

## **Pain scoring in Endometriosis: Entry Criteria and Outcome Measures for Clinical Trials. Report from The Art and Science of Endometriosis Consensus Meeting.**

K. Vincent, S. Kennedy, P. Stratton.

### **Introduction**

The meeting was convened by the National Institutes of Health (NIH), in collaboration with the American Society for Reproductive Medicine (ASRM), with the aim of establishing entry criteria and outcome measures for use in international clinical trials in endometriosis with regards to pain symptoms. The potential benefits of standardizing outcome measures include enabling trial results to be compared more easily and the production of systematic reviews to facilitate evidence-based practice.

A panel of invited scientists and clinicians from the UK, USA and Italy gave presentations throughout the day, and there was adequate time at the end of each session for questions and discussion (Appendix 1). Responses to 13 key questions, raised by the moderators following the earlier presentations, were then sought from the panel and the audience. A preliminary document summarizing the discussions and recommendations from the meeting was then circulated to all attendees and speakers, who were asked in a questionnaire (using both directed questions and free text) to comment on the document.

The document has been revised taking these responses into consideration and is now being presented to professional bodies such as the ASRM and European Society for Human Reproduction and Embryology (ESHRE) for further comments and approval before a final document is produced and published.

### **Overview of the Presentations**

Over the course of the day, all agreed that standardized entry and outcome criteria would be of benefit to clinicians, academics, industry and patients. This would allow for comparisons across trials and different treatment. It was readily apparent that ‘...perhaps the most common error committed by clinical researchers is to dismiss existing scales too lightly, and embark on the development of a new instrument with an unjustifiably optimistic and naïve expectation that they can do better.’ (1). After summarizing the latest developments in basic science and clinical research, the difficulties inherent in correlating symptoms and surgical findings were discussed. The remainder of the meeting focused on drawing on standardized outcome measurements from existing clinical research, and expressly stating that we should not try to “reinvent the wheel”.

The Biberoglu and Behrman (B&B) scale (2) was presented in detail by David Olive because it has been so widely used in clinical studies. However, it became apparent and all agreed that this scale had considerable limitations. The B & B scale principally asks questions about function and quality of life and, thus is not a pain scale. It includes three symptoms, dysmenorrhea, dyspareunia and chronic pelvic pain, and two signs, pelvic tenderness and induration, each of which are graded on a scale from 0 to 3 (or 4), with

higher numbers indicating more severe symptoms. The score is inaccurately inflated when the highest score of 4 is given to women without symptoms such as those with amenorrhea or who are not sexually active.

A second problem is that there is no standard for how or by whom this scale will be administered. Lacking is consistency as to whether it will be completed by the patient by herself, administered by the physicians and other study staff who collect the patient's information or collected as the clinician's impression of the patient's symptoms. Also lacking is a standard for the symptoms which will include women in studies or indicate that a treatment has succeeded.

Studies have inconsistently reported the results of the scale with some reporting only the individual scores for the symptoms and signs (3); some reporting only the sum of the symptoms(4), and others reporting the sum of all the signs and symptoms (5) In summary, it appears not to have been used or administered in a consistent manner, and it has never been validated or shown to be reproducible.

Paolo Vercellini presented information regarding the association between lesions and pain symptoms as background in determining both entry criteria and outcome measures. He presented data from a metanalysis he published of 1054 consecutive patients undergoing first-line conservative or definitive surgery in his unit between 1996 and 2002 (6). He concluded that the ASRM classifications stages or endometriomas are not associated with pain severity nor is the ASRM classification stage predictive of post-operative results or symptom recurrence. This is in agreement with data from other groups around the world, such as Szendei and colleagues in Hungary (7) and Mahmood et al. in Scotland (8) to name just two of the many such studies. Instead pain appears to be associated with deeply infiltrating lesions, the distance between nerve fibers and implants, and the number of nerve fibers within lesions.

The neuroscience of pain and endometriosis was presented by Karen Berkley using her translational research of a rat animal model (9). She pointed out that endometriosis and its associated pelvic visceral and muscle pain often occur alongside other painful conditions in widely disparate body regions. Her animal and human data suggest that the ectopic growths are innervated and may produce algogenic agents peripherally which contribute to engaging the CNS in generating pain symptoms. Importantly, she demonstrated that the experience of pain is a CNS phenomenon which arises from the intercommunication or matrix of connections in the brain. Since the CNS has a great deal of plasticity, the pain mechanisms related to endometriosis likely include central hormonal modulation, central sensitization, and remote central sensitization. These mechanisms are not unique to endometriosis and likely apply to other types of pain.

An overview of the different questionnaires used to assess quality of life (QoL) related to endometriosis was given by Crispin Jenkinson. While many studies have shown that the SF-36 has high internal consistency and has been shown to be useful in other medical conditions, there has not been a good correlation between QoL and pain intensity and/or use of medications in women with endometriosis (10-12) . Similarly, the EuroQOL EQ-

5D has been used but has not been validated in women with endometriosis (11). The only QoL scale that has been validated for use in women with endometriosis is the Endometriosis Health Profile 30 (13-15)

Charles Cleeland presented how to assess endometriosis pain from the perspective of a clinical expert in pain. In order to treat any type of pain, the following should be assessed: the severity, quality, location, and temporal pattern, as well as how it interferes with activities; its response to prior treatments; the adverse effects of treatment and whether the pain is somatic, neuropathic or visceral. Since pain is a subjective state, some have wondered whether they can trust their patient's ratings. Studies suggest that patient ratings are reliable.

Dr. Cleeland also presented different types of pain scales. The verbal rating scale or visual analogue scale is a scale in which 0 corresponds to no symptoms and 10 the worst pain imaginable. The strengths of scales are that they are easy to administer, easy to score, sensitive to treatment effects and they correlate with other intensity measures. Their weaknesses include that they have limited response categories, assume equal intervals between adjectives, and are not appropriate for low literacy patients. The brief pain inventory which is comprised of 11 questions and a pain drawing (16) is quick to complete, uses 0 to 10 scales so it is easy to complete, measures both pain severity and interference and is very sensitive to effective treatment.

Dr. Cleeland presented the Initiative on Methods, Measurement and Pain Assessment in Clinical Trials (IMMPACT) recommendations for clinical outcomes in pain trials (Table 1) to include: pain is measured in 0-10 scales, physical functioning (Brief Pain Inventory), emotional functioning (Beck, POMS), participant ratings of improvement or satisfaction (7 or 9 point scale), symptoms and adverse events, patient disposition (who drops out and why), and responder analysis (i.e. those with mild pain or 30 % reduction) (17). He advised that clinical studies should carefully define responders, use 0 to 10 scales, pick time points for measurement that are relevant to the condition under study, use latency of pain relief as a secondary outcome, and use pain symptoms as eligibility criteria.

Data analysis and interpretation were presented by Charles Cao. He presented the importance of collecting and analyzing data systematically. Mary Lou Ballweg played a couple of videos which describe the issues and pain experienced by women with endometriosis.

It was felt that there is no reason why the IMMPACT recommendations should not be applied to endometriosis. There followed a debate about QoL measures, which may be valuable in a disease such as endometriosis, even though they have been shown not to be useful in cancer trials. How to interpret QoL data and what constituted a clinically meaningful effect (particularly whether it should be defined by patients and/or doctors) were also topics that were debated.

The following 13 questions were then discussed in the light of the presentations.

## **The 13 questions**

### **1. How do we define 'endometriosis' for clinical trials?**

The majority of participants felt it was important to resolve this question before making decisions about outcome measures. However, following a heated debate, there were still marked differences of opinion between clinicians and scientists and between participants from different parts of the world. The definitions suggested were:

- Clinical picture: symptoms suggestive of endometriosis
- Surgical diagnosis: endometriosis seen at laparoscopy/laparotomy
- Histological diagnosis
- Response to treatment: pain relief with GnRH agonist

Although it was recognised that the guideline would need to be patient centered, it was still considered necessary for clinical trials in endometriosis to include only patients with a surgical diagnosis rather than those with a presumed diagnosis on the basis of either symptoms or a response to GnRH agonist treatment. However, histological confirmation is not required, which is consistent with US (18) and European (19) Guidelines. The recommendations could be adapted for pelvic pain trials in general (as was felt to be appropriate by a number of attendees), in which case entry criteria could be purely clinical.

Thereafter, there was considerable debate about how recently the surgery should have been performed. This proved problematic for two reasons: firstly, the natural history of endometriosis is uncertain. Secondly, it might be considered unethical within the context of a trial for medical therapy not to ablate or remove lesions found at an initial laparoscopy given the current recommendation (11).

To progress the meeting, it was decided that a surgical diagnosis of endometriosis within the last 12 months would be sufficient for the purposes of recruiting into clinical trials. However, a significant proportion of those who responded to the first document felt this time frame to be too narrow and a surgical diagnosis within the last 5 years was considered to be more realistic. This would also mean that a greater proportion of patients admitted to trials would have a return of symptoms rather than having symptoms persisting after surgery (i.e. failed surgical treatment).

### **2. What are the appropriate entry criteria?**

The entry and exclusion criteria need to be clearly defined in the trial protocol; however, it was agreed that the only essential entry criteria were a history of endometriosis-associated pain and surgical confirmation of the disease in the chosen time-period. These allow a wide variety of treatments to be tested on different stages of disease.

### **3. What are the baseline pain measurements?**

Dysmenorrhea and pelvic pain should be measured separately. An 11-point numerical rating scale (NRS) should be used, as agreed by consensus groups in other chronic pain conditions (20). These scales should be anchored by 0 = ‘no pain’ and 10 = ‘worst pain you can imagine’. Although it can be useful in laboratory pain studies to measure pain intensity and unpleasantness as separate entities, it may not be appropriate to do so in clinical trials as a single measure for each type of pain may be sufficient. Daily pain and the amount of vaginal bleeding should be recorded for at least one calendar month before treatment to ensure adequate baseline measurements are obtained.

#### **4. Should we:**

- a) persist with the Biberoglu and Behrman (B&B) scale?**
- b) adapt existing pain scale(s)?**
- c) develop a new (patient-derived) pain scale?**

The limitations and problems encountered in earlier trials using the B&B and its modified versions were discussed (2). Consensus was reached that the B&B should be rejected as a primary endpoint. However, it was felt that it should be retained (albeit in a standardised format) as a secondary endpoint in phase II/III trials for the time being.

There was considerable debate as to whether existing pain scale(s) should be adapted or a new (patient-derived) scale developed. As mentioned previously, caution was expressed earlier in the day about “reinventing the wheel”. It was stressed that many measures are already available which are validated and reproducible, and that developing a new scale takes a significant amount of time and resources. One group, however, who are already in the process of designing a new tool, argued convincingly for its utility although it is not yet in the public domain. Until the new tool can be assessed more widely, validated and shown to be sensitive to change, it seems sensible to adapt the IMMPACT Guidelines (17, 20) in the ways discussed below, because of their applicability to other chronic pain conditions.

#### **5. Should we take clinical signs into account or not?**

Consensus was easily reached that clinical signs should not be used as outcome measures.

#### **6. What about co-morbidity?**

In this context, it was agreed that co-morbidity meant other symptoms associated with endometriosis, rather than separate disease entities such as diabetes. It was felt that these could be assessed separately in a Quality of Life (QoL) tool; therefore, dysmenorrhea and pelvic pain should remain the two co-primary endpoints in every clinical trial. However, to understand the full impact of these symptoms, it was suggested that a woman could choose two or three which bother her the most and regularly report them as tertiary endpoints. This could usefully be adapted, for example for surgical trials involving resection

of rectovaginal disease when dyschezia and bowel function are particularly relevant, when the tertiary endpoints could also be trial-defined.

### **7. Should we use a single (summed) or separate pain assessments?**

It was agreed that separate pain assessments for dysmenorrhea and pelvic pain were superior to a single pain measure, principally because many treatments induce amenorrhea which makes a summed score unrepresentative of the true extent of improvement. There was marked disagreement, however, regarding how best to define dysmenorrhea. The simple definition “pain associated with menstrual bleeding” may not be sufficient. For example, it was not possible to agree whether pain associated with vaginal spotting at irregular times in the menstrual cycle constitutes dysmenorrhea or not. However, representatives from patient groups felt that most sufferers can tell when their bleeding is ‘menstrual’ and label this as dysmenorrhea as opposed to pelvic pain which might happen to be associated with some spotting. It was therefore agreed to rely upon the simple definition and the woman’s subjective impression of what menstrual means. The use of a daily diary, which records both pain scores and bleeding, should provide additional support for this approach.

### **8. Should we measure Quality of Life (QoL) as well? What about adverse events?**

The importance of measuring QoL was emphasized as well as the value of using a patient oriented, disease specific, measure with multiple domains, e.g. the Endometriosis Health Profile (EHP-30) (15). It was suggested that the scores from each separate domain should be reported in clinical trials rather than a single overall score, i.e. the sum of all the domains. However, it was pointed out that the value of using QoL tools in cancer trials had been questioned: to justify their use as outcome measures, there needs to be evidence that they can detect a change (in either direction) after treatment. Nevertheless, it was noted that measures of physical and emotional functioning, which appear in the EHP-30 are, at present, included in the IMMPACT recommendations.

Other reasons to use a QoL measure are that surgery can have a profound effect on body image and hormonal treatments can affect emotional functioning; in addition, the disease can impact on sexual function and feelings rating to infertility. The EHP-30 appears to meet all requirements and should therefore be included with reporting of domain scores.

It was suggested that a tool, such as the Patient Global Impression of Change (PGIC) (21), could be used for adverse event reporting. Caution was expressed about relying on this as it is no longer in common use (though still a part of IMMPACT) and it therefore does not form a part of these recommendations.

### **9. How often should we measure?**

It was felt that pain and bleeding should be reported on daily. It was believed that patients prefer to do this in the evenings, as mornings tend to be too busy. Suggestions

were made about the use of electronic diaries or web based scoring systems, but there were concerns about the applicability of these methods to international studies.

There was discussion about the frequency and time intervals for the measurement of secondary endpoints such as EHP-30 scores. However, it was acknowledged that transient side effects and/or early symptomatic improvements might be missed if data collection was not well timed. Therefore, it is suggested that these are measured weekly for the first 6 weeks, then monthly until 6 months and then at 9, 12, 18 and 24 months. Longer term data (e.g. up to 5 years post treatment) would be valuable and should be collected where possible, but may be difficult to obtain.

As many of the possible tertiary endpoints may have a cyclical component these will need to be scored daily in the same way as dysmenorrhoea and pelvic pain. There will need to be a 'Not Applicable' box as well as a scale for symptoms such as dyspareunia and dyschezia as a score of 0 could otherwise misleadingly suggest an improvement.

#### **10. How do we address cyclicity?**

The daily collection of information on both pain and bleeding was felt to be sufficient to capture cyclicity.

#### **11. Which rescue medications should be allowed/recorded and how?**

It was agreed that data on the use of rescue analgesia and complementary therapies should be captured. It was suggested that the pain score immediately before the use of such treatments should be recorded, as well as the indication, i.e. were they being taken for endometriosis-associated pain or an unrelated symptom? This would also allow information on adverse events to be captured if, for example, the analgesia was required for a headache that was secondary to hormonal treatment. A longitudinal analysis of rescue medication should also give an idea of the timescale required for the treatment to have an effect. Restricting the use of rescue medication would be both unethical and likely to increase the number of drop-outs from the trials.

#### **12. How do we define a responder?**

A clear definition of a responder should be provided in each trial. It is suggested that this be either a >30% or >50% reduction in symptoms; however, the precise definition will depend on the trial.

#### **13. What is a clinically meaningful effect?**

It was felt that the definition of a clinically meaningful effect should be patient-determined.

## **Specific recommendations:**

### **Entry Criteria:**

- Surgical diagnosis of endometriosis within the last 5 years
- Pain symptoms

### **Data capture at baseline:**

- ASRM classification (22)
- Baseline pain scores over at least 2 menstrual cycles
- EHP-30 (15)
- Previous treatments and responses

### **Primary Outcome Measures:**

- Daily ratings of pelvic pain
- Daily ratings of dysmenorrhea

Ratings on an 11 point NRS, anchored by 0 = ‘no pain’ and 10 = ‘worst pain you can imagine’ based on a recall of the worst pain experienced over the previous 24 hours. Daily record of bleeding as ‘none, spotting, light or heavy’ compared to a normal period.

### **Secondary Outcome Measures:**

- B&B (2) with separate scores for each domain  
Administered weekly for 6 weeks, then monthly until 6 months, then at 9, 12, 18 and 24 months
- EHP 30 (15) with separate and total scores  
Administered at the same time points as the B&B
- Use of rescue analgesia/therapies  
including an NRS before use and a record of the indication
- Study specific adverse event questionnaires with direct questions and free text  
Administered at the same time points as the B&B
- Detailed information as per the CONSORT Guidelines(23), including:
  1. The recruitment process.
  2. The number of candidate participants who were excluded and why.
  3. The number of candidates who chose not to enter the trial and why.
  4. The use of prohibited concomitant medications and other protocol deviations.
  5. The number and reasons for withdrawal from each treatment group.
  6. The types, rates and reasons for non-adherence with treatment in each group.

### **Tertiary Outcome Measures:**

- Daily NRS (or NA) of three symptoms the patient feels are important to her  
e.g. dyspareunia, dyschezia, fatigue etc.

**Table 1: Summary of IMMPACT recommendations (adapted from (20))**

| <b>Core Outcome Measures</b>  | <b>Tool</b>   |
|---|---|
| Pain  | 11-point (0-10) numerical rating scale of pain intensity<br>Usage of rescue analgesics<br>Categorical rating of pain intensity in circumstances in which numerical ratings may be problematic |
| Physical Functioning (Either measure)                                     | Multidimensional Pain Inventory Interference Scale<br>Brief Pain Inventory interference items   |
| Emotional Functioning (At least one measure)                              | Beck Depression Inventory<br>Profile of Mood States   |
| Participant ratings of global improvement and satisfaction with treatment | Patient Global Impression of Change   |
| Symptoms and Adverse events   | Passive capture of spontaneously reported adverse events and symptoms and use of open-ended prompts   |
| Participant disposition   | Detailed information regarding participant recruitment and progress through the trial, including all information specified in the CONSORT Guidelines  |

**Table 2: Examples of co-morbidities to be considered as tertiary endpoints**

|                           |
|---------------------------|
| Dyspareunia               |
| Bladder pain              |
| Dyschezia                 |
| Fatigue                   |
| Abnormal vaginal bleeding |

## References:

1. Streiner DL, Norman G. Health Measurement Scales - a practical guide to their development and use. 3rd ed: Oxford University Press, 2003.
2. Biberoglu KO, Behrman SJ. Dosage aspects of danazol therapy in endometriosis: short term and long term effectiveness. American Journal of Obstetrics and Gynecology 1981;139:645.
3. Lemay A, Maheux R, Hout C, Blanchet J, Faure N. Efficacy of intranasal or subcutaneous luteinizing hormone-releasing hormone agonist inhibition of ovarian function in the treatment of endometriosis. American Journal of Obstetrics and Gynecology 1988;158:233 - 6.
4. Hornstein MD, Yuzpe AA, Burry KA, Heinrichs LR, Buttram-VL J, Orwoll ES. Prospective randomized double-blind trial of 3 versus 6 months of nafarelin therapy for endometriosis associated pelvic pain. Fertility and Sterility 1995;63:955 -62.
5. Shaw RW. An open randomized comparative study of the effect of goserelin depot and danazol in the treatment of endometriosis. . Fertility and Sterility 1992;58:265 - 72.
6. Vercellini P, Fedele L, Aimi G, Pietropaolo G, Consonni D, Crosignani PG. Association between endometriosis stage, lesion type, patient characteristics and severity of pelvic pain symptoms: a multivariate analysis of over 1000 patients. Hum Reprod 2007;22:266-71.
7. Szendei G, Hernadi Z, Devenyi N, Csapo Z. Is there any correlation between stages of endometriosis and severity of chronic pelvic pain? Possibilities of Treatment. Gynecological Endocrinology 2005;21:93 - 100.
8. Mahmood T, Templeton A, Thomson L, Fraser C. Menstrual symptoms in women with pelvic endometriosis. British Journal of Obstetrics and Gynaecology 1991;98:558 - 63.
9. Berkley KJ, Rapkin AJ, Papka RE. The pains of endometriosis. Science 2005;308:1587-9.
10. Bodner CH, Garratt AM, Ratcliffe J, Macdonald LM, Penney G. Measuring health-related quality of life outcomes in women with endometriosis-results of the Gynaecology Audit Project in Scotland. Health Bulletin 1997;55:109 - 17.
11. Abbott JA, Hawe J, Clayton RD, Garry R. The effects and effectiveness of laparoscopic excision of endometriosis: a prospective study with 2-5 year follow-up. Human Reproduction 2003;18:1922 - 7.
12. Marques TEA, Bahamondes L, Aldrighi JM, Petta CA. Quality of life in Brazilian women with endometriosis assessed through a medical outcome questionnaire. The Journal of Reproductive Medicine 2004;49:115 - 20.
13. Jones G, Jenkinson C, Kennedy S. The impact of endometriosis upon quality of life: a qualitative analysis. Journal of psychosomatic obstetrics and gynaecology 2004;25:123-33.
14. Jones G, Jenkinson C, Kennedy S. Development of the Short Form Endometriosis Health Profile Questionnaire: the EHP-5. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2004;13:695-704.

15. Jones G, Kennedy S, Barnard A, Wong J, Jenkinson C. Development of an Endometriosis Quality-of-Life Instrument: The Endometriosis Health Profile-30. *Obstetrics and Gynecology* 2001;98:258-64.
16. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Annals of the Academy of Medicine, Singapore* 1994;23:129-38.
17. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB *et al.* Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2003;106:337-45.
18. ASRM. Endometriosis and infertility. *Fertility and Sterility* 2004;81:1441-6.
19. Kennedy S, Bergqvist A, Chapron C, D'Hooghe T, Dunselman G, Greb R *et al.* ESHRE guideline for the diagnosis and treatment of endometriosis. *Human Reproduction* 2005;20:2698-704.
20. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP *et al.* Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005;113:9-19.
21. Guy W. ECDEU assessment manual for psychopharmacology. Washington, DC: US Government Printing Office, 1976.
22. ASRM. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Fertility and Sterility* 1997;67:817-21.
23. Moher D, Schulz KF, Altman DG, for the CG. The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *Ann Intern Med* 2001;134:657-62.

## **Appendix 1: Programme for the day**

08.00 Pamela Stratton (NIH)  
Welcome

### **Pamela Stratton - Moderator**

08.15 Stephen Kennedy (University of Oxford)  
*What is the problem?*

08.30 Karen Berkley (Florida State University)  
*Fundamental mechanisms of pain and multiple system effects in endometriosis*

09.00 David Olive (University of Wisconsin-Madison)  
*Pain assessment in clinical trials – a review of the literature*

09.30 Paolo Vercellini (University of Milan)  
*Correlation between lesions and pain – a review of the literature*

10.00 Panel discussion

10.30 Refreshments

### **Stephen Kennedy - Moderator**

11.00 Crispin Jenkinson (University of Oxford)  
*Quality of life measurement*

11.30 MaryLou Ballweg (Endometriosis Association)  
*What's important to patients*

12.00 Panel discussion

12.30 Lunch

### **Stephen Kennedy and Pamela Stratton - Moderators**

13.30 Charles Cleeland (MD Anderson Cancer Center)  
*Pain measurement*

14.00 Charles Cao (Takeda Pharmaceuticals)  
*Data analysis and interpretation*

14.30 Panel discussion

15.00 Recommendations

16.30 Conclusions