Identification of early symptoms of endometriosis through the analysis of online social networks: A social media study.

Mathilde Fruchart, Fatima El Idrissi, Antoine Lamer, Karim-Ali Belarbi, Mohamed Lemdani, Djamel Zitouni, Benjamin Guinhouya

▶ **To cite this version:**

HAL Id: hal-04541899
https://hal.univ-lille.fr/hal-04541899

Submitted on 11 Apr 2024

Original research

DIGITAL HEALTH

Identification of early symptoms of endometriosis through the analysis of online social networks: A social media study

Mathilde Fruchart[1,2], **Fatima El Idrissi**[1,3], **Antoine Lamer**[1,2], **Karim Belarbi**[3,4], **Mohamed Lemdani**[2,3], **Djamel Zitouni**[2,3] and **Benjamin C Guinhouya**[1,2] (ID)

Abstract

Objective: Endometriosis is a complex full-body inflammation disease with an average time to diagnosis of 7–10 years. Social networks give opportunity to patient to openly discuss about their condition, share experiences, and seek advice. Thus, data from social media may provide insightful data about patient's experience. This study aimed at applying a text-mining approach to online social networks in order to identify early signs associated with endometriosis.

Methods: An automated exploration technique of online forums was performed to extract posts. After a cleaning step of the built corpus, we retrieved all symptoms evoked by women, and connected them to the MedDRA dictionary. Then, temporal markers allowed targeting only the earliest symptoms. The latter were those evoked near a marker of precocity. A co-occurrence approach was further applied to better account for the context of evocations.

Results: Results were visualised using the graph-oriented database Neo4j. We collected 7148 discussions threads and 78,905 posts from 10 French forums. We extracted 41 groups of contextualised symptoms, including 20 groups of early symptoms associated with endometriosis. Among these groups of early symptoms, 13 were found to portray already known signs of endometriosis. The remaining 7 clusters of early symptoms were limb oedema, muscle pain, neuralgia, haematuria, vaginal itching, altered general condition (i.e. dizziness, fatigue, nausea) and hot flush.

Conclusion: We pointed out some additional symptoms of endometriosis qualified as early symptoms, which can serve as a screening tool for prevention and/or treatment purpose. The present findings offer an opportunity for further exploration of early biological processes triggering this disease.

Background

Endometriosis is a complex full-body inflammation disease, affecting women of reproductive age, as well as cisgender, transgender, and non-binary people in a currently unknown number.[1–3] Simply stated, the disease affects anyone with a body part classified as 'female'.[1] Approximately 200 million people worldwide, including 10–15% of women of reproductive age and 2.5% of postmenopausal women are affected by endometriosis.[4,5] The most recognised

[1]Univ. Lille, UFR 3S, Faculté Ingénierie et Management de la Santé, Lille, France
[2]Univ. Lille, CHU Lille, ULR 2694 – METRICS, Lille, France
[3]Univ. Lille, UFR 3S, Faculté de Pharmacie, Lille, France
[4]Univ. Lille, Inserm, CHU-Lille, Lille Neuroscience & Cognition, Lille, France

Corresponding author:
Benjamin C. Guinhouya, Univ. Lille, ULR 2694 – METRICS, UFR 3S, 42 rue Ambroise Paré, F-59120 Loos, France.
Email: benjamin.guinhouya@univ-lille.fr

symptoms of endometriosis are chronic pelvic pain, dyspareunia, dysmenorrhea, menorrhagia, bowel symptoms, and infertility. Currently, not only is there no effective treatment for endometriosis, but the time between the development of the first lesions and diagnosis can be by about 7–10 years.[6,7] Thus, two major challenges must be met: identification of the earliest symptoms, which could support investigations on biological processes, and then biomarkers for potential therapeutic targets. The biological complexity under endometriosis has been addressed using computational biology approaches applied on endometriosis-related alterations (development and progression),[8] and endometriosis-related core symptoms.[9]

Nevertheless, the question of how to capture early symptoms of endometriosis, so that potential prevention strategies (as no relevant treatment currently exists for endometriosis) may have the highest impact on women still remains a matter of concern. Indeed, the identification unknown symptoms or early signs of endometriosis (i.e. those not clearly or frequently reported in the scientific literature), can be useful for healthcare professionals in their screening and diagnosis approach. The emergence of the Internet and the popularity of medical forums have provided additional health information about the patient experience.[10] For example, studies on the detection of COVID-19 symptoms on Twitter have emerged and supplemented clinically known symptoms.[11] Sarker et al.[11] reported the expression of anosmia and/or ageusia as symptoms not documented in the literature by matching the word of the corpus to a dictionary of medical concepts. The advantages of web-based data are its volumetric nature and level of detail regarding the patient's experience.

In view of the current dearth of data on the patient-reported experience on early signals of endometriosis, the aim of this study was to identify some of them, by collecting patient's experiences on online social networks platforms.

Methods

Study design and data sources

This was an observational and cross-sectional study consisting in the extraction and analysis of open and patient-centred discussions retrieved from online forums, which involved French-speaking communities. The selection of forums was made for posts released between February 2005 and March 2021, with the query 'endométriose inurl: forum' via the Google search engine. The collection and reviewing of URLs was made in March 2021 by two independent authors. Resources that were not forums, which presented a newsletter, which did not give rise to an exchange, or which required access rights were excluded. The collection of URLs containing the keyword 'endometriosis' accounted for the various threads on this topic. All pages on each thread were crawled, that is the URLs of the websites were scanned and retrieved.

As shown in Figure 5 (Supplementary material), the frequency of use of endometriosis forums was not constant over the selected study period (2005–2021). The number of posts was higher in the period 'before 2015' (wave 1) when compared to period 'after 2015' (wave 2). We then divided and compared data collected according to these two waves. We found with a kappa test that there is a huge agreement ($k = 87\%$) in the occurrence of words between the considered two waves, and that >80% of the evoked symptoms observed in the period 'after 2015' were similar to those of the period 'before 2015'. Consequently, data were pooled together in analyses.

We used the French edition of MedDRA (Medical Dictionary for Regulatory Activities) v.24. MedDRA is a dictionary of medical concepts to explore the symptomatology related to endometriosis.[12] The medical concepts ontology are hierarchically arranged according to five levels of detail, from system organ class (SOC) to lowest level term (LLT). We retained the LLT, which includes medical terms such as symptoms, risk factors, and quality of life, for example 'reproductive organs and breast disorders'.

The workflow was carried out in five steps (Figure 1): (i) the automated extraction of posts, and extraction of the vocabulary of medical concepts chosen for the analysis; (ii) cleaning the data from the 10 French online forums and the dictionary of medical concepts at the preprocessing step; (iii) detection of symptoms present in the corpus of posts from the dictionary; (iv) contextualisation of the list of symptoms detected in the text; and (v) identification of early symptoms with the building of a dictionary of temporal markers for the identification of early symptoms.

Data extraction and preprocessing

The posts of each discussion thread were scraped using BeautifulSoup, Selenium, Pandas and Urllib libraries to extract the data from the HTML and XPATH trees in Python language. They were then gathered into a common corpus with the same structure. Finally, both posts and MedDRA LLTs followed the same preprocessing process with the Re and SnowballStemmer libraries in Python.

The following phases allowed retaining only a clean text:

(i) the suppression step allowed removing duplicate/empty posts, automatic replies, URLs and special symbols on the corpus. We also removed medical concepts containing less than 3 letters (e.g. 'PA' a French acronym for blood pressure, which can also have a negative connotation like 'pas');

(ii) the homogenisation steps removed accents, punctuation marks and changed texts to lowercase;
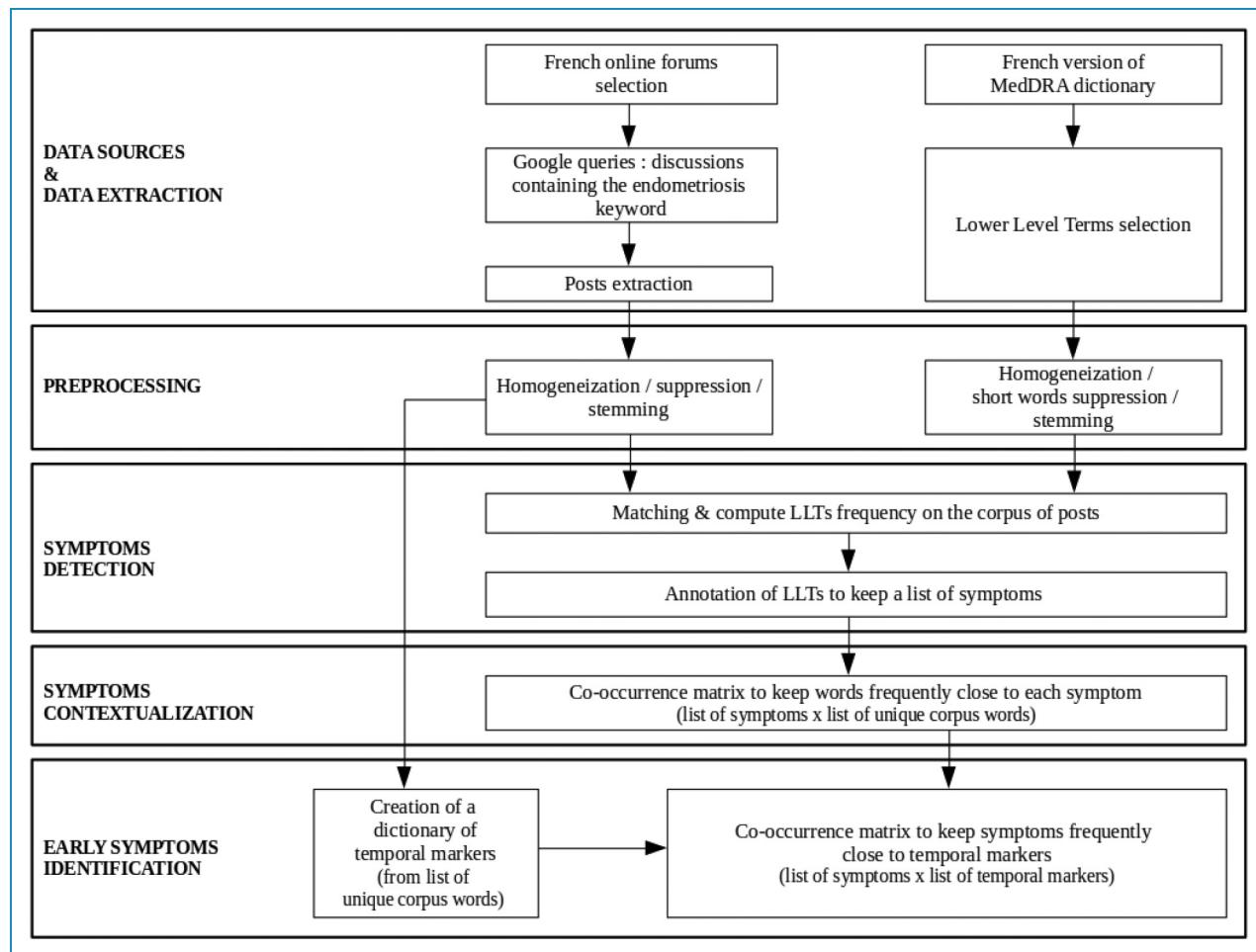
Figure 1. Study design: steps of data sources and data extraction, preprocessing, symptoms detection, symptoms contextualisation step and early symptoms identification.

(iii) the stemming process tokenised texts, in order to remove the stop words and bring the words back to their root-words. Tokenisation aims to split a message into a list of words.

A unique ID in our corpus to guarantee anonymisation replaced the user's nickname. The final information was organised in a 9-column table: forum name; discussion ID; post date; post time; user ID; scraped message; cleaned posts; stem posts; post ID. The final corpus included new columns corresponding to the message without punctuation, without characters and without capitalisation ('message_clean') and to the stemming process ('message_stem'), is available in Supplementary Table 3.

Symptom detection

A clinical symptom is defined as a finding that is reported by patients or by someone close to them.[13] For the purpose of the current study, a symptom was then deemed a signal perceived and reported by women, and

which does not require medical expertise (e.g. cysts are signs and cannot be considered as a symptom since their identification requires the intervention of a health professional). We identified the presence of LLTs in the corpus by exact matching each concept of the MedDRA dictionary on each post. We then calculated frequency of occurrence of each retained symptom. A manual annotation of each LLT with frequency greater than 10 was carried out according to five categories: symptoms, (risk or protective) factors, outcome, diagnosis and treatment. Two independent annotations were performed. A Kappa test was used for the analysis of annotators' agreement. The difference in the annotations were discussed and solved during a third annotation phase.

Contextualisation of symptoms

Contextualisation of symptoms helped providing more information on signals. Clarifying the meaning of each early symptom is helpful in understanding their semantics.

Table 1. Examples of original posts and preprocessed posts, showing the proximity between a temporal marker and an early symptom.

Message	Message stemmed
« Bonjour a toutesJai vu que qq une dentre vous avaient des **douleurs** *a la fesse, la jambe etc..* **Depuis** *plusieurs mois jai une* **douleur** *lancinante qui part de la fesse pour irradier ds le pelvis en passant par le bas du ventre, le vagin et les cuisses […] »*	*« toutesj dentr* **douleur** *fess jamb* **depui** **douleur** *lancin fess irradi pelv ventr* **vagin** *cuiss […] »*
« BonjourJe mexplique rapidement.Durant ma **jeunesse** *jai toujours eu les* **regles douloureuses** *mais cela ne minquetait pas […] »*	*« bonjourj mexpliqu rapid* **jeuness** **regl doulour** *minquet […] »*

A. Blue: temporal marker. B. Yellow: symptom identified in the window of 10 words around a temporal marker.

The tokenisation of the sub-corpus of the message (including the temporal precocity markers) allowed the words of the corpus to be crossed with the dictionary of symptoms previously established. A co-occurrence matrix was applied with a four-word window to generate the occurrence of word pairs found in proximity.[14] This number was set before the analysis, and can be adapted as appropriate. The matrix enabled to retrieve the words next to each symptom. The result of the co-occurrence matrix gave, for each symptom, a list of words found nearby. A common annotation of each association of a symptom with a context word was made to remove words that did not provide any additional information about the symptom. For example, the symptom 'spotting' associated with the word 'pill' leads to thinking that spotting is a consequence of taking the pill. An association of a symptom with a context word with similar meaning (e.g. synonyms), was grouped under a representative medical term. The visualisation of the results was carried out in the form of relational graphs with the graph-oriented database Neo4j. Each symptom was linked to the medical term by a link representing the contextualisation word.

Early symptoms identification

The unique terms of the corpus were studied individually by two readers, for empirical construction of a lexical field of temporal markers. Inside each post, proximity of a 10-word window between each temporal marker and each detected symptom was set.[15,16] This threshold was selected empirically, after a comprehensive reading of all the posts. The production of the co-occurrence matrix allowed the detection of symptoms close to a temporal marker evocative of precocity (Supplementary Table 1). The early symptoms selected were the most frequently found in co-occurrence with a temporal marker. A temporal marker was often introduced at the beginning of a sentence while a clinical sign can be found later in the sentence (Table 1). In a second step, the contextualisation of the all symptoms was filtered to keep only the elements of these early symptoms.

Ethical considerations

All of the data collected in this study came from public discussions. Information from public sources represents a public act, and is made available for passive data collection analysis. Such a type of study does not need any approval from an ethical committee, as in France, these committees are mainly involved in the assessment of studies that are intended to collect data de novo, and which may require patients' information and/or consent to be carried out. In the present case, as far as people have accepted the general conditions of use of the selected forums before partaking in exchanges, so their posts are made publicly available and can be retrieved and used for research, but not for commercial purposes. Nonetheless, to enable people to exert their full information right, a summary of the project, its results, and the coordinates of the corresponding author are displayed on the website of the academic
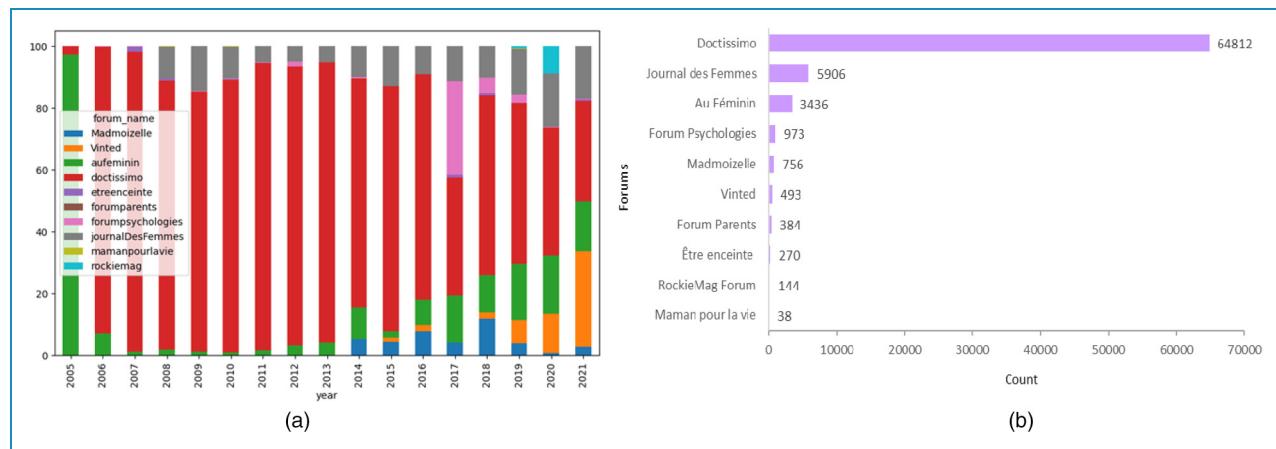
Figure 2. (a) Trend in the use of forums between 2005 and 2021. (b) Number of posts per forum.

laboratory (ULR 2694-METRICS, University of Lille). According to the privacy policy of the selected forums, which are in line with the General Data Protection Regulation (GDPR), the use of these publicly available posts does not require individual consent from users.[17,18] We further completely de-identified posts, and pseudonyms were replaced with unique identifiers and messages were not fully quoted.

Results

Data sources

The query 'endométriose inurl: forum' via the Google search engine resulted in 68 different websites. After applying the inclusion and exclusion criteria, 10 forums were selected: 'Doctissimo', 'Journal des Femmes', 'Au Féminin', 'Forum Psychologies', 'Madmoizelle', 'Vinted', 'Forum Parents', 'Être Enceinte', 'RockieMag Forum', and 'Maman pour la vie' (Figure 2a and b). These forums were about general health, women's media and specialised maternity. From 2006 to 2016, Doctissimo was the preferential forum used for endometriosis in these online communities (Figure 2a).

Data extraction and preprocessing

Overall, we identified 7148 URL discussion threads related to endometriosis (Supplementary Table 2) including 78,905 posts. The preprocessing step removed 1585 posts (2%). The cleaned corpus was finally composed of 77,320 unique posts. A total of 9390 users are distributed over the 10 forums. The top 3 forums by number of posts were 'doctissimo.fr' (82.1%, $N = 64,812$ posts), 'journaldesfemmes.fr' (7.5%, $N = 5906$ posts) and 'aufeminin.com' (4.3%, $N = 3436$ posts) (Figure 2b).

Regarding the extraction of medical concepts, 83,217 unique LLTs from MedDRA were used in the symptom detection analysis. The grouping by LLTs of this result gave the total number of occurrences of each LLTs. This yielded 2064 LLTs detected in the corpus. The 630 LLTs with more than 10 occurrences in the corpus were manually annotated according to symptoms, factors, outcomes of the disease, diagnostic methods, and treatments classes. One annotator obtained 117 selected symptoms, 365 other annotations (e.g. factor, outcome, diagnosis, and treatment) and 148 unreferenced (e.g. death, marriage, divorce, unemployment). A second annotator yielded 248 symptoms, 269 other categories, and 113 unreferenced. The difference in annotation revealed a Kappa score of 43%. This important discrepancy led to a third annotation in common. The final annotation (Supplementary Table 4) translated into 195 symptoms, 120 factors, 53 outcomes, 33 diagnoses and 35 treatment terms (Figure 3). Symptoms that can only be detected after a medical examination were removed (e.g. ovarian cysts).

Contextualisation of symptoms

The resulting dictionary of the most relevant unique terms of the corpus was used for the co-occurrence matrix. After the filtering columns that contain symptoms, the matrix was formed of the 8002 words in row and 167 symptoms (Supplementary Table 6). The word occurrences were kept for each symptom, and then checked manually. For the entire symptoms, 353 symptom/word context pairs were obtained. After the common annotation, 82 symptoms were contextualised grouped into 41 representative medical terms (Supplementary Table 7).

Early symptoms identification

The corpus included 8126 unique terms that were studied to extract the time markers. We identified 26 temporal markers: 'amont', 'ancien', 'adolescent', 'anteced', 'antecedent', 'anterieur', 'apparu', 'auparav', 'avant', 'debut',
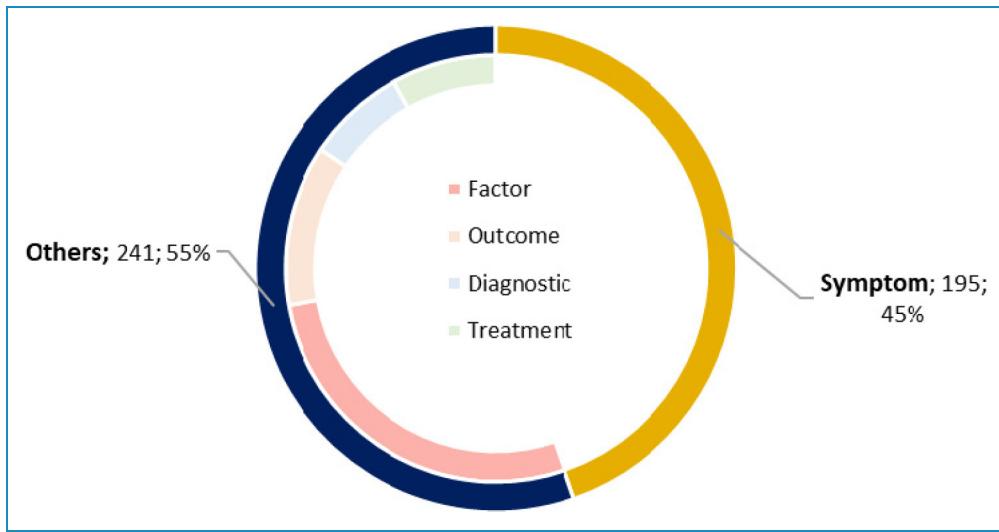
Figure 3. Distribution of the final annotation of LLTs in five categories: symptom, factor, outcome, diagnostic, and treatment. Symptoms. A. Factor: containing identified risk and protective factors. B. Outcome: containing identified symptoms (clinical signs detected pre-diagnosis). C. Diagnostic: containing diagnosis methods. D. Treatment: containing treatment methods.
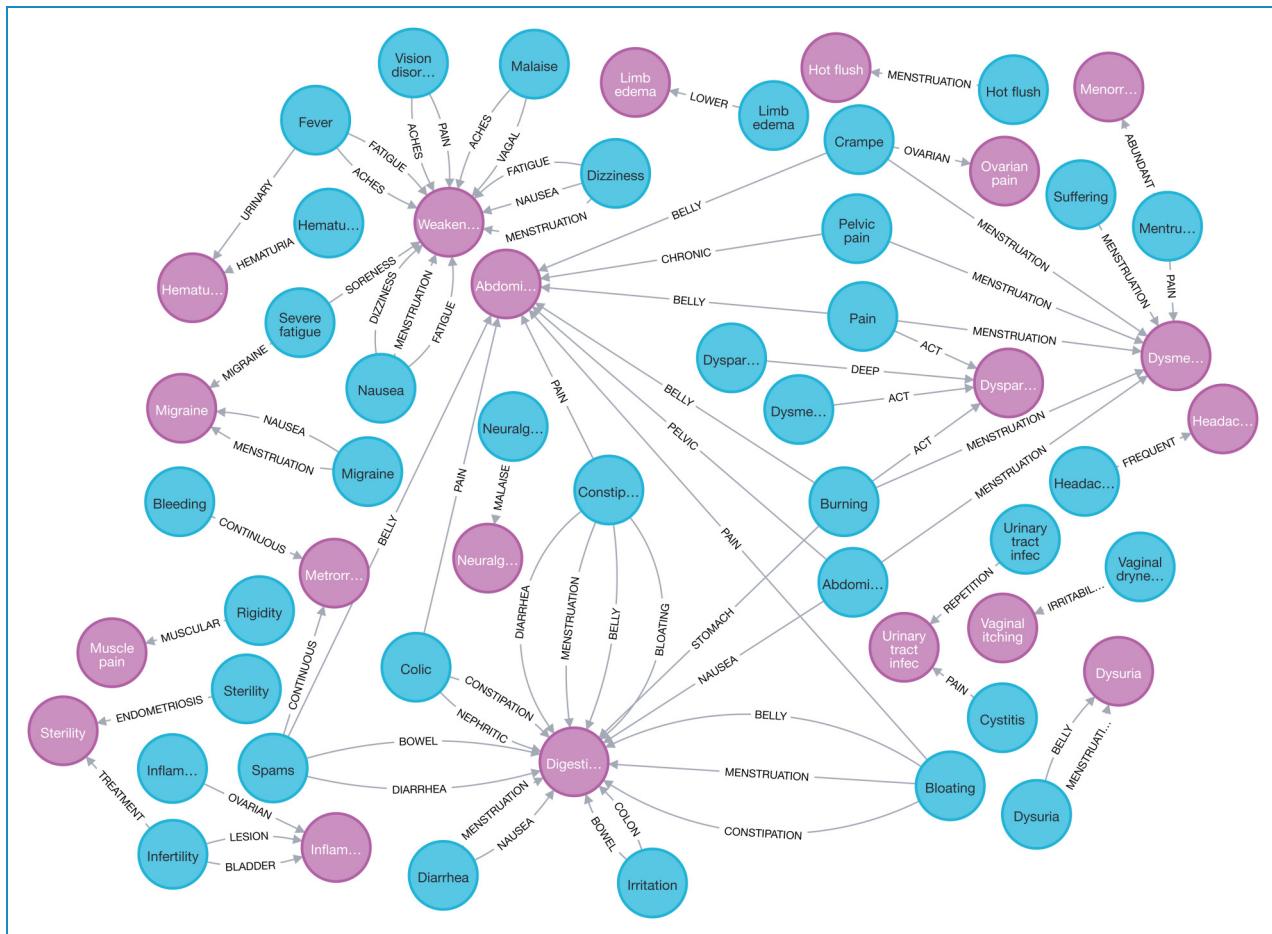
Figure 4. Network representing the detected early symptoms associated with the symptoms annotated with the contextualisation words (visualisation with Neo4j). A. Node blue: early symptoms detected in the corpus. B. Edges: contextualisation words associated with the symptom. C. Node orange: annotated symptom class.

'depui', 'enfanc', 'jeun', 'jeuness', 'lenfanc', 'apparaissent', 'premi', 'premier', 'quauparav', 'reapparaiss', 'reapparaissent', 'reapparaitr', 'reapparit', 'reapparu', 'vecu', 'vecus'. The messages located by temporal markers consisted in 15,032 unique posts. These posts were used for the rest of the analysis. From the co-occurrence matrix, the symptom occurrences were kept for each temporal marker, and then checked manually. A list of 53 unique early symptoms was identified (Supplementary Table 5).

This annotation was conducted to keep these 53 early symptoms contextualised. Seventy-four symptom/word context pairs were kept (Supplementary Table 8) with 20 general symptoms: 'Weakened general condition', 'Hot flush', 'Headache', 'Vaginal itching', 'Abdominal pain', 'Muscle pain', 'Ovarian pain', 'Dysuria', 'Dysmenorrhea', 'Dyspareunia', 'Haematuria', 'Urinary tract infection', 'Inflammation', 'Metrorrhagia', 'Menorrhagia', 'Migraine', 'Neuralgia', 'Limb oedema', 'Infertility', 'Digestive disorders' (Figure 4).

Discussion

This is a text-mining approach, based on the exploration of exchange platforms, with the goal to investigate the 'early' symptomatology of endometriosis. We collected 41 groups of symptoms, including 20 groups of them considered as 'early symptoms' associated with endometriosis, which were recontextualised based on the content of forums. An exploratory patient-centred approach was used in order to take the best advantage from free rich posts released by interested parties on endometriosis, just like a 'big' focus group on endometriosis. As such, forums can also be viewed as an interesting space for people to actually have sufficient time to update, correct or even contradict their own initial ideas on a given subtopic. By using the MedDRA, each symptom and its synonyms, were translated into a unified lexicon, which then eased the retrieve of only terms referred to as symptoms from the built corpus. Our study allowed pinpointing symptoms freely reported by women on forums since social media now appear as the preferred space for women with endometriosis to express themselves and share experiences/advices about their condition.[19] To the best of our knowledge, this approach has so far never been adopted in clinical and/or epidemiological studies.

Some of the symptoms identified in our analysis are in agreement with previous findings in the literature. Their underlying biological mechanisms in the pathophysiology of endometriosis have already been comprehensively discussed. The dysregulation of factors involved in the pathological process of endometriosis is directly associated with migraines, pelvic pain and more specifically with dysmenorrhea, dyspareunia, and painful bladder syndrome and irritable bowel syndrome.[20] Pelvic and abdominal pains reach the nervous system, and lead through nociception, to neuropathic or neuroinflammatory pain, which may explain the migraines and headaches experienced by patients.[20] The inflammatory environment of the disease may explain the infertility experienced by the patients, since these dysregulations can cause an alteration of the embryonic implantation and the inability of women with endometriosis to get pregnant.[21] Endometriotic lesions have been strongly associated with irritation causing abdominal pain, pain on defecation, urinary tract infection, dysuria, and digestive disorders.[22] Finally, endometrial lesions can cause pelvic pain, menorrhagia and metrorrhagia.[23] Other extracted symptoms, such as urinary pain, urinary tract infection, or gastrointestinal disorders are in agreement with the literature. Newly detected symptoms herein, whose relationship with endometriosis may be unknown need to be more extensively studied. Practitioners, who have not yet looked at, can now address these potentially new symptoms. Nonetheless, oedema for example would be associated with some pain and digestive disorders, which could guide the diagnosis towards endometriosis. Since peripheral nerves are directly associated with endometrial tissue,[20] earlier authors have assumed that implanting endometrial tissue around the sciatic nerve would cause an inflammatory reaction, severe pain and neuralgia.[24] Furthermore, it was argued that early diagnosis of endometriosis could avoid permanent nerve damage, which may be associated with bladder incontinence, muscle weakness and fatigue.[20]

Many patients complain of vaginal itching caused by endometrial lesions, with no clear physiological or biological explanations. The identified dizziness in the alteration of the patient's general state was found to be an adverse effect of endometriosis treatments (e.g. opioid, postoperative effects of laparoscopic surgery).[25,26] Fatigue is associated with somatic pain syndrome and is significantly related to endometriosis[27,28] or can be associated with the side effects of some treatments (e.g. contraceptive pill).[26] Patients with gastrointestinal disturbances report the presence of nausea.[29] Acute pain can justify this nausea, for abdominal pain being possibly a source of nausea, as well as dysmenorrhea. Furthermore, attempted treatments can also yield side effects such as nausea and vomiting.[30] The main information concerning a possible link between endometriosis and hot flashes suggested a post hormonal treatment effect, which may translate into artificial menopause.[31,32] Symptoms not biologically explained and/or explored in the literature may be consequences of other symptoms and/or due to treatments. Therefore, it would be interesting to further check and investigate these signals during consultations. Because some patients can self-report some of the symptoms found in this study during the consultation interview, such potentially novel signs can complete those already known and/or used by practitioners in their routine.

It is important to note that the text-mining approach has a number of limitations. Manual data extraction is time consuming and limited to the selected sources. An adaptation of the data extraction method to each source within a learning of each HTML code is required. This step can be facilitated by using an API to automatically extract data from all sources using a list of keywords. Regarding the identification of early symptoms, a dictionary was created in an empirical way. A more accurate method could be developed and applied at the place of the built timeline in the speech. The analysis of verb tenses can be used to classify and identify the symptoms that occur first. The lack of information about the contextualisation of some symptoms can add noise to the results. The context of the entire sentence in which the symptom is evoked can also cause confusion. Furthermore, without a formal or declared diagnosis by people using the selected forums for this study, we have purposely merged posts released by people with endometriosis and any third party (e.g. caregiver, relatives) interested by this condition. Nevertheless, the findings from this study would need to be compared with those of a future study involving only people with endometriosis. Moreover, the temporal markers are not rigorously linked to the earliness of symptoms as described in this study although they are in close proximity. An analysis of the context, taking into account the tenses of the verbs, might help to identify more precisely the temporality/earliness of the evoked symptoms. Moreover, since different symptoms may correspond to different stages of the disease, due the heterogeneous manifestation of endometriosis, it is difficult, once again, to ascertain that the novel symptoms obtained in this study are absolutely in the earlier phases of the disease. Nonetheless, these symptoms should be considered as additional signs that can inform clinical decision-makers and/or practitioners on endometriosis. In future studies, it is noteworthy that a sentiment analysis can also be applied on such sentences to make sure that the symptom is not evoked as a negation (e.g. '*I don't have a headache*'). Finally, at a methodological standpoint, our approach can be tested on another corpus of posts in order to assess its validity and reproducibility.

Conclusions

This study showed the relevance of forum discussions in detecting endometriosis symptoms from patient experiences. Symptoms completing previous findings (e.g. headache, neuralgia) were pointed out by our analysis. However, some of the detected symptoms in the current study have not yet been linked to any biological mechanism of the disease, and this invites further studies. Some of these potentially new symptoms highlighted could for example be used to inform a protein network analysis analogous to what was recently published, using the core symptoms of endometriosis.

ORCID iD: Benjamin C. Guinhouya https://orcid.org/0000-0003-1415-0301

References

1. Jones CE. Queering gendered disabilities. *J Lesbian Stud* 2021; 25: 195–211.
2. Rogers PAW, D'Hooghe TM, Fazleabas A, et al. Priorities for endometriosis research. *Reprod Sci* 2009; 16: 335–346.
3. Shim JY, Laufer MR and Grimstad FW. Dysmenorrhea and endometriosis in transgender adolescents. *J Pediatr Adolesc Gynecol* 2020; 33: 524–528.
4. Prodromidou A, Machairas N, Paspala A, et al. Diagnosis, surgical treatment and postoperative outcomes of hepatic endometriosis: a systematic review. *Ann Hepatol* 2020; 19: 17–23.
5. Zhu J, Xue X, He Z, et al. Using network pharmacology and molecular docking to explore the underlying anti-inflammatory mechanism of Wuyao-Danshen to treat endometriosis. *Ann Transl Med* 2022; 10: 98.
6. García-Gómez E, Vázquez-Martínez ER, Reyes-Mayoral C, et al. Regulation of inflammation pathways and inflammasome by sex steroid hormones in endometriosis. *Front Endocrinol* 2020; 10: 935. https://doi.org/10.3389/fendo.2019.00935
7. Smolarz B, Szyłło K and Romanowicz H. Endometriosis: epidemiology, classification, pathogenesis, treatment and genetics (review of literature). *Int J Mol Sci* 2021; 22: 10554.
8. Peng Y, Peng C, Fang Z, et al. Bioinformatics analysis identifies molecular markers regulating development and progression of endometriosis and potential therapeutic drugs. *Front Genet* 2021; 12. https://doi.org/10.3389/fgene.2021.622683
9. El Idrissi F, Fruchart M, Belarbi K, et al. Exploration of the core protein network under endometriosis symptomatology using a

computational approach. *Front Endocrinol (Lausanne)* 2022; 13: 869053. eCollection 2022.

10. van Uden-Kraan CF, Jansen F, Lissenberg-Witte BI, et al. Health-related and cancer-related internet use by patients treated with total laryngectomy. *Support Care Cancer* 2020; 28: 131–140.

11. Sarker A, Lakamana S, Hogg-Bremer W, et al. Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. *J Am Med Inform Assoc* 2020; 27: 1310–1315.

12. Schäfer F, Faviez C, Voillot P, et al. Mapping and modeling of discussions related to gastrointestinal discomfort in French-speaking online rorums: results of a 15-year retrospective infodemiology study. *J Med Internet Res* 2020; 22: e17247.

13. Cox AP, Ray PL, Jensen M, et al. Defining 'sign' and 'symptom'. In: *Second International Workshop on Definitions in Ontologies (IWOOD 2014)*. Houston, TX, USA, 2014, pp. 42–48.

14. Li P, Burgess C and Lund K. The acquisition of word meaning through global lexical co-occurrences. In: *Proceedings of the Thirtieth Annual Child Language Research Forum*. Centre for the Study of Language & Information, Stanford, USA, 2000, pp. 167–178.

15. Osadchiy V, Jiang T, Mills JN, et al. Low testosterone on social media: application of natural language processing to understand patients' perceptions of hypogonadism and its treatment. *J Med Internet Res* 2020; 22: e21383.

16. Zhang L, Hall M and Bastola D. Utilizing Twitter data for analysis of chemotherapy. *Int J Med Inform* 2018; 120: 92–100.

17. Eysenbach G and Wyatt J. Using the internet for surveys and health research. *J Med Internet Res* 2002; 4: e862.

18. Mann C and Stewart F. *Internet communication and qualitative research: a handbook for researching online*. 1st ed. London, UK: Sage Publications Ltd, 2000, pp.272.

19. Holowka EM. Mediating pain: navigating endometriosis on social media. *Front Pain Res (Lausanne)* 2022; 30: 889990. eCollection 2022.

20. Bajaj P, Bajaj P, Madsen H, et al. Endometriosis is associated with central sensitization: a psychophysical controlled study. *J Pain* 2003; 4: 372–380.

21. Veenstra van Nieuwenhoven AL, Heineman MJ and Faas MM. The immunology of successful pregnancy. *Hum Reprod Update* 2003; 9: 347–357.

22. Roman H, Ness J, Suciu N, et al. Are digestive symptoms in women presenting with pelvic endometriosis specific to lesion localizations? A preliminary prospective study. *Hum Reprod* 2012; 27: 3440–3449.

23. DiVasta AD, Vitonis AF, Laufer MR, et al. Spectrum of symptoms in women diagnosed with endometriosis during adolescence vs adulthood. *Am J Obstet Gynecol* 2018; 218: 324.e321–324.e311.

24. Chen S, Xie W, Strong JA, et al. Sciatic endometriosis induces mechanical hypersensitivity, segmental nerve damage, and robust local inflammation in rats. *Eur J Pain* 2016; 20: 1044–1057.

25. Chiuve SE, Kilpatrick RD, Hornstein MD, et al. Chronic opioid use and complication risks in women with endometriosis: a cohort study in US administrative claims. *Pharmacoepidemiol Drug Saf* 2021; 30: 787–796.

26. Hong J-Y. The effect of preoperative Ketorolac on WBC response and pain in laparoscopic surgery for endometriosis. *Yonsei Med J* 2005; 46: 812–817.

27. Álvarez-Salvago F, Lara-Ramos A, Cantarero-Villanueva I, et al. Chronic fatigue, physical impairments and quality of life in women with endometriosis: a case-control study. *Int J Environ Res Public Health* 2020; 17: 3610.

28. Ramin-Wright A, Schwartz ASK, Geraedts K, et al. Fatigue – a symptom in endometriosis. *Hum Reprod* 2018; 33: 1459–1465.

29. Armour M, Middleton A, Lim S, et al. Dietary practices of women with endometriosis: a cross-sectional survey. *J Altern Complement Med* 2021; 27: 771–777.

30. Gonadotropins. *LiverTox: clinical and research information on drug-induced liver injury [Internet]*. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases, 2012.

31. Rodgers AK and Falcone T. Treatment strategies for endometriosis. *Expert Opin Pharmacother* 2008; 9: 243–255.

32. Saltiel E and Garabedian-Ruffalo SM. Pharmacologic management of endometriosis. *Clin Pharm* 1991; 10: 518–531.