STATISTICAL ANALYSIS PLAN

THE *ACTIVE* TRIAL: A PROSPECTIVE RANDOMISED CONTROL TRIAL OF THE H1 IMPLANT VERSUS TOTAL HIP REPLACEMENT

A PROSPECTIVE, RANDOMISED (1:1), DOUBLE-BLINDED, MULTI-CENTRE STUDY

Study Sponsor: Embody Orthopaedic Limited

Trial Registration Numbers: ISRCTN10665654, NCT06162195

IRAS ID: 327954

REC Approval Number: TBA

Revision history

See below for descriptions of changes. Refer to the header stamp for the date of approval of the most recent version.

Version	Study Protocol Version	Description and reason for change(s)
1.0	1.0	First issue
1.01	1.03	Addition of trial registration numbers

Roles & responsibilities

This document was written, reviewed and electronically approved by the following people:

Name	Role
Dr Camilla Halewood	SAP writer
Dr Susannah Clarke	Sponsor

Contents

1 Background & rationale

The H1 Implant is a cementless, ceramic-on-ceramic hip resurfacing arthroplasty (HRA) device (Figure 1). The H1 Implant that has so far been used exclusively in a Clinical Investigation sponsored by Embody Orthopaedic Limited (Embody) (ISRCTN91554748, NCT03326804). The H1 Implant is set to receive CE marking in early 2024. This statistical analysis plan (SAP) relates to a randomised control trial (RCT) being sponsored by Embody to ascertain whether the composite clinical success of the H1 Implant is non-inferior to cementless total hip replacement (THR): "The ACTIVE Trial". The RCT is being managed by Imperial College London, acting as a contract research organisation (CRO).

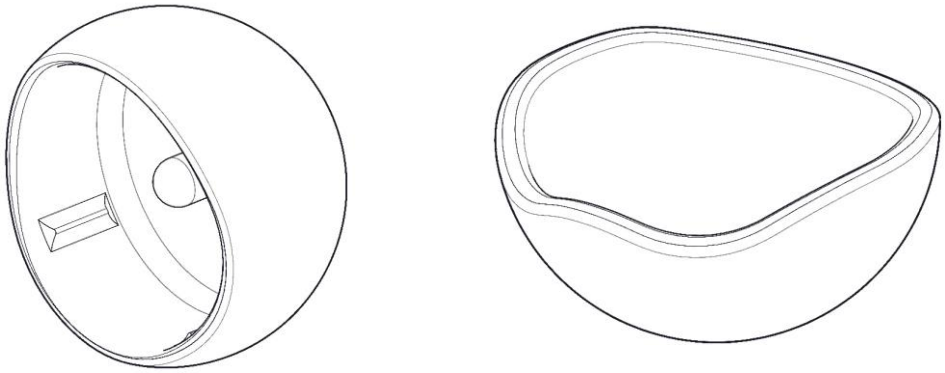

Figure 1. Schematic drawings of the H1 femoral head (left) and acetabular cup (right)

2 Specific objectives and hypotheses

2.1 Research hypothesis

The H1 Implant is non-inferior to cementless total hip replacement (THR), in terms of composite clinical success (CCS) at 24 months post-op.

2.2 Study objectives

The primary objective of this trial is to determine whether the H1 Implant is non-inferior to cementless THR in terms of CCS.

The key secondary objective is to determine whether the H1 Implant is superior compared to cementless THR in terms of physical activity and patient report outcome measures (PROMs).

Other secondary objectives are to compare the H1 Implant and cementless THR with respect to:

- Safety, through collection of all procedure and device related adverse events
- Noise generation, through patient survey.

3 Trial methods

3.1 Trial design

The ACTIVE Trial is designed as a randomised (1:1), controlled, assessor- and patient-blinded multicentre non-inferiority trial with two parallel groups and a primary endpoint of composite clinical success 24 months after surgery. The parallel groups will be:

1) **Experimental:** The H1 Implant;
2) **ACTIVE comparator:** cementless total hip replacement (THR).

3.2 Randomisation and blinding

Randomisation will be performed using variable block randomisation with block sizes of 4, 6 and 8, with a 1:1 allocation to the 2 groups, with stratification across sites. This will ensure approximately equal numbers across groups and approximately the same number per group at each site (to control for differences in the trial population because of environmental, social and demographic factors), while allowing different total numbers at each site. It will also ensure that at any given time, the numbers in each group will be approximately equal, allowing the interim analyses to take place. The randomisation will be managed by the CRO via an electronic data capture (EDC) system, which uses a validated randomisation algorithm (Castor EDC). Once a patient has been consented, they will be assigned to a treatment group according to the pre-determined order inside the block. Stratification by site is recommended for multi-centre studies[1, 2]. The patients, clinical assessors and data analysts will be blinded to the intervention each patient has received for the 24 month post-operative period. The surgeon, operating theatre team, radiographic analysts and CRO cannot be blinded but will not be involved in the care of the patient after the intervention unless the patient requires revision surgery.

3.3 Sample size

The sample size was calculated on the basis of the research hypothesis. Using pilot data from two sources, the CCS rates for cementless THR and the H1 Implant were estimated to be 91%[3*] and 92.7%[4], respectively. With a non-inferiority margin of 8% and setting the probability of a type I error to 5% and power to 80% and using the following formula:

$$n = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{P_{H1} - P_{THR} - \delta}\right)^2 (P_{H1}(1 - P_{H1}) + P_{THR}(1 - P_{THR}))$$

where:

n = required sample size
P_{H1} = probability of success for H1
P_{THR} = probability of success for THR
α = probability of a type I error
β = probability of a type II error
δ = non-inferiority margin
z = is the cumulative distribution function of a standardised normal deviate

*This CCS for the cementless total hip was calculated from the publication referenced, but excluded intra-operative ceramic liner fractures because these have been practically eliminated with modern day ceramic (BIOLOX®*delta)* and do not occur with polyethylene cups. The control in the RCT will be cementless BIOLOX®*delta*-on-BIOLOX®*delta* or cementless BIOLOX®*delta*-on-PE THRs.

The required sample size is 93 per group. Allowing for drop-outs, this gives a sample size of 100 per treatment arm, therefore a total of 200 patients will be recruited to the study. Assuming the cementless THR group goes on to match this pilot data of 91% success, 90% of H1 patients would be required to demonstrate composite clinical success to be non-inferior to the THR patients, with the margin of 8%.

3.4 Framework

The study protocol states that the key secondary objective is to determine whether the H1 Implant is significantly different compared to cementless THR in terms of physical activity (measured via pervasive activity monitoring and physical performance assessment) and PROMs (the University of California at Los Angeles (UCLA) Activity Score and the Hip Outcome Score (HOS)). Therefore, the key secondary outcomes will be tested for significant differences rather than non-inferiority as for the primary outcome.

For the pervasive physical activity monitoring, with 100 patients in each group (determined by the primary outcome) and with THR patients expected to achieve a daily average of approximately 258 minutes of low, moderate and vigorous activity at the 12 months post-op timepoint[5], the H1 patients will have to achieve 287 minutes to demonstrate a significant difference (using the using the standard deviation of 100 minutes for the general population[6]). If the H1 patients reach 338 minutes (thereby achieving maximum mortality risk reduction compared to the referent[6]), significant difference will be demonstrated with >99% power.

In addition, periods of “bouted” MVPA are of interest as there is some evidence that, below a certain amount of MVPA per day, if a greater percentage of MVPA is conducted in 10-minute bouts, there is a greater risk reduction in overall mortality[7]. With 100 patients in each group, H1 patients would have to achieve 60 minutes of 10-minute bouted MPVA per day to demonstrate a significant difference (90% power) compared to the approximately 48 minutes (SD 42 minutes) of 10-minute bouted MVPA per day estimated to be achieved by THR patients at the 1 year post-op timepoint[5].

The other secondary outcomes (safety and results of the noise assessment) will just be reported for both groups, no statistical analysis will take place.

3.5 Statistical interim analyses and stopping guidance

3.5.1 Interim analyses

Three formal statistical interim analyses are planned on the primary and key secondary endpoints. These interim analyses are planned to take place when at least half the number of patients recruited for the final analysis have reached the 6 month, 12 month and 24 month post-operative time points (i.e. 50 H1 patients and 50 cementless THR patients have reached those time points, the CCS has been calculated and they have completed 2 weeks of activity monitoring).

3.5.2 Planned adjustment of significance level due to interim analysis

Because multiple statistical comparisons will be performed (interim analyses and final analysis), the risk of false positive (type I) error will be inflated. Therefore a group sequential α-spending function shall be used, calculated using the O’Brien-Fleming method, with two-sided symmetric bounds[8]. This gives p-values of 0.00005, 0.0039, 0.0184 and 0.0412 at the three interim analyses and the final analysis, respectively.

3.5.3 Details of guidelines for stopping the trial early

Both the experimental arm (the H1 Implant) and the active comparator arm (cementless THR) are CE marked devices with proven safety and efficacy and by the time the first interim analysis is carried out,

most of the patients will have received their intervention. Therefore, the statistical results from the interim analyses alone will not be used to stop the trial early. However, if the interim analyses are cause for concern, safety information, as measured by adverse events and adverse device effects, will be closely assessed by the CRO and the sponsor to determine whether it can be justified to continue follow-up of the patients in the study.

3.6 Timing of final analysis

The final analysis and preparation of the final report for the H1 Implant versus cementless THR comparison is planned to take place when all 200 patients have reached the 24 month post-operative time point and data for the primary and key secondary endpoints have been received. This is anticipated to be in June 2026.

3.7 Timing of outcome assessments

The schedule of all of the study procedures is given in the study protocol[9].

4 Statistical principles

4.1 P-values and confidence intervals

4.1.1 Level of statistical significance

All applicable statistical tests will be either 1-sided or 2-sided depending on the outcome and will be performed using a 5% significance level.

4.1.2 Planned adjustment for multiplicity

4.1.2.1 Multiple secondary endpoints

In order to control the overall Type I error related to multiple secondary endpoints, multiplicity adjustment shall be made using the multi-step, step-up Hochberg procedure[10] for the likely-correlated secondary endpoints that will be undergoing statistical analysis (Table 1).

4.1.2.2 Interim analyses

As described in Section 3.5.2, p-values will also be adjusted to allow for interim analyses.

4.1.2.3 Subgroup analyses

The subgroup analyses described in Table 1 are designed to investigate the sensitivity of the conclusions drawn from the overall analysis, thereby increasing the confidence in the results obtained from the primary analysis and therefore will not be subjected to adjustment for type I error due to multiplicity concerns[11].

4.1.3 Planned adjustment for stratification

Because stratification will be used to ensure that at each surgical site there are equal numbers of patients in each group, the Cochran-Mantel-Haenszel (CMH) test will be used to generate an estimate of the association between implant type (H1 or THA) and achievement of CCS (achievement or not) while taking into account the possible confounding variable of surgical site. The data will be stratified into 4 levels for the 4 sites to create a series of 2x2 tables showing the association between implant type and CCS for each site and a weighted average of the odds ratios. The CMH test statistic will be calculated and p-value determined, with the null hypothesis being that all the odds ratios are equal to 1. Following this test, Woolf's heterogeneity test will be used to test the null hypothesis that all the odds ratios are equal to each other (even if they are not all equal to 1).

4.1.4 Confidence intervals to be reported

All confidence intervals presented will be 95% and two-sided.

4.2 Adherence and protocol deviations

4.2.1 Definition of adherence to the intervention and how it will be presented

Compliance will be assessed based on the % of patients who have attended all follow-up visits, completed the required questionnaires and assessments, and worn the activity tracker for the requested amount of time. Individual % compliance figures shall be calculated for each type of activity at each time point. Descriptive statistics on the percent compliance (N, mean, SD, median, minimum, maximum) will be summarised by randomisation group, time-point and type of activity.

4.2.2 Definition of protocol deviation for the trial

A protocol deviation is defined as a failure to adhere to the study protocol. Examples are:

- Errors in applying inclusion/exclusion criteria
- The wrong intervention being administered. If this occurs, the site must inform the CRO immediately so that the randomisation block can be adjusted accordingly.
- Patient not wearing the activity tracker at all
- Incorrect/lack of data being collected and documented
- Missed follow-up visits
- Patient is accidentally unblinded.

A protocol deviation should be defined as major or minor. A deviation should be considered to be major if it affects efficacy, the safety, physical or mental integrity of the participants in the trial, or the scientific value of the trial. More details about protocol deviations and a full list are contained in the study protocol[9].

4.2.3 Description of which protocol deviations will be summarised

The number (and percentage) of patients with major and minor protocol deviations will be summarised by treatment group with details of type of deviation provided. No formal statistical testing will be undertaken on protocol deviations.

4.3 Analysis populations

4.3.1 Definition of analysis populations

Once a patient receives the intervention into which they were randomised, their group is fixed, because this study concerns an implantable device. There are possible protocol deviations, but none that should affect the analysis populations except for the wrong intervention being administered. If this happens, another random patient's intervention should be switched, to avoid unequal sized groups. Therefore the analysis populations will include all randomised patients according to the treatment that they received.

5 Trial population

5.1 Screening data

Full screening details are contained within the CIP[9]. The following summaries will be presented:

- the number of days spent recruiting;
- the number of patients screened;
- the number of patients recruited;
- the number of patients recruited per day;
- the number of screened patients not recruited and the reason for non-recruitment.

This summary will be provided overall and by each centre.

5.2 Eligibility

The trial inclusion and exclusion criteria are specified in the CIP[9].

5.3 Recruitment

A CONSORT[12] diagram will be created, comprising the number of people screened, eligible, consented, randomised, receiving treatment, withdrawing and lost to follow-up, with reasons provided (Figure 2).

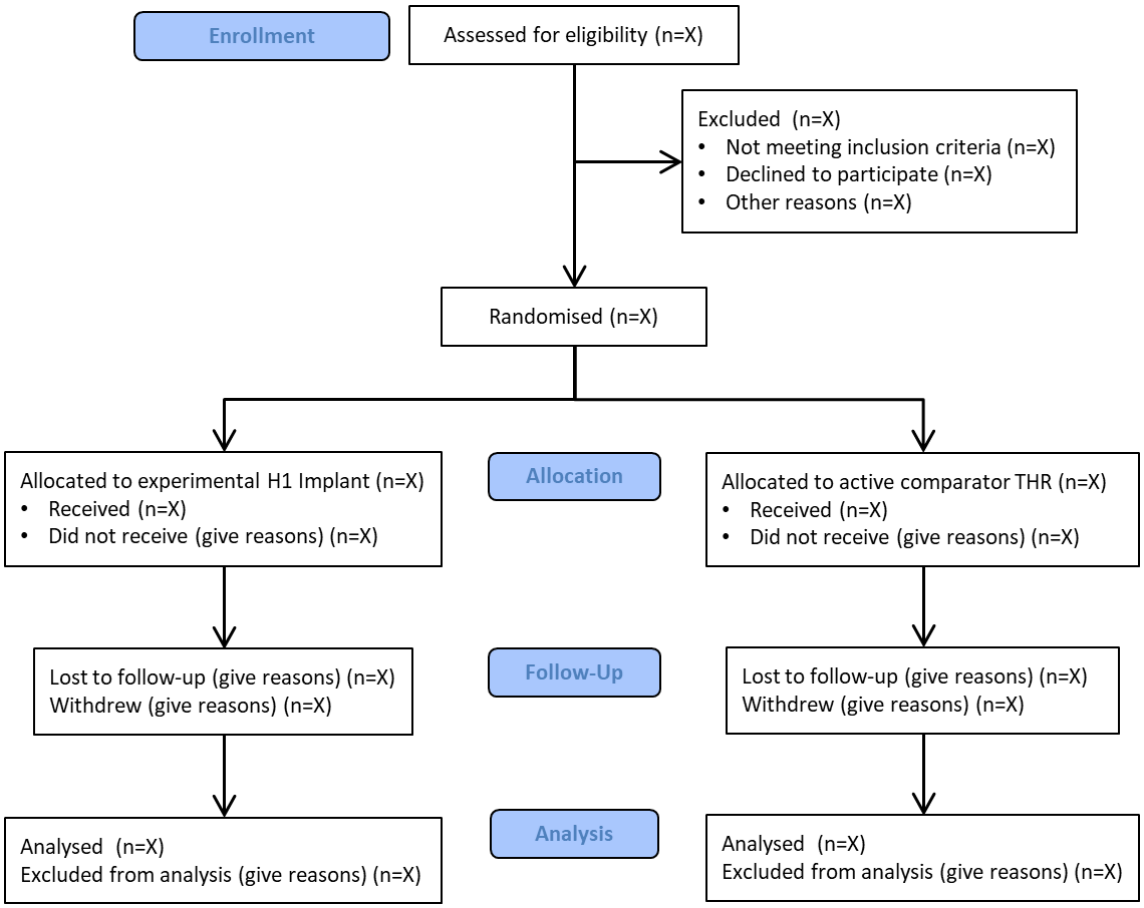

Figure 2. CONSORT diagram template

5.4 Withdrawal/follow-up

The level of consent withdrawal shall be tabulated and classified. Because this study involves implantable devices, patients cannot elect to withdraw from the intervention once it has been administered. Therefore the following types of withdrawal are possible:

- withdraw from the intervention after consent but before surgery;
- withdraw from follow-up but allow data collected to date to be used;
- withdraw from follow-up and withdraw consent for data collected to date to be used;
- lost to contact/follow-up;
- withdraw due to being revised (that is, the device has to be removed);
- patient death.

The classification, timing of and reason for each withdrawal or lost to follow-up shall be tabulated.

5.5 Baseline patient characteristics

Patients will be described with respect to age at surgery, sex, hip side being treated, current and historical physical activity, Modified Harris Hip Score, height, weight, both overall and separately for the two randomised groups. In addition, type and size of device shall be recorded once the intervention has taken place. Categorical data will be summarised by numbers and percentages. Continuous data will be summarised by mean, standard deviation and range if data are normal and by median, interquartile range and range of data are non-normal. Statistical tests shall not be undertaken for baseline characteristics, but the clinical importance of any imbalance between groups will be noted.

6 Analysis

6.1 Outcome definitions

The primary outcome is "composite clinical success". This is a binary outcome measure. Patients are deemed to have achieved CCS if they meet all of the following criteria at the 24 month timepoint:

- Modified Harris Hip Score[13, 14] ≥80.
- No revision or pending revision.
- Acetabular radiolucencies: not in all zones.
- Femoral radiolucencies: not in all zones.
- Absence of subsidence/migration of the acetabular or femoral component >5mm with clinical findings.
- Absence of serious adverse device effect (SADE).

The secondary outcomes are:

- Total activity count (AC), calculated as the total time spent in non-sleep and non-sedentary activity (in minutes per week) and minutes (in bouts of at least 10 minutes) of moderate to vigorous physical activity (MVPA) achieved during a week, as measured by the GENEActiv activity tracker over a period of 14 days. Patients will be asked to wear the GENEActiv for the full 14 days, but some non-wear is expected, so for the individual patient's data to be valid, they must wear the device for at least 2 weekdays and 2 weekend-days, for a minimum of 2/3 of their waking hours on each of those days[15]. Total weekly AC and MVPA will be calculated: [5× mean daily weekday AC/MVPA time + 2× mean daily weekend AC/MVPA time], to standardise across all patients no matter what their wear-time, as long as it meets the minimum requirements.
- Physical performance assessment. This will take the form of 2 assessments (30-second chair to stand (CTS) test and stair climb test (SCT)) recommended by Osteoarthritis Research Society International (OARSI)[16] pre-operatively and at each post-operative time point.
- The UCLA Activity Score[17]. This is a patient reported outcome measure (PROM) and takes the form of a self-administered questionnaire. The patient will score a mark between 1 and 10 pre-operatively (historical activity level), pre-operatively (current activity level) and at each post-operative time point (current activity level).
- Hip Outcome Score (HOS)[18]. This is a PROM and takes the form of a self-administered questionnaire focussing on activities of daily living and sports/exercise. The patient will score a mark between 0 and 100 pre-operatively and at each post-operative time point.
- Noise generation. This will be assessed via a self-administered questionnaire that has been developed by the sponsor based on the literature[19]. There is no overall score for the questionnaire, but it does include a single visual-analogue scale (VAS) question that will give a mark between 0 and 10 for each patient at each post-operative time point.

All outcomes will be measured at 6, 12 and 24 months post-operatively. In addition, all (serious) adverse events and (serious) adverse device effects will be recorded as and when they occur.

6.2 Analysis methods

Analysis methods for primary and secondary outcomes and subgroup analyses are summarised in Table 1. Outcomes shall be reported for each treatment group. No adjustments will be made for baseline variables. Categorical data will be summarised by numbers and percentages and 95% confidence intervals.

Continuous data will be summarised by mean, standard deviate and range if data are normal and by median, interquartile range and range of data are non-normal.

6.3 Additional analyses

Kaplan-Meier (KM) survival analysis shall be carried out on each treatment group, where survival is 1-revision (removal of the device). KM survival (with 95% confidence intervals) shall be compared between the groups but no formal statistical analysis will take place.

6.4 Missing data

Minimal losses to follow–up for the primary outcome is anticipated. Where there are missing data points for certain variables/outcomes at certain time points, this shall be indicated in results tables. However, if more than 5% of missing data is found for the primary outcome, a sensitivity analysis using multiple imputations and estimating–equation methods will be carried out. Multiple imputation will consider imputation models based on prognostic baseline and post-baseline variables under a missing at random assumption.

6.5 Harms

Safety management is described fully in the study protocol[9]. The proportion of patients with AEs, SAEs, ADEs and SADEs will be compared descriptively across treatments and differences assessed for clinical significance but no formal statistical testing will take place.

6.6 Statistical software

The analysis will be carried out using the up to date version of SPSS.

6.7 Data management plan

There is no data management plan (DMP). All trial data will be entered and stored on the EDC (Castor). No data cleaning will be required.

6.8 Trial master file

The CRO will hold the trial master file, including details of the randomisation process and protocol deviations.

Table 1. Summary of outcomes and analysis methods

Variable/outcome	Hypothesis	Outcome measure	Method of analysis
Primary: CCS	H1 Implant is non-inferior to THR	% patients achieving CCS	Blackwelder's non-inferiority test
Secondary: physical activity	H1 Implant is significantly better than THR	Average daily minutes of activity and mean bouted daily MVPA, minutes	Student's t-test (normal, equal variance) or Welch's t-test (normal, unequal variance) or Wilcoxon Rank-Sum (non-normal)
Secondary: physical performance assessment	H1 Implant is significantly better than THR	Number of chair stands in 30 seconds Time to perform stair climb test	
Secondary: HOS	H1 Implant is significantly better than THR	HOS Questionnaire	
Secondary: UCLA Activity Score	H1 Implant is significantly better than THR	UCLA Activity Score	
Secondary: safety	N/A	Adverse events and adverse device effects	N/A
Secondary: noise	N/A	Noise questionnaire	N/A
Adjustment for stratification	Odds ratios are equal across sites	% patients achieving CCS	Cochran-Mantel-Haenszel test followed by Woolf's heterogeneity test
Subgroup analysis: Low v High pre-op activity/physical assessment score/UCLA	N/A	Average daily CPM Mean bouted daily MVPA, minutes	Student's t-test or Welch's t-test or Wilcoxon Rank-Sum
Subgroup analysis: Low vs high "historical UCLA" score			
Subgroup analysis: male v female	N/A	% patients achieving CCS	Fisher's exact test
Subgroup analysis: small H1 Implants v all THRs			
Subgroup analysis: large H1 Implants v all THRs			
Subgroup analysis: H1 Implants v device type in the THR group			
Subgroup analysis: H1 Implants v bearing type in the THR group			

7 References

1. CPMP/ICH/363/96. ICH Topic E 9 Statistical Principles for Clinical Trials Step 5 NOTE FOR GUIDANCE ON STATISTICAL PRINCIPLES FOR CLINICAL TRIALS 1998.
2. CPMP/EWP/2863/99. Points to consider on adjustment for baseline covariates. 2003.
3. P050016. Summary of Safety and Effectiveness Data: Corin Hip Resurfacing System. 2007.
4. Data on File, Embody Orthopaedic Limited.
5. Bahl JS, Millar SC, Fraysse F, et al. Changes in 24-Hour Physical Activity Patterns and Walking Gait Biomechanics After Primary Total Hip Arthroplasty: A 2-Year Follow-up Study. *J Bone Joint Surg Am* 2021; 103: 1166-1174. 2021/05/28. DOI: 10.2106/jbjs.20.01679.
6. Ekelund U, Tarp J, Steene-Johannessen J, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. *bmj* 2019; 366.
7. Saint-Maurice PF, Troiano RP, Matthews CE, et al. Moderate-to-vigorous physical activity and all-cause mortality: do bouts matter? *Journal of the American Heart Association* 2018; 7: e007678.
8. O'Brien PC and Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979: 549-556.
9. Study Protocol: The Active Trial. 2023.
10. Series Editor. *Multiple Endpoints in Clinical Trials Guidance for Industry.* Report no 2022
11. CPMP/EWP/908/99. Points to consider on multiplicity issues in clinical trials. 2002.
12. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *International journal of surgery* 2012; 10: 28-55.
13. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969; 51: 737-755. 1969/06/01.
14. Chahal J, Van Thiel GS, Mather III RC, et al. The patient acceptable symptomatic state for the modified Harris Hip Score and Hip Outcome Score among patients undergoing surgical treatment for femoroacetabular impingement. *Am J Sports Med* 2015; 43: 1844-1849.
15. Menai M, Van Hees VT, Elbaz A, et al. Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: results from the Whitehall II study. *Scientific reports* 2017; 7: 1-9.
16. Dobson F, Hinman RS, Roos EM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013; 21: 1042-1052.
17. Zahiri CA, Schmalzried TP, Szuszczewicz ES, et al. Assessing activity in joint replacement patients. *J Arthroplasty* 1998; 13: 890-895.
18. Martin RL, Kelly BT and Philippon MJ. Evidence of validity for the hip outcome score. *Arthroscopy* 2006; 22: 1304-1311. 2006/12/13. DOI: 10.1016/j.arthro.2006.07.027.
19. Noise Survey: The Active Trial. 2023.