J Anus Rectum Colon*. 2020;4(1):1–13. <https://doi.org/10.23922/jarc.2019-003>
- 17 Japan Intractable Diseases Information Center. The number of recipients of specified medical expenses (designated intractable diseases), by age group and target disease. (accessed February 12, 2024) (In Japanese). Available from: <https://www.nanbyou.or.jp/entry/5354>
- 18 Reich J, Guo L, Groshek J, Weinberg J, Chen W, Martin C, et al. Social media use and preferences in patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2019;25(3):587–91. <https://doi.org/10.1093/ibd/izy280>
- 19 Reich J, Guo L, Hall J, Tran A, Weinberg J, Groshek J, et al. A survey of social media use and preferences in patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2016;22(11):2678–87. <https://doi.org/10.1097/MIB.0000000000000951>
- 20 Timms C, Forton DM, Poullis A. Social media use in patients with inflammatory bowel disease and chronic viral hepatitis. *Clin Med*. 2014;14(2):215–6. <https://doi.org/10.7861/clinmedicine.14-2-215>
- 21 Guo L, Reich J, Groshek J, Farraye FA. Social media use in patients with inflammatory bowel disease. *Inflamm Bowel Dis*. 2016;22(5):1231–8. <https://doi.org/10.1097/MIB.0000000000000713>
- 22 Groshek J, Basil M, Guo L, Parker Ward S, Farraye FA, Reich J. Media consumption and creation in attitudes toward and knowledge of inflammatory bowel disease: web-based survey. *J Med Internet Res*. 2017;19(12):e403. <https://doi.org/10.2196/jmir.7624>
- 23 Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. *J Med Internet Res*. 2019;21(8):e12610. <https://doi.org/10.2196/12610>
- 24 Lerrigo R, Coffey JTR, Kravitz JL, Jadhav P, Nikfarjam A, Shah NH, et al. The emotional toll of inflammatory bowel disease: using machine learning to analyze online community forum discourse. *Crohns Colitis* 360. 2019;1(2):otz011. <https://doi.org/10.1093/crocol/otz011>
- 25 Keller MS, Mosadeghi S, Cohen ER, Kwan J, Spiegel BMR. Reproductive health and medication concerns for patients with inflammatory bowel disease: thematic and quantitative analysis using social listening. *J Med Internet Res*. 2018;20(6):e206. <https://doi.org/10.2196/jmir.9870>