Statistical Analysis Plan

Study Title: Assessment of effectiveness and safety of a novel pessary for the non-surgical management of pelvic organ prolapse

NCT Number: NCT04508335

Document Date: 8/16/2023

1.0 OBJECTIVE

The Statistical Analysis Plan outlines the analysis groups, statistical tests, and success criteria for endpoints in the Collapsible Pessary Feasibility Trial.

2.0 SCOPE

Both primary and secondary endpoints are outlined.

- 2.1 Primary endpoints
  - 2.1.1 PFDI-20
  - 2.1.2 Adverse events
- 2.2 Secondary endpoints
  - 2.2.1 PFIQ-7
  - 2.2.2 PISQ-IR
  - 2.2.3 Objective assessment of ability to support the prolapse
  - 2.2.4 Global satisfaction
  - 2.2.5 Pain associated with insertion and removal
  - 2.2.6 Success and ease of fitting
  - 2.2.7 Frequency of self-management
  - 2.2.8 Ease of insertion and removal

3.0 REFERENCES

- 3.1 CP-002-01 Collapsible Pessary Feasibility Trial Protocol
- 3.2 CP-002-03 Collapsible Pessary Feasibility Trial Data Collection Form
- 3.3 SOP-1003 Design Control
- 3.4 F-3006 Test Report Template
- 3.5 SOP-1002 Quality Records
- 3.6 F-3010 Trial Adverse Event Log

4.0 DEFINITIONS

- 4.1 PFDI-20 – Pelvic Floor Distress Inventory-20
- 4.2 PFIQ-7 – Pelvic Floor Impact Questionnaire-7
- 4.3 PISQ-IR – Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire, IUGA-Revised
- 4.4 POP-Q – Pelvic Organ Prolapse Quantification
- 4.5 VAS – Visual Analogue Scale
- 4.6 CTCAE – Common Terminology Criteria for Adverse Events

5.0 RESPONSIBILITIES

5.1 Reia personnel and approved suppliers will be responsible for analyzing data.

5.2 Reia must maintain all quality records, including copies of data analysis performed by approved suppliers, according to SOP-1002 Quality Records.

6.0 PFDI-20 (PRIMARY)

6.1 Endpoint

6.1.1 Pelvic Floor Distress Inventory-20 (PFDI-20) of study pessary compared to current pessary

6.2 Data collection method

6.2.1 The PFDI-20 is a validated instrument to assess the presence and bother of symptoms that pelvic floor disorders have on health-related quality of life in women. It is a 20-item questionnaire and the short form of the 46 question PFDI. There are three subscales: Urinary Distress Inventory 6 (UDI-6), Colorectal-Anal Distress Inventory 8 (CRADI-8), and the Pelvic Organ Prolapse Distress Inventory 6 (POPDI-6). The overall score range is 0-300. The higher the score, the greater the perceived impact that pelvic floor dysfunction has on a patient's life.

6.3 Analysis group

6.3.1 In the primary analysis, all subjects who complete visit 1 will be included in this intention to treat group. Subjects who drop out of the study prior to visit 1 are not included in this analytic group. Subjects who drop out and are included in the analytic group will be assigned the lower (i.e., worse) limit score of 18.3 for the study pessary.

6.3.2 A secondary per protocol analysis will not include subjects who drop out because of failure to be fit.

6.4 Data analysis plan

6.4.1 For each subject, the difference in score for the PFDI-20 for the current and the study pessary will be calculated.

6.5 Success criteria

6.5.1 Wiegersma et al. determined the minimal important change (MIC) in the PFDI-20 to demonstrate a clinical effect to be 13.5-18.3 points in women choosing conservative management of their prolapse.[1] Each analysis is a one-sample, equivalence design. We use 2, one-sided t-tests, each with alpha set at .05 to test the composite null hypothesis that the mean difference score (μreia-control) between baseline (current pessary) and study pessary on the PFDI-20, is greater than 18.3, the upper equivalence limit, or lower than -18.3, the lower equivalence limit:

H01: μreia-control > 18.3 and H02: μreia-control < -18.3. The alternative hypothesis is thus: HA: -18.3 ≤ μreia-control ≤ 18.3

We reject the null hypothesis when the maximum p value across both t-tests is less than .05.

7.0 ADVERSE EVENTS (PRIMARY)

7.1 Endpoint

7.1.1 Prevalence and severity of adverse events, as defined below, that subjects experience using the study pessary.

7.2 Data collection method

7.2.1 At each visit, subjects will report discomfort, displacement, discharge, bleeding, obstructed urination, constipation, and other potential adverse events. If present, each reported event is described subjectively as mild, moderate, or severe by the subject.

7.2.2 At each in-person visit, practitioners will assess for physical exam findings of perineal tearing, vaginal irritation, vaginal ecchymosis, vaginal abrasions/superficial cuts, deep vaginal erosion, vaginal bleeding, granulation tissue, and copious vaginal discharge.

7.2.3 At each additional visit or phone call, expected and unexpected adverse events are reported by patients and/or observed by practitioners.

7.2.4 Study personnel can make entries in their site-specific F-3010 Trial Adverse Event Log for events described above or unique events not otherwise captured.

7.2.5 The events above will be reviewed and adapted to conform to Common Terminology Criteria for Adverse Events (CTCAE) ratings. The CTCAE ratings will be determined by the PI at each site.

7.2.6 The events will be condensed to unique adverse events consisting of:

- New subjective reports of events related to the study pessary not present at baseline with the current pessary
- Worsening subjective reports of events related to the study pessary that were present at baseline with the current pessary
- New physical exam findings related to the study pessary not present at baseline with the current pessary
- Worsening physical exam findings related to the study pessary that were present at baseline with the current pessary
- Any additional entries in the F-3010 Trial Adverse Event Log

These unique adverse events will represent the data analysis set.

7.3 Analysis group

7.3.1 All subjects who complete visit 0 will be included in the analysis group.

7.4 Data analysis plan

7.4.1 A count of mild, moderate, and severe unique adverse events will be tallied from the data analysis set. The type of adverse event will also be tallied.

7.5 Success criteria

7.5.1 The absence of serious and related adverse events that could put the general population at risk.

8.0 ADDITIONAL QUESTIONAIRES (SECONDARY)

8.1 Endpoint

8.1.1 Pelvic Floor Impact Questionnaire (PFIQ-7) of study pessary compared to current pessary

8.1.2 Pelvic Organ Prolapse/Urinary Incontinence Sexual Survey (PISQ-IR) of study pessary compared to current pessary

8.2 Data collection method

8.2.1 PFIQ-7 is a health-related quality of life instrument, designed to assess the life impact of pelvic floor symptoms in women with pelvic floor disorders. The PFIQ-7 has a parallel structure to the PFDI with 3 Scales: Colorectal Anal Impact Questionnaire (CRAIQ), Pelvic Organ Prolapse Impact Questionnaire (POPIQ), and Incontinence Impact Questionnaire (IIQ).

8.2.2 PISQ-IR is a validated evaluation tool which can be used clinically as well as in research for assessment of female sexual function (FSF) in women with female pelvic floor disorders.

8.3 Analysis group

8.3.1 PFIQ-7 and PISQ-IR are recorded at visit 0 for the current pessary and visit 3 for the study pessary, so only subjects who complete data collections forms through visit 3 will be included in the analysis.

8.4 Data analysis plan

8.4.1 For each subject, the difference in score for the PFIQ-7 and PISQ-IR for the current and the study pessary will be calculated.

8.5 Reportable outcomes

8.5.1 Average and standard deviation of the PFIQ-7 and PISQ-IR for the current and study pessary will be reported.

8.5.2 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary on the PFIQ-7 is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

8.5.3 We will use a two-tailed t-test to test the null hypothesis that the mean difference between baseline (current pessary) and study pessary on the PISQ-IR is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

9.0 ABILITY TO SUPPORT PROLAPSE (SECONDARY)

9.1 Endpoint

9.1.1 Objective assessment of the study pessary's ability to support anterior prolapse compared to current pessary

9.1.2 Objective assessment of the study pessary's ability to support posterior prolapse compared to current pessary

9.2 Data collection method

9.2.1 Modified POP-Q points Ba and Bp (measurements of anterior and posterior prolapse) will be measured during physical exam, with the pessary in situ.

9.3 Analysis group

9.3.1 At most, for each subject two Ba and two Bp measurements will be collected for each of the current pessary and the study pessary.

9.3.2 Subject's Ba and Bp measurements will be included in the comparative analysis if the subject has at least 1 data point for both the study pessary and their current pessary and did not drop out because of failure to fit.

9.4 Data analysis plan

9.4.1 For each subject, average Ba and Bp scores for the current pessary and study pessary will be calculated.

9.4.2 For each subject, a difference (current pessary – study pessary) in Ba and Bp scores will be calculated.

9.5 Reportable outcomes

9.5.1 Average and standard deviation of the Ba and Bp with the current and study pessary will be reported.

9.5.2 Mean difference (current pessary – study pessary) in Ba and Bp scores will be reported.

9.5.3 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary on the Ba is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

9.5.4 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary on the Bp is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

9.5.5 Due to intra- and inter- observer variation in the measurements for Ba and Bp, a difference > 1.5 cm is considered meaningful change. This analysis will be a one-sample, equivalence design. We will use 2, one-sided t-tests, to test the composite null hypothesis that the mean difference score between baseline (current pessary) and study pessary on the Ba measurement is less than or equal to -1.5 cm and greater than or equal to 1.5 cm. The p-value will be reported.

9.5.6 We will use 2, one-sided t-tests, to test the composite null hypothesis that the mean difference score between baseline (current pessary) and study pessary on the Bp measurement is less than or equal to -1.5 cm and greater than or equal to 1.5 cm. The p-value will be reported.

10.0 GLOBAL SATISFACTION (SECONDARY)

10.1 Endpoint

10.1.1 Global assessment of study pessary satisfaction compared to current pessary

10.1.2 Global assessment of study pessary satisfaction

10.2 Data collection method

10.2.1 Satisfaction will be assessed using a Visual Analog Scale (VAS) with a 10 cm linear continuum in which 0 represents no satisfaction and 10 represents complete satisfaction.

10.3 Analysis group

10.3.1 Global satisfaction is recorded at visit 1 for the current pessary and visit 3 for the study pessary, so only subjects who complete data collections forms through visit 3, or provide a score for both the current and study pessary before exiting the study, will be included in the analysis.

10.4 Data analysis plan

10.4.1 For each subject, a difference (current pessary – study pessary) in satisfaction scores will be calculated.

10.5 Reportable outcomes

10.5.1 Average and standard deviation of the global satisfaction of current and study pessary will be reported.

10.5.2 Mean difference (current pessary – study pessary) in satisfaction scores will be reported.

10.5.3 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary for global satisfaction is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

10.5.4 Cundiff et al. dichotomized VAS scores for pessary satisfaction into high satisfaction (greater than or equal to 80 mm) and low satisfaction (less than 80 mm).[2] For this study, we similarly deemed a VAS of 80 mm would indicate high satisfaction. We will use a one-sided t-test to test the null hypothesis that the mean global satisfaction score for the study pessary is less than 80 mm. The p-value will be reported.

11.0 PAIN ASSOCIATED WITH INSERTION & REMOVAL (SECONDARY)

11.1 Endpoint

11.1.1 Pain associated with study pessary insertion compared to current pessary

11.1.2 Pain associated with study pessary removal compared to current pessary

11.2 Data collection method

11.2.1 Pain will be assessed using a VAS with a 10 cm linear continuum in which 0 represents no pain and 10 represents worst pain.

11.3 Analysis group

11.3.1 At most, for each subject two VAS pain scores will be collected for each of insertion of current pessary, insertion of the study pessary, removal of the current pessary, removal of the study pessary. Subject's VAS pain scores will be included in the comparative analysis group if the subject has at least 1 data point for both the study pessary and their current pessary and in the non-comparative analysis group if they have at least 1 data point for the study pessary.

11.4 Data analysis plan

11.4.1 For each subject, average pain scores for the current pessary and study pessary will be calculated.

11.4.2 For each subject, a difference (current pessary – study pessary) in average pain scores will be calculated.

11.5 Reportable outcomes

11.5.1 Average and standard deviation of the pain of current and study pessary insertion and removal will be reported.

11.5.2 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary for pain during insertion is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

11.5.3 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary for pain during removal is zero. The p-value will be reported.

12.0 SUCCESS & EASE OF FITTING (SECONDARY)

12.1 Endpoint

12.1.1 Proportion of subjects successfully fit with the study pessary

12.1.2 Number of study pessaries required to fit

12.2 Data collection method

12.2.1 If subject is unable to be fit with a study pessary it will be recorded

12.2.2 Size of each study pessary tried will be recorded

12.3 Analysis group

12.3.1 All subjects who complete the study will be considered successfully fit.

12.3.2 For those who do not complete the trial, if the study team indicates that the reason for withdrawal is failure to fit or failure to retain, the subject will be considered failed to fit. If the study team indicates that the reason for withdrawal is discomfort with the study pessary and multiple sizes of study pessary have been tried, the subject may be considered failed to fit. This may be indicated on the data collection form or in data safety monitoring meetings. All other subjects who received the study pessary and were not deemed fit failures will be considered successfully fit.

12.3.3 For subjects who are successfully fit, each size tried will be considered a study pessary required to fit.

12.4 Data analysis plan

12.4.1 A count of subjects who were successfully fit will be calculated.

12.4.2 A count of subjects who failed to be fit will be calculated.

12.4.3 A proportion of subjects successfully fit to total subjects (successfully fit + failed to fit) will be calculated.

12.4.4 A count of total pessaries used will be calculated for each subject who completed.

12.4.5 An average of number of pessaries required to successfully fit will be calculated.

12.5 Reportable outcomes

12.5.1 The proportion of subjects successfully fit will be reported.

12.5.2 The average pessaries required to fit will be reported.

13.0 PROPENSITY & FREQUENCY OF SELF-MANAGEMENT (SECONDARY)

13.1 Endpoint

13.1.1 Change in frequency of pessary removal for self-managers

13.1.2 Change in management style (self-managed or exclusively practitioner-managed)

13.2 Data collection method

13.2.1 At visit 0, the subject is asked to estimate the average number times per month their current pessary was removed/reinserted since their last visit.

13.2.2 At visit 1 and visit 3, the subject is asked to estimate the number of times the pessary (current or study) was removed/reinserted since the last visit.

13.2.3 During the visit 2 screen and at visit 2 if it occurs, the subject is asked to estimate the number of times the study pessary was removed/reinserted since visit 1.

13.3 Analysis group

13.3.1 Current pessary self-management frequency will be estimated at visit 0 from the subject's memory, while the number self-management events with current pessary between visits will be reported at visit 1. Therefore, only visit 1 data will be used in the frequency analysis of the current pessary.

13.3.2 Since visit 2 occurs 1-3 weeks after visit 1, time between visits is not sufficient to capture self-management. Therefore, visit 2 screen data and visit 2 in-person data will not be included in the frequency analysis. Instead, number of self-management events reported at visit 3, which occurs 11-13 weeks following visit 2, will be used in the analysis of the study pessary.

13.3.3 Since data from visit 3 is required in the frequency analysis, only subjects who complete the study will be included in the analysis group.

13.3.4 Visit 0 and visit 1 data will be used in the management style analysis for the current pessary. Visit 2 and visit 3 data will be used in the management style analysis for the study pessary.

13.4 Data analysis plan

13.4.1 For visit 1 and 3 responses, number of self-management events since last visit will be converted to number of self-management events per month (defined as 28 days) using the number of days that passed between visits.

13.4.2 For each subject, a difference (current pessary – study pessary) in number of self-management events per month will be calculated.

13.4.3 A count of subjects who self-manage (number of self-management events per month > 0) will be calculated for the current pessary (visit 0 or 1) and the study pessary (visit 2 or 3).

13.4.4 In the case that the rate data is not normally distributed as required by the t-test, we may transform the data, so it is reasonably normally distributed (using transformations such as the log or square root). If transforming the data does not result in a normal distribution, a non-parametric t-test may be used.

13.5 Reportable outcomes

13.5.1 Average and standard deviation of the frequency of self-management with current and study pessary will be reported.

13.5.2 Mean difference (current pessary – study pessary) in number of self-management events per month will be reported.

13.5.3 The number of self-managers with the current pessary and study pessary will be reported.

13.5.4 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary in the frequency of self-management is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

13.5.5 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary in the number of self-managers is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

14.0 EASE OF INSERTION & REMOVAL (SECONDARY)

14.1 Endpoint

14.1.1 Ease of insertion of study pessary compared to current pessary for self-managers

14.1.2 Ease of removal of study pessary compared to current pessary for self-managers

14.1.3 Ease of insertion of study pessary by subject for self-managers

14.1.4 Ease of removal of study pessary by subject for self-managers

14.2 Data collection method

14.2.1 Ease will be assessed using a Visual Analog Scale (VAS) with a 10 cm linear continuum in which 0 represents very easy and 10 represents very difficult.

14.3 Analysis group

14.3.1 At most, for each subject two VAS ease of use scores will be collected for each of insertion of current pessary, removal of the current pessary, insertion of the study pessary, and removal of the study pessary. Subject's VAS ease of use scores will be included in the comparative analysis group if the subject has at least 1 data point for both the study pessary and their current pessary and in the non-comparative analysis group if they have at least 1 data point for the study pessary.

14.4 Data analysis plan

14.4.1 For each subject, average ease of use scores for the current pessary and study pessary will be calculated.

14.4.2 For each subject, a difference (current pessary – study pessary) in average ease of use scores will be calculated.

14.5 Reportable outcomes

14.5.1 Average and standard deviation of the ease of current and study pessary insertion and removal will be reported.

14.5.2 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary in the ease of insertion is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

14.5.3 We will use a two-tailed t-test to test the null hypothesis that the mean difference score between baseline (current pessary) and study pessary in the ease of removal is zero. This test is equivalent to a matched samples t-test. The p-value will be reported.

15.0 TEST REPORT

15.1 The data will be collected according to CP-002-01 Collapsible Pessary Feasibility Trial Protocol and the CP-002-03 Collapsible Pessary Feasibility Trial Data Collection Form and analyzed according to the procedure described above. Results of the analysis shall be documented in a written test report, as described in SOP-1003 Design Control. F-3006 Test Report Template may be used.

15.2 Deviations from the procedure will be reviewed for impact, approved by the project team and documented in the test report.

16.0 REVISION HISTORY

Revision	Summary of Changes/Affected Sections
01	Initial release of document
02	Updated to include mean differences described in data analysis plans in reportable outcomes; clarification provided on some analysis group populations

CITATIONS

1. Wiegersma M, Kollen BJ, Dekker JH, Berger MY, Panman CMCR, De Vet HCW. Minimal important change in the pelvic floor distress inventory-20 among women opting for conservative prolapse treatment. Am J Obstet Gynecol. 2016;216(4):397.e1-397.e7. doi:10.1016/j.ajog.2016.10.010
2. Cundiff GW, Amundsen CL, Bent AE, et al. The PESSRI study: symptom relief outcomes of a randomized crossover trial of the ring and Gellhorn pessaries. Am J Obstet Gynecol. 2007;196(4):405.e1-405.e8. doi:10.1016/j.ajog.2007.02.018