Self-report Pain and Symptom Measures for Primary Dysmenorrhea: A Critical Review

Chen C.X.*, Kwekkeboom K. L.*, Ward S. E.*

*University of Wisconsin-Madison School of Nursing, Madison, Wisconsin, USA

Running head: Review of Primary Dysmenorrhea Symptom Measures

Corresponding author:

Chen Xiao Chen,

University of Wisconsin-Madison School of Nursing,

K6/117 Clinical Science Center, 600 Highland Ave., Madison, WI 53792-2455, USA.

Phone: (608) 265-8569.

Fax: (608) 263-5458.

Email: xchen243@wisc.edu.

Submission category: Review

Disclosures: The authors have no conflicts of interest to disclose. No funding sources were involved in this study.

Database?

Databases including PubMed, PsychoINFO, Cumulative Index of Nursing and Allied Health Literature, and Health and Psychosocial Instruments were searched for self-report symptom measures that had been used among women with either primary dysmenorrhea or perimenstrual symptoms. A total of 15 measures met inclusion criteria and were included in the final analysis.

What does this review add?

This article addresses a void in the literature by critically reviewing content and psychometric properties of self-report pain and symptom measures for the common, but understudied problem of primary dysmenorrhea. The findings may direct further development and validation of a comprehensive symptom measure for descriptive and intervention research on primary dysmenorrhea.

This is the author's manuscript of the article published in final edited form as:
Chen, C. X., Kwekkeboom, K. L., & Ward, S. E. (2015). Self-report pain and symptom measures for primary dysmenorrhoea: A critical review. European Journal of Pain, 19(3), 377–391. https://doi.org/10.1002/ejp.556

Introduction

Primary dysmenorrhea (PD) is a prevalent and debilitating condition among women of reproductive age. Over 25% of women and up to 90% of adolescent girls experience PD; 15-20% report PD pain as severe or distressing (Andersch & Milsom, 1982; Banikarim *et al.*, 2000; Davis & Westhoff, 2001; Durain, 2004; Burnett *et al.*, 2005). In the US, PD is the leading cause of lost work hours in women, 10-30% of women miss work because of PD, which translates to a loss of 600 million working hours or up to 2 billion dollars annually (Dawood, 1988). In other countries, rates of absenteeism from work or school range from 10% to 17% (Svanberg & Ulmsten, 1981; Burnett *et al.*, 2005). Women with PD also experience lower physical activity (Chantler *et al.*, 2009), diminished work productivity (Ju *et al.*, 2014), and reduced quality of life (Unsal *et al.*, 2010; Iacovides *et al.*, 2013; Nur Azurah *et al.*, 2013).

Some far-reaching consequences of PD are less well-known. Studies show a high co-occurrence of PD and other chronic pain conditions, such as irritable bowel syndromes (IBS) (Altman *et al.*, 2006), migraine (Mannix, 2008), and fibromyalgia (Yunus *et al.*, 1989; Shaver *et al.*, 2006). PD can exacerbate symptoms of other pain conditions with increased sensitivity to pain both in and outside of the uterine referral area (Banikarim *et al.*, 2000; Vincent *et al.*, 2011; Tu *et al.*, 2013). For example, women with IBS or urinary calculosis experience more gastrointestinal or urinary symptoms if they have co-existing PD (Altman *et al.*, 2006). Scholars suggest that moderate-to-severe menstrual pain could be “a harbinger of more pain to come later in life" (p1940) (Vincent *et al.*, 2011). Women with PD are twice as likely to develop IBS as women without PD (Olafsdottir *et al.*, 2012). Reports show augmented sensitivity to pain and significant brain changes in women with PD, posing risk factors for future pain (Tu *et al.*, 2009, 2010; Vincent *et al.*, 2011).

Despite its prevalence and serious consequences, PD is understudied in the pain community, with only 0.1% of papers on pain dealing with PD (Berkley & McAllister, 2011). The belief that PD is a “normal” condition and that menstrual symptoms are “taboo”, could explain this lack of attention (Giamberardino, 2008; Berkley & McAllister, 2011). Historically, stereotyping dysmenorrhea as a psychogenic problem might contribute to the lack of research (Dawood, 1988).

Symptom assessment is a crucial step toward effective symptom management. To support future descriptive and PD intervention research, it is important to evaluate PD symptom measurement tools. Other reviews have focused on symptom assessment tools for premenstrual syndrome (PMS) (Budeiri *et al.*, 1994; Haywood *et al.*, 2002) which is distinct from PD (Booton & Seideman, 1989) (See Table 1 for differences between PD and PMS). The purpose of this paper is to describe and critically review the self-report tools that have been used to measure pain and other symptoms of PD. This review focuses on measurement for research rather than for clinical assessment of PD.

Methods

There are two types of dysmenorrhea: primary dysmenorrhea (PD), characterized by abdominal or pelvic pain occurring just before or during menstruation in the absence of pelvic pathological findings, and secondary dysmenorrhea, defined as painful menstruation in the presence of pelvic pathological findings (e.g., endometriosis, uterine fibroids) (International Association for the Study of Pain, 1994; Durain, 2004). This review focuses only on PD, and excludes tools designed for measuring symptoms of a specific condition related to secondary dysmenorrhea (e.g., endometriosis or uterine fibroids).

For this review, PD was conceptualized as a symptom-complex secondary to increased production of prostaglandins (particularly prostaglandin F2α) and other inflammatory mediators. This conceptualization is supported by 1) substantial evidence demonstrating a higher level of menstrual prostaglandins in women with PD compared to women without PD (Chan *et al.*, 1981; Dawood, 1981), 2) the great similarity between the symptoms of PD and the adverse effects observed in prostaglandin administration (Dawood, 2006), and 3) the effectiveness of prostaglandin inhibitors in decreasing menstrual prostaglandins and consequently relieving PD symptoms (Chan *et al.*, 1981; Dawood, 1984, 1988; Marjoribanks *et al.*, 2010). From the pathogenesis point of view, prostaglandins cause uterus muscle contraction. The contraction, along with the resulting ischemia, hypoxia, and sensitization of nerve endings in the nearby tissues, is responsible for menstrual pain (Lundstrom, 1981; Coco, 1999; Dawood, 2006). In addition to pelvic pain and its referred back and thigh pain, prostaglandins also may contribute to gastrointestinal symptoms, such as nausea, vomiting, bloating, and change in bowel

frequency (Dawood, 1981, 1984; Kinch, 1985; Jarrett *et al.*, 1996). Thus, our review includes tools that measure pain as well as other prostaglandin-related symptoms. Our review excludes measures designed to exclusively measure PMS/PMDD.

Literature Search Strategy

A literature search was performed through January, 2013 using PubMed, Cumulative Index of Nursing and Allied Health Literature (CINAHL), PsychoINFO, and Health and Psychosocial Instruments (HaPI). Search terms included ("menstrual pain" or "menstrual cramp" or "dysmenorrhea" or "menstrual symptoms" or "premenstrual symptoms") and ("questionnaires" or "instrument" or "measure" or "measurement" or "patient report outcome" or "assessment"). The search was not limited by year, but was limited to articles published in English. Relevant publications also were identified by using the "related articles option" on the databases. Any additional measurement tools identified from the reference lists of the retrieved publications were subsequently retrieved and reviewed. Papers were selected for further reading if the abstract contained any information related to measurement of symptoms of PD. A symptom measurement tool was excluded if it did not contain any menstrual pain-related symptom items. Once a measurement tool was identified, the name of the instrument was used as a search term to further identify studies reporting its use and its psychometric properties.

Measurement Tool Evaluation Criteria

Specific information regarding the purpose(s) of the tool, item content and scoring, symptoms measured, timeframe of measurement, reliability, validity, sensitivity (i.e., responsiveness) to change, and the populations under study were abstracted from the original studies and subsequent studies using that measurement tool. Information was summarized by the first author in a table format; the second and third authors checked a selection of articles to assess accuracy of reporting.

An overall summary of the strengths and weaknesses of the measures is provided in Table 2. The review criteria consisted of evaluation of the tools' validity, reliability, sensitivity to change, generalizability, multidimensionality, literacy requirement, and potential for recall bias. The quality of attributes for each tool was rated as poor (+), fair (++), good (+++). In scoring the attributes we

considered both the soundness of methodologies used and the actual resulting psychometric data. For this review, content validity was assessed as a) whether the items in the tool are comprehensive of PD symptoms based on the mechanism of prostaglandin release (i.e., pelvic, back and thigh pain, nausea, vomiting, bloating, change in bowel frequency) and b) whether the items are relevant or specific to PD. For construct validity, we sought evidence of known-groups validity, concurrent validity, convergent/discriminant validity, and/or stable factor structure established through factor analysis (Nunnally, 1994; Polit, 2012). Reliability was coded based on the overall consistency of a measure across time (test-retest reliability) and across items (internal consistency) (Nunnally, 1994; Polit, 2012). Sensitivity to change was determined by examining whether a measurement tool was able to identify changes in symptoms in response to treatment (Jensen, 2003). Multidimensionality was determined based on whether a tool measured dimensions of the pain and symptom experience beyond severity (e.g., pain quality, temporal aspects of symptoms, symptom distress, etc.). Generalizability was assessed based on the number of studies using a specific tool among women with PD and the diversity of the populations across studies in terms of age and country/culture/ethnicity.

Results

The review identified a total of 15 measurement tools in three categories: a) generic pain measures (n=5), b) tools designed specifically to measure PD symptoms (n=5), and c) tools designed to measure menstrual, premenstrual, or perimenstrual symptoms (n=5). Under each category, a description of each measurement tool and its psychometric properties is provided, followed by a critique of the tool.

Category I: Generic Pain Measures

Single-item Numerical Rating Scales (NRSs), Visual Analogue Scales (VASs), Verbal Rating Scales (VRSs), and Facial Pain Scales (FPSs)

Description.

The NRS, VAS, VRS, and FPS are single-item tools designed to measure pain intensity/severity in different pain populations and have been widely used in studies of PD. For the NRS, a person is asked to rate the pain from 0 to 10 (or 20, 100). While 0 represents “no pain”, the number of upper limit

represents the other extreme of pain intensity (e.g., "pain as bad as possible"). A VAS consists of a line, usually 10 cm long. Two ends of the line are labeled as the extremes of pain (e.g., "no pain" to "pain as bad as it could be") (Jensen & Karoly, 2011). A VRS consists of a list of adjectives describing different levels of pain intensity in rank order. The number of descriptors in a VRS ranges from 4 to 15 (Jensen & Karoly, 2011). The FPS measures pain by using illustrations of facial expressions of persons experiencing different levels of pain intensity (Jensen & Karoly, 2011). As generic pain measures, these tools have undergone extensive psychometric testing, and have evidence of validity , reliability, and sensitivity, and are generalizable across different populations with pain (Jensen, 2003; Jensen & Karoly, 2011), including women with menstrual pain (Tugay *et al.*, 2007; Ma *et al.*, 2010).

Critique

One major limitation of these single-item measures is that they are not designed to capture the full range of symptomology of PD. Women with PD may experience pain at more than one site, and they usually also experience other non-pain symptoms (Dawood, 1981, 1984; Kinch, 1985; Jarrett *et al.*, 1996). Another limitation of these single-item measures is the assumption that pain is a unidimensional experience. Only pain intensity is measured, while other dimensions of pain, such as affective components, quality, and temporal aspects, are not considered (Melzack, 1975). In addition, there is potential for recall bias when such scales are used retrospectively (Jensen, 2003). There is some suggestion that a VAS may be more difficult to understand than other single-item pain scales (Larroy, 2002; Jensen, 2003), however, Gagliese and Melzack (1997) reported that young participants (< 45 years old) with chronic arthritis pain (the age range where most women with PD fall) had little problem with the VAS.

The McGill Pain Questionnaire (MPQ)(Melzack, 1975) and the Short-Form MPQ (SF-MPQ)(Melzack, 1987)

Description

The MPQ is a widely used generic pain measure. It consists of 78 descriptors to measure four major dimensions of pain: sensory, affective, evaluative, and miscellaneous (Melzack, 1975). The

descriptors are read to a respondent who selects those words that best describe his/her feelings and sensations at that moment (Melzack, 1975). Three major indices are obtained: 1) the pain rating index based on the rank values of the words, 2) the number of words chosen, and 3) the present pain intensity, which is the number (0-5) and word ("no pain" to "excruciating") combination chosen as the indicator of overall pain intensity (Melzack, 1975). The SF-MPQ consists of 15 descriptors from the sensory and affective categories of the standard MPQ (Melzack, 1987). Each descriptor is rated on an intensity scale of 0 ("none") to 3 ("severe"). Three scores are derived from the sum of the values of the words chosen for sensory, affective, and total descriptors. Two additional items measure overall pain intensity, present pain intensity, a number from 1 ("mild") to 5 ("excruciating"), and a VAS (Melzack, 1987). The validity, reliability, and sensitivity to change of the MPQ and SF-MPQ have been established among different populations including women with PD (Reading, 1979; Eccles, 2005; Chen & Chen, 2010; Katz & Melzack, 2011; Wu *et al.*, 2012).

Critique

Compared to single item pain measures, the MPQ and SF-MPQ offer a more complete approach to pain assessment, measuring not only the sensory component of pain, but also its affective and evaluative aspects (Melzack, 1975, 1987; Katz & Melzack, 2011). In addition, the MPQ and SF-MPQ capture the quality of pain (e.g., dull, cramping, sharp) and its temporal aspects (e.g., the continuity, periodicity) (Jensen, 2003). However, some problems of the MPQ and SF-MPQ may limit their use for PD. First, as a generic pain scale, the MPQ and SF-MPQ do not capture the full range of symptomology in women with PD. Second, some MPQ and SF-MPQ items may not be relevant to women with PD, for example, descriptors such as tingling, itchy, smarting, cold, and freezing are used very infrequently by women with PD (Reading, 1979). Third, the MPQ demands a sophisticated literacy level and an extensive understanding of English (Burckhardt & Jones, 2003). For example, differentiating between words such as "smarting" and "stinging" may be difficult. In addition, understanding and responding to the standard 78-item MPQ demands a sufficient attention span and can be time consuming.

Category II: Tools Designed Specifically to Measure PD Symptoms

The Menstrual Symptom Questionnaire (MSQ)(Chesney & Tasto, 1975a)

Description

The MSQ was designed to differentiate two separate types of PD based on descriptions in the literature: spasmodic PD (characterized by cramp-like pain beginning the first day of menstruation resulting from uterine contraction) and congestive PD (characterized by dull pain accompanied by lethargy and depression prior to the onset of menstruation) (Dalton, 1969; Chesney & Tasto, 1975a). The MSQ includes 25 items: 24 items are statements about symptoms (12 characteristics of spasmodic PD, 12 characteristics of congestive PD). The final item asks respondents to select which type of PD is the most accurate description of their experience. For each of the 12 spasmodic items, a score of 5 ("always") to 1 ("never") is selected based on how often the woman experienced the symptom. Conversely, for the 12 congestive items, scores of 5 to 1 are assigned in reverse order, such that high scores indicate never experiencing these symptoms. The final item has two response options: a woman is scored 5 if she responds that the description of spasmodic PD is most like her experience and 1 if she chooses the congestive PD experience. The total score is calculated by summing responses across the 25 items, with higher scores indicating spasmodic PD (Chesney & Tasto, 1975a). The MSQ has good content validity, but there is less evidence for construct validity and reliability (Chesney & Tasto, 1975a; Cox, 1977; Webster *et al.*, 1979; Nelson *et al.*, 1984; Negriff *et al.*, 2009). The MSQ is sensitive to change (Jay *et al.*, 1986) and generalizable to women of different ages and with different cultural backgrounds (Chesney & Tasto, 1975a, 1975b; Cox, 1977; Nelson *et al.*, 1984; Sigmon & Nelson, 1988).

Critique

The item content of the MSQ is comprehensive of PD, but not all items are relevant. With advances in understanding the prostaglandin-mediated pathogenesis of PD, it is now known that both spasmodic (e.g., "I have cramps that begin on the first day of my period") and congestive symptoms (e.g., "I am constipated during my period") co-occur in women with PD (Dawood, 1981; Kinch, 1985; Jarrett *et al.*, 1996). Both types of symptoms measured by the MSQ are relevant to PD, however, some of the items concerning what women do about their symptoms (e.g., "take a prescription drug" and "use a hot water

bottle") measure behavioral responses to symptoms, not symptoms per se. Construct validity of the MSQ is questionable as the dichotomy of congestive and spasmodic symptoms appears not to be sound (Cox, 1977; Webster *et al.*, 1979), and in subsequent psychometric testing of the MSQ the two-factor structure was not supported (Cox, 1977; Nelson *et al.*, 1984; Negriff *et al.*, 2009). Test-retest reliability was assessed over a 2-week period, which is problematic in that the reliability of the tool could be confounded by the stability/instability of the symptoms over this time span or by existence of recall bias. Another problem with the MSQ is that it is unidimensional, measuring only symptom frequency. Lastly, the recall period of the MSQ was not described. It is unclear whether women are asked to describe their current or recent state (e.g., 28-day recall) or if they are asked to recall symptoms over a longer period of time (e.g., one-year recall or recall since menarche).

The Verbal Multidimensional Scoring System (VMS)(Andersch & Milsom, 1982)

Description

The VMS is a grading system developed to assess the prevalence and severity of PD. Information on its development was lacking in the report of the original epidemiological study (Andersch & Milsom, 1982). The VMS includes four categories of severity from Grade 0 (none) to Grade 3 (severe), with each grade based on criteria regarding pain severity, effects of pain on daily activities, systemic symptoms, and analgesic requirements. For example, the Grade 1 (mild pain) descriptor reads as: "menstruation is painful, but seldom inhibits a woman's normal activity. No systemic symptoms. Analgesics are rarely required."(p 656) (Andersch & Milsom, 1982). The VMS has content validity and construct validity (as evidenced by significant correlations between the VMS scores and VAS pain severity scores) (Andersch & Milsom, 1982), but little evidence of reliability. It is sensitive to change and generalizable across various ages and cultures (Gharloghi *et al.*, 2012; Lindh *et al.*, 2012).

Critique

The VMS intends to measure symptoms beyond pelvic / abdominal pain, and thus items are comprehensive and relevant to PD. According to Andersch and Milsom, the VMS is advantageous in that it is multidimensional, considering four dimensions of dysmenorrhea: pain severity, effects of pain on

daily activities, systemic symptoms, and analgesic requirements (Andersch & Milsom, 1982). However, rather than scoring each dimension separately, the developers combined the four dimensions together into a single score, which constitutes the major limitation of this tool. Suppose a woman rates her menstruation as painful, but she does not require analgesics since she uses non-analgesic medications or non-medication treatments. Such a woman would not fit into any of the given grading categories. The second problem with the VMS is its lack of clarity regarding what the phrase "systemic symptoms" means. Without clarification, such wording could be confusing to some respondents. Third, the instrument has not been evaluated in terms of reliability. Fourth, the time frame of the measurement is not clear; that is, does the VMS measure premenstrual symptoms, menstrual symptoms, or both? Lastly, the recall period of the tool is unclear.

The Retrospective Symptom Scale (RSS, or the Cox RSS) and the Daily Symptom Scale (DSS)(Cox & Meyer, 1978)

Description

The RSS was developed to measure intensity and frequency (duration) of commonly reported physical and emotional dysmenorrheic symptoms (Cox & Meyer, 1978). Women are asked to rate the frequency and severity of 18 symptoms on the basis of the experience of their last menstruation. The frequency ratings range from 0 ("did not occur") to 4 ("lasted several days"), and the severity ratings range from 0 ("not noticeable") to 4 ("very severely bothersome"). Total scores are calculated by summing the products of frequency and severity ratings for the 18 symptoms (Cox & Meyer, 1978). The DSS is identical to the RSS except that it is a daily symptom measurement tool (Cox & Meyer, 1978). The construct validity of the DSS and RSS have been supported using the known groups technique (Cox & Meyer, 1978). Test-retest reliability of the RSS was assessed by correlating symptom scores in two menstrual cycles(Cox & Meyer, 1978). The RSS has been used among young women across cultures and is sensitive to change with PD treatment (Sigmon & Nelson, 1988; Harel *et al.*, 1996; Ma *et al.*, 2010; Liu *et al.*, 2011).

Critique

The RSS and DSS have some important merits. First, they offer relatively comprehensive lists of symptoms associated with PD based on its pathogenesis, including pain at different locations (i.e., cramps, abdominal pain, backache, leg aches, and general aching) and gastrointestinal symptoms (i.e., nausea, vomiting, loss of appetite, diarrhea). Second, the instruments take into account the temporal aspects of symptoms (i.e., frequency or duration). However, there are several problems with these tools. First, the descriptors for the frequency and severity ratings do not exactly correspond to what they intend to measure. For example, "lasted less than 3 hours" reads more like a description of symptom duration than a description of symptom frequency, and "moderate bothersome" reads more like a description of symptom bother or symptom distress than symptom severity per se. Second, some methodological flaws in psychometric testing were noticed. Concurrent validity of the RSS was tested using a newly developed instrument (i.e., the DSS), the validity of which is unknown. The method used to assess test-retest reliability of the RSS was also problematic in that the investigators correlated symptom scores in two menstrual cycles (Cox & Meyer, 1978), but symptoms of PD may vary from cycle to cycle (Jarrett et al., 1996). In addition, the internal consistency of either tool has not been reported. Third, evidence for generalizability of the RSS and DSS is still limited, because the RSS has been used only in young women with PD (Harel et al., 1996; Ma et al., 2010; Liu et al., 2011), and the DSS is rarely cited in the literature. Lastly, recall bias could be an issue for the RSS.

The Symptom Severity Scale (SSS)(Chesney & Tasto, 1975b)

Description

The SSS was developed to measure symptoms associated with dysmenorrhea. Adapted from a symptom rating scale developed by Mullen (Mullen, 1971), the SSS measures 15 symptoms, most of which are pain and gastrointestinal symptoms (Chesney & Tasto, 1975b). Women are asked to recall the degree to which they experienced discomfort during the last menstruation. Each symptom is rated on a five-point scale, ranging from 1 (symptom not present) to 5 (very severely). Ratings on each item are summed to yield an overall SSS score, with higher values indicating greater symptom severity (Chesney & Tasto, 1975b). The SSS has been used among women of a wide age range with PD (Chesney & Tasto,

1975b; Sigmon & Nelson, 1988). However, reliability and construct validity of the scale have not been reported. It has been reported that the SSS detected therapeutic effects brought about by interventions for PD (Chesney & Tasto, 1975b).

Critique

The SSS offers a relatively comprehensive list of symptoms associated with dysmenorrhea, including pain at different locations (i.e., cramps, abdominal pain, backache, leg aches, and general aching) and gastrointestinal symptoms (i.e., nausea, vomiting, loss of appetite, diarrhea). However, several problems are noted. First, reliability and construct validity of the scale have not been documented in the literature. Second, only one dimension of symptoms, symptom severity, is measured. Third, recall bias is a concern for this retrospective measurement tool. Fourth, the generalizability of the tool is limited due to the small number of publications and little diversity in culture and ethnicity among populations that have been studied.

Category III: Tools Designed to Measure Perimenstrual Symptoms

The Menstrual Distress Questionnaire (MDQ)(Moos, 1968)

Description

The MDQ is a multi-symptom tool designed to measure menstrual cycle symptomatology. The items were derived from 1) a review of research, 2) open-ended questionnaires and/or interviews with wives of university graduate students, and 3) symptoms from the Blatt Menopausal Index (Moos, 1968). The MDQ contains 47 symptoms grouped into eight subscales: pain, concentration, behavioral change, autonomic reactions, water retention, negative affect, arousal, and control (Moos, 1968). The pain subscale includes six symptoms: muscle stiffness, headache, cramps, backache, fatigue, and general aches and pains. The MDQ measures severity of each symptom: 1) one week before menstruation (i.e., premenstrual or PM), 2) during menstrual flow (i.e., menstrual or M), and 3) during the remainder of the cycle (i.e., inter-menstrual or IM). A respondent is asked to retrospectively recall her most recent menstrual cycle and her worst menstrual cycle. A six-point rating scale is used in which responses range from 1 for “no experience” to 6 for “acute or partially disabling experience”. Each woman receives a

score on each of the eight subscales in each menstrual phase (PM, M, or IM) by adding together her scores for each of the symptoms on that subscale (Moos, 1968). The MDQ has undergone psychometric evaluation and has been shown to be sensitive to change, and generalizable across various ages and cultures/ethnicities (Moos, 1968; Markum, 1976; Van der Ploeg, 1990; Chen & Chen, 2010).

Critique

The MDQ is one of the most widely used measurement tools in menstrual symptom research (Hawes & Oei, 1992), and is comprehensive of PD symptoms, however, several problems for use in PD should be noted. First, some items on the MDQ may not be relevant to PD (e.g. depression, chest pain, and blind spots) based on the prostaglandin-mediated conceptualization. In fact, some symptom items listed on the MDQ are not symptoms per se, but rather coping behaviors in response to symptoms (e.g., take naps and avoid social activities). Second, the methodologies used to assess psychometric properties are flawed (Hawes & Oei, 1992). To assess test-retest reliability, fifteen women were asked to record their symptoms daily on nine selected days for each menstrual cycle. Correlating daily symptom scores in two consecutive menstrual cycles is problematic due to symptom change across cycles. In addition, the small sample may be subject to selection bias (Hawes & Oei, 1992), and the Pearson correlation coefficient tends to be inflated for small sample sizes (less than 15) (Reinard, 2006). The construct validity of the eight-subscale structure is questionable. The subscales, rather than being grounded in theoretical justification, are based on post-hoc exploratory factor analysis (Markum, 1976). Results from studies which attempted to replicate the original factor structure obtained by Moos have been inconsistent (Van der Ploeg, 1990; Hawes & Oei, 1992). Third, the MDQ is unidimensional and symptom intensity is the only dimension being captured. Fourth, recall bias can be an issue if the MDQ is used as a retrospective measure. Fifth, the use of the MDQ could be limited by its length and poor readability. Each participant needs to rate symptom severity 141 times (47 items * 3 times), and some terms may be difficult to understand for non-health professionals (e.g., lowered motor coordination).

The Daily Symptom Rating Scale (DSRS) (Taylor, 1979)

Description

The DSRS was developed to measure perimenstrual symptoms, with items derived from a list of perimenstrual symptoms commonly identified in the literature. Symptoms related to "positive affect" (i.e., cheerfulness, outgoingness, and energy) were included because these symptoms were "of theoretical interest" to the developers (p. 88) (Taylor, 1979). The DSRS encompasses 17 symptoms grouped into two subscales: an affective subscale (10 items) and a somatic subscale (7 items). A six-point scale is used to measure the intensity of the symptom experience, in which responses range from 0 for "not at all" to 5 for "very large amount". Women are asked to record their symptoms daily for 5 weeks. A menstrual cycle is divided into three phases: premenstrual (PM, 7 days before menstruation), menstrual (M, day one until menstruation ceased), and intermenstrual (IM, the remainder of the cycle) phases. Scores on each symptom are summarized by averaging the daily scores within each phase (i.e., PM, M, or IM). The pelvic or abdominal pain rating in the menstrual phase is taken as the rating for dysmenorrhea. The DSRS has fair validity in terms of measuring perimenstrual symptoms and some evidence of reliability (Taylor, 1979).

Critique

As a daily symptom measure, the DSRS reduces recall bias and permits assessment over several days to capture daily variation of symptoms. In addition, the brief nature of the DSRS makes it a feasible tool for measuring symptoms on a daily basis. However, several problems of the DSRS can be recognized. First, The DSRS is not comprehensive of PD symptoms, as it takes only the pelvic or abdominal pain rating in the menstrual phase as the rating for dysmenorrhea. Second, some symptom items on the DSRS may not be relevant to PD. The ratings for several symptoms (e.g., hopelessness, cheerfulness, outgoingness, energy) did not change very much over the entire cycle, and the symptom scores for PM and M phases were strongly correlated with that of the IM phase (correlations between .82 to .93) (Taylor, 1979). One may question whether these symptoms are dispositional attributes (personality traits), rather than menstrual symptoms. Menstrual symptoms tend to change over a menstrual cycle, while dispositional attributes are relatively stable (McCrae *et al.*, 2002). Third, there has been no evidence to support the two-factor structure of the scale, thus construct validity of the somatic and emotional

subscales is open to question. Fourth, the methodology used to assess test-retest reliability over 2 cycles is flawed (as noted in the critique of the RSS). Fifth, the DSRS is unidimensional; symptom severity is the only dimension that is captured. Sixth, the DSRS is infrequently cited in the literature. There is no evidence for sensitivity to change, and the generalizability to other populations may be limited.

The Daily Rating Form (DRF)(Endicott et al., 1986)

Description

The DRF was developed to describe symptom change across the menstrual cycle and to summarize patterns of change between the pre- and post- menses periods. Symptom items were selected, according to the authors, based on evidence that they describe symptoms that are often more severe premenstrually than postmenstrually. In addition, the authors incorporated two social impairment items (i.e., impaired work and social withdrawal) and several items of authors' interest (e.g., alcohol and drug use) (Endicott *et al.*, 1986). The DRF includes 20 symptoms grouped into five subscales: "dysphoric mood", "physical discomfort", "low energy", "consumption", and "more alcohol, sex, active." The DRF measures severity of symptoms using a six-point scale of 1 for "none of the feature" to 6 for "extremely severe levels of the feature". A respondent is asked to recall her daily symptoms each evening for an entire menstrual cycle. Difference scores in each symptom between postmenstrual days and premenstrual days are calculated by subtracting the means of the 5 postmenstrual days from the means of the 3 highest premenstrual days. To get the summary score for each subscale, the symptom difference scores under each subscale are averaged. PD related symptoms include abdominal pain, back, joint or muscle pain, and bloating, all of which are under the subscale of physical discomfort. Construct validity of the DRF was established through factor analysis (Endicott *et al.*, 1986), while other psychometric properties have not been reported.

Critique

As a concurrent symptom measurement tool, the DRF is not subject to recall bias, however, there are several problems in using the tool to measure symptoms of PD. First, the DRF does not include a comprehensive list of symptoms related to PD. Gastrointestinal symptoms are underrepresented. Second,

the developers incorporated items (e.g., drink alcohol, use drugs) that are not relevant to measuring PD symptoms and that may not be valid in describing menstrual symptom change. Third, reliability of the DRF has not been addressed adequately. Fourth, the construct validity of the DRF was established through factor analysis (Endicott *et al.*, 1986), but the theoretical justification of the five-subscale structure is unclear. Fifth, the item "back, joint, or muscle pain" covers three types of pain, and it cannot differentiate women with only back pain from women with back pain plus joint pain. Sixth, the DRF is not a multidimensional scale; symptom severity is the sole dimension that is captured. Seventh, the scoring system of the DRF is complicated, which may limit its utility. Lastly, though the DRF has been used in different populations (Endicott *et al.*, 1986; Choi *et al.*, 2001; Sit *et al.*, 2011), its generalizability to women with PD is not clear due to its limited use in PD-specific populations.

The Washington Women's Health Diary (WWHD)(Woods, 1987) *and the Menstrual Symptom Severity List (MSSL)(Mitchell et al., 1991)*

Description

The WWHD is a daily symptom rating tool designed to measure perimenstrual symptoms, which are defined by the developers as symptoms occurring immediately before and during menstruation. The WWHD lists 57 symptoms (40 negative and 17 positive) that were developed based on the Moos MDQ (Moos, 1968), the Premenstrual Assessment Form (Halbreich *et al.*, 1982), other literature, and the authors' experience (Woods, 1987). Women are asked to rate their symptom experiences daily for two to three menstrual cycles on a 0-4 scale, where 0 represents "not present" and 4 represents "extreme". A total symptom severity score is calculated based on the 40 negative symptoms. Symptom severity scores are calculated for days 4 through 10 postmenses, and days -7 through -1 premenses. The means of the three most severe symptomatic days are calculated for both phases (Woods, 1987). The 33-item MSSL was derived from the WWHD to measure premenstrual symptom severity patterns (including low severity or LS pattern, PMS pattern, and premenstrual magnification or PMM pattern). Items on the WWHD were removed if (1) they had minimal variance, (2) they were redundant based on the intercorrelations of .80 or higher, (3) they were positive symptoms, and (4) they were not in one of the five negative symptom

clusters from the principal component analysis (Mitchell *et al.*, 1991). Premenses scores are calculated by totaling the scores of days -5 to -1, that is, 5 days before the menses, while postmenses scores are calculated by totaling scores of days 6-10 (Mitchell *et al.*, 1991). The WWHD and MSSL have good validity and reliability in terms of measuring premenstrual symptoms (Mitchell *et al.*, 1991; Woods *et al.*, 1995, 1998, 1999), but their sensitivity to change has not been reported. The tools are generalizable to women of various ages and culture/ethnicities (Woods, 1987; Woods *et al.*, 1998; Kim, 2004).

Critique

Both the WWHD and the MSSL measure a wide range of perimenstrual symptoms, including symptoms associated with dysmenorrhea, such as abdominal pain, discomfort (other than cramps), backache, uterine or pelvic cramps, general aches and pains, decreased appetite, diarrhea, and nausea. As a daily symptom measure, recall bias is reduced as compared with the retrospective symptom measures over a longer period of time (e.g., recall symptoms over the most recent menstrual period) (Woods, 1987). They are high quality tools for capturing perimenstrual symptoms in general, however, their applicability to women with PD may be limited. First, the WWHD and MSSL intend to capture a wide spectrum of premenstrual symptoms, including symptoms of PMS and premenstrual magnification (i.e., increasing intensity of the symptoms of nonmenstrual conditions) (Woods, 1987; Mitchell *et al.*, 1991). Some items on the WWHD may not be relevant to symptom measurement in PD, for example, skin disorders, suicidal thoughts, bursts of energy, and intentional self-injury. Second, previous studies selected heterogeneous samples that included women with different premenstrual conditions (e.g., premenstrual magnification and premenstrual syndrome). Transferability of the psychometric properties to the PD-specific population is unclear, because these two tools have not been evaluated specifically among women with PD. Moreover, these two measurement tools are unidimensional, covering only the dimension of symptom severity.

Discussion

Main Issues in PD Symptom Measurement

Symptoms to measure

Lower abdominal/pelvic pain is the most salient symptom among women with dysmenorrhea. While it is common to use lower abdominal pain as a key indicator of PD, the release of prostaglandins and other inflammatory mediators may produce multiple symptoms (Dawood, 1981, 1984; Kinch, 1985). A comprehensive tool should include a broad spectrum of PD symptoms related to the prostaglandin mechanism. Because interventions for PD may have differential effects on multiple symptoms, a multi-symptom measure may be more sensitive to therapeutic change than a tool measuring only pain.

The three categories of tools reviewed measure different symptoms. Generic pain measures only assess pain, and are not comprehensive of symptoms experienced in PD. Tools designed to measure perimenstrual symptoms cover a wide range of symptoms, but many are not relevant to dysmenorrhea (e.g., symptoms associated with PMDD, premenstrual magnification of psychological problems). In addition, because these instruments have been developed for women with various menstrual problems, their psychometric properties may not be suitable in PD. Items in tools designed specifically to measure PD are both comprehensive of and relevant to PD. Thus, these measures are a logical choice for PD research, however, they vary in terms of psychometric properties.

Psychometric properties of PD-specific tools

Among the PD-specific tools, the MSQ and the VMS are problematic in terms of validity. The validity of the "spasmodic/congestive" construct of the MSQ has been widely criticized by researchers (Webster *et al.*, 1979), and the VMS is flawed in using a single score to measure conceptually different constructs (i.e., pain severity, pain interference, existence of other symptoms, behavioral response to symptoms). The RSS, DSS, and SSS, are most consistent with the conceptualization of PD as a prostaglandin-mediated symptom-complex. Evidence of validity and reliability of the RSS was established in the original study (Cox & Meyer, 1978), and there is some evidence supporting the sensitivity of the RSS to treatment effects (Harel *et al.*, 1996). However, evidence of internal consistency reliability of the RSS has not been reported, and psychometric properties of the DSS and SSS have rarely been evaluated due to their infrequent use. Some methodological flaws were present in psychometric testing, such as testing concurrent validity of the DSS using a newly developed tool and assessing test-

retest reliability of the RSS across rather than within menstrual cycles. These measures have potential but need further development and validation. Given limited generalizability of the tools, an important step will be psychometric testing in the country of study's respective language, particularly where the meaning of pain quality descriptors may be lost in translation.

Relevant dimensions

Most of the instruments reviewed are unidimensional, evaluating only symptom intensity/severity. Measures of other symptom dimensions, such as frequency, duration, quality, and distress can provide valuable information in understanding PD and evaluating effects of treatment. The RSS and DSS (which measure symptom duration) and the MSD (which measures the time of symptom onset) are the only instruments that touch on temporal aspects of symptoms. Temporal aspects of pain and other symptoms may be particularly relevant in PD, as symptoms often occur in intermittent episodes with symptom-free intervals. Symptom frequency and duration may impact symptom interference with daily life, and play a role in selecting symptom interventions (Jensen, 2003). In addition, symptom duration is important in identifying symptoms occurring at other phases of the menstrual cycle, and thus due to other pelvic pathology (Hofmeyr, 1996).

Only the MSD and MPQ include items on pain quality (e.g., "not intense, the continuous dull aching", "cramps") that may be useful in comprehensively understanding a woman's experience of PD. Visceral pain is often described as vague, dull, or periodic, while somatic pain is commonly described as sharp and localized (Siddall & Cousins, 1998). Assessing pain quality allows better characterization of PD pain components and could facilitate the identification of changes that signal pain sensitization and increased risk for future pain conditions. Given the likelihood that some treatments will impact specific qualities of pain more than others (Jensen & Karoly, 2011), inclusion of pain quality measures may also help test differential effects of PD treatments.

Symptom distress is rarely evaluated with the tools reviewed. The RSS and the DSS phrase the descriptors for intensity in terms of bother (i.e., not noticeable to very severely bothersome), but none of the tools were specifically designed to measure PD symptom distress. In addition, it is unknown whether

a single global measure of PD symptom distress would be adequate for measuring affective components of the symptom experience, or whether a multi-item measure of symptom distress is needed.

Retrospective recall bias

The measures reviewed are designed to be used with different recall periods, such as daily (24-hours) recall (i.e., DSS, DSRS, DRF, WWHD, and MSSL), recall of the most recent menstrual cycle (i.e., RSS, SSS, MDQ), and recall of the worst menstrual cycle (i.e., MDQ). Research suggests that recurring pain (like menstrual pain) may be remembered less accurately than novel or acute pain (Erskine *et al.*, 1990). Investigators have suggested that stereotypes about menstruation and PMS can lead women to overestimate their menstrual symptoms during recall (Woods *et al.*, 1982; McFarland *et al.*, 1989). Underestimation of menstrual symptom intensity has also been reported (Jukic *et al.*, 2008). Though menstrual cramps and premenstrual backache were less susceptible to recall bias than were emotional symptoms, the concordance between daily recall and recall of the most recent menstrual cycle was low (Woods *et al.*, 1982). Generally speaking, longer recall periods are more susceptible to recall bias. An ideal measure would use a short recall period. Research comparing 24-hour recall and momentary assessment supports the validity of a 24-hour recall for post-operative pain (Jensen *et al.*, 2008) and chronic pain in rheumatology patients (Broderick *et al.*, 2008). Given these findings, a 24-hour timeframe may be reasonable for measuring PD symptoms, but future research should confirm the validity of this recall period.

Future research

Use of PD-specific tools is a logical choice, however, such tools could be strengthened by incorporating additional symptom dimensions (e.g., symptom frequency, duration, pain quality, and distress) and by conducting psychometric testing using sound methodology and diverse samples. While there is no consensus on the optimal set of symptoms for measuring PD, items should be both comprehensive of and relevant to current undertanding of PD mechanisms. Currently, prostaglandin-mediated symptoms should be measured, including pain at different locations (i.e., cramps , backache, upper thigh pain, and general aching) and gastrointestinal symptoms (i.e., nausea, vomiting, decreased

appetite, diarrhea). Future advances in understanding of PD pathophysiology may result in the need for further revision of PD symptom measurement tools. For example, Giamberardino *et al*. have hypothesized that viscero-somatic pain referral is involved in the pathophysiology of PD (Giamberardino, 2003). Their work, along with others, may broaden our understanding of PD, necessitating a revision in the symptoms that are critical for measurement.

Given the fact that recall bias is a major concern in retrospective measurement of PD symptoms, ecological momentary assessment (EMA) may be a promising method for future research. The symptoms of PD are episodic in nature, fluctuating even throughout a single day. In this sense, even a 24-hour recall may be subject to bias. Use of EMA would allow research participants to report symptoms at the time of occurrence, thereby reducing reliance on memory and improving accuracy. In addition, EMA provides multiple assessments over time, rather than a single recall rating. This allows for not only detection of symptom pattern over time but also aggregating multiple data points, which increases reliability of the measure (Jensen & McFarland, 1993). Moreover, compliance rates for symptom reporting can be improved by electronic cueing (Jensen & Karoly, 2011). EMA has not been used among women with PD; however, previous research on other pain conditions has shown high compliance and user satisfaction (Jensen & Karoly, 2011).

Limitations

Limitations of this review should be noted. First, we queried different databases and searched the bibliographies of relevant literature, however, our search may have missed qualifying instruments or reports of psychometric properties, especially if they were not published in English. Second, our search may have missed studies describing the use of a specific tool, especially if a different name for the tool was used. Third, we reported only the initial evidence for sensitivity to change (i.e., whether a measurement tool was able to show intervention effects in any interventional study), and we did not quantify sensitivity to change. Future work shuld evaluate senstitive to change using advanced statistical methods (Husted *et al.*, 2000). Despite these limitations, this work adds to the literature by critically reviewing self-report pain and symptom measures necessary to advance PD research. Further work on

measurement adaptation and validation is needed as outlined above. Following this step, researchers can more fully describe the PD symptom experience and develop and test interventions targeted to those specific symptom components and their underlying pathophysiology.

Acknowledgements

The authors wish to express appreciation to Patricia Becker, PhD, RN, FAAN for her helpful comments in the preparation of this paper.

Author Contributions

All authors have collaborated on the conceptualization, design, interpretation of the data and drafting the manuscript; and all have read and concur with the submitted version.

References

Altman, G., Cain, K.C., Motzer, S., Jarrett, M., Burr, R., & Heitkemper, M. (2006) Increased symptoms in female IBS patients with dysmenorrhea and PMS. *Gastroenterol. Nurs. Off. J. Soc. Gastroenterol. Nurses Assoc.*, **29**, 4–11.

American Psychiatric Association, American Psychiatric Association, & Task Force on DSM-IV (2000) *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR.* American Psychiatric Association, Washington, DC.

Andersch, B. & Milsom, I. (1982) An epidemiologic study of young women with dysmenorrhea. *Am. J. Obstet. Gynecol.*, **144**, 655–660.

Banikarim, C., Chacko, M.R., & Kelder, S.H. (2000) Prevalence and impact of dysmenorrhea on Hispanic female adolescents. *Arch. Pediatr. Adolesc. Med.*, **154**, 1226–1229.

Berkley, K.J. & McAllister, S.L. (2011) Don't dismiss dysmenorrhea! *Pain*, **152**, 1940–1941.

Booton, D.A. & Seideman, R.Y. (1989) Relationship between premenstrual syndrome and dysmenorrhea. *AAOHN J. Off. J. Am. Assoc. Occup. Health Nurses*, **37**, 308–315.

Broderick, J.E., Schwartz, J.E., Vikingstad, G., Pribbernow, M., Grossman, S., & Stone, A.A. (2008) The accuracy of pain and fatigue items across different reporting periods. *Pain*, **139**, 146–157.

Budeiri, D.J., Li Wan Po, A., & Dornan, J.C. (1994) Clinical trials of treatments of premenstrual syndrome: entry criteria and scales for measuring treatment outcomes. *Br. J. Obstet. Gynaecol.*, **101**, 689–695.

Burckhardt, C.S. & Jones, K.D. (2003) Adult Measures of Pain: The McGill Pain Questionnaire (MPQ). *Arthritis Rheum. Arthritis Care Res.*, **49**, S96–S97.

Burnett, M.A., Antao, V., Black, A., Feldman, K., Grenville, A., Lea, R., Lefebvre, G., Pinsonneault, O., & Robert, M. (2005) Prevalence of primary dysmenorrhea in Canada. *J. Obstet. Gynaecol. Can. JOGC J. Obstétrique Gynécologie Can. JOGC*, **27**, 765–770.

Chan, W.Y., Yusoff Dawood, M., & Fuchs, F. (1981) Prostaglandins in primary dysmenorrhea: Comparison of prophylactic and nonprophylactic treatment with ibuprofen and use of oral contraceptives. *Am. J. Med.*, **70**, 535–541.

Chantler, I., Mitchell, D., & Fuller, A. (2009) Actigraphy quantifies reduced voluntary physical activity in women with primary dysmenorrhea. *J. Pain Off. J. Am. Pain Soc.*, **10**, 38–46.

Chen, H. & Chen, C. (2010) Effects of acupressure on menstrual distress in adolescent girls: a comparison between Hegu-Sanyinjiao Matched Points and Hegu, Zusanli single point. *J. Clin. Nurs.*, **19**, 998–1007.

Chesney, M.A. & Tasto, D.L. (1975a) The development of the menstrual symptom questionnaire. *Behav. Res. Ther.*, **13**, 237–244.

Chesney, M.A. & Tasto, D.L. (1975b) The effectiveness of behavior modification with spasmodic and congestive dysmenorrhea. *Behav. Res. Ther.*, **13**, 245–253.

Choi, S.H., Kang, S.B., & Joe, S.H. (2001) Changes in premenstrual symptoms in women with schizophrenia: a prospective study. *Psychosom. Med.*, **63**, 822–829.

Coco, A.S. (1999) Primary dysmenorrhea. *Am. Fam. Physician*, **60**, 489–496.

Cox, D.J. (1977) Menstrual symptom questionnaire: further psychometric evaluation. *Behav. Res. Ther.*, **15**, 506–508.

Cox, D.J. & Meyer, R.G. (1978) Behavioral treatment parameters with primary dysmenorrhea. *J. Behav. Med.*, **1**, 297–310.

Dalton, K. (1969) Period Pain. In *The Menstrual Cycle*. Pantheon Books, New York, pp. 39–45.

Davis, A.R. & Westhoff, C.L. (2001) Primary dysmenorrhea in adolescent girls and treatment with oral contraceptives. *J. Pediatr. Adolesc. Gynecol.*, **14**, 3–8.

Dawood, M.Y. (1981) Prostaglandins, hormones and dysmenorrhea. In Dawood, M.Y. (ed), *Dysmenorrhea*. Williams & Wilkins, Baltimore, pp. 21–52.

Dawood, M.Y. (1984) Ibuprofen and dysmenorrhea. *Am. J. Med.*, **77**, 87–94.

Dawood, M.Y. (1988) Nonsteroidal anti-inflammatory drugs and changing attitudes toward dysmenorrhea. *Am. J. Med.*, **84**, 23–29.

Dawood, M.Y. (2006) Primary dysmenorrhea: advances in pathogenesis and management. *Obstet. Gynecol.*, **108**, 428–441.

Douglas, S. (2002) Premenstrual syndrome. Evidence-based treatment in family practice. *Can. Fam. Physician Médecin Fam. Can.*, **48**, 1789–1797.

Durain, D. (2004) Primary dysmenorrhea: assessment and management update. *J. Midwifery Womens Health*, **49**, 520–528.

Eccles, N. (2005) A randomized, double-blinded, placebo-controlled pilot study to investigate the effectiveness of a static magnet to relieve dysmenorrhea. *J. Altern. Complement. Med.*, **11**, 681–687.

Endicott, J., Nee, J., Cohen, J., & Halbreich, U. (1986) Premenstrual changes: Patterns and correlates of daily ratings. *J. Affect. Disord.*, **10**, 127–135.

Erskine, Morley, & Pearce (1990) Memory for pain: a review. *Pain 03043959*, **41**, 255–265.

Gagliese, L. & Melzack, R. (1997) Age differences in the quality of chronic pain: A preliminary study. *Pain Res. Manag.*, **2**, 157–162.

Gharloghi, S., Torkzahrani, S., Akbarzadeh, A.R., & Heshmat, R. (2012) The effects of acupressure on severity of primary dysmenorrhea. *Patient Prefer. Adherence*, **6**, 137–142.

Giamberardino, M. (2008) Women and visceral pain: are the reproductive organs the main protagonists? Mini-review at the occasion of the "European Week Against Pain in Women 2007". *Eur. J. Pain*, **12**, 257–260.

Giamberardino, M.A. (2003) Referred muscle pain/hyperalgesia and central sensitisation. *J. Rehabil. Med. Off. J. UEMS Eur. Board Phys. Rehabil. Med.*, 85–88.

Halbreich, U., Endicott, J., & Schacht, S. (1982) Premenstrual syndromes: A new instrument for their assessment. *J. Psychiatr. Treat. Eval.*, **4**, 161–164.

Harel, Z., Biro, F.M., Kottenhahn, R.K., & Rosenthal, S.L. (1996) Supplementation with omega-3 polyunsaturated fatty acids in the management of dysmenorrhea in adolescents. *Am. J. Obstet. Gynecol.*, **174**, 1335–1338.

Hawes, E. & Oei, T.P.S. (1992) The menstrual distress questionnaire: Are the critics right? *Curr. Psychol.*, **11**, 264.

Haywood, A., Slade, P., & King, H. (2002) Assessing the assessment measures for menstrual cycle symptoms: a guide for researchers and clinicians. *J. Psychosom. Res.*, **52**, 223–237.

Hofmeyr, G. (1996) Dysmenorrhea. In Bassin, J. (ed), *Topics in Obstetrics and Gynaecology*. Julmar Communications, Johannesburg, pp. 269–274.

Husted, J.A., Cook, R.J., Farewell, V.T., & Gladman, D.D. (2000) Methods for assessing responsiveness: a critical review and recommendations. *J. Clin. Epidemiol.*, **53**, 459–468.

Iacovides, S., Baker, F.C., Avidon, I., & Bentley, A. (2013) Women With Dysmenorrhea Are Hypersensitive to Experimental Deep Muscle Pain Across the Menstrual Cycle. *J. Pain Off. J. Am. Pain Soc.*, **14**, 1066–1076.

International Association for the Study of Pain (1994) Group XXIV: Diseases of the bladder, uterus, ovaries, and anexa. In *Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms*, 2nd ed. edn. IASP Press, Seattle, Washington, pp. 163–171.

Jarrett, M., Cain, K., Heitkemper, M., & Levy, R. (1996) Relationship between gastrointestinal and dysmenorrheic symptoms at menses. *Res. Nurs. Health*, **19**, 45–51.

Jay, M.S., Durant, R.H., Shoffitt, T., & Linder, C.W. (1986) Differential response by adolescents to naproxen sodium therapy for spasmodic and congestive dysmenorrhea. *J. Adolesc. Health Care Off. Publ. Soc. Adolesc. Med.*, **7**, 395–400.

Jensen, M. (2003) The validity and reliability of pain measures for use in clinical trials in adults: review paper written for the initiative on methods, measurement, and pain assessment in clinical trials (IMMPACT) meeting. Washington, DC.

Jensen, M., P., Mardekian, J., Lakshminarayanan, M., & Boye, M., E. (2008) Validity of 24-h recall ratings of pain severity: biasing effects of "Peak" and "End" pain. *Pain 03043959*, **137**, 422–427.

Jensen, M.K. & Karoly, P. (2011) Self-report scales and procedures for assessing pain in adults. In Turk, D.C. & Melzack, R. (eds), *Handbook of Pain Assessment*, 3rd ed. edn. Guilford Press, New York, pp. 19–44.

Jensen, M.P. & McFarland, C.A. (1993) Increasing the reliability and validity of pain intensity measurement in chronic pain patients. *Pain*, **55**, 195–203.

Ju, H., Jones, M., & Mishra, G. (2014) The prevalence and risk factors of dysmenorrhea. *Epidemiol. Rev.*, **36**, 104–113.

Jukic, A.M.Z., Weinberg, C.R., Baird, D.D., Hornsby, P.P., & Wilcox, A.J. (2008) Measuring menstrual discomfort: a comparison of interview and diary data. *Epidemiol. Camb. Mass*, **19**, 846–850.

Katz, J. & Melzack, R. (2011) The McGill Pain Questionnaire: Development, Psychometric Properties, and Usefulness of the Long Form, Short Form, and Short Form-2. In Turk, D.C. & Melzack, R. (eds), *Handbook of Pain Assessment*, 3rd ed. edn. Guilford Press, New York, pp. 45–66.

Kim, H. won (2004) Perimenstrual symptoms of Korean women living in the USA: applicability of the WDHD(Women's Daily Health Diary) on prospective report. *Taehan Kanho Hakhoe Chi*, **34**, 1395–1401.

Kinch, R. (1985) Dysmenorrhea: a historical perspective. In Dawood, M.Y., McGuire, J.L., & Demers, L.M. (eds), *Premenstrual Syndrome and Dysmenorrhea*. Urban & Schwarzenberg, Baltimore, pp. 79–85.

Larroy, C. (2002) Comparing visual-analog and numeric scales for assessing menstrual pain. *Behav. Med. Wash. DC*, **27**, 179–181.

Lindh, I., Ellström, A.A., & Milsom, I. (2012) The effect of combined oral contraceptives and age on dysmenorrhoea: an epidemiological study. *Hum. Reprod.*, **27**, 676–682.

Liu, C.Z., Xie, J.P., Wang, L.P., Zheng, Y.Y., Ma, Z.B., Yang, H., Chen, X., Shi, G.X., Li, S.L., Zhao, J.-P., Han, J.-X., Li, J.D., Wang, Y.X., Tang, L., Xue, X.O., Li, M., Wang, Y., Sun, A.P., Xing, J.M.,

Cao, H.J., Zhu, J., & Liu, J.P. (2011) Immediate analgesia effect of single point acupuncture in primary dysmenorrhea: a randomized controlled trial. *Pain Med. Malden Mass*, **12**, 300–307.

Lundstrom, V. (1981) Uterine activity during the normal cycle and dysmenorrhea. In Dawood, M.Y. (ed), *Dysmenorrhea*. Williams & Wilkins, Baltimore, pp. 53–74.

Ma, Y.X., Ma, L.X., Liu, X. lian, Ma, Y.X., Lv, K., Wang, D., Liu, J.P., Xing, J.M., Cao, H.J., Gao, S.Z., & Zhu, J. (2010) A comparative study on the immediate effects of electroacupuncture at Sanyinjiao (SP6), Xuanzhong (GB39) and a non-meridian point, on menstrual pain and uterine arterial blood flow, in primary dysmenorrhea patients. *Pain Med. Malden Mass*, **11**, 1564–1575.

Mannix, L. (2008) Menstrual-related pain conditions: dysmenorrhea and migraine. *J. Womens Health 15409996*, **17**, 879–891.

Marjoribanks, J., Proctor, M., Farquhar, C., & Derks, R.S. (2010) Nonsteroidal anti-inflammatory drugs for dysmenorrhoea. *Cochrane Database Syst. Rev.*, CD001751.

Markum, R.A. (1976) Assessment of the reliability of and the effect of neutral instructions on the symptom ratings on the Moos Menstrual Distress Questionnaire. *Psychosom. Med.*, **38**, 163–172.

McCrae, R.R., Costa, P.T.J., Terracciano, A., Parker, W.D., Mills, C.J., De Fruyt, F., & Mervielde, I. (2002) Personality trait development from age 12 to age 18: Longitudinal, cross-sectional and cross-cultural analyses. *J. Pers. Soc. Psychol.*, **83**, 1456–1468.

McFarland, C., Ross, M., & DeCourville, N. (1989) Women's theories of menstruation and biases in recall of menstrual symptoms. *J. Pers. Soc. Psychol.*, **57**, 522–531.

Melzack, R. (1975) The McGill Pain Questionnaire: major properties and scoring methods. *Pain*, **1**, 277–299.

Melzack, R. (1987) The short-form McGill Pain Questionnaire. *Pain*, **30**, 191–197.

Mitchell, E.S., Woods, N.F., & Lentz, M.J. (1991) Recognizing PMS when you see it: Criteria for PMS sample selection. In Taylor, D.L. & Woods, N.F. (eds), *Menstruation, Health, and Illness*. Hemisphere Publishing Corp, Washington, DC, pp. 89–102.

Moos, R.H. (1968) The development of a menstrual distress questionnaire. *Psychosom. Med.*, **30**, 853–867.

Mullen, F.G. (1971) Treatment of dysmenorrhea by professional and student behavior therapists. Presented at the Fifth Annual Meeting of the Association for the Advancement of Behavior Therapy, Washington, DC.

Negriff, S., Dorn, L., Hillman, J., & Huang, B. (2009) The measurement of menstrual symptoms: factor structure of the Menstrual Symptom Questionnaire in adolescent girls. *J. Health Psychol.*, **14**, 899–908.

Nelson, R.O., Sigmon, S., Amodei, N., & Jarrett, R.B. (1984) The menstrual symptom questionnaire: the validity of the distinction between spasmodic and congestive dysmenorrhea. *Behav. Res. Ther.*, **22**, 611–614.

Nunnally, J.C. (1994) *Psychometric Theory*, 3rd ed. edn, McGraw-Hill series in psychology. McGraw-Hill, New York.

Nur Azurah, A.G., Sanci, L., Moore, E., & Grover, S. (2013) The quality of life of adolescents with menstrual problems. *J. Pediatr. Adolesc. Gynecol.*, **26**, 102–108.

Olafsdottir, L.B., Gudjonsson, H., Jonsdottir, H.H., Björnsson, E., & Thjodleifsson, B. (2012) Natural history of irritable bowel syndrome in women and dysmenorrhea: a 10-year follow-up study. *Gastroenterol. Res. Pract.*, **2012**, 534204.

Polit, D.F. (2012) *Nursing Research: Generating and Assessing Evidence for Nursing Practice*, Ninth Edition. edn. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia.

Reading (1979) The internal structure of the McGill pain questionnaire in dysmenorrhoea patients. *Pain 03043959*, **7**, 353–358.

Reinard, J.C. (2006) Ensuring Reliability and Validity. In *Communication Research Statistics*. SAGE Publications, Thousand Oaks, Calif.

Shaver, J., Wilbur, J., Robinson, F., Wang, E., & Buntin, M. (2006) Women's health issues with fibromyalgia syndrome. *J. Womens Health 15409996*, **15**, 1035–1045.

Siddall, P.J. & Cousins, M.J. (1998) Introduction to pain mechanisms: implications for neural blockade. In Cousins, M.J. & Bridenbaugh, P.O. (eds), *Neural Blockade in Clinical Anesthesia and Management of Pain*, 3rd ed. edn. Lippincott-Raven, Philadelphia, pp. 675–699.

Sigmon, S.T. & Nelson, R.O. (1988) The effectiveness of activity scheduling and relaxation training in the treatment of spasmodic dysmenorrhea. *J. Behav. Med.*, **11**, 483–495.

Sit, D., Seltman, H., & Wisner, K.L. (2011) Menstrual effects on mood symptoms in treated women with bipolar disorder. *Bipolar Disord.*, **13**, 310–317.

Svanberg, L. & Ulmsten, U. (1981) The incidence of primary dysmenorrhea in teenagers. *Arch. Gynecol.*, **230**, 173–177.

Taylor, J.W. (1979) The timing of menstruation-related symptoms assessed by a daily symptom rating scale. *Acta Psychiatr. Scand.*, **60**, 87–105.

Tu, C.-H., Niddam, D.M., Chao, H.-T., Chen, L.-F., Chen, Y.-S., Wu, Y.-T., Yeh, T.-C., Lirng, J.-F., & Hsieh, J.-C. (2010) Brain morphological changes associated with cyclic menstrual pain. *Pain*, **150**, 462–468.

Tu, C.-H., Niddam, D.M., Chao, H.-T., Liu, R.-S., Hwang, R.-J., Yeh, T.-C., & Hsieh, J.-C. (2009) Abnormal cerebral metabolism during menstrual pain in primary dysmenorrhea. *NeuroImage*, **47**, 28–35.

Tu, F.F., Epstein, A.E., Pozolo, K.E., Sexton, D.L., Melnyk, A.I., & Hellman, K.M. (2013) A Noninvasive Bladder Sensory Test Supports A Role for Dysmenorrhea Increasing Bladder Noxious Mechanosensitivity. *Clin. J. Pain*,.

Tugay, N., Akbayrak, T., Demirtürk, F., Karakaya, I.C., Kocaacar, O., Tugay, U., Karakaya, M.G., & Demirtürk, F. (2007) Effectiveness of transcutaneous electrical nerve stimulation and interferential current in primary dysmenorrhea. *Pain Med. Malden Mass*, **8**, 295–300.

Unsal, A., Ayranci, U., Tozun, M., Arslan, G., & Calik, E. (2010) Prevalence of dysmenorrhea and its effect on quality of life among a group of female university students. *Ups. J. Med. Sci.*, **115**, 138–145.

Van der Ploeg, H.M. (1990) The factor structure of the Menstrual Distress Questionnaire—Dutch. *Psychol. Rep.*, **66**, 707–714.

Vincent, K., Warnaby, C., Stagg, C.J., Moore, J., Kennedy, S., & Tracey, I. (2011) Dysmenorrhoea is associated with central changes in otherwise healthy women. *Pain*, **152**, 1966–1975.

Webster, S.K., Martin, H.J., Uchalik, D., & Gannon, L. (1979) The Menstrual Symptom Questionnaire and spasmodic/congestive dysmenorrhea: measurement of an invalid construct. *J. Behav. Med.*, **2**, 1–19.

Woods, N., Mitchell, E., & Lentz, M. (1995) Social pathways to premenstrual symptoms. *Res. Nurs. Health*, **18**, 225–237.

Woods, N.F. (1987) Premenstrual symptoms: another look. *Public Health Rep.*, **102**, 106.

Woods, N.F., Lentz, M.J., Mitchell, E.S., Shaver, J., & Heitkemper, M. (1998) Luteal phase ovarian steroids, stress arousal, premenses perceived stress, and premenstrual symptoms. *Res. Nurs. Health*, **21**, 129–142.

Woods, N.F., Mitchell, E.S., & Lenz, M. (1999) Premenstrual symptoms: delineating symptom clusters. *J. Womens Health Gend. Based Med.*, **8**, 1053–1062.

Woods, N.F., Most, A., & Dery, G.K. (1982) Estimating perimenstrual distress: a comparison of two methods. *Res. Nurs. Health*, **5**, 81–91.

World Health Orgnization (WHO) (2013) 2012 International Classification of Diseases Dignostic Codes [WWW Document]. *2012 ICD-9 Data*,. URL http://www.icd9data.com/2012/Volume1/default.htm

Wu, L., Su, C., & Liu, C. (2012) Effects of Noninvasive Electroacupuncture at Hegu (LI4) and Sanyinjiao (SP6) Acupoints on Dysmenorrhea: A Randomized Controlled Trial. *J. Altern. Complement. Med.*, **18**, 137–142.

Yunus, M.B., Masi, A.T., & Aldag, J.C. (1989) A controlled study of primary fibromyalgia syndrome: clinical features and association with other functional syndromes. *J. Rheumatol. Suppl.*, **19**, 62–71.

Table 1. Major Differences Between PD and PMS/PMDD

	PD	PMS/ PMDD
Diagnostic Code	ICD-9-CM 625.3	ICD-9-CM 625.4 (World Health Orgnization (WHO), 2013) DSM 311 (for PMDD)(American Psychiatric Association *et al.*, 2000)
Predominant /classic symptoms	Pelvic cramps or pain(World Health Orgnization (WHO), 2013)	Mood symptoms (e.g., depressed mood, irritability, affective liability,anger)(American Psychiatric Association *et al.*, 2000; World Health Orgnization (WHO), 2013)
Secondary/ other symptoms	Referred back, thigh pain, GI symptoms (e.g. nausea, vomiting and change in bowel frequency) secondary to the release of PGs and other inflammatory substances (Dawood, 1981, 1984)	Symptoms related to salt and water retention (weight gain, swelling and bloating) (World Health Orgnization (WHO), 2013)
Timing of Symptoms	Start several hours before or during menstruation; May get worse once menstruation begins (Booton & Seideman, 1989)	Start several days before menstruation; Typically relieved with the onset of menstruation (Booton & Seideman, 1989; Coco, 1999)
Classic Medical Treatment	NSAIDs, OCs (Coco, 1999)	Antidepressants particularly SRIs (Douglas, 2002)

PD, primary dysmenorrhea; PMS, premenstrual syndrome; PMDD, premenstrual dysphoric disorder; NSAIDs, non-steroidal anti-inflammatory drugs; PGs, prostaglandins, OCs, oral contraceptives. SRIs, serotonin reuptake inhibitors; ICD: International Classification of Diseases; DSM, Diagnostic and Statistical Manual of Mental Disorders

Table 2. Summary of strengths of the self-report pain and symptom measures for PD

Category	**Generic Pain Measures**		**Tools Designed Specifically to Measure PD Symptoms**					**Tools Designed to Measure Perimenstrual Symptoms**				
	NRSs, VASs, VRSs, FPSs	MPQ/ SF-MPQ	MSQ	VMS	RSS	DSS	SSS	MDQ	DSRS	DRF	WWHD	MSSL
Valid (Content) • Comprehensiveness • Relevance	 ++ +++	 ++ ++	 +++ +++	 +++ +++	 +++ +++	 +++ +++	 +++ +++	 +++ ++	 ++ ++	 ++ ++	 +++ ++	 +++ ++
Valid (Construct)	+++	+++	++	+++	+++	+++	+	++	+++	++	+++	+++
Reliable (Test-retest)	+++	+++	++	+	++	+	+	++	++	+	+++	+++
Reliable (Internal consistency)	n/a	+++	+	+	+	+	+	+++	+++	+	+++	+++
Sensitive to change	+++	+++	+++	+++	+++	+++	+++	+++	+	+	+	+
Generalizable • # of publications • Wide age range • Diverse culture/ethnicity	 +++* +++ +++	 +++ +++ +++	 +++ +++ +++	 +++ +++ +++	 ++ ++ +++	 + + +	 ++ +++ ++	 +++ +++ +++	 + ++ +	 + +++ +++	 + +++ +++	 + +++ +++
Multidimensional	+	+++	+	+	+++	+++	+	+	+	+	+	+
Does NOT require high literacy	+++	+	+++	+++	+++	+++	+++	+	+++	+++	+++	+++

NOT subject to recall bias	depends on time frame specified	depends on time frame specified	+	+	+	+++	+	+	+++	+++	+++	+++
Does NOT combine symptoms and coping behaviors	+++	+++	+	+	+++	+++	+++	+	+++	+	+++	+++

NRSs, Numerical Rating Scales; VASs, Visual Analogue Scales; VRSs, Verbal Rating Scales; FPSs, Facial Pain Scales; MPQ, the McGill Pain Questionnaire (Melzack, 1975); SF-MPQ, the Short-form McGill Pain Questionnaire (Melzack, 1987); MSQ, the Menstrual Symptom Questionnaire (Chesney & Tasto, 1975a); VMS, the Verbal Multidimensional Scoring System (Andersch & Milsom, 1982); RSS, the Retrospective Symptom Scale (Cox & Meyer, 1978); DSS, the Daily Symptom Scale (Cox & Meyer, 1978); SSS, the Symptom Severity Scale (Chesney & Tasto, 1975b); MDQ, the Menstrual Distress Questionnaire (Moos, 1968); DSRS, the Daily Symptom Rating Scale (Taylor, 1979); DRF, the Daily Rating Form (Endicott *et al.*, 1986); WWHD, the Washington Women's Health Diary (Woods, 1987); & MSSL, the Menstrual Symptom Severity List (Mitchell *et al.*, 1991); n/a, not applicable.

*+ for FPSs

For "valid" and "reliable":

+, the criterion has NOT been tested OR the criterion is NOT satisfied.

++, the criterion has been tested. However, the testing method was flawed, OR the test results were either unsatisfactory or inconsistent.

+++, the criterion has been tested with sound methodology and acceptable results, OR the criterion is satisfied.

For "sensitive to change":

+, there is no intervention study using the instrument OR there is no intervention study showing positive results brought about by the intervention using the instrument.

+++, there is at least one study (the outcome of which was measured by the instrument) showing positive intervention effect.

For "generalizable" (# of studies):

+, ≤3 studies on PD.

++, 4-10 studies on PD.

+++, >10 studies on PD.

For "generalizable" (age):

+, previous study samples included only adolescent OR only college age women.

++, previous study samples included both adolescent AND college age women, BUT no middle age women.

+++, previous study samples included women of different ages across the reproductive spectrum.

For "generalizable" (diverse culture/ethnicity):

+, the tool has been used only in one country/race/ethnicity.

++, the tool has been in two countries/races/ethnicities.

+++, the tool has been used in ≥ 3 countries/races/ethnicities.

For "multidimensional":

+, the tool only measures one dimension (e.g., symptom intensity/severity) of PD symptom experience.

+++, the tool measures more than one dimension of symptom experience (i.e., measures symptom intensity, duration, and frequency, etc.).