STATISTICAL ANALYSIS PLAN v1.1

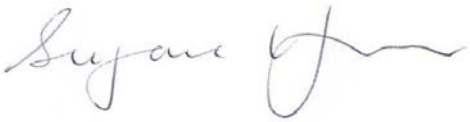
4 June 2018

Prepared by:

Andrew Elders

Senior Statistician

Approved by:



Suzanne Hagen

Chief Investigator

ISRCTN57746448

Table of Contents

1. Study Design

OPAL is a two−arm multicentre randomised controlled trial of the effectiveness of biofeedback−mediated intensive pelvic floor muscle training (PFMT) versus basic PFMT for female stress or mixed urinary incontinence (UI). Full details of the study design are published in the OPAL trial protocol.[1]

2. Sample Size and Power Calculation

There were no published long−term data on the primary outcome measure, the ICIQ−UI Short Form score, in a similar population which could have been used to inform a sample size calculation. Studies including women with stress/mixed UI report have reported mean scores at baseline of around 10 with SD of around 5[2,3]. It is possible that the SD at 24 month follow−up in OPAL will be greater than 5 and perhaps as large as 10. Assuming a clinically meaningful difference of 3 points on the ICIQ−UI−SF score (e.g. change from leaking urine “once a day” to “never”), which is similar to the minimal clinically important difference of 2.5 reported in a study of older women[4], and SD of 10, a sample size of 234 per group would detect this difference (standardised effect size of 0.3) with 90% power at the 5% level of significance (2−sided alpha). Allowing for 22% loss to follow−up means that 300 participants need to be recruited to each group (total sample size of 600).

3. Outcome Measures

All outcomes are measured using responses from participant follow−up questionnaires at 6, 12 and 24 months (unless otherwise indicated). Outcomes are also measured at baseline, where applicable.

3.1 Primary outcome

The primary clinical outcome is UI severity at 24 months, measured using the ICIQ−UI Short Form questionnaire[7], which has been developed by the International Consultation on Incontinence.

The ICIQ−UI−SF score is composed from responses to the following three questions in the participant questionnaire:

- Q1: How often do you leak urine? (A9 in questionnaire)
  [never=0, once a week=1, 2/3 times a week=2, once a day=3, several times a day=4, all the time=5]
- Q2: How much urine do you usually leak? (A10 in questionnaire)
  [none=0, small=2, moderate=4, large=6]
- Q3: Overall, how much does leaking urine interfere with your everyday life? (A16 in questionnaire)
  [not at all=0 through to 10 a great deal]

The ICIQ−UI−SF measurement is the sum of the three scores and will be treated as continuous data. Possible scores range from 0 to 21.

A missing response to any of these three questions will mean that the overall score will be treated as missing. However, if the response is 'never' to Q1 and missing for Q2, then Q2 will be assumed to be 'none'. Similarly, if the response is missing for Q1 and 'none' to Q2, then Q1 will be assumed to be 'never'. Also, if the response is 'never' to Q1, 'none for Q2 and missing for Q3, then Q3 will be assumed to be zero.

Scores from 0 to 12 can be considered as mild/moderate incontinence and scores from 13 to 21 as severe incontinence[8].

3.2 Secondary outcomes

3.2.1 Urinary outcomes

- ICIQ−UI Short Form questionnaire[7] score at 6 and 12 months (continuous data).
- Patient Global Impression of Improvement in UI (PGI−I)[10], ordinal measure scored on a 7−point scale (very much better through to very much worse). This is not asked at baseline, and instead the Patient Global Impression of Severity (PGI−S)[10] is asked and will be reported.
- Proportions of women with UI cured and UI improved, derived from the ICIQ−UI Short Form questionnaire[7]. Cure will be defined as a "never" response to the first question and a "none" response to the second. Improvement will be defined as reduction from baseline in the ICIQ−UI−SF score of three points or more, based on a minimum

clinically important difference of 2.5[4]. Missing item-level data will be handled in the same way as described in section 3.1 when deriving these outcomes.

Uptake of surgery for UI (i.e. a 'Yes' response to question G5b). Treatment rates will be described and analysed separately for months 0-6, months 7-12, and months 13-24 (along with an overall treatment rate for months 0-24 for women who respond to all three follow-up questionnaires). The description and date of the surgery (recorded in question G5c on the questionnaire) will be checked to ensure that events are not double counted as a result of a respondent reporting the same event on more than one questionnaire.

- Uptake of other treatment for UI: treatment rates will be determined separately for hospital admissions, outpatient consultant attendances (combining NHS hospital visits and private treatment), GP consultations, nurse appointments (combining GP practice visits, NHS hospital visits and private treatment), physiotherapy appointments (combining NHS hospital visits and private treatment), medication and other treatment/advice. Specifically, the rates being reported will relate to questionnaire responses as follows:
    - Hospital admission: 'Yes' to first part of question G5
    - Outpatient consultant appointment: 'Yes' in G3 or G4 only where it relates to a hospital doctor
    - GP consultation: A non-zero response to G1a only where is relates to urine leakage
    - Nurse appointment: : A non-zero response to G2a only where is relates to urine leakage, or 'Yes' in G3 or G4 only where it relates to a nurse
    - Physiotherapy appointment: 'Yes' in G3 or G4 only where it relates to a physiotherapist
    - Medication: 'Yes' to G6
    - Other treatment/advice: 'Yes' to G7

  The uptake rate of any treatment (i.e. if any of the aforementioned categories of treatment have been accessed) will also be measured, as a binary outcome. As for surgery for UI, treatment rates will be described and analysed separately for months 0-6, months 7-12, and months 13-24 (along with an overall treatment rate for months 0-24 for women who respond to all three follow-up questionnaires).

- Other urinary symptoms (ICIQ−FLUTS[11], with further details available at www.iciq.net/ICIQ.FLUTS.html). Each individual FLUTS response item comprises five levels of response and will be scored from 0 to 4. There are three subscales (all of which will be treated as continuous variables):
    - ICIQ-FLUTS Filling Score with possible range from 0 to 16 (4 response items: nocturia, urgency, bladder pain, diurnal frequency), therefore using the same approach as the INVESTIGATE study[12]
    - ICIQ-FLUTS Voiding Score with possible range from 0 to 12 (3 response items: hesitancy, straining, intermittency).
    - ICIQ-FLUTS Incontinence Score, constructed from the following 5 questionnaire items:
        - A8: Does urine leak before you can get to the toilet? (urgency)
        - A9: How often do you leak urine? (frequency)
        - A11: Does urine leak when you are physically active, exert yourself, cough or sneeze? (stress UI)
        - A12: Do you ever leak urine for no obvious reason and without feeling that you want to go? (unexplained UI)
        - A13: Do you leak urine when you are asleep? (enuresis)

      The ICIQ-FLUTS Incontinence Score is the sum of the scores from these five questions, with a possible range from 0 to 20. Responses to "How often do you leak urine?" will be recoded from the six levels required for the ICIQ-UI−SF score to the five levels required for the ICIQ-FLUTS measure, by merging “several times a day” and “all the time” responses into a single category.

3.2.2 Quality of life outcomes

4 UI-specific quality of life measured by the ICIQ−LUTSqol[13] (King’s Health Questionnaire, www.iciq.net/ICIQ.LUTSqolmodule.html). This is a 19 item scale with each of the questions from Q1 to Q19 scored from 1 to 4 and then summed to give an overall score with a possible range from 19 to 76. A score of 1 is applied to any question where the response is missing, and also in questions 7, 8 and 9 if a response of ‘not applicable is given. If there are missing values for the majority of

questions (i.e. at least 10), then the overall score will be missing. There is an additional 20th question (bother scale) which will be analysed as a stand-alone measure. The King's Health Questionnaire replaces the Wagner scale (ICIQ−UIqol)[14] which was proposed in the original OPAL protocol.

5 General health measured by the EQ−5D−3L score[15] (which ranges from −0.654 to 1) and the EQ−5D visual analogue score (which ranges from 0 to 100). EQ−5D will also be used in the economic analysis (Economic outcomes will also be measured and are described further in the Health Economic Analysis Plan).

All quality of life measures will be treated as continuous data.

3.2.3 Pelvic floor outcomes

- Prolapse symptoms measured by the Pelvic Organ Prolapse Symptom Score (POP−SS)[16]. There are seven frequency-based questions, each scored from 0 to 4. The POP-SS measure is the sum of these scores with a possible range from 0 to 28.
- Bowel symptoms based on an early version of the ICIQ−BS short form[17]. The development and validation of an ICIQ short bowel symptom questionnaire is no longer being carried out, so each of the six bowel symptom questions will be handled as separate measures.
- Pelvic floor muscle function (Oxford scale[18], ICS method[19]) using PERFECT measurements (Power, Endurance, Repetitions, Fast Contractions, Every Contraction Timed[16]) taken as part of the six-month clinic assessment. This contains the following five measures:
    - Power (maximum voluntary contractions) for both fast and slow contractions, for strongest side only. A higher value represents greater strength. Where a measurement is provided for both left and right sides, the strongest side is the maximum of the left and right measurement (NB the strongest side could be different for fast and slow, and between time points). Power is recorded as an ordinal measure (0, 1,2,3,4 and 5) and values of 3+, 4- etc can be recorded. Values appended with a +/-, however, will not be analysed as a separate level of response, but will be analysed as the nearest whole number. Power (slow) is also recorded on the baseline Clinical Assessment Form (CAF) (where any pelvic floor muscle function data is recorded on both

the CAF and the Therapy Assessment Form (TAF), then CAF data will be used). It should be noted that the Oxford Scale itself refers only the slow power measurement.

- Endurance (in seconds), for slow contraction only, and for strongest side only. A value of 10 seconds can mean 10 seconds or more. Any values greater than 10 seconds will be analysed as 10.
- Repetitions, during both fast and slow contractions, for strongest side only. A value of 10 can mean 10 repetitions or more. Any values of greater than 10 will be analysed as 10.

For endurance and repetitions, measurements should have only been recorded for the side with the strongest power. However, if a measurement is provided for both sides, then the data for the analysis should be taken from the side with the larger power measurement (even if the endurance/repetitions measurement is lower in this side). If both sides have equal power and the endurance/repetitions measurement is given for both sides, then the maximum of the left and right endurance/repetitions measurement should be used.

Only the slow power measure will be compared in the analysis, but descriptive summaries of the other outcomes will be reported.

- Self−efficacy for PFMT measured by the Pelvic Floor Muscle Exercise Self−Efficacy Scale)[20]. This is a 17 item scale with each item scored from 1 to 5 and then summed to give an overall score with a possible range from 17 to 85. All pelvic floor outcome measures will be treated as continuous data, other than the bowel symptoms which are categorical ordinal variables.

3.2.4 Adherence

There are several potential measures of adherence, including attendance at appointments and initial uptake rates (from clinic assessment data), adherence using data recorded in the exercise diary and biofeedback devices, and long-term adherence using follow−up questionnaire data.

The theory is that biofeedback is educational and motivational, and the content of the biofeedback protocol is underpinned by the Information-Motivation-Behavioral Skills (IMBS) model of behaviour change[21]. We hypothesise that the introduction to biofeedback use in the clinic can

provide the technique in terms of behavioural skill, and subsequent biofeedback use is evidence of motivation. There might also be a dose response relationship between the frequency of biofeedback use and UI severity.

We will therefore measure the following outcomes using data from the TAF completed by the therapist during appointments unless otherwise stated:

- Introductory teaching of pelvic floor muscle exercises (PFME) for those randomised to basic PFMT, or introduction to biofeedback use for those randomised to biofeedback, at either the first or second clinic appointment (single binary outcome). The data items to identify this in the basic PFMT group are a ‘Yes’ response to ‘Teach PFM contraction’ in the Visit 1 Checklist or a ‘Yes’ response to ‘1/2/3 sets of PFM contractions’ in the Visit 2 Checklist. The data items to identify this in the biofeedback group are a ‘Yes’ response to ‘Teach probe and electrode insertion/removal’ in the Visit 1 Checklist or a ‘Yes’ response to ‘Woman inserts/removes probe and electrode’ in the Visit 2 Checklist.
- PFMT / biofeedback use in clinic (binary outcome). This is skill rehearsal being practiced (under clinician supervision) during at least one clinic appointment attended after the introductory session(s). The checklist data item to identify this is a 'Yes' response to '1/2/3 sets of PFM contractions' (in the basic PFMT group) or a 'Yes' response to 'BF used throughout practice session' (in the biofeedback group). Completion rates for these checklist items will also be reported. If the checklist item is missing (in the biofeedback group), then either a 'Yes' response to 'Periform provided with instructions on use, cleaning etc' or 'Patient introduced to clinic biofeedback' will be used.
- PFMT / biofeedback use at home (binary outcome), confirmed during at least one appointment subsequent to the introductory appointment(s). The data item to identify this is a ‘Yes’ response to TAF question 4.1 for ‘Exercise programme followed’ (in the basic PFMT group) or a ‘Yes’ response to TAF question 4.5 for ‘BF programme followed’ (in the biofeedback group). If there are any missing values in these TAF data items, then home exercise will be determined by a non-zero integer for ‘number of PFM sessions (not using BF)’ in the exercise diary (in the basic PFMT group) or either a non-zero integer for ‘number of PFM sessions (using BF)’ in the

exercise diary or a ‘Yes’ to TAF question 4.6 for ‘BF home stats downloaded?’ (in the biofeedback group).

- Number of clinic appointments attended (count data, up to the six offered within the study).
- Frequency of PFME at 6, 12 and 24 months using questionnaire data relating to the previous:
    - day (ordinal data derived from E3 and E4)
    - week (count data derived from E5 and E6)
    - month (ordinal data derived from E1 and E2)

4. Statistical Methods

4.1 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, crossover or non-adherence.

Descriptive statistics will be tabulated by treatment group (see Tabulations section) 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.

The main analysis will be a complete case analysis, where a participant is included in an analysis of a particular outcome if the participant has observed data for that outcome. Estimates of treatment effect between groups (i.e. the effect of intensifying PFMT) will be estimated using generalised linear models for all outcomes. Statistical significance will be at the 5% level.

All models will adjust for the following factors (all of which are minimisation covariates except for the type of therapist):

- Centre number (categorical variable)
- Age in years
- Type of UI (categorical variable: stress UI / mixed UI)
- UI severity[7] at baseline. Data will be used from Clinical Assessment Form used for randomisation rather than baseline questionnaire (for

completeness), and the ICIQ-UI-SF score will be used (rather than the severity categorisation).

- Type of therapist (physiotherapist / nurse / other). 'Other' therapist type includes combination of therapist types during the period of treatment, or consultant, or midwife, or no therapist. If the therapist type is missing for one appointment but is the same for all the other appointments, then the missing appointment will be assumed to have same therapist type as the other appointments.

It is possible for a woman to receive, across the series of her appointments, treatment from a combination of a physiotherapist and a nurse. However, it is expected that the type of therapist will usually be the same for all appointments, because it is normal practice for women to be treated by the same therapist for all appointments and many centres only employ one type of therapist. The name of the therapist at each appointment will be collected on the TAF and the therapist type will be recorded on the database.

Treatment effect estimates between groups will be the mean difference for continuous outcomes and odds ratios for binary and ordinal outcomes. Effect sizes will be determined by the regression coefficient for randomised group from the generalised linear model, and will be presented with 95% confidence intervals. Statistical tests to compare groups at baseline will not be undertaken.

Questionnaire responses will be sought at 6, 12 and 24 months after randomisation. Six-month questionnaires will be excluded from the analysis if completed more than three months before or after the time point at which the questionnaire was issued, and 12-month and 24-month questionnaires will be excluded if completed more than six months before or after the questionnaire being issued. 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.

A single final analysis is anticipated 24 months after the last woman has been recruited. Statistical analysis will be conducted, where possible, using Stata (StataCorp, College Station, TX, USA).

4.2 Primary outcome

The ICIQ−UI−SF score at 24 months will be analysed using a linear mixed model. The model will regress the ICIQ−UI−SF score on an indicator variable for the treatment group (coded as 1 if randomised treatment was biofeedback and 0 for basic PFMT) and will adjust for minimisation covariates and therapist type, along with the baseline ICIQ−UI−SF score[22]. The baseline Clinical Assessment Form will be used to provide the baseline ICIQ−UI−SF scores for use in the models (as this data will be 100% complete and used for minimisation). Centre will also be fitted, where possible, as a random effect.

4.3 Secondary outcomes

The analysis of secondary outcomes will be undertaken in a similar manner to the analysis of the primary outcome, but using generalised linear models appropriate for the type of data: linear mixed models for continuous data, binary logistic regression for rates and proportions (e.g. for the proportions of women with UI cured/improved), ordered logit for ordinal data (e.g. for the Oxford Scale for pelvic floor muscle function), Poisson or negative binomial regression for count data (e.g. for the number of days that a woman has carried out PFME).

The PGI-S[10] will be included as a baseline covariate when the PGI-I is compared.

Comparisons between groups for the uptake rates both for UI surgery and for any other UI treatment will be analysed using binary logistic regression. Other comparisons relating to treatment sub-categories (e.g. GP appointments, nurse appointments etc) will not be modelled, but descriptive summaries will be reported for these outcomes.

For EQ−5D scores, if the distribution is approximately normal then a model similar to the analysis of the ICIQ−UI−SF will be used. It is, however, unlikely that EQ−5D will have a normal distribution in this population and significant ceiling effects are likely to be observed[23]. If the distribution is not normal, then a two-step estimation model will be considered to account for inflation in the number of scores with a value equal to one. This model would combine binary logistic and linear regression, with a Heckman correction for sample selection bias[24].

No comparisons will be modelled for bowel symptoms because a validated tool is not being used. Descriptive summaries of the data will be reported.

The analysis of pelvic floor muscle function will compare the Oxford Scale using an ordinal model. A descriptive summary of the distributions of the other PERFECT components at six months by treatment group will be reported but group comparisons will not be modelled for these outcomes (see section 3.2.3).

For adherence outcomes, the main comparisons of outcomes relating to the IMBS behaviour change theory[21] (i.e. introduction, clinic use, and home use) will be analysed using logistic regression. The other outcomes will not be modelled, but descriptive summaries will be reported. A descriptive summary of dose levels will be reported (Table 1.6, requested by the PMG in January 2018). We will also report a descriptive summary of adherence rates (i.e. introductory teaching/ any PFMT in clinic / any PFMT at home) broken down by attenders and non-attenders at the 6-month clinic assessment (as suggested by the HTA in April 2018).

4.4 Subgroup analyses

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

- Type of incontinence (SUI / MUI)
- Age (<50 / ≥50 years)
- UI severity at baseline (mild or moderate ICIQ−UI−SF score <13 / Severe ≥13)[7]
- Type of therapist (physiotherapist / nurse / other)

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 interactions[25].

4.5 Treatment received as allocated

The proportion of participants in each group who received treatment as allocated will be reported.

Treatment will be regarded as being received as allocated if a 'Yes' response has been given to each of the following:

- 'Teach PFM contraction', on the Visit 1 Checklist
- 'Teach probe and electrode insertion/removal', on the Visit 1 Checklist (applies to biofeedback group only)
- 'During VE give feedback on PFM contraction', on the Visit 1 Checklist
- 'PFM program written in home exercise diary and given to patient' (for at least one appointment), in section 6 of the TAF (for the basic PFMT group) or section 7 of the TAF (for the biofeedback group)

This definition will be used to define non-compliers in section 4.6. Our definition is consistent with how 'on-treatment' is defined in the analysis plan for the process evaluation. Stricter levels of treatment received were considered (including the maximum possible level of intervention), but these definitions will not be used as being on-treatment should relate to receiving a minimal dose of the randomised treatment. It should be noted that being on-treatment is different to being adherent.

There were four women randomised to biofeedback who were given basic PFMT from the outset (in error), and their data are recorded on therapy assessment forms for the basic PFMT group. The ID numbers are 12038, 13001, 13006 and 28008. Similarly, ID 26003 was randomised to the basic PFMT group but received biofeedback. The data for these women will still be analysed according to their randomised allocation, but they will be classed as non-compliers in section 4.6.

4.6 Analysis to address non-compliance

The statistical analysis of the RCT will be based on all women as randomised, irrespective of subsequent compliance with the treatment allocated. However, as is common in complex intervention trials, it is anticipated that not all participants will comply fully with the intervention to which they were randomised, and therefore further analysis of the primary outcome will be conducted which will take non-compliers into account, e.g. a complier average causal effect (CACE) analysis[26], (using the binary definition of non-compliance set out on section 4.5, i.e. having received introductory teaching (as defined in section 3.2.4) and having treatment received as allocated for at least one subsequent appointment).

Further analysis will be considered to determine if effectiveness is influenced not just by a binary measure of compliance, but also by dose level (level of PFMT / biofeedback use in the clinic and at home).

4.7 Additional analyses

The DMEC recommended an additional analysis to explore the differential learning which may occur (e.g. some women may use the device for a short time and then be confident while some may wish to use the device for longer as an aid). However, further discussions with the TSC and DMEC conceded that this would be a very difficult analysis to do, and would be unlikely to contribute a meaningful addition to the trial results, so this analysis will not be performed.

We will consider investigating whether the primary outcome is mediated by self-efficacy and adherence. A full quantitative mediation analysis[27] is being considered as a secondary analysis, which would be carried out after the main trial has been completed.

5. Missing Data

5.1 Loss to follow−up

Complete loss to follow−up is defined as a participant who has no information on outcomes at any follow−up timepoint, but has not withdrawn consent. These participants will not contribute data to any of the assessed outcomes. Partial loss to follow−up is defined as a participant contributing some follow−up data, but no further information is known on other follow−up outcomes. Such participants will contribute to the outcomes for which there are data.

5.2 Withdrawals

If a participant prospectively withdraws consent, no further data are captured or retained on or after the date of withdrawal of consent. Depending on when the consent is withdrawn, the above rules on loss to follow−up apply.

5.3 Post−randomisation exclusions

Participants who, after being randomised, are found to have been ineligible at the time of randomisation could be classed as post-randomisation exclusions

or protocol violations. There are five post-randomisation exclusions (22015, 27010, 27012, 27018, 29002). However, as this is a pragmatic trial, their data will not be excluded from any analyses. I

There were four women (13021, 13027, 28019, 31006) who became pregnant since randomisation. These cases will not be treated as protocol violations and will remain included in the analysis.

Seven participants (11022 15039 22003 28008 28024 30015 31001) consented to enter the trial but withdrew after randomisation and requested that their data be excluded. In these cases, their baseline data (if collected) will still be included in the analysis because, from a legal point of view, these participants consented to this at the start of the trial.

5.4 Imputation

Partially missing baseline data will be adjusted in order to improve efficiency[28]. There are 11 participants (1.8%) with completely missing baseline data. Centre mean imputation of missing baseline data for continuous variables will be undertaken in order to reduce bias. For categorical variables, an additional category for the missing data will be created.

Although no imputation of missing participant−level outcome data will be carried out in the main analysis, imputation of missing data from instruments (e.g. ICIQ−UI, EQ−5D, self-efficacy scale) will be undertaken at item−level according to the rules of the specific instrument.

5.5 Sensitivity analyses

It is recommended that sensitivity analyses are carried out where there are missing outcome data[29]. Specifically, the following sensitivity analyses will be performed, using multiple imputation (assuming data to be missing at random) and pattern mixture modelling (assuming data to be missing not at random) for missing primary outcome data at 24 months:

- Non−responders assumed to be missing 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).

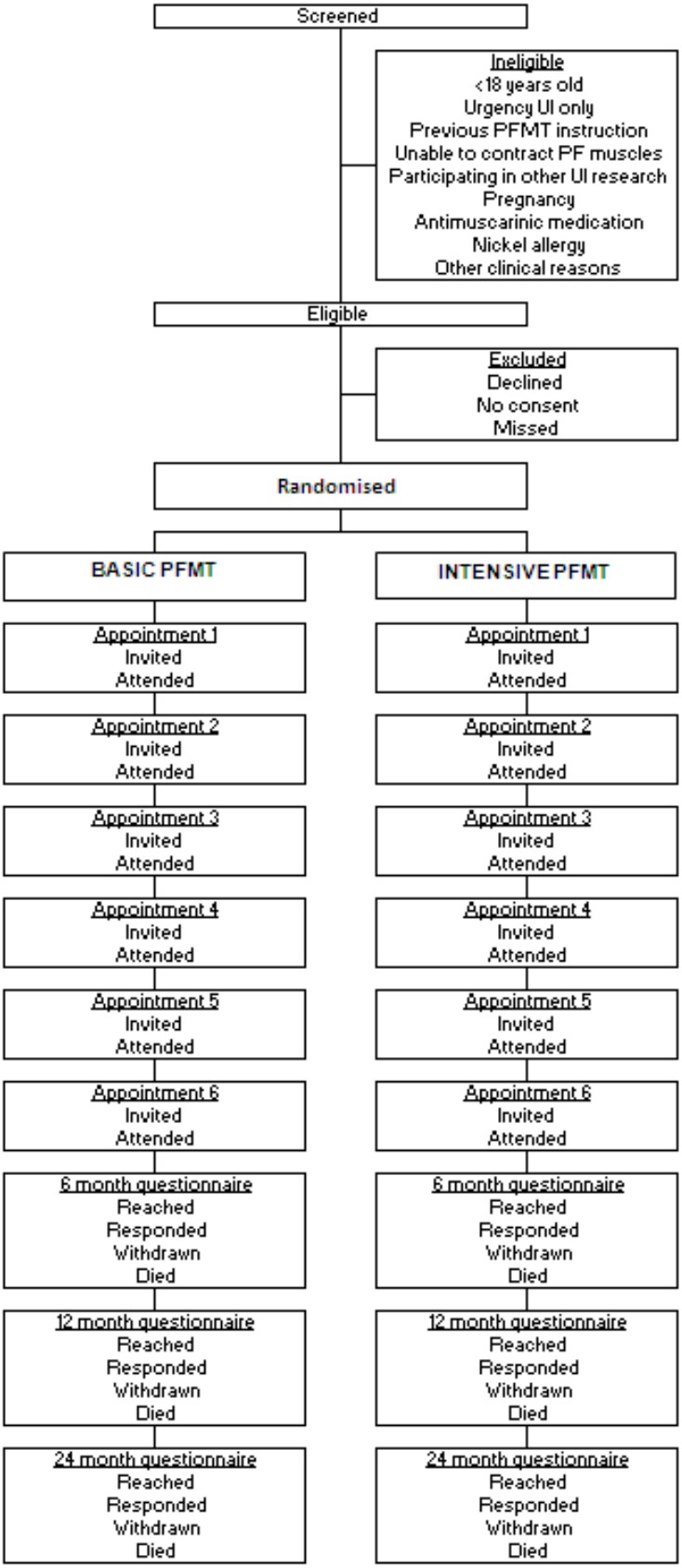
We will also undertake an exploratory investigation to describe particular characteristics of non-responders (e.g. differences in baseline ICIQ-UI−SF and levels of adherence to intervention)

The original protocol specified an exclusion criterion for women less than one year postnatal[1], but this was relaxed to six months in a protocol revision in 2015. A sensitivity analysis of the primary outcome will be carried out (as requested by the HTA) to test the effect of this protocol amendment, where randomised women who were less than one year postnatal will be excluded in the sensitivity analysis.

Improvement in UI is defined by a reduction of three points or more in the ICIQ_UI-SF (see section 3.2.1). A sensitivity analysis to test this assumption will be carried out using other thresholds (as agreed by the TSC in March 2017).

TABULATIONS

CONSORT Diagram



NB: The CONSORT diagram will also include proportions for overall treatment received as allocated.

Table 1.1: Baseline characteristics of participants

		BIOFEEDBACK N= 300			BASIC PFMT N=300		
Age	N, Mean, SD						
BMI	N, Mean, SD						
Number of previous Births:							
0	N, n, %						
1	N, n, %						
2	N, n, %						
3	N, n, %						
4 or more	N, n, %						
Delivery Mode History:							
Breech	N, n, %						
Caesareans Before	N, n, %						
Caesareans During	N, n, %						
Forceps	N, n, %						
Normal Vaginal	N, n, %						
Vacuum	N, n, %						
Type of incontinence:							
Stress	N, n, %						
Mixed	N, n, %						
Urinary measures:							
ICIQ-UI Short Form	N, Mean, SD						
ICIQ-FLUTS Filling Score	N, Median, IQR						
ICIQ-FLUTS Voiding Score	N, Median, IQR						
ICIQ-FLUTS Incontinence Score	N, Median, IQR						
How urine leakage is now (PGI−S scale):							
normal	N, n, %						
mild	N, n, %						
moderate	N, n, %						
severe	N, n, %						
Quality of life measures:							
ICIQ-LUTSqol	N, Mean, SD						
ICIQ−LUTSqol bother scale	N, Mean, SD						
EQ-5D	N, Mean, SD						
EQ-5D VAS	N, Mean, SD						
Pelvic floor measures:							
Prolapse Symptom Score (POP−SS)	N, Mean, SD						
Pelvic floor muscle function							
Power slow (Oxford scale)	N, Mean, SD						
Power fast	N, Mean, SD						
Endurance	N, Mean, SD						
Repetitions slow	N, Mean, SD						
Repetitions fast	N, Mean, SD						
Self−efficacy scale for PFMT	N, Mean, SD						
How often done PFMT last month:							
none	N, n, %						
few times a month	N, n, %						
once a week	N, n, %						
few times a week	N, n, %						
once a day	N, n, %						
few times a day	N, n, %						

Table 1.2: Baseline bowel symptoms

		BIOFEEDBACK N= 300			BASIC PFMT N=300		
Difficulty emptying bowels?							
never	N, n, %						
occasionally	N, n, %						
sometimes	N, n, %						
most of the time	N, n, %						
all of the time	N, n, %						
Rush to toilet?							
never	N, n, %						
occasionally	N, n, %						
sometimes	N, n, %						
most of the time	N, n, %						
all of the time	N, n, %						
Stool leak?							
never	N, n, %						
occasionally	N, n, %						
sometimes	N, n, %						
most of the time	N, n, %						
all of the time	N, n, %						
How often bowels open?							
3 or more a day	N, n, %						
twice a day	N, n, %						
once a day	N, n, %						
2/3 times a week	N, n, %						
1 a week or less	N, n, %						
Motions usually?							
watery	N, n, %						
sloppy	N, n, %						
soft and formed	N, n, %						
hard	N, n, %						

Table 1.3: Appointments attended

		BIOFEEDBACK			BASIC PFMT		
Number of appointments							
0	N, n, %						
1	N, n, %						
2	N, n, %						
3	N, n, %						
4	N, n, %						
5	N, n, %						
6	N, n, %						
More than 6	N, n, %						
Total number of appointments	N, Mean, SD						

Table 1.4: Treatment received at 1st appointment

		BIOFEEDBACK			BASIC PFMT		
Treatment received as allocated	N, n, %						
Daily exercise programme recommended							
Length of hold	N, Mean, SD						
Number of repetitions	N, Mean, SD						
Fast connections	N, Mean, SD						
Number of times per day	N, Mean, SD						

Table 1.5: Treatment received at 2nd/3rd/4th/5th/6th appointment

		BIOFEEDBACK			BASIC PFMT		
Treatment received as allocated	N, n, %						
Daily exercise programme recommended							
Length of hold	N, Mean, SD						
Number of repetitions	N, Mean, SD						
Fast connections	N, Mean, SD						
Number of times per day	N, Mean, SD						

Table 1.6: Number of PFMT / biofeedback sessions in clinic and at home

		BIOFEEDBACK			BASIC PFMT		
Number of clinic sessions							
0	N, n, %						
1	N, n, %						
2	N, n, %						
3	N, n, %						
4	N, n, %						
5	N, n, %						
6	N, n, %						
Number of home sessions:							
0	N, n, %						
1	N, n, %						
2	N, n, %						
3	N, n, %						
4	N, n, %						
5	N, n, %						
6	N, n, %						

Table 2.1: Outcomes at 6/12/24 months

		BIOFEEDBACK			BASIC PFMT		
Urinary outcomes							
ICIQ Urinary Incontinence Short Form Score	N, Median, IQR						
UI cured	N, n, %						
Improvement in UI	N, n, %						
Severe incontinence (≥13)	N, n, %						
Global impression of improvement (PGI−I scale)	N, Median, IQR						
ICIQ−FLUTS Filling Score	N, Median, IQR						
ICIQ−FLUTS Voiding Symptoms Score	N, Median, IQR						
ICIQ−FLUTS Incontinence Score	N, Median, IQR						
Treatment for UI in previous 6 months*							
Surgery	N, n, %						
Hospital admission	N, n, %						
Number of nights in hospital	N, Mean, SD						
Outpatient consultant appointment	N, n, %						
GP consultation	N, n, %						
Nurse appointment	N, n, %						
Physiotherapy	N, n, %						
Medication	N, n, %						
Advice	N, n, %						
Other treatment/advice	N, n, %						
Quality of life outcomes							
UI−specific quality of life (ICIQ−LUTSqol)	N, Mean, SD						
ICIQ−LUTSqol bother scale	N, Mean, SD						
EQ−5D score	N, Mean, SD						
EQ−5D visual analogue score	N, Mean, SD						
Pelvic floor outcomes							
Prolapse Symptom Score (POP−SS)	N, Mean, SD						
Pelvic floor muscle function							
Power slow (Oxford scale)	N, Mean, SD						
Power fast	N, Mean, SD						
Endurance	N, Mean, SD						
Repetitions slow	N, Mean, SD						
Repetitions fast	N, Mean, SD						
Self−efficacy scale for PFMT	N, Mean, SD						

* For the 24 months table, this will be treatment for UI in previous 12 months, and also previous 24 months

Table 2.2: Bowel symptoms at 6/12/24 months

		BIOFEEDBACK N= 300			BASIC PFMT N=300		
Difficulty emptying bowels?							
never	N, n, %						
occasionally	N, n, %						
sometimes	N, n, %						
most of the time	N, n, %						
all of the time	N, n, %						
Rush to toilet?							
never	N, n, %						
occasionally	N, n, %						
sometimes	N, n, %						
most of the time	N, n, %						
all of the time	N, n, %						
Stool leak?							
never	N, n, %						
occasionally	N, n, %						
sometimes	N, n, %						
most of the time	N, n, %						
all of the time	N, n, %						
How often bowels open?							
3 or more a day	N, n, %						
twice a day	N, n, %						
once a day	N, n, %						
2/3 times a week	N, n, %						
1 a week or less	N, n, %						
Motions usually?							
watery	N, n, %						
sloppy	N, n, %						
soft and formed	N, n, %						
hard	N, n, %						

Table 2.3: Adherence

	BIOFEEDBACK			BASIC PFMT		
Introductory teaching						
Any adherence in clinic						
Any adherence at home						
Number of clinic appointments attended						
Frequency of PFME at 6 months						
daily						
weekly						
monthly						
Frequency of PFME at 12 months						
daily						
weekly						
monthly						
Frequency of PFME at 24 months						
daily						
weekly						
monthly						

Table 2.4: Adverse events

		BIOFEEDBACK			BASIC PFMT		
Category of adverse event							
	N, n, %						
	N, n, %						
	N, n, %						
	N, n, %						
	N, n, %						
	N, n, %						

References

1. OPAL trial protocol. NIHR HTA Programme. www.nets.nihr.ac.uk/__data/assets/pdf_file/0017/81170/PRO−11−71−03.pdf. (Protocol paper submitted to BMJ Open, May 2018).

2. Hajebrahimi S, Corcos J, Lemieux C. International consultation on incontinence questionnaire short form: comparison of physician versus patient completion and immediate and delayed self−administration. Urology 2004; 63(6): 1076−1078.

3. Sherburn M. Bø K, Galea M. Evaluation of outcome measures for stress urinary incontinence in older women. Neurourology and Urodynamics 2009: 28(7); 715−716.

4. Sherburn M, Bird M, Carey M, Bø K, Galea MP. Incontinence improves in older women after intensive pelvic floor muscle training: an assessor−blinded randomized controlled trial. Neurourology and Urodynamics 2011; 30: 317−324.

5. Castro RA, Arruda RM, Zanetti MR, Santos PD, Sartori MG, Girão MJ. Single−blind, randomized, controlled trial of pelvic floor muscle training, electrical stimulation, vaginal cones, and no active treatment in the management of stress urinary incontinence. Clinics (Sao Paulo). 2008; 63(4): 465−72.

6. Imamura M, Abrams P, Bain C, Buckley B, Cardozo L, Cody J, Cook J, Eustice S, Glazener C, Grant A, Hay−Smith J, Hislop J, Jenkinson D, Kilonzo M, Nabi G, N'Dow J, Pickard R, Ternent L, Wallace S, Wardle J, Zhu S, Vale L. Systematic review and economic modelling of the effectiveness and cost−effectiveness of non−surgical treatments for women with stress urinary incontinence. Health Technology Assessment 2010; Vol. 14: No. 40.

7. Avery K, Donovan J, Peters TJ, Shaw C, Gotoh M, Abrams P. ICIQ: A brief and robust measure for evaluating the symptoms and impact of urinary incontinence. Neurourology and Urodynamics 2004; 23(4): 322−330.

8. Klovning, A., Avery, K., Sandvik, H., & Hunskaar, S. (2009). Comparison of two questionnaires for assessing the severity of urinary incontinence: The ICIQ-UI SF versus the incontinence severity index. Neurourology and urodynamics, *28*(5), 411-415.

9. Locher JL, Goode PS, Roth DL, Worrell RL, Burgio KL. Reliability assessment of the bladder diary for urinary incontinence in older women. Journal of Gerontology 2001; 56A: M32−M35.

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

11. Brookes S, Donovan J, Wright M, Jackson S, Abrams P. A scored form of the Bristol Lower Urinary Tract Symptoms questionnaire: data from a randomized controlled trial of surgery for women with stress incontinence. American Journal of Obstetrics and Gynecology 2004; 191(1):73−82.

12. Hilton et al. "INVESTIGATE-I (INVasive Evaluation before Surgical Treatment of Incontinence Gives Added Therapeutic Effect?): a mixed-methods study to assess the feasibility of a future randomised controlled trial of invasive urodynamic testing prior to surgery for stress urinary incontinence in women." *Health technology assessment (Winchester, England)* 19.15 (2015): 1.

13. Kelleher C, Cardozo L, Khullar V, et al. 1997. A new questionnaire to assess the quality of life of urinary incontinent women. British Journal of Obstetrics and Gynaecology 104:1374−9.

14. Wagner TH, Patrick DL, Bavendam TG, Martin ML, Buesching DP. Quality of life of persons with urinary incontinence: development of a new measure. Urology 1996;47(1):67−71.

15. The EuroQol Group. EuroQol−−a new facility for the measurement of health−related quality of life. Health Policy 1990; 16(3):199−208.

16. Hagen S, Glazener C, Sinclair L, Stark D, Bugge C. Psychometric properties of the pelvic organ prolapse symptom score. British Journal of Obstetrics and Gynaecology, 2009; 116(1): 25−31.

17. Abrams P, Avery K, Gardener N, Donovan J. The International Consultation on Incontinence Modular Questionnaire: www.iciq.net. Journal of Urology, 2006; 175: 1063−6.

18. Laycock J, Jerwood D. Pelvic floor muscle assessment: the PERFECT scheme. Physiotherapy 2001; 87(12): 631−642.

19. Messelink B, Benson T, Berghmans B, Bø K, Corcos J, Fowler C, Laycock J, Huat−Chye Lim P, van Lunsen R, Lycklama a¤ Nijeholt G, Pemberton J, Wang A, Watier A, Van Kerrebroeck P. Standardization of Terminology of Pelvic Floor Muscle Function and Dysfunction: Report From the Pelvic Floor Clinical Assessment Group of the International Continence Society. Neurourology and Urodynamics 2005; 24: 374−380.

20. Chen S−Y. The development and testing of the pelvic floor muscle exercise self−efficacy scale. Journal of Nursing Research 2004; 12(4): 257−265.

21. Fisher JD, Fisher WA. The information-motivation-behavioral skills model. In: DiClemente R, Crosby R, Kegler M, editors. Emerging theories in health promotion practice and research. San Fransisco: Jossy-Bass; 2002. p. 40-70.

22. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow−up measurements. British Medical Journal, 2001; 323: 1123−4.

23. Tincello D, Sculpher M, Tunn R, Quail D, van der Vaart H, Falconer C, Timlin L. Patient Characteristics Impacting Health State Index Scores, Measured by the EQ-5D of Females with Stress Urinary Incontinence Symptoms. Value in Health 2010; 13(1): 112-118.

24. Heckman JJ. Sample selection bias as a specification error. Econometrica: Journal of the Econometric Society 1979; 153-161.

25. 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.

26. 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.

27. Baron RM, Kenny DA (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

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

29. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analyses in randomised trials with missing outcome data. British Medical Journal, 2011; 342: d40.

Deviations from the Statistical Analysis Plan

The results presented reflect the full set of planned analyses set out in the Statistical Analysis Plan, with the following exceptions:

- Models were planned to adjust for therapist type (physiotherapist / nurse / other). The analyses however used a binary covariate (physio / not physio), and the use of this covariate was examined in a post-hoc sensitivity analysis.
- A post-hoc analysis was conducted whereby the definition of compliance in the biofeedback group was relaxed to allow for training in the use of the biofeedback device during either the first or second appointment, instead of the planned definition which did not include training during the second appointment.
- The planned sensitivity analysis to examine the effect of relaxing the exclusion criteria for postnatal women (from 12 months to 6 months) was not conducted as the data required for this analysis was not collected when women were screened for eligibility.
- A post-hoc analysis was conducted whereby the ordinal PGI-I scale was dichotomised, with a positive outcome being when the response was ‘much better’ or ‘very much better’
- The analysis of EQ5D data is not reported because these results are presented as part of the health economic analysis.
- In the analysis of further non-surgical treatment, nurse and physiotherapist data have been combined in order to be consistent with the health economic analysis..

HEALTH ECONOMICS ANALYSIS PLAN

7 May 2018

Prepared by:

M. Kilonzo

Health Economist

Approved by:

G. Scotland

Health Economist

S. Hagen

Chief Investigator

ISRCTN57746448

1.0 Objective of the study

The principal research question being addressed in the economic analysis is what is the cost effectiveness of a policy of basic PFMT compared with biofeedback intensified PFMT?

1.1 Study design

The main economic evaluation will be based on data collected alongside the RCT. The effect measure for the cost-effectiveness analysis will be quality-adjusted life years. An additional modelling analysis which considers a longer time horizon will also be conducted to provide additional information for policy makers.

1.2 Study population and perspective

The trial population is women presenting with urinary stress or mixed urinary incontinence. Both within trial and model analyses will assess the costs and cost-effectiveness of the interventions compared from the perspective of the NHS. The within trial analysis will also include a societal perspective that will consider the cost to the participants and their families. The methods for within trial and the modelling analyses are described below.

1.3 Follow-up period

Resource utilisation and quality of life will be measured over four time points (baseline, six, 12 and 24 months), over 24 months follow-up period using two sources (CRFs and patient reported questionnaires). The second year cost and benefits will be discounted using the NICE recommended 3.5% (NICE 2013). The primary outcome for the economic evaluation is the incremental cost-effectiveness ratio (ICER) of biofeedback intensified PFMT compared to PFMT.

2. Data collection

2.1 Resource use

Intervention resource use namely number of visits to the therapists will be captured through the appointment record collected using Therapist Assessment form. A basic PFMT protocol will be delivered during six therapy appointments over a 16 week period at around weeks 0, 1, 3, 6, 10 and 15. The first appointment is expected to last for an hour and each follow-up

appointment will last 30 mins. The intensive group intervention will have the additional resource of biofeedback in the clinic and a simple portable biofeedback machine for home use. Information on the biofeedback units will be source from the trial office. All consenting women will receive questionnaires at baseline, and at six, 12 and 24 months (24 months questionnaire including time and travel questions) post randomisation.

Further resource use will be recorded prospectively for every woman within the study (Table 1). Resource use incurred at personal cost to the participants (such as purchase of pads, medication) will also be collected using a questionnaire at 24 months. Resource-use data collected will include the use of primary (GP services) and secondary (hospital inpatient stay, surgical interventions for their incontinence) NHS services by the participants, including further referral for subsequent additional specialist management. Health service costs refer to those incurred directly by the NHS due to any surgery, subsequent appointments and procedures.

Table 1 Resource use data

	Resource	**Unit**	**Source**
Intervention resource use	Appointments with therapist	Number	CRF
	Biofeedback machine and consumables	Number	CRF
Other secondary care resource use	Outpatient visit	Number	CRF & PR Questionnaire
	Inpatient readmissions	Number	CRF & PR Questionnaire
Primary care resource use	Practice nurse visit	Number	PR Questionnaire
	GP visit	Number	PR Questionnaire
	Visit to other providers	Number	PR Questionnaire
Participant resource use	Medications	Number	PR Questionnaire
	Pads/catheters	Number	PR Questionnaire
	Visits to non NHS providers	Number	PR Questionnaire

PR patient reported

2.2 Unit costs

The unit costs will be applied in British Pound Sterling £. Unit costs/prices will be obtained using published estimates BNF (Joint Formulary Committee 2017), Reference costs (Department of Health 2017) and PSSRU Unit Costs of Health and Social Care (Curtis L. & Burns A. 2017) as outlined in Table 2.

Table 2 Average NHS unit costs

Area of resource use	Resource	Unit cost	Source	Notes
Intervention resource use	Portable biofeedback machines		Manufacturer	Current 2018 unit cost (excluding VAT) for Neurotrac Simplex with Bluetooth /Peritone single channel EMG
	Electrodes		Trial office	Average cost of electrodes purchased for the trial
	Probe		Manufacturer	Current 2018 unit cost for probe Peritone plus multilingual
	Therapist visit (physiotherapist or specialist nurse)		PSSRU	Based on average cost per hour of patient contact of Band 6 and Band 7 -nurse
Primary care	GP visit		PSSRU	Cost per surgery consultation lasting 9.22 minutes, including qualifications given in brackets, both are including direct care staff costs
	Nurse visit (GP practice)		PSSRU	Surgery consultation based on the 2006/7 UK general practice survey is 15.5 minutes (including qualifications given in brackets)
Secondary care (outpatient services)	NHS doctor visit		Reference costs	First non-admitted face to face appointment (follow-up appointments) Consultant-led (gynaecology)
	NHS nurse visit/physiotherapist		Reference costs	First non-admitted face to face appointment (follow-up appointments) non-consultant-led (gynaecology)
Secondary care (inpatient)	Overnight stay in hospital		Reference costs	A weighted average of HRG4 codes LB16G, LB16H,LB16J and LB16K for urinary incontinence

				without interventions CC 0-8+ non-elective inpatient excess bed days
Participant resource use	Medications	Various patient-reported	BNF	See supplementary table for more detail
	Surgical interventions		Reference costs	A weighted average of HRG4 codes LB51A and LB51B CC score 0-2+ for TVT
	Non-surgical interventions (Injections)		Reference costs	Intermediate endoscopic bladder procedures LB14Z
	Private care			
	Doctor (GP)		BUPA	Based on 15 minute appointment
	Nurse		BUPA	Initial consultation (follow-up consultations) pay as you go appointments (assumed equivalence with physiotherapist)
	Physiotherapist		BUPA	Initial consultation (follow-up consultations) pay as you go appointments

2.3 Estimation of cost per patient and average cost per patient by elements of resource use and total cost per patient

For each area of resource use, estimates of resource utilisation (Table 3) will be combined with unit costs (Table 2) to derive total costs for each item of resource use and each patient. These data will be averaged to provide estimates of the average cost per patient for each item of resource use.

Table 3 Average resource use per arm of treatment and difference

	Intensive PFMT N Mean SD	**Basic PFMT N Mean SD**	**Difference [95% CI]**
Appointments with therapist			
21 Portable biofeedback			
Outpatient visits			
Inpatient admissions			
Operation for incontinence N (%)			
Practice nurse visits			
GP visits			
Visits to other providers			
Medications Incontinence medications N (%) Antibiotics N (%)			
Pads/catheters			
Visits to non NHS providers			

*Number

The costs for each item of resource use for each patient will be summed to produce a total cost for each patient and an average total cost per patient (table 4) in each intervention arm.

Table 4 Average cost per arm of treatment and difference in cost

	Intensive PFMT N Mean SD £	**Basic PFMT N Mean SD £**	**Difference [95% CI]**
Appointments with therapist			
Portable biofeedback			
Outpatient visits			
Inpatient admissions			

Operation for incontinence			
Practice nurse visit s			
GP visits			
Visits to other providers			
Medications Incontinence Antibiotics			
Pads/catheters			
Visits to non NHS providers			

2.4 Participant- and companion-incurred costs and indirect costs

Personal costs to the participants (such as costs of travelling to appointments and work/social restrictions) will also be investigated. Participant resource utilisation comprises three main elements: self-purchased health care; travel costs for making return visit(s) to NHS health care (such as petrol, public transport and parking); and time costs of travelling and attending NHS health care (such as time involved away from usual activities or work). All self-purchased health care relates to treatment purchased for the management or treatment of urinary incontinence. Time and travel costs relate to time spent travelling to and attending hospital or primary care providers in relation to urinary incontinence. (Table 5). Estimation of travel costs will include information from participants about the number of visits to, for example, their GP or physiotherapist (estimated from the health-care utilisation questions at the various questionnaire time points) and the unit cost of making a return journey to each type of health-care provider (from the participant time and travel cost questions collected at 24 months). The cost of participant time will be estimated in a similar manner. The participant is asked how long they spent travelling to, and attending, their last visit to each type of health-care provider. Information will also be sought on the activity they would have been undertaking (e.g. paid work, leisure, housework) had they not attended the health-care provider. They are further asked if they were accompanied by a friend or a relative and their time and travel costs will also be incorporated into the analysis. These data will be presented in their natural units, for example hours, and also costed using standard economic conventions, using the Department of Transport estimates for the value of work and leisure

time. These unit time costs will then be combined with the number of health-care contacts derived from the health-care utilisation questions to elicit a total time and travel cost from a patient perspective. Details of unit costs applied to the various activities are included below.

Table 5 Participant time and travel cost

Activity	Unit cost (£)	Source and notes
Unit costs applied to participant and companion travel		
Cost per mile travelled by car		HMRC
Car parking charges	Various	As reported by participants
Cost of public transport (bus, train, taxi)	Various	As reported by participants
Cost of return journey by hospital car	Per trip	Torbay and South Devon NHS Foundation Trust
Cost of non-emergency patient transport service (via ambulance)		NHS reference costs
Unit costs applied to participant and companion time		
Paid work	Per hour	ONS annual survey of hours and earnings
Housework	Per hour	NHS pay review body
Child care	Per hour	ONS annual survey of hours and earnings (as paid work)
Caring for a friend/family member	Per hour	ONS annual survey of hours and earnings(as paid work)
Voluntary work	Per hour	ONS annual survey of hours and earnings(as paid work)
Retired	Per hour	TAG data book
Leisure	Per hour	TAG data book
Unemployed	Per hour	TAG data book
Ill/disabled (long term, unrelated to incontinence)	Per hour	TAG data book

2.5 Derivation of quality of life

A generic instrument EQ-5D -3L^{TM} will be used to measure the quality of life. Trial participants will be asked to complete the EQ-5D-3L^{TM} at baseline and at six, 12 and 24

months after their intervention. This instrument will provide the quality of life weights to compute the QALYs. The responses to the EQ-5D-3L™ questionnaire will be valued using UK general population tariffs, based on the time trade-off technique to generate a utility score for every participant within the trial (Dolan P 1997). QALYs (Table 6) will be calculated on the basis of these assumptions, using an area beneath the curve approach, assuming linear extrapolation of utility between time points. Quality of life data is collected using items from the condition specific tool (ICIQ urinary incontinence short form questionnaire) for comparison. ICIQ-UI SF data are collected at baseline, six, 12 and 24 months. These data will be converted into a utility index using a published algorithm (Brazier J 2004).

Table 6 Quality of life measures

Score	Intensive PFMT N Mean SD	Basic PFMT N Mean SD	Difference [95% CI]
Baseline EQ-5D-3L			
6 months EQ-5D-3L			
12 months EQ-5D-3L			
24 months EQ-5D-3L			
Total QALYs EQ-5D-3L			
Baseline ICIQ-UI SF			
6 months ICIQ-UI SF			
12 months ICIQ-UI SF			
24 months ICIQ-UI SF			
Total QALYs ICIQ-UI SF			

3.0 Data analysis

The economic analysis will be undertaken using the intention to treat principle. All components of costs will be described with the appropriate descriptive statistics where relevant: mean and SD for continuous and count outcomes; numbers and percentages for

dichotomous and categorical outcomes (e.g. numbers reporting problems on EQ-5D-3L). All analyses will be conducted using Stata® version 14.1 software (StataCorp LP, College Station, TX, USA). Depending on the results, investigations will be carried out for skewed cost data (i.e. a small proportion of participants incurring very high costs), using GLMs to test alternative model specifications for appropriate fit to the data. The GLM models allow for heteroscedasticity by selecting and specifying an appropriate distributional family for the data. This family offers alternative specifications to reflect the relationship between the mean and variance of the estimates under consideration (Glick 2007, Drummond 2005). Two diagnostic actions will be performed to identify the most appropriate distributional family: (1) a modified Park test (2) the Akaike information criterion (AIC) will be consulted.

Both cost and QALY difference analyses will be adjusted for baseline prognostic factors (all of which are minimisation covariates except for the type of therapist):

- Centre number
- Age in years
- Type of UI
- UI severity
- Type of therapist (physiotherapist / nurse / other)
- Baseline EQ5D

The first five factors are in line with the clinical effectiveness analyses and the baseline EQ5D will be included for the economic analysis. We will carry out standard parametric tests for differences in costs, with the robustness of the parametric tests confirmed using bias-corrected, nonparametric bootstrapping (Barber JA, Thompson SG. 2000).

3.1 Incremental cost per and QALYs gained

Incremental cost-effectiveness ratios will be computed comparing the cost of the interventions. The difference in effectiveness will be expressed in terms of quality adjusted life years. These data will be based on responses to EQ5D and questions from the ICIQ-UI SF relating to the loss of urine, retrieved from the participant questionnaire. Incremental cost-utility ratios will be computed comparing the interventions. The difference in utility will be expressed in terms of QALYs at 24 months.

The point estimate of the incremental cost-effectiveness ratio (ICER) will be calculated as:

$$ICER = \frac{Ci - Cj}{Ei - Ej} = \frac{\Delta C}{\Delta E}$$

where Ci and Cj are the mean costs among women in the PFMTBF arm and PMFT arm respectively. Similarly, Ei and Ej are the mean quality-adjusted life years in the PMFTBF arm and PMFT arm. The ICER will be assessed against the NICE recommended cost-effectiveness threshold £20,000-30,000 per QALY gained.

Measures of variance for NHS costs, incontinent participants and QALYs will be derived using bootstrapping. From the results of the bootstrapping cost-effectiveness acceptability curves (CEACs) will be created (Table 7). Cost-effectiveness acceptability curves will be used to display the inherent uncertainty surrounding cost-effectiveness at various threshold values for society's willingness to pay for r additional QALY. CEACs present results when the analysis follows a net benefit approach. This approach utilises a straightforward re-arrangement of the cost-effectiveness decision rule used when calculating ICERs (see below) to create the net monetary benefit for each bootstrapped iteration at increasing values of WTP per QALY:

$$NMB = \lambda.\Delta E - \Delta C > 0$$

Where λ is represents a decision maker's willingness to pay for incontinence avoided or a QALY gained. If the above expression holds true for a given iteration and threshold WTP value (λ), then the intervention is considered cost-effective for that iteration. As society's willingness to pay is unknown, the NMB will be calculated for a number of possible λ values including the usual £20-£30K range often adopted by policymakers within the NHS (NICE 2013). Table 7 shows the data that will be collected in relation to cost-effectiveness in order to calculate ICERs and, following on from this, the NMB of the interventions.

Table 7 Incremental cost effectiveness (replicated for both the QALY based analyses and for the number of participants who are incontinent)

	Cost	Effect	Δ Cost	Δ Effect	ICER (ΔC/ΔE)	Probability cost effective £20,000
Most costly trial arm						
Least costly trial arm						

A balance sheet approach will be used to report the costs and QALYs of women that are incontinent or not.

3.2 Missing data

Missing data are a frequent problem in economic evaluations undertaken within a randomised controlled trial setting. There are several possible methods that can be employed to account for such missing data: mean or multiple imputation. Imputation analysis will be conducted if more than 5% of the data needed is missing for the primary analysis. The handling of missing data will be dependent on the pattern of missing data. If the data is "Missing at Random (MAR)", multiple imputation will be used. Components of cost data will be imputed, based on linear regression models that were adjusted for minimisation variables, baseline utility and treatment allocation group. Missing utility values will be imputed using predictive mean matching. Chained equations will be used for the imputations.

4.1 Sensitivity and sub-group analysis

Sensitivity analysis will be performed to gauge the impact of varying key assumptions and/or parameter values in the base-case analysis.

1) Sensitivity analyses in relation to the source of information for length of appointment will be performed. The base-case analysis will utilise cost estimation based on therapist reported length of appointment. The first sensitivity analysis will be performed using

information using the recommended length of appointment (I hour for the first appointment and 30 minutes for follow-up appointments). Further to this, if it is possible to obtain costs from a leading trial centre, a further analysis which utilises these costs will be performed. These analyses will serve to highlight the differences in results when using national and centre-specific tariffs.

2) The base-case analysis in terms of utilities will be adjusted for baseline values to account for variability that may be present amongst the intervention groups. An unadjusted analysis will also be performed as a sensitivity analysis to highlight the importance of this base-case assumption.

3) Analysis exploring the impact of changing the discount rate used for second-year costs and QALYs in accordance with NICE best practice recommendations, varying the discount rate from 0% to 6% per annum will be undertaken.

4.2 Subgroup analysis

Depending on the availability of data, subgroup analysis similar to that described in the statistical analysis plan will be undertaken. This will be based on

- Type of incontinence (SUI or MUI)
- Type of therapist (physiotherapist or nurse)
- Age (<50/≥50 years)
- UI severity (ICIQ-UI SF score <13/≥ 13)

5.0 Modelling analysis

If relevant an economic model which considers a longer time horizon will also be developed to provide additional information for policy makers (Briggs A, Sculpher M, Claxton K 2006). Model analysis will be undertaken if the within trial analysis indicate that either treatment is cost effective at £20-£30,000 willingness to pay threshold. No modelling will be undertaken if one treatment dominates the other. In the model, the findings of the trial will be extrapolated to 10-20 years. The model will describe care pathways that people may follow and will include the initial therapy and any subsequent treatments. The structure of the model will be based on an existing model (Imarura M et al 2010) that was developed in collaboration

with the expert panel of service users, patients, clinicians and trial collaborators. Parameter estimates for relative effectiveness up to two years, costs and utilities will be derived from the trial data. Data from the trial will be supplemented with data from other sources (e.g. Cochrane review, other RCTs). These data will be assembled systematically and will follow guidelines for good practice (Philips Z. et al. 2004).

Outcomes in the model will be expressed in terms of an incremental cost per QALY. Parameter uncertainty will be integrated by the incorporation of probability distributions into the model and involve Monte Carlo simulation. Other forms of uncertainty such as that associated with choices made about the structure of the model, discount rate, etc. will be addressed through sensitivity analysis. The base case and sensitivity analyses will be presented as cost effectiveness acceptability curves.

6.0 References

NICE (2013) Guides to the methods of technology appraisal 2013. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case.

Joint Formulary Committee. British National Formulary (online). London: BMJ Group and Pharmaceutical Press. URL: www.medicinescomplete.com

Department of Health. NHS Reference Costs 2015–2016. URL: www.gov.uk/government/publications/nhs-reference-costs-2015-to-2016

Curtis, L. & Burns, A. (2017) Unit Costs of Health and Social Care 2017, Personal Social Services Research Unit, University of Kent, Canterbury. https://doi.org/10.22024/UniKent/01.02/65559

HM Revenue and Customs 2017. Expenses and benefits: business travel mileage for employees' own vehicles. https://www.gov.uk/expenses-and-benefits-business-travel-mileage/rules-for-tax

Office for National Statistics - https://www.ons.gov.uk/visualisations/dvc376/index.html

Department of Transport Guidance TAG book 2011 https://www.gov.uk/government/publications/tag-data-book

BUPA 2017 Pay as you go healthcare. https://www.bupa.co.uk/health/payg

Dolan P Modelling valuations for EuroQol health states. *Med Care* 1997; 35: 1095-108. http://dx.doi.org/10.1097/00005650-199711000-00002

Brazier J1, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire.

Glick H, Doshi J, Sonnad S, Polsky D. Economic Evaluation in Clinical Trials. 1st edn. New York, NY: Oxford University Press; 2007.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. 3rd edn. Oxford: Oxford University Press; 2005.

Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med. 2000;19:3219–36]. Med Decis Making. 2008 Jan-Feb;28(1):113-26. Epub 2007 Jul 19.

Briggs A, Sculpher M, Claxton K (2006) Decision modelling for health economic evaluation.

Oxford: Oxford University Press.

Imamura M, Abrams P, Bain C, Buckley B, Cardozo L, Cody J, Cook J, Eustice S, Glazener C, Grant A, Hay-Smith J, Hislop J, Jenkinson D, Kilonzo M, Nabi G, N'Dow J, Pickard R, Ternent T, Wallace S, Wardle J, Zhu S and Vale L Systematic review and economic modelling of the effectiveness and efficiency of non-surgical treatments for women with stress urinary incontinence (SUI) Health Technology Assessment 2010; 14, 40. Website http://www.hta.ac.uk/project/1612.asp

Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health technology assessment (Winchester, England) 2004; 8(36)

Qualitative Study and Process Evaluation Analysis Plan v1.0

23rd July 2018

Contents

1.0 Introduction

The OPAL study comprises:

1. a parallel group multicentre randomised controlled trial to compare effectiveness and cost-effectiveness of basic PFMT versus biofeedback-mediated intensive PFMT for women with stress UI and mixed UI;
2. a mixed-methods, nested process evaluation, and
3. a longitudinal semi-structured interview study with purposively selected women from both arms to explore experience of, and adherence to the trial interventions.

This analysis plan focusses on the process evaluation and interview study. The SAP (version 1.1, 04/06/2018) outlines the procedures for the main trial. Throughout this document, the quantitative data where referred to will be analysed in line with the conventions laid out in the SAP except where stated otherwise.

2.0 The research questions

The research questions that are relevant to this analysis plan are:

1) To identify and investigate, via process evaluation, the possible factors that impact upon the effectiveness of (a) the intervention and (b) intervention delivery fidelity

2) How these factors influence effectiveness

3) Whether the factors differ between randomised groups

4) To investigate women's experiences of the interventions, both (a) basic and (b) intensive PFMT,

5) To identify the (a) barriers and (b) facilitators which impact on adherence in the (c) short- and (d) long-term,

6) To explain the process through which barriers and facilitators influence adherence, and

7) To identify whether barriers and facilitators differ between randomised groups

3.0 Tables of research questions and methods used to answer the questions

The following tables summarise the analysis. They build on the framework developed by Grant and colleagues (Grant et al, 2013) and break the research questions down into their component parts in line with their mapping onto the Grant Framework.

Table One: Analysis conducted when blind to outcomes

	Type of data and data analysis – see key	**Qt & ql**	**Qt & ql**	**Ql**	**Ql**	**Qt & ql**	**Qt**	
RQ nos	**RQs**	**Checklists (TAF)**	**Consultation recordings**	**HCP interviews**	**Women interviews**	**Exercise diaries**	**Main study Qnaire**	**Mixing Methods**
	DELIVERY TO INDIVIDUALS							
1b	What treatment is delivered? (enactment treatment delivery)	P Qt	P Qt	S	S	S		Sequential Design – quant 1st
1b, 4	What do participants learn? (enactment treatment receipt)		P Ql & Qt	S	P Ql	P Ql		Triangulation
1b	Does treatment delivery match intended protocols?	P Qt	P Qt	S	S	P Qt		Sequential Design – quant 1st
5, 7	What are the barriers and facilitators to treatment delivery?	S	S	P Ql	P Ql			Sequential Design – qual 1st
5, 7	What are the barriers and facilitators to treatment receipt?		S	P Ql	P Ql			Sequential Design – qual 1st
1a,2,3	What elements of delivery influence outcome?	S	S	P Ql	P Ql	S	S	Qualitative hypothesis generating to test post trial outcome analysis

	Type of data and data analysis – see key	**Qt & ql**	**Qt & ql**	**Ql**	**Ql**	**Qt & ql**	**Qt**	
RQ nos	**RQs**	**Checklists (TAF)**	**Consultation recordings**	**HCP interviews**	**Women interviews**	**Exercise diaries**	**Main study Qnaire**	**Mixing Methods**
	RESPONSE OF INDIVIDUALS							
1a,4	Of what is delivered and learned, what is used (enactment)?	P Qt	P Qt		S	P Qt	P Qt	Sequential Design – quant 1st
1a, 4	How do those receiving treatment respond?		S	P Ql	P Ql		S	Sequential Design – qual 1st
4,5	What are the barriers and facilitators to treatment response in individuals?		S	P Ql	P Ql	S		Sequential Design – qual 1st
3,7	Are there/ what are the differences in response between basic and BF arms (from PE data)?		S	P Ql	P Ql			Qualitative analysis
1,2,3	What factors might explain different patterns of response between intervention and control groups?	P Qt	P Qt	S	S	P Qt		Sequential Design – quant 1st
1a), 4,	How do women adhere to exercise during active treatment?	S	S	S	P Ql	P Ql	P Qt	Triangulation
5, 6, 7	What factors explain adherence during active treatment?	S	S	S	P Ql	S	P Qt	Triangulation

	Type of data and data analysis – see key	**Qt & ql**	**Qt & ql**	**Ql**	**Ql**	**Qt & ql**	**Qt**	
RQ nos	**RQs**	**Checklists (TAF)**	**Consultation recordings**	**HCP interviews**	**Women interviews**	**Exercise diaries**	**Main study Qnaire**	**Mixing Methods**
	MAINTENANCE							
4	How do women adhere to exercise post treatment?				P Ql		S	Qualitative analysis
5, 6	What factors explain adherence post intervention?	S	S	S	P Ql		S	Qualitative analysis
5, 6	What are the barriers and facilitators to adherence during active treatment?	S	S	P Ql	P Ql	S		Sequential Design – qual 1st
5, 6	What are the barriers and facilitators to adherence post treatment?			P Ql	P Ql			Qualitative analysis
2,6,7	Are there/ what are the differences between PFMT and BF groups in adherence?	S	S	P Ql	P Ql	S	S	Sequential Design – qual 1st
2,7	What factors explain those differences (if they exist)?	S	S	P Ql	P Ql	S	S	Sequential Design – qual 1st

	Type of data and data analysis – see key	Qt & ql	Qt & ql	Ql	Ql	Qt & ql	Qt	
RQ nos	**RQs**	**Checklists (TAF)**	**Consultation recordings**	**HCP interviews**	**Women interviews**	**Exercise diaries**	**Main study Qnaire**	**Mixing Methods**
	THEORY							
1b, 4,5	What influence has theory had on the adoption of treatment protocols?	S	S	P Ql	P Ql			Qualitative analysis
1b,4,5	What influence has theory had on treatment delivery?	S	S	P Ql	P Ql			Qualitative analysis
5,6	What influence has theory had on adherence?	S	S	P Ql	P Ql	P Qt		Triangulation
2,3	Is BF a treatment intensifier?	S	S	P Ql	P Ql	S		Sequential Design – qual 1st
	Longitudinal Qualitative study emergent findings / further newly identified questions			P Ql	P Ql			

Table Two: Analysis conducted after trial outcome known

	Type of data and data analysis – see key	**Qt & ql**	**Qt & ql**	**Ql**	**Ql**	**Qt & ql**	**Qt**	
RQ nos	**RQs**	**Checklists (TAF)**	**Consultation recordings**	**HCP interviews**	**Women interviews**	**Exercise diaries**	**Main study Qnaire**	**Mixing Methods**
1a, 2,3	What elements of delivery influence outcome?	S	S	P Ql	P Ql	S	S	Qualitative hypothesis generating to test post trial outcome analysis
1,2, 3	What factors might explain different patterns of response between intervention and control groups?	P Qt	P Qt	S	S	P Qt	S	Sequential Design – quant 1st
	EFFECTIVENESS							
1a, 6,7	What factors explain the links (if they exist) between treatment and outcome?	P Qt	P Qt	P Ql	P Ql	P Qt	S	

Key:
Qt = Quantitative (predominates); qt = quantitative supplements
Ql = Qualitative (predominates); ql = qualitative supplements
P= primary data source
S= Supplementary source

4.0 Our definition of on treatment

Our **definition** of PFMT (on treatment) and this applies to both arms:

- Checklist – tick next to teach PFMT (yes)
- Checklist: During VE give feedback on PFM contraction (yes)
- Q6 in TAF – exercise prescription (yes to PFM programme written in home exercise diary and given to patient)

To be on treatment these three things in **any** one appointment or ticked at least once across appointments. If someone does not get all three of these things across time at least once (at any one time) this is a protocol deviation.

BF on treatment is as defined in the SAP.

To be on treatment:

	On treatment	**Protocol deviation**
PFMT	Get PFMT as defined	No PFMT as defined
BF	Get PFMT as defined PLUS get BF as defined	No BF as defined no PFMT as defined No BF as defined but get PFMT as defined BF as defined but no PFMT as defined

5.0 Analysis methods for each data source (column headings in tables one and two)

All statistical analysis will be undertaken in line with the SAP (version 1.1; 04/06/2018).

5.1 Checklist

Data sample: The checklist is completed by the therapist for all appointments for all women; it is contained within the TAF.

To analyse the 'tick box' items on the checklists we will:

1. Descriptively summarise item use (total number used/ percentage of all items) for each visit for core and for additional items; and by group (basic and biofeedback).
2. Descriptively summarise item use (total number used / percentage of all items in category) in each category as defined in the checklist (e.g. beliefs, emotions and information n out of 4 per participant and describe by mean/SD) for core and for additional items, and by group (basic and biofeedback).
3. Descriptively summarise item use (total number used / percentage of all items) of items used in each visit from core set and from additional set; and by group (basic and biofeedback).
4. Descriptively summarise item use across time per item (e.g. action planning is repeated in 3 visits), for core set, (for additional set) and by group (basic and biofeedback).
5. Descriptively summarise item use (total number used / percentage of all items in category) within IMB categories (information, motivation, behaviour) for core and for additional and by group (basic and biofeedback).

Post report analyses:

6. Ditto point 5 above for items within behaviour change technique category per visit and across visits

To analyse the open questions on the checklist we will:

1. Extract a sample of 10% of the data ensuring equal representation from BF and PFMT arms.
2. Use content analysis to develop a coding frame from that data.
3. Apply the coding frame to the data set.
4. Describe the codes narratively.

Finally we will interpret the findings through discussion.

5.2 Consultation recordings

Sample as shown in table below.

Table Three: Sample profile of audio-recordings of consultations

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Total
Basic	8	6	7	6	6	12	45
Intensive	10	6	5	6	4	11	42
Total	18	12	12	12	10	23	87

To analyse consultation recordings we will:

1. Develop a framework for the recordings that references the checklist components and allows assessment of component quality.
2. Listen to recordings and complete the framework.
3. A 10% sample of audio-recording frameworks completed by experienced analysts to assess inter-rater agreement.
4. Enter data into SPSS.

If appropriate proceed to

5. Describe the number (percentage) of use of each component in the framework and the component quality for each visit.
6. Score per behaviour change technique category per visit and across visits
7. Score per IMB category (information, motivation, behaviour).
8. Interpret the analysis.

Cross checking checklists and consultation recordings

Sample: women where we have audio-recording and the checklist that matches to the recording.

To assess the links between the audio and the checklist we will:

1. Identify the relevant checklists and link the data using participant ID numbers.
2. Check differential coder functioning
3. Match the relevant checklist items to the equivalent components on the audio-recording framework
4. Count the matches, report according to core, additional, by group, by therapist, by session
5. Interpret the analysis.

5.3 Interviews with Healthcare Professionals

Sample: Interviews with intervention therapists where possible and where not possible with other trial staff at sites. Thirty interviews conducted: 26 therapists, one nurse delivering OPAL and two nurses and one administrative person involved in a variety of tasks such as recruitment, consenting patients and dealing with IT issues. (Across 21 sites)

Data from interviews with therapists will be analysed using the Framework Approach [Ritchie and Spencer 1994] as follows.

1. Familiarisation with the data
2. Development of a thematic framework using a priori and inductive codes
3. Thematic framework applied across the dataset.
4. Framework matrices developed
5. Conceptual maps used to make links between themes.

Analysis will be discussed by analyst and grant holders to support the interpretation of the data.

5.4 Interviews with Women (case study)

Interviews are longitudinal with a purposive sample of women who were part of the intervention.

Table Four: Sample of women recruited into the case study

	Baseline	**6 months**	**12months**	**24months**	**Case complete**
Basic	20	16 (4 lost)	15 (1 Lost)	14 (1 Lost)	14 (6 lost)
Intensive	20	17 (3 lost)	13 (4 Lost)	11(2 Lost)	11 (9 lost)
Total	40	33 (7 lost)	28 (5 Lost)	25 (3 Lost)	25 (15 lost: 7 only baseline data; 4 only baseline and 6 month data)

The process of analysis for the case study data will follow the following principles (Yin, 2014):

1. Each individual interview transcribed and entered into Nvivo
2. Individual interviews read and coded according to a priori coding scheme
3. 10% of interviews will be coded by a second analyst
4. Inductive codes developed as seen within the data
5. Case summaries documented for complete cases, further inductive codes developed.
6. A framework table completed to summarise the cases.
7. Interim reports written that explain the data in relation to the key research questions.
8. Theoretical propositions developed.
9. Key features of data within a tail (trial arm) are explained
10. Tails are compared based on theoretical propositions.

5.6 Exercise Diaries

Data sample: 'open' questions in exercise diaries (qualitative data) and diary entries relating to number and/or frequency of exercising (quantitative data).

For the qualitative data we will:

1. Extract a sample of 10% of the data ensuring equal representation from BF and PFMT arms.
2. Use content analysis to develop a coding frame from that data.
3. Apply the coding frame to the data set.
4. Describe the codes narratively.
5. Analyse as above for 'ticklist' items in checklist.
6. Interpret the data

To analyse the quantitative data diary entries we will:

7. Descriptively summarise entries (dates exercised / sessions per date) for each diary issued / for all diaries; and by group (basic and biofeedback).
8. Descriptively summarise number of additional agreements (and whether signed by patient / clinician) per diary issued and for all diaries, and by group (basic and biofeedback).
9. Descriptively summarise exercise programme prescribed (hold-relax duration and repetitions; quick release repetitions; sessions per day / days per week; positions used for exercising (lying, sitting, standing), per diary issued / for all diaries and by group (basic and biofeedback).
10. Interpret the data

5.7 Main study questionnaires

The analysis of data arising from the main study questionnaires that will be used within the process evaluation analysis are as documented within the SAP. The main SAP documents that "we will consider investigating whether the primary outcome is mediated by self-efficacy and adherence. A full quantitative mediation analysis is being considered as a secondary analysis, which would be carried out after the main trial has been completed" and is therefore not part of the initial process evaluation report. For completeness however a draft mediational analysis plan is included as an appendix to this plan (please see Appendix 1).

6.0 Data synthesis across data sources

The synthesis will focus on the research questions as defined in the table above (*i.e.* the rows). Synthesis will occur as follows (O'Cathian et al, 2010):

1. All analysis for the relevant RQ from each individual data source will be re-read.
2. A matrix will be constructed that outlines the data sources at the top and the appropriate RQ as the row for each main construct (e.g. delivery to individuals) – see example below.
3. Each source data will summarised in the appropriate cell (term silence will be used when there are no relevant data from that source).
4. The agreement/ disagreement across sources will be documented in the final column.
5. A summary of key findings outlining explanation and meaning for each concept will be documented.

Table Five: Example triangulation matrix for delivery to individuals

RQ nos	**RQs**	**Checklists (TAF)**	**Consultation recordings**	**HCP interviews**	**Women interviews**	**Exercise diaries**	**Biofeed back down load**	**Main study Qnaire**	**Agree/ disagree & commentary**
1b	What treatment is delivered? (enactment treatment delivery)	*Each cell contains a summary of the data from that source*	*If there are no data from the source 'silence' will be documented*						*Extent to which there is agreement across sources*
1b, 4	What do participants learn? (enactment treatment receipt)								
1b	Does treatment delivery match intended protocols?								
5, 7	What are the barriers and facilitators to treatment delivery?								
5, 7	What are the barriers and facilitators to treatment receipt?								
1a, 2,3	What elements of delivery influence outcome?								

7.0 Summary

This document outlines the analysis planned for the qualitative study and process evaluation associated to the OPAL trial. It demonstrates how each data source will be managed and how the data will be brought together in the final synthesis.

8.0 References

GRANT A, TREWEEK S, DREISCHULTE T, FOY R, GUTHRIE B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. *Trials.* 2013. 14:15

O'CATHIAN A, MURPHY E, NICHOLL J. Three techniques for integrating data in mixed methods studies. *BMJ.* 2010. 341:4587-4595

RITCHIE, J. and SPENCER, L., 1994. Qualitative data analysis for applied policy research. In: A. BRYMAN and R. BURGESS, eds, *Analysing Qualitative Research.* London: Routledge, pp. 173-194.

YIN, R.K., 2014. *Case Study Research: Design and Methods. 5th Ed.* London: Sage Publications.

Appendix 1: Draft example mediational analysis

This element of the process evaluation aims to provide insight regarding whether the proposed theoretical mechanisms of change (the mediating variables) help explain the trial outcomes. We will undertake a mediational analysis using statistical modelling techniques (e.g. based on Item Response Theory or Structural Equation Modelling) and bootstrapping techniques. As the main underpinning theoretical mechanism by which change might occur was postulated as self-efficacy the remaining text in this section will focus on describing an example mediational analysis plan using self-efficacy as the hypothesised primary mediating variable.

Data requirements

The main trial Statistical Analysis Plan (SAP) provides a description of all the measures collected that may be used in the mediational analyses.

In our chosen example the mediating variable will require raw score data for self-effciacy at all time points. The outcome variable(s) will be the trial primary outcome (and if appropriate secondary outcomes). The moderating variables will be demographic / baseline assessment data (e.g. age, condition severity). Data for the mediational analyses will have already been entered, cleaned and any missing values imputed as appropriate. Data to be transferred after the main SAP has been actioned.

Analysis procedure

1. We will provide a summary describing any remaining missing data, for each item and for each of the constructs, within the self-efficacy data set at each time point.
2. In order to confirm that the self-effciacy scores provide reliable data scores for the mediational analysis, the psychometric properties of the self-effciacy baseline and follow-up data will be examined and analysed (using factor analysis, and if appropriate, Rasch analysis).
3. Reliability tests will also be carried out on the baseline self-effciacy data to check the internal consistency of each construct before use in the subsequent mediational analysis.
4. A descriptive summary will be given of self-efficacy scores at each time point as well as change scores across time. Depending on the trial result (see step 5) the summaries will be presented as either a whole cohort or split to reflect the two trial groups.
5. A parsimonious model will be built to answer pre-specified hypotheses, in this example if there is a positive trial result the main hypothesis will be:

   Do changes in women's self-efficacy between baseline and 2 years explain the difference in outcomes between the BF PFMT group and the Basic PFMT group at 2 years? (see Figure 1 for diagramtic representation)
6. If there is a null trial result, the data from the two groups will be pooled and the hypothesis will be:

   Do changes in women's self-effciacy between baseline and 2 years explain their outcomes at 2 years?

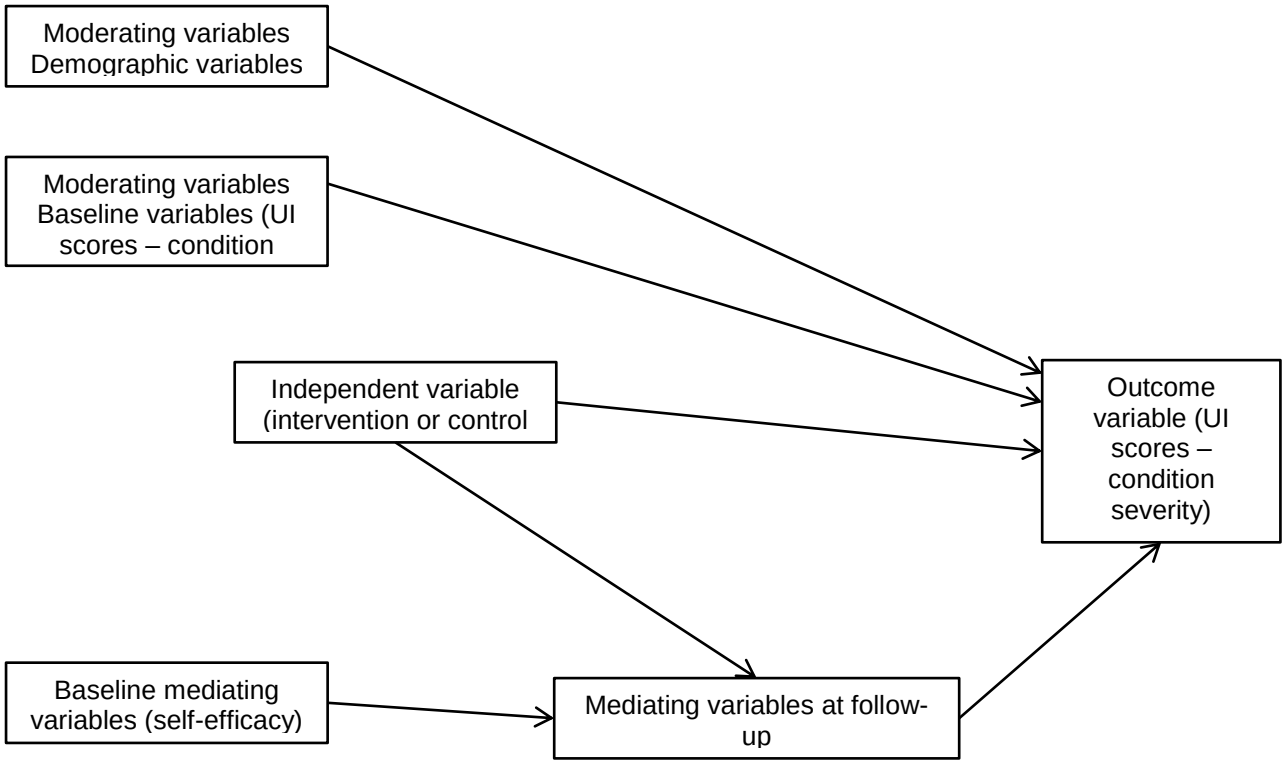


Figure 1 Diagramatic representation of the mediational model for hypothesis (point 5): Do changes in women's self-efficacy between baseline and 2 years explain the difference in outcomes between the BF PFMT group and the Basic BFMT group at 2 years?

Additional modelling options will be explored as appropriate, for example attendance and adherence data may be included depending whether a Complier Average Causal Effect (CACE) analysis is undertaken in the main statistical analysis. Our intention will be to add complexity or nuance to the modelling but avoid duplication of any analysis already undertaken.