Clinical Study Synopsis

This Clinical Study Synopsis is provided for patients and healthcare professionals to increase the transparency of Bayer's clinical research. This document is not intended to replace the advice of a healthcare professional and should not be considered as a recommendation. Patients should always seek medical advice before making any decisions on their treatment. Healthcare Professionals should always refer to the specific labelling information approved for the patient's country or region. Data in this document or on the related website should not be considered as prescribing advice.
The study listed may include approved and non-approved formulations or treatment regimens. Data may differ from published or presented data and are a reflection of the limited information provided here. The results from a single trial need to be considered in the context of the totality of the available clinical research results for a drug. The results from a single study may not reflect the overall results for a drug.

2. Synopsis

Date of report:	9th October, 2013
Study title:	Validation Study to assess psychometric properties of a newly developed Patient Reported outcome tool for Endometriosis (Short title: **Val**idation Study for **E**ndometriosis **PRO** (VALEPRO))
Sponsor's study number:	15849
NCT number:	NCT01643122
EudraCT number:	Not applicable
Sponsor:	Bayer HealthCare
Clinical phase:	Observational
Study objectives:	To assess the content validity of the Endometriosis Symptom Diary (ESD) and determine the most relevant items for measuring pelvic pain, dysmenorrhea, dyspareunia and analgesic use To assess the content validity of the Endometriosis Impact Scale (EIS) and determine the most relevant concepts for measuring impact of endometriosis pain To evaluate the psychometric properties of the ESD and the EIS To determine the validity of a 7-day recall period for the EIS
Test drug:	Not applicable
Reference drug:	Not applicable
Indication:	Endometriosis

Diagnosis and main criteria for inclusion:	The inclusion criteria for subjects to be eligible for inclusion in the study are listed below: • women aged between 18 and 45 years of age, inclusive • women with endometriosis confirmed by laparoscopy or laparotomy within five years before the baseline visit • good general health (except findings related to endometriosis) as proven by medical history, • patient has experienced endometriosis symptoms (i.e. pain) in the past 4 weeks, as assessed by the numerical rating scale
Study design:	Prospective observational study conducted in the US and Germany in women with surgically confirmed endometriosis

Methodology	ESD, a newly developed ePRO instrument with a 24-hour recall assessing patient experience of endometriosis symptoms (measure being validated). EIS, a newly developed ePRO instrument with a 7-day recall assessing impact of endometriosis pain (measure being validated). Endometriosis Daily Impact Diary (EDID), a daily diary with a 24-hour recall assessing impact of endometriosis pain on physical functioning. The EDID is made up of selected items of the EIS that have been modified to use a 24-hour recall and will be used to explore the validity of the 7-day recall for the EIS. Biberoglu and Behrman Scale (B&B), an instrument assessing pelvic pain, dysmenorrhea, and dyspareunia. The B&B will be used to explore convergent validity of ESD and EIS. Endometriosis Health Profile-30 (EHP-30), a disease-specific health-related quality of life instrument. The EHP-30 will be used to explore convergent validity of the ESD and EIS. Numeric Rating Scale (NRS) for pelvic pain with a 4-week recall. The NRS will be used to describe the clinical severity of the study population, to group patients for known-group comparisons, and to identify stable/unstable patients for the purpose of exploring responsiveness of the ESD and EIS. Visual Analogue Scale (VAS) for pelvic pain with a 4-week recall. The VAS will be used to describe the clinical severity of the study population, to group patients for known-group comparisons, and to identify stable/unstable patients for the purpose of exploring responsiveness of the ESD and EIS. European Quality of Life-5 Dimensions (EQ-5D), a generic preference-based measure of health status. The EQ-5D will be used to describe the study population. Short Form-36 Questionnaire (SF-36v2), a generic measure of health-related quality of life and well-being. The SF-36v2 will be used to describe the study population. Patient Global Impression of Severity (PGI-S). The PGI-S will be used to describe the clinical severity of the study population, to group patients for known-group comparisons, and to identify stable/unstable patients for the purpose of exploring responsiveness of the ESD and EIS. Patient Global Impression of Change (PGI-C). The PGI-C will be used to identify stable/unstable patients for the purpose of exploring responsiveness of the ESD and EIS.

	Clinical Global Impression of Severity (CGI-S). The CGI-S will be used to describe the clinical severity of the study population, to group patients for known-group comparisons, and to identify stable/unstable patients for the purpose of exploring responsiveness of the ESD and EIS. Clinical Global Impression of Change (CGI-C). The CGI-C will be used to identify stable/unstable patients for the purpose of exploring responsiveness of the ESD and EIS.
Study center(s):	32 sites in two countries: Germany (15), USA(17)
Publication(s) based on the study (references):	None at the time of report creation
Study period:	First subject, first visit: 31 Aug 2012 Last subject, last visit: 29 Jul 2013 Study Completion Date: 23 Sep 2013
Early termination	Not applicable
Number of subjects:	Planned: 250 patients Analyzed: 268 patients (USA: n=133; Germany: n=135)
Criteria for evaluation	
Efficacy / clinical pharmacology:	Not applicable
Safety:	AE reporting was performed by the investigator according to local legislation
Other:	Endometriosis Symptom Diary (ESD) evaluation, Endometriosis Impact Scale (EIS) evaluation, Assessment of 7-day Recall Period for the EIS

Statistical methods:	1) Descriptive statistics • Demographic and clinical characteristic of the study population as well as supplementary PRO measures 2) Correlation analysis to explore the validity of the 7-day recall period of the EIS • Comparison of response to the EIS (physical activities items, a 7-day recall) with the equivalent EDID items 3) Analysis of item performance, item reduction, and scaling of the ESD and EIS • Descriptive item statistics, floor and ceiling effects, item-to-item correlations, confirmatory factor analysis (CFA), exploratory factor analysis (EFA), Rasch model analysis, and differential item functioning (DIF)
Substantial protocol changes:	Protocol Version 2.0, dated 11 JUL 2012 introduced the following changes: Newly developed Endometriosis Daily Impact Diary (EDID) was added

Study subjects:

Two hundred and sixty eight (268) women from 18 and 45 years of age (mean 31.6 years, SD 6.7 years, median 31.0 years) with endometriosis confirmed by laparoscopy or laparotomy within five years before the baseline visit were enrolled in the study.

Efficacy / clinical pharmacology evaluation

Not applicable

Safety evaluation

AEs were to be reported by investigator only if requested by local legislation. ***[Other evaluations]***

<u>Endometriosis Symptom Diary (ESD) evaluation</u>

Assessment of pelvic pain: The ESD v4.0 comprises three items designed to assess pelvic pain severity: a global item assessing ‘worst pain’ in the past 24 hours (ESD item 1) and two items assessing ‘worst constant’ and ‘worst short-term’ pain in the past 24 hours (ESD items 2 and 3, respectively). The three items had similar response distributions at Day 8. All response options on the 0-10 Numerical Rating Scale (NRS) were endorsed indicating that the response options are valid. Observation of inter-item correlations at day 8 revealed that all three items correlate highly with one another, exceeding pre-specified cut-offs typically used to indicate statistical redundancy ($r \geq 0.80$). As would be expected, observed inter-item

correlations were highest (r≥0.90) between global assessments of 'worst pain (ESD item 1) and assessments of 'constant pain' and 'short-term pain' (ESD items 2 and 3, respectively) indicated consistency between responses to these items by women in this study sample at this time-point. Inter-item correlations between item 2 (Constant pain) and item 3 (Short-term pain) were r=0.80, indicating that these two items may measure similar but not completely equivalent concepts. Consideration of explained variance (r2) of item 2 by item 3 at day 8 was 0.64, highlighting overlap between participant responses to the two items but suggesting that they are not measures of equivalent concepts. The variable map produced for the exploratory Rasch analysis demonstrated that there were no significant gaps in the item hierarchy, indicating that each of the three items can adequately cover the severity continuum individually and also provides assurances that no further items are needed to characterize the pain experience among women in this population. The rating scale structure showed no evidence of step threshold disordering for all three items which again supports the adequacy of the 0-10 NRS as a valid means of assessing pain severity in this population. Similarly, the observation that the response categories cover the full range of person severities suggests that the three items are able to differentiate between patients across a range of symptom severities. Finally, comprehensive investigations revealed no evidence of DIF by participant age, country and level of education, further supporting the validity of responses options for these items.

Findings from the VALEPRO study support the content validity of the ESD items assessing pelvic pain severity. While evidence indicated high correlation between scores for the three items assessing pelvic pain in the VALEPRO study population at day 8, insights from prior qualitative research and consideration by a panel of clinical experts support the retention of all three items in as part of the ESD. Because of the high correlation between the scores of the three items, a total score calculated based on a combination of these items is not proposed, rather these concepts are to be treated as separate individual endpoints.

Assessment of pain due to menstrual bleeding: The ESD v4.0 comprised a number of items designed to facilitate the assessment of pain due to menstrual bleeding. These include two items designed to measure the presence of menstrual bleeding in the form of respondent-reported presence of a menstrual period (ESD item 4) and respondent-reported menstrual bleeding severity (ESD item 5) and a further item assessing 'worst pain' that respondents attributed to their period (ESD item 6). Evidence from the VALEPRO study and subsequent insights from exert clinicians supports prior findings from qualitative research regarding ambiguities surrounding the definition of menstrual pain and non-menstrual pain. Therefore, it is proposed that assessment of pain due to menstrual bleeding by the ESD not rely upon respondent attribution but instead be defined according to concurrent bleeding, as assessed by a daily assessment of menstrual bleeding severity (e.g. ESD item 5). In accordance, ESD item 4 (self-reported presence of a menstrual period) and ESD item 6 ('worst pain' that respondents attributed to their period) are considered redundant and are recommended for removal. Comparison of number of bleeding days and average pain on bleeding days revealed similar results when a simple criterion of days of ≥ light bleeding on ESD item 5 and a more complex criterion of days within a bleeding episode (defined as day(s) of bleeding (of which at least one day is ≥ light bleeding) preceded and followed by at least 2 consecutive days of no bleeding or 'spotting') were used. In the absence of evidence for the additional benefit of the more complex criterion, it is recommended that pain on bleeding days be determined by use of the simpler criterion. Following consensus opinion among expert clinicians it is

proposed that pain due to menstrual bleeding be defined as responses to ESD item 1 ('worst pain') on days of ≥ light bleeding (as assessed by ESD item 5).

Assessment of dyspareunia and avoidance of sexual intercourse: The ESD v4.0 includes three items designed to assess whether the respondent has had sexual intercourse during the preceding 24 hours (ESD item 10), the worst pain experienced either during or after sexual intercourse (ESD item 11) and whether the respondent avoided sexual intercourse due to pain or expected pain (ESD item 12). All three items performed satisfactorily (in term of response distribution) during the VALEPRO study. Together with qualitative findings and expert clinician opinion, these data support the inclusion of these items to assess the two concepts of dyspareunia and avoidance of sexual intercourse.

Endometriosis Impact Scale (EIS) evaluation

EIS v3.0 comprised 32 draft items with six hypothesized domains describing impact on physical activities, emotions, social and leisure activities, paid work and study, household activities, sexual activities and three stand-alone items assessing the impact on ability to concentrate, eat, and sleep. The performance of EIS items in the VALEPRO study, together with qualitative findings and expert clinician inputs, suggested that following items could be removed from EIS v3.0:

- Item 22 and 23 (social/ leisure activities), 24 and 25 (work/ study), 26 and 27 (household activities) due to a high level of overlap (high inter-item correlations) with proximal physical activities items and their global items assessing these concepts. In addition, feedback from expert clinicians indicated that the global items were sufficient to assess these concepts
- Item 31 (difficulty eating) due to a high floor effect evident at day 8 in the VALEPRO study (55%) which indicated that the concept was not relevant to the majority of patients. Qualitative findings and input from expert clinicians supported the deletion of this item

The physical activities and emotional well-being domains were clearly defined, and further analyses were performed to ascertain which items should be retained in these sub-domains.

Physical Activities Domain: The item response distributions of the eight items demonstrated that all response options were endorsed by patients at this single time point, thus supporting the validity of the response options developed for the scale. Findings relating to the statistical performance of the EIS physical activity items in the VALEPRO study indicated that several items may be considered for deletion. However, qualitative findings and input from expert clinicians were carefully considered before making the following decisions:

- Item 11 (Lift) was deleted due to evidence of redundancy with item 12 (Carry). Qualitative findings and clinical input indicated that item 11 (Lift) was less clinically relevant than item 12 (Carry).
- Whilst it is acknowledged that there is some evidence of redundancy between item 8 (Walk) and item 9 (Stand), the qualitative findings and clinical input suggested that both items assess clinically relevant concepts and should be retained.

- Item 10 (Sit) demonstrated misfit to the Rasch model, but as this was only marginal and qualitative findings and clinical input indicated that this item was clinically relevant, it is recommended that the item should be retained.

A secondary Rasch analysis which was undertaken on the seven retained items demonstrated that the removal of item 11 (Lift) enhanced the performance of the scale. No significant positive item residuals as well as acceptable item and person separation and reliability were found. The removal of item 11 (Lift) improved the fit of the remaining items with all items demonstrating acceptable fit to the Rasch model. The validity of items and their response options are supported by their alignment with the study sample in terms of severity and the fact that there are no significant gaps in the measurement continuum. Category functioning also showed that there was no evidence of category disordering and sufficient differentiation of subject endorsement at each step, thus supporting the validity of the response options for the items. Support for the internal reliability of the physical activities scale is provided by the very high Cronbach's α (α = 0.96) which was observed on day 8 of the VALEPRO study.

Emotional-wellbeing Domain: The item response distributions of the nine items demonstrated that all response options were endorsed by patients at this single time point, thus supporting the validity of the response options developed for the scale. Findings relating to the statistical performance of the EIS emotional-wellbeing items in the VALEPRO study indicated that several items may be considered for deletion. However, qualitative findings and input from expert clinicians were carefully considered before making the following decisions:

- Item 15 (Depressed) was deleted due to evidence of redundancy with item 14 (Sad). Qualitative findings and clinical input indicated that item 15 (Depressed) was less clinically relevant than item 14 (Sad).
- Item 19 (Annoyed) was deleted due to evidence of redundancy with item 16 (Irritable) and item 17 (Bad mood). Qualitative findings and clinical input indicated that item 19 (Annoyed) was less clinically relevant than items 16 (Irritable) and 17 (Bad mood).
- Item 20 (Anxious) demonstrated misfit to the Rasch model, and also had a high floor effect (43.16%). However, qualitative findings from concept elicitation and cognitive interviews indicate that this concept is highly relevant to many patients. Expert clinicians supported the clinical relevance of this item and recommended that it be retained as it is distinct from all other concepts included in the domain, and an important aspect of emotional wellbeing.

A secondary Rasch analysis which was undertaken on the seven retained items demonstrated that the removal of the item 15 (Depressed) and item 19 (Annoyed) improved the fit to the Rasch model. There was no evidence of significant local dependence and acceptable item and person separation and reliability. Item 20 (Anxious) still demonstrated misfit to the Rasch model, although the misfit was less pronounced with infit and outfit mean square values of 1.77 and 1.61 respectively. The validity of the 7-item scale is supported by the variable map which shows that the items closely align with the persons, with no significant gaps in the severity measurement continuum. Finally, the response options of the items are supported by category functioning which showed that there was no evidence of category disordering and sufficient differentiation of subject endorsement at each step. The internal consistency reliability of the revised 7-item scale is supported by the high Cronbach's α observed

(α=0.95). Although it is acknowledged that the removal of item 20 (Anxious) would improve Cronbach's α marginally (from 0.951 to 0.952), this was not considered sufficient a change to warrant the removal of this item, particularly given the relevance of this concept to patients.

Assessment of 7-day Recall Period for the EIS

Findings relating to the comparisons between EIS (7-day recall) scores and EDID (24-hour recall) scores support the validity of the 7-day recall period of the EIS.

Overall conclusions

Analyses conducted using data collected via the VALEPRO study support the content validity of a shortened 10-item version of the ESD (version 5.0) for the assessment of pain (both due to menstrual bleeding and independent of menstrual bleeding), dyspareunia and avoidance of sexual intercourse.

Evaluation of EIS items revealed a degree of redundancy in item content, particularly for items assessing time lost and reduced ability to complete work/study, household activities and social leisure activities; with evidence suggesting that these concepts could adequately be captured by global items assessing impact on these aspects of daily life. Analyses conducted provided strong support for the validity of the EIS physical activities and emotional wellbeing domains and the items comprising these domains. Evidence of redundancy for some items of the EIS, however, has resulted in a shortened scale comprising 22 items (version 4.0).

Following the confirmation of content validity and determination of scoring algorithms specified in this report, further psychometric testing is planned to evaluate reliability, validity, and responsiveness of the ESD and EIS.