PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Acupuncture Point InjectionTreatment of Primary Dysmenorrhea: A randomised, double blind, controlled study
AUTHORS	Wade, Christine; Wang, Li; Zhao, Wen; Cardini, Francesco; Kronenberg, Fredi; Gui, Sui; Ying, Zhi; Zhao, Nai; Chao, Maria; Yu, Jin

VERSION 1 - REVIEW

REVIEWER	Pallavi Latthe Birmingham Women's NHS Foundation Trust UK
REVIEW RETURNED	07-May-2015

GENERAL COMMENTS	references 4 and 5 are superseded by more uptodate references (october 2014 and Jan 2010); please update them It would be useful to have a schematic diagram of the points of injections on the legs if possible.

REVIEWER	Liu Cun-Zhi Beijing TCM Hospital affiliated with Capital Medical University,China.
REVIEW RETURNED	11-May-2015

GENERAL COMMENTS	1. There are many problems in the writing, and you should seek a professional to help you modify the paper. 2. The design of control group is not appropriate, it can't answer the hypothesis accurately. 3. Abstract, Results, Exclusion criteria does not belong to results. 4. Participants randomized to group B received a vitamin K3 injection in the right buttock and saline in a non-acupuncture point near but not SP6. How did you locate this non-acupuncture point?

REVIEWER	Mark W Strudwick University of Queensland Australia
REVIEW RETURNED	15-May-2015

GENERAL COMMENTS	This work is based on a clinical protocol, already in use, according to this report. Since the clinical results are anecdotal, the research design is both important and relevant. Overall, the research as presented appears to fulfil these criteria. Some areas are not clear, however. 1. Was the injection a "once only" treatment? If so, then the results

	would seem appropriate as any effect of a single instance treatment would dissipate with time. 2. Was the treatment applied on each subsequent visit after the initial injection? If so, then the results may indicate habituation or attenuation. I have some reservations about the results. In Table 3, the standard deviations appear quite high for treatment cycle 2. This causes some doubt as to the validity of these measurements. Are these the best analytic tools to use? How well have they been validated? And how often are they cited in the literature? I cannot easily find use of the Cox or MMDQ. The use of vitamin K has not be fully explained, especially in light of the statement that K1 has a better safety profile than K3. While I understand this research is investigating the efficacy of a clinical protocol that has been in use for some time, I should like to see why vit K was chosen rather than sterile water or a local anaesthetic - both of which I would assess more appropriate in this instance.

REVIEWER	Andrew Hinde Southampton Statistical Sciences Research Institute University of Southampton United Kingdom
REVIEW RETURNED	09-Jul-2015

GENERAL COMMENTS	This paper describes a well-designed (and clearly quite demanding) study of the effect of vitamin K3 injections on primary dysmenorrhea. The description of the study makes it clear that similar studies are unlikely to be common, so the results of this one should definitely be published. Before publication, however, there are one or two revisions needed, mainly to clarify the methods used. 1. I am a statistical reviewer, not an expert in the substantive topic, so please forgive my ignorance, but why was the saline injection to Group B given to a point near the SP6 acupuncture point but not right at SP6? Does this not potentially compromise this injection as a control? 2. Somewhat facetiously, can I ask whether any women were excluded because they did not have a working phone (p. 6, l. 20)? 3. (This is my main recommendation.) Can you add a couple of sentences describing your models in more detail? You refer to 'mixed effects' models, but this covers a pretty wide range. First, the way I would have approached the model is to note that observations are clustered within women, so I need a woman-level random effect to control for this. I guess (hope) this is how your reasoned but you do not say. Second, how do you accommodate time since treatment in your models? Are the results reported in Table 2 some kind of average of the effects at different times? Or did you include time since treatment as a covariate? Figure 2 suggests that there is a widening of the difference between Group C and Groups A and B over time, though this effect is not huge, and might not be statistically significant. Can youi add a sentence explaining how you dealt with time since treatment? Finally, Table III and the text are inconsistent. On p. 12, ll. 2-5 you

	state that '[a]ll groups had significant differences ... between the baseline cycle (Cycle 1) and the post treatment cycle (Cycle 2) on the Cox Duration and Cox Intensity scores.' But Table III indicates a significant difference for Group C but not for Groups A and B. Similarly, you write that '[a]ll groups also showed a significant difference ... between the baseline scores of menstrual distress and Cycle 7 scores at the six month follow up report.' But again Table III suggests that only Group C had such a difference. This should be corrected.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Reviewer Name Pallavi Latthe

1 References 4 and 5 are superseded by more uptodate references (october 2014 and Jan 2010); please update them.

The references have been updated.

2. It would be useful to have a schematic diagram of the points of injections on the legs if possible.

We have added an acupuncture point schema which shows SP6.

Reviewer: 2 Reviewer Name Liu Cun-Zhi
Beijing TCM Hospital affiliated with Capital Medical University,China.

1. There are many problems in the writing, and you should seek a professional to help you modify the paper.

The reviewer does not indicate the problems that he finds with the writing so it is difficult for the authors to address them.

2. The design of control group is not appropriate, it can't answer the hypothesis accurately.

We planned the study to control for both the site of injection and also the agent with two active control groups, maintaining comparability of maneuver between the groups by having all patients receive three injections. When we controlled for the injection site GROUP B and GROUP C received the same dose of Vitamin K3. When we controlled for the agent GROUP A received saline instead of Vitamin K3. We hypothesized that the treatment developed by the clinicians, and used successfully for many years at a high volume menstrual disorder clinic would have a more robust clinical effect than the two active control groups.

Group A Saline acupuncture point injection: Study participants randomized to this group received normal saline injection in the San Yin Jiao or Spleen 6 (SP6) acupuncture point (both legs) and saline in right buttock. This served as an active control for vitamin K. Inactive agent (saline) in the classical SP6 acupuncture point. The authors have read the literature on injection treatments for pain and understand that saline injection controls may have limitations.

Group B Vitamin K deep muscle injection: Study participants randomized to this group received a vitamin K3 injection in the right buttock and saline in a non-acupuncture point near but not SP6 (both legs). This served as an active control for the acupuncture point injection site.) Active agent(vitamin K) and inactive injection site (off point on the inside of the ankle.) The authors have read the literature on sham acupuncture controls and understand the limitations of using sham acupuncture, which is an

active not an inert treatment, such as might be proposed in “placebo controlled” trial. We have added a sentence on page 6 to clarify our intention with GROUP B.

The treatment that was hypothesized to be optimal and which had been used in a large urban menstrual disorder clinic for many years included an active agent, vitamin K3, in the classical SP6 acupuncture point.

3. Abstract, Results, Exclusion criteria does not belong to results.

Exclusion criteria have been moved to the “Participants” section in the abstract. The exclusion criteria are listed in the Materials and Methods section on pages 6 and 7.

4. Participants randomized to group B received a vitamin K3 injection in the right buttock and saline in a non-acupuncture point near but not SP6. How did you locate this non-acupuncture point?

The clinicians located the sham point at 1 cm above Sp6, 1 cm towards the back of the leg from Sp6, and penetrated the needle 1 cm beneath the skin.

Reviewer: 3 Reviewer Name Mark W Strudwick
Institution and Country University of Queensland Australia

1. Was the injection a "once only" treatment? If so, then the results would seem appropriate as any effect of a single instance treatment would dissipate with time.

See response to #2 below.

2. Was the treatment applied on each subsequent visit after the initial injection? If so, then the results may indicate habituation or attenuation.

We have clarified the text to indicate that it was a single treatment in both the abstract (page 2) and in three places in the main body of the text (page 5 ,6 and 7).

3. I have some reservations about the results. In Table 3, the standard deviations appear quite high for treatment cycle 2. This causes some doubt as to the validity of these measurements. Are these the best analytic tools to use? How well have they been validated? And how often are they cited in the literature? I cannot easily find use of the Cox or MMDQ.

Pain perception and report of symptoms of menstrual distress are highly variable, both amongst women and between cycles, so we expected that standard deviations could be high.

The Cox Retrospective Symptom Score and the Moos Menstrual Distress Questionnaire were the instruments most commonly used in the English language science literature for studies of menstrual disorders, at the time the study was planned in 2003. The authors administered both questionnaires with the idea of validating results from one to the other and determining the best instrument to used in future research.

The Moos Menstrual Distress Questionnaire (MMDQ) is a validated instrument which is appropriate

for measurement of the dysmenorrhea component of menstrual symptomatology. The questionnaire has been validated using the responses of over 800 English-speaking women with a mean age of 25 and education level of 15 years.(Moos, 1985) Original data were collected on both the most recent and the subjectively "worst" menstrual cycle. The questionnaire has been used in many studies of dysmenorrhea, and in translated versions, in many non-English speaking populations including Spanish-speaking (Moos, 1985) and more recently in Chinese for use in Hong Kong.(Chang, 1999) The full form of the MMDQ contains 47 symptoms which women are asked to rate on a five-point (0-5) scale; ratings can be done during the premenstrual, menstrual and intermenstrual phases of the cycle. The MMDQ includes a sub-scale for pain associated with menstruation. The pain sub-scale contains 6 items: muscle stiffness, headache, cramps, backache, fatigue, and general aches and pains which can be tallied in a score. One approach is to use these questions to assess (retrospectively) the participant's subjective assessment of the last menstrual period. Such an approach may be prone to recall bias. Retrospective scores have been compared, however, with scores collected for just the days before and after the onset of menstruation, and found that daily scores were comparable to retrospective scores.(Moos, 1985) We administered 2 pain scales (Cox Retrospective Sympton Scale (Cox RDD) and the MMDQ. The MMDQ is more commonly used than the Cox RSS, but we preferred the Cox RSS because it has two dimensions: pain intensity and pain duration. Data collected retrospectively has been validated against items collected daily.(Cox, 1978)

Moos, R. H. Perimenstural symptoms: A manual and overview of research with the menstrual distress questionnaire. 1985. Stanford, Social Ecology Laboratory, Stanford University School of Medicine.

4. The use of vitamin K has not be fully explained, especially in light of the statement that K1 has a better safety profile than K3. While I understand this research is investigating the efficacy of a clinical protocol that has been in use for some time, I should like to see why vit K was chosen rather than sterile water or a local anaesthetic - both of which I would assess more appropriate in this instance.

We agree with the reviewer that the use of vitamin K is not fully explained and
that K1 may be a better choice than K3 in future studies. The mechanism for vitamin K's influence on menstrual pain is not known so we cannot report it.

Treatment with K3 had, however, been used in the treatment in a high volume clinic for over 20 years without observed adverse clinical effects. A water-soluble preparation of vitamin K3 (menadione) is available for adults in the U.S. The safety data that privileges K1 over K3 is from studies in infants and prompted the American Academy of Pediatrics to recommend K1 over K3 in the routine administration of vitamin K in newborns to prevent hemorrhage.

As the authors did not know whether K1 would have different clinical effects than K3, and as K3 has a very reasonable safety profile, we decided to test the treatment as it had been standardized at the Menstrual Disorder Clinic, rather than introduce another variable that could confuse the interpretation of the results of the proposed study. Since the inception of this study other studies, using the same clinical protocol, have been conducted in the U.S. and Italy with K1 and have shown that acupuncture point injection with K1 has a clinical effect similar to K3.

The researcher who developed the treatment, Dr. Jin Yu, rationalized the use of point injection with vitamin K on the basis of several influences: ethnobotany, innovation in acupuncture technique (acupuncture point injection) and clinical observations and research studies conducted in the 1980's.

She collected and chemicallly analysed plants used in traditional herbal medicine for painful menses and found that they contained high amounts of vitamin K. She administered vitamin K orally to her patients with dysmenorrhea in a busy urban menstrual disorder clinic. Oral administration was not effective. Studies of oral vs injection administration of vitamin K indicate that injection may be more

effective.

Acupuncture point injection was developed as an experimental innovation in Chinese Medicine and clinicians and researchers began to inject various agents (saline, analgesics, herbal extracts and vitamins) in specific points as experimental treatments in the 1980's. Dr. Yu tried injecting Vitamin K3 (which was commonly available and used worldwide for blood disorders) into the acupuncture point commonly used for gynecological conditions, including dysmenorrhea. She observed a robust and rapid clinical effect of the treatment. Women who were bent over and sweating from severe pain would recover in only a few minutes. She then conducted studies of the treatment in animals and women at the Menstrual Disorder Clinic at the Obstetrics and Gynecology Hospital in Shanghai, China.

While the author's find these developments over many years extremely interesting, it was not possible to include these details in a research report, as many journals are reducing word count allowances. We refer the reviewer to papers we cited on various topics of vitamin K use from a leading laboratory in vitamin K research at Tufts. We sought the expertise of this laboratory seeking an answer to just the question the reviewer asked and published a preliminary study indicating vitamin K levels before and after treatment in 3 subjects. Documenting absorption of vitamin k administered by acupuncture point injection is a preliminary step in the pharmacokinetics of vitamin K.

Booth SL. Roles for vitamin K beyond coagulation. Annu. Rev. Nutr. 2009;29:89-110
Truong JT, Booth SL. Emerging issues in vitamin K research. J Evid Based Complement Altern Med 2011;16(1):73-79
Chao M, Wade C, Booth SL. Plasma phylloquinone concentrations increase following acupoint injection for primary dysmenorrhea. J Acupunct Meridian Stud 2014;7(3):151-54

Reviewer: 4 Reviewer Name Andrew Hinde
Southampton Statistical Sciences Research Institute
University of Southampton,United Kingdom

This paper describes a well-designed (and clearly quite demanding) study of the effect of vitamin K3 injections on primary dysmenorrhea. The description of the study makes it clear that similar studies are unlikely to be common, so the results of this one should definitely be published. Before publication, however, there are one or two revisions needed, mainly to clarify the methods used.

We thank the reviewer for recognizing the design and the demands of our study. An international team worked together for over a decade in various clinical settings to provide clinical evidence to increase treatment options for a common health condition.

1. I am a statistical reviewer, not an expert in the substantive topic, so please forgive my ignorance, but why was the saline injection to Group B given to a point near the SP6 acupuncture point but not right at SP6? Does this not potentially compromise this injection as a control?

For Group B we used a sham acupuncture point to control for the site of injection. In order to limit this group to a control for the site of injection we administered a single dose of vitamin K3 in the buttock in the same amount as 2 doses of vitamin K3 in each ankle as administered in the GROUP C. Please see clarification of the control GROUP B above in response to Reviewer 2 and modification of the text on page 6.

2. Somewhat facetiously, can I ask whether any women were excluded because they did not have a working phone (p. 6, l. 20)?

At the time the study was planned (2003) not every young person had a mobile phone. By the time the study was initiated and certainly by the time the study ended it was much more common, so no women were excluded because they did not have a phone.

3. (This is my main recommendation.) Can you add a couple of sentences describing your models in more detail? You refer to 'mixed effects' models, but this covers a pretty wide range. First, the way I would have approached the model is to note that observations are clustered within women, so I need a woman-level random effect to control for this. I guess (hope) this is how your reasoned but you do not say. Second, how do you accommodate time since treatment in your models? Are the results reported in Table 2 some kind of average of the effects at different times? Or did you include time since treatment as a covariate? Figure 2 suggests that there is a widening of the difference between Group C and Groups A and B over time, though this effect is not huge, and might not be statistically significant. Can youi add a sentence explaining how you dealt with time since treatment?

We supplied additional details (response variables, covariates, random effect) in the statistical model specifications at the bottom of page 9 and on page 10. The model accommodates the "time effect" by including two intercept terms for the two time points, 0 and 60 minutes. Note that we have only included the pain measurements at 0 and 60 minutes. This is because, as the reviewer noticed, the treatment effect takes place gradually and it is more appropriate to use the last time point as the proper measurement for the treatment effect.

Finally, Table III and the text are inconsistent. On p. 12, ll. 2-5 you state that '[a]ll groups had significant differences ... between the baseline cycle (Cycle 1) and the post treatment cycle (Cycle 2) on the Cox Duration and Cox Intensity scores.' But Table III indicates a significant difference for Group C but not for Groups A and B. Similarly, you write that '[a]ll groups also showed a significant difference ... between the baseline scores of menstrual distress and Cycle 7 scores at the six month follow up report.' But again Table III suggests that only Group C had such a difference. This should be corrected.

We thank the reviewer for indicating that results in Table III were not correctly noted. We have correctly noted the significance of all between cycle differences within groups on Cox scores and between group differences on Moos scores in Table III. Unlike the primary outcome between group differences were not large enough to be significant on the long term outcomes as the between cycle differences were. This is not unexpected as all groups received active treatments and that measurement over multiple cycles may not be as specific as pain scores within an hour after treatment.

VERSION 2 – REVIEW

REVIEWER	Mark W Strudwick University of Queensland Australia
REVIEW RETURNED	02-Sep-2015

GENERAL COMMENTS	The authors appear to have addressed all issued raised in the first review. I recommend publication.

REVIEWER	Andrew Hinde Southampton Statistical Sciences Research Institute University of Southampton United Kingdom
REVIEW RETURNED	23-Sep-2015

GENERAL COMMENTS	In my report on the original submission of this paper, I raised the following three points. (1) More explanation was needed of why Group B had a saline injection in a non-acupuncture point near the SP6 acupuncture point but not in the SP6 acupuncture point itself. (2) More explanation was needed of the models underlying the results in Table II. (3) There were inconsistencies between the text and the results presented in Tables III and IV. I have read the revised version and I am happy with your response to points (1) and (2) but I am afraid that there are still problems with Tables III and IV and the associated text. I have looked in detail at these tables and the more I look at it the more questions I have. Here are my outstanding concerns. You do not say what tests you used to compare the means in the baseline cycle with the means in the six treatment cycles. There is nothing in your methods section on pp. 9-10 to explain this. I would have used t-tests for dependent samples to compare women in the same group in different cycles (as you effectively have repeated measures), and you may have done this. However if you did it is not possible to check your results as you do not present the standard deviation of the paired differences (instead you present the standard deviation of the measures at each cycle, which is not the same thing). The formula for the t-statistic is $X_D/(s_D/\sqrt{n})$, where X_D is the mean difference between the baseline cycle and the cycle being tested, s_D is the standard deviation of the difference, and n is the sample size. Note c to table 3 refers to a 'significant difference between groups A and C' and 'between groups B and C' on pain reduction in the MMDQ columns. What test did you use for this? Another issue here is that MMDQ measures more than just pain. When you say there is a significant difference 'on pain reduction' do you mean just the pain element of the MMDQ score? A similar question arises in Table IV. What tests did you do in order to reach the conclusion that '[m]ean analgesic consumption was significantly reduced in Groups A and C when compared to Group B ($p = 0.013$)'. And does this conclusion refer just to Visit 7 or all

	Visits? Actually, I find the reported statistical significance levels in Table III rather implausible, and perhaps they detract from the main story, which is that all three treatments appear to lead to a substantial and reduction in menstrual distress which persists through at least six cycles, and to a similarly persistent reduction in analgesic tablet consumption. Your samples are small, and whether or not any specific comparison reaches statistical significance is, in my view, not such a crucial issue. It is the overall pattern of the results which is most important. In fact, in Table III you are making a total of 90 comparisons (6 cycles x 3 treatment groups x 5 measures of distress) so you would expect 4-5 statistically significant effects at the 95% significance level even under the null hypothesis of no effect. But the evidence from the pattern in the means for the Cox duration, the Cox intensity and the daily activity restriction seems persuasive to me. I recommend that you compute the statistical significance of all the comparisons in Tables III and IV using the paired sample t-test. You can then either (i) present the standard deviations of the differences (the s_Ds in the formula above) in the STD columns of Tables III and IV and indicate the statistical significance of each comparison from the paired sample t-test (use $p < 0.05$ for such small samples); (ii) omit the standard deviation columns completely, and simply add a note giving the number of comparisons that were statistically significant ($p < 0.05$) in each column. In the text, the statements on p. 15, ll. 8-10 are contradictory. You say there was no significant difference in tablet consumption between the three groups, but note b in Table IV indicates that there was such a difference! Your conclusion on p. 17,ll. 3-6 is a good statement of the limitations of the analyses in Tables III and IV yet it omits to mention the persistent nature of the impact of all three treatments.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewer 2
Corresponding Author: Christine Wade
Manuscript ID bmjopen-2015-00816 revision #2

Acupuncture Point Injection Treatment of Primary Dysmenorrhea: A randomized, double blind, controlled study

We thank the reviewer for his careful reading of our paper, and for his patience and persistence. His questions about the secondary long term outcomes allowed us to clarify our results and to discuss them more clearly.
The methods and results of 2 types of comparisons are better described in Table III and IV and the corresponding text. We did not use a linear model with adjustments; we report paired comparisons on the secondary outcomes. We have removed the standard deviations and clarified the notes again on Tables III and IV.
We understand why the reviewer was confused about the conflicting statements about significance in the text and the tables. Please see clarifications in responses below. The first type of test we reported on is a WITHIN group test that determines whether menstrual distress was reduced between two different time points 1.(Baseline and Cycle 2) and 2. (Baseline and Cycle 7) WITHIN each group. WITHIN each group we observed a significant reduction in menstrual distress at these time points when testing the means, and we interpret this as a reduction in menstrual distress over the long-term in each and all of the three groups.
The second type of test was a comparison between the menstrual distress means of each group. However, the difference in the reduction BETWEEN groups was not significant, except between the Baseline Cycle and Cycle 2 on the MMDQ score. The MMDQ Scale may be more sensitive than the COX Scales. We don't know. We collected both scales to help us determine the most appropriate measure for future research, as well as for clinical outcomes.
We have described the methods (ANOVA) used to test the secondary outcomes on page 10 in the Methods Section. We have modified the reporting of the results on page 13 in the Results Section and we have emphasized the pattern of reduction of menstrual distress in all three groups on page 17 of the Discussion Section.

Below are the reviewer's outstanding concerns. Our responses are bulleted below in large text.
You do not say what tests you used to compare the means in the baseline cycle with the means in the sixtreatment cycles. There is nothing in your methods section on pp. 9-10 to explain this. I would haveused t-tests for dependent samples to compare women in the same group in different cycles (as youeffectively have repeated measures), and you may have done this. However if you did it is not possibleto check your results as you do not present the standard deviation of the paired differences (instead you present the standard deviation of the measures at each cycle, which is not the same thing). The formulafor the t-statistic is XD/(sD/√n), where XD is the mean difference between the baseline cycle and the cycle being tested, sD is the standard deviation of the difference, and n is the sample size.
•We have described the methods (ANOVA) used to test the secondary outcomes on page 10 in the Methods Section.
Note c to table 3 refers to a 'significant difference between groups A and C' and 'between groups B and C' on pain reduction in the MMDQ columns. What test did you use for this? Another issue here is that MMDQ measures more than just pain. When you say there is a significant difference 'on pain reduction' do you mean just the pain element of the MMDQ score?
•We used ANOVA. We corrected the footnote and changed "pain" to "menstrual distress" as we

tested on the whole scale not just the pain dimension. We removed the standard deviations from the table.
A similar question arises in Table IV. What tests did you do in order to reach the conclusion that '[mean
analgesic consumption was significantly reduced in Groups A and C when compared to Group B (p = 0.013)'.

• We removed the comparisons from the analgesic data and report only the descriptives.

Actually, I find the reported statistical significance levels in Table III rather implausible, and perhaps they
detract from the main story, which is that all three treatments appear to lead to a substantial and reduction in menstrual distress which persists through at least six cycles, and to a similarly persistent reduction in analgesic tablet consumption. Your samples are small, and whether or not any specific comparison reaches statistical significance is, in my view, not such a crucial issue. It is the overall patternof the results which is most important. In fact, in Table III you are making a total of 90 comparisons (6cycles x 3 treatment groups x 5 measures of distress) so you would expect 4-5 statistically significant effects at the 95% significance level even under the null hypothesis of no effect. But the evidence from the pattern in the means for the Cox duration, the Cox intensity and the daily activity restriction seems persuasive to me.
•We agree with the reviewer that the overall pattern of results is more important than the difference between groups on the secondary outcomes. We also agree that there was a trend in reduction of menstrual distress over the six month follow-up. This effect may be related to the treatment but possibly not. A study with a different control group and more rigorous modeling would be more suitable for determining the long-term effect. Women were screened prior to the study and reported they had severe dysmenorrhea only partially relieved by any other treatment for 6 months prior to the study, but we did not do any pre-treatment measurements of menstrual cycle distress except for our baseline measure. We also note the possible interaction between analgesic consumption and menstrual distress scores. We have emphasized (the Discussion Section on page 17) the trends of pain reduction and decrease of analgesic use.
I recommend that you compute the statistical significance of all the comparisons in Tables III and IV using the paired sample t-test. You can then either
(i) present the standard deviations of the differences (the sDs in the formula above) in the STD columns

of Tables III and IV and indicate the statistical significance of each comparison from the paired sample ttest (use $p < 0.05$ for such small samples);

(ii) omit the standard deviation columns completely, and simply add a note giving the number of

comparisons that were statistically significant ($p < 0.05$) in each column.

• We have eliminated the standard deviations from the Cox and MMDQ outcomes in Table III.
In the text, the statements on p. 15, ll. 8-10 are contradictory. You say there was no significant difference in tablet consumption between the three groups, but note b in Table IV indicates that there was such a difference!

• We have corrected this inconsistency on page 15.

Your conclusion on p. 17,ll. 3-6 is a good statement of the limitations of the analyses in Tables III and IV yet it omits to mention the persistent nature of the impact of all three treatments.
•We agree with the reviewer's comments that the overall pattern of reduction of menstrual distress in

all the groups should be emphasized. The difference between the three groups which the Cox and MMDQ scales may not be sensitive enough to capture over such a long follow-up and with the degree of variation of menstrual cycle symptoms and consumption of analgesics. We report the trend in the abstract and in the results pages 14 and 15. We expected all three injection treatments to be active and they were on both short-term and long-term outcomes. We have noted the pattern of relief in all three groups in the Discussion on page 17.