The ALMA-CKD trial STATISTICAL ANALYSIS PLAN

Study title	An electronic triggering decision-support system to improve detection, management and nephrologist referral of patients with chronic kidney disease in primary care: A pragmatic cluster randomized controlled trial
Short title	ALMA-CKD
Registration	NCT06386172
Principal Investigators	Juan Jesus Carrero, Professor of Kidney Epidemiology, Karolinska Institutet Stefan Jacobson, Professor of Nephrology, Karolinska Institutet
Randomization date	4th September 2023
SAP version	1.0
Date	20 November 2025
Written by	Stephen Nash (Statistician, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet)

Summary of versions and changes

Version	Description	Date approved
1	Initial version	2025-11-20

Contents

1 Background

One in ten adults in Region Stockholm have chronic kidney disease (CKD), which dramatically increases healthcare costs and the risk of poor patient outcomes, including cardiovascular disease and death. Identification and early management of these patients is done in primary care settings. However, most adults with CKD in our region are under detected, undiagnosed and undertreated, with low rates of referral to nephrology-specialist care.

In this pragmatic cluster-randomized controlled trial (RCT) among primary healthcare centers in Region Stockholm we seek to test the effect of a new electronic clinical decision support (CDS) triggering system to assist general practitioners through the guideline-recommended processes of CKD care of this large population segment.

2 Objectives

The objective of the ALMA study is to evaluate the effectiveness of the CDS trigger to ensure that people who meet the criteria for CKD testing are invited to test and are actually tested.

2.1 Outcomes

There are four key outcomes for this trial; one primary outcome (A) and three secondary outcomes (B, C, D). All four are binary outcomes.

	Comparison
A	**Undergoing a creatinine test within 12 months of a visit at which time the patient was eligible for annual kidney function screening**
B	Undergoing an albuminuria test within 12 months of a visit at which time the patient was eligible for annual kidney function screening
C	Undergoing a creatinine re-test within 6 months of the results of a creatinine test which indicated a re-test
D	Undergoing an albuminuria re-test within 6 months of the results of an albuminuria test which indicated a re-test

Eligibility for annual kidney function screening (outcomes A and B) is defined as a diagnosis of any one of the following:

- CKD
- hypertension
- diabetes
- cardiovascular disease
- nephrectomy

An indication for re-testing creatinine (outcome C) is having eGFR<60 ml/min/1.73 m2 in the original test.

An indication for re-testing albuminuria (outcome D) is having a dipstick test denoting KDIGO A2+ or a urinary albumin to creatinine (uACR) test >3 mg/mmol in the original test.

2.2 Other secondary outcomes

There are four non-test related secondary outcomes:

Outcome 2.1 CKD diagnosis: Proportion of patients meeting indication who receive a diagnosis of CKD (according to ICD-10) within six months of eGFR / A2 re-testing. By "indication" we mean eGFR<60 ml/min/1.73 m^2 and/or dipstick A2+ or >3 mg/mmol at re-test. Only patients with at least six months of follow-up after their re-test and who are CKD-negative at the first visit will be included in this analysis (that is, re-test must occur on or before 30 June 2025).

Outcome 2.2 Initiation of new CKD-treatment: The proportion of patients who start new CKD-modifying treatment within six months of re-test indicating a CKD diagnosis (as defined above).

Outcome 2.3 Referral for ultrasound: The proportion of patients meeting the criteria for nephrologist care who receive a referral for ultrasound of the kidneys.

Outcome 2.4 Accepted to nephrologist care: The proportion of patients meeting the criteria for nephrologist care (according to Swedish guidelines) who are newly accepted into care. By "newly accepted" we mean we will exclude patients who have had a visit with a nephrologist in the two years prior to their first visit to a participating healthcare center during the trial period.

3 Study Methods

3.1 Trial design

This is a cluster-randomized controlled trial, with 66 healthcare centers randomized in a 1:1 ratio to either the new CDS trigger system (the intervention arm) or usual practice (the standard arm). The study includes all 66 Primary Healthcare centers in Region Stockholm, and is embedded in the Stockholm CREAtinine Measurements project (SCREAM), which will enable us to have access to all relevant data for patients at these healthcare centers without need to modify clinical practice or perform active data collection.

There will be a four-month run-in period following randomization to allow clinicians to adjust to the new system. For data analysis purposes, recruitment into the trial will start on 1st January 2024.

3.2 Randomization

Randomization was restricted to ensure (approximate) balance on two factors, based on 2019 (pre-Covid) levels:

- Number of people eligible to be tested for albuminuria
- Proportion of people actually tested for albuminuria

We restricted the randomization as follows:

- There were four centers with an eligible population much larger than the other 62 centers

    - Husläkarmottagning Handen = 5434
    - Ektorp = 4712
    - Liljeholmens vårdcentral = 3762
    - Jakobsberg vårdcentral = 3286
    - No other center had a 2019 eligible population greater than 3000
- We first created 10,000 randomizations of the 62 “non-large” centers, such that 31 were in each arm.
- We then created all possible permutations of the four large centers such that two were in one arm and two were in the other arm. There are six possible permutations which achieve this (AABB, ABAB, ABBA, BAAB, BABA, BBAA)
- We combined these two sets to give us 60,000 potential randomizations
- We then mirrored these allocations to give us 120,000 randomizations (that is, reversed the allocation of intervention and standard arms)
- For each potential allocation we calculated the average proportion of people tested in 2019. (This was calculated on a center level, not a population level, so the denominator was 33 in each arm.)
- If the difference in the average proportion between the arms was greater than 3.7% we discarded the allocation. The figure of 3.7% was chosen because, under different randomly drawn sets of the initial 10,000 allocations, 3.7% ensured that around, but less than, 10% of the final 120,000 allocations were discarded. And further, it was felt that 3.7% was small enough such that if there was indeed a between-arm difference of 3.7% we would still be confident that the two arms were similar enough to be confident of drawing unbiased conclusions form the results.
- Finally, one allocation was randomly selected from the remaining allocations

The code to perform this randomization was written by the trial statistician (SN) and then given to an independent statistician, outside of the study team, who changed the random seed which selected the 10,000 allocations, and the random seed which selected the final allocation. When the code was run, 10,348 (8.6%) allocations were removed due to a difference between the arms in 2019 test rates, and hence the final allocation was chosen from 109,652 possible allocations.

3.3 Post-randomization change

Four days after randomization the study team were informed that two of the 66 centers had been merged into one administrative unit (although still on separate sites). These two centers were Bredäng and Sätra; their new name is “Bredäng Sätra vårdcentral”. Due to the merger we will not be able to retrieve outcome data at a lower level than the merged center. As by chance these two centers had been assigned to the same arm (the intervention) we decided to proceed with the trial under the original randomization. We note here that the two original restrictions still held after the merger of these centers:

1. the population of the new center is 1,455, which means it ranks 24th (of 65) in size
2. the difference between the two arms in terms of 2019 test rates is now 3.4% (< 3.7%)

This SAP is written to reflect that change. We will analyze the study with 65 centers (33 in the standard arm and 32 in the intervention).

3.4 Sample size

The study was adequately powered to detect an absolute improvement in our primary study outcome of 13% with an alpha of 5%. To calculate the power of our trial, we evaluated the processes of care in the 66 primary healthcare centers during 2019 as collected in the SCREAM project. We chose this year prior to the COVID-19 pandemic to have more objective estimations of health use not affected by and policies during the pandemic. During 2019, 662 995 unique individuals visited those primary healthcare centers at least once. Of those, 92 011 were eligible for screening/testing of kidney function: 66% took a creatinine test, and 34% took a creatinine test. Of people with first detected eGFR<60 ml/min/1.73 m2 or albuminuria, 59% and 28% were retested for creatinine or albuminuria within 3-6 months.

We assumed a recruitment period of 12 months and a total of 33 health center (clusters) per arm with the same number of patients as in 2019. We estimated the intraclass correlation for the outcomes to be 0.1 (as quantified in the UK trial by Major et al, J Am Soc Nephrol. 2019 Jul;30(7):1261-1270.). The table below gives the originally calculated power, and the power we now have, conservatively assuming 32 clusters per arm.

Comparison	Original power (33 clusters per arm)	New power (32 clusters per arm)
A. Primary outcome: Detect an increase of people having a creatinine test from 66% to 79%	86%	**85%**
B. Detect an increase of people having a albuminuria test from 34% to 48%	84%	**83%**
C. Detect an increase of people having a re-test of creatinine from 59% to 73%	86%	**82%**
D. Detect an increase of people having a re-test of albuminuria from 28% to 42%	86%	**84%**

3.5 Framework

This will be a superiority study, with two-sided p-values and confidence intervals for all comparisons.

3.6 Statistical interim analysis and stopping guidance

There is no planned interim analysis, and no pre-specified stopping guidance.

3.7 Timing of final analysis

The final analysis will be performed 24 months after the start of the recruitment phase (that is, after the initial run-in period). Hence the cutoff point for measuring outcomes is 31st December 2025.

4 STATISTICAL PRINCIPLES

4.1 Confidence intervals and p-values

We will calculate 95% confidence intervals throughout. No formal statistical significance level will be set.

4.2 Adherence and Protocol deviations

The trial intervention is implemented at the level of the cluster (health center). As such, we do not anticipate any protocol deviations or lack of adherence. If other clusters are merged during the duration of the study we will handle this on a pragmatic, case-by-case basis. If the two centers are in the same arm we will continue without change. If the two centers are in different arms, the decision as to which diagnostic system to use will be made by us in collaboration with Region Stockholm.

4.3 Definition of estimands

Following the estimands framework defined in the European Medicines Agency "ICH E9 (R1) addendum", we here define estimands for the main trial objectives. For each outcome we will report risk ratios as the primary population-level summary (effect size), and risk difference as a secondary measure.

4.3.1 Primary estimand

Attribute	Description
Population	Adults who visit a primary healthcare center and are indicated for annual kidney function screening
Treatment	Intervention: New clinical decision support triggering system designed to manage chronic kidney disease, implemented at healthcare centers using the ALMA platform Control: Standard ALMA CDS platform
Endpoint	Undergoing a creatinine test within 12 months of first visit in 2024
Summary measures	1. Ratio of proportions (risk ratio) 2. Absolute difference of proportions (risk difference)
Intercurrent events	1. Death (from any cause): Exclude from analysis 2. Patient transfers to (has visit with) another healthcare center in this trial: use outcome value, and analyze according to the randomisation group of the original healthcare center

4.3.2 Secondary estimands

Outcome B: Albuminuria testing

Attribute	Description
Population	Adults who visit a primary healthcare center and are indicated for annual kidney function screening
Treatment	Intervention: New clinical decision support triggering system designed to manage chronic kidney disease, implemented at healthcare centers using the ALMA platform Control: Standard ALMA CDS platform
Endpoint	Undergoing an albuminuria test within 12 months of first visit in 2024

Summary measure	1. Ratio of proportions (risk ratio) 2. Absolute difference of proportions (risk difference)
Intercurrent events	As above

Outcome C: Creatinine re-testing

Attribute	Description
Population	Adults who receive a creatinine test result at a primary healthcare center suggesting low eGFR (<60 ml/min/1.73 m^2) and thus are indicated for a re-test within six months
Treatment	Intervention: New clinical decision support triggering system designed to manage chronic kidney disease, implemented at healthcare centers using the ALMA platform Control: Standard ALMA CDS platform
Endpoint	Undergoing a creatinine re-test within six months of a test result which indicated a re-test
Summary measure	1. Ratio of proportions (risk ratio) 2. Absolute difference of proportions (risk difference)
Intercurrent events	As above

Outcome D: Albuminuria re-testing

Attribute	Description
Population	Adults who receive an abnormal albuminuria test result at a primary healthcare center of dipstick test denoting KDIGO A2+ or a urinary albumin to creatinine (uACR) test >3 mg/mmol and are thus indicated for a re-test within six months
Treatment	Intervention: New clinical decision support triggering system designed to manage chronic kidney disease, implemented at healthcare centers using the ALMA platform Control: Standard ALMA CDS platform
Endpoint	Undergoing an albuminuria re-test within six months of a test result which indicated a re-test
Summary measure	1. Ratio of proportions (risk ratio) 2. Absolute difference of proportions (risk difference)
Intercurrent events	As above

4.4 Analysis populations

The unit of observation is the individual patient. Each patient will only be included once in each analysis.

For outcomes A and B, each patient will be included at the time of their first visit during the recruitment phase: we will refer to this as their Index date for that outcome. This will be used as the start time for the 12 months window for testing.

For outcomes C and D, each patient will be included at the time of their results from their creatinine (C) or albuminuria (D) test: we will refer to this as their Index date for that outcome. This will be used as the start time for the 6 months window for re-testing.

There will be an analysis population associated with each of the four outcomes listed above (A-D). Randomisation occurred in September 2023; however, we will allow a four-month run-in period for the new system. Recruitment of individual participants will begin on 1st January 2024. For outcomes A and B, recruitment will close on 31st December 2024, to allow 12 months of follow-up for all patients. For outcomes C and D, recruitment will close on 30th June 2025, to allow six months of follow-up for all patients.

Modified ITT population A: All patients who attend a visit at one of the study health centers during 2024 and who remain alive and registered at that center for a further 12 months, and are eligible for annual kidney function screening.

Modified ITT population B: Same as Modified ITT population A (above)

Modified ITT population C: All patients who receive a test result for creatinine from one of the study health centers between 1st January 2024 and 30th June 2025, who remain alive and registered at that center for a further six months, and have an indication for a re-test.

Modified ITT population D: All patients who receive a test result for albuminuria from one of the study health centers between 1st January 2024 and 30th June 2025, who remain alive and registered at that center for a further six months, and have an indication for a re-test.

5 TRIAL POPULATION

5.1 Screening data

Individual patients are not screened for entry other than requirement to be eligible for screening / re-test defined above. All consenting Primary Healthcare Centers in Region Stockholm are included in the trial.

5.2 Eligibility

There are no eligibility checks for individual patients.

5.3 Withdrawal / Follow-up

It is not anticipated that any center will withdraw from the study; if it happens we will decide what to do on a pragmatic, case-by-case basis. Using the SCREAM database we will be able to access relevant patient data for all patients in the 12 months of the study recruitment, up to 12 months after the final patient has been registered.

5.4 Baseline patient characteristics

We will report baseline patient characteristics for each analysis population, both at an individual level (summarized by arm) and summarized for each cluster.

6 ANALYSIS

6.1 Analysis methods

The same analysis methods will be used for all outcomes defined above. In all analyses we will always adjust for the randomization factor of cluster size (either "large" or "not-large") and the cluster-level proportion with that outcome in the 12 months preceding the start of the trial. For the two centers which merged we will combine the data and treat as one entity. The precise means of adjustment will depend on the analysis method, see below for details.

We will perform both an individual-level analysis (to estimate risk ratios) and a cluster-level analysis (to estimate risk difference).

6.1.1 Individual-level analysis

For the individual-level analysis we will use a mixed-effects regression model using a log link and binomial distribution, with the healthcare centers as random intercepts to estimate a risk ratio. Fixed effects will include randomized arm, and *a priori* covariates for cluster size ("large" vs "non-large" healthcare center), the outcome proportion from 2019 (which was used in the randomization procedure; included as a linear continuous variable). Other covariates will be included if there is a "large" difference between the arms as defined below:

- For binary variables, the difference between the proportion in the arms is greater than 20% (pooling all data and ignoring clusters)
- For continuous variables, $M_1 > 1.2 \times M_2$, where M_1 is the mean in the arm with the higher mean, and M_2 is the mean in the arm with the lower mean (pooling all data and ignoring clusters).

6.1.2 Cluster-level analysis

For the cluster-level analysis we use a method described by Hayes & Moulton (Cluster randomised trials, second edition. CRC Press, pp 197-198) to estimate a risk difference. This method allows for the adjustment of individual-level covariates using a two-stage method. In the first stage, an individual-level logistic regression analysis is performed, ignoring the clustering, and excluding the treatment allocation. For each cluster, we then calculate the difference between the number of events (tests, re-tests) expected by the model and the observed number. In the second stage, we compare the cluster summaries from each arm using a standard two-sample t-test. If the difference in expected number of events is systematically different between the arms, we conclude this is due to the effect of the triggering system.

We will adjust, *a priori*, cluster size ("large" vs "non-large") and outcome proportion in 2019, plus any other baseline characteristics which have a large imbalance between trial arms, as defined above (§6.1.1).

This analysis will be performed using the user-contributed Stata command `clan` (https://ideas.repec.org/c/boc/bocode/s458844.html). For full transparency, the trial

statistician Stephen Nash was a co-developer of this command. The command is free to use and he does not receive any financial benefit from it.

6.2 Sub-group analysis

Two sub-group analyses are planned: frequency of visits and sex. These will be conducted on the four testing outcomes (A-D) only.

6.2.1 Sub-group analysis 1: Repeat visits

We will categorize each participant according to whether they attended one visit, or more than one visit, during the recruitment phase (12 months for outcomes A and B, 18 months for outcomes C and D). We will include this variable in the individual-level analysis model as an interaction with the treatment variable to assess if the intervention is more effective among patients who attend more frequently.

6.2.2 Sub-group analysis 2: Sex

We will assess whether the intervention works differently among men and women by including sex in the individual-level analysis model as an interaction with the treatment variable.

6.3 Exploratory analyses

Exploratory analysis will use a time-to-event outcome for each participant, instead of a simple binary variable calculated at 12 or six months. This analysis will be performed on an individual-level dataset, using time from Index date as the timescale. All patients will be censored at death or on 31st December 2025.

6.4 Missing data

We will be using routinely collected data from each healthcare center, as recorded on their computer systems. As such, we do not anticipate any missing data, nor do we plan to impute any missing values.

6.5 Sensitivity analyses

One of the centers was used to test the functionality of the CDS-triggers, and in the randomization process this center was randomized to standard care, thus not receiving the intervention. There is a small risk that the doctors at this center became more alert to the national guidelines for CKD and as a result perform more retesting, issues more diagnoses and more referrals. This will be investigated with a sensitivity analysis, where we will remove that center and compare the remaining 64 centers (32 vs 32).

6.6 Process measures

We will report process measures to assess the usage of each system.

1. Number of active users per month
2. Number of times the algorithms have been activated and the top ones per month
3. Number of times that the algorithms have been opened (i.e. there has been a click)

We may report other measures in addition to those listed above, as process measures will be reported as descriptive statistics only, with no formal comparisons between trial arms. We

may choose to investigate the effect of specific measures of clinician interaction with the system, either as a mediator of the effect of the intervention or for its direct effect on the outcome. Such *ad hoc* analysis will be exploratory in nature and not pre-specified in this plan.

6.7 Harms

We do not anticipate any harms due to the implementation of this new triggering system. Clinicians will maintain decision-making control over all aspects of patient care at all times.

6.8 Statistical software

Analyses will be performed in Stata (version 18 or above).

Appendix: Shell tables

Shell Table 1: Baseline characteristics in mITT

Characteristic	Standard Arm	Intervention Arm
Number of eligible patients (N)		
Sex = female: N (%)		
Age: median (IQR)		
Attained education		
Income category		
Comorbid history		
Ongoing medications		
Creatinine(umol/L): mean (SD)		
Albuminuria (g/dL): mean (SD)		
eGFR (ml/min/1.73 m^2)		
KDIGO G category		
KDIGO A category		

Shell Table 1. Baseline characteristics of eligible participants in each trial arm for CKD screening
Patients eligible for an annual kidney function screening.

Shell Table 2a: Main results from individual analysis

Outcome	Measure	N	n with outcome (%)	Effect size	95% CI	P-value
A	Standard arm					
	Intervention arm					
B	Standard arm					
	Intervention arm					
C	Standard arm					
	Intervention arm					
D	Standard arm					
	Intervention arm					

Shell table 2a. Results of primary analysis (individual-level)
The effect size column will show estimated risk ratio for the individual-level analysis (RR > 1 in favor of the new triggering system)

Shell Table 2b: Main results from cluster-level analysis

Outcome	Measure	Mean proportion	Effect size	95% CI	P-value
A	Standard arm				
	Intervention arm				
B	Standard arm				
	Intervention arm				
C	Standard arm				
	Intervention arm				
D	Standard arm				
	Intervention arm				

Shell table 2b. Results of primary analysis (cluster-level)
The effect size column will show estimated difference in proportions (risk difference) for cluster-level analysis (a difference greater than 0 indicates an effect in favor of the new triggering system)

Shell Table 3: Outcomes at a cluster level

	A		B		C		D	
	N	**n with outcome (%)**	**N**	**n with outcome (%)**	**N**	**n with outcome (%)**	**N**	**n with outcome (%)**
Standard arm centers								
1								
etc								
Intervention arm centers								
34								
etc								

Shell table 3. Outcomes in each center