TOPSY STATISTICAL ANALYSIS PLAN
FINAL Version 2.0
12th May 2021

Prepared by:

Andrew Elders [redacted] Date: 12/05/2021
Senior Statistician

Cath Best [redacted] Date: 12/05/2021
Trial Statistician

Approved by:

Carol Bugge [redacted] Date: 12/05/2021
Chief Investigator

Table of Contents

1. Study Design

TOPSY is a parallel two−arm, multi-centre, pragmatic, superiority, randomised controlled trial to evaluate the clinical effectiveness and cost-effectiveness of pessary self-management compared with standard care for women with pelvic organ prolapse. Full details of the study design are available in the TOPSY trial protocol (Hagen 2020).

1.1 Research Questions

1. What is the clinical and cost-effectiveness of self-management of vaginal pessaries to treat pelvic organ prolapse, compared to standard pessary care on condition-specific quality of life? (RQ1)
2 What are the barriers and facilitators to intervention acceptability, intervention effectiveness, fidelity to delivery, and adherence for women treated with vaginal pessary and the health professionals who treat them, and how does this differ between randomised groups? (RQ2)

2. Scope

This document details the planned statistical analysis and presentation of data for the main paper and final study report for the TOPSY study. Any deviations from this Statistical Analysis Plan (SAP) will be described and justified in the final report of the trial. Subsequent post-hoc analyses of a more exploratory nature (i.e. analyses not in the protocol nor the final report, but that may be carried out in the future) will not be bound by this analysis plan, although they are expected to follow the same broad principles. The SAP will be available when the study findings are submitted for publication by the funder (NIHR HTA) and when submitted to a journal. This SAP has been developed according to recently published guidelines (Gamble 2017). The analysis specified in this SAP will be conducted by the trial statistician after the trial database has been locked.

This document does not cover the analysis of the internal pilot data (now completed), the cost-effectiveness analysis element of RQ1 or the process evaluation (RQ2). Separate analysis plans exist for each of these elements of the study: Health Economics Analysis Plan (HEAP) and Process Evaluation Analysis Plan (PEAP) respectively.

3. Sample Size

The protocol specifies a sample size of 330 women (165 per group) to ensure 90% power to detect a difference of 20 points in the PFIQ-7 score at 18 months, assuming a standard deviation of 50, two-sided alpha of 0.05, and 20% loss to follow-up. In order to detect this standardised effect size of 0.4 SDs (20/50 points), 132 women per group are required, or 165 per group to allow for loss to follow-up.

Two recent trials which used PFIQ-7 in populations of women using pessaries reported SDs at 12 months and 24 months between 25 and 40 (Wiegersma 2014, Panman 2016). These studies however were relatively small, conducted in only a few centres, and neither measured PFIQ-7 at 18 months. Given this uncertainty, a conservative assumption was made that the SD could be as high as 50.

The trial has recruited to target. The final participant was recruited in February 2020 and the total number randomised is 340. It was agreed that women who had already completed eligibility by the time 330 women had been recruited would be consented and randomised.

4. Outcome Measures

All outcomes are measured using responses from participant follow-up questionnaires at 6, 12 and 18 months (unless otherwise indicated). Outcome measures are also collected at baseline, where applicable. All primary and secondary outcome measures will be analysed in the statistical analysis, with the exception of uptake of additional treatment for prolapse which will be analysed as part of the health economic analysis.

4.1 Primary outcome measure

The primary outcome measure is the Pelvic Floor Impact Questionnaire (PFIQ-7). The PFIQ-7 is a reliable, valid and responsive short form of the PFIQ which measures condition-specific quality of life in women with pelvic floor disorders including urinary incontinence, prolapse and faecal incontinence (Barber 2005, Barber 2011). The participant-completed instrument includes items asking about the effect of bladder, bowel and vaginal symptoms on the woman's activities, relationships and feelings. Each item in the PFIQ-7 has four response levels (0=not at all; 1=somewhat; 2=moderately; 3=quite a bit). There are three subscales (Urinary Impact Questionnaire UIQ-7, Colorectal-Anal Impact questionnaire CRAIQ-7, Pelvic Organ Prolapse Impact Questionnaire POPIQ-7). The mean score is calculated from complete responses within each domain. This is then scaled up (by multiplying by 100/3) to become the domain score out of 100. The overall PFIQ-7 score is the sum of the three domain scores (out of 300) (Barber 2005). Higher scores indicate worse quality of life.

The subscale scores are calculated as the mean of completed items in the subscale, so missing data is in effect imputed at the subscale mean. If more than 4 of the 7 items in any subscale are missing, then that subscale will be treated as missing. If a subscale score is missing, then it is not possible to calculate the PFIQ-7 score for that woman and it will be treated as missing (and imputed in all analyses other than the observed cases analysis, see section 6.3).

4.2 Secondary outcomes (validated measures)

4.2.1 EQ-5D-5L

The analysis of EQ-5D-5L is specified in the Health Economic Analysis Plan and is therefore not included in statistical analysis.

4.2.2 Pelvic Floor Distress Inventory (PFDI-20)

This will measure the severity of prolapse-related symptoms. This was developed and validated in parallel with the PFIQ-7 (Barber 2005). It contains 20 questions about the presence of bladder, bowel and pelvic symptoms, and how bothersome these are. There are three subscales (UDI-6, CRADI-8, POPDI-6), with each subscore ranging from 0-100 and a total score of 0-300.

4.2.3 Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-IR))

The PISQ-IR (Rogers 2013,) will be used to assess women's sexual symptoms and the following steps will be used to calculate a summary score. The PISQ-IR summary score can be calculated only for sexually active women, as it does not meet criterion validity in women who are not sexually active (Constantine 2017).

a) If a respondent indicates not having a sexual partner at Q12, verify that 9 or more of the following 18 Question items: Q7, Q8a, Q8b, Q8c, Q9, Q10, Q11, Q15, Q16, Q17, Q18, Q19a, Q19b, Q19c, Q20a, Q20b, Q20c, Q20d have a response provided.
b) If a respondent indicates having a sexual partner, verify that 11 or more of the following 21 Question items: Q7, Q8a, Q8b, Q8c, Q9, Q10, Q11, Q13, Q14a, Q14b, Q15, Q16, Q17, Q18, Q19a, Q19b, Q19c, Q20a, Q20b, Q20c, Q20d have a response provided.
c) Calculate the reverse response value for Question items Q8b, Q8c, Q9, Q11, Q14a, Q14b, Q16, Q17, Q18, Q19a, Q19b, Q19c using the formula provided.
d) For any respondent indicating no sexual activity in response to Q12, delete any response that may be provided for Q13, Q14a or Q14b.
e) Add the score values of valid items to give a total score.
f) Divide the total score by the number of valid items for which a response was provided to calculate the mean summary score.

In addition, there are ten subscales. The following four subscales can be calculated for women who are not sexually active (reverse response values apply to Q3, Q4a, Q4b and Q6):

- partner related (Q2a, Q2b)
- condition specific (Q2c, Q2d, Q2e)
- condition impact (Q3, Q5b, Q5c)
- global quality (Q4a, Q4b, Q5a, Q6)

The following six subscales are for women who are sexually active:

- arousal/orgasm (Q7, Q8a, Q10, Q11)
- condition specific (Q8b, Q8c, Q9)
- partner related (Q13, Q14a, Q14b)
- desire (Q15, Q16, Q17)
- condition impact (Q18, Q20b, Q20c, Q20d)
- global quality (Q19a, Q19b, Q19c, 20a)

We will also report the number of women who change from sexually active at baseline to not active at follow-up (and vice versa).

4.2.4 General Self Efficacy scale (GSE)

The GSE (Schwarzer 1995, Chen 2001) will be used to assess a woman's general self-efficacy (hypothesised to be a moderator of quality of life). This is a ten item scale with score ranging from 10 to 40. Although the trial registration mentions the Self-Efficacy Scale for Practising PFEs (SESPPFE, Sacomori 2013), there was an agreed change and the trial protocol indicates that the GSE will be the only validated measure for self-efficacy. However, see section 4.3.3 for Pessary Confidence Questionnaire which will be used to assess pessary specific self-efficacy.

4.2.5 Patient Global Impression of Improvement (PGI-I)

The PGI-I is a single-item ordinal measure asking the individual to rate the change in their condition since having treatment, which has been validated for urogenital prolapse (Yalcin 2003, Srikrishna 2010). An amended version asking women to describe how they feel about their pessary care since taking part in the study will be used, with response options ranging from very much better to very much worse.

4.3 Secondary outcomes (non-validated measures)

These outcome measures may require further validation using the data collected in TOPSY, although this is beyond the scope of the SAP.

4.3.1 Pessary Complications Questionnaire

This is a new pessary questionnaire developed for TOPSY (with 15 possible types of complication relating to pessary use e.g. discharge, odour, pain, discomfort, bleeding). Two items (questions 9 and 10) are applicable to self-management only. Each complication type will be reported separately. The proportion of complication types reported will also be calculated for each participant. Only the thirteen categories applicable to both standard care and self-management will be used in this calculation. For women who are not sexually active, only the relevant subset of complication types will be included in the calculation (questions 11 & 12 will therefore be excluded).

4.3.2 Pessary Use Questionnaire

A new questionnaire (includes nine questions) developed for TOPSY, will be used to assess the pattern of a woman's pessary use, including perceived acceptability and benefit along with continuation of pessary use. This will include questions that ask women: whether or not they are still using a pessary as treatment for prolapse; when they last removed and re-inserted their pessary; reasons for pessary removal; interference of the pessary with everyday life and if they find the pessary an acceptable treatment.

4.3.3 Pessary Confidence Questionnaire

In the absence of any suitable condition-specific measure, we developed the Pessary Confidence Questionnaire as a set of questions relating to pessary self-efficacy based on existing guidance (Bandura 1977), which will complement the validated measure of general self-efficacy (GSE). The questionnaire consists of six visual analogue questions ranging from 0 (not confident) to 100 (highly confident). Data for each will be presented separately, no summary score will be reported as each measure is a separate construct.

4.3.4 Uptake of additional telephone support related to pessary use

Question 16 in the Pessary Complication Questionnaire asks participants if they contacted their local clinic for advice in relation to pessary complications during the previous 6 months. The proportions of women who take up this additional support will be calculated for both groups. A secondary definition will be used (in descriptive summaries only) where additional support will be identified from the Clinic Visit Log (CRF07, questions B1 to B6), Only *additional* calls will be included in this definition, which can be determined from the reason recorded on the log. For example, a standard care appointment by telephone during the pandemic is not additional support).

4.3.5 Adherence to randomised protocol ('on treatment')

The proportions of women adhering to the self-management or standard care protocols for the duration of the 18-month intervention period will be reported in each group. Adherence will be analysed further as part of the process evaluation.

Women randomised to self-management:

- IF they have NOT discontinued pessary care (determined by the change of status form);
- IF they have received the intervention (identified either on the intervention checklist or from confirmation by the site);
- AND if they have answered YES to question Q2d (Pessary use questionnaire) "Have you inserted your pessary yourself in the last 6 months" at either 6, 12 or 18 months.

Women randomised to standard care:

- IF they have NOT discontinued pessary care (determined by the change of status form);
- AND if they have answered Yes to Q1 (Pessary use questionnaire) "Have you used a pessary for prolapse at any time in the last 6 months?" at least once at either 6, 12 and 18 months
- AND they have ticked NO to Q2d (Pessary use questionnaire) "Have you inserted your pessary in the last 6 months" at ALL time-points (6, 12 and 18 months).

Any analyses using 'on treatment' data will be complete case analyses as questionnaire responses are required at all follow-up timepoints in order to determine adherence. However, the following additional analyses will be carried out to deal with missing data required to decide if standard care women are on treatment:

- where a response of NO will be assumed if the response to Q2d is missing at a particular time point.
- those who have answered YES to Q2d at any time point (in either group (standard care and self-management) will be assumed to have a YES response at any other time point where the Q2d response is missing
- where a response of YES will be assumed if the response to Q2d is missing at a particular time point.

There are various reasons why a woman might stop being on treatment, including the uptake of other treatments (e.g. prolapse surgery) which could potentially effect outcomes). Uptake of further treatment will be assessed as part of the health economic analysis.

4.3.6 Health of vaginal tissues

At baseline and 18 months, women have a vaginal examination undertaken at the clinic by a healthcare professional to assess the health of vaginal tissues. Proportions of women with vaginal tissue problems associated with pessary use will be calculated. This includes four categories: tissue granulation, ulceration, inflammation or other. The proportion of women with any vaginal tissue problem will also be calculated.

5 Statistical Methods

This section sets out the general approach to the statistical analysis

5.1 General methods

Analyses will be conducted according to the intention-to-treat principle such that randomised participants will be analysed according to the treatment group to which they were originally assigned, regardless of treatment received, non-adherence or crossover.

Descriptive statistics will be tabulated by treatment group (see the dummy tables in Section 9) showing means and standard deviations for continuous and count data (or median and

interquartile range if data are skewed) and frequency and percentages for binary and categorical data.

Treatment effect sizes between groups will be estimated using generalised linear models appropriate to the type of outcome measures (repeated measures mixed effects models for continuous outcome measures, mixed effect logistic regression for binary outcomes and mixed effect ordinal regression for ordered categorical outcomes). Effect sizes will be reported as mean differences for continuous outcome measures and odds ratios for binary and ordinal outcomes, presented with 95% confidence intervals. Statistical significance will be at the 5% level accordingly.

All models will adjust for the following minimisation factors:

- Age (in years, as a linear variable)
- Pessary user type (new user / existing user)
- Centre – as a random effect. Multiple sites within a centre (e.g. Manchester) will be treated as a single unit, as typically the same staff work across multiple sites. Note that some centres (e.g Ayrshire & Arran) have multiple centre codes.

Models will also adjust for the baseline measure as a covariate, where applicable. A sensitivity analysis of the primary outcome will be conducted with previous hysterectomy included as an additional fixed effect.

A single final analysis is anticipated 18 months after the last participant has begun receiving the intervention. Statistical analysis will be conducted, where possible, using Stata (StataCorp, College Station, TX, USA). The trial statistician will be unblinded to allocation.

5.2 Baseline characteristics

Baseline characteristics will be tabulated by randomised group. Given that TOPSY includes two potentially different populations (new and existing pessary users), an additional tabulation of baseline characteristic will be tabulated by pessary user type. No inferential tests will be undertaken when comparing participant characteristics between randomised groups, but comparisons of participant characteristics will be tested between new and existing users (chi-squared tests for categorical data, unpaired t-tests for continuous data.

5.3 Timing of bassline assessment of prolapse

Baseline data for prolapse (stage and type) were obtained either from the most recent existing assessment recorded in the clinical notes before the initial pessary was fitted or from an assessment carried out immediately prior to randomisation. Neither method is completely optimal as the former can be quite historic and the latter can underestimate the extent of the prolapse if conducted immediately after pessary removal. We will therefore report the proportion of women assessed by each method, and for the former method, we will report a

summary measure for the duration (e.g. mean number of months) from assessment to randomisation. Any assessment data collected after randomisation should not be regarded as baseline data and will be treated from the outset as missing. If assessment data have been collected via both methods, then the assessment prior to initial pessary fitting will be used when summarising stage and type.

5.4 Intervention received

The proportion of women randomised to self-management who are self-managing at 2 weeks and at 18 months will be reported. Other data relating to the delivery of the interventions will be reported through the process evaluation, the primary analysis of which will be conducted by an investigator other than the trial statistician and blinded to the results of the statistical analysis.

5.5 Primary outcome measure

Additional descriptive data by group (skewness and kurtosis) will be reported for the PFIQ-7 at each time point.

In the longitudinal analysis of covariance (as described by Twisk 2018), the value of the three follow-up measurements of the outcome variable will be employed as the dependent variable. The value of dependent variable measured at the different follow up points will be adjusted for the baseline value of the dependent variable. The model will include 'time' (or measurement point) as dummy variables because a non-linear development of the outcome over time is anticipated. Interaction effects between treatment and time will be included in the model and the fixed part of the model can be written as:

$$Y_t=\beta_0+\beta_1 X + \beta_2 Y_{t0} + \beta_3 \text{dummy\_time}_1+\beta_4 \text{dummy\_time}_2+\beta_5 \text{dummy\_time}_3 +\beta_6 \text{dummy\_time}_1*X+\beta_7 \text{dummy\_time}_2*X+\beta_8 \text{dummy\_time}_3*X$$

Where Y is PFIQ-7 score, X is the treatment condition, β represents the model coefficients and dummy time will be dummy variables for the measurement time points, Y_{t0} is the baseline value of the dependent variable.

The treatment effect is estimated by the sum of the regression coefficient for the treatment variable and the coefficient for the interaction between the treatment and dummy for time at the primary end point ($\beta_1 + \beta_8$). The model will incorporate age and pessary user type as fixed effects and participant and recruitment centre as random effects. A random effect of participant is included at the level of the individual to account for the non-independence of observations under repeated measures. The covariance structure will be modelled as unstructured. The repeated measures model will also estimate mean differences in PFIQ−7 at 6 and 12 months. The three PFIQ-7 subscales will also be analysed separately using equivalent models.

Following advice previously given by the DMEC, further analysis will be considered if distributions of residuals are highly skewed. To assess whether the data meet the assumptions behind the mixed-effects model we will investigate normal quantile plots of residuals and standardized residuals. If the underlying assumptions behind the mixed-effects model analysis are violated, then we will explore data transformation and mixed effects negative binomial or zero-inflated Poisson models.

5.6 Secondary outcome measures

Secondary outcomes that have continuous measures will be analysed in a similar way to the primary outcome measure. However, the reporting of subscales will include only descriptive summaries. Binary and ordinal outcomes will estimate odds ratios from longitudinal logistic regression models with age, pessary type previous hysterectomy and centre as fixed effects.

For the Pessary Complications Questionnaire, all 15 individual items will be summarised and a mean difference will be estimated for the overall measure. Individual items will also be summarised for the two other non-validated questionnaires (pessary use and confidence), but overall scores will not be calculated and between-group comparisons will not be made for any items within these questionnaires.

For the uptake of additional support, separate binary logistic regression models will estimate odds ratios for telephone support and for additional clinic appointments. Both outcomes will consider the uptake of additional support over the 18-month follow-up period as a single time-point. Descriptive summaries will also be reported for the reasons for additional support.

Table A: Summary of inferential testing of secondary outcome measures

Secondary outcome	Time points reported descriptively by treatment arm	Time points for inferential test and method
Pelvic Floor Distress Inventory (PFDI-20)	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by longitudinal analysis of covariance
UDI-6 (subscale of PFDI-20)	Baseline, 6 months, 12 months and 18 months	No test
CRADI-8 (subscale of PFDI-20)	Baseline, 6 months, 12 months and 18 months	No test
POPDI-6 (subscale of PFDI-20)	Baseline, 6 months, 12 months and 18 months	No test
Pelvic Organ Prolapse/Urinary Incontinence Sexual Questionnaire (PISQ-IR)	Baseline, 18 months	Difference between arms at 18 months adjusted for baseline by longitudinal analysis of covariance
Number of women sexually active	Baseline, 18 months	No test

General Self-Efficacy	Baseline, 18 months	Difference between arms at 18 months adjusted for baseline by longitudinal analysis of covariance
Patient Global Impression Generated Index of Improvement (PGI-I)	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months. Ordinal regression
Pessary complications Bothersome Discharge	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Bothersome Smell	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Vaginal Pain	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Urine Infection	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Urine Incontinence	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Difficulty Emptying bladder	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Bowel Incontinence	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Difficulty Emptying Bowel	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Unable to remove pessary	Baseline, 6 months, 12 months and 18 months-	No test, self-management group only
Pessary complications Difficulty removing pessary	Baseline, 6 months, 12 months and 18 months-	No test, self-management group only
Pessary complications Difficulty Having Sex	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Pain During Sex	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Pessary Fell Out	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Non-Menstrual Bleeding	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications Other problem	Baseline, 6 months, 12 months and 18 months	No test
Pessary complications proportion	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by longitudinal analysis of covariance
Pessary use- used last 6 months	Baseline, 6 months, 12 months and 18 months	No test
Pessary use- You removed	Baseline, 6 months, 12 months and 18 months	No test
Pessary use- Often removed	Baseline, 6 months, 12 months and 18 months	No test

Pessary use- Why removed	Baseline, 6 months, 12 months and 18 months	No test
Pessary use- You inserted	Baseline, 6 months, 12 months and 18 months	No test
Pessary use- Continue pessary	Baseline, 6 months, 12 months and 18 months	No test
Pessary changes comfortable	Baseline, 6 months, 12 months and 18 months	No test
Pessary change convenient	Baseline, 6 months, 12 months and 18 months	No test
Pessary care acceptable	Baseline, 6 months, 12 months and 18 months	No test
Pessary confidence- improve symptoms	Baseline, 6 months, 12 months and 18 months	No test
Pessary confidence- avoid surgery	Baseline, 6 months, 12 months and 18 months	No test
Pessary confidence-benefit health and well-being	Baseline, 6 months, 12 months and 18 months	No test
Pessary confidence-manage problems	Baseline, 6 months, 12 months and 18 months	No test
Pessary confidence-remove pessary	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression
Pessary confidence-insert pessary	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression
Number of additional telephone calls for support	Baseline, 6 months, 12 months and 18 months	Difference between arms in total number of calls between baseline and 18 months
Number of additional clinic visits	Baseline, 6 months, 12 months and 18 months	Difference between arms in total number of additional appointments between baseline and 18 months
Adverse events (see section 5.11)	Within 18 months	No test

5.7 Subgroup analyses

Subgroup analyses of the primary outcome will be carried out within the following groups identified at baseline:

- Age (<65/65+)
- Pessary user type (new/existing)
- Previous hysterectomy (yes/no)

Stricter levels of overall statistical significance (P<0.01) will be sought, reflecting the exploratory nature of these analyses. Heterogeneity of treatment effects amongst subgroups will be tested for using the appropriate subgroup by treatment group by time interactions. (Wang 2007)

5.8 Crossover

Crossover will be initially identified if either of the following options has been recorded as a change of status:

- Woman was randomised to self–management but has reverted to standard care
- Woman was randomised to standard care but is now self-managing her pessary

A stricter definition of crossover for women randomised to standard care will require the same change of status as above and to have received TOPSY self-management teaching.

Sensitivity analyses of the primary outcome will be conducted which will take non-compliers into account using a complier average causal effect (CACE) analysis (Dunn 2003). This will be carried out using the above definitions for crossover. A further per protocol analysis of the primary outcome measure using the 'on treatment' definitions set out in section 4.3.5.

Centres (prior to TOPSY) offered different protocols in relation to self-management. For this reason, we will also summarise the numbers of crossovers at each centre.

5.9 Questionnaires retuned late

Questionnaire return times will be plotted for each follow up time point to describe the variation in return times. Separate plots will be produced for online and postal returns. There may be cases where participants responded late to a follow-up questionnaire. If a questionnaire is completed more than 3 months late (e.g. more than 9 months post-randomisation for a 6-month questionnaire), then that questionnaire will be excluded. Questionnaires that have been excluded because of being returned late may still be attributed to a later time point if the questionnaire for the later time point has not been received.

Cumulative distribution graphs will be created to summarise data received in relation to the following elements of data collection:

- By time point: 6, 12 and 18 months
- By response mode: online versus paper questionnaires
- Data collection pre-COVID versus during COVID

5.10 Post-randomisation exclusions

Participants who, after being randomised, are later found not to have met the eligibility criteria, could be classed as post-randomisation exclusions or protocol violations on the Change of Status log. However, as this is a pragmatic trial, their data will not be excluded from any analyses but we will report the number of such instances along with the reasons recorded.

5.11 Safety data

The number of serious adverse events (SAEs) reported and the proportion of women with an SAE will be summarised by treatment group. Information on severity, expectedness and causality will be presented.

6 Missing Data

6.1 Baseline data

Missing baseline data will not be imputed when reporting baseline characteristics, but imputation will be carried out prior to analysis in order to improve efficiency (White 2005). Centre mean imputation will be undertaken for continuous variables and an additional category will be created for categorical variables. However, we anticipate the impact of this to be minimal as the baseline data required for the analysis are mostly minimisation variables which will be complete.

6.2 Validated instruments

Imputation of missing data from validated outcome measures will be undertaken at item−level according to the rules of the specific instrument and carried out prior to analysis. Where an instrument has subscales (e.g. PFIQ-7), imputation will be carried out at the overall score level rather than subscale level. The trial office follows up partially completed questionnaires with sites, so the amount of missing data at subscale level is anticipated to be low.

6.3 Sensitivity analyses

Our primary analysis is by longitudinal analysis of covariance. Although this approach makes the assumption of data at the follow up points being missing at random, it is recommended that sensitivity analyses are carried out under differing assumptions of missingness (White 2005). Specifically, we will perform the following sensitivity analyses of the primary outcome:

- Observed cases at 18 months only (missing completely at random)
- Pattern mixture (repeated measures) models (missing not at random)
    - Non-responders assumed to have worse outcomes in both groups (and also in each group only).
    - Non-responders assumed to have better outcomes in both groups (and also in each group only).

  The missing data in the pattern mixture models will adjust imputed values (determined under a missing at random mechanism) by a value of 20 points where possible (equivalent to the size of the important difference used in the sample size calculation).

We will also undertake an exploratory investigation to describe particular characteristics of non-responders (e.g. differences in baseline PFIQ-7).

7 Additional analyses

7.1 Additional sensitivity analysis

When the protocol paper was submitted to The BMJ Open, a reviewer questioned how we were handling baseline PFIQ-7. We will therefore conduct a sensitivity analysis where the baseline response is included as part of the outcome vector rather than a covariate (Dinh 2011).

7.2 Data collected at clinic appointments

We will report a descriptive summaries of data recorded on the CRFs for baseline and 18 months, and also 6 and 12 months in the standard care group. This will include pessary type (and material) used and how this changes during the intervention period, with a sub-category to indicate use post-hysterectomy to address an evidence gap relating to pessary types used in hysterectomised women.

8 COVID-19 pandemic

The UK entered a prolonged period of lockdown in March 2020 at a time when recruitment had ended but the majority of participants were still in the active/intervention phase of the trial and being followed up. The lockdown resulted in the widespread cancellation of pessary clinics and difficulties for some participants to receive, complete and return questionnaires at the right time. This led to changes to both intervention delivery and data collection, which in turn has some implications for the analysis.

8.1 Crossover

A participant in the standard care group could miss both 6-month and 12-month appointments and therefore essentially be self-managing. This in turn could lead to an underestimate of effectiveness of self-management, although this issue will be mitigated as many centres have replaced their face-to-face appointments with telephone consultations. Contrastingly, the standard care group during lockdowns could have little or no care and no self-management, which could in turn lead to an overestimate of the treatment effect. The analysis outlined in section 5.8 relating to crossover and not being 'on treatment' is designed to address these issues.

8.2 Mode of data collection

The pandemic has led to a change in the way some questionnaire data are collected, which can lead to biased results if not addressed in the analysis (Hood 2012). Some outcome data will now be collected by telephone and a revised 18-month questionnaire has been developed to facilitate this. Sensitivity analysis of the primary outcome will be conducted by adding mode of collection to the model, with and without an interaction with treatment allocation. Principal stratification will be used as the models will include post-randomisation covariate data (actual rather than planned mode of collection).

8.3 Timing of data collection

It is anticipated that many questionnaires could be returned late, either because participants are unable to return them due to the lockdown, or because they received them late as a result of staff being unable to access the trial office to send out questionnaires on time. Restrictions to accessing the trial office may also necessitate some questionnaires being sent early. It is also anticipated that many vaginal examinations at 18-months will be delayed. Section 5.9 sets out rules for handling and excluding data collected outside a 3-month window, but an additional sensitivity analysis of the primary outcome will be conducted where these rules will be relaxed.

8.4 18-month clinic appointment

During the COVID-19 restrictions, the 18-month clinic appointment was split into two sections. The first part is a telephone call when the participant reaches 18 months and the second part (including the vaginal examination) is a face-trace appointment when restrictions are lifted. This has led to two new CRFs with some data collected at each time point, specifically Section B (pessary related symptoms) and Section E (planned care). Where there has been duplication of data collection, the descriptive summary of Section B will use data from the preliminary telephone call (as we are predominantly interested in symptoms specifically at 18 months) but the descriptive summary of Section E will use data from the later appointment (as we are predominantly interested in the plan for long-term care rather than interim measures).

8.5 Uptake of telephone support

A sensitivity analysis for the uptake of telephone support restricting the analysis to outcomes prior to the beginning of the first lockdown on 23 March 2020.

8.6 COVID survey

A survey of a subsample of participants was conducted in 2020, a retrospective addition to the project, due to the pandemic. The survey was sent to women who missed a clinic appointment due to COVID. The objective was to investigate the extent to which the pandemic had an impact on attitudes to pessary care. Descriptive summaries of the results will be tabulated, along with demographic characteristics of those participants who responded (see Table 7).

8.7 Impact on Primary outcome measure

The primary outcome measure asks about impact of symptoms on activities. Some of these activities were unavailable during lockdown. For example, “Entertainment activities such as going to a movie or a concert”. We will summarize responses to individual items on the PFIQ-7 by lockdown status to determine whether there was an effect on responses. We will also examine rates of missing data for individual PFIQ-7 items in and out of lockdown.

9 Dummy tables

Table 1: Participant characteristics by minimisation factors

Table 1 shows the distribution of participant characteristics at baseline by treatment condition. These variables are taken from the baseline CRF.

		n	% of N
Recruitment centre	Manchester (St. Mary's)		
	Middlesbrough		
	Birmingham		
	Norwich		
	Taunton		
	Croydon		
	Addenbrookes		
	Ayrshire and Arran		
	Basingstoke		
	Fife		
	Yeovil		
	NHS Lanarkshire		
	Newcastle		
	Lothian		
	Sheffield		
	County Durham		
	Liverpool		
Pessary User Type	New user		
	Existing user		
Age Group	< 65 years		
	≥ 65 years		

Table 2: Participant characteristics at baseline by treatment condition

	Treatment condition					
	Self-management		Standard care		Total	
	n	%	n	%	n	%
Hormone Therapy						
Yes						
No						
Total		100.0%		100.0%		100.0%
Systemic HRT						
Yes						
No						
Total						
Oestrogen						
Yes						
No						
Total						
Chronic Cough						
Yes						
No						
Total						
Diabetes						
Yes						
No						
Total						
Arthritis						
Yes						
No						
Missing						
Total						
Constipation						
Yes						
No						
Total						
Recurrent UTIs						
Yes						
No						
Total						
Vulvodynia						
Yes						
No						
Total						
Hysterectomy						
Yes						
No						

Missing						
Total						
Pelvic Floor Surgery						
Yes						
No						
Total						
Other Co-Morbidity						
Yes						
No						
Missing						
Total						
Inflammation Of Tissues						
Yes						
No						
Missing						
Total						
Ulceration						
Yes						
No						
Missing						
Total						
Granulation						
Yes						
No						
Missing						
Total						
Other Clinical Concerns						
Yes						
No						
Missing						
Total						
Bothersome Discharge						
Yes						
No						
Total						
Bothersome Smell						
Yes						
No						
Total						
Vaginal Pain						
Yes						
No						
Total						
Other Pain						

Yes						
No						
Missing						
Total						
Urine Infection						
Yes						
No						
Total						
Urine Incontinence						
Yes						
No						
Total						
Difficulty Emptying bladder						
Yes						
No						
Total						
Bowel Incontinence						
Yes						
No						
Missing						
Total						
Difficulty Emptying Bowel						
Yes						
No						
Total						
Sexually Active						
Yes						
No						
Missing						
Total						
Difficulty Having Sex						
Yes						
No						
Not applicable						
Missing						
Total						
Pain During Sex						
Yes						
No						
Not applicable						
Missing						
Total						
Pessary Fell Out						
Yes						

No						
Total						
Non-Menstrual Bleeding						
Yes						
No						
Total						

Table 3: Distribution of continuous variables at baseline by treatment condition

Table 3 shows the distribution of continuous variables at baseline by treatment condition from the CRF

	Treatment condition					
	Self-management		Standard care		Total	
	n Mean (SD)	Median (IQR) range	n Mean (SD)	Median (IQR) range	n Mean (SD)	Median (IQR) range
Number of births (parity)						
Age						
BMI						
N						

Table 4: Distribution of sociodemographic variables at baseline by treatment condition

Table 4 shows the distribution of continuous variables at baseline by treatment condition from the participant details table. ***These are likely to contain moderate levels of missing data**

	Treatment condition					
	Self-management		Standard care		Total	
	n	Col %	n	Col %	n	Col %
Educational Qualification						
No formal qualifications						
Secondary/further qualifications						
Higher education						
Missing						
Total						
Current Employment Status						
Full time employment						
Part time employment						
Student						
Housework						
Seeking work						

Other						
Missing						
Total						
Ethnic Group						
Indian						
Caribbean						
African						
Any other Black background						
British						
Irish						
Any other White background						
White and Black Caribbean						
would prefer not to say						
Missing						
Total						

Table 5: Primary analysis: longitudinal analysis of covariance on PFIQ-7 at 6, 12 and 18 months

Variable	**Coefficient**	**95% CI**	**P value**
Fixed effects:			
Random effects:			

Table 6: Sensitivity analyses of the primary outcome

Type	SAP section	Sensitivity analyses	Effect size (MD, 95% CI)
Covariates	5.1	A sensitivity analysis of the primary outcome will be conducted with previous hysterectomy included as an additional fixed effect.	
Per protocol (Cross over)	5.8	Sensitivity analyses of the primary outcome will be conducted which will take non-compliers into account using a complier average causal effect (CACE) analysis. This will be carried out using the definitions for crossover in 5.7	
Per protocol (on treatment)	4.3.5/5.8	A further per protocol analysis of the primary outcome measure using the 'on treatment' definitions set out in section 4.3.5.	
Missing data assumptions	6.3	Observed cases at 18 months only (missing completely at random)	
Missing data assumptions	6.3	Pattern mixture (repeated measures) models (missing not at random) Non-responders assumed to have worse outcomes in both groups (and also in each group only). Non-responders assumed to have better outcomes in both groups (and also in each group only).	
Mode of data collection	8.2	With mode of data collection added to the model as a fixed effect with and without an interaction with treatment allocation. Principal stratification will be used as the models will include post-randomisation covariate data (actual rather than planned mode of collection).	
Time frame for exclusion of late returns	8.3	Section 5.9 sets out rules for handling and excluding data collected outside a 3-month window, but an additional sensitivity analysis of the primary outcome will be conducted where these rules will be relaxed.	
Analysis model	7.1	Baseline response is included as part of the outcome vector rather than a covariate-	

Table 7: COVID Survey

	Treatment condition					
	Self-management		Standard care		Total	
	n	%	n	%	n	%
When your pessary appointment was cancelled were you given clear instructions on what to do if there was a problem with your pessary?						
Yes						
No						
Total		100.0%		100.0%		100.0%
Type of contact						
Phone						
NHS connections						
MS Teams						
Letter						
Zoom						
Other						
Total						
How worried?						
Very high						
High						
Moderate						
Low						
Very low or none at all						
Total						
Seek advice?						
GP						
Practice nurse						
Physiotherapist						
Dietician						
Total						
Bothersome Discharge						
Yes						
No						
Total						
Bothersome Smell						
Yes						
No						
Total						
Vaginal Pain						
Yes						
No						
Total						
Other Pain						
Yes						

No						
Missing						
Total						
Urine Infection						
Yes						
No						
Total						
Urine Incontinence						
Yes						
No						
Total						
Difficulty Emptying bladder						
Yes						
No						
Total						
Bowel Incontinence						
Yes						
No						
Missing						
Total						
Difficulty Emptying Bowel						
Yes						
No						
Total						
Difficulty Having Sex						
Yes						
No						
Not applicable						
Missing						
Total						
Pain During Sex						
Yes						
No						
Not applicable						
Missing						
Total						
Pessary Fell Out						
Yes						
No						
Total						
Non-Menstrual Bleeding						
Yes						
No						
Total						

Did you remove your pessary yourself while waiting on your appointment?						
Yes						
No						
Total						
Did the delay in having a clinic appointment change your attitude to pessary use?						
Yes						
No						
Total						
Did the delay in having a clinic appointment change your attitude to self- management?						
Yes						
No						
Total						
Has the covid-19 pandemic changed how you feel about your prolapse symptoms?						
Yes						
No						
Total						
Has the covid-19 pandemic changed how you feel about returning to your usual clinic based hospital care?						
Yes						
No						
Total						

10 References

1. Hagen S, Kearney R, Goodman K, Melone L, Elders A, Manoukian S, Agur W, Best C, Breeman S, Dembinsky M, Dwyer L, Forrest M, Graham M, Guerrero K, Hemming C, Khunda A, Mason H, McClurg D, Norrie J, Karachalia-Sandri A, Thakar R, Bugge C (2020). Clinical and cost-effectiveness of vaginal pessary self-management compared to clinic based care for pelvic organ prolapse: protocol for the TOPSY randomised controlled trial. *Trials* [in press].

2. Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., Williamson, P.R., Altman, D.G., Montgomery, A., Lim, P. and Berlin, J., 2017. Guidelines for the content of statistical analysis plans in clinical trials. JAMA, 318(23), pp.2337-2343.

3. Wiegersma, M., Panman, C.M., Kollen, B.J., Berger, M.Y., Lisman-Van Leeuwen, Y. and Dekker, J.H. (2014) Effect of pelvic floor muscle training compared with watchful waiting in older women with symptomatic mild pelvic organ prolapse: randomised controlled trial in primary care. British Medical Journal, 349: g7378.

4. Panman, C.M., Wiegersma, M., Kollen, B.J., Berger, M.Y., Leeuwen, Y.L., Vermeulen, K.M. and Dekker, J.H. (2016) Effectiveness and cost-effectiveness of pessary treatment compared with pelvic flor muscle training in older women with pelvic organ prolapse: 2-year follow-up of a randomized controlled trial in primary care. Menopause. 23(12), 1307-1318.

5. Barber, M.D., Walters, M.D. and Bump, R.C. (2005) Short forms of two condition specific quality of life questionnaires for women with pelvic floor disorders (PFDI-20 & PFIQ-7). American Journal of Obstetrics and Gynecology. 193, pp.103-113.

6. Barber, M.D., Chen, Z., Lukacz, E., Markland, A., Wai, C., Brubaker, L., Nygaard, I., Weidner, A., Janz, N.K. and Spino, C., 2011. Further validation of the short form versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ). Neurourology and Urodynamics, 30(4), pp.541-546.

7. Rogers RG, Rockwood TH, Constantine ML, Thakar R, Kammerer-Doak DN, Pauls RN, Parekh M, Ridgeway B, Jha S, Pitkin J, Reid F. A new measure of sexual function in women with pelvic floor disorders (PFD): The Pelvic Organ Prolapse/Incontinence Sexual Questionnaire, IUGA-Revised (PISQIR). International Urogynecology journal. 2013 Jul 1;24(7):1091-103.

8. Constantine ML, Pauls RN, Rogers RR, Rockwood TH (2017). Validation of a single summary score for the Prolapse/Incontinence Sexual Questionnaire–IUGA revised (PISQ-IR). International Urogynecology Journal, 28(12), 1901-1907.

9. Schwarzer, R. and Jerusalem, M. (1995) Generalized Self-Efficacy scale. In J. Weinman, S. Wright, and M. Johnston. Measures in health psychology: A user's portfolio. Causal and control beliefs. Windsor, England: NFER-Nelson, 35-37.

10. Chen G, Gully SM, Eden D (2001). Validation of a new general self-efficacy scale. Organizational Research Methods, 4(1), 62-83.

11. Sacomori, C., Cardoso, F. L., Porto, I. P., & Negri, N. B. (2013). The development and psychometric evaluation of a self-efficacy scale for practicing pelvic floor exercises. Brazilian Journal of Physical Therapy, 17(4), 336-342.

12. Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. American Journal of Obstetrics and Gynecology 2003; 189: 98−101.

13. Srikrishna, S., Robinson, D. and Cardozo, L. (2010) Validation of the Patient Global Impression of Improvement (PGI-I) for urogenital prolapse. International Urogynaecology Journal. 21(5), 52-528.

14. Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84 (2), 19215.

15. Twisk J, Bosman L, Hoekstra T, Rijnhart J, Welten M, Heymans M (2018). Different ways to estimate treatment effects in randomised controlled trials. Contemporary Clinical Trials Communications, 10: 80-85.

16. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine − reporting of subgroup analyses in clinical trials. New England Journal of Medicine 2007; 357: 2189−94.

17. Dunn G, Maracy M, Dowrick C, Ayuso-Mateos JL, Dalgard OS, Page H, Lehtinen V, Casey P, Wilkinson C, Vázquez-Barquero JL, Wilkinson G (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. The British Journal of Psychiatry, 183(4), 323-331.

18. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Statistics in Medicine 2005; 24(7): 993-1007.

19. Dinh P & Yang P (2011). Handling baselines in repeated measures analyses with missing data at random. Journal of Biopharmaceutical Statistics, 21(2), 326-341.

20. Hood K, Robling M, Ingledew D, Gillespie D, Greene G, Ivins R, Russell I, Sayers A, Shaw C, Williams (2012). Mode of data elicitation, acquisition and response to surveys: a systematic review. J. Health Technol Assess;16(27):1-162.

TOPSY STATISTICAL ANALYSIS PLAN LIST OF DEVIATIONS

Version 1

23rd March 2022

This document lists deviations to the main Trial SAP and additional data issues that occurred since data lock

Prepared by:

Andrew Elders [redacted] Date: 23/03/2022
Senior Statistician

Cath Best [redacted] Date: 23/03/2022
Trial Statistician

Approved by:

Carol Bugge [redacted] Date: 23/03/2022
Chief Investigator

1. **SAP Deviations**

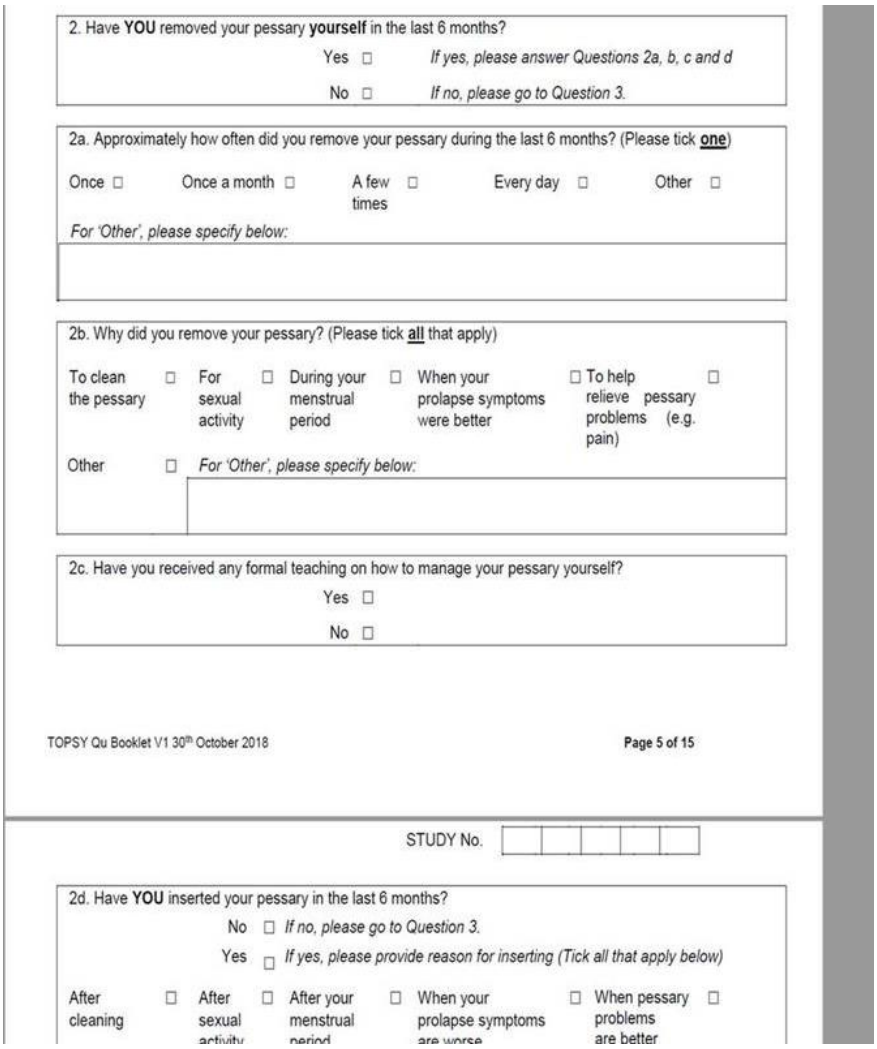
Section 4.3.5 of TOPSY SAP v2.0 dated 12th May 2021.

The statistician highlighted the following issue she noticed during the analysis (on 06th October 2021);

ON treatment definition for the standard care group states;

'AND they have ticked NO to Q2d (Pessary use questionnaire) "Have you inserted your pessary in the last 6 months" at ALL time-points (6, 12 and 18 months). '

However on analysis that there is a very large amount of missing data for 2d because if they answer No to Q2 "Have YOU removed your pessary yourself in the last 6 months?" they are routed straight to Q3. (see screen shot below).



2. Have **YOU** removed your pessary **yourself** in the last 6 months?

Yes ☐ *If yes, please answer Questions 2a, b, c and d*

No ☐ *If no, please go to Question 3.*

2a. Approximately how often did you remove your pessary during the last 6 months? (Please tick **one**)

Once ☐ Once a month ☐ A few times ☐ Every day ☐ Other ☐

For 'Other', please specify below:

2b. Why did you remove your pessary? (Please tick **all** that apply)

To clean the pessary ☐ For sexual activity ☐ During your menstrual period ☐ When your prolapse symptoms were better ☐ To help relieve pessary problems (e.g. pain) ☐

Other ☐ *For 'Other', please specify below:*

2c. Have you received any formal teaching on how to manage your pessary yourself?

Yes ☐

No ☐

TOPSY Qu Booklet V1 30th October 2018 **Page 5 of 15**

STUDY No.

2d. Have **YOU** inserted your pessary in the last 6 months?

No ☐ *If no, please go to Question 3.*

Yes ☐ *If yes, please provide reason for inserting (Tick all that apply below)*

After cleaning ☐ After sexual activity ☐ After your menstrual period ☐ When your prolapse symptoms are worse ☐ When pessary problems are better ☐

If we strictly adhere to the definition of 'On treatment' then there are only 5 women in the standard care group on treatment.

It was agreed that the data would be recoded for all women who tick 'no' to Q2 as 'no' to Q2d (since if they have not removed it they won't have reinserted it either) and this would be a post-hoc analysis on the basis of that we only viewed the whole database after data lock.

Section 5.6 of TOPSY SAP v2.0 dated 12th May 2021; Table A

In regards to tests on the pessary confidence questionnaire data. In the SAP it stated that the difference between arms would ONLY be done in the last 2 questions (snap shot of table A below).

Pessary confidence-manage problems	Baseline, 6 months, 12 months and 18 months	No test
Pessary confidence-remove pessary	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression
Pessary confidence-insert pessary	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression

As a post Hoc analysis, the difference between arms at 18 months should be applied to Q4 "You can manage problems related to using a pessary" This was a simple oversight at the time of writing the SAP.

Pessary confidence-manage problems	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression
Pessary confidence-remove pessary	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression
Pessary confidence-insert pessary	Baseline, 6 months, 12 months and 18 months	Difference between arms at 18 months adjusted for baseline by regression

2. Data changes after the TOPSY data cut

a. Blank forms

The following CRF 08s are being line listed but no data is included (sometimes this happens if centre goes to enter a form then realises it is not that form etc). We have checked these and confirmed not a true entry. These won't be counted in any numbers and no further action required.

- 12055-
- 15335
- 16018
- 18006
- 33014

b. Amendments to analysis files (no change to original database)

The following were removed from the analysis files post data lock:
15145 has an entry for additional telephone support on 1/6/20 (CRF 07 which is same as 18 month CRF call. Thus CRF 07 in this instance was entered to the database in error. Additional call removed from analysis

12003: clinic visit log for 12003 dated 08/03/2021 but their 18-month clinic visits was 19/02/2020. This should have been deleted and has not been. Clinic visit removed from analysis

22053 – SAE and AE added to database on 23/11/2021 for hip replacement and CVA post 18 months follow up. Not included in analysis

33013 – CRF 18 – was only entered to database in Dec 2021. Centre didn't send before datalock. Not included in analysis.

In response to post-data lock queries the following data fields were changed:
RandomisationDate changed to 02jan2019 for StudyNo 16055
RandomisationDate changed to 01may2019 for StudyNo 26013
DateCompleted for PFIQ7 changed to 28mar2019 for StudyNo==16027 & Timepoint==1
DateCompleted for PFIQ7 changed to 29dec2019 for StudyNo==22024 & Timepoint==1
DateCompleted for PFIQ7 changed to 03oct2019 for StudyNo==16027 & Timepoint==2
DateCompleted for PFIQ7 changed to 14sep2020 for StudyNo==15357 & Timepoint==3
DateCompleted for PFIQ7 changed to 04nov2019 for StudyNo==19007 & Timepoint==0

DateCompleted for PFIQ7 changed to 10dec2019 for StudyNo==26016 & Timepoint==1
DateCompleted for PFIQ7 changed to 09oct2019 for StudyNo==24030 & Timepoint==0
DateCompleted for PFIQ7 changed to 01oct2020 for StudyNo==34031 & Timepoint==1

SAE DateofEvent changed to 9/12/2020 for 21017