Statistical Analysis Plan

Title: Efficacy of a novel sensory discrimination training device for the management of phantom limb pain: protocol for a randomised placebo-controlled trial.

Authors: Sarah Oatway, Denis Martin, Andrew Graham, Alan Batterham, Alasdair MacSween, Sally Smith, Dr Deepak Ravindran, Cormac Ryan

Version: 21
IRAS Project ID: 250337
Sponsor Study Number: 16692

Statistical analysis

The primary outcome data (SF-MPQ-2 total score) will be analysed using an ANCOVA model, conditioning on the baseline value of the outcome, in a linear mixed model with restricted maximum likelihood and Satterthwaite degrees of freedom. This model is one of the principled approaches to dealing with missing outcome data under a plausible missing at random assumption; provided a participant has outcome data for at least one of the two timepoints then they will be included in the analysis. We do not plan to conduct any missing-not-at-random sensitivity analysis via multiple imputation, as there is no strong argument to suggest that participants lost to follow up would have fared worse with respect to the primary outcome. Diverse reasons for loss to follow up may include withdrawal on safety grounds (inability to feel the stimulation), unrelated adverse events, adverse device effects (or perception thereof), lack of motivation, competing caring or time constraints, treatment burden, complexity of device set up, and impact of unrelated medical issues. Given the nature of this trial, we do not anticipate missing baseline data but will apply a suitable approach if so.[21] Fixed effects will be time (3 weeks and 3 months), treatment group, and the time-by-treatment interaction, adjusting for baseline SF-MPQ-2. Alongside the fixed effect for baseline, we shall include a baseline-by-time interaction, to allow for a different association between baseline and outcome at each timepoint. To preserve the desired Type 1 error control, an indicator variable (fixed factor) for membership of the four strata defined by the two, two-level stratification factors will be included in the model.[3] We specify the most general and flexible 'unstructured' residual correlation matrix accounting for the repeated measurements nested within participant, imposing no constraints and estimating a variance for each timepoint and their covariance. Therefore, the random participant intercept is omitted ('noconstant'). The primary endpoint will be the estimate of the effect of the intervention at 3 weeks. We shall present the point estimate together with 95% confidence interval and P-value, via t-based inferences using Satterthwaite degrees of freedom. All analyses will be conducted using Stata software (current version: (StataCorp, 2025. Stata Statistical Software: Release 19, College Station, TX, USA: StataCorp LP). Sample draft Stata code for this primary analysis is provided below:

```
mixed sfmpq2 i.group##i.time c.baseline##i.time i.indicator||id:, noconstant residuals(unstructured, t(time)) reml dfmethod(satterthwaite)
```

The behaviour of the model residuals will be inspected visually using residuals plots, with data transformations applied if indicated. An intention-to-treat analysis (primary; participant included if they have outcome data for at least one of the two timepoints) and a per-protocol analysis (secondary) will be undertaken. In the per-protocol analysis the participant will be required to have undertaken at least 10 sessions of sensory discrimination training totalling (600 minutes) to be considered suitable for inclusion in the analysis. The primary analysis will be carried out by a statistician blinded to group assignment.

The three secondary outcome measures will be treated as continuous and analysed in the same way, but obviously with no baseline value for the General Subjective Outcome Score (GSOS). The secondary outcomes are considered equally important, after a significant primary. Type I error for the analysis of the secondary outcomes will be

preserved using the Holm multiple testing algorithm.[6 14 16] Study diary data will be summarised using descriptive statistics.

Secondary analysis: Secondary analysis (non-blinded) will be carried out to explore treatment heterogeneity. By inclusion of appropriate interactions with treatment group, we will explore putative effect modifiers, including baseline value of outcome, age, sex, total treatment time (minutes) with the device, chronicity, and whether the amputation was due to trauma or planned surgery. The trial is not powered for quantifying treatment heterogeneity, so we emphasise that this secondary analysis will be purely exploratory/ descriptive.

Qualitative analysis: A purposive sample of participants from the active intervention group will be invited to undergo a semi-structured interview. During these interviews, participants will be asked about the usability and acceptability of the SP1X device. The interviews will be audio recorded, transcribed verbatim, and analysed thematically. Thematic analysis will be undertaken by the research assistant (RA) using the method of Braun and Clarke.[1]

Sample size

We regard the maximum feasible sample size within the constraints of this trial as a total of 100 participants. With an estimated loss to follow-up of 20%, the effective sample size is 80 participants (40 per group in a 1:1 allocation ratio). Our targeted effect size is a difference in follow-up (3-week timepoint) SF-MPQ-2 total score between arms of 1 unit – a moderate clinically relevant effect size. The between-subjects variability (SD) was derived from a small reliability study with 37 participants drawn from the same population as the proposed trial (unpublished observations). The observed SD was 2.13 units. The pre-post correlation (with one week between test and retest) was 0.82. The pre-post correlation (the reliability of the measure) is expected to decrease over time. Estimates for the decline in correlation with more distant pairs of timepoints post-randomisation range from -0.003 per month[20] to -0.009 per month.[7] We have assumed a conservative value for this correlation of r=0.7 at the 3-week primary endpoint. An effective sample size of 40 participants per arm gives c.84% power at 2-sided P=0.05 to detect the targeted effect size. This sample size estimate was derived assuming an ANCOVA analysis model, adjusting for the baseline value of the outcome measure. To allow reproducibility of our sample size estimate, the relevant Stata code is provided herein (StataCorp, 2023. Stata Statistical Software: Release 18, College Station, TX, USA: StataCorp LP): *sampsi 5 4, sd1(2.13) sd2(2.13) method(ancova) pre(1) post(1) r01(0.7) n(40) n2(40).*

Randomisation

The RA will use the Sealed Envelope™ remote randomisation software system to randomise participants (1:1) to one of two trial groups: 1) Intervention Group (SP1X) 2) Placebo Group (SP1X7). A unique participant identification code (e.g., 001, 002) will be allocated to all participants. The allocation will be stratified by sex (male/ female) and chronicity (<6 months/ ≥6 months post-surgery), generating four separate strata. Within these strata, participants will be allocated using a mixed (unrestricted and restricted) randomisation approach. This approach uses an initial simple (unrestricted) randomisation to provide unpredictability and guard against guessing[18] followed by allocation using randomly permuted blocks to achieve reasonable balance, resulting in approximately 50 participants per arm overall. More explicit details of the randomisation scheme are not provided herein, to avoid the facilitation of the deciphering of the allocation sequence.

Intervention

There will be two parallel groups, an intervention group (SP1X) and a placebo group (SP1X7). Both the intervention and placebo device will be provided to participants for use in their own homes. During an online video call the participant will be instructed how to set up and use the device by the RA (an HCPC registered Physiotherapist). The RA will observe the first full treatment session (60 minutes) and up to 30 minutes of the second treatment, and provide any additional advice needed. Participants in both groups will be asked to use the device for 60 min/day, as one block, or as multiple shorter sessions of ≥20 minutes duration on at least 15 days of the 21-day treatment period. They will be asked to spread out the use of the device, over the treatment period and to record their device use in a study diary (figure 1).

Primary and secondary outcome measures

Primary Outcome

The primary outcome will be the total score of the McGill Pain Questionnaire revised (SF-MPQ-2)[5 12] at the 3-week time point. The SF-MPQ-2 is a commonly used questionnaire to assess pain levels in a range of pain conditions. There are 22 items/ pain descriptors across 4 pain sub-scales/ domains: continuous, intermittent, neuropathic, and affective. Participants rate each item on an 11-point (0-10) scale, where 0 = none and 10 = worst possible pain. The mean of the 22 items provides the SF-MPQ-2 total score. Two or more missing responses on any sub-scale results in an invalid outcome. A targeted effect size of 1 unit difference between groups will be used.

Secondary Outcomes

The following secondary outcomes are specified at the 3-week timepoint (immediately post intervention):

- Overall Pain Intensity Score: Visual Analogue Scale (100mm)[19]
- Frequency adjusted pain score: (0-100)
- General Subjective Outcome Score (GSOS)[11]

Exploratory Outcomes

No inferences will be drawn from these outcomes.

- SF-MPQ-2 total score at the 3-month time point.
- TAPES (modified)[8 9]
- Quality of life [EQ-5D-5L][4 13]
- Sleep Disturbance [the PROMIS Short Form v.1.0 – Sleep Disturbance 4a questionnaire][10]
- Participant satisfaction [15]
- Device Usability [an adapted version of the 10 question System Usability Scale (SUS)].[2]
- Study Diary of device and medication usage
- Concordance with Protocol
- Success of blinding

References

1. Braun, V. and Clarke, V., 2006. Using thematic analysis in psychology. Qualitative research in psychology, 3(2), pp.77-101.
2. Brooke, J. (1996). Sus: a “quick and dirty’ usability. *Usability evaluation in industry*, *189*(3), 189-194.
3. Bugni FA, Canay IA, Shaikh AM. 2018. Inference under covariate-adaptive randomization. Journal of the American Statistical Association. 113(524):1784-1796.
4. Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & Van Hout, B. (2018). Valuing health-related quality of life: An EQ-5 D-5 L value set for E ngland. *Health economics*, *27*(1), 7-22
5. Dworkin, R. H., Turk, D. C., Revicki, D. A., Harding, G., Coyne, K. S., Peirce-Sandner, S., ... & Melzack, R. (2009b). Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). *Pain®*, *144*(1-2), 35-42.
6. Food and Drug Administration. Multiple end points in clinical trials guidance for industry. https://www.fda.gov/media/162416/download. Accessed February 10, 2025
7. Frison, L., & Pocock, S. J. (1992). Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Statistics in medicine*, *11*(13), 1685-1704.
8. Gallagher, P., MacLachlan, M. (2000). Development and psychometric evaluation of the Trinity Amputation and Prosthesis Experience Scales (TAPES). Rehabilitation Psychology, 45, 130-154
9. Gallagher, P., Franchignoni, F., Giordano, A., & MacLachlan, M. (2010). Trinity amputation and prosthesis experience scales: a psychometric assessment using classical test theory and rasch analysis. American Journal Of Physical Medicine & Rehabilitation, 89(6), 487-496.
10. Gershon, R. C., Rothrock, N., Hanrahan, R., Bass, M., & Cella, D. (2010). The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *Journal of applied measurement*, *11*(3), 304.

11. Harland, N. J., Dawkin, M. J., & Martin, D. (2015). Relative utility of a visual analogue scale vs a six-point Likert scale in the measurement of global subject outcome in patients with low back pain receiving physiotherapy. *Physiotherapy*, *101*(1), 50-54.
12. Hawker, G. A., Mian, S., Kendzerska, T., & French, M. (2011). Measures of adult pain: Visual analog scale for pain (vas pain), numeric rating scale for pain (nrs pain), mcgill pain questionnaire (mpq), short-form mcgill pain questionnaire (sf-mpq), chronic pain grade scale (cpgs), short form-36 bodily pain scale (sf-36 bps), and measure of intermittent and constant osteoarthritis pain (icoap). *Arthritis care & research*, *63*(S11), S240-S252.
13. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research, 20(10), 1727-1736.
14. Holm SA, 1979, A simple sequentially rejective multiple test procedure, Scandinavian Journal of 52 Statistics, 6(2): 65-70.
15. NHS Choices (2020) Friends and Family Test (FFT), NHS. Available at: http://www.nhs.uk/using-the-nhs/about-the-nhs/friends-and-family-test-fft/.England, N. H. S. (2014). Friends and Family Test in general practice guidance.
16. Pocock SJ, Rossello X, Owen R, Collier TJ, Stone GW, Rockhold FW. Primary and Secondary Outcome Reporting in Randomized Trials: JACC State-of-the-Art Review. J Am Coll Cardiol. 2021 Aug 24;78(8):827-839.
17. Ryan, C., Harland, N., Drew, B. T., & Martin, D. (2014). Tactile acuity training for patients with chronic low back pain: a pilot randomised controlled trial. BMC Musculoskeletal Disorders, 15(1), 1-11
18. Schulz, K. F., & Grimes, D. A. (2002). Unequal group sizes in randomised trials: guarding against guessing. *The Lancet*, *359*(9310), 966-970.
19. Scott, J. and Huskisson, E.C., 1976. Graphic representation of pain. *pain*, *2*(2), pp.175-184.
20. Walters, S.J., Jacques, R.M., dos Anjos Henriques-Cadby, I.B., Candlish, J., Totton, N. and Xian, M.T.S., 2019. Sample size estimation for randomised controlled trials with repeated assessment of patient-reported outcomes: what correlation between baseline and follow-up outcomes should we assume? Trials, 20, pp.1-16.
21. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. Stat Med, 2005;24:993–1007.