UiO : Faculty of Medicine
University of Oslo

STATISTICAL ANALYSIS PLAN (SAP)

Study title	Confident management of shoulder pain in general practice – implementing an evidenced based guideline for shoulder pain, a hybrid design cluster randomized study
Short title	The confident shoulder study
Study protocol version	v1.0 SAFE 120919
Clinical trial number (NCT)	NCT04806191
Funding	Western Norway Regional Health Authority and Norwegian Research Fund for General Practice
SAP version	Final

SAP Signatures

We, the undersigned, certify that we have read this SAP and give our approval for its use in this clustered randomized study on shoulder pain management

Principal Investigator

[1]Stein Jarle Pedersen

Name:	Signature:	Approval date:

List of collaborators and contributors

[2]Ole Marius Ekeberg

Name:	Signature:	Approval date:

[3]Niels Gunnar Juel

Name:	Signature:	Approval date:

[4]Jens Ivar Brox

Name:	Signature:	Approval date:

[5]Pål Joranger

Name:	Signature:	Approval date:

[6]Eva Bieringer

Name:	Signature:	Approval date:

[7]Kaia Engebretsen

Name:	Signature:	Approval date:

[8]Silje Reme

Name:	Signature:	Approval date:

[9]Bård Natvig

Name:	Signature:	Approval date:

Trial statistician

[1]Ibrahimu Mdala

Name:	Signature:	Approval date:

Affiliations

1. Department of General Practice, University of Oslo, Norway
2. Department of Research and Innovation, Helse Fonna, Norway
3. Department of General Practice, University of Oslo, Norway
4. Department of Clinical Medicine, University of Oslo, Norway
5. Department of Nursing and Health Promotion, OsloMet - metropolitan university
6. Department of Research and Innovation, Fonna Hospital Trust, Haugesund, Norway
7. Department of Physical Medicine and Rehabilitation, Oslo University Hospital, Ullevål
8. Department of Psychology, Department of Pain Management, Oslo University Hospital
9. Department of General Practice, University of Oslo, Norway

List of abbreviation

EQ5D-5L	EuroQual 5 level generic quality of life measure
FABQ-A	Fear avoidance belief questionnaire activity
ICC	Intra-cluster correlation
ICD-PC	
ITT	Intention-to-treat
MI	Multiple imputation
NB	Negative binomial
Orebro-10	
PP	Per protocol
SAP	Statistical analysis plan
SPADI	Shoulder Pain and Disability Index

Table of Contents

1.0 Introduction, study design and study objectives

1.1 Introduction

The confident shoulder study is a stepped wedge cluster randomized controlled trial that will assess the effectiveness of implementing an evidence-based guideline for shoulder pain in general practice in Norway. Detailed background information of this study are currently available in the study's protocol.

This statistical analysis plan (SAP) provides detailed description of the statistical analyses methods and techniques as outlined in the protocol. The description will also include the analysis population sample such as whether the research data will be analyzed on the basis of intention-to-treat (ITT) or per-protocol (PP). In addition, ways of dealing with missing data, which may cause data attrition and hence may affect the statistical power, are described. In particular, the focus will be given to multiple imputation of multilevel data.

1.2 Study design

A stepped wedge cluster randomized trial (SWCRT) with a focus on assessing the efficacy and feasibility of an evidence-based clinical guideline on shoulder pain management will be taken. The guideline, confident care will be delivered in sequential steps to each GP surgery (cluster) at a pre-determined time point and will be delivered to all the GP surgeries by the end of the study period. This two-year follow-up study will recruit 10 GP surgeries with each GP surgery initially acting as a control (usual care) until it crosses over to receive the confident care.

1.3 Primary and secondary objectives

The primary objective of the study is to assess whether implementing an evidence-based guideline for shoulder pain in general practice in Norway using an electronic decision tool will improve the patients' shoulder pain and function at 12 weeks compared to usual care.

The secondary objectives are to:

1. evaluate the feasibility of using an electronic decision support tool to implement the guideline

2. assess the health resource use and conduct a health economic evaluation of the treatment model

1.4 Hypotheses

For the main research question, our hypotheses are

- Null hypothesis: Confident care of shoulder pain management does not give better clinical outcomes than usual care
- Alternative hypothesis: Confident care in shoulder pain management offers better clinical outcomes than the usual care.

The hypotheses of the secondary research question are

- Null hypothesis: Using an electronic decision support tool in the implementation of the guideline is not feasible
- Alternative hypothesis: Using an electronic decision support tool in the implementation of the guideline is feasible.

2.0 Sample size

The primary outcome measure for the statistical power calculation is the yearly incidence rate of shoulder pain in general practice. Based on the ICD-PC codes for shoulder pain, approximately 67 patients on average are given a primary shoulder diagnosis in GP surgeries in Østfold County. Most of these GP surgeries have an average of three to four GPs. The incidence of shoulder pain in general practice has been reported to be 2-4%. Therefore, in estimating the sample size for this study, we used a conservative assumption of 3% as the incidence of shoulder pain diagnosis in this population. Setting the desired statistical power of the study at 80%, alpha at 0.05 and an intra-cluster correlation (ICC) of 0.002, we estimated that we needed to recruit 250 patients to detect a clinically important difference of 8 points between the groups, taking into account a possible dropout rate of 10%.

3.0 Study populations

When analyzing the research data, we shall consider two populations; intention-to-treat (ITT) and per protocol (PP) to account for the fact that there may be delays between the expected start date and the actual start date. However, we expect all general practices to be randomized to receive the confident care guideline and implement it.

3.1 Intention-to-treat (ITT) population

For the intention-to-treat analysis, a binary indicator variable, which denotes whether an observation occurred before or after randomization shall be constructed and used in the regression models. The binary factor shall represent the study months in each general practice pre-randomization and post-randomization. This means that in the ITT population, the general practices shall be analyzed according to their randomized crossover time regardless of whether this was achieved.

3.2 Per Protocol (PP) population

In the per-protocol (PP) analysis, the effect of the confident care guideline shall be assessed by considering three categories in the regression analysis: 1. pre-randomization 2. post-randomization but pre-implementation and 3. post-implementation. This would allow adjustment due to delays in the implementation in the implementing general practices.

4.0 Outcomes

The research data for all patients diagnosed with shoulder pain will be collected from 10 general practices in Østfold County and in the health region of Helse Fonna. The outcomes will be obtained during the entire period of study between January 2020 and 2022 at all pre-specified time points. Therefore, multiple observations per patient are possible.

4.1 Primary outcome

The primary outcome measure is the Shoulder Pain and Disability Index (SPADI), which is a self-reported questionnaire with 13 items regarding shoulder pain and disability. These are patient level observations obtained at time points 0, 1, 2 and 3 denoting baseline, 6 weeks, 12 weeks and 52 weeks respectively.

4.2 Secondary outcome

Several secondary outcome measures, also at patient level, shall be obtained. These will be measured either on four different measurement scales: Likert, ordinal, categorical (binary measures) and as counts.

The outcomes that will be measured on a Likert scale are

- EQ5D-5L (Health related quality of life) with scores ranging from 0 to 1.00
- Orebro-10 with scores ranging from 0 to 100
- Fear avoidance belief questionnaire activity (FABQ-A) with scores ranging from 0 to 24
- Pain self-efficacy Questionnaire with scores ranging from 0 to 60
- Brief Illness Perception Questionnaire (B-IPQ) with scores ranging from 0 to 90
- Patient experience Questionnaire with scores ranging from 17 to 88

The following secondary outcomes will be measured on an ordinal scale

- Participant global improvement
- Participant satisfaction with treatment

The binary outcomes to be categorized into (0/1) for (No/ Yes) will include the following secondary outcomes

- Work
- Work disability (sick leave)
- Medication usage such as over the counter pain medications, corticosteroid injections

Two count variables that we will consider are

- The number of sick leave days
- Medication usage frequency

With the exception of participant global improvement and participant satisfaction with treatment, all secondary outcomes will have baseline measurements.

5.0 Potential adjusting variables

The following demographic and baseline characteristics of the patients will be collected at the time of screening: patient age, gender, marital status, years of education and occupational status. Information about patients' comorbidities and other shoulder related information will be self-reported at baseline. These will include areas of bodily pain, previous shoulder pain, duration of shoulder pain and pain intensity.

We will also collect GPs demographic information such as age, gender, years since graduation, specialty, type of practice, number of GPs in the practice, the number of weekly working hours, and number of people on the GPs' lists at recruitment.

6.0 Statistical methods

Depending on the scale of measurement of both primary and secondary outcomes, appropriate mixed effects regression models will be fitted to these data. For the main outcome variable, SPADI, a linear mixed effects regression model will be considered. All binary secondary outcomes described above will be analyzed using mixed effects binary logistic regression models whereas mixed effects ordered logistic regression models will be considered for ordinal outcomes. Count data relating to the number of sick leave days and medication usage frequency will be analyzed using mixed effects Poisson regression models if the data are not equi-dispersed. The assumption of equi-dispersion, which states that the distributional mean and the distributional variance are equal, is important in selecting the type of Poisson regression model to use. In the case where this assumption is violated, we shall consider extensions of the Poisson regression model such the negative binomial (NB) regression model. In all these mixed effects regression models, general practices and GPs within general practices will be used as random effects in the models.

To account for expected delays between expected start date and the actual start date, ITT and PP analyses shall be carried out. We shall construct a binary variable denoting the study months in each general practice before and after randomization and use it in the ITT analysis. On the other hand, we shall assess the effect of confident care in the management of shoulder pain using a three-level factor denoting pre-randomization, post-randomization but pre-implementation and post-implementation.

All results from the mixed effects binary logistic regression models shall be presented as estimated odds ratios with 95% CIs. For both linear and binary/ ordered mixed effects regression models, estimates of intra-cluster correlation (ICC), which describes the amount of variability in the outcome measures that can be attributed to differences between the general practices and the GPs within the general practices shall be obtained from the models. However, results from the count regression models shall be presented as incidence rate ratios (IRR) with their 95% CIs. For example, the IRR shall represent the **changes** in the number of sick day leave at a certain time point relative to the referent group. IRR estimates that are significantly above 1 shall represent significant increases in the number of sick leave days whereas IRR estimates that are significantly less than 1 shall represent significant decreases in the number of sick leave days.

We shall use descriptive statistics to summarize categorical data in the form of frequencies (counts) and percentages. Normally distributed continuous variables shall be summarized using the mean and standard deviation (SD) whereas non-normal continuous data shall be summarized using medians with first and third quartiles.

All analyses shall be performed using StataSE statistical package and the significance level shall be set at $\alpha = 0.05$.

6.1 Model building for the primary outcome

To explore the impact of the intervention on SPADI over time, we shall consider building the linear mixed effects regression models in four separate steps. We shall initially ignore the time effect in an unadjusted before and after the intervention model (model 1). Therefore, the SPADI for any given patient *i* seen by GP$_j$ in clinic$_k$ (general practice$_k$) is given as:

Model 1

$$\text{SPADI}_{ijk} = \beta_0 + \beta_1 I_{ijk} + \alpha_{ijk} + \varepsilon_{ijk}$$

The parameter estimates β_0 and $\beta_1$represent the fixed effects associated with the intercept and the effect of the intervention *I* (1 = exposed and 0 = unexposed (control) for patient *i* seen by GP *j* in clinic *k*. α_{ijk} is the random effect of patients nested by their GPs, who are in turn nested

in clinics and ε_{ijk} is the random error assumed to be homogeneous and uncorrelated with mean 0 and variance δ^2.

Model 2

To assess if any potential effect of the intervention on SPADI is also related to time, we shall introduce a time component *T* in model 2 and explore how the impact of the intervention develops over time.

$$\text{SPADI}_{ijk} = \beta_0 + \beta_1 I_{ijk} + \beta_2 T_t + \alpha_{ijk} + \varepsilon_{tijk}$$

Here, the parameter estimate β_2 represents the effect of time.

Model 3

Model adjustment of patient level characteristics presented in section 5 above shall be considered in model 3 as follows:

$$\text{SPADI}_{ijk} = \beta_0 + \beta_1 I_{ijk} + \beta_2 T_t + [\beta] X_{ijk} + \alpha_{ijk} + \varepsilon_{tijk},$$

where $[\beta]$ is a matrix of the patient level covariates. The covariates that we shall consider include the following: patient age, gender, marital status, years of education and occupational status. In addition, GP level covariates such as age, gender, years since graduation, specialty, type of practice, number of GPs in the practice, the number of weekly working hours, and number of people on the GPs' lists at recruitment shall be considered for model adjustment by introducing an additional term, $[\beta_1] X_{ijk}$ if necessary.

Model 4

We also believe that the intervention may either decrease SPADI score over time as the GPs gain experience and are confident with the intervention or SPADI scores may increase over time if, for example, the GPs become less enthusiastic about it or if the intervention is not effective. Therefore, we will consider a two-way interaction between time and exposure to the intervention in model 4.

$$\text{SPADI}_{ijk} = \beta_0 + \beta_1 I_{ijk} + \beta_2 T_t + [\beta] X_{ijk} + \beta_3 T_t I_{ijk} + \alpha_{ijk} + \varepsilon_{tijk},$$

The parameter estimate β_3 is the interaction between time and exposure to the intervention *I* (1 = exposure and 0 = control).

6.2 Model building of secondary outcomes

We shall also consider similarly structured models for secondary outcomes if the measurement scale is either continuous or categorical. In a binary logistic regression setting with patient, GP and clinic random effects and adjusted for patient level covariates and time, the binary responses Y_{tijk} shall be modelled by considering a model of the form:

$$Y_{tijk} \sim \text{Binomial}\ (1, p_{tijk})$$

$$\text{logit}\ (p_{tijk}) = \beta_0 + \beta_1 I_{ijk} + \beta_2 T_t + [\beta] X_{ijk} + \beta_3 T_t I_{ijk} + \alpha_{ijk} + \varepsilon_{tijk}$$

Y_{tijk} is 1 if patient *i* at time *t* seen by GP *j* in clinic *k* had for example, sick leave and 0 otherwise. The parameter estimates are defined as in model 4 above.

However, for count data such as the number of sick leave, a model of the form below shall be considered;

$$\log\ [E(Y|x_{ijk})] = \beta_0 + \beta_1 I_{ijk} + \beta_2 T_t + [\beta] X_{ijk} + \beta_3 T_t I_{ijk} + \alpha_{ijk}$$

where $E\left(Y\middle|X_{ijk}\right)$ is the expectation of *Y*, which is a vector of count response variables given the covariates X_{ijk} and the parameter estimates β_0, $\beta_1, \beta_2, [\beta]$ and β_3 are as described above.

7.0 Missing data

Missing data have the potential of decreasing the power of the study resulting in either under or over estimation of the effect of the confident care guideline of shoulder pain management. This means that missing data may potentially lead to biased estimates. For example, the bias in model estimates may be introduced if the missing data for the outcome measure, SPADI is related to the reason for the data missing. This may occur if, for some reasons more younger patients decide to withdraw from the study leaving older patients who are more likely to be at higher risk of shoulder pain.

Rubin first described three forms of missingness in the data based on the reasons for the missingness [ref]. The methods for handling data missingness in research depend on the missingness mechanism. The mechanisms are missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). MCAR describes a situation where the data missing are independent of the observed outcomes and the covariates. For example, if a patient

decides to attend a wedding instead of completing the questionnaire on health related quality of life. MNAR is missingness that is neither MCAR nor MAR. It occurs when data missingness is dependent on the unobserved data. For example, patients with chronic shoulder pain are less likely to report the number of sick leave days if they think that this may reveal the amount of support they are getting from NAV. However, a more practical assumption that we shall assume in this study, is MAR. This occurs when the missingness depends on the observed outcomes and possibly on the observed covariates. For example, data on SPADI may be missing if a patient is withdrawn from the study due to a suspected case of inflammatory joint disease.

To account for the fact that clustered data are dependent, we shall use multilevel multiple imputation using the **miceadds** package in R. The imputed data shall be saved and the data shall be analyzed using StataSE. However, we shall only consider multilevel MI if 10% or more of the data are missing.

8.0 Reporting of adverse events

We do not believe that there shall be adverse events associated with the implementation of the confident care nor adverse event related to the usual care. Therefore, no analysis and reporting of adverse events is planned.

9.0 References

1. Rubin DB. Inference and missing data. Biometrika. 1976;63(3):581–592. doi: 10.1093/biomet/63.3.581.
2. Robitzsch A, Grund S (2020). *miceadds: Some Additional Multiple Imputation Functions, Especially for 'mice'*. R package version 3.9-14, https://CRAN.R-project.org/package=miceadds.