Mastering Effectiveness Research by Meta-Analysing Trial and Real World Evidence in Immune-mediated inflammatory Diseases (MERMAID)

Protocol V1

Title	Mastering Effectiveness Research by Meta-Analysing Trial and Real World Evidence in Immune-mediated inflammatory Diseases (MERMAID)
Research question & Objectives	This project aims to conduct treatment effectiveness research using data from RCTs and registries, with a goal of generating evidence that supports the optimal use of immune-modifying therapies among people with immune-mediated inflammatory diseases (IMIDs). Four main objectives are: 1) To emulate the trials of treatments for IMIDs using data from the registries in the United Kingdom (UK) and evaluate benchmarking success. 2) To compare subgroup strata based on patient characteristics and identify common effect modifiers in each disease across IMIDs. 3) To conduct transportability analyses by combining datasets obtained from trial repository and IMID registries. 4) To predict treatment effect outcomes of ongoing IMID trials using the common factors identified for a successful benchmarking of a trial.
Protocol version	V1.0
Last update date	11 July 2025
Contributors	**Principal investigator:** Zenas Yiu **Collaborators:** Kimme Hyrich, David McAllister, Mia Rodziewicz, Sarah Dyball, Jimmy Limdi, Ian Bruce, Richard Warren, Evan Kontopantelis, Stella Kim, Binh Phan, Georgia Reeves
Study registration	**Site:** University of Manchester **Identifier:** NA
Sponsor	MRC Clinician Scientist Fellowship (Grant no: MR/Z504026/1)
Conflict of interest	NA

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A-STAR	UK-Irish Atopic Eczema Systemic Therapy Register
BADBIR	British Association of Dermatologists Biologics and Immunomodulators Register
BILAG-BR	British Isles Lupus Assessment Group Biologics Register
BSRBR-RA	British Society for Rheumatology Biologics Register-Rheumatoid Arthritis
IBD	Inflammatory bowel disease
IL	Interleukin
IMIDs	Immune-mediated inflammatory diseases
JAK	Janus Kinase
RA	Rheumatoid arthritis
RCT	Randomised controlled trial
RWE	Real world evidence
TMLE	Targeted maximum likelihood estimation
TNFi	Tumour necrosis factor inhibitors
UK	United Kingdom
UK IBD	United Kingdom Inflammatory Bowel Disease Register

AMENDMENTS AND UPDATES

Version date	Version number	Section of protocol	Amendment or update	Reason
11/07/2025	1.0	First draft	NA	NA

MILESTONES

Milestone	Date
Workstream I) Benchmarking	
Systematic search of clinical trials for target trial emulation	25 Apr 2025
Finalise protocol V1	18 Aug 2025
Application to IMID registries for data access	18 Feb 2025
Data cleaning – IMID registry datasets	18 Aug 2025
Feasibility counts of samples size	18 Aug 2025
Data analysis – Target trial emulation	18 Aug 2026
Workstream II) Stratification	
Identify potential effect modifiers based on expert opinion	20 May 2025
Identify effect modifiers based on systematic review and data analysis	18 Jan 2026
Application to trial repositories for data access	18 Aug 2026
Conduct re-analyses of both registry and trial data in different strata of the pre-identified effect modifiers	18 Feb 2027
Workstream III) Transportability	
Application to trial repositories for data access	18 Aug 2025
Data cleaning – combine trial and registry datasets	18 Aug 2027
Data analysis – joint analysis, network meta-analysis	18 Aug 2028
Workstream IV) Prediction	
Develop a list of criteria for successful benchmarking examples from workstreams I-III	18 Aug 2028
Systematic search of ongoing trials to predict results	18 Aug 2028
Data cleaning	18 Aug 2028
Data analysis – predict outcomes from trials	18 Feb 2029

1 RATIONALE AND BACKGROUND

What is known about the condition:

Immune-mediated inflammatory diseases (IMIDs) are a group of conditions affecting different areas of the body with common immune-pathomechanisms, overlapping genetic predisposition, and therapeutic options.[1] Up to 7% of people globally, and approximately 4.7 million people in the United Kingdom (UK) suffer from one or more IMIDs.[2] Common IMIDs include rheumatic diseases such as rheumatoid arthritis (RA) and lupus, skin diseases such as psoriasis and atopic eczema, and inflammatory bowel diseases (IBD) such as Crohn's disease and ulcerative colitis. These conditions can lead to complications and other health problems, which may impose significant physical and psychological burden on individuals.[3] IMIDs are chronic conditions that can hinder individuals from achieving their full life potential or result in premature mortality, thereby contributing to the societal burden through decreased workforce productivity and engagement.[4-8] In the United States, an approximate cost incurred by productivity loss due to psoriasis was reported to be $16.6 billion. In addition, the total societal cost of RA was estimated to be $39.2 billion per year.[9]

What is known about the exposure of interest:

Immunomodulators are the mainstay of treatment for IMIDs. Before the advancement of biotechnologies, conventional broad acting immunosuppressants were mainly used to reduce inflammation. Because these treatments are not designed to target specific components of the immune system based on the disease pathogenesis, many of people with IMIDs did not respond to these therapies and rarely achieved complete remission, and were at risk of suffering from multiple adverse effects.[1]

Over the past few decades, novel immunomodulators that precisely target key cytokines and signalling pathways became available, allowing better control of the symptoms and treatment associated adverse events.[1] Common novel immunomodulators include biologic agents such as tumour necrosis factor inhibitors (TNFi), interleukin (IL) inhibitors (e.g. IL-1, IL-6, IL-12, IL-17 , IL-23 inhibitors), and small molecule drugs such as Janus Kinase (JAK) inhibitors. These treatments have been tested for clinical use through large international multi-centre explanatory randomized controlled trials (RCTs), forming the main evidence base to inform clinicians of the benefits and risks of these medications.[10-15]

Previous systematic reviews of RCTs that examined efficacy of novel immunomodulators have reported favourable treatment outcomes compared to non-targeted conventional treatments.[10,12,14,15] For example, higher proportion of patients who received biologic treatments for psoriasis (e.g. anti-IL17, anti-IL12/23, TNFi) showed improvements in disease severity compared to those who received non-biologic treatments. When compared to placebo arms, infliximab, bimekizumab, ixekizumab, and risankizumab significantly improved disease severity in people with moderate to severe psoriasis.[10] In treatment of RA, a combination of biologic agent and methotrexate had significantly better control of disease activity than oral methotrexate alone in patients with inadequate response to methotrexate. Adalimumab, etanercept, certolizumab, or infliximab with methotrexate had slower disease progression compared to methotrexate alone, however, no clinically meaningful differences were observed.[14] For patients with systemic lupus erythematosus, anifrolumab and belimumab had a better control of disease activity compared to the standard of care alone.[15] Despite the improved efficacy of novel immunomodulators shown in previous studies, current evidence on the safety profiles are still under debate. Some studies demonstrated comparable safety profiles of novel biologic therapies, while people with certain baseline conditions were associated with significantly higher risk of serious adverse events and infections.[10,11]

Gaps in knowledge:

While the advent of novel immunomodulators have transformed the lives of people with IMIDs, current clinical practices are mostly based on RCTs which may lack generalisability to the real-world population. One of the most significant limitations of RCT is the underrepresentation of minority populations. Individuals may not meet the restrictive trial eligibility criteria, face barriers to research participation, or be less likely to consent to participate in trials. Examples include pregnant women, older adults, and individuals from ethnic minority backgrounds. This potentially leads to an over-estimation of benefit and under-estimation of risk for some of these patients, a phenomenon termed the efficacy-effectiveness gap[16], especially if these drugs have a different effect in these populations.

What is the expected contribution of this study:

This is the first project to combine data from trials and registries across different IMIDs and implement the target trial framework. It will enable cross-IMID learnings, for example of potential consistent subgroup effects of common treatments across different diseases, or

consistent differences between trial and UK real-world populations. The aim of this project is to understand whether findings from observational studies can serve as evidence of reliable treatment effects in IMIDs. Learnings from this project may contribute to the advancement of target trial framework, particularly in the context of observational studies involving IMIDs. For example, we may identify situations where observational studies are insufficient or unreliable for clinical decision making. Our recommendations for potential considerations to add to the target trial framework could be used as a guideline for future studies that aims to emulate clinical trials using routinely collected data. Findings of this project may also provide insights into personalised treatment for IMIDs. For example, significant subgroup treatment effects based on clinical characteristics could be used as evidence in developing personalised treatment strategies.

2 RESEARCH QUESTION AND OBJECTIVES

This project aims to conduct treatment effectiveness research using data from RCTs and registries, with a goal of generating evidence that supports the optimal use of immune-modifying therapies among people with IMIDs.

Four main objectives of this project are:

1. To emulate the trials of treatments for IMIDs using data from the registries in the UK and evaluate benchmarking success.
2. To compare subgroup strata based on patient characteristics and identify common effect modifiers in each disease and across IMIDs.
3. To conduct transportability analyses by combining datasets obtained from trial repository and IMID registries.
4. To predict treatment effect outcomes of ongoing IMID trials using the common factors identified for a successful benchmarking of a trial.

3 METHODS

3.1 Data sources

Longitudinal observational data will be obtained from the following IMID registries: British Association of Dermatologists Biologics and Immunomodulators Register (BADBIR) for psoriasis, UK-Irish Atopic Eczema Systemic Therapy Register (A-STAR) for atopic eczema, British Isles Lupus Assessment Group Biologics Register (BILAG-BR) for lupus, British Society for Rheumatology Biologics Register-Rheumatoid Arthritis (BSRBR-RA) for RA, and UK IBD Register (UK IBD) for IBDs. These are based in the UK and the Republic of Ireland and are designed as prospective pharmacovigilance registries to follow people with IMIDs starting systemic therapies from secondary care. Randomised controlled trials data will be used for subgroup analyses and transportability analyses. Data will be requested through Vivli, a private non-profit organisation that act as a neutral broker between data custodians and researchers.[17]

3.1.1 Context and rationale for data sources

Reason for selection: These data sources are considered as national registries in the UK and the Republic of Ireland and have a much larger sample size than RCTs, which is critical for investigation of rare adverse effects or subgroup effects of medications. In addition, these data sources allow the inclusion of a diverse research population. Selection of a target trial will be based on the ability to emulate the participant eligibility criteria and outcome measures using the registry data.

Strengths of data source(s): These registries include baseline clinical information from all participants and scheduled follow-up data, and benefits from data enrichment through linked secondary care hospitalisation data. Obtaining individual participant data from clinical trials will allow for secondary subgroup analysis.

Limitations of data source(s): Limitations of registry data are missing information, measurement errors arising from data entry errors, lack of data standardisation due to

inconsistencies in data collection methods, and potential recall bias with patient-reported characteristics.

Data source provenance/curation: Data are available from the British Association of Dermatologists, the British Society for Rheumatology, UK IBD Registry and Vivli upon approval.

Table 1. Metadata about data sources and software.

	Data 1	Data 2	Data 3	Data 4	Data 5
Data Source(s):	**BADBIR (Psoriasis registry)**	**BILAG-BR (Lupus registry)**	**BSRBR-RA (Rheumatoid arthritis)**	**A-STAR (Atopic eczema)**	**IBD (Inflammatory bowel disease)**
Study Period:	Start date: Initiation of exposure End date: Target trial dependent	Start date: Initiation of exposure End date: Target trial dependent	Start date: Initiation of exposure End date: Target trial dependent	Start date: Initiation of exposure End date: Target trial dependent	Start date: Initiation of exposure End date: Target trial dependent
Eligible Cohort Entry Period:	Same as study period	Same as study period	Same as study period	Same as study period	Same as study period
Data Version (or date of last update):	TBC	TBC	TBC	TBC	TBC
Data sampling/extraction criteria	People exposed to one of the treatments listed below: Adalimumab, Methotrexate, Adalimumab, Secukinumab, Ustekinumab, Fumaric Acid Esters, Etanercept, Apremilast, Acitretin, Ixekizumab, Bimekizumab, Brodalumab, Certolizumab, Deucravacitinib, Infliximab	People exposed to one of the treatments listed below: Belimumab, Rituximab, Anifrolumab, SLE standard therapy	People exposed to one of the treatments listed below: Rituximab, Etanercept, Infliximab, Adalimumab, Certolizumab, Tocilizumab, Anakinra, Tofacitinib, Baricitinib, Upadacitinib, Filgotinib, Azathioprine, Hydroxychloroquine, Leflunomide, Methotrexate, Sulfasalazine	People exposed to one of the treatments listed below: Dupilumab, Upadacitinib, Methotrexate, Ciclosporin	People exposed to one of the treatments listed below: TBC
Type(s) of data:	Registry	Registry	Registry	Registry	Registry

Data linkage:	HES	NA	Death, cancer data	NA	NA
Conversion to CDM*:	NA	NA	NA	NA	NA
Software for data management:	Stata/R	Stata/R	Stata/R	Stata/R	Stata/R

*CDM = Common Data Model

3.2 Research Methods

3.2.1 Study design

This project involves four interrelated workstreams: benchmarking, stratification, transportability, and prediction.

3.2.1.1 Workstream1: Benchmarking

RCTs as targets for benchmarking will be identified through a systematic search of explanatory trials using medical databases, systematic reviews and trial registries. Trials will be considered for emulation if they report outcomes that align with those captured in observational registry data and involve head-to-head comparisons of treatments. Selected trials will be emulated using the IMID registry data where the registry sample sizes for the comparator arms are equal or exceed that of the trial. We will emulate the pragmatic trial equivalent of the target RCT. Target trial protocols will include eligibility criteria, treatment strategies, treatment assignment, outcome measures, and follow-up period. Effect estimates obtained from trial emulation will be compared with the estimates from target trials to evaluate benchmarking success (agreement between trial and registry outcomes).

3.2.1.2 Workstream 2: Stratification

Patient-level data will be obtained from the trial repositories. Potential effect modifiers will be identified based on the patient data from the explanatory trial with literature review, and expert opinion, and using data driven approaches involving machine learning methods. Identified effect modifiers may either be different or consistent across IMIDs for the same therapy. Treatment effect estimates will be analysed in various subgroup strata of the identified effect modifiers for the conditional average treatment effect. Benchmarking success will be evaluated for the identified subgroup across both trials and registries using the same metrics as in the benchmarking workstream.

3.2.1.3 Workstream 3: Transportability

Data of trials that showed successful emulation will be combined with registry data to conduct transportability analyses. The combined dataset will be weighted to create a pseudo-population that reflects the baseline characteristics of the patient population in the registry and trial data, separately. Treatment effect estimates will be compared before and after

application of weights to identify any differences in the efficacy of treatments. In the case where there is i) a small difference in clinical significance between trial and registry estimates after reweighting, or ii) substantial regulatory agreement and low standardised differences, joint analyses will be conducted by combining data from trial and registry and reweighting to the real-world population.

Sensitivity analyses will be conducted using the registry data for those who do not meet the trial eligibility criteria. The treatment effect estimates will be compared between individuals who would not be eligible for the trial and those who would be.

3.2.1.4 Workstream 4: Prediction

Learnings from previous workstreams will be used to provide recommendations for potential considerations to add to the existing target trial framework for observational studies in IMIDs using treatment registries. The updated framework will be applied to emulate on-going trials and predict the results. The predicted results will be compared to the actual results to test how closely our results can match the results from the RCTs.

3.3 Setting

3.3.1 Definition of time 0 for entry to the study population

People with IMIDs and have received at least one of the exposures of interest for each disease will be included. The cohort entry date (i.e. time 0) will be defined as the date of treatment initiation. A graphical representation of the study design is shown in Figure 1.

.

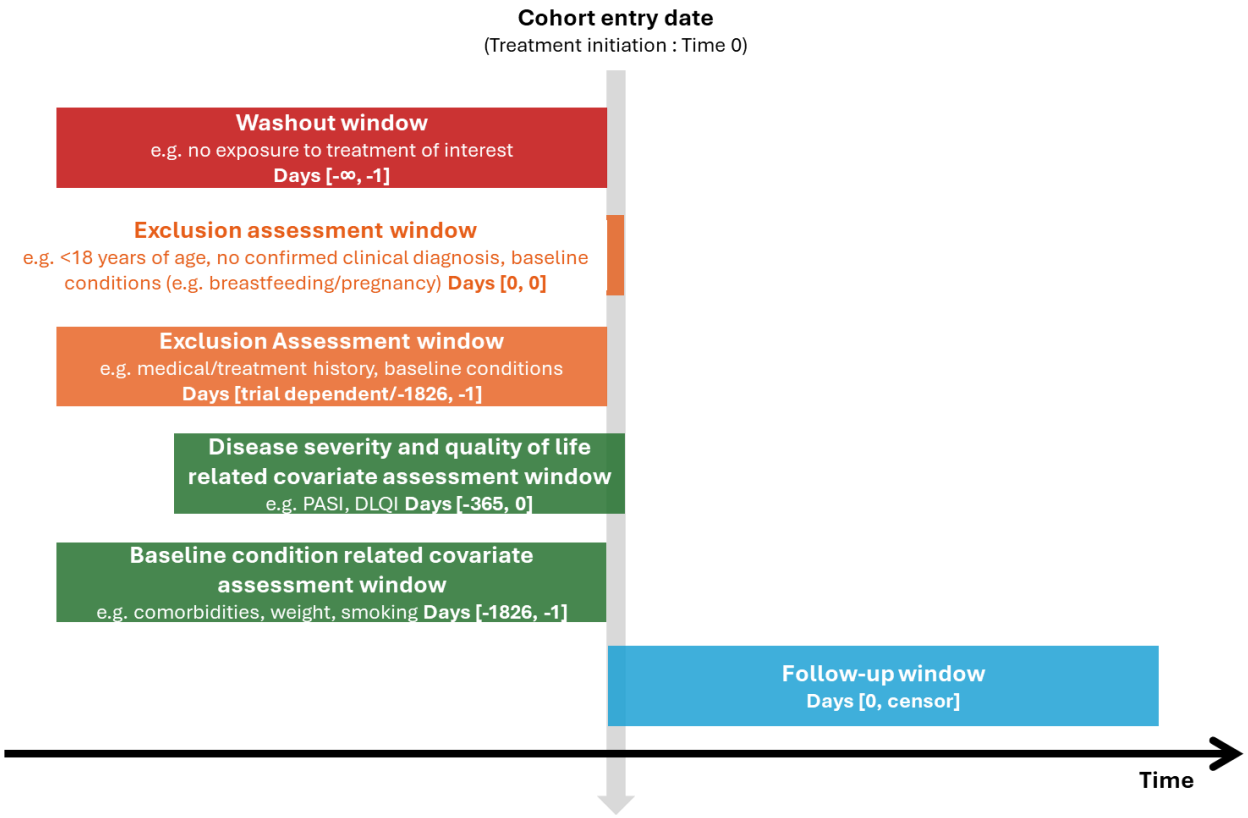

Figure 1. Study design

3.3.2 Inclusion and exclusion criteria

Inclusion and exclusion criteria for all workstreams will follow the eligibility criteria of the selected target trials for benchmarking across different IMIDs. The duration of all assessment windows will follow the target trials. Depending on the study population (e.g. new treatment users), there will be a washout period prior to the cohort entry date. In cases where the duration of the washout window is not defined, the history of exposure will be identified from all available records prior to the cohort entry date.

Other exclusion criteria will be applied during the exclusion assessment window defined in the target trials. In cases where look-back period is not specified, the duration of the exclusion assessment window will follow the covariate assessment window. For example, exclusion criteria on pre-existing comorbid conditions will be identified up to five years prior to the cohort entry date, whereas disease severity or quality of life measures (e.g. Psoriasis Area and Severity Index scores, Dermatology Life Quality Index) will be identified up to one year prior to the cohort entry date.

3.4 Variables

3.4.1 Context and rationale for exposure(s) of interest

3.4.1.1 Algorithm to define duration of exposure effect

People with a confirmed diagnosis of IMIDs who have met disease classification criteria of the trial and initiated one of the treatments listed in Table 1 will be included in this study. Exposure of interest will be the same as the target trials across different IMIDs. The date of treatment initiation will be defined as time 0. Continuous treatment will be defined as continuous administration of the same treatment throughout the follow-up period. To align with current clinical practice, patients who restart the same treatment within one month of discontinuation will also be considered as receiving continuous treatment. Duration of exposure effect will be defined as the time from cohort entry (i.e. time 0) until treatment discontinuation, end of study period, or death, whichever occurs earlier. (Figure 2)

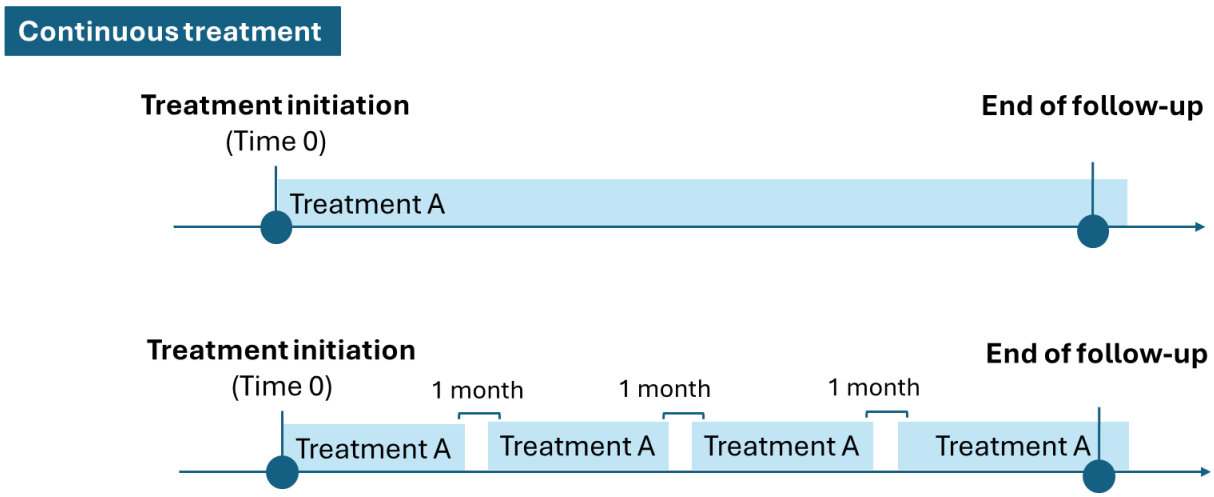

Figure 2. Graphical illustration of continuous treatment.

3.4.2 Context and rationale for outcome(s) of interest

3.4.2.1 Agreement metrics

Throughout the project, the alignment of effect estimates between target trials and emulated observational studies will be evaluated using three agreement metrics[18]: regulatory agreement, estimate agreement, and standardised differences.

Regulatory agreement examines whether the outcomes obtained from observational studies are in concordance with target trials by comparing the direction and statistical significance of the effect estimates. Since RCTs provide evidence to determine regulatory approvals, this metric aims to assess the ability of observational studies to draw the same conclusions as the target trials. For a successful emulation, observational studies should be able to show that their effect estimates are in the same direction and statistical significance as the target trials.

Estimate agreement aims to assess whether the effect estimates from emulated observational studies lie within the 95% confidence interval of the estimates reported in target trials. Observational studies are likely to have more statistical power than RCTs. There may be cases where the effect estimates are in the same direction, but the statistical significance is achieved in observational studies only. Therefore, the estimate agreement can be used as a secondary metric as it only considers the ratio of the variances between emulated observational studies and target trials.

Standardised differences can be used as an exploratory measure by quantifying the differences in the effect estimates and associated variances between emulated observational studies and target trials. The standardised differences are calculated as below[18]:

$$SD = \frac{RWE\hat{\theta} - RCT\hat{\theta}}{\sqrt{RWE\hat{\sigma}^2 + RCT\hat{\sigma}^2}}$$

A successful emulation will be determined based on the 95% confidence interval (CI) for the difference. CIs are calculated as below:

$$95\%\ CI = d \pm 1.96 \times SE(d)$$

3.4.2.2 Algorithm to define outcome events

All primary and secondary endpoints will follow the benchmarking target trials. Identification of outcome events will allow maximum of two months grace period before and after the end of follow-up period to minimise missing data.

Expected outcomes from each workstream are described below:

Workstream 1: Benchmarking

- Overall correlation between the paired trial-registry results

- Number and percentage of studies meeting the trial emulation outcomes.
- Evaluation of the study characteristics with an emphasis on discrepancies between trial and emulations in the protocol, including sample size; population characteristics; failure to emulate exposure (e.g. dosing regimen); missing data; attrition; number of important unmeasured confounders; and their association with trial emulation outcomes.
- List of trials which were successfully emulated (which will be most relevant to 'prediction' workstream)
- List of trials which were unsuccessfully emulated (which will be most relevant to 'stratification' workstream).

Workstream 2: Stratification

- Evaluate both *a priori* defined effect modifiers and data-driven prognostic variables in both trial and registry data
- Evaluate benchmarking success in stratified groups
- Identification of effect modifiers / biomarkers where trial benchmarking became successful in the strata.

Workstream 3: Transportability

- Quantification of important trial vs real-world population differences across IMIDs
- Identify trials where patient consent should incorporate information from estimates that have substantially changed in a clinically significant way after re-weighting for a real-world population.
- Identify the effect estimate in people who would have been ineligible for trials for better patient consent.
- Improve precision of estimate of effect for the comparators evaluated in the joint analysis.

Workstream 4: Prediction

- List of additional criteria to add to the existing target trial framework.
- Overall correlation between the paired trial-registry results.
- Number and percentage of studies meeting the trial emulation outcomes.
- Evaluation of the successful criteria and whether this should be used in future observational analyses by researchers and regulatory authorities.

3.4.3 Context and rationale for follow up

Follow-up periods for each cohort study will be the same as the target trial. Each cohort will be followed up from the date of treatment initiation (i.e. time 0) until the last known follow-up date (e.g. end of study period, event of discontinuation, event of outcome, or event of death, which ever occurred earlier). All patients will be censored at the end of the follow-up. Discontinuation due to treatment or adverse events will be considered as treatment failures regardless of the actual outcome at the time of discontinuation. Patients who were discontinued due to any other reason will be censored on the date of treatment discontinuation. (Figure 3) The date of add-on or switch will follow the definitions from the target trials. Depending on the target trials, any observations that deviate from treatment protocol may be censored, or dropped out of denominator, or considered as failure.

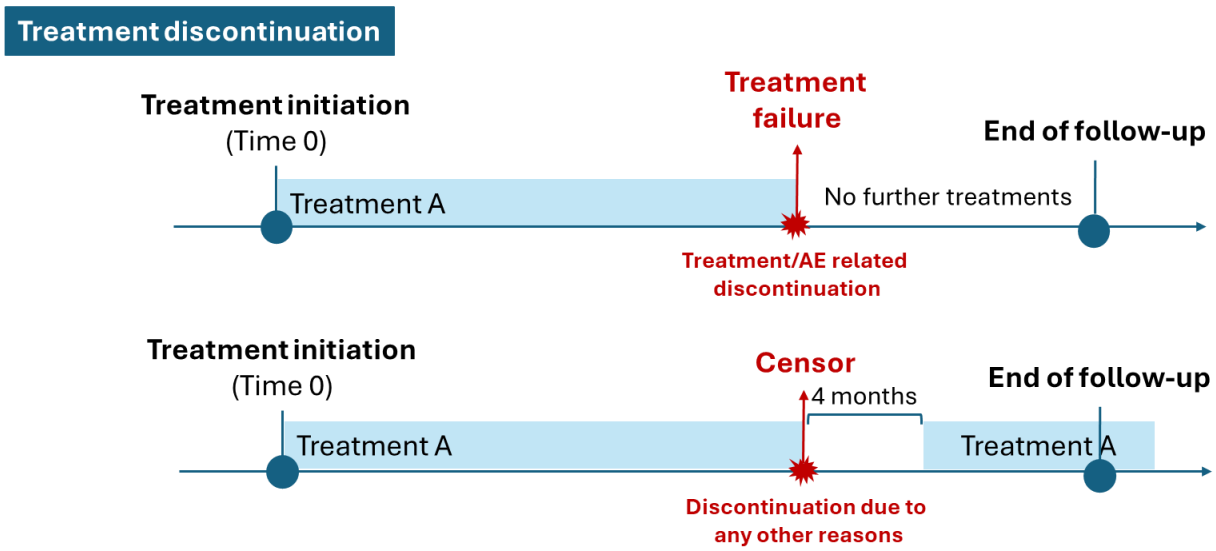

Figure 3. Graphical illustration of treatment discontinuation.

Table 2. Operational definitions of follow-up.

	IMIDs
Follow up start	Date of treatment initiation (i.e. time 0)
Follow up end[1]	Last known follow-up date (e.g. End of study period, event of death, event of treatment discontinuation, event of outcome)
Date of outcome	Date of outcome event (e.g. infection diagnosis, PASI score at week 24)
Date of death	Date of death
End of observation in data	Event of discontinuation, no following observations from the last observation
Day X following index date *(specify day)*	End of study period will be the same as target trials

End of study period (specify date)	All patients will be followed up for N weeks from the index date : censor
End of exposure (specify operational details, e.g. stockpiling algorithm, grace period)	Event of discontinuation due to treatment : failure Event of discontinuation due to adverse event : failure Event of discontinuation due to any other reason : censor
Deviation from protocol *(specify algorithm)*	Target trial dependent Observational analogue of modified intention to treat analysis: include outcome at specified timepoint, no censoring Observational analogue of per-protocol analysis: censor at timepoint at which participant deviates from allocated assignment and protocol
Other date *(specify)*	NA

[1] Follow up ends at the first occurrence of any of the selected criteria that end follow up.

3.4.4 Context and rationale for covariates

Covariates for each workstream will be dependent on the outcomes of the identified benchmarking trials. A summary of the baseline characteristics of the trial populations will be presented to describe the distribution of covariates. Additionally, important prognostic factors of IMIDs or confounding factors will be informed by clinical opinion and identified from previous literature. For stratification workstream, additional covariates that might be relevant to each subgroup strata based on the effect modifiers will be included. For transportability workstream, sociodemographic characteristics or baseline conditions that show different distribution in registry and trial data will be prioritised to balance out the differences between two populations. Directed acyclic graphs will be presented to outline the assumptions.

Relevant covariates will be identified during the covariate assessment windows. (Figure 1) Disease-related covariates such as disease severity index and quality of life index scores will be looked up to one year prior to the cohort entry date. Baseline conditions such as comorbidities, weight, and smoking status will be identified up to five years prior to the cohort entry date.

3.5 Data analysis

3.5.1 Context and rationale for analysis plan

The treatment effects for each IMID will be assessed depending on the type of outcome measures identified from the target trials. For example, generalised linear models will be used

to assess binary outcome measures, whereas for time-to-event outcomes, selection of appropriate models, such as semi-parametric or parametric models, will be considered.

Benchmarking success will be evaluated by estimating the overall correlation between the paired trial-registry using the main and the secondary criteria of regulatory agreement. The main criterion of regulatory agreement refers to the full agreement between the trial and registry and examines whether the estimates and 95% (Confidence Interval) CIs are in the same direction. The secondary criterion of regulatory agreement is defined by whether the registry estimate lies within the 95% confidence interval of the trial estimate. As an exploratory criterion, the standardised difference of the effect estimates between trial and registry should achieve statistical significance. Pearson correlation coefficient and Cohen's Kappa coefficient will be used to measure reliability of the agreement between trial and registry estimates. Improvements in clinical practice over time may be an important factor in determining treatment outcomes. For cases where advancements in treatment strategies have significantly impacted the outcomes, sensitivity analyses will be conducted by restricting the observational periods (e.g. follow-up duration) or using historical controls. For transportability analyses, multivariable logistic regression model will be fitted to the combined datasets to identify baseline covariates that significantly predict enrolment in a trial or registry, respectively. The concordance statistics (C-statistic) will be used to evaluate the predictive power of a model.

To adjust for potential confounding factors in benchmarking and stratification workstreams, different techniques such as propensity score-based weighting, g-estimation, and targeted maximum likelihood estimation (TMLE) will be applied. Doubly robust methods such as augmented inverse probability treatment weighting and TMLE will be favoured to reduce the chance of model misspecification leading to bias. The most appropriate method will be selected based on covariate balance and whether this leads to a closer emulation of the trial effect estimate. For g-estimation, regression or survival models can be used to estimate the probability of receiving a treatment depending on the study design. In cases where dynamic treatment strategies need to be emulated, methods such as clone censoring weighting will be applied.

To detect unmeasured confounders, negative controls, E-values, and quantitative bias analysis models will be used. Negative control refers to any exposure or outcome that do not have causal relationship with the outcome of interest. Therefore, if the negative control leads to similar results as the main exposure of interest, it may indicate that the observed effect from the main analysis is likely to be due to confounding. Potential negative controls will be identified across IMIDs with clinical opinion and input.

The magnitude of association to unmeasured confounding will be assessed using the E-values. The E-value estimates the minimum strength of association between confounder, treatment and outcome while accounting for the measured covariates included in the model.[20] This method will measure the robustness of the observed results.

In cases where the results are more sensitive to bias, quantitative bias analysis will be performed. Appropriate bias models will be selected depending on the type of bias and the availability of external information on the bias parameters. This approach quantifies the impact of unmeasured bias on the observed effect estimates and generates bias-adjusted estimates.

Sensitivity analyses will explore how emulation design and population differences may influence the closeness of the target trial emulation. For example, differences in effect estimates between RCTs and observational studies may be related to emulation designs such as differences in eligibility criteria (e.g. number of key eligibility criteria able to be emulated, small sample size), treatment strategies (e.g. incomplete emulation of concomitant treatments, dose titration), assignment procedures (e.g. number of days measured prior to cohort entry, inability to adjust for potential confounders), quantification of look-back or follow-up periods (e.g. strict or unspecified look-back periods, duration of follow-up), quality of outcome measures (e.g. using operational definition to measure outcome, missing data, loss to follow-up), and data analysis (e.g. missing data in baseline covariates). Meta-regression method will be used to examine the association between these characteristics and the closeness of emulation across studies. This approach is an extension of meta-analysis that are conceptually similar to the weighted linear regression model. It includes study-level covariates to explain variability in emulation success from each study.[19]

Table 3. List of *a priori* identified measures for closeness of emulation between trials and registries

Target trial framework element	Gaps between RCT and RWE		
	Identified measures	**Categorisation**	**Description**
Eligibility criteria	Number of key eligibility criteria able to be emulated	Continuous	Number of key eligibility criteria where there is a difference between the way trial and RWE measure the covariates
	Sample size	Small sample size/moderate-large sample size; or sample size as a continuous variable	Small sample size: Below X participants Moderate-large sample size: Between X-X participants Number of participants as continuous variable
Treatment strategies	Concomitant treatments	Incomplete/Complete;	Incomplete: Not all concomitant treatments identified in RCT are captured in registry data Complete: All concomitant treatments identified in RCT are fully captured using registry data
	Dose titration after treatment initiation	Continuous	Proportion of incomplete emulation for dose titration
	Time of treatment initiation	Continuous	Average year/month of the treatment initiation
Assignment procedures	Covariate measurement quality for adjustment of confounding	Continuous	Number of days measured prior to cohort entry
	Number of identified potential confounders not adjusted for	Continuous	Number of major (considered influential) and minor (considered non-influential) potential confounders not adjusted for
Follow-up period	Strict or unspecified look-back periods	Yes/No	Yes: Used different look back period from RCT (or RCT did not mention specific look back period) No: Same look back period as the RCT
	Duration of follow-up	Continuous	Closeness of the measured outcome to the desired timepoint, measured by days
Outcome	Definition of outcome measures	Yes/No	Yes: Operational definition was used to identify outcome measures in registry data. No: Same definition from RCT was used to identify outcome measures.

	Missing data in outcome measures	Good/moderate-poor; or use proportion as a continuous variable to plot against	Good: No missing or minimum (e.g. <X%) missing data Moderate: Moderate level of missing data X%-X% Poor: High proportion of missing data (>X%) Proportion of missing data
	Measurement of attrition	Continuous	Proportion of patients lost to follow-up
Causal contrasts of interest			
Data analysis	Missing data in baseline covariates	Continuous	Proportion of missing baseline covariate data

3.5.2 Data management

Data will be transferred and stored in a secured virtual platform where only the authorised researchers will be allowed to access the data. All datasets will be anonymised, and some information will be abbreviated to comply data privacy regulations.

3.5.3 Quality control

To ensure data quality, the proportion of missing data will be identified. After registration, patient is followed up at 3 months, 6 months, and 12 months post-treatment. Due to the lags in data capture, we will allow 1 month grace period for defining continuous treatment and ±2 months grace period for identifying outcome measures. Codes for data cleaning and statistical analyses will be written and cross checked between two researchers.

3.5.4 Study size and feasibility

No sample size calculation will be conducted as the cohort identification

3.6 Limitations of the methods

Benchmarking the RCTs using the real-world registry data assumes that the treatment allocation in RCTs do not affect the outcomes and the treatment given in trials are assumed to be identical to the treatments given in real-world settings. Violation of this assumption may lead to selection bias. It also assumes that all baseline covariates measured in the

observational data are adjusted for their confounding effects, allowing similar baseline characteristics between treatment groups and between populations. This overlap will allow different populations or treatment groups to be comparable. Transportability analyses also take above assumptions into account but with stronger requirements. Positivity assumption is particularly important in transportability analyses as it assumes non-zero probability of receiving any treatments. This will allow some degree of diversity in study population, and violation of this assumption may impose challenges in generalising the observed effect estimates to the UK population. We will use simulation and extrapolation methods to account for any positivity assumptions.

4 PROTECTION OF HUMAN SUBJECTS

This study does not directly involve human participants, and the data obtained from registries are used for non-interventional cohort studies.

5 REPORTING OF ADVERSE EVENTS

NA

6 REFERENCES

1. McInnes IB, Gravallese EM. Immune-mediated inflammatory disease therapeutics: past, present and future. *Nat Rev Immunol*. Oct 2021;21(10):680-686. doi:10.1038/s41577-021-00603-1
2. El-Gabalawy H, Guenther LC, Bernstein CN. Epidemiology of immune-mediated inflammatory diseases: incidence, prevalence, natural history, and comorbidities. *J Rheumatol Suppl*. May 2010;85:2-10. doi:10.3899/jrheum.091461
3. Monteleone G, Moscardelli A, Colella A, Marafini I, Salvatori S. Immune-mediated inflammatory diseases: Common and different pathogenic and clinical features. *Autoimmunity Reviews*. 2023/10/01/ 2023;22(10):103410. doi:https://doi.org/10.1016/j.autrev.2023.103410
4. Kimball AB, Gieler U, Linder D, Sampogna F, Warren RB, Augustin M. Psoriasis: is the impairment to a patient's life cumulative? *J Eur Acad Dermatol Venereol*. Sep 2010;24(9):989-1004. doi:10.1111/j.1468-3083.2010.03705.x

5. Warren RB, Kleyn CE, Gulliver WP. Cumulative life course impairment in psoriasis: patient perception of disease-related impairment throughout the life course. *Br J Dermatol*. May 2011;164 Suppl 1:1-14. doi:10.1111/j.1365-2133.2011.10280.x
6. Ogdie A, Haynes K, Troxel AB, et al. Risk of mortality in patients with psoriatic arthritis, rheumatoid arthritis and psoriasis: a longitudinal cohort study. *Ann Rheum Dis*. Jan 2014;73(1):149-53. doi:10.1136/annrheumdis-2012-202424
7. Voss A, Laustrup H, Hjelmborg J, Junker P. Survival in systemic lupus erythematosus, 1995-2010. A prospective study in a Danish community. *Lupus*. Oct 2013;22(11):1185-91. doi:10.1177/0961203313498796
8. Bultink IEM, de Vries F, van Vollenhoven RF, Lalmohamed A. Mortality, causes of death and influence of medication use in patients with systemic lupus erythematosus vs matched controls. *Rheumatology (Oxford)*. Jan 5 2021;60(1):207-216. doi:10.1093/rheumatology/keaa267
9. Jacobs P, Bissonnette R, Guenther LC. Socioeconomic burden of immune-mediated inflammatory diseases--focusing on work productivity and disability. *J Rheumatol Suppl*. Nov 2011;88:55-61. doi:10.3899/jrheum.110901
10. Sbidian E, Chaimani A, Guelimi R, et al. Systemic pharmacological treatments for chronic plaque psoriasis: a network meta-analysis. *Cochrane Database Syst Rev*. Jul 12 2023;7(7):Cd011535. doi:10.1002/14651858.CD011535.pub6
11. Janke K, Biester K, Krause D, et al. Comparative effectiveness of biological medicines in rheumatoid arthritis: systematic review and network meta-analysis including aggregate results from reanalysed individual patient data. *Bmj*. Jul 7 2020;370:m2288. doi:10.1136/bmj.m2288
12. Ernst D, Krüger K, Witte T. [Head-to-head trials in rheumatoid arthritis]. *Z Rheumatol*. Oct 2024;83(8):609-619. Head-to-Head-Therapiestudien zur rheumatoiden Arthritis. doi:10.1007/s00393-024-01517-8
13. Felten R, Scherlinger M, Mertz P, Chasset F, Arnaud L. New biologics and targeted therapies in systemic lupus: From new molecular targets to new indications. A systematic review. *Joint Bone Spine*. Mar 2023;90(2):105523. doi:10.1016/j.jbspin.2023.105523
14. Hazlewood GS, Barnabe C, Tomlinson G, Marshall D, Devoe DJ, Bombardier C. Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying anti-rheumatic drugs for rheumatoid arthritis: A network meta-analysis. *Cochrane Database Syst Rev*. Aug 29 2016;2016(8):Cd010227. doi:10.1002/14651858.CD010227.pub2
15. Kostopoulou M, Mukhtyar CB, Bertsias G, Boumpas DT, Fanouriakis A. Management of systemic lupus erythematosus: a systematic literature review informing the 2023 update of the EULAR recommendations. *Annals of the Rheumatic Diseases*. 2024;83(11):1489-1501. doi:10.1136/ard-2023-225319
16. Nordon C, Karcher H, Groenwold RH, et al. The "Efficacy-Effectiveness Gap": Historical Background and Current Conceptualization. *Value Health*. Jan 2016;19(1):75-81. doi:10.1016/j.jval.2015.09.2938
17. Vivli. About Vivli: Overview. Accessed April, 2025. https://vivli.org/about/overview/
18. Franklin JM, Pawar A, Martin D, et al. Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. *Clin Pharmacol Ther*. Apr 2020;107(4):817-826. doi:10.1002/cpt.1633
19. Higgins J, López-López J, Aloe A. Meta-Regression. In: Christopher Schmid, Theo Stijnen, Ian White, eds. *Handbook of Meta-Analysis*. First ed. Chapman and Hall/CRC; 2020:chap 7.

20. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Annals of Internal Medicine*. 2017/08/15 2017;167(4):268-274. doi:10.7326/M16-2607