Implementation of Comprehensive Geriatric Assessment based perioperative medicine services to improve clinical outcomes for older patients undergoing elective and emergency surgery with cost effectiveness.

Perioperative medicine for Older People undergoing Surgery Scale Up ***(POPS-SUp)***

Statistical Analysis Plan

Version No	V1.0
Date Finalised	23 July 2023
Author(s)	Professor John Norrie, Queen's University Belfast
CI Name	Professor Jugdeep Dhesi, King's College London
CI Email address	Dhesi Jugdeep <Jugdeep.Dhesi@gstt.nhs.uk>

Signatures	
Trial Statistician: Prof. John Norrie 	**Date: 2024-07-23**
Chief Investigator: Professor Jugdeep Dhesi	**Date:**
Co-Chief Investigator:	**Date:**

Document Control		
Version No	**Date**	**Summary of Revisions**
1.0	2024-07-23	Initial Creation

Contents

1. Introduction

This is a hybrid implementation-effectiveness interrupted time series study using mixed-methods to examine the use of a coproduced implementation strategy, to support implementation of perioperative medicine (proactive care) for older people undergoing surgery (POPS) services and evaluate clinical and cost effectiveness across the NHS. Two sequential cohorts of nine hospitals will be supported over consecutive twelve-month periods, to implement and evaluate POPS services using a piloted, feasible network model (NHS Elect POPS programme).

Eighteen NHS hospitals providing emergency and/or elective perioperative services for older people located across England, Scotland, Wales, and Northern Ireland, with representation of rural and urban NHS services, serving diverse populations in terms of socioeconomic circumstances, race, and ethnicity.

POPS-SUp is a hybrid implementation-effectiveness study, which will examine two inter-linked interventions:

- the trimodal implementation strategy (designed to support implementation of a POPS service)

- POPS services delivering perioperative Comprehensive Geriatric Assessment (CGA)-based care (and their impact on clinical and cost effectiveness)

The POPS-SUp study will have co-primary outcomes across the implementation and effectiveness stages.

The implementation stage will be assessed through a process evaluation, which will be covered in a separate Process Evaluation Analysis Plan (PEAP).

The effectiveness intervention (perioperative CGA-based care delivered through a POPS service) involves a holistic assessment across medical, functional, social, and psychological domains, using objective measures to inform multidisciplinary optimisation to improve outcomes in older people living with frailty and/or multimorbidity.

Perioperative CGA delivered through a multidisciplinary, geriatrician led service has shown clinical and cost effectiveness in the research setting for patients being considered/undergoing elective and emergency surgery. This proposal will study the impact of the CGA-based POPS intervention on clinical and cost effectiveness, when delivered at scale in the NHS setting using our trimodal implementation strategy.

2. Scope

The objective of this SAP is to describe the statistical analyses contributing to the final report and publication(s) of the clinical effectiveness stage of the POPS-SUp study. The quantitative aspects of POPS-SUp (the hybrid implementation-effectiveness interrupted time series, quasi-experimental design) will be analysed according to a comprehensive Statistical Analysis Plan (SAP), authored by the study statistician, and agreed by the independent Project Oversight Committee.

This SAP will specify all the statistical analyses, all aspects of model validation, how missing data will be approached, what sensitivity analyses will be included, and details of statistical progress reporting to the funder and the study oversight committees, and periodic safety reporting.

This Statistical Analysis Plan deals exclusively with the effectiveness stage – with length of hospital stay as the primary clinical effectiveness outcome.
The cost-effectiveness analyses will likewise be covered in a separate Health Economics Analysis Plan (HEAP). In addition, there will be a complementary Plan for the qualitative work/process evaluation (PEAP).

This SAP will full integrate with the comprehensive Data Management Plan (DMP), which together will cover all aspects of the data collection and quantitative analyses. The DMP will define all the required databases, all the electronic Case Report Forms, all the modes of data collection, the training required by task-based authorised data operators, the quality assurance processes, all security and fidelity processes, database validation steps and so on.

This document has been written based on information contained in the POPS-SUp study protocol [V.0.1 5th Dec 2023].

3. Statistical Issues section from the grant proposal

3.1. ***Research Subjects' (Page 9):*** *POPS-SUp will be recruiting at hospital site as opposed to at an individual patient level. Patient metrics will be collected from routinely available data and therefore individual consent will not be required, except in a small subset of patients in whom informed consent will allow qualitative evaluation and quality of life data for health economic evaluation.*

3.2. ***Clinical data collection (page 11-12):*** *will be captured from routinely recorded hospital data and through HES linkage. Recruited sites will collect outcome measures at baseline, during implementation and following implementation, to address the study aims through providing control data for internal comparison at site level:*

- *Baseline data collection (3 months): Implementation outcome measures, clinical outcome measures and metrics for health economic appraisal will be collected for three months preceding the start of the POPS service implementation.*

- *Implementation phase data collection (6 months). The same implementation, clinical and health economic outcome measures, as at baseline, will be collected throughout the six-month implementation phase.*

- *Post-implementation data collection (3 months). The same implementation, clinical and health economic outcome measures, as at baseline and during implementation, will be collected in the three-month period following implementation.*

- *National linkage of data. Data examining days alive and out of hospital at 90 days and mortality (90 day and 12-month mortality) will be obtained through HES linkage.*

3.3. ***Sample size (Page 9):*** *… an unweighted one sample t-test with each of 18 sites contributing a single after-before data point would have 90% power at 5% level of significance to detect an effect size of 0.8 i.e., a difference in mean length of stay after over before implementation of 0.8 standard deviations. If the standard deviation at the site level was five days, this would equate to being able to detect a mean difference of four days, which would be sufficient to impact clinical practice.*

The true power of the study will be greater (for example, allowing us to detect a smaller mean difference, of two-three days) and will be accurately assessed using simulation informed by early data from the study, and using the statistical model to be employed, which will depend on individual level data within each site, potentially at multiple time points (eg each week for around 12 times in each of the three month before and after periods), and fully account for dependency across participants within sites, and adjusting for known prognostic factors.

At present, we are expecting each site to contribute on average 15 participants per month, with NELA data showing a median length of stay of ten days in those without return to theatre.

3.4. Primary outcome (clinical effectiveness) (Page 14)

The primary outcome for clinical effectiveness will be length of hospital stay.

3.5. ***Secondary clinical effectiveness outcomes I (Page 14)*** *(clinician reported and patient reported measures)*

- *30-day readmission*
- *Comprehensive Complication Index*
- *Return to preoperative place of residence.*
- *Days alive and out of hospital 90 days,*
- *90-day mortality, 12-month mortality (HES linkage)*
- *Operative or non-operative management*

3.6. ***Secondary clinical effectiveness outcomes II (Page 14)*** *(on a selected subset with consent; target n=84)*

- *HRQoL*
- *Clinician defined 'medically fit and mentally fit for discharge.'*
- *Shared decision making (SDMQ9 from CQIN)*
- *Decisional regret Scale*

4. Statistical Considerations for the Analyses

4.1. **General:** The analysis of the interrupted time series will adopt the recommendations of Cruz 2017 (in particular providing flexibility to model evolving variability and correlation between the before and after periods) and follow the useful guidance of Bernal 2018.

4.1.1. **Rationale for ITS design:** The randomised controlled trial (RCT) is seen as the gold standard design, allowing causal interpretation of the estimated intervention effect. The RCT design relies on being able to randomise individuals or groups of

individuals (clusters) to intervention or control. Given the current status of CGA deployment in the NHS, such randomisation would not be feasible across clinicians and participants.

We did consider a stepped wedge design, often seen as useful when evaluating the performance of an intervention which is going to be implemented. However, there are substantial logistical challenges that made such a stepped wedge design not likely workable.

Instead, we chose the Interrupted Time Series (ITS) design that would allow evaluation of CGA by comparing the level and trend of carefully specified outcome after the intervention compared with that before the intervention. We did also consider a Differences-in-Differences design but felt the ITS design is particularly suited to interventions introduced at a population level over a clearly defined time period that target population level outcomes (Bernal 2017).

In addition, the ITS design here uses routine recorded data (primary outcome length of hospital stay, with data aggregated at the site level) on a largely unselected cohort (there is no consent process, and the intervention will be delivered by existing staff doing their usual jobs) and hence external generalisability should be strong (Barnighausen 2017a and 2017b).

See also Bhaskaran 2013, Bernet 2013, Zhang 2024, Kontopantelis 2015, Hudson 2019 for additional details on various design considerations for the ITS. See Morales 2023 for an instructive example of the reporting of an ITS designed study.

4.1.2. **Analysis population**. The primary analysis will include all participants recruited in the study where possible (akin to an 'intention to treat' analysis, consistent with a treatment policy estimand (ICH E9(R1) – see EMA 2020 and Kanan 2024).

4.1.3. **Trial Periods:** There will be 2 periods, each of 12 months, each recruiting from 9 sites (no site in the first period will contribute to the second period).

4.1.4. **Time Structure:** The intended structure of each period will be 3:6:3 months of Before: Implementation: After, and the intention of the analysis will be to estimate the effect of the intervention by comparing the After – Before.

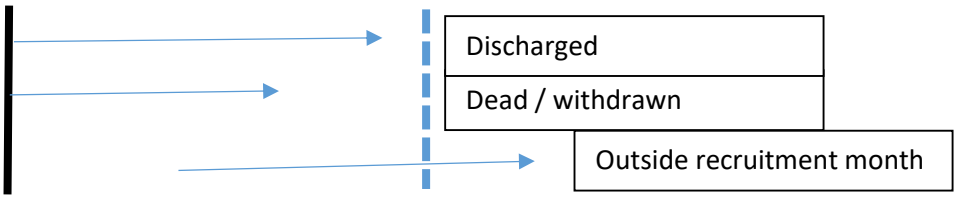
4.1.5. **Compliance/Fidelity:** This is a statistical analysis plan for the effectiveness part of a hybrid implementation-effectiveness study. Compliance with the intervention and/or fidelity of the implementation of the intervention is being measured and assessed separately in the POPS-sUP process evaluation. There will not be any statistical modelling (e.g., causal effect modelling) to adjust the treatment effects for any measure of compliance.

4.1.6. **Statistical Reporting**. In general terms, categorical data will be presented using counts and percentages, whilst continuous variables will be presented using the mean, median, standard deviation (SD), minimum, maximum, Q1, Q3, inter-quartile range (IQR) and number of patients with an observation (n).

4.1.7. **Graphical analysis.** We will produce boxplots, i.e., a graphical summary of the distribution including mean, median, first and third quartile, minimum and maximum values, for before and after the intervention, and by site, and period.

4.1.8. **Recruitment rate:** The expected recruitment rate is average 15 recruits per month per site. So, a site will be expected to contribute 180 patients over 12 months; and over a period (12 months) the 9 sites are expected to contribute 1620 patients; and over the 2 periods (the total for the study) will be expected to be 3240 patients.

4.1.9. **Follow up:** The length of stay of all those recruited will be followed up, including follow up occurring outside the recruitment month.



4.2. **Primary Outcome:** The primary outcome (for clinical effectiveness) is length of hospital stay.

4.2.1. **Measurement:** This will be measured on an individual basis as the total length of hospital stay, including any re-admission within 30 days of discharge.

4.2.2. **Shape of treatment response**. Following Cruz (2017) we will estimate the shape of the treatment response from the data, but also from Bernal (2018) we anticipate that the shape of the treatment response will incorporate (a) a gradual change in the slope (trend) and (b) a gradual change in the 'step' (intercept) and (c) it is a possible there may be a delay (lag) in either or both of these effects (slope and intercept).

4.2.3. **Statistical model.** A times series model with a continuous outcome (on either length of stay or $\log_{10}$(length of stay) to address skewness, with estimates and 95% confidence intervals back-transformed to the original untransformed scale, days) – see Turner 2020 and 2021

4.2.3.1. **Model terms:** We will estimate the effect of the intervention on the co-primary outcome of (untransformed or transformed) length of hospital stay, with terms for treatment (after – before periods, within site) accounting for any deaths, and adjusting for local site effects (including staggered times of intervention) and any temporal trends (potentially non-linear), and adjust for either site or individual level covariates strongly related to the outcome.

4.2.3.2. **Periods:** We will adjust for the two cohorts (first cohort n=9 sites, first period of 12 months; and second cohort n=9 sites, second period of 12 months; a total of 18 sites over a two 12-month periods).

4.2.3.3. **Model assumptions**: All the assumptions regarding the statistical model will be assessed, including (a) autocorrelation structure, and (b) non-stationarity and (c) seasonality, if appropriate.

4.2.4. **Units of measurement and analysis**: The length of stay will be measured at the individual level and analysed at the site level, in a time unit of one week (to be confirmed at the sample size re-estimation step at the end of the first period) i.e., all those recruited in a site in a specific week, meaning that the expectation is that a site will contribute 13 before and 13 after data points, each aggregating on average 3-4 patients.

4.3. **Secondary Outcomes:** The secondary outcomes (e.g., 30-day re-admission, Comprehensive Complications Index, return to pre-op residence, days alive out of hospital at 90 days, mortality at 90 days and 1 year, and the quality of life measures) will be analysed in a similar way to the primary outcome with a statistical model appropriate for the specific secondary outcome (eg binary or ordinal logistic regression, time-to-event (Cox) regression, linear regression).

4.4. **Subgroup Analyses:** Pre-defined Subgroup analysis will be restricted to the primary clinical effectiveness outcome alone. Any further subgroup analysis (e.g., if suggested later by new data external to the study) will be labelled exploratory. Pre-specified subgroup analyses will be unlikely to be adequately powered.

4.5. **Missing data:** We do not anticipate much missing data on the primary outcome (length of stay). Nonetheless, we will check the robustness of the findings to any patterns of missing data using sensitivity analyses (including multiple imputation under an assumption of missing at random, or possibly pattern mixture type models for informative missingness) (Little 2012)

4.5.1. A multiple imputation approach will be used assuming the data are missing at random. In addition, and probably more consistent with the likely missing data generating mechanisms, sensitivity type analyses assuming the data are missing not at random (i.e., informatively missing) will be explored e.g., using pattern mixture models, or tipping-point type approaches.

4.5.2. These sensitivity analyses would attempt to identify different types of missing data by an underlying reason or reasons, and then imputing values that capture plausible measurements for those missing data.

4.5.3. The (gamma) γ-adjustment approach (van Buuren 2018) will be followed, and also the recommendations on sensitivity analyses (Molenburghs 2007).

4.6. **Safety:** The safety data (e.g., medical and surgical complications, factors around delayed discharge, delirium, acute coronary syndrome, cardiac failure, arrythmias, pneumonia, wound infection, urinary tract infections, faecal incontinence, falls, acute post-op complications (cardiac, pulmonary, infections, bowel/bladder, vascular), level 2/3 care post-surgery; and other adverse events) will be presented descriptively.

4.7. **Interim and Final Analysis: This a**nalysis plan describes the end of trial statistical analyses to be performed for POPS-sUP.

There will be no formal interim analyses.

There will be a sample size check / re-estimation step at or around the end of the first cohort of 9 sites followed for the first 12 months, which will validate the assumptions behind the power calculation (specifically the assumed common standard deviation) and in particular upgrade the estimation of the actual power of the study using simulation, using the appropriate statistical model for the primary outcome of length of stay, instead of the simple approximation using a 1-sample t-test as above. The timing of the sample size re-estimation coincides with the end of the first period of 12 months. See Zhang 2011 and Hawley 2019 for further details on sample size estimation in ITS designs.

4.8. **Control series.** Although including a control condition can be useful in the estimation and interpretation of the modelled primary outcome, there was no obvious candidate outcome here and for simplicity this option was not pursued (see Bernal 2018 for further details).

5. **Tables and Figures:** *for the Statistical Report – to be specified.*

6. **Statistical software:** All analysis and data manipulation will be carried out using SAS or R for Windows or Stata unless otherwise stated.

7. **Data sharing:** A file, or set of files, containing the final data will be prepared, along with a data dictionary. These will be made available to the Chief Investigator at the end of the analysis phase.

References

Barnighausen T, Oldenburg C et al. Quasi-experimental study designs series – paper 7: assessing the assumptions. J Clin Epidemiol 2017; 89; 53-66

Barnighausen T, Tugwell P et al. Quasi-experimental study design series – paper 4 uses and values. J Clin Epi 2017; 89; 21-29

Bernal JL, Cummins S et al. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017; 46(1); 348-355

Bernal J, Soumerai S et al. A methodological framework for model selection in interrupted time series studies. J Clin Epidemiol 2018; 103; 82-91.

Bernal J, Cummins S et al. The use of controls in interrupted time series studies of public health interventions. Int J Epidemiol 2018; 2082-2093.

Bernet AC et al. Effectiveness-implementation hybrid designs: implications for quality improvement science. Implementation Sci 2013; 8; S2

Bhaskaran K et al. Time series regression studies in environmental epidemiology. Int J Epidemiol 2013; 42; 1187-1195

Cruz M, Bender M et al. A robust interrupted time series model for analysing complex healthcare intervention data. Stat Med 2017; 36(29); 4460-76

European Medicines Agency ICH E9[R1] addendum on estimands and sensitivity analyses in clinical trials to the guideline on statistical principles for clinical trials 2020; www.ema.europa.eu\en\documents\scientific-guidelines (accessed 2024-07-25)

Hawley S et al. Sample size and power considerations for OLS interrupted time series analysis: a simulation study. Clin Epidemiol 2019; 11; 197-205

Hudson J, Fielding S, Ramsay CR. Methodology and reporting characteristics of studies using interrupted time series designs in healthcare. BMC Med Res Methodol 2019; 19(1); 137.

Kontopantelis E et al. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analyses. BMJ 2015; 350; h2750.

Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT et al. The prevention and treatment of missing data in clinical trials. N Engl J Med; 367:1355-60.

Kanan BC et al. The estimands framework: a primer on the ICH E9(R1) addendum. BMJ 2024; 384; e076316

Molenburghs G, Kenward M. Missing Data in Clinical Studies, Wiley, 2007; Section 19ff on sensitivity analyses

Morales DR et al. Estimated impact from the withdrawal of primary care financial incentives on selected indicators of quality of care in Scotland: controlled interrupted time series. BMJ 2023; 380; e072098.

Partridge JSL, Harari D et al. Randomised clinical trial of comprehensive geriatric assessment and optimization in vascular surgery. Br J Surg 2017l 104; 679-87.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

SAS® Institute Inc. SAS for Windows. SAS Institute Inc.: Cary, NC, U.S.A

StataCorp 2023. Stata Statistical software Release 18. College Station, TX. StataCorp LLC.

Turner S, Karahalios A et al. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. J Clin Epidemiol 2020; 122; 1-11

Turner S, Karahalios A et al. Comparison of six statistical methods for interrupted time series studies: empirical evaluation of 190 published series. BMC Medical Research Methodology; 2021; 21(1); 134.

Van Burren. Flexible Imputation of Missing Data, Chapman and Hall, 2018, section 3.8ff

Waring J, Martin GP. Implementing a perioperative care of older people undergoing surgery (POPS) service: findings from a multisite qualitative implementation study. Age and Ageing 2023; 52; 1-8

Zhang F, Wagner AK. Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. J Clin Epidemiol 2011; 64; 1252-1261

Zhang Y, Ren Y et al. Design and statistical analysis reporting among interrupted time series studies in drug utilisation research: a cross-sectional survey. BMC Medical Research Methodology 2024; 24; 62.