REMORA 2 Stepped Wedge Cluster Randomised Trial (REMORA2_SWT)

Statistical Analysis Plan (SAP)

17 Feb 2026

Table of contents

1 Approvals

This statistical analysis plan (SAP) has been prepared, reviewed, and approved by:

	Name	Role In Trial	Date	Signature (or 'by email')
Prepared by	Matthew Parkes, Dr	Supervising Statistician	20 Nov 2025	
	Yahya Salimi, Dr	Trial Statistician (former)	31 Jul 2025	by email
	Colin Everett, Mr	Trial Statistician	09 Dec 2025	by email
Reviewed by	William Dixon, Prof	Co-Lead Investigator	24 Nov 2025	by email
	Sabine van der Veer, Prof	Co-Lead Investigator	24 Nov 2025	by email
	Deb Griffiths-Jones, Dr	REMORA2 Programme Manager	08 Aug 2025	by email
	Chris Linsley, Mr	REMORA2 Data Manager	03 Jul 2025	by email
	Peter Bower, Prof	Co-Investigator (Primary Care & Trials Methodology)	15 Jul 2025	by email
	Sean Gavan, Dr	Co-Investigator (Health Economics)	14 Jul 2025	by email
	Nick Bansback, Dr	Co-Investigator (Health Economics)	14 Jul 2025	by email
Approved by	Lucy Yardley, Prof	PSC Chair	20 Jan 2026	by email
	Karla Hemming, Prof	PSC Statistician	17 Feb 2026	by email
	Duncan Porter, Prof	PSC Clinician	16 Jan 2026	by email
	Savia de Souza, Dr	PSC Patient Member	20 Jan 2026	by email

Statistical Analysis Plan (SAP)

Version 1.0 Date 17 Feb 2026

2 Revision History

Version	Date of Issue	Summary of Revisions	Timing with Respect to Analyses
v1.0	17 Feb 2026	N/A (first issue)	Before final analysis and final data lock

3 List of Abbreviations

Brief IPQ	Brief Illness Perception Questionnaire
CACE	Complier-Average Causal Effect
CDAI	Clinical Disease Activity Index
CI	Confidence Interval
COM-B	Capabilities, Opportunities, Motivations for Behaviour change questionnaire
CONSORT	Consolidated Standards of Reporting Clinical Trials
CONSORT-SW-CRT	Consolidated Standards of Reporting Clinical Trials and Stepped Wedge Cluster Randomised Trials
CRSE	Continuous Recruitment, Short Exposure
CRSE-SWCRT	Continuous Recruitment, Short Exposure Stepped Wedge Cluster Randomised Controlled Trial
CRP	C-Reactive Protein
CRT	Cluster Randomised Trial
CRXO	Cluster Randomised Cross-Over Trial
DAG	Directed Acyclic Graph
DAS28	Disease Activity Score for 28 joints
EHR	Electronic Health Record
EQ-5D-5L	EuroQol five-dimension scale questionnaire
ESR	Erythrocyte Sedimentation Rate
EULAR	European League Against Rheumatism
GLMM	Generalised Linear Mixed Model
HAQ	Health Assessment Questionnaire
HEAP	Health Economic Analysis Plan
ICC	Intra-class correlation
ISRCTN	International Standard Randomised Controlled Trials Number
IMD	Index of Multiple Depriviation
IQR	Interquartile Range
IRAS	Integrated Research Application System
ITT	Intention to Treat
MAR	Missing at Random
MAR	Missing Completely at Random
MCID	Minimal Clinically Important Difference
MI	Multiple imputation
MICE	Multiple imputation by chained equations
MNAR	Missing not at random
NHS	United Kingdom National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health and Care Research
PEI	Patient Enablement Instrument
PGfAR	Programme Grant for Applied Research

PSC	Programme Steering Committee
QALYs	Quality-adjusted life years
RA	Rheumatoid Arthritis
RAID	Rheumatoid Arthritis Impact of Disease (RAID) score (7 items)
RAPID-3	Routine Assessment of Patient Index Data 3
RCT	Randomised Controlled Trial
REMORA	Remote Monitoring in Rheumatoid Arthritis
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SIV	Site Initiation Visit
SJC28	Swollen Joint Count for 28 joints
SoC	Standard of Care
SW-CRT	Stepped Wedge Cluster Randomised Trial
TJC28	Tender Joint Count for 28 joints
UoM	University of Manchester
WPAI-RA	Work Productivity and Activity Impairment Questionnaire

4 Administrative information

4.1 Trial and Trial registration information

4.1.0.1 Study Full Title and Trial Acronym

Transforming Outpatient Consultations by Integrating Regular Symptom Tracking into Clinical Care: A **S**tepped **W**edge Randomised Controlled **T**rial (RCT) of **Re**mote **Mo**nitoring of **R**heumatoid **A**rthritis, Compared to Usual Care (REMORA-SWT)

4.1.0.2 Trial registration number: 51539448

4.1.0.3 IRAS Identifier: 322792

4.1.0.4 Protocol Version 3.0 Date 15 May 2025

4.1.0.5 Data Management Guidebook [In Draft]

Names and roles of SAP contributors

Name	Role(s) in Trial	Date Joined the Trial
Professor William Dixon	Co-Lead Investigator	(from inception)
Professor Sabine van der Veer	Co-Lead Investigator	(from inception)
Dr Deborah Griffiths-Jones	REMORA Programme Manager	(from inception)
Mr Chris Linsley	REMORA2_SWT Data Manager	Sep 2024
Dr Matthew Parkes	Co-Investigator & Supervising Statistician	(from inception)
Dr Yahya Salimi	REMORA2_SWT Trial Statistician	Oct 2024 - Sep 2025
Mr Colin Everett	REMORA2_SWT Trial Statistician	Oct 2025

5 Introduction

This document details the planned statistical analyses for the REMORA2 Stepped Wedge Cluster Randomised Trial (SW-CRT), "REMORA2_SWT" (NIHR PGfAR Ref NIHR202030; ISRCTN51539448; IRAS ID 322792).

5.1 Background & Rationale

The **Re**mote **Mo**nitoring of **R**heumatoid **A**rthritis (REMORA) system is an intervention designed to improve the care of people living with rheumatoid arthritis (RA) by enabling them to track their symptoms using a smartphone app and integrate the resultant data into the electronic health record (EHR) to be viewed at forthcoming consultations with their rheumatology team. In earlier phases of the REMORA programme, we conducted a single-site pilot study among 20 RA participants who tracked symptoms over a three-month period, and their clinicians. We showed that the REMORA system is beneficial with regular symptom reporting identifying changes in condition, which would otherwise have been missed, promoting shared conversations about disease management and altering treatment decisions. Next, we developed scalable infrastructure to ensure that the use of the REMORA system can be extended and rolled out across multiple NHS sites. We then tested our scalable infrastructure to deploy our intervention and identified and implemented several modifications to support its use in future.

REMORA2_SWT is a mixed-methods study, incorporating quantitative, qualitative, and health economics components to the analyses. This document (the statistical analysis plan, SAP) describes the definition, derivation, and reporting of quantitative elements of the study, and how these contribute to the overall objectives of REMORA2_SWT. For analysis of the health-economic-related elements of the study, please refer to the Health Economics Analysis Plan (HEAP).

5.2 Primary Research Question

The REMORA2_SWT study's overall aim is to identify whether rheumatology clinics using integrated symptom tracking (IST) are superior compared with usual standard of care (SoC) clinics in improving care and outcomes among patients with RA.

5.3 Assumed Mechanism of Action

Figure 1 outlines the underpinning theoretical pathway relating the mechanism of action of the study intervention and the outcome. This pathway is not known truly, but assumed, and illustrated here to outline the theoretical rationale behind objectives and outcomes selected for study in the trial.

In brief, we assume that the randomised intervention (clinic type: either IST or SoC clinics) will lead to treatment decisions that arise from patient/clinician discussion during rheumatology clinic visits. These decisions about medication or non-pharmacological treatments then lead to changes in disease activity, which are further mediated through (unmeasured) health behaviour that occurs in participants between visits (e.g. actual medication use, appointment attendance, health behaviours, lifestyle changes, etc.). Treatment decisions are made based on discussions between clinicians and participants. This is reliant on understanding disease activity, which is mediated through the engagement that clinicians and participants have during clinic visits. For participants in the IST arm, we assume that participants' engagement with the REMORA system (app) in between study visits and clinicans' update of viewing symptom tracking graphs will influence the information quality, and thus, the outcome.

The precise mechanism of engagement will differ between conditions, hence the overall comparison is a simple head-to-head comparison of disease activity at follow-up between the two clinic types, irrespective of these mediating factors (see Table 2 for a more detailed definition of the comparison of interest.

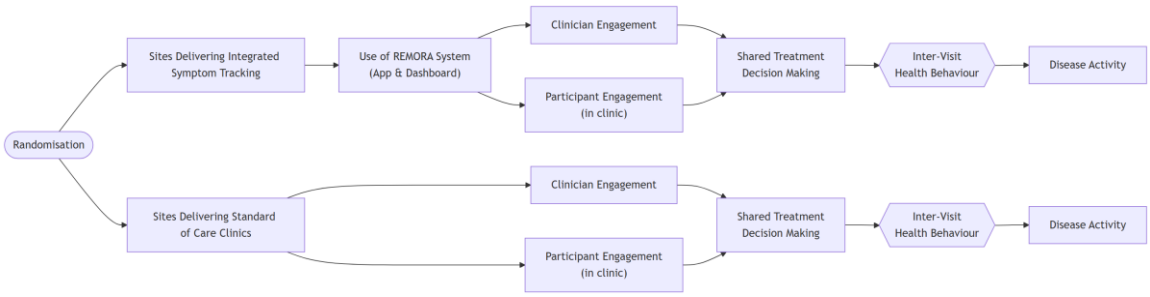

Figure 1: Directed Acyclic Graph (DAG) of Assumed Relationships in the REMORA2 Stepped Wedge Trial (REMORA2-SWT). Hexagonal boundaries indicate factors not measured in REMORA2-SWT.

5.4 Study Objectives

Study objectives are defined in Table 1.

Note: The objectives listed in Table 1 relate to REMORA2_SWT as a whole (a mixed-methods study), and therefore encompasses qualitative, quantitative, and health economic-specific objectives. Where objectives are directly addressed by analyses defined in this document, they are stated explicitly.

Table 1: REMORA2 Stepped Wedge Trial (REMORA2_SWT) Objectives

Objective	Outcome/end products relating to the objective	Endpoint/origin/source of outcome
Primary Objective		
1. Evaluate the effect of integrated symptom tracking, compared with usual standard of care (SoC) on improvement in disease activity in patients with RA attending rheumatology clinics in the secondary care setting.	Disease Activity Score as measured on the DAS28/Clinical Disease Activity Index (CDAI) taken from EHRs at baseline and 12 (±3) months[1].	Difference in marginal mean disease activity score at 12 (±3) months follow-up between usual standard of care (SoC) and integrated symptom tracking conditions, after accounting for measured covariates (including baseline DAS28).
Secondary Objectives		
2a. Evaluate the effect of clinic visits with integrated symptom tracking, compared with clinic visits using usual standard of care (SoC) on secondary quantitative indicators of disease activity and impact, such as pain, disability, quality of life, and work productivity and activity impairment.	European League Against Rheumatism (EULAR) Response Categories: Good, Moderate, Non-Responders, derived from disease activity score from EHRs.	Difference in proportion of participants meeting good/moderate/non-response at 12 (±3) months follow-up between standard of care (SoC) and integrated symptom tracking conditions.
	Patient-reported disease activity	Difference in RAPID-3 items total score at 12 (±3) months follow-up relative to initial (baseline) visit between standard of care (SoC) and integrated symptom tracking conditions.
	Difference in patient global assessment from the REMORA web survey at 12 (±3) months follow-up relative to initial (baseline) visit between standard of care (SoC) and integrated symptom tracking conditions.	
	Difference in Rheumatoid Arthritis Impact of Disease (RAID) score from the REMORA web survey at 12 (±3) months follow-up relative to initial (baseline) visit between standard of care (SoC) and integrated symptom tracking conditions.	
	Joint involvement and severity via the Swollen Joint Count for 28 joints (SJC28) and the Tender Joint Count for 28 joints (TJC28) score taken from the EHRs.	Difference in Swollen Joint Count for 28 joints (SJC28) total joint count at 12 (±3) months follow-up relative to initial (baseline) visit between standard of care (SoC) and integrated symptom tracking conditions.

[1] As clinic visits are not scheduled to occur at exactly 12 months, DAS28 scores are eligible within 3 months either side of the 12-month timepoint.

	Difference in Tender Joint Count for 28 joints (TJC28) total joint count at 12 (±3) months follow-up relative to initial (baseline) visit between standard of care (SoC) and integrated symptom tracking conditions.	
	Medication Use	Difference in quantity and type of medication use extracted from EHR at 12 (±3) months relative to initial level at first completion ('baseline') between standard of care (SoC) and integrated symptom tracking conditions.
	Work impairment	Difference in Work Productivity and Activity Impairment Questionnaire (WPAI-RA) score as reported in the REMORA web survey at 12 (±3) months relative to initial level at first completion ('baseline') between standard of care (SoC) and integrated symptom tracking conditions.
	Quality of life	Difference in the EuroQol five-dimension scale questionnaire (EQ-5D-5L) health profiles, index score and quality-adjusted life years gained as reported in the REMORA web survey at 12 (±3) months relative to initial level at first completion ('baseline') between standard of care (SoC) and integrated symptom tracking.
	Disability	Difference in the Health Assessment Questionnaire (HAQ) score as reported in the REMORA web survey at 12 (±3) months follow-up relative to initial (baseline) visit between standard of care (SoC) and symptom tracking conditions.
2b. Evaluate the effect of clinic visits with integrated symptom tracking, compared with clinic visits using usual standard of care (SoC) on shared decision making, self-management and the consultation experience, using mixed methods.	Patients' experience of shared decision-making during clinic visits.	Difference in CollaboRATE score as reported in the REMORA web survey at 12 (±3) months relative to initial level at first completion (baseline) between standard of care (SoC) and integrated symptom tracking conditions.
	Researcher-observed shared decision-making processes in clinic.	Difference in OPTION score as collected during clinic observations between standard of care (SoC) and integrated symptom tracking conditions; conversation analysis of recorded consultations compared between standard of care (SoC) and integrated symptom tracking conditions.
	Patients' experience of consultation quality.	Difference in Patient Enablement Instrument (PEI) score as reported in the REMORA web survey at 12 (±3) months relative to initial level at first completion (baseline) between standard of care (SoC) and integrated symptom tracking conditions.
	Patients' capabilities, opportunities, and motivations for self-management behaviour.	Difference in Capabilities-Opportunities-Motivations-Behaviour (COM-B) score as reported in the REMORA web survey at 12 (±3) months relative to initial level at first completion (baseline) between standard of care (SoC) and integrated symptom tracking conditions.
	Patients' understanding of their RA.	Difference in brief Illness Perception Questionnaire (Brief IPQ) score as reported in the REMORA web survey at 12 (±3) months relative to initial level at

		first completion (baseline) between standard of care (SoC) and integrated symptom tracking conditions.
	Perceived differences, benefits and disadvantages regarding self-management and consultation and decision-making processes of using integrated symptom tracking compared with standard of care (SoC).	Comparative analysis between standard of care (SoC) and integrated symptom tracking conditions based on interviews with health care professionals and patients, and observations and conversation analyses of clinic visits.
3. Identify barriers to behaviour change, intervention uptake and wider NHS implementation, and ways to address these barriers.	Patient and health care professional expectations, experiences and views regarding the acceptability and usefulness of (the support for) integrated symptom tracking.	Interviews with: patients and health care professionals taking part in the trial, health care professionals not taking part in the trial, patients declining to take part or who do not download the app, and other professionals and volunteers involved in the implementation of REMORA; observations of clinic visits; informal feedback from patient and professional participants and others (e.g., other staff, carers) via other routes, e.g., helpdesk queries.
	Facilitators and barriers to patient and clinical behaviour change, and uptake and wider NHS implementation of integrated symptom tracking.	
4. Evaluate the incremental cost-effectiveness of integrated symptom tracking in RA compared with usual standard of care (SoC).	Quality of life	Difference in EuroQol five-dimension scale questionnaire (EQ-5D-5L) health profiles, index score and quality-adjusted life years gained as reported in the REMORA web survey at 12 (±3) months relative to initial level at first completion ('baseline') between standard of care (SoC) and integrated symptom tracking.
	Duration of consultations	Difference in cumulative consultation length at 12 (±3) months between standard of care (SoC) and integrated symptom tracking conditions.
	Medication Use	Difference in quantity and type of medication use extracted from the EHR at 12 (±3) months between standard of care (SoC) and integrated symptom tracking conditions.
	Resource use	Difference in cumulative resource use as extracted from the EHR at 12 (±3) months related to initial level at first completion ('baseline') for standard of care (SoC) and integrated symptom tracking conditions.
		Difference in cumulative resource use and travel time as reported in the REMORA web survey at 12 (±3) months between standard of care (SoC) and integrated symptom tracking conditions.
	Health care professionals' engagement with integrated symptom tracking.	Number and duration of REMORA dashboard views in the EHR by health care professionals.
	Patient engagement with integrated symptom tracking.	Number of completed daily, weekly and monthly questionnaires in the REMORA app in the integrated symptom tracking condition.
5. Explore possible mechanisms that might explain any observed improvement in disease activity (or lack thereof) in relation to integrated symptom tracking	All quantitative and qualitative outcomes as listed above for objectives 1-4.	Relationships between primary and secondary indicators of disease activity (O1-2a) and other outcomes: shared decision making, self-management and consultation experience (O2b); patients' and professionals' views and experi-

using mixed methods.		ence of integrated symptom tracking and factors influencing successful implementation (O3); and patient and health care professionals' engagement with integrated symptom tracking (O4).
Tertiary Objectives		
6a. Examine representativeness and diversity of study participants compared to the wider RA population.	Characteristics (demographics, comorbidities, etc.) and medication/resource use of consented patient participants, linked to regional shared primary and secondary care health records of external reference RA populations.	Differences in REMORA participants' characteristics and medication/resource use compared to the wider RA population.
6b. Examine longitudinal trajectories of symptoms and other aspects of RA, including assessment of treatment response.	All quantitative outcomes as listed above for objectives 1-2 that were collected at more than one time point, linked to regional shared primary and secondary care health records of external reference RA populations.	Trajectories of outcomes through time for integrated symptom tracking and SoC condition, including associations with initiating a range of RA treatments.

6 Methods

6.1 Trial Design

REMORA2_SWT is a continuous recruitment, short exposure stepped wedge cluster randomised controlled trial (CRSE-SWCRT, as per the definition by Copas et al.[1]). In this study, 'clusters' are individual hospital sites (16 in total) which are randomised to switch from SoC to IST. Two sites will switch at the end of each 3-week period simultaneously, with two sites being referred to hereon as a 'sequence'. There are 16 study sites (clusters) in total, therefore 8 sequences. With 8 sequences, there will be 9 periods in total to complete the stepped wedge design. A schematic diagram of the study design is shown in Figure 2.

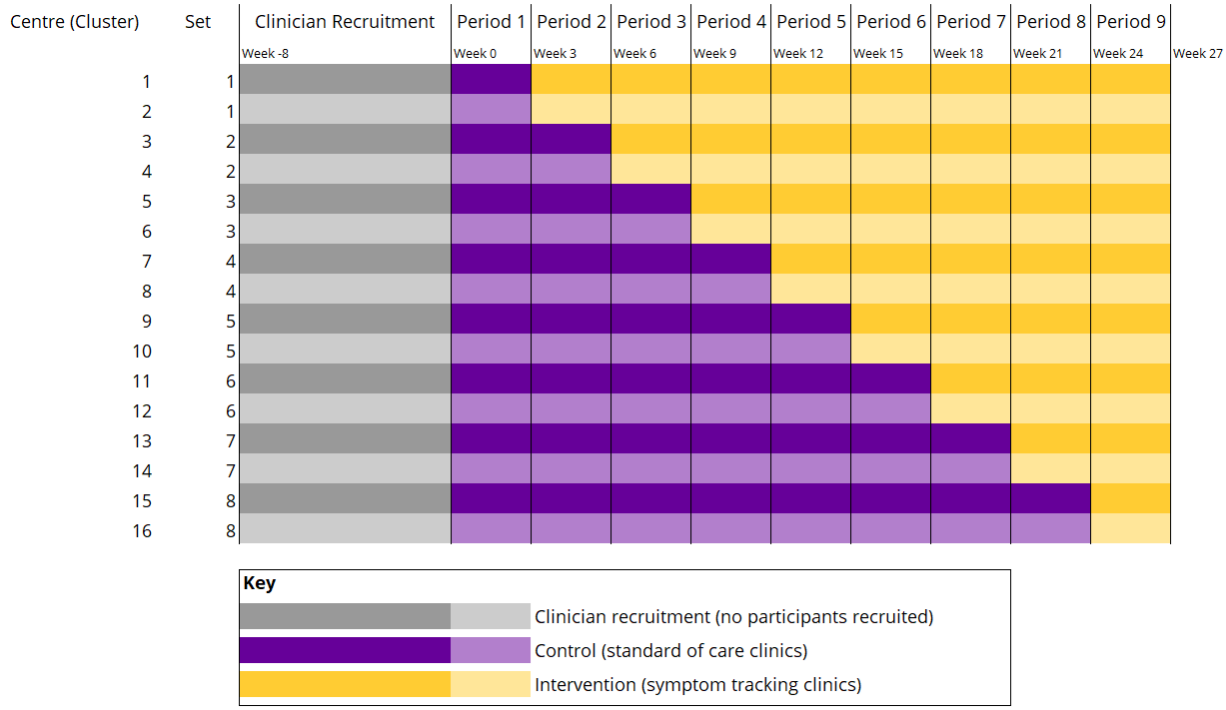

Figure 2: REMORA2 Stepped Wedge Cluster Randomised Trial (REMORA2_SWT) Schematic

6.1.1 Justification for Use of the Stepped Wedge Design

The CRSE-SWCRT design was considered the most appropriate design to answer the research objectives outlined in Section 5.4. Justification for the design is provided here, in-keeping with current recommendations[2].

The intervention for REMORA2_SWT necessitates the use of a pragmatic clinical trial design that can be implemented in a real-world setting, embedded in actual clinical pathways in the NHS. This favours pragmatic clinical trial designs, of which stepped wedge trials (SWTs) and cluster randomised trials (CRTs) are commonly used, and perhaps also cluster cross-over trials (CRXO), although these are relatively less common.

Since we anticipate that the implementation of IST clinics in NHS hospital sites will be challenging, and to an extent, unique to each hospital site, the research team are keen to maximise the opportunities to implement IST during the trial to increase learning and data capture of the implementation process. Furthermore, all of the sites were keen to experience this novel intervention. A CRT

comparing a new intervention versus SoC will only implement the novel intervention at 50% of sites, which is much more limited than an SWT which implements the intervention at *all* study sites.

Using an SWT design allows the research team to have a sequential or staggered implementation of the novel intervention as the trial progresses, whereas a CRT design requires all sites (clusters) to be fully implemented and operational at the start of the trial. However, as the SWT design requires all sites to open to participant recruitment simultaneously, clinician training and recruitment in REMORA2_SWT was deemed to be more practical to implement if done simultaneously before trial commencement (see grey region in Figure 2).

Of the SWT subtypes, the CRSE-SWCRT was considered the most appropriate, as it allows measurement of a baseline observation that can be incorporated into the analysis to improve estimate precision (compared to a cross-sectional SWT which has only one time point observation per participant). However, unlike open and closed cohort designs, the CRSE-SWCRT also only allows participants to experience one intervention, which reduces the risk of contamination of the treatment effect between interventions. In our study specifically, for examples, participants allocated to SoC will never experience the IST condition, and vice versa. Of the available SWT subtypes, the CRSE-SWCRT design was considered the best balance between required sample size and avoiding contamination.

While a CRXO design would also allow implementation of the new intervention at all study sites, the requirement for 50% of sites to 'deimplement' IST (a technically burdensome intervention to initially implement) was deemed impractical, and of limited benefit, given that this would leave 50% of sites at trial completion without the intervention, with the added risk of carry-over effects (contamination) from clinicians and sites that had used IST and then reverted to SoC clinics. Also, the CRXO design precludes staggered implementation of the novel intervention, unlike an SWT design.

6.2 Intervention(s)

6.2.1 Active/experimental condition

The active condition in REMORA2_SWT is rheumatology clinics using IST, which are enabled to display data from the REMORA app via the 'REMORA dashboard'. Participants in this condition are given access to the REMORA symptom tracking app, which is a bespoke app created for this project containing a series of daily, weekly, and monthly questionnaires to collect data about participants' RA symptoms (for precise items collected, see Section 6.6.2. This data is integrated into the EHRs for review during participants' rheumatology clinic visits, via the REMORA dashboard. Clinicians taking part in the trial are given mandatory training on how to access and interpret symptom tracking data in the dashboard prior to commencement of the trial (in weeks -8 to 0, see Figure 2, grey shaded region), with the opportunity to attend refresher training throughout the study period.

The app is designed to provide enhanced symptom information sharing for clinicians and participants during the consultation. However, the REMORA system provides no guidance about treatment decisions, meaning that decision making is left to the discretion of clinicians as per usual care.

6.2.2 6.2.2 Control/comparator condition

The control condition is SoC clinics. Participants receive usual care, i.e. typical rheumatology clinics as currently delivered in the participating study sites. The REMORA dashboard is not activated for clinicians to review during the SoC condition, and participants recruited during this condition are not provided with access to the REMORA app. Symptoms reported to clinicians in the SoC condition are therefore using the same communication methods as in typical rheumatology clinics (i.e. via oral questioning and recall). Treatment decision-making is left to the discretion of clinicians, as per usual care. We consider this control an 'absent' comparator (rather than active, or placebo comparator),

as the active element of the intervention is not present in this arm, and all other conditions remain identical.

6.3 Randomisation

Prior to commencement of the study, the 16 participating sites will be randomly assigned a site letter (A-P) by the programme manager. This will mask the site's identity from the trial and supervising statisticians (MP & YS), who will remain blinded to site name for the duration of the study until the primary and secondary pre-specified analyses detailed in this SAP are completed. Independently of the University of Manchester (UoM) study team, the supervising statistician (MP), using computer-generated simple randomisation via a pseudorandom number generator, will randomly allocate the site letters into 8 sequences of 2 clusters (sites). The pseudorandom number generator will be initiated with a random 5-digit integer which is used as the 'seed', taken from a website that generates random integers taken from atmospheric weather data (https://www.random.org/).

Following this, the programme manager will randomly allocate the 8 sequences of sites using simple randomisation into an order 1-8 dictating the order in which each sequence will switch over from SoC to IST. The code and seed used to generate the randomisations will be stored in a password-protected secure folder that only the supervising statistician will have access to until recruitment to the trial has completed. The list of site names and letter allocations will be held securely by the UoM study team; the supervising statistician will not have access to this information.

The code generating the randomisation scheme will be shared on the REMORA Open Science Framework Repository after participant follow-up has completed - see Section 10.

6.4 Sample Size

Sample size was based on the ability to detect a minimal clinically important difference (MCID) of 0.6 points in the Disease Activity Score for 28 Joints (DAS28) score between conditions at 12 months. A total of 736 participants (46 per cluster) were required to achieve 90% power with a two-sided alpha of 0.05, assuming an intraclass correlation coefficient (ICC) of 0.05, a standard deviation (SD) of 1.2, and a discrete time decay correlation structure with a cluster autocorrelation (CAC) of 0.8. We assumed a coefficient of variation in cluster size of 0.15 to allow for some variation in cluster sizes. The sample size calculation also includes 30% attrition rate in the intervention group and 15% dropout rate in both conditions[3].

6.5 Framework

6.5.1 Primary Outcome Analysis Framework

The primary outcome will be analysed using a *superiority* hypothesis testing framework, evaluating whether the IST condition leads to a statistically significant improvement in disease activity compared to the SoC condition at 12-month follow-up (±3 months).

6.5.2 Secondary Outcome Analysis Framework(s)

The framework for the secondary outcomes in this trial varies based on the specific objective and nature of the outcomes:

All quantitative secondary outcomes described in this SAP will be analysed using a *superiority* hypothesis testing framework, assuming the superiority of IST compared with SoC. Secondary qualitative outcomes described in Table 1 are not detailed in this plan.

These include:

- Disease activity measures, such as European League Against Rheumatism (EULAR) response categories, Routine Assessment of Patient Index Data 3 (RAPID-3), Rheumatoid Arthritis Impact of Disease (RAID), and joint involvement and severity (SJC28, TJC28).
- Patient-reported outcomes, including Work Productivity and Activity Impairment (WPAI-RA), quality of life (via the EuroQol 5 dimensions, 5 levels questionnaire, EQ-5D-5L), medication use, Resource use, and disability (using the Health Assessment Questionnaire, HAQ).
- Shared decision-making and consultation experience, using the CollaboRATE questionnaire.
- Self-management and illness perception, as measured by the Capabilities, Opportunities, Motivations for Behaviour change (COM-B) and Brief Illness Perception Questionnaire (IPQ-B) scores.

For these outcomes, the null hypothesis assumes no difference between conditions, while the alternative hypothesis posits that IST leads to superior outcomes.

6.6 Outcome Definitions

6.6.1 Primary Outcome

6.6.1.1 Definition

The primary outcome measure that fulfils objective 1 (O1) is the DAS28 score[4], assessed at 12 months (±3 months to allow for variation in clinic scheduling). The DAS-28 score is extracted from EHRs by study staff at the participating study sites, and recorded on a dedicated REDCap form.

The DAS28 is a composite outcome measure consisting of the following components:

- A tender joint count over 28 joints (TJC28)
- A swollen joint count over 28 joints (SJC28)
- A patient-reported global assessment of RA activity, and
- One inflammatory marker blood test: Either C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR).

The DAS28 is expressed on a continuous scale, with more positive score values indicating worse disease activity.

In the literature, the term 'DAS28' is used to refer collectively and/or interchangeably between the DAS28-CRP and DAS28-ESR - DAS28 scores derived using one or other blood test. There is variation in practice between hospital trusts - often due to local trust policy reasons (e.g. CSR is the test offered by that hospital site). REMORA2-SWT has been designed to allow capture of alternative disease activity measures if they are the only ones available at the site, which can then be used to infer the DAS28-CRP score (see Section 6.6.1.2).

The specific version of DAS28 considered the primary outcome in this study is the DAS28-CRP score at 12-months. In pre-recruitment site surveys, the DAS28-CRP was the most commonly used derivation of the DAS-28 across the 16 participating sites in REMORA2_SWT. Study sites are preferentially encouraged to collect the DAS28-CRP version of the DAS28, although overriding factors may prevent this from occurring in all sites at all time points.

The blood test used in deriving the DAS28-CRP/ESR is ideally collected/requested at the clinic visit where the other components are recorded. However, in practice, previous recent blood tests are sometimes used to derive a DAS28 score, for example to avoid an unnecessary blood draw. For the analysis, blood tests collected within 3 months of the visit date were considered acceptable, with blood tests occuring outside this window considered missing and not used to derive the score for that visit.

Since follow-up appointments are scheduled according to clinical need, the 12 month ('final follow-up') assessment will be defined as the closest DAS28-CRP score to 12 months between months 9 and 12 or, if no visit has happened in that preceding 3 month interval, the closest DAS28-CRP score to 12 months between months 12 and 15 (i.e. the subsequent 3 month interval). All secondary outcome measures for objectives O2a and O2b (as defined in Table 1) will be collected at baseline and this final follow-up visit. Additional clinic visits may be scheduled during the study period between the baseline and final visit, and EHR data on disease activity will be extracted for these 'interim' follow-up visits. However, the final analysis will only include the baseline (initial) and final follow-up data in the model.

Planning of the REMORA-SWT took place during the COVID-19 pandemic period when movement restrictions were in place, leading to fewer face-to-face appointments. An alternative measure of disease activity, the Clinical Disease Activity Index (CDAI)[5,6] has been included in both the REMORA web survey, and when follow-up appointments take place over the telephone, to provide insight on disease activity during a period when face-to-face appointments, and thus acute phase reactants, could not be captured. The CDAI combines the TJC28, SJC28, physician global, and patient-reported global assessment of RA activity.

As the COVID-19 pandemic restrictions were lifted during the study, we anticipate relatively few CDAI observations to be collected. The number collected will be reported in the trial reports, but as this frequency is anticipated to be very low, the CDAI will not be considered a secondary outcome of interest (due to the high proportion of missing observations relative to the total sample size).

Collectively, the disease activity measures for REMORA2-SWT are calculated as shown in Note 1.

Note 1: Calculation of DAS28 Composite Outcomes

Primary Outcome:

DAS28-CRP score (from EHRs)[4]: $DAS28Score_{CRP} = 0.56 \times \sqrt{TenderJointCount} + 0.28 \times \sqrt{SwollenJointCount} + 0.014 \times PatientGlobalScore + 0.36 \times \ln(CRP + 1) + 0.96$

Alternative Disease Measures:

DAS28-ESR score (from EHRs): $DAS28Score_{ESR} = 0.56 \times \sqrt{TenderJointCount} + 0.28 \times \sqrt{SwollenJointCount} + 0.014 \times PatientGlobalScore + 0.70 \times ln(ESR + 1)$

DAS28-CRP 3-component score (from EHRs) (DAS28-CRP-3)[7]: $DAS28Score_{CRP} = 0.56 \times \sqrt{TenderJointCount} + 0.28 \times \sqrt{SwollenJointCount} + 0.36 \times \ln(CRP + 1) * 1.10 + 1.15$

DAS28-ESR 3-component score (from EHRs) (DAS28-ESR-3)[7]: $DAS28Score_{ESR} = 0.56 \times \sqrt{TenderJointCount} + 0.28 \times \sqrt{SwollenJointCount} + 0.70 \times \ln(ESR + 1) * 1.08 + 0.16$

CDAI score (from EHRs)[5]: $CDAIScore = TenderJointCount + SwollenJointCount + PatientGlobalScore + PhysicianGlobalScore$

A form of CDAI is also collected on the REMORA web survey, where no physician is present. A 'modified CDAI score' which does not feature the physician rating is derived as:

Modified CDAI score (from web survey): $mCDAIScore = TenderJointCount + SwollenJointCount + PatientGlobalScore$

This modified CDAI score can also be calculated from the EHR data from clinic visits, allowing a direct comparison of disease activity between the web survey and EHR datasets.

The EHR case report form (CRF) contains a field (`"Das28Score"`) which allows capture of the DAS28 score calculated in clinic by the healthcare professional. If available, this score then may be used for treatment decision making in that visit. As it is unclear how this has been derived in relation to the source components and their timing, this variable is captured, but will *not* be used in the primary analysis, with the primary analysis taking place on the derived score calculated directly from the components as per the formulae in Note 1, but will be considered in primary outcome sensitivity analyses.

A comparative analysis will be performed to check for consistency between the score calculated in clinic, and the scores calculated from the components in the database and reported in the trial output. Notably, treatment decisions are often made based on this in-clinic-calculated score, and so the capture of this information, and how this relates to other disease activity measures, was felt to be useful as a check against data quality and consistency in this outcome (as one might expect high consistency between these two measures).

6.6.1.2 Missing components

The primary outcome analysis will take place on the DAS28-CRP score (with missing observations observations of this outcome being imputed as detailed in Section 9.2). There are planned analysis investigating the effect of the use of these alternative measured of disease severity, which are described in Section 9.1.1.2.

There are 3-component versions of the DAS28-CRP and DAS28-ESR[7], available when a patient global score is not collected, but the other components are. These will be derived using the formulae listed in Note 1.

Guidance on how to derive DAS28-CRP, DAS28-ESR, and CDAI scores when one or more other component(s) are missing is unclear in the literature. For example, if a blood test is unavailable, there is no guidance on whether or how a DAS-28 score can be derived.

If any one or more component(s) of any of the disease activity measures is missing for a given participant and study visit, with the exception of the patient global score, then we consider it not possible to calculate a valid score, and will consider this observation of the outcome missing.

Of the available measures, we consider the following heirarchy the order of importance for the disease activity outcome measures collected:

1. DAS28-CRP score (*Primary Outcome Measure*)
2. DAS28-ESR score
3. DAS28-CRP [3-component] score
4. DAS28-ESR [3-component] score
5. Clinic-calculated DAS28 Score only (referred to as `"Das28Score"` in the study datasets)

These alternative measures of disease activity are intended to be used to infer disease activity status, should the DAS28-CRP not be available, or should some component(s) not be collected.

Aside from omission or substitution, a third way to consider this missing component issue is to use the other available disease activity measures to infer plausible values of the missing DAS28-CRP score using multiple imputation (MI). This will also be investigated, and details on the multiple imputation model used are outlined in Section 9.2.

6.6.1.3 Primary Outcome Analysis Estimand

The estimand of interest to address objective O1 is defined in Table 2.

Table 2: Primary Outcome Analysis Estimand

Attribute	Definition	As Defined in REMORA2_SWT
Population	Patients for whom researchers want to estimate the treatment effect	Adult patients with active definite or active probable RA (as judged by the treating clinician), who are under the care of a consented health care professional.
Treatment conditions	Different intervention strategies being compared in the treatment effect definition	Intervention: Rheumatology clinics featuring IST, where patient symptom data collected daily via the REMORA app are visible to clinicians. Control: SoC rheumatology clinics - typical rheumatology clinics, unchanged from normal clinical practice, where participants are not given access to the REMORA app
Endpoint	Outcome for each participant that is used in the treatment effect definition	DAS28-CRP at 12 months follow-up (±3 months to allow for typical variation in date setting reflective of normal practice)
Summary Measure	Method used to summarise and compare the endpoint between treatment conditions (eg, risk ratio, odds ratio)	Marginal mean difference between IST and SoC at 12 (±3) months, after accounting for baseline values
Handling of postrandomisation (intercurrent) events	Strategies used to handle each postrandomisation in the treatment effect definition; different strategies could be used for different types of intercurrent events	Failure to register (in intervention condition only): participants included in analysis regardless of registration status Death (both conditions): participants included in analysis up to point of death (anticipating a very low number of deaths in REMORA2) Loss-to-follow-up[2] (both conditions): Missing observations post-lost-to-follow-up out are filled using multiple imputation in the primary outcome analysis Withdrawal from study[3] (both conditions): Missing observations post-withdrawal are imputed using multiple imputation in the primary outcome analysis Poor intervention fidelity (intervention condition only): All participants included in primary outcome irrespective of fidelity. Additional analyses added to consider effect of intervention fidelity on diease activity (in intervention arm only)

Note: The above definitions are for the primary outcome analysis only. Sensitivity analyses described in Section 9.1.1.2 may target different estimands, depending on the assumptions made.

[2] 'Loss-to-follow-up' refers to a circumstance where outcome data has not been collected at a time point without a specified reason, for example where a participant not attended a clinic visit, been discharged from service, moved out of area, etc. 'Withdrawal' refers to participants who actively discontinue the study and notify the research team, or are withdrawn by the research team/chief investigator through an active decision. Different withdrawal classifications are described in Section 8.2.

[3] 'Loss-to-follow-up' refers to a circumstance where outcome data has not been collected at a time point without a specified reason, for example where a participant not attended a clinic visit, been discharged from service, moved out of area, etc. 'Withdrawal' refers to participants who actively discontinue the study and notify the research team, or are withdrawn by the research team/chief investigator through an active decision. Different withdrawal classifications are described in Section 8.2.

6.6.1.4 Reporting

The treatment effect will be reported as the adjusted mean difference between conditions (IST vs. SoC), with a 95% confidence interval (CI). Graphical representations of mean DAS28/CDAI over time, stratified by treatment group, will also be presented.

Dummy tables and figures outlining the content and presentation of outcome reporting is included in the appendix.

6.6.2 Secondary Outcomes

6.6.2.1 Definition

The secondary outcomes of the REMORA trial are designed to evaluate a broad range of effects of the integrated symptom tracking intervention beyond disease activity alone. These include additional disease activity measures, validated patient-reported outcomes, measures of shared decision-making and consultation experience, self-management capacity, illness perception, and indicators of medication use and healthcare resource utilisation. The definitions summaries for each outcome are as follows:

Additional Disease Activity Measures:

- **EULAR response categories**[4,8]: Categorical (ordinal) outcome. Derived from changes in DAS28-CRP scores, patients are categorised as having a Good, Moderate, or No response based on predefined criteria of change from baseline and absolute DAS28 levels. From EHR dataset.
- **Patient Global Assessment scores**: Continuous, single-item 0–10 numerical rating scale capturing the patient's overall perception of disease activity. From both EHR. This is also collected in the web survey, but this version is not considered a secondary outcome of interest.
- **Swollen Joint Count (SJC28)**: Clinician-assessed count (frequency) of swollen joints out of 28 specified joints. A higher score (range 0–28) indicates greater disease activity. From EHR dataset.
- **Tender Joint Count (TJC28)**: Clinician-assessed count (frequency) of joints out of the 28 specified joints that are tender on palpation or pressure, with scores ranging from 0 (no tenderness) to 28 (maximal tenderness), reflecting inflammatory burden and pain perception. From EHR dataset.
- **Routine Assessment of Patient Index Data 3 (RAPID-3)**[9]: A composite score (0–30) based on patient-reported physical function, pain, and patient global estimate. Higher scores indicate greater disease impact. RAPID-3 has been validated for use in RA settings where formal joint counts may not be feasible. From web survey dataset.
- **Rheumatoid Arthritis Impact of Disease (RAID)**[10]: A 0–10 composite score assessing seven domains (pain, function, fatigue, sleep, physical and emotional well-being, and coping). Higher scores indicate worse disease impact. From web survey dataset.

Patient-Reported Outcomes:

- **Work productivity and activity impairment (WPAI-RA)**[11]: Evaluates the impact of RA on work productivity and daily activities over the past seven days. Domains include absenteeism, presenteeism, overall work productivity loss, and activity impairment. Continuous out-

come, reported as a percentage (0–100%), with higher scores indicating worsening impact of disease on productivity/activities. From web survey dataset.

- **Quality of life, via the EuroQol 5 dimensions, 5 levels questionnaire (EQ-5D-5L)**[12]: Assesses health-related quality of life across five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each rated on five levels. An index score (utility) is generated, treated as a continious variable, typically ranging from <0 (worse than death) to 1 (perfect health). From web survey dataset.
- **Health Assessment Questionnaire (HAQ)**[13]: Measures physical functioning across eight domains (e.g., dressing, walking, hygiene), scored from 0 (no disability) to 3 (severe disability). From web survey dataset.

Shared Decision-Making and Consultation Experience:

- **CollaboRATE**[14]: A three-item measure of the patient's perception of shared decision-making during a consultation, each item scored 0–9. Higher scores reflect better shared decision-making. From web survey dataset.

Self-Management and Illness Perception:

- **Capabilities, Opportunities, Motivations for Behaviour change (COM-B)**[15]: A brief validated tool assessing self-reported readiness for health-related behaviour change. Responses are recorded on a Likert scale across components representing the COM-B modes. COM-B produces scores on 3 components: a Capabilites Score, Opportunities Score, and Motivations Score. From web survey dataset.
- **Brief Illness Perception Questionnaire (IPQ-B)**[16]: A 9-item questionnaire assessing emotional and cognitive representations of illness (e.g., consequences, timeline, control, identity), scored individually or summarised into a single score. From web survey dataset.

Medication Use and Resource Utilisation:

- **Frequency of medication use (by type)**: Extracted from EHRs, includes drug type, start and stop dates, and number of medications prescribed throughout follow-up. From EHR dataset.
- **Cumulative healthcare resource utilisation**: Captures hospital admissions, outpatient visits, investigations, and other RA-related healthcare usage derived from EHR and web survey datasets.

6.6.2.2 Reporting

Descriptive statistics of secondary outcomes will be summarised across both time points of interest (baseline and final visit), using statistics appropriate to each outcome: frequency-type outcomes (e.g. synthetic DMARD frequency, EULAR response status) will be described with frequencies and proportions; continuous outcomes (e.g. RAPID-3 score, SJC28) will be summarised using means and SDs if normally distributed, and median and interquartile ranges (IQRs) if substantial deviation from normality is apparent upon visual inspection of histograms. Visualisation methods, including bar charts for proportions and line plots for trajectories of continuous outcomes over time, will accompany the tabular results.

EULAR response status of participants will be described both at baseline and follow-up, with visualisations of the within-person changes in status will be produced (e.g. Sankey diagram). See the appendix for further details on how outcomes will be reported and presented.

6.6.3 Tertiary/Exploratory Outcomes

6.6.3.1 Definition

Tertiary and exploratory outcomes aim to explore broader impacts of the intervention and mechanisms underlying observed changes. These include: - Representativeness and diversity of study participants: Comparisons of participant demographics, comorbidities, medication use, and resource utilisation with shared health records from the wider RA population. - Longitudinal symptom trajectories: Patterns of change in disease activity (DAS28-CRP), RAPID-3, RAID, SJC28, TJC28, and EQ-5D-5L scores over time, analysed across the 12-month observation window. - Engagement metrics: Frequency of app usage, completion rates of symptom-tracking questionnaires, and healthcare professional interactions with the REMORA dashboard.

6.7 Interim Analysis & Early Stopping

6.7.1 Number and timing of interim analysis/es

There are no planned formal interim analyses for REMORA2_SWT, and so no additional adjustments to the sample size to account for this are necessary.

6.7.2 Analysis adjustment(s) due to interim analysis/es

Not applicable - no formal interim analyses planned.

6.7.3 Guidance on early trial closure ("Stopping Rules")

The Programme Steering Committee (PSC) retain the rights to review accruing trial data and the ability to recommend early trial closure independently of the trial team (for example for safety concerns), but these recommendations do not form the basis of any formal planned interim analyses. No formal stopping rules are given, and any early trial closure recommendations from the PSC will be made as a consensus based on a combination of clinical and scientific judgement based on the accrued trial data. The final decision will be made by the Sponsor. The Sponsor can over-rule the recommendation of the PSC and may stop the trial independently of the recommendation of the PSC, if appropriate.

6.8 Timing of Final Analysis

The final analysis will be conducted after the following conditions are met:

- Completion of Follow-Up: All participants have completed their 12(±3) month (final) follow-up assessment, or are otherwise accounted for (e.g., lost to follow-up, withdraw).
- Data Cleaning and Verification: All study data, including primary and secondary outcome measures, have been entered into the database, cleaned, and verified for accuracy and completeness.
- Database Lock: The trial database will be locked after resolving all outstanding data queries and verifying consistency across data sources (e.g., EHRs, REMORA web survey).
- The final analysis is expected to begin within 3 months of the last participant's final follow-up visit and following confirmed database lock.

6.9 Timing of Outcome Assessments

Primary Outcome: The DAS28-CRP will be assessed at baseline and at 12(±3) months.

Secondary Outcomes: Other measures of disease activity, patient-reported outcomes, and shared decision-making metrics will be collected at baseline and 12 months. Timing of qualitative secondary outcomes will be collected during study visits but are not described in this SAP as they do not contribute to the quantitative analyses.

Tertiary Outcomes: Longitudinal outcomes (engagement metrics and data on symptoms via the REMORA app) will be collected continuously throughout follow-up, with analyses conducted after study completion. Additional web surveys collecting the secondary outcomes listed in Section 6.6.2 are collected at 3, 6, and 9 months in addition to baseline and 12 months, and so longitudinal analyses will consider change in these additional visits over time. Data from additional interim clinic visits (between baseline and 12 (±3) months) are also collected, but will not be included in the primary or secondary outcome analyses: as these additional visits make models more complex and requiring additional assumptions about trajectories, the additional time points in the secondary outcomes will be considered tertiary/exploratory analyses. Qualitative tertiary outcomes (e.g. clinician feedback) will be collected during study visits, but are not described in this SAP as they do not contribute to the quantitative analyses.

7 Statistical Principles

7.1 Confidence Intervals and *P* Values

7.1.1 Level of Statistical Significance

The alpha level for this trial used to assess statistical significance ("significance level") is set at 0.05 (5%), therefore p values below this threshold will be considered *statistically significant*.

7.1.2 Confidence Intervals

Consistent with the alpha level, all effect estimates will be presented as 95% CIs, calculated assuming the normal distribution (z-score-derived 95% CIs). CIs excluding a null effect (typically zero for continuous outcomes, and one for ratio-type outcomes e.g. odds-ratios) will be considered *statistically significant*.

7.1.3 Adjustment(s) for multiplicity and/or other error control methods

No multiplicity adjustment is applied to the primary outcome (DAS28-CRP), as it is a single hypothesis test.

For secondary outcomes, unadjusted p-values will also be reported, to allow readers to manually calculate their own preferred type-I error-rate adjustment if desired, in accordance with recommendations from Hooper et al. (2025)[17].

7.2 Intervention Fidelity: Adherence, Compliance, Contamination

Intervention fidelity in the REMORA-SWT has two principal aspects: continuing use of the REMORA app by participants ('participant engagement') and use of the REMORA dashboard during consultations ('clinician engagement').

With respect to clinician engagement, number of times the REMORA dashboard is accessed, as well as total time viewing the dashboard are collected. Both of these are quantitative outcomes, but make strong assumptions that merely opening the app means that it is being looked at and/or used. An additional self-report questionnaire is given to clinicians, asking them to rate how much they used the dashboard, and how useful they found it. In combination, these outcomes should give an indiciation of the level of clinician engagement throughout the study.

Regarding participant engagement, reminder notifications are provided to participants via the REMORA app to help promote completion of the daily, weekly, and monthly questionnaires. Daily alarms for unanswered questionnaires will be triggered every day at 18:30, weekly alarms will be triggered every 7 days at 15:30 and monthly alarms will be triggered every 28 days at 12:00. If a participant completes and submits a questionnaire between the time the question set becomes available (e.g., midnight for a daily question) and the scheduled alarm, the notification will not alert.

The team also send email reminders when engagement falls below 50% for the duration of tracking (after an initial run-in period). These are sent monthly, up to a maximum of 5 reminders within the first 12 months.

A specific level of engagement is not defined *a priori* and there is no set 'target' for engagement for participants (e.g. a set number of days with symptom data defining adequate use of the app). Participants are instead encouraged to engage with the app as much as possible. The engagement level is explicitly examined in (exploratory) analyses as a potential mediator of the treatment effect.

Quantifying app engagement is important but considered an exploratory outcome given that the literature on defining what constitutes 'sufficient' engagement in studies using data collection apps to inform shared decision making is scant. We use two types of methodology to define app engagement by participants, and will report both as findings from the trial. The first, simpler, method ("completeness of symptom tracking") is defined as the proportion of study days on which a participant has recorded *at least one* symptom in the REMORA app, out of the total possible days that they were in the study. Cutoffs for high, medium, and low engagement are defined as follows:

- 'High engagement': >50% of all possible days
- 'Medium engagement': 25–50% of all possible days
- 'Low/No engagement': <25% of all possible days

A second, more complex definition of engagement is based on previous methodology used by members of the REMORA2 research team[18] examining time-varying engagement patterns in a different smartphone study. This methodology makes use of Hidden Markov Models to derive assumed latent 'states' of engagement, based on the observed number of study days on which a participant has recorded *at least one* symptom in the REMORA app. This is an exploratory method, and details of the modelling process will be included in the relevant publications.

The engagement level of participants (continuing use of the REMORA app) and clinicians (use of the REMORA dashboard during consultations) is explicitly examined in (exploratory) analyses as a potential mediator of the treatment effect, through estimation of treatment effects at different hypothetical levels of participant engagement (see Section 9.1.1).

Contamination risk is negligible, as participants in the control condition do not receive access to the REMORA app. Potential contamination (e.g., clinicians modifying their approach to consultations for SoC participants based on experiences in the IST condition, albeit without access to tracked symptoms) will be monitored via qualitative interviews and clinic observations, with findings reported descriptively.

7.2.1 Definition of fidelity of the intervention and how this is assessed including extent of exposure

7.2.2 Fidelity Reporting

The following participant engagement metrics will be reported:

- A dot plot visualizing engagement patterns during study period.
- The frequencies and proportions of participant engagement, using both the simple and complex definitions of engagement (see Section 7.2).
- Mean number (and SD) of completed daily symptoms by IST participants.

The following clinician engagement metrics will be reported:

- Mean number (and SD) of REMORA dashboard views per clinician, per participant, out of the total number of clinic visits - split by study site and region (Greater Manchester/North West London).

7.2.3 Protocol Deviations

Protocol deviations are collected in a study log throughout trial conduct. Protocol deviations can occur in the following categories: Deviations for consent, eligibility, timing/visit checks, endpoint da-

ta and Serious Adverse Event (SAE) reporting. When they occur, deviations are reviewed by the study team (programme manager, chief investigator plus relevant other staff) and are classified into one of four categories: Serious breach, Major deviation, Minor deviation or No non-compliance (i.e. not considered a deviation, after review).

Serious breaches are defined as any departures from the study protocol that impact *to a significant degree* the safety, or physical or mental integrity, of trial participants; or the scientific value of the trial. Major deviations are defined as any departures from the study protocol that *may* impact the safety, or physical or mental integrity, of trial participants; or the scientific value of the trial. Minor deviations are defined as any departures from the study protocol that *do not* impact the safety, or physical or mental integrity, of trial participants; or the scientific value of the trial.

Failure to register on the REMORA app is not considered a protocol deviation due to the anticipated potential high frequency of these occurring, and because collecting information about failure to register is integral to answering the primary research question (effectiveness of the app when implemented in a real-world setting). Participants who fail to register on the REMORA2 app will be included in the primary outcome analysis model.

With respect to the analysis, protocol deviations will be included/excluded from the analysis population as described in Table 3:

Table 3: How Participants with Protocol Deviations will be Included/Excluded from the Analysis Sample

Protocol Deviation with Respect to:	How to Include in Analysis Sample
Consent process *(e.g. participant was wrongly consented)*	If major deviation, exclude in analysis If minor devia include from analysis
Study eligibility *(e.g. participant does not meet criteria, but is recruited regardless)*	If major deviation, exclude in analysis If minor devia include from analysis
Time point *(e.g. clinic visit occurs outside of study window)*	Include in the analysis in all cases
End point (outcome) *(e.g. component missing from DAS-28 score)*	Include the participant, treat observation as missing and handle as per Section 9.2
SAE occurs	Include in all cases
Failure to register with REMORA2 app	Include in all cases (not considered a protocol devia but included for reference)

7.2.4 Protocol Deviation Reporting

Frequencies and types of protocol deviations will be summarised by study site and condition (SoC/IST).

7.3 Analysis Populations

The primary outcome analysis uses and intention to treat (ITT) analysis approach, therefore the analysis population is: all eligible participants, included in the analysis model according to their allocated intervention, regardless of intervention fidelity (adherence) level, and according to the protocol deviation handling specification described in Table 3 and withdrawal/lost-to-follow-up handling specified in Section 8.2.

8 Trial Population

8.1 Screening & Eligibility Information

Eligibility criteria are detailed in the protocol[3] and include adults (aged ≥18) with active RA, under the care of a consented healthcare professional, with access to an app-compatible device and ability to follow app setup instructions. Screening and eligibility data will be reported, including numbers screened, eligible, and recruited.

All eligibility information described above will be broken down by study site.

8.2 Withdrawal and Lost to Follow-Up ("Attrition")

Participants may leave REMORA2-SWT for a range of different reasons, as defined below:

- **Withdrawal**: A participant may choose to withdraw from different aspects of the study:
  - *Withdrawal from Active Participation*: The participant requests to discontinue all active study procedures (e.g., using the app, completing web surveys).
  - *Withdrawal from All Data Collection*: The participant withdraws consent for all aspects of the study, and all data collection, including from the EHR, will cease from the time of withdrawal onwards.
  - *Total Withdrawal*: The participant withdraws consent for all aspects of the study, and all data collection up to this point. All data collection, including from the EHR will cease from the time of withdrawal onwards, and all currently collected data for that participant up to that time point will also be removed.
  - *Investigator-Initiated Withdrawal*: A participant may be withdrawn at the discretion of the investigator. This typically occurs if a participant is retrospectively found to be ineligible after study commencement (e.g., due to a change in diagnosis or not meeting device eligibility criteria).
- **Non-Registration**: Defined as IST participants who consent, but fail to register for the REMORA app within 30 days of recruitment. We anticipate a 30% non-registration rate (see Section 6.4). These participants will still be included in ITT analyses.
- **Lost to Follow-Up**: Defined as participants missing their 12-month (±3 months) DAS28-CRP assessment without explicit withdrawal, for example where a participant not attended a clinic visit, been discharged from service, moved out of area, etc. Participants that cease to engage with the REMORA app will not be considered lost to follow-up, but instead their engagement will be considered as per definitions described in Section 7.2. Participants lost to follow-up will still be included in ITT analyses.

'Withdrawal' refers to participants who actively discontinue the study and notify the research team, or are withdrawn by the research team/chief investigator through an active decision.

How participants' data is handled post-withdrawal differs prior to protocol substantial amendment SA03:

- Before SA03: If a participant withdraws/is withdrawn, then all data collection from this time point forward ceases, and 'exposure' to the REMORA2 app/IST clinics stops, reverting to SoC clinics.
- After SA03: If a participant withdraws/is withdrawn, then all EHR and REMORA2 app data collection can continue, however 'exposure' to the IST clinics stops, reverting to SoC clinics.

In alignment with the ITT desired in the primary outcome analysis, post-withdrawal observations from participants who withdraw after SA03 was implemented will continue to be included in the primary outcome analysis model.

The study database has a tag to differentiate between visits which occurred before/after SA03 implementation, to differentiate between when participants are/are not exposted to IST clinics in the IST arm.

8.2.1 Reporting

A CONSORT-SW-CRT diagram (as described in the CONSORT SWT extension[19]) will detail screening, recruitment, non-registration, withdrawals, and lost-to-follow-up by cluster and time step.

A table will summarise reasons for withdrawal and lost-to-follow-up by condition (IST vs. SoC).

8.3 Baseline Characteristics

All participant baseline characteristics will be summarised as means with SDs for continuous, normally distributed variables and medians with interquartile ranges (IQRs) for continuous variables exhibiting skew in their distribution. Categorical variables (both binary and ordinal) will be summarised as frequencies with percentages, broken down by item category.

The following variables will be summarised:

- Age at baseline visit
- Gender
- Ethnicity (all 16 ethnicity categories used in the CRF; white vs. non-white)
- Time since RA diagnosis (in years)
- Probable / definite RA (as judged by the treating clinician)
- Number of comorbidities
- Smoking status (current/former/never)
- BMI
- Index of Multiple Deprivation Score (IMD)

Dummy tables outlining how this baseline information will be presented is included in the appendix.

9 Analysis

9.1 Analysis Methods

9.1.1 Primary Outcome Analysis

9.1.1.1 Method (including any covariate adjustment and assumptions required to consider)

The primary analysis of REMORA_SWT will be conducted using a generalised linear mixed-effects regression model (GLMM), assuming a Gaussian distribution and an identity link function, to compare disease activity (via the DAS28-CRP) in IST vs. SoC. The outcome variable is the DAS28-CRP score at 12 months (±3 months), treated as a continuous measure of disease activity. The model will include fixed effects for baseline DAS28-CRP score to adjust for initial disease severity, study period (implemented as a dummy-coded categorical variable [1-9], which accounts for period effects (time trends) whilst making no assumptions about the shape (linearity) of such effects - i.e. assuming a piecewise secular trend), treatment condition (IST vs. SoC as a categorical variable), and the following covariates: age at baseline visit (in years), gender (male; female; non-binary/transgender/other; prefer not to say), ethnicity (white; non-white), time since diagnosis (in years), whether the diagnosis was probable or definite RA, number of comorbidities, smoking status (current/former/never), BMI (as a continuous variable), and IMD score of the postcode of the participant (as a continuous variable) to broadly control for socioeconomic factors from the participants. These covariates were selected based on clinical relevance and expertise from the study team. Given the number of covariates, should the model have issues with convergence, then a model dropping the ethnicity, time since diagnosis, probable or definite RA, number of comorbidities, and smoking status terms will be used instead. The exact model terms used will be reported in the relevant publication.

Observations of categorical variables with optional completion (i.e. a 'prefer not to say' option) will be considered as missing values, anticipating that these will occur at a low frequency.

Random intercepts will be included in the model to account for within-site correlation, and will use robust (sandwich) standard errors[20] to account for potential misspecification of the covariance structure and heteroscedasticity. This modelling approach reflects the trial's stepped-wedge structure and accounts for both temporal and hierarchical clustering[21,22]. Cluster-by-period interaction effects will also be included. As the period variable is dummy coded, this would result in a large number of additional terms being included in the model, and so, should the model fail to converge with this effect, initially, the interaction effect terms will be omitted from the model, and if still failing to converge, the period effect will be reconsidered as a linear term, rather than dummy-coded terms.

Any model convergence issues and steps taken to counteract this will be described in the corresponding results paper and trial report, with full details of the final imputation model used, and the steps taken to achieve this model

The model will apply robust (sandwich) standard errors[20] to account for potential misspecification of the covariance structure and heteroscedasticity, particularly given the small number of clusters and the likelihood of unbalanced data across sites. The linearity of continuous predictors and the normality and homoscedasticity of residuals will be assessed during model diagnostics.

This model specification reflects current best practice for SW-CRT design, including the recommendations from Li et al. (2020)[22] for mixed models in cluster randomised settings. Treatment effects will be reported as adjusted mean differences with 95% CIs and corresponding p-values for statisti-

cal inference. The use of this modelling approach ensures that both the clustering of participants within sites and the temporal structure of the trial are appropriately accounted for, providing robust and unbiased estimates of the intervention effect.

The IMD score for each participant is calculated via the postcode lookup tool (https://imd-by-postcode.opendatacommunities.org/imd/2019), matched to the postcode provided by each participant in the baseline web survey CRF.

9.1.1.2 Sensitivity Analysis/es

9.1.1.2.1 Effect of Considering Time as a Continuous Variable (Rather than Categorical)

Hooper et al.[23] suggest that, for the trial design used by REMORA2-SWT, time effects should be considered as a continuous outcome, rather than as a categorical variable. We therefore intend to repeat the primary outcome analysis using time as a continuous (linear) variable to investigate the effect of considering time in this manner. Interaction effects will also be modified to reflect this different implementation of the time effect.

9.1.1.2.2 Effect of Limited Visits on the Treatment Effect Difference

The intervention (IST) has greatest potential to affect the final outcome if the REMORA dashboard is reviewed after baseline once tracked symptom data is available, with treatment decisions at that visit then affecting disease activity scores at subsequent visits. Due to increased service pressure in the UK NHS and resource limitations, there is a risk that participants may have fewer clinical visits than desirable for patients with active disease. Participants will have only one baseline and one follow-up visit during the study, limiting the opportunity for treatment changes informed by the intervention.

An additional sensitivity analysis which only includes participants who have three or more clinical contacts (applied to both conditions), will be undertaken to examine whether the effect of the REMORA system is more pronounced.

9.1.1.2.3 Effect of Missingness in the Primary Outcome, in the Primary Outcome Analysis

As outlined in more detail in Section 9.2, multiple imputation by chained equations (MICE) will be used in the primary analysis to generate plausible values for missing DAS28-CRP data, assuming that data is missing at random (MAR) (using Rubin's definitions of missingness mechanisms[24]). We assume that, given that multiple disease activity measures are available it is reasonable to assume that plausible values for missing observations of the primary outcome can be generated.

Sensitivity analyses will be conducted to investigate the impact of different missingness assumptions (missing completely at random [MCAR] using a complete case analysis, and missing not-at-random [MNAR], considering different values of the missing observations).

9.1.1.2.4 Effect of Assumptions Around Engagement in the Primary Outcome Analysis

To assess the robustness of the primary outcome analysis (comparing DAS28-CRP between conditions at 12(±3) months, after accounting for covariates), the following sensitivity analyses will be conducted:

1. Complete Case Analysis: Including only participants with complete data (in all analysis variables – both DAS28-CRP and other model covariates) at baseline and 12 months to evaluate the impact of missing data.

2. Per-Protocol Analysis: Includes only participants who had >50% engagement ('high engagement as per definition in Section 7.2), plus at least one REMORA dashboard use per clinic visit.

3. Best-Case and Worst-Case Engagement Scenarios: Given the anticipated variability in engagement with integrated symptom tracking among participants and clinicians, we will perform sensitivity analyses to explore the following hypothetical extreme scenarios:

    - Best-Case Scenario: Marginal means will be extracted to quantify the treatment effect under the an assumption of perfect (100%) engagement levels across all participants, and assuming clinicians use the REMORA dashboard at every study visit, which will assess the *maximum* potential benefit of the intervention.

    - Worst-Case Scenario: Conversely, marginal means will also be extracted under a model assuming low engagement levels of <25% ('Low/No engagement' as per definition in Section 7.2), and also assuming that the REMORA dashboard was not used at any visit by the clinician.

9.1.1.2.5 Effect of Partially-Reported Laboratory Values

In clinical practice, CRP and ESR values are sometimes reported partially by trust laboratories. For example, rather than the exact values, they are sometimes reported as "<4" or "<5" on the report forms. For the primary outcome analysis, these values will be assumed to be the middle of the range between the upper and lower limits provided (e.g. if "<4" is reported, the score will be calculated using a value of 2; if "<5" is reported, the score will be calculated using a value of 2.5, etc.). Sensitivity analyses will investigate the effect of altering this value to the upper/lower limits of the range.

9.1.1.3 Subgroup Analysis/es

The REMORA2_SWT trial has no planned subgroup analyses for the primary outcome analysis.

9.1.2 Secondary Outcome Analysis

9.1.2.1 Method (including any covariate adjustment and assumptions required to consider)

The analysis of secondary outcomes in REMORA2_SWT will be conducted using statistical models appropriate to the structure and distribution of each outcome, while accounting for the trial design.

Much like the primary outcome analysis model, continuous secondary outcomes will be analysed using a generalised linear mixed-effects regression model (GLMM), assuming a Gaussian distribution and an identity link function, to compare the continuous outcome between the IST and the SoC condition. The model will include fixed effects for baseline value of the continuous outcome taken from the baseline visit (either baseline web survey or baseline clinic visit, as appropriate to the outcome - defined in the list below), study period (implemented as a dummy-coded categorical variable [1-9]), treatment condition (IST vs. SoC as a categorical variable), and the following covariates: age at baseline visit (in years), gender (male; female; non-binary/transgender/other; prefer not to say), ethnicity (white; non-white), time since diagnosis (in years), whether the diagnosis was probable or def-

inite RA [probable/definite], number of comorbidities, smoking status (current/former/never), BMI (as a continuous variable), and IMD score of the postcode of the participant (as a continuous variable) to broadly control for socioeconomic factors from the participants. These covariates were selected based on clinical relevance and expertise from the study team. Secondary continuous outcomes to be analysed using this model are:

- Routine Assessment of Patient Index Data 3 (RAPID-3) score (from web survey)
- Patient global assessment score (from web survey)
- Rheumatoid Arthritis Impact of Disease (RAID) score (from web survey)
- Tender Joint Count for 28 Joints (TJC28) (from web survey)
- Swollen Joint Count for 28 Joints (SJC28) (from web survey)
- EQ-5D-5L index score (from web survey)
- Work Productivity and Activity Impairment Questionnaire (WPAI-RA) score (from web survey)
- HAQ disability index (from web survey)
- CollaboRATE score (from web survey)
- Patient Enablement Instrument (PEI) total score (from web survey)
- The three Capabilities-Opportunities-Motivations-Behaviour (COM-B) subscale scores (from web survey)
- Brief Illness Perceptions Questionnaire (IPQ-B) (from web survey)

Categorical outcomes will be analysed using a generalised linear mixed-effects regression model (GLMM), assuming a binomial distribution and a logit-link function, to compare the categorical outcome between the IST and the SoC condition. The model will include fixed effects for study period (implemented as a dummy-coded categorical variable [1-9]), treatment condition (IST vs. SoC as a categorical variable), and the following covariates: age at baseline visit, gender, ethnicity (coded into two categories: white vs. non-white), time since diagnosis (in years), whether the diagnosis was probable or definite RA, number of comorbidities, smoking status (current/former/never), BMI (as a continuous variable), and IMD score of the postcode of the participant (as a continuous variable) to broadly control for socioeconomic factors from the participants. No baseline value of the categorical outcomes will be included in this model, meaning that this model estimate is of the marginal probability of response to treatment, after controlling for the other covariates specified. These covariates were selected based on clinical relevance and expertise from the study team. Secondary categorical outcomes to be analysed using this model are:

- EULAR response categories
- Medication use: Proportion of participants currently taking multiple (>1) conventional DMARDs at 12 months
- Medication use: Proportion of participants switching medication category (from 1 conventional DMARDs/>1 conventional DMARD/Biologic DMARD monotherapy/targeted synthetic DMARD monotherapy/combined conventional + synthetic DMARD)
- Medication use: Proportion of participants currently taking biologic DMARDs at 12 months
- Medication use: Proportion of participants currently taking targeted synthetic DMARDs at 12 months

Random intercepts will be included in all models to account for within-site correlation, and will use robust (sandwich) standard errors[20] to account for potential misspecification of the covariance structure and heteroscedasticity. This modelling approach reflects the trial's stepped-wedge structure and accounts for both temporal and hierarchical clustering[21,22]. As with the primary outcome analysis model, cluster-by-period interaction effects will be included. As the period variable is dummy coded, this would result in a large number of additional terms being included in the model, and so, should the model fail to converge with this effect, initially, the interaction effect terms will be omitted from the model, and if still failing to converge, the period effect will be reconsidered as a

linear term, rather than dummy-coded terms. Any model convergence issues and steps taken to counteract this will be described in the corresponding results paper and trial report.

Should any model have issues with convergence, then a model dropping the ethnicity, time since diagnosis, probable or definite RA, number of comorbidities, and smoking status terms will be used instead. The exact model terms used will be reported in the relevant publication.

In cases of probable skewed distributions (e.g., healthcare resource use, work impairment), alternative link functions or appropriate distributions (e.g., gamma or log-normal) will be explored within the mixed-effects framework.

9.1.2.2 Sensitivity Analysis/es

The REMORA2_SWT trial has no planned sensitivity analyses for the secondary outcome analyses.

9.1.2.3 Subgroup Analysis/es

The REMORA2_SWT trial has no planned subgroup analyses for the secondary outcome analyses.

9.1.3 Tertiary/Exploratory Outcome Analysis

Exploratory analyses will investigate broader impacts and potential mechanisms behind observed treatment effects. Planned exploratory analyses include the following:

1. Representativeness: Comparisons between trial participants and the broader RA population from Greater Manchester and North West London will be made using descriptive statistics to assess generalisability. We anticipate the data for the wider RA population will come from the two regional sub-national Secure Data Environments.

2. Longitudinal Trajectories: Using mixed-effects models to examine changes in outcomes such as DAS28-CRP, RAPID-3, and EQ-5D-5L over time.

3. Engagement Metrics: Descriptive statistics with appropriate visual representations will summarise app usage rates and healthcare professional interactions with the dashboard through time.

9.2 Missing Data Considerations

9.2.1 Missing Data Reporting

Summary tables will describe the frequency and proportion of missing observations for all primary (DAS28-CRP) and secondary outcomes (e.g., EULAR response categories, RAPID-3, EQ-5D-5L, SJC28, TJC28, WPAI-RA, HAQ, CollaboRATE, COM-B, Brief IPQ, medication use, resource use), and alternative disease activity measures (DAS28-ESR, 3-component versions, etc.) by time point, treatment condition (IST vs. SoC), and cluster (Study sites). Missing data patterns will be visualised using heatmaps to illustrate the distribution of missingness across variables and participants, aiding in the identification of systematic patterns.

9.2.2 9.2.2 Missing Data Mechanisms

To enable the planned primary and secondary outcome analysis, multiple imputation by chained equations (MICE) will be used to generate 50 imputed datasets (i.e. m = 50), assuming that missing

observations are MAR. This number is considered sufficient to produce stable estimates and ensure reproducible results across a wide range of missing data).

The imputation model is intended to estimate plausible values for all primary and secondary outcomes, making one 'master dataset' that can be used consistently across all primary and secondary analyses. This ensures consistency in the data generating mechanisms across the study, and also maximises use of what we assume will be relatively highly-correlated auxiliary variables to estimate plausible values for missing observations of variables included in each analysis (i.e. following an 'inclusive strategy').

The MICE model will include the following variables, taken from the EHR and web survey datasets:

- Study site
- Treatment condition (IST/SoC)
- Time since study commencement until follow-up appointment
- Disease activity measures:
    - DAS28-CRP score
    - DAS28-ESR score
    - DAS28-CRP [3-component] score
    - DAS28-ESR [3-component] score
    - CDAI score
    - Clinic-calculated DAS28 Score
    - EULAR response categories
    - All components constituting the above
- Age at baseline visit
- Gender
- Ethnicity (all 16 ethnicity categories used in the CRF; white vs. non-white)
- Time since RA diagnosis (in years)
- Probable / definite RA (as judged by the treating clinician)
- Number of comorbidities
- Smoking status (current/former/never)
- BMI
- Index of Multiple Deprivation Score (IMD)
- Study site
- Questionnaires:
    - RAPID-3 score
    - RAID score
    - WPAI-RA score
    - EQ-5D-5K index score
    - HAQ score
    - CollaboRATE score
    - The three COM-B indices (Capabilities score, Opportunities score, Motivations score)
    - Brief Illness Perception Questionnaire (IPQ-B)

Should any of the above variables have greater than one-third (33.3%) of the observations missing, they will be omitted from the model, as they will be considered to have too much missing information to be beneficial to the modelling.

We will include no interaction terms between any of the included variables, a trade-off that attempts to keep MI model complexity lower, and giving a better chance of successful model convergence,

given the large number of included variables. We assume that using a large number of interrelated auxiliary variables, which are observed longitudinally, is sufficient to provide good estimates of plausible values without including additional interaction terms to constrain more complex relationships between particular variables.

We will initially attempt to include all observed study visits, however if a large number of interim observations (those not at the baseline or final visits) are missing, these will be omitted from the model in order to increase chance of successful model convergence.

Should model convergence not be able to be achieved using MICE using the above model specification, multiple, simpler, more parsimonious models using only the baseline and final visit observations and/or smaller subsets of the above covariates (e.g. disease activity variables only, participant demographics only), and/or will be created and appended to create the full imputed datasets, accepting that this is suboptimal, and modelling all relevant variables in the same model is more desirable. The full details of the imputation model will be described, including issues with convergence and the unsuccessful models attempted will be included in the trial report.

Each imputed dataset will be analysed using the models specified in Section 9. Results will be pooled using Rubin's rules[24,25] and Wald testing to produce combined estimates, 95% CIs, and p-values, accounting for within- and between-imputation variability.

In the sensitivity analyses for the primary and secondary outcomes, we will also examine estimates using:

- **Complete Case Analysis**, using only participants with complete sets of observations for each analysis model, assuming that missing observations are MCAR.
- **MNAR Exploration**: Additional mixed-effects models will be run for the primary outcome analysis, assuming missing that DAS28-CRP scores are systematically higher by adding a constant offset of 1.2 points [i.e. 1 MCID], to imputed values for participants with missing observations of the primary outcome to test sensitivity of the primary outcome analysis estimates assuming that missing observations of the primary outcome are MNAR, and due to participants leaving the study due to increased disease activity.

9.3 Safety/Harms Analysis

SAEs will be monitored and reported per the protocol. Given the non-invasive nature of the REMORA app, significant safety concerns are not anticipated.

- **Coding and Categorisation**: SAEs will be coded using MedDRA terminology, categorised by severity (mild, moderate, severe), expectedness (expected vs. unexpected per protocol), and causality (unrelated, possibly related, related to app use).
- **Analysis**: Incidence rates of SAEs will be summarised by condition (IST vs. SoC) and cluster. Descriptive statistics (frequencies, percentages) will be reported, with narrative descriptions of SAEs. No formal hypothesis testing is planned due to expected low event rates.
- **Reporting**: A table will summarise SAE frequencies by category, severity, and condition. SAEs will be listed individually with details (e.g., outcome, causality).

9.4 Statistical Software

All statistical analyses will be conducted using R (version 4.5.0 or later). The exact R version and package versions used will be documented in the final analysis report to ensure reproducibility.

10 Data Sharing/Open Research Practices

The REMORA Project (which encompasses the REMORA feasibility study, and the REMORA2-SWT) has an Open Science Framework Repository, located at http://osf.io/sj9ha/. Here, the trial team will store:

- General information about the project, including links to any relevant university pages
- Links to trial registry entries about the project
- Approved versions of the trial protocol
- Approved versions of the statistical analysis plan
- Links to the published trial protocol
- Consent forms and participant information sheets
- Healthcare professional training materials
- Code used to generate the randomisation scheme
- Data preparation (cleaning) code used to prepare the dataset(s) for the primary and secondary analyses (including scripts used to score any derived variables e.g. questionnaire scoring)
- Data analysis code for the primary and secondary analyses
- Prepared datasets (once fully anonymised and deidentified), if possible within IP and other governance agreements REMORA2_SWT complies with.
- Main trial reports
- Publications arising from the trial
- Conference abstracts, posters, and slides from oral presentations arising from the trial

Access to identifiable datasets will be considered upon arrangement of a suitable data sharing agreement - contact the trial team email address (REMORA2@manchester.ac.uk) for further details.

10.1 Dataset Availability

The REMORA trial is committed to promoting transparency and facilitating further research through comprehensive data sharing and open research practices. If possible, synthetic, de-identified versions of the trial analysis datasets will be made publicly available following completion of the study. If this is not possible due to resource constraints and/or due to the need to comply with IP or other governance or data protection agreements, at a minimum, data dictionaries will be shared, describing the full structure of datasets used in the analyses. Researchers can access these datasets/data dictionaries through our OSF trial repository.

For access to original participant-level data intended for ancillary studies, researchers are required to submit a formal request to the trial team. Such requests will be evaluated on a case-by-case basis by the trial's co-principal investigators. This process ensures that data sharing aligns with ethical standards and respects participant consent. To initiate a request, please use the trial team email address.

10.2 Analysis Code Availability

To support reproducibility and additional analyses, the complete statistical analysis code utilised in the REMORA2_SWT trial will be made publicly accessible. This includes all scripts and functions developed for data processing, analysis, and visualisation. The code will be available in our OSF trial repository.

By providing access to our protocol, SAP, synthetic datasets/data dictonaries, and analysis code, we aim to promote an open research environment that encourages collaboration, verification of findings, and the advancement of knowledge in RA care.

11 References

1. Andrew J. Copas, James J. Lewis, Jennifer A. Thompson, Calum Davey, Gianluca Baio, James R. Hargreaves. Designing a stepped wedge trial: Three main designs, carry-over effects and randomisation approaches. *Trials*. 2015;16(1):352. doi:10.1186/s13063-015-0842-7

2. Karla Hemming, Monica Taljaard. Reflection on modern methods: When is a stepped-wedge cluster randomized trial a good study design choice? *International Journal of Epidemiology*. 2020;49(3):1043-1052. doi:10.1093/ije/dyaa077

3. Sabine N. Van Der Veer, Deb Griffiths-Jones, Matthew Parkes, Katie L. Druce, Paul Amlani-Hatcher, Christopher J. Armitage, Nicholas Bansback, Peter Bower, Dawn Dowding, Benjamin Ellis, Jill Firth, Sean Gavan, Elaine Mackey, Caroline Sanders, Charlotte A. Sharp, Karen Staniland, William G. Dixon, the REMORA team. Remote monitoring of rheumatoid arthritis (REMORA): Study protocol for a stepped wedge cluster randomized trial and process evaluation of an integrated symptom tracking intervention. *Trials*. 2024;25(1):683. doi:10.1186/s13063-024-08497-9

4. G Wells, J-C Becker, J Teng, M Dougados, M Schiff, J Smolen, D Aletaha, P L C M Van Riel. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. *Annals of the Rheumatic Diseases*. 2009;68(6):954-960. doi:10.1136/ard.2007.084459

5. Daniel Aletaha, Valerie Pk Nell, Tanja Stamm, Martin Uffmann, Stephan Pflugbeil, Klaus Machold, Josef S Smolen. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: Validation of a clinical activity score. *Arthritis Research & Therapy*. 2005;7(4):R796. doi:10.1186/ar1740

6. Cécile Gaujoux-Viala, Gaël Mouterde, Athan Baillet, Pascal Claudepierre, Bruno Fautrel, Xavier Le Loët, Jean-Francis Maillefert. Evaluating disease activity in rheumatoid arthritis: Which composite index is best? A systematic literature analysis of studies comparing the psychometric properties of the DAS, DAS28, SDAI and CDAI. *Joint Bone Spine*. 2012;79(2):149-155. doi:10.1016/j.jbspin.2011.04.008

7. Bert Vander Cruyssen, Stijn Van Looy, Bart Wyns, Rene Westhovens, Patrick Durez, Filip Van Den Bosch, Eric M Veys, Herman Mielants, Luc De Clerck, Anne Peretz, Michel Malaise, Leon Verbruggen, Nathan Vastesaeger, Anja Geldhof, Luc Boullart, Filip De Keyser. DAS28 best reflects the physician's clinical judgment of response to infliximab therapy in rheumatoid arthritis patients: Validation of the DAS28 score in patients under infliximab treatment. *Arthritis Research & Therapy*. 2005;7(5):R1063. doi:10.1186/ar1787

8. Isabel Castrejón, Ana M. Ortiz, Esther Toledano, Santos Castañeda, Alberto García-Vadillo, Esther Patiño, Isidoro González-Álvaro. Estimated Cutoff Points for the 28-Joint Disease Activity Score Based on C-reactive Protein in a Longitudinal Register of Early Arthritis. *The Journal of Rheumatology*. 2010;37(7):1439-1443. doi:10.3899/jrheum.091333

9. Theodore Pincus, Yusuf Yazici, Martin J. Bergman. RAPID3, an Index to Assess and Monitor Patients with Rheumatoid Arthritis, Without Formal Joint Counts: Similar Results to DAS28 and CDAI in Clinical Trials and Clinical Care. *Rheumatic Disease Clinics of North America*. 2009;35(4):773-778. doi:10.1016/j.rdc.2009.10.008

10. Karen Holten, Nina Paulshus Sundlisæter, Joseph Sexton, Lena Bugge Nordberg, Till Uhlig, Tore K Kvien, Daniel H Solomon, Espen A Haavardsholm, Siri Lillegraven, Anna-Birgitte Aga. Performance of the Rheumatoid Arthritis Impact of Disease (RAID) score in relation to flares in disease activity. *RMD Open*. 2024;10(1):e003486. doi:10.1136/rmdopen-2023-003486

11. Wei Zhang, Nick Bansback, Annelies Boonen, Adam Young, Amitabh Singh, Aslam H Anis. Validity of the work productivity and activity impairment questionnaire - general health version in patients with rheumatoid arthritis. *Arthritis Research & Therapy*. 2010;12(5):R177. doi:10.1186/ar3141

12. M. Herdman, C. Gudex, A. Lloyd, Mf. Janssen, P. Kind, D. Parkin, G. Bonsel, X. Badia. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x

13. Leann Maska, Jaclyn Anderson, Kaleb Michaud. Measures of functional status and quality of life in rheumatoid arthritis: Health Assessment Questionnaire Disability Index (HAQ), Modified Health Assessment Questionnaire (MHAQ), Multidimensional Health Assessment Questionnaire (MDHAQ), Health Assessment Questionnaire II (HAQ-II), Improved Health

Assessment Questionnaire (Improved HAQ), and Rheumatoid Arthritis Quality of Life (RAQoL). *Arthritis Care & Research*. 2011;63(S11). doi:10.1002/acr.20620

14. Glyn Elwyn, Paul James Barr, Stuart W. Grande, Rachel Thompson, Thom Walsh, Elissa M. Ozanne. Developing CollaboRATE: A fast and frugal patient-reported measure of shared decision making in clinical encounters. *Patient Education and Counseling*. 2013;93(1):102-107. doi:10.1016/j.pec.2013.05.009

15. Chris Keyworth, Tracy Epton, Joanna Goldthorpe, Rachel Calam, Christopher J. Armitage. Acceptability, reliability, and validity of a brief measure of capabilities, opportunities, and motivations ("COM-B"). *British Journal of Health Psychology*. 2020;25(3):474-501. doi:10.1111/bjhp.12417

16. Elizabeth Broadbent, Keith J. Petrie, Jodie Main, John Weinman. The Brief Illness Perception Questionnaire. *Journal of Psychosomatic Research*. 2006;60(6):631-637. doi:10.1016/j.jpsychores.2005.10.020

17. Richard Hooper. To adjust, or not to adjust, for multiple comparisons. *Journal of Clinical Epidemiology*. 2025;180:111688. doi:10.1016/j.jclinepi.2025.111688

18. Katie L Druce, John McBeth, Sabine N Van Der Veer, David A Selby, Bertie Vidgen, Konstantinos Georgatzis, Bruce Hellman, Rashmi Lakshminarayana, Afiqul Chowdhury, David M Schultz, Caroline Sanders, Jamie C Sergeant, William G Dixon. Recruitment and Ongoing Engagement in a UK Smartphone Study Examining the Association Between Weather and Pain: Cohort Study. *JMIR mHealth and uHealth*. 2017;5(11):e168. doi:10.2196/mhealth.8162

19. Karla Hemming, Monica Taljaard, Joanne E. McKenzie, Richard Hooper, Andrew Copas, Jennifer A. Thompson, Mary Dixon-Woods, Adrian Aldcroft, Adelaide Doussau, Michael Grayling, Caroline Kristunas, Cory E. Goldstein, Marion K. Campbell, Alan Girling, Sandra Eldridge, Mike J. Campbell, Richard J. Lilford, Charles Weijer, Andrew B. Forbes, Jeremy M. Grimshaw. Reporting of stepped wedge cluster randomised trials: Extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ (Clinical research ed)*. 2018;363:k1614. doi:10.1136/bmj.k1614

20. Bingkai Wang, Xueqi Wang, Fan Li. How to achieve model-robust inference in stepped wedge trials with model-based methods? *Biometrics*. 2024;80(4):ujae123. doi:10.1093/biomtc/ujae123

21. Michael A. Hussey, James P. Hughes. Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*. 2007;28(2):182-191. doi:10.1016/j.cct.2006.05.007

22. Fan Li, James P Hughes, Karla Hemming, Monica Taljaard, Edward R. Melnick, Patrick J Heagerty. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Statistical Methods in Medical Research*. 2021;30(2):612-639. doi:10.1177/0962280220932962

23. Richard Hooper, Andrew Copas. Stepped wedge trials with continuous recruitment require new ways of thinking. *Journal of Clinical Epidemiology*. 2019;116:161-166. doi:10.1016/j.jclinepi.2019.05.037

24. Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. 1st ed. Wiley; 1987. doi:10.1002/9780470316696

25. Stef van Buuren. *Flexible Imputation of Missing Data*. 0th ed. Taylor & Francis Group; 2012. doi:10.1201/b11826