# - Methods

This chapter outlines the rationale for, implementation of, and some reflections on the methodology selected for use in this study. The first section outlines the decision to apply a mixed-methods Critical Realist (CR) approach to effectively address the research questions and meaningfully compare the *realist/ positivist* biomedical with *social constructivist* critical epistemological perspectives associated with PMS research. The second section describes the various data collection processes involved in the qualitative part of this study; a series of semi-structured interviews with 16 of the world’s top PMS experts, and 12 self-identified PMS ‘patients’. The third section outlines the five steps involved in the analysis of the interview data. This includes data transcription, three iterative analytical steps associated with CRDA (Critical Realist Discourse Analysis), and the triangulation of participant descriptions of PMS with robust epidemiological data. The fourth and final section discusses some of the main benefits, challenges, and limitations of the methodology.

## 3.1 Methodological approach

### Research design

The idea to compare expert with patient descriptions of PMS was partly inspired by a paper that did so regarding the symptoms of ‘endometriosis’ (Fauconnier et al., 2013). It found that twelve endometriosis symptoms were described by patients only (mainly pain-related), and five symptoms by expert gynaecologists only (mainly much rarer experiences). Some gynaecologists also listed depression as a core symptom, yet *none* of the forty-one patients mentioned it (Fauconnier et al., 2013, p. 2690). While the significance of these particular discursive differences was not discussed within the endometriosis study, I thought they might possibly reflect documented clinical (and wider societal) tendencies to minimise female experiences of pain (Fitzgerald & Hurst, 2017; Knight, Bunch, Kenyon, Tuffnell, & Kurinczuk, 2020), and to overemphasise female experiences of emotional distress (Sayers, 2015; Ussher, 2011). I was, therefore, keen to see if any similar patterns might emerge in expert versus patient descriptions of PMS, given its highly contested existence as a biomedical condition.

The way in which to conduct such a comparison was less easily decided, especially because there are advantages and limitations to both qualitative and quantitative research approaches. Quantitative methods can provide a more generalisable account of the ‘regularities that characterise a population’ (Downward & Mearman, 2007, p. 90), but are inherently limited by the typically ‘closed’ questions asked of participants (subject to pre-existing theoretical assumptions) as well as the size and composition of the population sample (Bryman, 2012, p. 408). For instance, nearly all PMS population studies use a predetermined list of symptoms rather than letting participants describe their experiences without any ‘priming’. The findings, therefore, may reflect and strengthen the researchers’ own assumptions about what counts as a premenstrual symptom rather than (as is often implied) providing a ‘neutral’ account of participant experiences.

In contrast, qualitative methods are typically more embedded in the experiences and perspectives of participants and can provide a more detailed explanatory account of why and how an observable phenomenon may occur (Bryman, 2012, p. 408). The problem being that the intensive nature of this type of research significantly reduces the number of participants, which prevents statistical analysis and the generalisation of findings to a wider population. For example, the findings of Laws’ qualitative study of 14 explicitly ‘anti-sexist’ men’s attitudes towards menstrual health issues (including PMS) could not be positioned as generalisable to a larger male population, “*no claim is made that the men I interviewed are representative of other men*” (Laws, 1992, p. 120).

Of greatest concern was that the methodological approach selected was relevant, effective, and appropriate for answering the research questions (and within the scope of doctoral research) (Bryman, 2012, p. 41). Three research questions were developed in response to the various contradictions identified in the literature on societal and biomedical definitions and descriptions of PMS.

RQ1. How and why are certain premenstrual changes prioritised over others in expert and patient descriptions of PMS?

RQ2. How do expert and patient descriptions of PMS compare with robust epidemiological data on premenstrual changes?

RQ3. What are the clinical, academic, and social implications of these expert and patient descriptions of premenstrual changes?

RQ1 was best addressed through qualitative inquiry because it required rich in-depth data to analyse *how* and *why* individuals may describe PMS in a certain way. In contrast, RQ2 and RQ3 demanded a comparison of these (*qualitatively* derived) participant discourses with robust *quantitative* epidemiological data about the type, frequency, and prevalence of problematic premenstrual changes in the general menstruating population (the rationale for doing so is discussed below). As a result, a mixed-methods research design was selected as the most appropriate way in which to answer these specific research questions.

As outlined in the previous chapter, PMS has typically been studied using *either* qualitative *or* quantitative methods. Biomedical PMS research is almost exclusively derived from quantitative epidemiological or experimental methods, whereas critical PMS studies tend to be either qualitative participant interviews or discourse analysis, or quantitative attitudinal surveys or experimental approaches. As far as I am aware, this must be one of the first PMS studies to apply a mixed methods research approach.

### Theoretical Framework

While the distinction between qualitative and quantitative research paradigms is far from binary, or exclusive (Bryman, 2012, p. 409), their typical features do also tend to align with different epistemological and ontological positions (Danermark et al., 2019; McEvoy & Richards, 2006). The previous chapter outlined why taking a Critical Realist (CR) meta-theoretical perspective is, thus, of value when trying to compare, and/or integrate, biomedical (realist/ positivist) and critical (social constructionist) descriptions of PMS. In addition, taking a CR position can also overcome some of the analytical issues associated with the use of mixed methods (Danermark et al., 2019; Olsen, 2014). For instance, the epistemological and ontological differences between qualitative and quantitative research paradigms, necessarily produce different types of data. It may, therefore, seem difficult or even inappropriate to integrate highly contextualised interpretative data with quantitative data that establish statistically significant empirical generalisations (Cresswell & Plano Clark, 2011).

CR transcends this problem by positioning both types of data as necessarily *interpretive* (informed by discourse and the embodied, material, and institutional constraints influencing the researcher’s experience and knowledge of the world); *partial* (in that they can only ever account for observable empirical phenomena, which may also be different facets of the same, partly unobservable, mechanism e.g., psychosocial versus biological factors in, or experiences of, premenstrual change); and *transitive* (in that these empirical observations are made at a certain point in time and context and, thus, subject to these conditions) (Owens, 2011). By positioning all knowledge as interpretive, partial, and transitive, both types of data and findings are rendered as *contingent* descriptions of the ‘domain of the empirical’, rather than direct or impartial accounts of the domain of the ‘real’ (Owens, 2011, p. 8). Using this theoretical framework actively prevents the common methodological epistemic fallacy of conflating epistemology with ontology i.e., the assumption that what is observed of a phenomenon represents its objective, whole, or intransitive (unchanging) ontological reality (Danermark et al., 2019).

In practical terms, a CR theoretical framework allows the mixed-methods researcher to ‘sensibly combine’ and compare the relative (albeit contingent) explanatory power of both quantitative and qualitative data, regarding how well they can account for what can be observed of any given phenomenon (Danermark et al., 2019; Downward & Mearman, 2007, p. 77). In this case, comparing qualitative expert and patient descriptions of PMS with robust quantitative epidemiological data, to identify any notable similarities or differences in their epistemological claims. Indeed, by comparing these different types of data, CR asserts that new theoretical insights into the mechanisms behind any given phenomenon are made possible, through the application of a type of inferential logic known as ‘retroduction’ (Danermark et al., 2019; Downward & Mearman, 2007). This type of mixed-methods data triangulation is, therefore, useful in two ways; to (contingently) ‘verify’ epistemological claims made about ‘what PMS is’, as well as reveal ‘different features of the same layered reality’ of problematic premenstrual changes, which may be indicative of their underlying mechanisms (Downward & Mearman, 2007, p. 92).

In short, a CR theoretical framework provided a means by which to effectively compare, integrate, and make informed *theoretical* generalisations about, the different epistemological and ontological approaches and data associated with a mixed-methods study of PMS (e.g., realist/ positivist versus social constructivist, psychosocial versus biological, qualitative versus quantitative).

### Critical Realist Discourse Analysis (CRDA)

While CR is a practice-focused philosophy, there are no specific methods associated with it (Danermark et al., 2019; McEvoy & Richards, 2006). Indeed, it is possible to employ most research methods within a CR theoretical framework. As a result, while the main research question clearly implicated some sort of Discourse Analysis (DA) in order to assess *how* experts and patients describe PMS, it was not immediately clear which methods of data collection and analysis would be of most value.

A review of the various methods available was conducted, with semi-structured interviews emerging as probably the most suitable and convenient way in which to elicit the rich discursive data required for DA. For example, participant observation and other ethnographic methods would have been difficult to implement given the global dispersal of the PMS experts. Focus groups would also have been difficult to arrange for the same reason, and the presence of others could have resulted in ‘peer pressure’ to conform with group consensus positions (Hollander, 2004). Archival or other text-based analyses were limited by a lack of sufficient documentation of the ISPMD consensus-building meetings regarding the current formal biomedical definitions of PMS/ PMDD[[1]](#footnote-1). Another strength of the semi-structured interview was that it enabled the use of a detailed schedule of questions, to ensure comprehensive topic coverage and to allow for more direct comparison between participant responses. This would not have been possible in less structured data gathering approaches such as participant observation, or conversation analysis (of naturally occurring talk).

The literature review process indicated that some form of ‘critical’ discourse analysis (CDA) would be of value, especially given the highly contested, stigmatised, and gendered nature of PMS discourses. Initially, Foucauldian genealogical discourse analysis (Jager & Maier, 2016) and the Discourse-Historical Approach (DHA) (Reisigl & Wodak, 2016) were both under serious consideration, due to their interdisciplinary orientations and attention to the historical and political context of discourse. In fact, the Critical Realist Discourse Analysis (CRDA) methodology that was ultimately selected (described below), directly builds on the tenets of Foucauldian discourse analysis. The main reason why CRDA was found to be a better fit was its explicit focus on integrating data about, and analysis of, the ‘extra-discursive’ embodied, material, and institutional contexts within which individuals choose/ are compelled to use certain discourses.

CRDA was first described as a distinct methodology by Sims-Schouten, Riley and Willig in their example study ‘Women’s Talk of Motherhood, Childcare and Female Employment’ (2007). The methodology was later refined by Sims-Schouten and Riley (2014) and presented again in another example study, ‘Presenting Critical Realist Discourse Analysis as a Tool for Making Sense of Service User’s Accounts of the Mental Health Problems’ (2019). I had to slightly adapt the full CRDA methodology (designed for larger multi-researcher projects) to fit the scope of doctoral study (described in more detail in section 3.4) but was able to implement all of the key steps outlined in the table below [Table 3.1].

Table . The three phases of Critical Realist Discourse Analysis (CRDA) (Sims-Schouten & Riley, 2014)

Shape

Description automatically generated with medium confidence

### Mixed-Methods Data Triangulation

As mentioned above, the second and third research questions relied on some form of ‘data triangulation’ to compare participant discourses with *robust* quantitative epidemiological data. The triangulation of expert and patient PMS discourses with the available epidemiological data on premenstrual symptoms was conducted for the purposes outlined by Olsen (2007, p. 4):

Triangulation is something we do in order to generate a dialectic of learning. Triangulation means mixing approaches to get two or three viewpoints upon the things being studied. The resulting dialectic of learning thrives on the contrasts between what seems self-evident in interviews, what seems to underlie the lay discourses, what appears to be generally true in surveys, and what differences arise when comparing all these with official interpretations of the same thing.

This step went beyond the CRDA method outlined by Sims-Schouten and Riley (2014), but was in line with other CR data triangulation approaches (Danermark et al., 2019; Downward & Mearman, 2007; McEvoy & Richards, 2006; Modell, 2009; Olsen, 2014). Due to time constraints and the known existence of largescale, accessible, and reasonably robust quantitative data, it was decided that secondary data would be sufficient for this purpose. In total, only four secondary data sources were used to triangulate the epistemological claims made by the interview participants [Table 3.2]. They were selected based on their suitability (accessible raw data on symptom type, timing, prevalence, and severity within a menstruating population) and quality (large sample size, randomised over selected population sample, cross-cultural over one country context, theoretically robust and transparent sampling method/ study inclusion criteria, prospective over retrospective data, some measure of symptom severity included, and systematic reviews of multiple studies over single studies).

I started with Mallia’s (2015) systematic review of PMS population studies since it is the most recent of only two systematic reviews ever conducted of the epidemiological data (the other one being Dennerstein et al. (2012)). This provided robust quantitative data regarding the prevalence of premenstrual symptoms in samples of the wider menstruating population. I then read through all 26 studies included in Mallia’s review and found that *only* Dennerstein *et al*. (2011) qualified as a largescale cross-cultural population study that had also published the raw data regarding premenstrual symptom type, relative prevalence, duration, and severity, along with comprehensive demographic information about the randomised sample population. The Romans *et al.* (2012) data was included as the *only* systematic review of prospective epidemiological studies on the timing and prevalence of mood-related symptoms in the wider menstruating population (plus some data on non-menstruating ‘control’ populations). Finally, the Gold *et al.* (2016) study was included as the *only* largescale randomised and ethnically diverse population study that has identified a statistically significant association between a form of biomarker and premenstrual symptoms. As always, and especially because these data sources were often the ‘only’ robust largescale quantitative data available, they were also subject to various methodological and epistemological limitations [outlined in Table 3.2]. Overall, however, I was satisfied that the quality of data provided was sufficient for the purposes of discourse triangulation.

Table . The selection criteria for, and limitations of, the four epidemiological datasets used to triangulate the epistemological claims made in participant discourses.

Shape

Description automatically generated with medium confidence

Data triangulation enhanced this study by enabling the comparison and ranking of participant discourses based on their relative ability to describe observed patterns in quantitative epidemiological datasets regarding premenstrual symptoms. This is a form of ‘judgemental rationalism’, a core tenet of CR, which derives from its commitment to ontological realism. In short, just because all knowledge is necessarily interpretive, “*it does not follow that any interpretation is as good as another”* (Owens, 2011, p. 10)*.* For instance,inferences that are replicable or corroborated by numerous additional data sources are more robust than those that are not. This is not to imply that the quantitative epidemiological data with which the participant discourses were compared, were ‘more’ objectively descriptive of what PMS ‘really is’ (i.e., ontological rather than epistemological data), just that they provide a reasonably robust interpretive, contingent, and partial description of what has been ‘generally’ observed of the phenomenon across numerous studies, albeit subject to the methodological and epistemological limitations of quantitative approaches.

## 3.2 Data collection

### Participant selection

All participants were selected using a purposive sampling approach to ensure that reasonably diverse yet qualified perspectives on PMS were represented within the small sample size required for in-depth qualitative analysis (Palinkas et al., 2015). Sampling in this way typically involves identifying individuals that are especially knowledgeable about (i.e. the experts) or experienced with the research topic (i.e. the patients) (Cresswell & Plano Clark, 2011), which fits the research questions very well. Plus, since there are only very few PMS research specialists in the world, the use of more random or ‘probabilistic’ sampling methods were simply not feasible for this group.

While the patient group had a much larger population from which to sample, efforts were also made to purposefully select as diverse a group as possible within the parameters of the study. This was primarily to explore potential ‘extra-discursive’ factors in patient experiences and perspectives on PMS, such as certain embodied and material characteristics previously identified in the literature. For example, age (Dennerstein et al., 2011b; Quintana-Zinn et al., 2017), sexuality and/ or marital status (Cortese & Brown, 1989; Dennerstein et al., 2010; J. M. Ussher & Perz, 2013a), socio-economic status (de Carvalho et al., 2018; Schmelzer et al., 2015), educational attainment (Dennerstein et al., 2011b; Marván & Escobedo, 1999), cultural background (J. Chrisler, 2012; J. M. Ussher et al., 2012, 2017), and parity (Dennerstein et al., 2010; Johnson, 1987; Warner & Bancroft, 1990).

Taking a purposive approach avoided some of the limitations of non-probabilistic sampling purely by *convenience*, which would have likely resulted in the recruitment of predominantly young, white, single, middle-class, heterosexual patients and far fewer, mainly male, UK-based PMS specialist gynaecologists through my existing academic and professional networks. For example, given the known influence of various embodied, material, and institutional factors in experiences of and attitudes towards PMS, more homogenous participant groups would risk producing more limited, and even potentially inaccurate, findings (Emmel, 2014). Purposive sampling was thus important for two reasons: to enable a more nuanced qualitative analysis of participant discourses beyond the main analytical group affiliations (expert/patient, and psychiatrist/ gynaecologist/ feminist academic), and to include a more diverse range of patient perspectives and experiences than have typically been canvassed in qualitative PMS research to date, i.e., young, white and middle-class women attending PMS clinics (Hawkey et al., 2017).

In the case of the clinical PMS experts (‘the clinicians’), the names of all those involved in the five ISPMD (International Society for Premenstrual Disorders) diagnostic consensus meetings held between 2010 and 2016, were collated. This group was responsible for the current formal biomedical definitions of PMS and PMDD. All non-specialist attendees were then excluded, including generalist epidemiologists, statisticians, administrators, and facilitators. This left a long list of 21 potential study participants. The contact details for three of these ISPMD members could not be traced. The remaining 18 individuals were initially approached by email (n=17) or telephone (n=1). Five people did not respond to either the initial or reminder emails, two declined to participate since they had retired, one because they were no longer working in this field, and another due to time constraints. In total, nine members of the ISPMD (half of those contacted) agreed to participate in the study.

To boost the number of experts involved, and to include critical (feminist) academic as well as clinical PMS experts, I also emailed an outline of my study to three relevant professional networks; the UK National Association for PMS (NAPS), the Society for Menstrual Cycle Research (SMCR), and the International Association of Premenstrual Disorders (IAPMD). As a result, seven additional PMS experts (four clinical and three academic) were recruited to the study, taking the total to 16. I considered this number to be satisfactory given the qualitative nature of this part of the study, and the very small number of PMS experts currently publishing on this topic; approximately 35-40 biomedical specialists and seven critical PMS experts, globally. The entire expert recruitment period lasted four months, from October 2019 to January 2020.

Ten patients were purposively recruited through existing personal and professional networks within London. Three of the patients selected were known to me prior to recruitment and were approached for their specific health, cultural, or professional experiences, which were thought to (and did) provide uniquely valuable insights into lay experiences of, and attitudes towards, PMS. Seven patients were recruited in person, after I had presented a talk or participated in a local community or patient group meeting (in my role as founder of Menstrual Matters). I simply described my doctoral research topic to the groups and asked if anyone wanted to take part. The only strict inclusion criterion was that they must identify as someone who ‘experiences PMS’. All those who volunteered were recruited to the study.

Since I was keen to recruit as diverse a group of patients as possible, I ensured early engagement with Black, Muslim, Orthodox Jewish, and other marginalised community groups. This was a relatively straightforward process because I live in a diverse part of London and had existing professional and personal links to various local groups. A further two North American patients were recruited through the above engagement with international professional networks, after they presented as willing to participate in the study but lacking ‘expert’ knowledge of PMS. Data analysis was carried out alongside the data collection process, enabling the cessation of on-going patient recruitment once an indication of data saturation had occurred (in this case when no new discourse codes had been generated in six consecutive transcript analyses). The patient recruitment period lasted five months, from November 2019 to March 2020.

Despite having access to some highly marginalised communities, it was only university-educated ‘patients’ who volunteered to participate. Without further investigation (beyond the scope of this study), this could indicate any of several phenomena previously identified in the literature; that knowledge of, attribution of experiences to, and help-seeking behaviour for, PMS may vary with socio-economic status and/ or educational attainment (Dennerstein et al., 2011b; Petta et al., 2010), that university-educated individuals may be more able to attend health-related community meetings or talks, or to volunteer in general (Donahue et al., 2020; Parboteeah et al., 2004), that associated stigma may prevent some people from wanting to publicly discuss their (pre)menstrual experiences (Johnston-Robledo & Chrisler, 2013; J. M. Ussher et al., 2017), or that my relatively privileged embodied status as a six-foot tall, white, middle-class, university-affiliated menstrual health researcher and relative ‘expert’ (especially after giving a talk) dissuaded some qualifying, and perhaps even interested, individuals from coming forward (Sharma et al., 2009).

### Ethical considerations and data management

This study received ethical approval from the King’s College London research ethics panel; reference number MRSP-19/20-14792 [Appendix A]. Although PMS is a highly stigmatised social label (Johnston-Robledo & Chrisler, 2013), the study was categorised as ‘minimal risk’ because the participants were non-vulnerable adults and the main focus of the interviews was to elicit definitions of PMS, rather than potentially distressing accounts of personal experiences. To alleviate any concerns, participants were reminded before and during the interview that they could withdraw from the study at any point up until April 2020 (writing up of findings), refrain from answering questions without explanation, and that the data would be made anonymous.

The only other ethical risk identified related to the anonymity of the expert participants. As there are so few PMS experts in the world, it is entirely possible that certain demographic characteristics or experiences could be identified as belonging to a particular individual. To mitigate this risk, pseudonyms were used for all participants from the transcription process onwards, the demographic data collected were anonymised and compiled by group rather than individual characteristics, and any potentially identifiable data were removed from the interview transcripts before analysis.

A comprehensive information sheet detailing the research purpose, interview protocol, data protection plan, anonymization process and researcher contact details was shared with participants before interviews took place [Appendix B]. Informed consent was obtained from all participants, either signed by hand or electronic signature (n=19), or through explicit email confirmation (n=9) [Appendix C]. One expert did not consent to two specific elements; to be contacted regarding future related studies (element six), or for the use of their anonymised data in future research or publications (element seven). As a result, their raw interview data will not be publicly shared. All other participants consented to all eight elements listed.

Before their interviews began, the patients were reminded about the £10 gift voucher they would be given to compensate for their time (up to an hour), regardless of their answers or on-going participation. This is in line with research that suggests it is only fair to pay lay participants for their time, so long as the payment is not too much, too little, or provided in a way that might coerce consent or participation, or overly influence responses (Head, 2009).

Only one of the experts asked to see and edit their interview transcript, despite this being listed as an option on the ‘expert’ information sheet [point seven in Appendix B]. They were concerned about an anecdote involving a former colleague, which could be construed as inappropriate information, or as potentially identifiable data. The anecdote was the only significant section of text deleted from the transcript and its omission made very little difference to the overall analysis.

Another expert was asked to clarify a few partially transcribed responses, caused by internet connection issues during the online interview. They subsequently made dozens of editorial changes to the full transcript and changed the meaning of a couple of responses mentioning disagreements between psychiatry and gynaecology, to imply unproblematic professional consensus. Since informed consent had been provided in advance of these changes, and several of the other experts had also mentioned these disputes, I decided to only accept the text changes that did not substantially change the meaning of the original oral interview response. This experience provided useful insight into some of the methodological, ethical, and research credibility issues that can arise when sharing interview transcripts with participants for the purpose of validation or research transparency (Mero-Jaffe, 2011).

The short demographic survey sent to participants after interview contained some questions relating to what may be considered personal or private information (income, marital status, sexuality, religious affiliation, ethnicity, and parity) [Appendix D]. Their inclusion was based on previously identified correlations between certain embodied ‘extra-discursive’ factors and PMS (as outlined above). To ensure informed consent and in line with ethical good practice (Sieber et al., 2009), the survey data are shared anonymously either in generalised (group) form, or in relation to one specific characteristic at a time, only. Participants were also given the option ‘prefer not to say’ in response to all non-mandatory questions.

All but one of the interviews were audio recorded but once transcribed, the audio files were deleted, in accordance with the data management plan [Appendix E]. All potentially identifiable data were removed from the interview transcripts before analysis. The transcripts were labelled with pseudonyms and stored securely in three separate digital and two physical locations [Appendix D]. Any non-sharable data, such as participant contact information and the pseudonym list, will be deleted two years after thesis submission [Appendix E]. This data management plan is compliant with the 2018 GDPR (General Data Protection Regulations) and ESRC studentship funding conditions.

### Participant characteristics

A total of 28 participants were interviewed across the two main participant groups. This sample size is in line with what is considered ‘normal’ within most qualitative research methodologies (Baker & Edwards, 2012; Creswell, 2013; Saunders et al., 2018). A short (5 minute) online participant background survey was sent to all participants after interview [Appendix D]. Its purpose and rationale for asking personal questions was explained during the interviews and included in the survey information [Appendix D]. The same survey was used for both the expert and patient groups, with one additional question later emailed to the experts regarding their direct involvement in the creation of formal definitions of PMS/ PMDD in either the DSM, or ICD-11. The anonymised and generalised participant group data is shown in Table 3.3.

Table . Participant characteristics and information

Shape

Description automatically generated with medium confidence

Considering the small number of participants, diversity within the groups was very good across several embodied demographic characteristics. Some notable exceptions were; the sex of the expert participants (unfortunately, eight of the nine ISPMD members who declined or failed to respond to the email invitation were male, leading to twice as many female than male ISPMD participants, even though the society has slightly more male than female expert members); the age and employment status of the experts (predominantly close to or beyond retirement age); the ethnicity of the experts (which, while somewhat representative of the ISPMD membership, was entirely White); and the high educational attainment level of the patients (as previously mentioned, all were university educated, or to a similar professional level, with most holding a postgraduate qualification). Some unavoidable participant homogeneity related to the sex and limited age range of the patients, which reflected the embodied yet temporary phenomenon of the female reproductive cycle.

### Participant interviews

Separate expert and patient participant group interview schedules were developed, based on their anticipated knowledge and experience of PMS [Appendix F]. Open questions were used to reduce interviewer-influence on participant discourses, and to encourage more extensive answers. The expert schedule included 26 questions and the patient schedule, 31 questions. Around half of the questions (n=15) were the same across both schedules to enable more straightforward comparison between participant groups. Over time, a couple of adjustments were made to each of the schedules, largely in response to observed participant difficulties in understanding the meaning of three of the questions (E20, E25 and P26) and in one instance, the addition of a question to the patient schedule after the first patient interview (P18) *“Do you think doctors have enough knowledge/ training on menstrual cycle-related symptoms?”*. Only one of these questions affected the participant group comparative analysis. The patients’ extremely limited knowledge of Premenstrual Exacerbation (PME) meant that their responses to the question (E14/ P26) “*Many chronic health conditions get worse at certain times in the menstrual cycle. Would you count the expression of these as premenstrual symptoms? Why?”* could not be meaningfully compared with those of the expert group.

The interviews were undertaken according to participant preference and geographical limitations, either in person (n=8), online video (n=19), by telephone (using Skype) (n=1), or email (n=1). The differences observed between the email interview and the others are discussed in more detail in section 3.4. The location of the face-to-face interviews were determined by the participant; their home (n=3), a university room (n=2), a religious building (n=1), an art studio (n=1), and a medical consulting room (n=1). All but the email interview were audio recorded, using two separate smartphone devices when conducted in person, or online audio recording software plus one smartphone device when conducted virtually/ by telephone.

The length of the expert interviews ranged from 21 to 83 minutes, with an average duration of 52 minutes. The interviews with patients ranged from 23 to 81 minutes, with a slightly longer average duration of 56 minutes. Two of the patients (Faith and Gemma) were interviewed at the same time (for their convenience) and one patient (Ria) was interviewed over two separate calls due to time constraints. Three of the experts (Fran, Celia and Laura) were asked a reduced selection of questions due to time constraints (only having half an hour with which to speak to me). Brief written notes were made during the interviews to provide prompts for follow-up questions, but they were not included for analysis.

## 3.3 Data analysis

### Transcription

The audio recordings were transcribed as soon as possible after each interview, using an automated transcription software called Sonix AI. The quality of the verbatim transcription varied between 80% and 90% accuracy, mainly depending upon the accent of participant (UK and European accents were less accurately transcribed than North American or Australian accents). I then made manual corrections and notations to the transcription text using the editing feature of the same software during repeated listening to the audio recordings (approx. four hours per interview on average). The software automatically identified and annotated the length of pauses and changes in speaker. Further annotation adapted from the transcription conventions of interactional discourse analysis was added manually [Appendix G]. For example, utterances such as ‘um’ or ‘er’, sighs, audible intakes or exhalations of breath, stuttering, extreme volume changes, emphasis on certain words, and the reported speech of others, were annotated using an adapted glossary of symbols, based on those first described by Jefferson (2004) [Appendix G]. The interview transcripts were then anonymised (as described above) before being uploaded to the qualitative data analysis software (NVIVO) for further annotation and coding.

### CRDA Step 1: Discursive Psychology (DP) annotation and analysis

In line with the CRDA methodology outlined by Sims-Schouten and Riley, a ‘*three -level synthesised discourse analysis’* was applied to the interview data (2014, p. 57). First, a preliminary Discursive Psychology (DP) annotation and analysis of the data was carried out before applying any thematic coding. This approach allowed me to “*examine the rhetorical strategies participants employ that allow them to do certain interactional work*” (Sims-Schouten & Riley, 2014, p. 57). In short, DP approaches are more interested in what the talk is ‘doing’ than what the person is ‘saying’. There is no strict step-by-step methodology for DP, since it is a highly iterative rather than linear process, but I did follow the basic analytical guidelines provided by Wiggins (2017), depicted below [Fig. 3.1].

Diagram

Description automatically generated

Figure . The Analytical Stages of Discursive Psychology (Wiggins, 2017, p. 115)

I began by annotating each interview transcript according to the ‘discursive devices’ employed. All 20 discursive devices listed by Wiggins (2017) were annotated [see Appendix H for the full list]. Seven discursive devices were found to be of most analytical value in this study; pronoun shifts (e.g., from ‘I’ to ‘we’); silences, pauses and hesitations; hedging talk (e.g., repeated or stuttering speech and/or utterances of ‘uh’ or ‘er’); minimisation (e.g., language such as ‘a bit’, ‘only’, or ‘just’); affect displays (e.g., laughter, sighs, or audible inhalation/ exhalation); metaphor usage; and category entitlements (e.g., ‘women’, ‘psychiatrists’ or ‘gynaecologists’).

For example, ten of the sixteen expert participants displayed a distinct ‘stuttering’ speech pattern, typically when becoming aware of a possible contradiction in what they had just said. The following excerpt demonstrates the way in which this type of discursive ‘disfluency’ can sometimes signal a speaker trying to rapidly ‘hedge’ their words to absolve themselves of personal responsibility for the perceived inconsistency (Fairclough, 2014). Also note the associated use of pronoun shifts between ‘I’ and ‘they’ (lines 1 to 3) to distance Andrew from the current DSM definition of PMDD [Excerpt 3.1].

Excerpt 3.1

**Andrew:** (E22)[...] I regard this as a kind of compromise […] but still they do include them [typically mild physical symptoms in the diagnostic criteria for PMDD] (.) not to annoy the gynaecologists too much (.) I assume (.) and I (.) if I (.) if I were king, I would (.) I would have not regarded them as part of this at all […] and I think so (.) so (.) so (.) er (.) I (.) I (.) I believe that it's a (.) I think we should regard item number eleven in the DSM as a compromise (.) and it's perhaps not that (.) er (.) there’s no robust scientific basis for that

### CRDA Step 2: (Rhetorical) Discourse coding and analysis (DA)

Signs of discursive difficulty such as these multiple or extended pauses, pronoun shifts, repetition, and/ or hedging talk, became highly *visible* during the process of DP annotation, which provided a valuable starting point for the second phase of discourse analysis, as recommended by Jorgensen and Phillips (2002, p. 118). Thus, after a transcript had been fully annotated according to the conventions of DP, I went back to the beginning and read it again to provide an overall impression of the rhetorical content, paying particular attention to areas of possible discursive crisis.

The purpose of this second analytical approach was to identify the different discourses participants used to describe their experiences of, or epistemological/ ontological claims about, PMS. As suggested by Sims-Schouten and Riley (2014, p. 57), these included ‘common-sense discourses’ (e.g., the crazy or irrationally angry female ‘PMS stereotype’, or menstruation as ‘dirty’ or ‘pathological’) as well as more ‘institutionally-oriented discourses’ (e.g., feminist discourses of PMS as socially constructed, or biomedical discourses of PMS as a ‘hormonal’ phenomenon). The idea is that the ‘discursive repertoire’ of any given participant reflects their particular construction of ‘what PMS is’, as well as providing insight into the “*possibilities and limitations of using these discourses in relation to what… [they can] say, think, and do*” (Sims-Schouten & Riley, 2014, p. 57).

Inductive open coding was applied to ‘chunks of text’ (rather than line by line, as can be the case for thematic coding) (Willig, 2015), creating or attributing existing codes to each discrete discourse encountered in the data. Such coding is in line with Grounded Theory Methods, which derive theory from data in a systematic, yet interpretive way, as originally outlined by Glaser and Strauss (Glaser & Strauss, 1967). The reason for using an exclusively inductive approach to the rhetorical coding was to minimise the influence of any preconceptions I had about discursive themes I ‘expected’ to find in the data.

The discourse analysis was initially carried out by hand on printed transcripts (using written notes, colour coding and a code book), before being organised into a hierarchical coding frame and transferred to a digital format using NVIVO coding software. Initially, over 120 discourse codes were created, but during the digitisation process, these were refined and combined into 75 codes [Appendix I]. Over five months of iterative and overlapping data collection and analysis, these codes were further refined into 42 *axial* codes, defined as ‘aggregates of the most closely interrelated (or overlapping) open codes for which supporting evidence is strong’ (Corbin & Strauss, 2012, p. 109) [Appendix I]. In turn, these axial codes were then grouped into the four major discursive themes described in detail over the following four chapters. Each theme was composed of roughly the same number of axial codes and associated excerpts of text.

Conducting iterative and combined DP and DA methods on each transcript, as well as across the gradually increasing dataset, provided unique and extremely valuable insights into the data. This approach enabled a more nuanced and empirically robust analysis of what, how, and for what purpose, discourses were being constructed in the text than would have been possible through DA, alone. For example, by annotating all potential ‘metaphors’ (as part of DP), most of the participants were found to use of the concept of a ‘normal curve’ or ‘spectrum of experiences’ when describing PMS [Excerpt 3.2].

Excerpt 3.2

**Ria (patient):** (P13) Um (.) like looking at the normal curve and distribution (.) a lot of PMS quote unquote (.) [depicts quotation marks with hands] 'symptoms' (.) are just things (.) ha! [laugh] (.) that happen to us during that time (.) which are like beautiful and magical (pause) and also a huge struggle?

It was only by reflecting on the usage of the term ‘normal’ as a metaphor within specific discursive contexts that one of the four main discursive themes (outlined in chapter five) evolved from being ‘differentiating typically mild (i.e., ‘normal’) from problematic premenstrual change is difficult’ (DA alone) to ‘the use of the normal curve analogy enables debilitating cyclical symptoms (indicative of underlying health issues) to be positioned as simply a ‘more severe’ version of typically mild premenstrual changes and vice versa’ (DP plus DA). In short, it was possible to move beyond the identification of a recurring rhetorical trope in the data and gain valuable insight into *why* it may exist and *how* it is constructed through a particular discursive mechanism.

### CRDA Step 3: Extra-discursive Analysis

The third level of ‘Critical Realist’ Discourse Analysis (CRDA), involves an analysis of the ways in which embodied, material or institutional factors may influence the participants’ discursive repertoires (Sims-Schouten & Riley, 2014). The idea being that ‘extra-discursive’ contextual information (e.g., the sex, cultural background, socio-economic status, or professional training) of an individual may “*provide the conditions of possibility that allows their sense-making to make sense”* (Sims-Schouten & Riley, 2014, p. 58). While the main focus was on comparing expert with patient descriptions of PMS, these broad participant group categories were further disaggregated by various demographic characteristics established through the ‘participant background survey’ [Appendix D]; clinical/ academic discipline, sex, age, ethnicity, sexual orientation, religious affiliation, nationality, marital status, parity/ current childcare responsibilities, educational attainment, and socio-economic status (based on household income).

In practical terms, the interview data were split by question and participant group (experts and patients) in an Excel workbook. The discourses utilised by each participant were then compared within and across the participant groups. Around half of the participant questions allowed for direct comparison between the expert and patient groups, the other half provided group-specific data. Discursive similarities and differences in the data were then identified and analysed with reference to the above extra-discursive characteristics. This was to assess the level of discursive agreement within and between the participants across multiple intersecting embodied, material, and institutional factors.

For example, the use of certain metaphors in participant descriptions of premenstrual experiences was found to vary between the expert and patient groups [Table 3.4].

Table . The top 10 metaphors used to describe premenstrual experiences and their relative usage by the expert and patient participant groups

Shape

Description automatically generated with low confidence

In some instances, if an embodied/ material/ institutional factor was found to be of analytic relevance for a particular discourse, further comparative analysis was conducted using the NVIVO software, especially regarding the frequency of specific terms or discursive device usage. For instance, when one of the patient-specific metaphors (dirty/ toxic) was further analysed, it was found to be used *only* by patients with some sort of religious affiliation (discussed in more detail in chapter seven). While qualitative research methods prohibit the generalisation of claims made about populations beyond the participants directly involved in the study, the identification of a potential relationship between a specific metaphor’s usage and an extra-discursive factor such as an individual’s religious affiliation, could provide useful insight into its possible discursive *origins* and act as a prompt for future research. Not least because explicit descriptions of menstruation as an ‘unclean’ and spiritually, morally, and physically ‘impure’ phenomenon are found within all the major religion’s sacred texts (Cohen, 2020; Tan et al., 2017).

### Data Triangulation

Finally, wherever possible, the epistemological and ontological claims made in the interview data about ‘what PMS is’, what causes it, and how prevalent it is within the population, were compared with robust (quantitative) empirical data of relevance to these questions. As described above, the four epidemiological data sources used for triangulation were selected based on being the best quality data available and accessible regarding each claim. Participant discourses with high levels of discursive agreement within, but differences between, specific demographic groups were then ‘ranked’ according to how closely they described the patterns identified in quantitative epidemiological data.

For example, the most common PMS symptoms described by the participants were compared to those generated from a large-scale cross-cultural epidemiological study (Dennerstein et al., 2011), to assess the relative de/prioritisation of certain symptoms below/over others by the expert and patient participant groups (described in detail in chapter five). Similarly, the common claim that PMDD (Premenstrual Dysphoric Disorder) is akin to ‘severe PMS’ was compared with a systematic review of 47 prospective studies, which found no conclusive evidence of a specific ‘premenstrual’ negative mood change in the general menstruating population, thus contradicting this popular ‘common-sense’ discourse (described in more detail in chapter four).

## 3.4 Some methodological reflections

### Conducting interviews for CRDA

To ensure that the interview data would cover the main topics of interest identified during the literature review, schedules (lists of questions) were developed for the expert and patient participant groups [Appendix F]. Care was taken to phrase the questions in an open and non-leading manner to elicit responses that would be as free flowing as possible and only minimally influenced by my own perspective. I also set out to apply what is described by Smith and Elger, as “*a CR approach to the design, conduct and analysis*” of interviews, which relies on the consciously *active* role of the interviewer (2014, p. 109). Their work builds on that of (Holstein & Gubrium, 1997, p. 125), who proposed that the active interviewer, "*sets the parameters for responses, constraining as well as provoking answers that are germane to the researchers’ interest*”. These combined priorities (adequate topic coverage, non-leading phrasing, and provocative questioning) gave rise to a relatively long list of probing questions for both groups (26 for the experts and 31 for the patients), given that the average length of interview was around an hour.

As a result, the interviews were much more structured and directive (in terms of provoking participants to make, and reflect upon, epistemological claims) than is typical for a ‘semi-structured’ approach. To alleviate potential participant concern that they were being ‘tested’ in some way (especially the patient group), I explained that I was mainly interested in the way in which they described their experiences and personal perspectives, at the start of each interview. A couple of the experts (both extremely experienced qualitative interviewers) mentioned the unusually full interview schedule and provocative question format when providing feedback at the end. One implied that it had been an unusual (lines 7, 9, & 11) and possibly even mildly uncomfortable experience (line 4), based on their use of the words ‘interesting’, ‘different’ and ‘interrogation’ [Excerpt 3.3].

Excerpt 3.3

**Interviewer:** (E26)And that's it (.) so the last one is how do you feel about the interview (.) do you have any questions or comments you'd like to add at this point

**Zoe:** Um (.) no (.) no (.) no questions or comments on the interview itself (.) no (.) I did (.) it did feel a little bit like an interrogation [hahah! laugh]?

**Interviewer:** Oh [with intake of breath] (.) I'm sorry (.) I was trying to keep to time (.) but actually (.) we were fine in the end [heh! Short laugh]

**Zoe:** Yeah (.) it's just (.) it's interesting […] you were asking me more about my knowledge and my opinions rather than my experiences (.) and so I just found it quite interesting that (.) I'm used to doing more of the experiential interviewing than this (.) um (.) opinion-based interviewing (.) so (.) but that's not a (.) that's not a comment about the interview (.) it's just that it was a different form of interview for me

The relatively directive and structured question/ answer format also highlighted the way in which the interview situation produced quite different patterns in speech than those typically observed in ‘natural talk’, the usual focus of DP analysis. For example, overlapping speech very rarely occurred in the interview data (except during the joint patient interview with Faith and Gemma). Similarly, ‘turn taking’ was not of much analytic value within this study, since this was largely dictated by the question/ answer format and associated interviewer/ participant roles, rather than ‘natural’ conversational interactions. This is not to say that the application of DP to the interview data was inappropriate, just that some of the core ‘interactional’ elements of the method were not applicable in this case. Indeed, despite the interactional limitations outlined above, most of the resulting data were of high quality and relevance to the research questions and methods employed. Moreover, a less structured interview format might not have allowed the relatively straightforward data comparison between the expert and patient participant groups, which was of substantial value in this study.

Another methodological insight into the importance of interview approach was provided by one of the expert participants. A clinical psychologist (Barbara) wished to participate in the study but preferred to respond to my questions by email, since she was based overseas and lacked confidence in using online communications software. The striking contrast between the written responses and those elicited from the other participants, provided useful insight into the differences between email versus oral interview approaches. For example, the written responses were extremely short, strictly in line with formal biomedical definitions and guidelines (rather than personal reflections on these) and lacking in discursive devices such as metaphor or pauses/ hesitations [Fig. 3.2]. This substantially limited the analytical value of the data. Interestingly, some signifiers of affect did manage to be communicated through written text, most notably Barbara’s use of capitalisation to signal ‘emphasis’ or ‘strength of feeling’ in respect to certain epistemological claims [Fig. 3.2]. This is the same notation used to represent ‘volume’ in DP audio transcription conventions [Appendix G]. Barbara’s comparative example helpfully justified the use of oral interviews by indicating what might have been lost if alternative research methods had been employed (such as email interviews or online surveys).

A screenshot of a computer

Description automatically generated with medium confidence

Figure . The use of capitalisation to depict strength of feeling in Barbara’s written responses

### Adapting CRDA to doctoral level research

I really enjoyed using the CRDA methodology outlined by Sims-Schouten et al. (2007; 2014; 2019). It was the first time I have ever conducted a discourse analysis of any kind, despite over a decade of applied social research experience within the context of international development and human rights policy and intervention evaluation. The multiple, iterative, and interwoven levels and types of analyses involved in the method resulted in a more rigorous and systematic analytical process than I have ever previously applied to qualitative data. Such an approach yielded additional benefits, especially regarding the identification of *how* and *why* participants may use certain discourses due to extra-discursive (as well as discursive) constraints. I also appreciated the way in which the methodology actively encouraged *retroductive* thinking (moving between knowledge of one thing to another) to reveal possible discursive *mechanisms* underpinning the participant discourses (Danermark et al., 2019).

Indeed, taking an explicitly Critical Realist philosophical position allowed me to compare, and even integrate, allegedly ‘oppositional’ or ‘incompatible’ academic and biomedical expert PMS discourses. Its emphasis on the contingent, interpretive, and transitive nature of all human knowledge, also resolved some of the methodological/ epistemological/ ontological issues that can arise when using mixed methods. For example, CR posits that the participant discourses (qualitative data) and even the robust epidemiological (quantitative) data used to triangulate them, are mediated by societal discourses, and subject to embodied, material, and institutional constraints, but that it is still possible to meaningfully compare these different types of data (despite their ontological differences) to get ‘empirical feedback’ on the *relative* explanatory power of the different discourses (Danermark et al., 2019). In this way, it was possible to rank participant discourses according to how well they could account for the findings of robust epidemiological data, without implying that the quantitative data are an intransient or objective representation of ‘what PMS really is’.

The full CRDA methodology, however, had to be adapted to fit within the available timeframe and scope of doctoral research. For example, I was unable to undertake more substantial participant background research, which might have provided additional potentially relevant contextual data previously identified in the literature, and thus, provided even more nuance to the analyses. For example, regarding patient body image or self-objectifying psychological profiles (Jappe & Gardner, 2009; Kleinstäuber et al., 2016; Muljat et al., 2007; J. M. Ussher & Perz, 2020b), quality of personal relationships (Coughlin, 1990; Kuczmierczyk et al., 1992; J. M. Ussher & Perz, 2013b), level of menstruation/ PMS knowledge (Marván & Cortés-Iniestra, 2001; Marván & Escobedo, 1999), or adherence to gender norms (Brooks et al., 1977; J. M. Ussher, 2004).

In fact, due to the already large number of potentially relevant extra-discursive contextual factors included in the analysis (age, sex, ethnicity, marital status, parity, sexual orientation, socio-economic status, religious affiliation, educational attainment level, and clinical/ academic discipline) and the necessarily small number of participants involved in in-depth qualitative research, only very few of these factors were found to be of analytic value regarding the participant discourses. Of course, this is not to say that they did not have an influence, just that due to the relatively high number of intersecting factors, it was difficult to identify correlations between them and specific discourses that might only be used by three or four individuals. It could even be argued that the demographic diversity of the participants effectively limited the number of observable correlations between demographic characteristics and discursive content/ devices used in such a small study population.

It was, therefore, remarkable to find *any* demographic patterns in the data, which could indicate that either these associations are relatively important, or simply the most frequently expressed, extra-discursive influences upon lay and expert descriptions of PMS in this study. Beyond the two main factors found to have a consistent influence upon participant discourses (patient/ expert status and expert academic/ clinical discipline), discursive associations were only made in relation to expert participant sex, and patient ethnicity and religious affiliation. Male sex was found to be a potential factor in defining ‘pure’ or ‘core’ PMS as predominantly mood-based (described in more detail in chapter six), racialised ethnic minority patients appeared to be subject to greater degrees of familial/ societal/ clinical disbelief in their experiences (described in more detail in chapters four and five), and only patients with a religious affiliation metaphorically positioned (pre)menstrual experiences as ‘dirty’ (described in more detail in chapter seven).

The above ‘mathematical’ issues regarding the identification of extra-discursive factor correlations in the data were not only a limitation, however, since they did at least reduce some of the analytical burden involved in the CRDA methodology. While 28 hour-long interviews may be ‘average’ for more typical discourse analysis approaches, the multiple interacting levels of analyses of CRDA resulted in a complex and long analytic process. In fact, the data analysis phase took around eight months to complete, although, the first four months did overlap with the collection phase. It was fairly time-consuming to annotate and code all of the Jefferson transcription conventions commonly used in DP (Jefferson, 2004). Early in the data analysis process, I decided to omit four of the most labour-intensive elements relating to speed of speech and intonation [Appendix G]. Similarly, discursive devices that were infrequently used or not obviously associated with the main rhetorical themes or participant groups were coded but not comprehensively analysed, resulting in the prioritisation of seven out of the twenty listed by Wiggins (2017) [Appendix H].

While these methodological compromises undoubtedly reduced the analytic potential of the data to some extent, I do not think that they substantially affected the main findings. Overall, I feel that the adapted CRDA methodology was an effective approach with which to address the research questions. It enabled an insightful and grounded analysis that I feel accurately reflects what the participants said, as well as some of the underlying discursive and extra-discursive factors in why and how they did so (as demonstrated in the following four empirical chapters). I would caution against undertaking CRDA lightly, however, especially at the doctoral level, because there are very few ways to reduce the time and effort required to apply these multiple analytical methods in an iterative manner. Topics with fewer potential demographic/ extra-discursive factors identified in the literature or those without such a strong comparative element between participant groups, may permit more manageable timescales.

1. In contrast to the substantial documentation of enabling and oppositional processes and discourses surrounding the original inclusion of PMDD in the DSM, as analysed by Figert (1996). [↑](#footnote-ref-1)