Statistical Analysis Plan

Study Name	Evaluation of the National tool for observation of infection prevention measures in the healthcare (NOST) – a cluster randomized trial
Registration	ClinicalTrials.gov: NCT05721183
SAP Version	1.1
SAP Version Date	October 30, 2024
Study Statistician	Petter Elstrøm
Protocol Version (SAP associated with)	1.6 (June 16, 2023)
Study Principal Investigator	Petter Elstrøm
SAP Author(s)	Chris Rose, Petter Elstrøm

1 Table of Contents

2 Abbreviations and Definitions

Abbreviation	Definition
HAI	*Healthcare-associated infection*
ICC	*Intraclass correlation coefficient*
IPC	*Infection prevention and control*
NOIS	*Norwegian Surveillance System for Antibiotic Use and Healthcare-Associated Infections*
NOST	*National Tool for Observation of Infection Prevention Measures*
NPR	*Norwegian Patient Registry*
POSI	*Postoperative site infection*
RD	*Risk difference*
RR	*Risk ratio*
VESUV	*Norwegian Institute of Public Health's web-based outbreak notification system*

3 Introduction

3.1 Preface

Effective infection prevention and control (IPC) is essential to ensure high-quality healthcare services. Infections that occur during hospitalization and because of services provided may interfere with the outcome of needed medical treatments. To prevent healthcare-associated infections (HAIs), it is essential that all healthcare personnel are well trained in and follow standard IPC measures during patient care. Standard precautions include hand hygiene, use of personal protective equipment and more.

The Norwegian Institute of Public Health (NIPH) is introducing a new electronic tool and national template for direct observation of compliance with recommended IPC measures in healthcare. The solution is called the National Tool for Observation of Infection Prevention Measures (NOST). NOST is a quality improvement tool that includes a web-based solution for observing and recording the degree of compliance with recommendations for hand hygiene and other IPC measures.

The trial protocol is available at https://zenodo.org/records/7648821

3.2 Research Questions

Would risk of compliance with hand hygiene recommendations by employees who perform patient-related work in Norwegian hospital wards (*population*), measured at one-year postimplementation (*primary outcome*), be different if wards were to implement and adhere to NOST (*intervention*) compared to if NOST was not implemented and not adhered to (*control*)?

Would risk of outbreaks and infections among inpatients at Norwegian hospitals, measured over one-year postimplementation (*example secondary outcomes*), be different if wards were to implement and adhere to NOST (*intervention*) compared to if NOST was not implemented and not adhered to (*control*)?

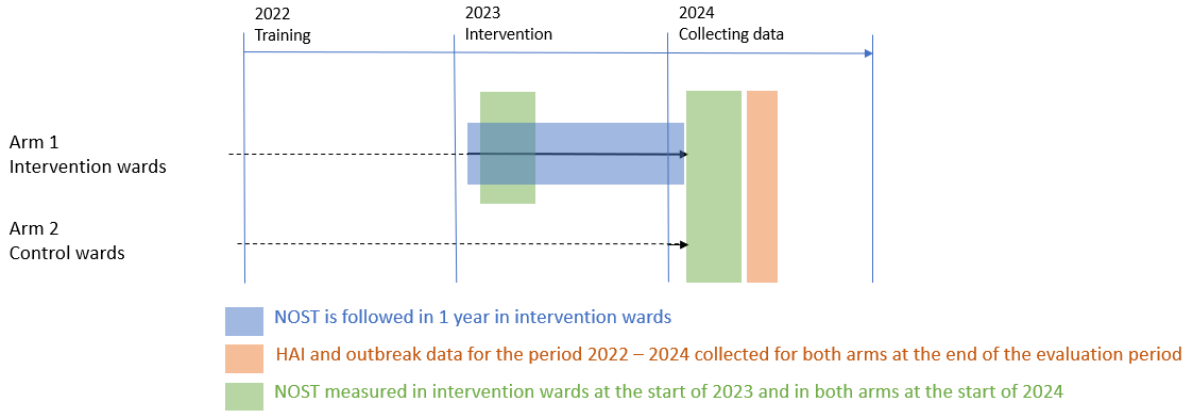
These questions are about the effect of implementing and adhering to NOST, not the effect of a policy that a ward should implement NOST. This has important implications for the estimands and estimators.

4 Study Methods

4.1 General Study Design and Plan

Populations	1. Employees in Norwegian hospitals who perform patient-related work. 2. Inpatients in Norwegian hospitals. 3. Norwegian hospital wards.
Intervention	Implementing NOST
Control	Not implementing NOST
Primary outcome	Compliance with hand hygiene recommendations by hospital employees who do patient-related work, measured at one year after randomization.
Design	Cluster-randomized parallel two-arm superiority trial.
Blinding	Trial participants and other personnel cannot be blinded to treatment allocation. The trial statistician will be blinded to treatment allocation.
Treatment allocation	Hospital wards will be randomized 1:1. Randomization will be stratified by hospital to ensure approximately equal allocation of wards within hospital.

The following diagram illustrates the sequence and duration of the study periods:



The following table shows the structure of data registered in NOST:

Place			**Observation of hand hygiene**			**Person**
Hospital	*Ward*	*Section number*	*Observation number*	*Indication*	*Compliance*	*Profession*
1	5	1	1	A	1	Nurse
			2	B	1	
			3	D	0	
1	4	2	4	B	1	Physician
2	2	3	5	A	0	Nurse
			6	B	1	
			7	C	0	
			8	C	0	
			9	E	1	
1	5	4	10	A	1	Nurse
			11	E	1	assistant

Description of the variables:

- Hospital: a unique ID-number for each hospital. The data will also include information about the actual health trust and health region
- Ward: a unique ID-number for each ward within the actual hospital. The wards will be allocated to intervention and control arms
- Section number: a unique running number for a section of observations where the observer has followed a person and registered all hand hygiene indications happening within a specific setting or task. A section will most often include several observations where the person has or should have performed hand hygiene.
- Observation number: a unique running number for an observation where hand hygiene was or should have been performed
- Indication: A code for why hand hygiene was advised, based on the guidelines by the World Health Organization
- Compliance: A code displaying whether the person performed hand hygiene or not (1 or 0) when the indication arose
- Profession: The health profession of the observed person. Other personal information is not registered. Only one person is observed within a section number, but the same persons may be observed in several sections

4.2 General Study Populations and Inclusion-Exclusion Criteria

There are three general study populations (see above and the study protocol). This SAP introduces a third population not specified in the protocol, hospital ward. This was introduced because one of the secondary outcomes (outbreaks) cannot be measured at the level of employee or inpatient but can be measured at the level of ward.

We will exclude hospitals and their associated randomized wards from all analyses if the hospital after randomization decides not to implement or to terminate the execution of NOST, and the decision is communicated to the Norwegian Institute of Public Health.

4.3 Treatment Allocation

Stratified randomization will be used to allocate wards to 1:1 intervention and control arms. Randomization will be stratified by hospital. Randomization will be performed using a computer-based system and the investigators will not be able to manipulate treatment allocation. The approach is outlined in the protocol.

4.4 Blinding/Masking

It is not possible to blind trial participants or observers to treatment allocation. While much of the outcome data will be obtained from registries (e.g., to which events are notified), the people who report data may not be blinded to treatment allocation. However, the trial statistician who will analyze the data will be blinded to treatment allocation.

When all data have been collected, a member of the research group will apply a blind to the treatment allocation variable, so that the statistician will not know which arms correspond to the intervention and control treatments. The analysis of the primary outcome will be performed blinded. Before unblinding, we will publish a short document on Zenodo describing how the result for the primary outcome was interpreted by the project team. The blind will then be removed and the secondary analyses will be performed unblinded.

5 Outcomes

5.1 Primary outcome

The primary outcome is compliance with hand hygiene recommendations by hospital employees who do patient-related work. Compliance will be assessed one year after randomization by trained and experienced observers. The observers will record the number of observations made and the number of these observations that meet the criteria for compliance. The number of observations and compliances will be measured at the level of hospital ward. Compliance will therefore be aggregated at the level of hospital ward and the primary outcome is a count variable.

5.2 Secondary outcomes

As for the primary outcome, all secondary outcomes will be aggregated at the level of hospital ward. Unlike the primary outcome, all secondary outcomes are measured over the one-year trial period rather than at the end of the trial.

Secondary Outcome	Type	Denominator	Data source
Infectious disease outbreaks			
1. Outbreak	Count	Wards	VESUV
2. Inpatient infection	Count	Inpatients in wards with outbreaks	
3. Employee infection	Count	Employees in wards with outbreaks	
Surveillance of HAI			
4. Inpatient infection	Count	Inpatients included in the surveillance	NOIS
5. Inpatient antibiotic treatment	Count	Inpatients included in the surveillance	
Surveillance of postoperative site infections			
6. Inpatient treatment for postoperative site infection	Count	Surgical inpatients included in the surveillance	NOIS-POSI
HAI diagnoses			
7. Inpatient HAI diagnosis	Count	Inpatients in the study period	NPR
Length of hospital stay			
8. Bed-days per inpatient per stay	Continuous		NPR

6 Sample Size

Assuming $\alpha = 0.05$, $\beta = 0.8$, a control risk of 0.4, an intervention risk of 0.6 (i.e., the treatment effect is a risk difference of at least 0.2), an ICC of 0.1, an average cluster size of 30 observations among hospital employees who do patient-related work, a loss of 2 d.f. for adjustment of cluster-level covariates, and stratification by hospital, it was estimated that at least 52 hospital wards should be recruited (26 in each arm).

7 General Analysis Considerations

7.1 Timing of Analyses

Analyses will be performed after the database is locked. No interim or follow-up analyses are planned.

7.2 Analysis Sets

Although this SAP defines three populations, there is only a single analysis set that will contain all wards randomized and not excluded (see exclusion criteria, above). In practical terms, the data set will have one ward per row and there will be 9 outcome variables (columns), plus covariates. Outcomes will be analyzed in the arms to which the wards were randomized.

7.3 Covariates and Subgroups

Because randomization is stratified by hospital, we originally planned to adjust for hospital as a fixed effect in all analyses. However, the planned adjustment would require estimating one parameter for each hospital (minus the reference hospital), in addition to treatment effect. The large number of parameters that would therefore need to be estimated could lead to imprecision on the effect estimates and possibly also estimation problems. Further, the trial is unlikely to include all hospitals in Norway and hence the sample will not exhaust all possible levels of the hospital variable. For these reasons, and the chosen regression model (see section 9), we will account for the likely clustering of outcomes within hospital using cluster-robust standard errors, with hospital as the cluster variable.

We will obtain data on the following variables to support four subgroup analyses:

1. Hospital department (e.g., psychiatry, orthopedics, neurology, …)
2. Employee profession (e.g., physician, nurse, …)
3. Indications of hand hygiene (e.g., before patient contact, after exposure to body fluid, ...)

These subgroups are based on the hypothesis that baseline hand hygiene compliance may differ by type of ward or type of profession with different levels of education or work tasks, and compliance may vary depending on the situations in which hand hygiene is recommended. It may be the case that treatment effect is different in these subgroups. (For example, if some ward types already have exceptionally high baseline compliance with hand hygiene recommendations, then only a small beneficial treatment effect is possible for this subgroup. Conversely, if a type of ward or profession has exceptionally low baseline compliance for all or some of the indications then a large beneficial treatment effect may be achievable in this subgroup.

7.4 Multiple Centers

We will not combine outcomes by ward or hospital (i.e., we will not define “pseudo-centers”). We will not explore treatment-by-hospital interactions. Between-hospital differences in risk of compliance, for example, will be accounted for using cluster-robust standard errors with hospital as the cluster variable.

7.5 Missing Data

We will examine the data for spurious values and seek to verify or obtain correct values if data are miscoded (e.g., negative counts, nonintegral values entered for counts, or counts that appear to be unrealistically large). We will report any data processing decisions that need to be made at the time of analysis.

It is very unlikely that data will be missing for the treatment, hospital, or ward variables, which are needed in all analyses.

Data may be missing for the primary outcome variable, which is measured by observers. It is unlikely

that data will be missing for the secondary outcome variables, which will be obtained from registries, but data will be missing if hospitals do not provide data to the registries (Norwegian Institute of Public Health).

If wards do not provide outcome data for the primary outcome, we will contact them and ask them to provide the missing data. If more than 5% of wards do not provide primary outcome data, we will use Little's test to assess if the data are unlikely to be missing completely at random (MCAR) but missing at random (MAR). If we assess the data to be MAR, we will use multiple imputation by chained equations (MICE) and perform the prespecified analyses including all wards.

The primary outcome comprises counts of observations and counts of those observations that meet the criteria for compliance. We anticipate that both counts will be missing for wards that do not provide outcome data. It will therefore be necessary to impute missing values for these variables in a way that ensures both are valid counts, and that the number of compliances is not greater than the number of observations imputed for a given ward. We will do this as follows.

For each ward with non-missing primary outcome data, we will compute a point estimate of the logit risk of compliance as $w_i = \text{logit}\, r_i = \log r_i - \log(1 - r_i)$, where $r_i = n_i/N_i$. Within the MICE framework, we will impute logit risks for wards with missing primary outcome data (e.g., using a linear model with normal errors), and hence risks for wards with missing primary outcome data as $r_i = 1/(1 + e^{-w_i})$, where w_i is an imputed logit risk. Missing observation counts (N_i) will be modelled using a Poisson model. This may impute that no observations were made in some wards. We will address this issue by replacing zero observation counts with ones. Finally, we will impute the number of compliances for the wards with missing primary outcome data as the passive variable $\lfloor N_i r_i + {}^1/_2 \rfloor$ (i.e., "rounding half up" a possibly non-integral number of imputed compliances to an integer), where N_i and r_i are imputed as described above. The r_i will not be used in subsequent analyses.

The imputation model for the primary outcome will include:

- Treatment assignment
- Adherence (if available; see section 7.6)
- Hospital (assuming data are not missing for all wards in a hospital)
- Hospital region
- Type of department
- The secondary outcomes (on the basis that the primary outcome is likely to be highly correlated with other outcomes)

If we cannot assess that the data are likely MAR and cannot rule out the possibility that the data are missing not at random (MNAR), we will perform complete case analyses, but will not draw strong conclusions about the causal effect of NOST and report the missing data as a possible limitation.

The secondary outcomes depend on registry data. It is possible that some wards or hospitals will either never report data or will intermittently report data. Both possibilities represent missing outcome data problems. We will therefore follow the same general approach as for missing primary outcome data. In the case that secondary outcome data is known to be intermittently reported (i.e., data for some collection periods are available for certain wards, but no data have been reported by the wards for other periods), we will use MICE to impute for the collection periods with missing data and then compute the total number of inpatients infected (for example). The imputed outcome is therefore a function of imputed variables (a "passive" variable), which should be handled carefully in the analysis. In Stata, it is necessary to use the syntax "**`mi passive: generate`**…" or "**`mi register passive`**…", prior to "**`mi estimate:`**…" to ensure the passive variables are

generated correctly in the imputed data sets. In R, passive variables can be handled using the approach described in section 3.4 of van Buuren and Groothuis-Oudshoorn [1].

Unless legally or ethically prohibited, we will report any missing data and wards or hospitals that cannot be included in analysis, for example due to withdrawal from the trial.

7.6 Intercurrent Events

Recall from section 3.2 that the research question is about the effect of implementing and adhering to NOST. We anticipate that some wards may not adhere or fully adhere to the treatment to which they were randomized. For example, a ward may have been randomized to implement and adhere to NOST for the one-year period of the trial, but either did not implement NOST or did not fully adhere. Similarly, a ward may have been randomized to the control arm (i.e., to not implement NOST), but NOST was partially or fully implemented. Because such events would occur after randomization and would change the interpretation of the outcome from that originally intended and hence give a treatment effect estimate that does not have the intended interpretation, non-adherence can be treated as an intercurrent event (ICE) [2].

All estimands (see section 9) for which statistical analyses will be performed utilize collapsible measures of effect (risk ratio or mean difference). This allows any non-adherence to be addressed via regression adjustment. We will measure adherence using the proportion of time after randomization and up and including the final data collection period (i.e., adherence will be measured using a value between 0 and 1, inclusive).

If 5% or more wards do fully not adhere to their randomized treatment, we will estimate the effect of being randomized and adhering to NOST as the sum (on the log scale for ratio effect measures) of two effects: randomization to NOST and adherence to NOST. Note that adherence to NOST will also be measured in the control arm to facilitate the estimation of the effect of full versus no adherence to NOST. We will then exponentiate the sum to obtain the desired estimate of treatment effect.

If fewer than 5% of wards do not adhere to their randomized treatment we will not account for the non-adherence, but report it is a possible limitation.

8 Summary of Study Data

A CONSORT trial flow diagram will be presented.

A table of baseline characteristics will be presented following CONSORT guidelines, as follows. Continuous variables will be summarized using the following descriptive statistics: median and interquartile range (IQR). The frequencies and percentages of observed levels will be reported for categorical variables. Baseline characteristics will be reported in a table like the following:

		Control (N = XXX)	**NOST (N = XXX)**
Hospitals	N (%)	XXX	XXX
Number of beds	Median [IQR]	XXX [XXX to XXX]	XXX [XXX to XXX]
Health region			
North	N (%)	XXX	XXX
Central	N (%)	XXX	XXX
…	…	…	…
South-East	N (%)	XXX	XXX
Wards	N (%)	XXX	XXX
Health region			

North	N (%)	XXX	XXX
Central	N (%)	XXX	XXX
…	…	…	…
South-East	N (%)	XXX	XXX
Department			
Oncology	N (%)	XXX	XXX
Orthopedics	N (%)	XXX	XXX
…	…	…	…
Neurology	N (%)	XXX	XXX
Profession			
Physician	N (%)	XXX	XXX
…	…	…	…
Nurse	N (%)	XXX	XXX

Effect estimates for the primary and secondary outcomes will be reported in a table like the following:

	Control	**Intervention**	**Risk Ratio (95% CI)**	***p*-value**	**Risk Difference (95% CI)**
Primary outcome					
Compliance	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
Secondary outcomes					
Outbreaks	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
Patients infected	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
…	…	…	…	…	…
			Mean Difference (95% CI)	***p*-value**	
Bed days/stay	XXX (SD = XXX)	XXX (SD = XXX)	XXX [XXX to XXX]	0.XXX	

Effect estimates for the subgroup analyses for the primary outcome will be reported in a table like the following:

	Control	**Intervention**	**Risk Ratio (95% CI)**	***p*-value**	**Risk Difference (95% CI)**
Department					
Psychiatry	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
Orthopedics	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
…	…	…	…	…	…
Neurology	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
Profession					
Physician	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
…	…	…	…	…	…
Nurse	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
Indication					
Before patient contact	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]
…	…	…	…	…	…
After exposure to body fluid	XXX/XXX (XXX%)	XXX/XXX (XXX%)	XXX [XXX to XXX]	0.XXX	XXX [XXX to XXX]

8.1 Derived Variables

The secondary outcomes are derived from ongoing notifications or from reports that hospitals should submit for specific periods, e.g. four times a year. For example, the number of HAIs (NOIS) over the one-year trial period would be the sum of such reports. This has implications for imputation, if some hospitals do not report data for one or more of the four-month periods (see section 7.5).

8.2 Protocol Deviations

Any deviations from the original protocol or this SAP will be reported and justified.

9 Estimation and Analyses

While this trial is cluster-randomized (ward is the cluster), outcome data will be measured at the level of ward. The unit of randomization and the unit of analysis are therefore identical. It is not necessary to account for cluster randomization as in trials where clusters of units are randomized, and outcomes are measured on the units.

9.1 Risk Ratio Estimation

If fewer than 5% of wards do not adhere to their randomized treatment, we will estimate marginal risk ratios (RRs) using binomial regression without accounting for the non-adherence. For the primary and most of the secondary outcomes, this is:

$$n_i \sim B(N_i, p_i) \text{ where } \log p_i = \beta_0 + \beta_X x_i$$

With respect to the primary outcome, n_i is the number of compliances in the ith hospital ward, N_i is the number of observations made in the ith hospital ward, β_0 estimates log risk in control, x_i indicates if the ith hospital ward was randomized to NOST, β_X estimates the effect of being randomized to NOST as a marginal log risk ratio. The marginal log risk ratio of being randomized to and adhering to NOST is estimated by β_X.

If 5% or more wards do fully not adhere to their randomized treatment, we will estimate marginal risk ratios (RRs) using binomial regression as follows:

$$n_i \sim B(N_i, p_i) \text{ where } \log p_i = \beta_0 + \beta_X x_i + \beta_A a_i$$

Here, $0 \leq a_i \leq 1$ models the degree of adherence by the ith hospital ward to NOST (see section 7.6) with $\alpha_i = 0$ corresponding to non-adherence and $\alpha_i = 1$ corresponding to full adherence, and β_A estimates the effect of full adherence to NOST. The marginal log risk ratio of being randomized to and adhering to NOST is estimated by $\beta_X + \beta_A$. Note that the x_i and a_i may be highly correlated (almost collinear) because we expect that most wards that implement NOST will also adhere to NOST. For this reason, we will not interpret β_X and β_A separately. We will assume that the statistical software will cope well with the high correlation but will address this issue during analysis in the case of nonconvergence.

Estimation can be performed using Stata as follows (see footnote 1 on page 14):

```
binreg compliances i.treat c.adh, n(observations) rr vce(cluster hospital)
```

where data are arranged with one ward per row, **`compliances`** is a variable containing the numbers

of observations where the employees comply, **treat** is a treatment indicator where 1 indicates randomization to NOST, **adh** is a variable containing adherences (which would be omitted if fewer than 5% of wards do not adhere to their randomized treatment), **observations** is a variable containing the numbers of observations made, and **hospital** is a factor variable that codes for the hospital to which the wards belong. Treatment effect can then be estimated on the log scale using:

```
lincom _b[1.treat] + _b[adh]
```

Estimation can be performed using R as follows:

```
model <- glm(cbind(compliance_success, compliance_failure) ~ treat + adh,
                                data = df,
                        family = binomial(link = "log")
           coeftest(df, vcov. = vcovCL(model, cluster = hospital)
              lincom <- glht(model, linfct = c(treat + adh = 0)
```

If the analysis is performed in R after imputation (see section 7.5), the script will be run in a loop through all imputed data sets before estimates and standard errors are combined using the pool() function in the "mice" package.

The variable names may differ in the actual data set.

For each risk ratio estimate, we will also compute an indicative risk difference (RD; see the tables in section 8). This will be done by taking the difference between a point estimate of the control risk and the product of that control risk and the estimated risk ratio. The indicative RD will be presented by with a 95% confidence interval.

9.1.1 Sensitivity analyses

The estimator proposed for the case that 5% or more wards do fully not adhere to their randomized treatment involves accounting for an intercurrent event (incomplete adherence to NOST due to, for example, late start or cessation) by estimating the log risk ratio of being randomized to and adhering to NOST as $\beta_X + \beta_A$, where a_i, the variable for coefficient β_A, is a continuous variable in [0, 1] with $a_i = 0$ modelling no adherence to NOST and $a_i = 1$ modelling full adherence to NOST. Because this analysis requires a modelling assumption, we will perform sensitivity analyses for each analysis that uses the model proposed.

In the first sensitivity analysis, we will estimate marginal risk ratios using a restricted estimation sample that will exclude wards that did not adhere to the treatment to which they were assigned (i.e., we will exclude wards randomized to NOST which did not fully implement NOST, and exclude wards randomized to control that partially or fully implemented NOST). The restricted estimation sample will be analyzed using the following model:

$$n_i \sim B(N_i, p_i) \text{ where } \log p_i = \beta_0 + \beta_X x_i$$

where the log risk ratio of fully adhering to NOST is estimated by β_X.

In the second sensitivity analysis, we will model adherence using a factor variable, with $k + 1$ levels that model degree of adherence to NOST. We will estimate marginal risk ratios (RRs) using binomial regression as follows:

$$n_i \sim B(N_i, p_i) \text{ where } \log p_i = \beta_0 + \beta_X x_i + \beta_{A1} a_{i1} + \beta_{A2} a_{i2} + \cdots + \beta_{Ak} a_{ik}$$

where $\{a_{i1}, a_{i2}, \dots, a_{ik}\}$ are binary variables that indicate the non-reference levels of the factor variable that models adherence (these variables will be created automatically in software), and $\{\beta_{A1}, \beta_{A2}, \dots, \beta_{Ak}\}$ are log risk ratios for the non-reference levels. We will define the reference (i.e., base) level to be no adherence to NOST and estimate the effect of being randomized and adhering to NOST as $\beta_X + \beta_{A1}$, where β_{A1} is the log risk ratio for the level of the factor variable corresponding to full adherence to NOST.

Having obtained the three estimates of the effect of being randomized and adhering to NOST (i.e., one from the analysis specified in section 9.1 and one for each of the two sensitivity analyses), we will assess the sensitivity of our estimates to the assumptions underpinning the three models using seemingly unrelated estimation[1] (SUE) as follows. Let $\{\gamma_1, \gamma_2, \gamma_3\}$ denote the three estimates on the log risk ratio scale. Sensitivity will be assessed using a two-sided Wald test of the hypothesis $\gamma_1 = \gamma_2 = \gamma_3$. We will report the result of this test and interpret the p-value according to the following table:

P-value	**Estimates γ_1, γ_2, and γ_3 all statistically exclude no effect**	**Interpretation**
<0.05	Yes, same directions	• The magnitude of the effect of being randomized and adhering to NOST <u>is sensitive</u> to modelling assumptions. • The direction of the effect of being randomized and adhering to NOST <u>is not sensitive</u> to modelling assumptions.
<0.05	Yes, different directions	• One or both sensitivity analyses disagreed with the main estimator of the effect of being randomized and adhering to NOST. • The effect of the intervention is unclear.
<0.05	No	• One or both sensitivity analyses disagreed with the main estimator of the effect of being randomized and adhering to NOST. • The effect of the intervention is unclear.
≥0.05	Yes, same directions	• The magnitude of the effect of being randomized and adhering to NOST <u>is not sensitive</u> to modelling assumptions. • The direction of the effect of being randomized and adhering to NOST <u>is not sensitive</u> to modelling assumptions.
≥0.05	Yes, different directions	*This condition is nonsensical: it should not be possible for the estimates to indicate different directions of effect and for the test that they are equal to suggest they are.*
≥0.05	No	• The magnitude of the effect of being randomized and adhering to NOST <u>is not sensitive</u> to modelling assumptions. • The direction of the effect of being randomized and adhering to NOST <u>is not sensitive</u> to modelling assumptions.

[1] Note that, in Stata, use of SUE requires that clustering within hospital be specified as an option to **`suest`** rather than **`binreg`** for all models, and that the variance-covariance matrix of the estimates (VCE) be computed using the observed, rather than the expected, information matrix (i.e., using the **`vce(oim)`** option of **`binreg`**).

9.2 Mean Difference Estimation

If fewer than 5% of wards do not adhere to their randomized treatment, we will estimate marginal differences in means using linear regression with cluster-robust standard errors, with hospital as the cluster variable:

$$y_i \sim N(\beta_0 + \beta_X x_i, \sigma)$$

where for the bed days per inpatient stay outcome, y_i is bed days per inpatient stay for the ith hospital ward, β_0 estimates the mean for controls, x_i indicates if the ith hospital ward was randomized to NOST, β_X estimates the effect of being randomized to NOST as a marginal mean difference, and σ is the residual standard deviation.

If 5% or more wards do fully not adhere to their randomized treatment, we will estimate marginal differences in means using the model:

$$y_i \sim N(\beta_0 + \beta_X x_i + \beta_A a_i, \sigma)$$

where $0 \leq a_i \leq 1$ models the degree of adherence by the ith hospital ward to NOST (see section 7.6 and the definition of a_i in section 9.1) and β_A estimates the effect of full adherence to NOST. The marginal mean difference of being randomized to and adhering to NOST is estimated by $\beta_X + \beta_A$. We will use the same strategy for handling possible nonconvergence as described in section 9.1.

Estimation can be performed using Stata as follows:

```
regress bed i.treat c.adh, vce(cluster hospital)
```

where data are arranged with one ward per row, **bed** is a variable containing bed days per inpatient stay, and the other variables are defined as in section 9.1.

Estimation can be performed using R as follows:

```
model = glm(bed ~ treat + adh, data = df,
   family = gaussian(link = "identity")
coeftest(df, vcov. = vcovCL(model, cluster = hospital)
```

9.2.1 Sensitivity analysis

We will perform sensitivity analyses similar to those for risk ratios (see section 9.1.1).

10 Estimands

This section presents all estimands in terms of the objective, a statement of the estimand in plain language, the target population (i.e., who or what the estimate can be applied to), the analysis set (i.e., what subset, if any, of the full analysis set will be used to perform the analysis), the outcome variable, strategies for handling ICEs and missing data, the effect measure that will be estimated, and the estimator used to perform the analysis.

10.1 Primary Outcome Estimand

Objective To assess the effect of NOST with respect to compliance with hand hygiene recommendations.
Estimand The risk of compliance with hand hygiene recommendations, measured at one year, by employees who perform patient-related work in Norwegian hospitals if NOST was introduced and

adhered to versus if NOST was not introduced and not adhered to.	
Target population Employees	**Analysis set** All
Outcome Variable Compliance with hand hygiene recommendations	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2 Secondary Outcome Estimands

10.2.1 Outbreaks (VESUV)

Objective To assess the effect of NOST with respect to outbreaks.	
Estimand The risk of outbreak, measured over one year, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Hospital wards	**Analysis set** All
Outcome Variable Outbreaks (VESUV)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.2 Inpatient Infection (VESUV)

Objective To assess the effect of NOST with respect to inpatient infection (using data from VESUV).	
Estimand The risk of inpatient infection defined as part of an outbreak, measured over one year, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Inpatients	**Analysis set** Wards with reported outbreak
Outcome Variable Inpatient infections (VESUV)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.3 Employee Infection (VESUV)

Objective To assess the effect of NOST with respect to employee infection.	
Estimand The risk of employee infection as defined part of an outbreak, measured over one year, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Employees	**Analysis set** Wards with reported outbreak
Outcome Variable Employee infections (VESUV)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.4 Inpatient Infection (NOIS)

Objective To assess the effect of NOST with respect to inpatient infection (using data from NOIS).	
Estimand The risk of inpatient infection, measured over one year in repeated prevalence studies, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Inpatients	**Analysis set** All
Outcome Variable Inpatient infection (NOIS)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.5 Inpatient Antibiotic Treatment (NOIS)

Objective To assess the effect of NOST with respect to inpatient antibiotic treatment.	
Estimand The risk of inpatient antibiotic treatment, measured over one year in repeated prevalence studies, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Inpatients	**Analysis set** All
Outcome Variable Inpatient antibiotic treatment (NOIS)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.6 Inpatient Postoperative Site Infection (NOIS-POSI)

Objective To assess the effect of NOST with respect to inpatient treatment for postoperative site infection.	
Estimand The risk of postoperative site infection, measured over one year in periods of incidence surveillance of patients undergoing certain surgical procedures, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Inpatients undergoing surgical procedures monitored in NOIS-POSI	**Analysis set** Wards with relevant surgical patient groups (included in the NOIS-POSI surveillance)
Outcome Variable Inpatient postoperative site infection (NOIS-POSI)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.7 Inpatient HAI Diagnosis (NPR)

Objective To assess the effect of NOST with respect to inpatient HAI diagnosis.	
Estimand The risk of inpatient HAI diagnosis, measured over one year, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Inpatients	**Analysis set** All
Outcome Variable Inpatient HAI diagnosis (NPR)	
ICE Strategy See sections 7.6, 9.1, and 9.1.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) RR and indicative RD	**Estimator** Binomial regression with cluster-robust standard errors

10.2.8 Bed-days per Inpatient per Stay (NPR)

Objective To assess the effect of NOST with respect to bed-days per inpatient per stay.	
Estimand The mean number of bed-days per inpatient per stay, measured over one year, in Norwegian hospital wards if NOST was introduced and adhered to versus if NOST was not introduced and not adhered to.	
Target population Inpatients	**Analysis set** All
Outcome Variable Bed-days per stay (NPR)	
ICE Strategy See sections 7.6, 9.2, and 9.2.1	**Missing Data Strategy** See section 7.5
Effect Measure(s) MD	**Estimator** Linear regression with cluster-robust standard errors

11 Reporting Conventions

In general, percentages in the table of baseline characteristics will be rounded to whole numbers, and summaries of continuous variables will be presented to one decimal place.

Statistical estimates will be presented as points with two-sided 95% confidence intervals and *p*-values.

Point estimates and confidence intervals will be reported to 2 decimal places, and *p*-values will be reported to three decimal places or as $p < 0.001$.

12 Quality Assurance of Statistical Programming

The statistical code will be written and tested on fictitious data. Testing will include producing all results, tables, and figures as they will appear in the final manuscript prior to analyzing the trial data.

Analysis code and data will be versioned using an appropriate system to be chosen by the analyst and may be published alongside the trial results.

13 Summary of Changes to the Protocol and/or SAP

This section is to be completed to document and justify any changes to a published version of this SAP.

14 Acknowledgements

The template used to produce this SAP was derived with permission from a template developed by the Michigan Institute for Clinical and Health Research (MICHR; grant UL1TR002240), which was based in turn on version CCTU/TPLV2 of the template produced by Cambridge University Hospitals Clinical Trials Unit.

15 References

1. van Buuren, S. and K. Groothuis-Oudshoorn, *mice: Multivariate Imputation by Chained Equations in R.* Journal of Statistical Software, 2011. **45**(3): p. 1 - 67.
2. International Council for Hamonisation, *E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials, May 2021*. 2021.