Emotional Reactions to Infertility Diagnosis in Patients and Partners: Thematic and Natural Language Processing Analyses of Textual Data from the 1,000 Dreams Infertility Survey

Running title: Qualitative and NLP analyses of infertility survey

Jacky Boivin[a*], Mustafa Oguz[b], Mai Duong[b], Owen Cooper[b], Dina Filipenko[b], Marie Markert[c], Carl Samuelsen[d], William R Lenderking[e]

[a]Cardiff Fertility Studies Research Group, School of Psychology, Cardiff University, 70 Park Place, Cardiff, Wales, UK, CF10 3AT; [b]Evidera, The Ark, 201 Talgarth Road, 2nd Floor, London, UK W6 8BJ; [c]Ferring Pharmaceuticals A/S, Amager Strandvej 405, DK - 2770 Kastrup, Denmark; [d]Outcomes Analytica, Vetlandsveien 66, 0685 Oslo, Norway; [e]Evidera, 500 Totten Pond Rd, Fifth Floor, Waltham, MA, US 02451

***Corresponding author:** Jacky Boivin (boivin@cardiff.ac.uk)

ABSTRACT

Research question: What are the emotional impacts of infertility on patients and/or partners, and how can qualitative thematic analyses (TA) and natural language processing (NLP) help evaluate textual data?

Design: This was a cross-sectional, multi-country survey conducted from March to May 2019. A total of 1,944 patients and/or partners from nine countries responded to the open-ended question asking about their initial feelings related to an infertility diagnosis. A mixed-method approach that integrated NLP topic modelling and TA was used to analyse responses. Sentiment polarity, which expresses the valence of respondent sentiments in NLP, was quantified for each response. Linear regression was used to evaluate the association between patient characteristics and sentiment negativity.

Main results: NLP and TA showed that the most common emotional reactions to infertility diagnoses were sadness, depression, stress, disappointment, anxiety, frustration, confusion, and loss of self-confidence. NLP topic modelling found additional reactions such as shared feelings with partners, recollections about causes of infertility, and treatment experience. Responses to the open-ended question were brief (median: three words) with 71.8% conveying negative sentiments. Some respondent characteristics showed small but significant associations with sentiment negativity, such as country (Spain, China, and France were more negative than the United States), treatment engagement (no treatment was more negative than one or more treatment), and marital status (missing/other was more negative than divorced).

Conclusion: Infertility diagnoses create an emotional burden for patients and partners. The mixed-method approach provides a compelling synergy in support of the validity of these findings and shows potential for these techniques in future research.

Keywords: Infertility, emotional burden, sentiment, natural language processing, thematic analysis, multi-country survey

INTRODUCTION

Infertility—the failure to establish a clinical pregnancy after 12 months of regular and unprotected sexual intercourse—affects approximately 48.5 million couples in 2010 (Mascarenhas et al., 2012), approximately 8% to 12% of reproductive-aged couples worldwide (Mascarenhas et al., 2012; Vander Borght, Wyns, 2018). A diagnosis of infertility can cause psychological and emotional stress to patients who are infertile and their partners. A recent literature review concluded that 25% to 60% of infertile people reported psychiatric symptoms and that their levels of anxiety and depression were significantly higher than in fertile controls (De Berardis et al., 2014). Typical reactions to infertility include shock, sadness, depression, anger and frustration, loss of self-esteem and self-confidence, and a general loss of sense of control (Vander Borght, Wyns, 2018; Simionescu et al., 2021). The potential impact of psychological factors on pregnancy rates is a controversial area. Some studies show that the more distressed women are prior to and during treatment, the lower pregnancy rates are, while others did not (Rooney, Domar, 2018). Regardless of the causal relationship, psychological interventions for couples with infertility can reduce anxiety and depression and significantly increase pregnancy rates (Simionescu et al., 2021). Marital satisfaction in patients with infertility was influenced by their and their partners' perceived stress. Therefore, psychological interventions that target a reduction in perceived stress and enhancement of marital satisfaction in the context of infertility should treat the couple as a unit (Maroufizadeh et al., 2019).

Although infertility is a couple diagnosis, one partner is often deemed to be the patient and the other the partner of the patient. Understanding the experience of both patients and partners is crucial to alleviating the burden and meeting the needs of patients with infertility and their partners. The 1,000 Dreams survey was conducted to better understand the perceptions and emotions of patients and partners during their infertility journey, what motivates couples seeking treatment, and treatment barriers (Domar et al., 2021; Boivin et al., 2022). This global, cross-cultural online survey was developed in English and translated into French, German, Italian, Spanish, and Mandarin. Translations were validated by national linguists.

Massive amounts of textual data obtained from social media and large online surveys raise a challenge to thematic analysis (TA), the most common method of analysis in qualitative research. Advances in natural language processing (NLP) provide exciting applications in qualitative research, due to NLP's ability to 1) analyse massive amounts of textual data obtained from social media sources (e.g., online chat forums) that may not be well represented in the medical literature (Osadchiy et al., 2020a; Osadchiy et al., 2020b); and 2) be automated for regular, repeat evaluations to compare changes over time (Himmel et al., 2009). TA and NLP approaches are grounded in the data but involve different procedures. TA involves a complex understanding of the context while NLP focuses exclusively on statistical regularity in word usage. NLP uses various methods to extract information from large textual data sets: a) word frequency—to summarise the number of times each word occurs in a text; b) topic modelling—to discover abstract topics in a collection of textual replies using a probabilistic model, and c) sentiment analysis—a computational approach to measuring the feeling that a text conveys to the reader (Reagan et al., 2017). The most common method for sentiment analysis is to use a pre-developed dictionary (lexicon) which includes a set of word stem ratings or scores to determine the overall sentiment of all words in the document. A sentiment score can be in binary form (positive or negative) or continuous (different lexicons use different score ranges). Other methods for categorising sentiment include supervised learning methods and unsupervised (or deep) learning methods (Socher et al., 2013).

A previous online survey study that utilised both topic modelling and grounded theory qualitative analysis (Baumer et al., 2017) showed that results from the mixed-methods analysis provided additional value when investigating free-text survey responses. The current study combined the TA and NLP approach to gain a better understanding of reactions to diagnosis from the perspective of the patient and partner, and to explore the potential of the NLP methodology in understanding the experiences of people with infertility.

MATERIALS AND METHODS

Study Design and Population

A total of 1,944 participants (self-reported patients and/or ~~their~~ partners to patients) completed the survey question in nine countries: Australia, Canada, France, Germany, Italy, China, Spain, the United Kingdom (UK), and the United States (US). ~~Participants were recruited after the infertility diagnosis, a~~Participants were recruited after their medical infertility diagnosis (following self-report of a professional infertility diagnosis, medical consultation, or one or more cycles of fertility treatment). All participants reported having been directly diagnosed with infertility by a medical professional (patient) or being a partner of someone diagnosed with fertility issues (partner). Participants were invited to the survey on an individual basis, it is unlikely that both a~~the~~ patient~~s~~ and his/her~~their~~ partner were~~are~~ invited. A quota was applied in relation to their stage of the treatment journey to ensure that both early diagnosed and treatment-experienced patients and partners were well presented in the study population. More details about the survey design have been described in other publications (Domar et al., 2021; Boivin et al., 2022). The current study analysed the responses to the open-ended question: '*At the time of your/your partner's infertility diagnosis, how did you initially feel? Please share all thoughts/feelings you may have had related to your/your partner's initial infertility diagnosis.*' Responses were processed through data cleaning steps, including removing redacted responses and duplicates and correcting misspellings. A total of 1,795 individuals contributed data to the final study population.

Overall Approach

A mixed-method approach, which combined NLP and TA, was used to analyse responses to the survey question (Figure 1). Several NLP methods were used in this study, including word frequency, topic modelling, and sentiment analysis. Figure 1 describes the process of NLP analyses. Word frequency, a simple NLP method, was utilised to collect the most frequent words in all responses. Topic modelling identified the main topics from the responses through Latent Dirichlet Allocation (LDA) (Blei et al., 2003), ~~which~~a generative probabilistic model that estimates how important each

word is to each topic and how important each topic is to each of the responses. Each topic identified by LDA contained a range of word probabilities and response probabilities (higher values indicated more important words and responses). ~~Two~~The last step of the topic modelling analysis involved two researchers independently ~~reviewed~~ reviewing the words and responses with the highest probability and suggested a label for each topic. These labels were discussed, and any differences were resolved. Similar topics were grouped to form overarching themes. ~~Figure 1 describes~~The last component of the NLP analyses was the sentiment analysis (Figure 1). Sentiment in the survey responses was quantified using the AFinn (Nielsen, 2011), a manually compiled lexicon ~~used to quantify sentiment in the survey responses. AFinn~~that contains around 2,500 words rated for scores with an integer between –5 (most negative) and 5 (most positive) regarding their polarities. Words such as 'outstanding,' 'thrilling,' and 'superb' have a score of 5, while score of –5 is mainly curse words. In a validation study (Ozdemir, Bergler, 2015b), AFinn was shown to be the best performer among some of the more widely used lexica such as the multi-perspective question answering (MPQA) (Wilson et al., 2005), opinion lexicon by Bing Liu (Hu, Liu, 2004; Ozdemir, Bergler, 2015a), and other automatically compiled lexica (Mohammad et al., 2013). Other lexica classify words' polarities to positive, negative, and neutral, while AFinn uses a score-based approach, which provides more flexibility for analyses. In this study, after assigning a score to each word, the sentiment score of each response was calculated as the sum of the scores of the most positive and negative word in the response. Responses that did not contain a word in the lexicon were dropped from the analysis. As the lexicon has been validated, it is unlikely that these responses would contain important words. The sentiment score of words followed by 'not,' 'no,' 'never,' and 'without' was multiplied by –1 to reverse the polarity of the sentiment score. Responses were classified into negative (<0), positive (>0) and neutral (0) sentiments based on their scores. Since it was expected that most responses would have negative sentiment, scores on the negativity scale were also calculated based on the most negative word in the responses (i.e., scores ranged from –5 to 0, with a score of 0 characterised as neutral and/or positive). All NLP analyses were conducted using R version 3.6.1. For topic modelling, 'topicmodels' package (version 'topicmodels_0.2-8.tar.gz') was used for the analyses.

The qualitative portion of the analysis was structured using Braun and Clarke TA methodology (Braun, Clarke, 2006). Data were cleaned and analysed using the software ATLAS.ti version 8.4 (Friese, Ringmayr, 2018). The TA took an inductive bottom-up approach to identify themes and patterns as driven by the data (Braun, Clarke, 2006). All survey responses were reviewed and analysed, regardless of whether they were relevant to the study's objectives. There was no interpretation beyond the survey responses (a response was defined as any text entered in the response box). The multiple steps of TA included the following: 1) the researcher read the response to each survey question and considered meanings and patterns before coding began; 2) the researcher identified selections of the data that were interpreted as important to the survey respondent and generated the initial codes; 3) the researcher identified subthemes and overarching themes between the codes, and 4) subthemes were grouped to create main themes.

The NLP and qualitative analyses were carried out by different teams to prevent the results of the two approaches from influencing each other. The topics identified by the NLP and the themes identified by the TA were then compared to identify concepts shared between each unique method.

Statistical Analysis

The characteristics of respondents at the time of the survey were described. Mean, standard deviation (SD), median, 25th and 75th percentiles and minimum/maximum values were reported for continuous variables; frequencies and percentages were reported for categorical variables.

Overall sentiment scores were quantified by mean, SD, median and minimum/maximum values. A linear multivariate regression model was used to explore factors associated with a significantly lower score on the negativity scale (–5 to 0). The first model included all baseline characteristic variables; variables were then selected for the final model using a backward elimination approach (Heinze et al., 2018), which involved removing the most insignificant independent variable, then re-estimating the model and repeating the process until no insignificant predictor was left. The final model included variables with p-value ≤ 0.05). A sensitivity analysis was also conducted using the overall sentiment score (–5 to 5).

Subgroup Analyses

Emotions identified in the responses were characterised by frequencies of themes in the TA and frequencies of words in the NLP word count analysis. The overall sentiment scores of the responses were summarised for subgroups by age, sex, patient-partner status, sexual preference, employment, country and treatment engagement and outcome at the time of the survey. Due to the brevity of responses, the topic modelling analyses were not repeated for any subgroup.

RESULTS

Descriptive Statistics for Respondent Characteristics

The characteristics of the 1,795 respondents are described in Table 1. The median length of responses to the open-ended questions was three words (Q1–Q3: 2–5), and the maximum was 32 words. Approximately 43% of participants were enrolled in at least one treatment by the time of the survey, compared with 57% without any treatment engagement or consultation. The mean age of respondents at the time of enrolment in the survey was 36 years (SD: 9.7 years), and 57.2% of the respondents were female. The percentages of patient and partner respondents were 53.6% and 46.4%, respectively~~,~~ ~~and 91.5% were heterosexual.~~ Most participants self-identified as heterosexual (91.5%), while 5.5% self-identified as homosexual and 3.1% as 'other' (e.g., bisexual). The number of respondents was distributed evenly across countries. Couples averaged three years trying to conceive before the infertility diagnosis. Social and psychological support was most provided by partners (20.4% of all respondents), followed by a specialist (e.g., psychologist, therapist or social worker, 19.8%).

NLP Analyses

Word Frequency

The most frequently mentioned word in all responses was 'sad' (n=440, 24.5%). 'Anxious' and 'worried' were mentioned in 14.0% of responses (n=251), followed by 'angry' or 'frustrated' (n=228, 12.7%), 'depressed' (n=187, 10.4%), and 'disappointed' (n=172, 9.6%) (Figure 2). Subgroup analyses showed some minor differences by sex and patient-partner status. Sadness and disappointment were

more common in partners than in patients, while depression and frustration were more common in patients.

NLP Topic Modelling

The 14 topics (the optimal number in LDA) identified by NLP topic modelling and lists of words with the highest probability from each topic are provided in Supplementary Table 1 and Supplementary Figure 1. The NLP-identified topics were grouped to form three overarching topics, and labelled by two researchers:

1. Acceptance (i.e., feeling sad but relieved to know the cause of the problem and hoping for future treatment or considering adoption) or unacceptance of the diagnosis (i.e., initially feeling shocked, confused, disappointed, devastated, angry, guilty, inadequate as a woman or man, fear and hopeless for the future). Some started to look for solutions and treatments after they processed their feelings about the situation. This corresponds to topic number 1, 2, 3, 4, 6, 7, 9, 12, and 13 (Supplementary Table 1).
2. Relational reflections involving shared feelings with partners, concerns about the relationship with partners, and worried about the impact on life, relationships, and family. This corresponds to topic number 8 and 10 (Supplementary Table 1).
3. Recollections about causes of infertility, treatment experience, and concerns about treatment cost and safety. This corresponds to topics 5, 11, and 14 (Supplementary Table 1).

These three overarching topics were not mutually exclusive (i.e., a response that had a high probability mass in topic 1 could also have a high probability in topic 2).

Sentiment Analyses

Responses that did not contain a word that was scored in the lexicon were dropped from the analysis (n=236), which left 1,708 responses for the sentiment analyses. The overall sentiment analysis showed that 71.8% of the responses were negative (score <0), 12.2% were positive (>0), and 16.0% were neutral (0). Supplementary Table 2 shows overall sentiment score frequencies by the

composition of the most positive and most negative words. Of all responses, 89.1% contained at least one negative word, 37.1% had at least one positive word, and 28.7% comprised positive and negative words. The mean sentiment score for all responses was –1.32 (SD 1.5, min –4, max 4) There was no response with a score of –5 (e.g., curse words) or 5 (e.g., 'superb'). The overall sentiment score and sentiment score by respondent characteristics are shown in Table 2.

Associations of respondent characteristics with sentiment scores on the negativity scale (–5 to 0) were explored in a multivariable linear regression model. The final multivariate model included country, treatment engagement, time spent trying to conceive, sexual preference, and marital status (Table 3). These were also all variables that showed significant association with sentiment scores in the univariate regression. Although R-squared was low (multiple R-squared: 0.029), the F-test of overall significance showed that our model provides a better fit than a model with no predictor variables (F-statistic: 3.187 on 16 and 1691 degree of freedom [DF], p-value: 1.945e-05). Some characteristics such as country, treatment engagement, sexual preference, marital status, and time spent trying to conceive showed a significant impact on sentiment scores (Table 3). On a scale of –5 to 0, US respondents had significantly higher (more positive) sentiment scores than respondents from Spain (0.37, $p<0.001$), China (0.25, $p=0.0032$), the UK (0.24, $p=0.0045$) and France (0.23, $p=0.0094$). Respondents with no treatment engagement had a significantly lower (more negative) score compared with respondents enrolled in at least one treatment (0.12, $p=0.02711$). Respondents whose marital status was missing or listed as 'other' (e.g., were in a relationship, civil partnership or concubinage) had significantly lower scores compared with those who were divorced (0.35, $p=0.0034$). Time spent trying to conceive and sexual preference had very small but statistically significant associations with the sentiment score. Other characteristics did not show any significant association with the sentiment score ($p>0.05$). In the sensitivity analysis (Supplementary Table 3), only country, treatment engagement, and sexual preference had a significant association with the overall sentiment score (–5 to 5). The same trend was observed for these characteristics.

TA

The emotional impacts reported by participants varied, and many responses were categorised into more than one candidate theme. Six themes (or emotion categories) emerged from the TA: 1) sad, depressed and disappointed (40.6%); 2) anxious, worried, scared and afraid (21.6%); 3) frustrated, angry and upset (17%); 4) stressed, shocked and overwhelmed (9.2%); 5) lost, confused and unsure (8.6%); and 6) hopeful, optimistic and positive (5.8%). These six themes were classified into three categories: self-appraisal or event-related emotions (theme 1, 3 and 4) and future appraisals (theme 2, 5 and 6) (Figure 3). Supplementary Table 4 provides illustrative quotes. Subgroup analyses were conducted for all themes found in the TA. Very minimal differences in theme distribution were found between subgroups by sex and patient-partner status (Figure 3).

DISCUSSION

NLP methods are increasingly being utilised to extract information from unstructured health-related texts, such as electronic health records and posts on social media and patient platforms. This study combined the classic TA and NLP methods to characterise the reactions of patients and partners with the infertility diagnosis. The most common emotions among respondents were sadness, depression, stress, disappointment, anger, frustration, anxiety, scared, confusion, loss of self-confidence and feeling lost. This finding was consistent with the previous study (Vander Borght, Wyns, 2018). Almost three-quarters of responses were negative, which suggests a significant emotional burden among survey respondents when recollecting diagnosis. Sentiment score analyses showed that both female and male respondents felt equally negative, and partners felt slightly more negative than patients about the infertility diagnosis. Word analysis and TA in this study suggested gender and partner effects; low self-esteem was a common reaction to diagnosis in female patients and female partners showed emotional empathy;~~,~~ ~~whereas~~ male patients were more concerned about the future, ~~while~~ and male partners felt more disappointed. Emotions seemed to be driven by internal sources for patients and external sources for their partners.

Lazarus and Folkman (Biggs et al., 2017)described a transactional model of stress, appraisal and coping which has been highly influential in research into how people react to potentially stress-inducing situations. According to this theory, cognitive appraisals of events (in this case, diagnosis) determine whether people experience the event as a stressor. When an event is perceived as a threat but coping resources are considered insufficient, then people may react to events with considerable stress reactions (e.g., emotional, cognitive, physiological or behavioural). In the present study, emotions reported in relation to diagnosis were mainly threat emotions (e.g., anxious, worried and scared) and harm emotions (sad, depressed and disappointed), typically reported for events perceived as highly threatening to one's well-being (or the well-being of a loved one).

This study showed that topics modelled using NLP identified similar emotion sets to TA, but the computer-driven grouping of textual data provided a structure that allowed reviewers to identify topics without examining a large amount of text. ~~The last step of~~Although topic modelling can be automated to regularly analyse large, updated data sets, the labelling component still needs human input. This last step is ~~to review and assign a label to each topic, which is like~~similar to TA, but less time-consuming since only the most important responses ~~are~~, probabilistically, need to be reviewed. Our study showed that topic modelling can reveal equally valuable insights as TA. For example, while reviewing the top responses in NLP topic 1, reviewers found that the feelings of stress and shock were common among respondents who did not expect the diagnosis. These patients felt too overwhelmed to think about potential treatment and solutions. In contrast, for those who suspected the diagnosis, there was a sense of acceptance and hope for future treatment. Topic modelling helped to identify less expected topics, such as recall about the cause of infertility and treatment experiences and shared feeling with partners. Recollections of prior reproductive events could be triggered because people try to explain what could cause their diagnosis (Koert et al., 2018).~~Topic modelling can be automated to regularly analyse large, updated data sets; however, the labelling component still needs human input.~~ ~~The~~ In conclusion, the mixed approach (NLP and TA) provides a compelling synergy in support of the validity and accuracy of these findings and shows potential for future research using large textual datasets.

Osadchiy et al., 2020 (Osadchiy et al., 2020b) also employed a mixed-method approach, utilising classic qualitative analysis and NLP to understand patient anxieties in an online male infertility forum. The study used semantic-based analysis to translate the text into quantitative metrics related to different psychological processes (e.g., affective, social, cognitive and perceptual) and linguistic dimensions (e.g., parts of speech and grammar). The semantic analysis found that posts by men were more honest but had a more tentative or anxious style of writing compared with posts by women. Another study conducted by the same group of authors used NLP to understand patients' perceptions of hypogonadism and its treatment (Osadchiy et al., 2020a). These studies and our own suggested that NLP shows promise as an additional tool for qualitative and mixed-method research.

The current study is the first to measure the impact of an infertility diagnosis on the emotions of patients and partners using sentiment analysis. The overall sentiment score (–5 to 5) expresses the polarity and magnitude of the emotion of textual data and the sentiment score was found to be a potentially powerful tool to quantify and compare the overall sentiment of the textual data among subgroups. The overall sentiment scores were negative and similar across respondent subgroups by age, sex, patient-partner and employment status. ~~It~~Although patients and partners were not likely to be related in this survey, it might be worth noting that the couple may influence/mirror each other and therefore have similar responses (e.g., '*we both love children, so we felt lost and aimless*'). On the negative scale (–5 to 0), the sentiment score expresses the negativity of responses. Our study found that on the negative scale, sentiment score was slightly but significantly lower (more negative) for Spain, China, France and the UK compared with the US, while respondents in Canada, Australia and Germany had similar scores to respondents from the US, which may point to cultural differences. Hynie, Burns, 2006 (Hynie, Burns, 2006) suggested that stressors in individualist cultures (e.g., US or UK), that involve personal loss might be more significant and profound. In collectivist cultures (e.g., China or India), on the other hand, social pressure can often result in strained marital and social relationships for women and men who are infertile. Recent studies have shown that more women in China who are infertile experienced clinically significant symptoms of depression and anxiety than those in Hungary (Lakatos et al., 2017; Ni et al., 2021). Ni et al., 2021 suggested that observed

differences could be due to pressure caused by the traditional ideology and societal gender orientation in China. Our study aligned with the findings from Hynie et al. 2006, which showed the impact of the individualist-collective cultures on the emotions of women and men who are infertile. The multivariable linear regression analysis in our study showed that participants who self-identified as homosexual or 'other' felt more positive than heterosexual participants (p=0.03306). This may be linked to the desire to have children. Two studies in Italy and in the US found that gay men and lesbians were less likely than their heterosexual peers to express desire for parenthood hence may have been affected less negatively by a diagnosis (Riskind, Patterson, 2010; Baiocco, Laghi, 2013). Participants who were enrolled in at least one treatment were feeling more positive than those who were not enrolled in any treatment (p=0.02711). In Boivin et al., 2022, a correlation between seeking mental health support and receiving fertility treatment was observed. A significantly higher proportion of respondents seeking mental health support had received fertility treatment (48.4%) than those who did not seek support (38.7%). Mental health support and hope for parenthood through a positive treatment outcome could have a positive impact on patients and partners' emotions. One limitation of our sentiment analysis by patient-partner status is that the sentiment score might reflect the feelings of the respondent's partners, instead of their own feelings (e.g., '*she felt painful and really sad about it'*). Although the regression model had a low R-squared, our model was used for descriptive purposes, i.e., for capturing the association between respondents' characteristic and sentiment scores rather than for causal inference or prediction (Shmueli, 2010), where the interest is to predict new or future observations.

This study showed that additional emotional support is needed for both patients and partners, not only at the time of the infertility diagnosis but also during the subsequent, lengthy treatments. Only 43.6% of respondents received any type of social and psychological support post-diagnosis. According to the survey conducted in four European countries, 46% of women in treatment wish they had received more information about the emotions they were likely to experience (Domar et al., 2012). In van Empel et al., 2010 (van Empel et al., 2010), emotional support was one of two main categories of weakness identified in infertility care. Qualitative interviews with men found that participants had a

311 weak relationship with the medical professionals involved in their care. Participants also felt
312 dismissed from the treatment process and perceived a sense of blame from their healthcare providers
313 (Arya, Dibb, 2016). Since the 1,000 Dreams survey was not designed to provide insight into clinical
314 consequences or policy implications, additional research is needed to inform health policy.

315 There were some limitations to this study. First, the brevity of responses (a median of three words)
316 may have negatively impacted the interpretability of the topics, measurement of sentiment score and
317 characterisation of differences in emotions by respondent characteristics. Second, the sentiment score
318 in this study only summarised polarity (or negativity) of a response, it does not fully represent the
319 various categories and complexity of emotions (e.g., feeling sad, depressed and disappointed have the
320 same score of –2) or the composition of the same sentiment score (i.e., neutral sentiment or very polar
321 opposite sentiments). A machine learning or deep learning approach might be considered to detect
322 such types of feelings. The two above-mentioned limitations could have negative impacts on the
323 interpretation of results.

324 The potential limitations in the qualitative methodology pertained to whether the data included in the
325 analysis had enough depth to provide coverage of the relevant themes. TA centred on words
326 describing emotions since these were the predominant words in participant responses. Some inherent
327 subjectivity is present in all qualitative research if the researcher applies an unconscious bias to the
328 coding; therefore, the codes and results were reviewed by more than one scientist. The survey
329 question was asked at the time of enrolment into the survey, not at the time of diagnosis; therefore,
330 respondents might be subject to recall bias or be affected by the treatment outcome after the diagnosis.
331 Finally, the survey question was not worded to understand the causes of emotions; therefore, more
332 research will be needed to understand the reasons that drive patients' and partners' feelings and better
333 inform healthcare professionals.

334

CONTRIBUTORS

All authors contributed to data interpretation and contextualisation; all authors reviewed and approved the final version of the manuscript. JB, MO, MD and CS contributed to conceptualisation, design and methods of the study. JB, MO, MD, OC, DF, MM and CS were responsible for data analysis and quality control of the results. CS and WRL provided critical review of the results. JB, MD, DF, MM and WRL contributed to the writing of the manuscript.

ACKNOWLEDGEMENTS

The authors would like to thank the researchers at W2O Group who developed the survey and collected the results for analysis and Dave A. Andrae, Senior Research Scientist at Evidera for his advices on the linear regression model.

ROLE OF THE FUNDING SOURCE

Funding for this project was provided by Ferring Pharmaceuticals. Evidera received funding from Ferring Pharmaceuticals to conduct the study and develop this manuscript.

DECLARATION OF INTEREST

JB has received consultancy fees from Ferring Pharmaceuticals; received personal honoraria for lectures from Ferring Pharmaceuticals, Merck Group, Theramex, and MedThink; has received consultancy fees from UK Parliament; and has received travel expenses from the European Society of Human Reproduction and Embryology. MO, MD, OC, DF, and WRL were salaried employees of Evidera at the time of this work, and are not allowed to accept direct remuneration from any clients for their services. MM is an employee of Ferring Pharmaceuticals. CS has received consulting fees and has worked as an in-house consultant for Ferring Pharmaceuticals.

REFERENCES

Arya ST, Dibb B. The experience of infertility treatment: the male perspective. *Hum Fertil (Camb).* 2016;19(4):242-248.

Baiocco R, Laghi F. Sexual orientation and the desires and intentions to become parents. *J Fam Stud.* 2013;19(1):90-98.

Baumer E, Mimno D, Guha S, Quan E, Gay GK. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology.* 2017;68(6):1397-1410.

Biggs A, Brough P, Drummond S. Lazarus and Folkman's psychological stress and coping theory. In: *The handbook of stress and health: A guide to research and practice.* Hoboken, NJ, US: Wiley Blackwell; 2017:351-364.

Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research.* 2003;3(January):993-1022.

Boivin J, Vassena R, Costa M, et al. Tailored support may be required to reduce the impact of the infertility journey on mental health, relationships and daily lives of infertile patients and partners to infertile patients. *Reprod Biomed Online.* 2022.

Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology.* 2006;3(2):77-101.

De Berardis D, Mazza M, Marini S, et al. Psychopathology, emotional aspects and psychological counselling in infertility: a review. *Clin Ter.* 2014;165(3):163-169.

Domar A, Gordon K, Garcia-Velasco J, La Marca A, Barriere P, Beligotti F. Understanding the perceptions of and emotional barriers to infertility treatment: a survey in four European countries. *Hum Reprod.* 2012;27(4):1073-1079.

Domar A, Vassena R, Dixon M, et al. Barriers and factors associated with significant delays to initial consultation and treatment for infertile patients and partners of infertile patients. *Reprod Biomed Online.* 2021;43(6):1126-1136.

383 Friese S, Ringmayr T. *ATLAS. ti 8 Windows user manual.* Berlin: ATLAS. ti Scientific Software
384 Development GmbH; 2018.
385 Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the
386 practicing statistician. *Biom J.* 2018;60(3):431-449.
387 Himmel W, Reincke U, Michelmann HW. Text mining and natural language processing approaches
388 for automatic categorization of lay requests to web-based expert forums. *J Med Internet Res.*
389 2009;11(3):e25.
390 Hu M, Liu B. Mining and summarizing customer reviews. Paper presented at: ACM SIGKDD
391 International Conference on Knowledge Discovery and Data Mining2004; Seattle, WA.
392 Hynie M, Burns LH. Cross-cultural Issues in Infertility Counseling. In: Covington SN, Burns LH, eds.
393 *Infertility Counseling - a Comprehensive Handbook for Clinicans.* Cambridge: Cambridge
394 University Press; 2006:61-82.
395 Koert E, Harrison C, Bunting L, Gladwyn-Khan M, Boivin J. Causal explanations for lack of
396 pregnancy applying the common sense model of illness representation to the fertility context.
397 *Psychol Health.* 2018;33(10):1284-1301.
398 Lakatos E, Szigeti JF, Ujma PP, Sexty R, Balog P. Anxiety and depression among infertile women: a
399 cross-sectional survey from Hungary. *BMC Womens Health.* 2017;17(1):48.
400 Maroufizadeh S, Hosseini M, Rahimi Foroushani A, Omani-Samani R, Amini P. The Relationship
401 between Perceived Stress and Marital Satisfaction in Couples with Infertility: Actor-Partner
402 Interdependence Model. *Int J Fertil Steril.* 2019;13(1):66-71.
403 Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S, Stevens GA. National, regional, and global
404 trends in infertility prevalence since 1990: a systematic analysis of 277 health surveys. *PLoS*
405 *Med.* 2012;9(12):e1001356.
406 Mohammad S, Kiritchenko S, Zhu X. NRC-Canada: Building the state-ofthe-art in sentiment analysis
407 of tweets. Paper presented at: Seventh International Workshop on Semantic Evaluation2013.

Ni Y, Tong C, Huang L, Zhou W, Zhang A. The analysis of fertility quality of life and the influencing factors of patients with repeated implantation failure. *Health Qual Life Outcomes.* 2021;19(1):32.

Nielsen FA. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. *CEUR Workshop Proceedings.* 2011;No. 718:93-98.

Osadchiy V, Jiang T, Mills JN, Eleswarapu SV. Low Testosterone on Social Media: Application of Natural Language Processing to Understand Patients' Perceptions of Hypogonadism and Its Treatment. *J Med Internet Res.* 2020a;22(10):e21383.

Osadchiy V, Mills JN, Eleswarapu SV. Understanding Patient Anxieties in the Social Media Era: Qualitative Analysis and Natural Language Processing of an Online Male Infertility Community. *J Med Internet Res.* 2020b;22(3):e16728.

Ozdemir C, Bergler S. ~~A Comparative Study of Different Sentiment Lexica for Sentiment Analysis of Tweets. Paper presented at: Recent Advances in Natural Language Processing2015b; Hissar, Bulgaria~~CLaC- ¨ SentiPipe: SemEval2015 Subtasks 10 B, E, and Task 11. Paper presented at: International Workshop on Semantic Evaluation2015a; Denver, CO.

Ozdemir C, Bergler S. A Comparative Study of Different Sentiment Lexica for Sentiment Analysis of Tweets. Paper presented at: Recent Advances in Natural Language Processing2015b; Hissar, Bulgaria.

Reagan AJ, Danforth CM, Tivnan B, Williams JR, Dodds PS. Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. *EPJ Data Science.* 2017;6(1):28.

Riskind RG, Patterson CJ. Parenting Intentions and Desires Among Childless Lesbian, Gay, and Heterosexual Individuals. *J Fam Psychol.* 2010;24(1):78–81.

Rooney KL, Domar AD. The relationship between stress and infertility. *Dialogues Clin Neurosci.* 2018;20(1):41-47.

Shmueli G. To Explain or to Predict? *Statistical Science.* 2010;25(3):289-310.

Simionescu G, Doroftei B, Maftei R, et al. The complex relationship between infertility and psychological distress (Review). *Exp Ther Med.* 2021;21(4):306.

Socher R, Perelygin A, Wu J, et al. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. Paper presented at: EMNLP2013.

van Empel IW, Nelen WL, Tepe ET, van Laarhoven EA, Verhaak CM, Kremer JA. Weaknesses, strengths and needs in fertility care according to patients. *Hum Reprod.* 2010;25(1):142-149.

Vander Borght M, Wyns C. Fertility and infertility: Definition and epidemiology. *Clin Biochem.* 2018;62:2-10.

Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phraselevel sentiment analysis. Paper presented at: Conference on Human Language Technology and Empirical Methods in Natural Language Processing2005; Vancouver, Canada.

448 **TABLES**

449 **Table 1. Characteristics of Respondents (N=1,795)**

	N=1,795
Length of responses (number of words)	
Median (p25%, p75%)	3 (2, 5)
Minimum–maximum	1–32
Age (years), mean (SD)	36.0 (9.7)
Median (p25%, p75%)	35 (29, 41)
Minimum–maximum	18–77
Age categories (years), n (%)	
18–30	555 (30.9)
30–40	784 (43.7)
40-77	456 (25.4)
Sex, n (%)	
Female	1,026 (57.2)
Male	767 (42.7)
Other	2 (0.1)
Group, n (%)	
Patients	963 (53.6)
Partners	832 (46.4)
Sex, patient-partner status, n (%)	
Female, partner	378 (21.1)
Female, patient	648 (36.1)
Male, partner	453 (25.2)
Male, patient	314 (17.5)
Other partner	1 (0.1)
Other patient	1 (0.1)
Sexual preference, n (%)	
Heterosexual	1,642 (91.5)
Homosexual	98 (5.5)
Other	55 (3.1)
Marital status, n (%)	
Divorced	70 (3.9)
Married	1,031 (57.2)
Single	338 (18.8)
Other	87 (4.8)
Missing	264 (14.7)
Employment, n (%)	
Full-time	1,435 (79.9)

	N=1,795
Part-time	197 (11.0)
Unemployed	163 (9.1)
Country, n (%)	
Australia	191 (10.6)
Canada	183 (10.2)
China	239 (13.3)
France	176 (9.8)
Germany	173 (9.6)
Italy	186 (10.4)
Spain	203 (11.3)
United Kingdom	182 (10.1)
United States	262 (14.6)
Year(s) spent trying to conceive prior to infertility diagnosis (mean [SD])	3.23 (2.44)
Treatment engagement and outcome at the time of the survey, n (%)	
Enrolled in 1 treatment, successful	214 (11.9)
Enrolled in 1+ treatment, unsuccessful	237 (13.2)
Enrolled in 2+ treatments, successful	325 (18.1)
Infertility diagnosis, no consultation	644 (35.9)
Infertility diagnosis, no treatment enrolment	375 (20.9)
Psychological history of psychological disorders by the time of diagnosis, n (%) yes	952 (52.8)
Eating disorder	129 (7.2)
Obesity	272 (15.1)
Depression	566 (31.4)
Anxiety	602 (33.4)
Post-traumatic stress disorder	166 (9.2)
Obsessive-compulsive disorder	133 (7.4)
Social and psychological support provided, n (%) yes	
Specialists, such as psychologists, therapists or social worker	356 (19.8)
Local support groups	167 (9.3)
Online support groups, online forums	247 (13.7)
Family without experience with fertility treatment(s)	152 (8.4)
Friends without experience with fertility treatment(s)	155 (8.6)
Partner	368 (20.4)
Healthcare provider	322 (17.9)
Friends or family with fertility treatment(s) experience	276 (15.3)
Any support	744 (43.6)

Abbreviation: SD = standard deviation

Table 2. Overall Sentiment Score and Sentiment Score by Respondent Characteristics (n=1,708)

Sentiment Score	Mean (SD)	Median (Minimum, Maximum)
Overall	-1.32 (1.5)	-2 (-4, 4)
Age (years)		
17–25	-1.15 (1.59)	-2 (-3, 3)
25–30	-1.33 (1.43)	-2 (-3, 3)
30–35	-1.33 (1.45)	-2 (-4, 4)
35–40	-1.36 (1.52)	-2 (-3, 4)
40–45	-1.38 (1.62)	-2 (-3, 3)
45–77	-1.35 (1.47)	-2 (-3, 3)
Sex		
Female	-1.32 (1.47)	-2 (-3, 4)
Male	-1.33 (1.54)	-2 (-4, 4)
Other	-2.5 (0.71)	-2.5 (-3, -2)
Group		
Partner	-1.37 (1.5)	-2 (-3, 4)
Patient	-1.28 (1.5)	-2 (-4, 4)
Sex, patient-partner status		
Female partner	-1.42 (1.46)	-2 (-3, 4)
Female patient	-1.25 (1.48)	-2 (-3, 3)
Male partner	-1.32 (1.54)	-2 (-3, 3)
Male patient	-1.35 (1.53)	-2 (-4, 4)
Sexual preference		
Heterosexual	-1.35 (1.49)	-2 (-4, 4)
Homosexual	-1.06 (1.69)	-2 (-3, 3)
Other	-0.94 (1.5)	-1 (-3, 4)
Employment		
Full-time	-1.31 (1.52)	-2 (-4, 4)
Part-time	-1.37 (1.47)	-2 (-3, 4)
Unemployed	-1.38 (1.37)	-2 (-3, 2)
Country		
Australia	-1.2 (1.48)	-2 (-3, 3)
Canada	-1.2 (1.6)	-2 (-3, 3)
China	-1.3 (1.31)	-2 (-3, 3)
France	-1.49 (1.48)	-2 (-3, 3)
Germany	-1.27 (1.66)	-2 (-4, 3)
Italy	-1.37 (1.65)	-2 (-3, 3)
Spain	-1.65 (1.39)	-2 (-3, 3)
United Kingdom	-1.51 (1.43)	-2 (-3, 3)
United States	-1.04 (1.49)	-2 (-3, 4)
Treatment engagement and outcome at the time of the survey, n (%)		
Enrolled in 1 treatment, successful	-1.12 (1.63)	-2 (-3, 3)
Enrolled in 1+ treatment, unsuccessful	-1.37 (1.47)	-2 (-3, 2)
Enrolled in 2+ treatments, successful	-1.21 (1.68)	-2 (-3, 3)

Sentiment Score	Mean (SD)	Median (Minimum, Maximum)
Infertility Dx, no consultation	-1.35 (1.45)	-2 (-3, 4)
Infertility Dx, no treatment enrolment	-1.47 (1.35)	-2 (-4, 3)

452 Abbreviations: Dx = diagnosis; SD = standard deviation

Table 3. Association of Respondent Characteristics and Sentiment Score on Negativity Scale (-4 to 0)—Multivariable Linear Regression

Variables	Estimate	p-value
Canada vs the United States	-0.11	0.18317
Australia vs the United States	-0.09	0.27521
Germany vs the United States	-0.09	0.32458
China vs the United States	-0.25	0.00317
Italy vs the United States	-0.09	0.32816
France vs the United States	-0.23	0.00940
The United Kingdom vs the United States	-0.24	0.00446
Spain vs the United States	-0.37	<0.001
Infertility Dx, no treatment enrolment vs. enrolled in 1+ treatment	-0.12	0.02711
Infertility Dx, no consultation vs. enrolled in 1+ treatment	-0.02	0.61583
Time spent trying to conceive prior to infertility diagnosis ≥2 years vs. <2 years	-0.12	0.01000
Sexual preference Homosexual or others vs. Heterosexual	0.16	0.03306
Marital status Married vs. Divorced	-0.17	0.12276
Marital status Single vs. Divorced	-0.22	0.05888
Marital status Widowed vs. Divorced	-0.16	0.55874
Marital status Missing or Others vs. Divorced	-0.35	0.00339

Residual standard error: 0.8619 on 1691 degrees of freedom (DF); Multiple R-squared: 0.02927, Adjusted R-squared: 0.02009; F-statistic: 3.187 on 16 and 1691 DF, p-value: 1.945e-05

Abbreviation: Dx = diagnosis

Figure legends

Figure I. Topic Modelling and Sentiment Analysis Process

Figure II. NLP Analyses - Top Words Frequently Mentioned in All Responses - Overall and by Patient-Partner Status

Figure III. Thematic Analysis - Theme Distribution - Overall and by Patient-Partner Status

Supplementary Figure I. Words with the Highest Probability from Each Topic

FIGURES

Commented [MG1]: Figures will need to be submitted as high-res images with the Caption and footnotes left in the main document. Would need originals for this.

Figure I. Topic Modelling and Sentiment Analysis Process

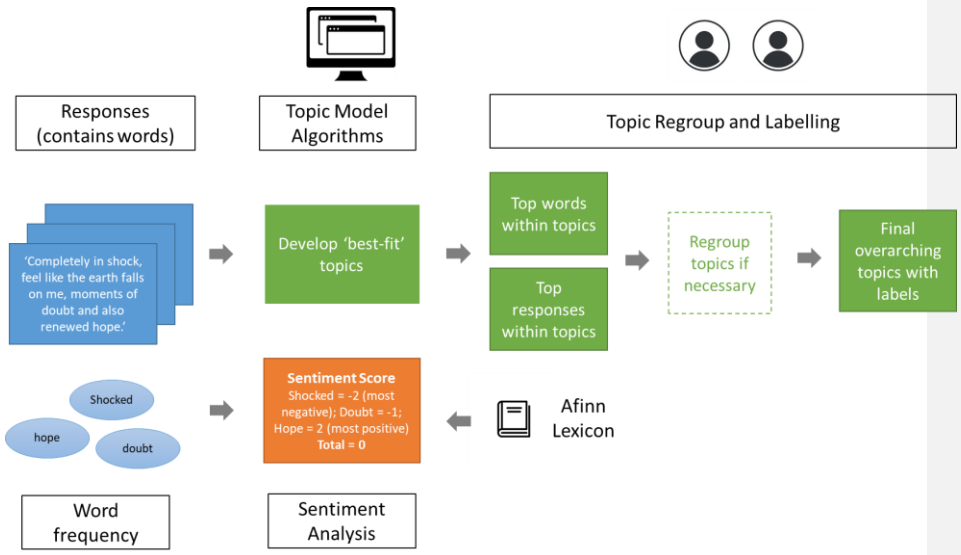


Figure II. NLP Analyses ~~- Top Words~~ Most Frequently Mentioned ~~in All Responses -~~ Words Overall and by Patient-Partner Status

Commented [MD2]: I added 'overall' and organize in the same way as the TA results (figure III)

Commented [MM3R2]: The new heading does not say what is reported in the figures. Either add to heading or a note to the table

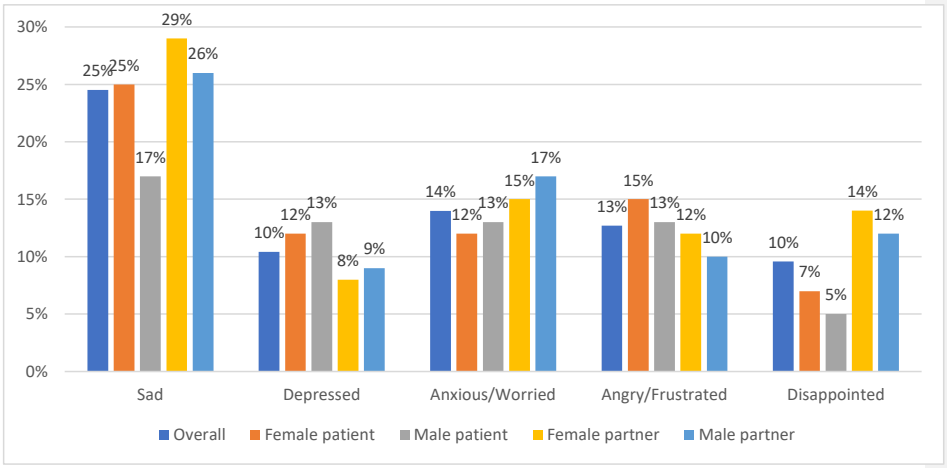
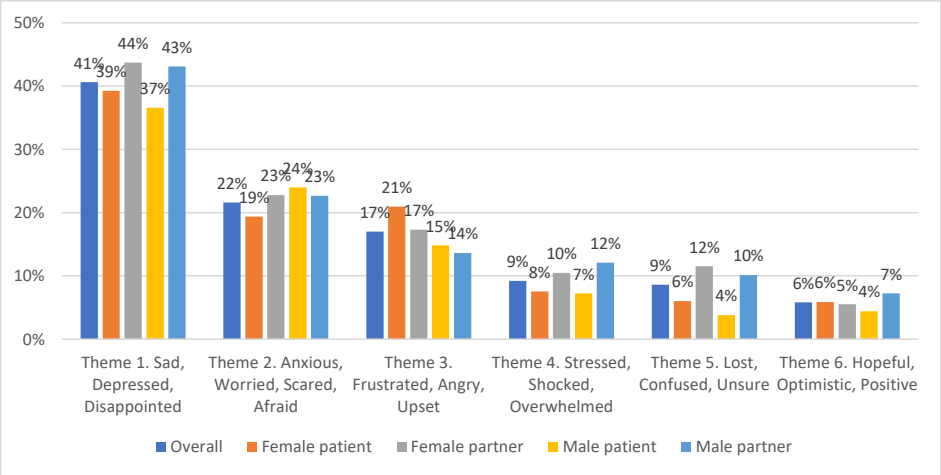

30%
25%
20%
15%
10%
5%
0%
25%
25%
17%
29%
26%
10%
12%
13%
8%
9%
14%
12%
13%
15%
17%
13%
15%
13%
12%
10%
10%
7%
5%
14%
12%
Sad
Depressed
Anxious/Worried
Angry/Frustrated
Disappointed
Overall
Female patient
Male patient
Female partner
Male partner

Figure III. Thematic Analysis—Theme Distribution—Overall and by Patient-Partner Status



SUPPLEMENTARY ~~TABLES AND FIGURES~~ MATERIAL

Commented [MG4]: Submit file separately

Supplementary Table I. Topics Identified by NLP Topic Modelling

Inserted Cells

Topics	Overarching topics	Example of Highest Probability Responses
1. Relief and acceptance as infertility being suspected, fear but motivation for future treatment OR infertility diagnosis not expected, feeling devastated, unacceptance and fear for the future.	Acceptance or unacceptance of the diagnosis	'My sister suffers from the same condition (polycystic ovary syndrome) so I was prepared...' 'It was at last proved something was wrong and it is not the end. Had some medications and some diet to increase it.' 'Relieved, it was treatable so it gave us an answer.'
2. Feeling hurt, shocked, depressed, loss of appetite, loss of confidence toward partner, hopeless about the future	Acceptance or unacceptance of the diagnosis	'Hurt, frustrated, I had lost my courage and my confidence to continue living with him.' 'Anxious, loss of appetite, gained weight, loss of sleep, unhappy, lack of understanding.' 'How could this be, I don't believe it. Afterwards I often lost my temper, and lost confidence and enthusiasm in my life.'
3. Unexpected results, feeling confused, disappointed about the unexplained infertility, looking for solutions	Acceptance or unacceptance of the diagnosis	'Disappointed in lack of knowledge from health care providers regarding reasons for not being able to conceive as we were told we are part of the 20% unexplained group due to insufficient medical testing available at this time, wife suffered a period of major depression due to inability to conceive.' 'Shocked and confused about the new information coming to light and the shock from it.' 'This diagnosis was a great burden for us, since at that time we had a strong wish for children and afterwards we considered getting more detailed advice and researching other options for fulfilling the wish for children.'
4. Unacceptance, anger against oneself and the world, feeling worthless, impotent	Acceptance or unacceptance of the diagnosis	'Struck down, sad and angry against this misfortune.' 'Pessimistic, feeling of not understanding. Why do I suffer from this disease? I couldn't get my head around it.' 'Feeling that the world is falling apart, that we are different from others…'
5. Desire to have a baby, feeling terrible about the diagnosis	Recollections about causes of infertility, treatment experience, and concerns about treatment cost and safety	'…I struggled with my mental health, avoided going to baby showers and cried when I saw other moms with their babies sometimes. I even convinced myself I didn't really want kids and I'd be a horrible mother, so it would hurt so much that I couldn't.' 'Horrible and sad at the fact I couldn't carry a baby because of the shape of my uterus.'
6. Self-frustration, anxiety, stress, uncertainty, looking for solutions	Acceptance or unacceptance of the diagnosis	'Very stressed out, while waiting and uncertainty, the pain prevails, guilt' 'A lot of uncertainty and confusion…' 'Extremely anxious and extremely nervous. I didn't know what to do except to look for an effective way to solve the problem.'
7. Feeling difficult to accept, hopeless	Acceptance or unacceptance of the diagnosis	'Difficult to accept, but nothing can be done about it…' 'Hopeless and as if there is nothing we can do. worried about expenses of IVF and difficulty of adopting.' 'It is difficult to accept, I feel like the sky is falling, there is absolutely no way to accept or imagine.'
8. Shared feelings with partners, concerns about relationship with partners, self-blaming	Relational reflections involving shared feelings with partners, concerns	'Anxious, restless, disappointed, I was scared our marriage would change…' 'She was very upset, very hurt. We didn't know what to do, I also didn't know how to comfort her.'

Topics	Overarching topics	Example of Highest Probability Responses
	about relationships, and impact on life	'I thought I was going to collapse, I felt guilty for being with my partner.'
9. Feeling disappointed but thinking about potential treatment and adoption	Acceptance or unacceptance of the diagnosis	'Disappointment, anger, incomprehension, but also desire to overcome this ordeal somehow, using scientific tools such as medical investigation, tests and specialists' help, but also with my faith in God.' 'We both felt very uncomfortable, since we actually wanted to start a family. For some time we were emotionally almost at rock bottom, but then we pulled ourselves back up and then adopted children - one boy and one girl. Now we are very happy and it doesn't bother us at all anymore...'
10. Feelings towards partner after diagnosis, worried about impact on life, relationships, family	Relational reflections involving shared feelings with partners, concerns about relationships, and impact on life	'I felt sorry for him that he had to go through the depression but I was angry that it was not only affecting our relationship but it was affecting our sex life our family life and our social life so we all suffered and missed out on so much.' 'I still feel devastated but have friends in similar situations going through IVF so it is reassuring to know that it doesn't have to be the end of our dreams for a family.'
11. Recall about causes of infertility (medical history), treatment journey	Recollections about causes of infertility, treatment experience, and concerns about treatment cost and safety	'I expected that. My periods have always been irregular. Sometimes I would only have 2 per year. I knew long ago that I was suffering from polykistiques ovaries, but it would have taken 9 years and several doctors to receive a written diagnosis.' 'I knew that it would happen because of his kidney failure and dialysis...' 'I found out I was infertile after I was diagnosed with breast cancer…'
12. Hoping for a treatment	Acceptance or unacceptance of the diagnosis	'Completely in shock, feel like the earth falls on me, moments of doubt and also renewed hope.' 'Panicked, to have children, I will pro-actively cooperate with the hospital's primary care physician.' 'I'm still very young, I will work with my primary care physician to treat my illness.'
13. Feeling inadequate as a woman/man, guilt~~, eating disorder~~	Acceptance or unacceptance of the diagnosis	'… I thought my husband no longer saw me as a "proper" woman…' 'Like I couldn't do the one thing that a is supposed to happen naturally for a woman.' 'Devastated. I felt less of a woman and like I was letting my husband down.' '…Knowing I could not give my wife the one thing she desired most killed me inside as I felt it was all my fault…'
14. Fertility treatment experience, cost, safety	Recollections about causes of infertility, treatment experience, and concerns about treatment cost and safety	'…We didn't preserve eggs or sperm due to the cost. We have no ability to reproduce with our dna now. Work insurance now covers thus basic need…' 'Anxious, alone, and overwhelmed by how we could possibly afford such expensive treatments since insurance deemed infertility treatments an "unnecessary luxury" ' 'I was devastated as our first fertility doctor had matter of fact told me that I may have gone through premature menopause before conducting any tests and put doubt in my mind that we would ever have children…'

Inserted Cells

476

477

Supplementary Table II. Overall Sentiment Score by Composition of Most Positive and the Most Negative Words

Sentiment Score	Most Negative Word's Score	Most Positive Word's Score	Frequencies
-4	-4	0	0.1%
-3	-4	1	0.1%
-3	-3	0	20.5%
-2	-4	2	0.1%
-2	-3	1	1.8%
-2	-2	0	38.2%
-1	-4	3	0.1%
-1	-3	2	4.4%
-1	-2	1	4.9%
-1	-1	0	1.7%
0	-3	3	1.1%
0	-2	2	12.2%
0	-1	1	0.2%
0	0	0	2.5%
1	-2	3	2.4%
1	-1	2	1.1%
1	0	1	1.6%
2	-1	3	0.3%
2	0	2	4.5%
3	0	3	2.2%
4	0	4	0.1%

Note. Overall sentiment score range -4 to 4. The sentiment score = sum of the most negative and positive word score. The table shows that values of sentiment scores (-4 to 4) can be achieved through different compositions of positive and negative words. For example, sentiment score 0 could be due to highly negative and highly positive words that cancel each other out (of varying intensities) or to neutral emotions.

Supplementary Table III. Association of Respondent Characteristics and Overall Sentiment Score (-5 to 5)—Multivariable Linear Regression

Variables	Estimate	p-value
Intercept	-0.94	< 0.001
Canada vs United States	-0.19	0.21063
Australia vs United States	-0.14	0.32674
Germany vs United States	-0.21	0.15917
China vs United States	-0.23	0.09642
Italy vs United States	-0.32	0.03681
France vs United States	-0.43	0.00440
United Kingdom vs United States	-0.47	0.00136
Spain vs United States	-0.63	<0.001
Infertility Dx, no treatment enrolment vs. enrolled in 1+ treatment	-0.28	0.00370
Infertility Dx, no consultation vs. enrolled in 1+ treatment	-0.18	0.02779
Sexual preference Homosexual or others vs. Heterosexual	0.34	0.00827

Residual standard error: 1.506 on 1696 degrees of freedom (DF); Multiple R-squared: 0.0246; Adjusted R-squared: 0.01828; F-statistic: 3.889 on 11 and 1696 DF, p-value: 1.404e-05

Abbreviation: Dx = diagnosis

Supplementary Table IV. Thematic Analysis Candidate Themes and Frequencies Identified using Thematic Analysis

Candidate Theme	Illustrative Quote and Respondent	Frequency of Respondents in Each Candidate Theme (%) (N=1,795)
1. Sad, Depressed and Disappointed	*'Sorrow, I felt really bad, very sad, as if life ended right there.' Heterosexual female patient*	40.6%
2. Anxious, Worried, Scared and Afraid	*'Scared and worried that I may never be able to have a child. Concerned about the cost of treatment and whether the relationship would hold up through all the pressure.' Heterosexual female patient*	21.6%
3. Frustrated, Angry and Upset	*'Angry against myself and against my body and against the whole world, why me?' Homosexual male patient*	17%
4. Stressed, Shocked and Overwhelmed	*'it was a surprise because we did not expect it at all. We have a teenage son and we did not have any problems with him' Heterosexual male partner*	9.2%
5. Lost, Confused and Unsure	*'I did not know how to communicate with my other half, I did not know how to comfort her.' Heterosexual male partner*	8.6%
6. Hopeful, Optimistic and Positive	*'Felt hopeful that the treatments and procedures would work and work quickly for us.' Homosexual female patient*	5.8%

Supplementary Figure I. Words with the Highest Probability from Each Topic