

# Innovative Trial Approaches in Immune-Mediated Inflammatory Diseases: Current Use and Future Potential

**Michael J Grayling**

Newcastle University

**Theophile Bigirumurame**

Newcastle University

**Svetlana Cherlin**

Newcastle University Institute for Health and Society

**Luke Ouma**

Newcastle University

**Haiyan Zheng**

Newcastle University Institute for Health and Society

**James Wason** (✉ [james.wason@newcastle.ac.uk](mailto:james.wason@newcastle.ac.uk))

Newcastle University <https://orcid.org/0000-0002-4691-126X>

---

## Research article

**Keywords:** Adaptive design, Basket design, Bayesian design, Composite endpoint, High-dimensional data, Routinely collected data, SMART trial, Umbrella design.

**DOI:** <https://doi.org/10.21203/rs.3.rs-63613/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Despite progress that has been made in the treatment of many immune-mediated inflammatory diseases (IMIDs), there remains a need for improved treatments. Randomised controlled trials (RCTs) provide the highest form of evidence on the effectiveness of a potential new treatment regimen, but they are extremely expensive and time consuming to conduct. Consequently, much focus has been given in recent years to innovative design and analysis methods that could improve the efficiency of RCTs. In this article, we review the current use and future potential of these methods within the context of IMID trials.

**Methods:** We provide a review of several innovative methods that would provide utility in IMID research. These include novel study designs (adaptive trials, Sequential Multi-Assignment Randomised Trials, basket, and umbrella trials) and data analysis methodologies (augmented analyses of composite responder endpoints, using high-dimensional biomarker information to stratify patients, and emulation of RCTs from routinely collected data). IMID trials are now well-placed to embrace innovative methods. For example, well-developed statistical frameworks for adaptive trial design are ready for implementation, whilst the growing availability of historical datasets makes the use of Bayesian methods particularly applicable.

To assess whether and how these innovative methods have been used in practice, we conducted a review via PubMed of clinical trials pertaining to any of 51 IMIDs that were published between 2018-20 in five high impact factor clinical journals.

**Results:** Amongst 97 articles included in the review, 19 (19.6%) used an innovative design method, but most of these were relatively straightforward examples of innovative approaches. Only two (2.1%) reported the use of evidence from routinely collected data, cohorts, or biobanks. Eight (9.2%) collected high-dimensional data.

**Conclusions:** Application of innovative statistical methodology to IMID trials has the potential to greatly improve efficiency, to generalise and extrapolate trial results, and to further personalise treatment strategies. Currently, such methods are infrequently utilised in practice. New research is required to ensure that IMID trials can benefit from the most suitable methods.

## Background

Immune-mediated inflammatory diseases (IMIDs) consist of many distinct conditions that share common inflammatory pathways. They range in prevalence from more common conditions such as rheumatoid arthritis (0.5-1% prevalence in western populations(1)) and psoriasis (2% prevalence in North America(2)), to much rarer conditions such as Behçet's disease (estimated 0.005% prevalence in the US(3)). Overall, around 5–7% of the population of western societies has at least one IMID(4), with co-occurrence of multiple IMIDs common(5). IMIDs are associated with significant, chronic, morbidity

affecting quality of life and leading to premature death. As many IMIDs develop later in life, the prevalence is likely to increase as the world population ages.

Despite substantial progress in treatment of IMIDs with newly developed disease-modifying anti-rheumatic drugs and biologics, a substantial proportion of patients fail to respond to treatment or eventually relapse after successful treatment(6). Consequently, a considerable number of new drugs are in the clinical development pipeline(7) that require demonstration of efficacy and safety. Additionally, with the number of treatments currently available, there is substantial scope for optimising present use through the development of 'treat-to-target' approaches(8) and the tailoring of treatment according to patient subgroups(9). Any such optimised approach also requires demonstration of efficacy and safety, however.

The highest form of evidence is generated by randomised controlled trials (RCTs). For a new drug they provide the most compelling confirmation of benefit over standard therapies. For compare different treatment optimization strategies, RCTs avoid biases that may occur in an evaluation via a retrospective or prospective observational study. Despite the benefits of RCTs, there are important drawbacks too. RCTs are very expensive to conduct, especially large phase III trials with longer-term follow-up(10). Accordingly, there has been a strong focus on developing innovative methods for increasing the efficiency of clinical trials. These may have the aim of providing more information from the same number of patients (e.g., by increasing the power to find significant treatment effects), or to reduce the average number of patients recruited to trials without sacrificing power.

In this paper we provide an overview of several innovative methods for increasing the efficiency of clinical trials, framing our discussions within the context of potential benefits to IMID research. We also present a review of recently published IMID trials to investigate how often these approaches have been used in practice.

## **Overview Of Innovative Methods For Immune-mediated Inflammatory Disease Trials**

### **Emulating trials from observational data**

Given the large costs associated with prospective RCTs, an important question to consider is whether one is needed to answer a research hypothesis. This question has received particular attention in recent years, given the increasing amount of routinely collected data available, from sources such as CALIBER(11). Furthermore, there are now an array of patient cohorts and registries, with IMID-Bio-UK(12) an example of a UK initiative to bring these together for various IMIDs.

These data sources allow comparisons of different treatment strategies to be conducted through retrospective observational studies. Results from such analyses can be valuable, but are subject to confounding and other flaws such as selection bias and immortal-time bias(13). This is especially true if inappropriate analyses are applied.

An example, from outside of IMIDs, of where inappropriate analyses gave a misleading answer is presented by Dickerman et al(14). The effect of statins on the risk of developing cancer was assessed from retrospective data by comparing individuals who had received multiple years of statin therapy against those who had not. Even after adjustment for potential confounders, this approach was severely biased: a consequence of the fact that individuals who received multiple years of statin therapy could not have done so if they had died from cancer before or during that time. Within IMIDs, a recent paper(15) reviewed retrospective comparative effectiveness evaluations in rheumatoid arthritis; it was found most analyses had some flaws that would potentially lead to biases.

Instead, an approach called emulation of a target trial(16) can address many biases and result in more reliable answers. This involves specifying the 'target trial' that one would have liked to have done (i.e., which patient population, intervention, comparator, and outcomes) and analysing the data in a way that emulates this as closely as possible. Each timepoint in the retrospective data is then examined to identify which patients would have been eligible for randomisation in the target trial. The probability that they could have received intervention or comparator is modelled in a way that emulates random assignment from a trial as closely as possible. Dickerman et al(14) demonstrate how this approach, applied to data from CALLIBER, yields the same conclusions as a large meta-analysis of RCTs for the (lack of) effect of statins on reducing risk of cancer.

With many IMIDs being chronic conditions, RCTs are often used to compare different strategies for employing treatments known to be efficacious. Examples may include testing different 'treat-to-target' strategies(8) that may employ more aggressive treatment until a measure of disease activity is below a set threshold. When different strategies are already being employed in practice, and frequent measures of disease activity are recorded in routine data, emulation of target trials may be an efficient approach for evaluating different strategies.

It is important to note, however, that target trial emulation is still subject to bias. This is especially true if the routine dataset does not record sufficient information on potential confounding variables (or if there is a lot of missing data). Consequently, there may still be a need for prospective RCTs of treatment strategies. Nonetheless, target trial emulation could play an important role in prioritising which strategies should be tested and whether an RCT is likely to be successful in finding a significant effect.

### **Adaptive trial designs**

An adaptive design is one "that offers pre-planned opportunities to use accumulating trial data to modify aspects of an ongoing trial while preserving the validity and integrity of that trial"(17). Adaptive designs consist of a wide range of approaches that can improve efficiency in trials. Unlike the other innovative methodologies we discuss here, they have been discussed at length in other recent articles. There are both papers that have provided an overview of adaptive designs in general(18) and for specific clinical areas such as rheumatology(19). We refer the reader to these articles for a comprehensive introduction to adaptive designs.

However, we do provide in Table 1 a brief summary of several available types of adaptation and their potential advantages. We also highlight one key factor that influences the added efficiency provided by an adaptive design: the ratio between the recruitment length of the trial and the time taken to observe the primary endpoint(20). If it takes a long time to observe the primary endpoint, then at an interim analysis there will be a proportion of patients who do not contribute information and who don't benefit from an adaptation. As an example, if the primary outcome takes one year to observe and all patients are recruited in six months, then by the time the first patient's one-year outcome has been observed, all patients have been recruited and the adaptive design cannot provide any utility. A more quickly observed 'intermediate' outcome can be used to make adaptations, but it must be sufficiently informative for the primary outcome to be useful.

Given the amount of well-developed methodology now available for adaptive trial design, it is this consideration on the choice of primary outcome and its observation time relative to the anticipated recruitment rate, which we believe may principally influence whether an adaptive approach would provide efficiency advantages for a given IMID trial.

### **Basket and umbrella trial designs**

Because of rapid advancements in biological and genomic understanding during the past few decades, an increasing number of new therapies are being formulated to target specific molecular or immune aberrations. Given that many IMIDs share common mechanisms, these targeted therapies may perform equally well for multiple distinct IMIDs.

Originating in oncology settings, basket and umbrella trial designs have recently emerged as new types of efficient approaches for testing treatment efficacy in potentially heterogeneous subgroups(21). These novel designs are administratively efficient as they investigate multiple treatments or diseases, sometimes both, in a single study under an overarching protocol. Figure 1 gives conceptual illustrations of basket and umbrella trial designs with components (sub-studies) defined by biomarkers or genetic mutations, to which the new treatment(s) for evaluation are matched.

(Figure 1 here)

While traditional oncology trials focus on a single treatment for a specific cancer histology, basket trials can involve multiple histologies and enrol patients with a common mutation that the new therapy targets. As shown in Figure 1, an oncology basket trial consists of a number of sub-studies, with each specific to a histology or disease subtype. The principle aim is to test the treatment efficacy in various sub-studies simultaneously. As examples, Drilon et al(22) evaluated the efficacy of Larotrectinib, a tropomyosin receptor kinase inhibitor, in diverse TRK fusion positive tumours. Hyman et al(23) evaluated the BRAF inhibitor vemurafenib, finding significant activity in some tumours (e.g., non-small cell lung carcinoma (NSCLC) and Erdheim-Chester disease), yet inactivity in pancreatic cancer and multiple myeloma.

Efforts have been made to translate the idea of basket designs to disease areas outside of oncology. For example, patients can be stratified to enter a trial with multiple sub-studies by biological characteristics, such as disease stage, number of prior therapies, specific genetic/epigenetic changes, or demographic characteristics(24). There is also precedent for a basket-type approach having been used in IMiD research. Although not officially labelled a basket trial, TRANSREG(25) is a multicentre open-label trial involving 11 IMiD patient subgroups evaluating the safety, biological and clinical effects of low-dose interleukin-2. The broad eligibility criteria allow patients with rare IMiD diseases to participate in the trial.

Early strategies for analysing basket trials regard the sub-studies in isolation. Although this fully acknowledges the heterogeneity between responses to the same treatment observed in the various patient subgroups, this inevitably leads to low-powered tests due to small sample sizes. Several sophisticated approaches have been developed to enable sharing of information across sub-studies(26–29), among which the proposal by Zheng and Wason(26) can be readily applied to non-oncology basket trials with covariates. With necessary extension or modification, these approaches could lead to the efficient design and analysis of IMiD basket trials.

By contrast, umbrella designs, illustrated in Figure 1, offer the possibility to efficiently test multiple targeted therapies in a single disease population(24). To date, umbrella designs have only been implemented in oncology(30): patients of the same tumour type, as screened by an array of biomarkers, receive the treatment specific to their genetic aberration. The ongoing ALCHEMIST trial(31) represents an early example of an umbrella trial. It enrolls NSCLC patients and evaluates therapies targeting two types of genetic changes, EGFR mutations and ALK translocations, which are hypothesised as key factors to tumour growth and disease progression.

The increased understanding in pharmacogenomics and pharmacogenetics of IMiDs, especially rheumatoid arthritis(9,32), makes umbrella designs a suitable approach to answering more treatment-related questions efficiently in a single trial. The identification of specific genes and epigenetic changes involved in the development of rheumatoid arthritis, which may be predictive of the response to treatment, could potentially lead to the initiation of an umbrella trial.

With the multi-biomarker approach of umbrella trials, more patients are likely to meet eligibility criteria for at least one of the biomarker-defined subgroups. This is particularly beneficial compared to an alternative ‘enrichment’ trial that tests one targeted treatment in a subgroup. However, there are unresolved issues in how best to allocate patients who test positive for more than one biomarker, or to no biomarker, in an umbrella trial. Allocating the most suitable treatment to such patients is not straightforward.

Umbrella designs are flexible and can possibly be integrated with various adaptive designs to make them more efficient. Biomarker adaptive randomization could be incorporated to assign patients to the most promising biomarker-linked treatments using accruing trial data (e.g., as in the recent BATTLE trials(33)); a MAMS type approach could be used when a number of treatments are available for evaluation within a cohort; and if promising treatments unavailable at the start of the trial become available, protocol amendments could be made to allow addition of trial arms.

Ultimately, both basket and umbrella designs allow investigators to test more research questions in the same trial. Basket trials help assess whether a new therapy works in distinct patient subgroups (or related diseases) and to what extent(34), while umbrella trials identify whether biomarker-treatment pairs are valid and which one(s) can best improve outcomes.

### **Sequential multiple assignment randomised trial (SMART) designs**

Therapy of chronic conditions or rapidly fatal diseases often requires several lines of treatment with different drugs or interventions used as the disease progresses. In each line, the treatment may achieve the required clinical objective (e.g., response), or not (e.g., non-response). When treatment fails for a patient at a certain line, it is common medical practice to switch to a different treatment or strategy for the next line. The type or dose of the treatment/intervention may be adjusted repeatedly according to a patient's ongoing clinical information, including their treatment history and response to previous treatments(35,36).

An adaptive intervention is a treatment strategy that personalises treatment through established decision rules that recommend when and how the treatment changes, taking into account the history of previous treatments and response to those treatments(37). A Sequential Multiple Assignment Randomised Trial (SMART) is a multistage trial design that is used to construct effective dynamic treatment regimens (DTR), also known as adaptive interventions (AIs) or adaptive treatment strategies(38). Figure 2 depicts an example of a SMART design in which only non-responders to first stage intervention are re-randomised in the second stage. This would provide information to inform an AI that chooses which first-line intervention to use, and how to subsequently treat patients who do not respond to the first-line treatment.

(Figure 2 here)

An AI consists of four key elements: critical decision point(s), intervention component(s), tailoring variable(s), and decision rule(s). The first element, a sequence of *critical decision point(s)*, comprises the intervention to begin with, when and how to measure signs of response/nonresponse, how to maintain the success of the initial intervention, and what interventions may be used for non-responders. The second element, the *intervention components*, is a set of intervention/treatment options at each critical decision point. From Figure 2 we can see that there are two treatments options in the first stage (treatment A and B), and six treatment options in the second stage (two options for responders, and four options for no-responders). The third element is the *tailoring variable(s)*. A tailoring variable is an early indicator of the overall outcome (success or failure of the intervention). The response status at week 24 plays the role of the tailoring variable in the example shown in Figure 2. Lastly, the *decision rules* occurring at each critical decision point link the tailoring variable(s) to the intervention components. Each stage in a SMART corresponds to one of the critical decisions involved in the adaptive intervention. Each participant moves through the multiple stages, and at each stage the participant is randomly (re)assigned to one of several intervention options(35,39). Each AI can be summarized in the form (X1;X2:X3) where X1 is the recommended first-stage treatment, X2 the recommended second-stage treatment for responders, and X3 the recommended second-stage treatment for non-responders. There are four

different adaptive interventions embedded in the SMART depicted in Figure 2: (A,A,C),(A,A,D),(B,B,E), and (B,B,F).

SMARTs have been used for a wide range of chronic conditions, including some IMIDs. Recent studies that have used them include the CATIE study of treatments for schizophrenia(40), the EXTEND trial of treatments for alcohol dependence(41), and studies of treatments for metastatic renal cell carcinoma(42), depression(43), HIV infection(44,45), ulcerative colitis(46), autoinflammatory recurrent fever syndromes(47), psoriasis(48–50), and rheumatoid arthritis(51).

An alternative design to a SMART study is the use of “multiple one-stage-at-a-time” randomised trials. This design considers each critical decision point as an independent trial(39). For instance, from the SMART in Figure 2, there are three different “one-stage-at-a-time” trials. The first trial would correspond to the first stage treatment options. The second trial would study treatment in non-responders to treatment A, and the third trial would study treatment in non-responders to treatment B. One advantage of the SMART design over the “multiple one-stage-at-a-time” is that it uses information from all stages to find the best AI. To do this, it uses Q-Learning; a multistage regression method that can use data from a SMART study to examine whether and how certain variables are suitable to develop an AI or improve an existing one.

SMARTs are not without limitation, however. In particular, some issues arise from modelling data from SMARTs when the estimation of the optimal AI is of interest. These include model building, missing data, statistical inference, and choosing an outcome when only non-responders are re-randomised(36). The fact that the re-randomisation depends on the evolving patient status, along with the sequential design nature of the SMART, bring more complexities to the handling of missing data compared to classical clinical trials. For instance, in a SMART study where only non-responders are re-randomised at the second stage, a patient who is lost to follow-up during the first stage will have missing information on their intermediate response status, second stage treatment, and outcome. It is not possible to know whether the information in the second stage is truly missing or is missing by design since it depends on an unobserved patient response status. Furthermore, the use of flexible regression approaches to avoid complex functions in the Q-learning approach can also make it difficult to acquire interpretable results and valid statistical inference due to potential high variability(36).

SMARTs provide a lot of potential utility to chronic IMIDs, where the most suitable AI is of interest.

### **Use of high-dimensional data to stratify patients: Adaptive signature trial designs**

It is common in clinical trials that only a subgroup of treated patients may benefit from an experimental therapy(52–55). Identifying these subgroups would allow tailoring of treatment, avoiding costly or toxic treatment of individuals who will not benefit. To identify such subgroups, predictive biomarkers are required. Predictive biomarkers are biomarkers (objective characteristics associated with some aspect of a patient’s function or health) that are associated with the response to treatment. If one has a predictive biomarker, this can be used to predict the likely response to treatment. Some clinical areas, such as



oncology, have strong availability of predictive biomarkers. For example, the RAS-mutation identified a subgroup of patients with a significant benefit across all efficacy endpoints associated with treatment for colorectal cancer(56).

However, predictive biomarkers are lacking for most IMIDs, meaning predicting response to treatment is more difficult(57–59). For example, in rheumatoid arthritis although genetic variants associated with response to methotrexate have been identified(60–63), there is a lack of consensus on the predictive utility of these variants.

In the absence of predictive biomarkers, alternative methods that utilise high-dimensional information could be used. With the rapid development of new next generation sequencing, proteomics, and medical imaging technologies, a large amount of high-dimensional data about patients is starting to be collected in clinical trials. There is the potential for this information to be informative for identifying subgroups of patients who are likely to benefit from a new treatment.

To utilize high-dimensional information in RCTs, a method has been developed known as the adaptive signature design (ASD). The aim of the ASD is to allow a single RCT to both test the overall treatment effect in all patients and to form a predictive biomarker signature that predicts a subgroup of patients who strongly benefit from the treatment. Although the ASD has 'adaptive' in its name, it is not actually an adaptive design as it does not change anything about the trial.

The original method(64,65) utilised (high-dimensional) gene expression data in an oncology setting, but it can be used in any case where heterogeneity in the treatment effect is expected and there is high-dimensional information available. Which of the high-dimensional data should be included in the signature is determined by imposing a threshold on the significance level, odds ratios, and number of biomarkers. Further papers have proposed modifications of the original ASD(66–68) to provide improved performance (in terms of correctly identifying a subgroup who benefit from treatment). In these methods, the high-dimensional data is used to form a signature that is computed based on the interaction between these data with the treatment. The adaptive signature is represented by a single score for each patient. The scores can then be utilised to divide the patients into subgroups using a variety of clustering techniques, or as covariates in the tests of association with the outcome. The test for the overall comparison between the arms can be performed by testing for the difference between the arms in the trial population (at the significance level  $\alpha_1$ ) and testing for the difference between the arms in the subgroup (at significance level  $\alpha_2$ ). The overall significance level of the trial is then controlled at the  $\alpha = \alpha_1 + \alpha_2$  level (Figure 3).

Figure 3 here

In conclusion, ASDs are a novel methodology that can develop and validate predictive signatures in a single trial. They have the potential to increase the efficiency of clinical trials by finding the group of patients benefiting from particular treatments. However, when the clinical benefit for a subgroup is minimal, a large sample size might be required to detect it with sufficient power. Additionally, the

performance of the designs deteriorates if there are many covariates that are not associated with patient benefit. To address this issue, an additional pre-filtering of the covariates might be required. This family of designs may also benefit from exploring different methods of interaction of treatment with high dimensional covariates(69), and from considering multiple trial endpoints(70). These considerations notwithstanding, ASDs offer a potential route to identifying patient subgroups that will benefit from treatment in IMIDs for which predictive biomarkers are currently lacking.

### **Composite responder endpoints and augmented analysis methods**

Clinical trials specify primary and secondary outcomes that measure how patients respond to a treatment or intervention. The primary outcome should be chosen as a measurement that will be more favourable if the treatment being tested is efficacious or effective. As many IMIDs have complex manifestations and multiple symptoms, it can be difficult to specify a single measurement as being the most important. For this reason, it is common that primary outcomes in IMID trials combine multiple relevant measurements into a single composite outcome. A specific type of composite endpoint is a responder endpoint, which divides patients into responders and non-responders based on different measurements, or components. Some of these components can be binary and others may be whether continuous measurements are above a threshold.

The standard method of analysis for composite responder endpoints is to treat them as binary variables (responder or non-responder). The analysis then estimates the proportion of patients who are responders and whether there is a significant difference between arms: this is done with a suitable binary method such as Fisher's exact test or logistic regression, amongst many others.

Responder endpoints have the appealing property of summarising very complex information into an easy-to-interpret single quantity. This is also a limitation when applying analysis methods that treat the outcome as binary: much information is discarded, especially from continuous components when dichotomising (see, e.g.,(71,72)) which can lead to a reduction in power(73).

Assuming that the responder endpoint is clinically relevant, there are alternative ways of estimating the proportion of patients who are responders. For endpoints that define response based on a single continuous component, methods were proposed in the 1990s to more precisely estimate the proportion of responders(74,75). For composite responder endpoints that are a mixture of continuous and binary components, the augmented binary method has been proposed to provide higher efficiency. This was originally proposed for response criteria endpoints used in phase II oncology trials(76) but has since been extended to endpoints used in IMIDs such as rheumatoid arthritis(77) and systemic lupus erythematosus (SLE)(78). The method has also been extended to endpoints that are formed from the time until a composite event occurs(79) (e.g., time until relapse, where relapse involves a continuous biomarker being above a certain level), although further work in this area is needed.

The augmented binary method requires no additional data to be collected; it simply fits a more complex statistical model to the data collected on the different components and uses this model to estimate the

difference between arms in the proportion of responders (together with a confidence interval and p-value). It has been shown in various papers(77,78,80,81) to provide large gains in efficiency, equivalent to applying the traditional binary analysis with a sample size of 30% or more higher. The extent of the increase of efficiency depends on to what extent the continuous component(s) distinguish between responders and non-responders(82).

A previous review(83) found that several IMID conditions used composite responder outcomes. We show some examples of these in Table 2.

## **Current Use Of Innovative Methods In Immune-mediate Inflammatory Disease Trials**

### **Review methods**

To investigate the frequency with which innovative methods have been used in IMID trials in recent years, we searched PubMed on June 18 2020. We restricted our evaluation to clinical trial publications that have appeared since 2018 in any of five high impact factor journals relevant to IMIDs (*New Engl J Med, Lancet, Ann Rheum Dis, Arthritis Rheumatol, J Am Acad Dermatol*). To provide a comprehensive evaluation, we included articles containing any of 51 IMID disease terms. See the Supplementary Materials for the search term. This search returned 160 articles for review.

Each article was reviewed by JMSW to establish whether it met the inclusion criteria: that the article was a primary report of the results of a clinical trial conducted to evaluate the efficacy of one or more treatments for one or more IMIDs. Retrospective trial analyses were thus excluded, as our focus was on how innovative methods have been used in practice in the design and analysis of IMID trials. For each article deemed eligible for inclusion, data was extracted by JMSW for 21 questions relating to the trial's design and analysis, and in particular the use of innovative methods (see Supplementary Table 1). Owing to the objective nature of the extraction questions, high reproducibility on evaluation of inclusion and subsequent data extraction was anticipated. Nonetheless, ten articles were randomly chosen for duplicate review by MJG. The authors agreed on inclusion for all ten articles. Agreement on extracted data was 95%. See the Supplementary Materials for further details.

### **Findings**

Ninety-seven articles were deemed to be eligible for inclusion. A summary of the extracted data for these 97 articles is given in Table 3.

While more than 20 distinct conditions were evaluated in the eligible trials, the plurality (31%) found were in rheumatoid arthritis. Notable numbers were also found in psoriatic arthritis, psoriasis, and SLE. The majority of trials (75%) were funded and sponsored by industry.

Most (65%) eligible trials had two arms. Some rarer conditions used single arm trials with no prospective control arm. In other cases, more than two arms were included: in most instances this was for industry-funded trials of a new drug, with different doses or regimens included as distinct arms. We did not identify any MAMS trials.

There was some reported use of innovative approaches (19.6%). These consisted predominantly of group-sequential designs (or a futility analysis), sample-size re-assessment, and re-randomising some participants as in a SMART design. For re-randomisations, we did not find any examples where an analysis was performed to determine the best AI. The median recruitment length was 96 weeks and primary endpoint length was 24 weeks. This indicates that for a majority of trials the ratio of endpoint length to recruitment length would be sufficiently low for an adaptive design to provide efficiency(84).

In a majority of trials (60%), patients with other autoimmune diseases were not eligible for the trial. In other cases, this was not an explicit exclusion criteria but it is likely that such patients would be indirectly excluded through criteria such as being naïve to therapies that are commonly used for other IMIDs.

We found very few examples where collection of high-dimensional data was reported (8.2%). In the eight trials that did report this, the most common approach was to analyse each variable separately. Reported use of routinely collected data in the design of the trial was also low.

The use of responder endpoints (involving dichotomization of continuous measurements) was very high. The majority of trials (68%) had a primary endpoint that was defined in this way; an even higher proportion (84%) had a responder endpoint as a secondary outcome. These endpoints were routinely analysed using standard methods, such as a Cochrane-Mantel-Haenszel or Fisher exact test.

## Discussion

In this paper we have provided an overview of innovative methods that could provide utility to IMID trials. We have also shown that few recently reported trials are utilizing innovative approaches through a literature review.

Although 19.6% of included trials used some approach that we classified as innovative, most of these were relatively straightforward approaches, such as a futility analysis or having a second randomization of non-responding patients (without applying techniques for analysing SMARTs.) There is a high potential for more advanced innovative approaches to be used in future IMID trials, but this requires improved awareness, education, and software.

One notable finding was that it was very common, amongst multiple distinct IMIDs, for trial endpoints to be responder endpoints. Over two-thirds of trials had such an endpoint as the primary, and almost 90% had a secondary endpoint. In every case the endpoint was analysed as if it were binary. As we have described, there are much more efficient analysis methods available and it is important for them to be made available for use in practice. Some freely-available software is currently available(85) but there is

the need for more generic software and methods that can be used across all such endpoints used in IMID trials.

Presently, it appears that collection of high dimensional information and use of routinely collected data is rare in IMID trials. A limitation of our review is that we may have missed use of this from just examining primary reports of RCTs. For example it may be common for high-dimensional information to be collected but reported in secondary analysis papers. In addition it may not be felt a worthwhile use of space in a primary report of an RCT to discuss how routinely collected data was used to inform the trial design.

The majority of trials were sponsored and funded by industry. Although there were uses of innovative approaches in industry sponsored trials, use of more advanced methods that we have discussed in this paper could be hampered by regulatory issues (either actual or perceived). For use of some more advanced designs and analysis approaches in confirmatory trial settings, it will be important to ensure they are supported by regulators.

In conclusion, IMID trials could substantially benefit from use of more innovative approaches that we have reviewed in this paper. Further research, better software, and more dissemination is needed to ensure all IMID trials, that could benefit, do so.

## List Of Abbreviations

AI: Adaptive intervention

ASD: Adaptive Signature Design

DTR: Dynamic treatment regime

IMID: Immune-mediated inflammatory disease

MAMS: Multi-arm multi-stage

NSCLC: Non-small cell lung carcinoma

RCT: Randomised controlled trial

SLE: Systemic Lupus Erythematosus

SMART: Sequential multiple assignment randomised trial

## Declarations

**Ethics approval and consent to participate:** Not applicable.

**Consent for publication:** Not applicable.

**Availability of data and materials:** All data generated or analysed during this study are included in this published article [and its supplementary information files].

**Competing interests:** The authors declare that they have no competing interests.

**Funding:** JMSW is funded by the Medical Research Council (MC\_UU\_00002/6 and MR/N028171/1). Funding bodies played no role in the design, collection, analysis or interpretation of the data.

**Authors' contributions:** JMSW and MJG conceived the idea for the article. JMSW performed the literature review. MJG performed the analysis of the review data. All authors contributed to writing the manuscript. All authors read and approved the final manuscript.

**Acknowledgements:** Not applicable.

## References

1. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. Vol. 4, Autoimmunity Reviews. Elsevier; 2005. p. 130–6.
2. Langley RGB, Krueger GG, Griffiths CEM. Psoriasis: Epidemiology, clinical features, and quality of life. In: Annals of the Rheumatic Diseases. BMJ Publishing Group; 2005. p. ii18.
3. Calamia KT, Wilson FC, Icen M, Crowson CyS, Gabriel SE, Kremers HM. Epidemiology and clinical characteristics of behcet's disease in the us: A population-based study. *Arthritis Care Res.* 2009 May 15;61(5):600–4.
4. Kuek A, Hazleman BL, K O AJ. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. *Postgr Med J.* 2007;83:251–60.
5. El-Gabalawy H, Guenther LC, Bernstein CN. Epidemiology of immune-mediated inflammatory diseases: Incidence, prevalence, natural history, and comorbidities. Vol. 37, *Journal of Rheumatology.* 2010. p. 2–10.
6. Winthrop KL, Weinblatt ME, Bathon J, Burmester GR, Mease PJ, Crofford L, et al. Unmet need in rheumatology: Reports from the Targeted Therapies meeting 2019. Vol. 79, *Annals of the Rheumatic Diseases.* BMJ Publishing Group; 2020. p. 88–93.
7. Blaess J, Walther J, Gottenberg JE, Sibilia J, Arnaud L, Felten R. AB0332 immunosuppressive and immunomodulating agents in rheumatoid arthritis: a systematic review of clinical trials and their current development stage. *Ann Rheum Dis.* 2020 Jun 1;79(Suppl 1):1464–5.
8. Solomon DH, Bitton A, Katz JN, Radner H, Brown EM, Fraenkel L. Review: Treat to target in rheumatoid arthritis: Fact, fiction, or hypothesis? Vol. 66, *Arthritis and Rheumatology.* John Wiley and Sons Inc.; 2014. p. 775–82.
9. Ktak A, Paradowska-Gorycka A, Kwiatkowska B, Raciborski F. Personalized medicine in rheumatology. *Reumatologia.* 2016;54(4):177–86.

10. Sertkaya A, Wong H-H, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin Trials*. 2016 Apr 8;13(2):117–26.
11. CALIBER | UCL Institute of Health Informatics - UCL – University College London [Internet]. [cited 2020 Jul 2]. Available from: <https://www.ucl.ac.uk/health-informatics/caliber>
12. Research - Research units A-Z - Immune-Mediated Inflammatory Disease Biobanks - UK [Internet]. [cited 2020 Jul 2]. Available from: <https://www.gla.ac.uk/research/az/imid/>
13. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: Example using statins for preventing progression of diabetes. *BMJ*. 2010 Apr 24;340(7752):907–11.
14. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med*. 2019;25(10):1601–6.
15. Zhao SS, Zhao SS, Lyu H, Lyu H, Solomon DH, Solomon DH, et al. Improving rheumatoid arthritis comparative effectiveness research through causal inference principles: Systematic review using a target trial emulation framework. *Ann Rheum Dis*. 2020;79(7):883–90.
16. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758–64.
17. Dimairo M, Pallmann P, Wason J, Todd S, Jaki T, Julious SA, et al. The Adaptive designs CONSORT Extension (ACE) statement: a checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ*. 2020 Jun 17;369.
18. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson L V., et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med*. 2018 Dec 28;16(1):29.
19. Buch MH, Pavitt S, Parmar M, Emery P. Creative trial design in RA: Optimizing patient outcomes. Vol. 9, *Nature Reviews Rheumatology*. 2013. p. 183–94.
20. Wason JMS, Brocklehurst P, Yap C. When to keep it simple – adaptive designs are not always useful. *BMC Med*. 2019 Dec;17(1).
21. Renfro LA, Sargent DJ. Statistical controversies in clinical research: Basket trials, umbrella trials, and other master protocols: A review and examples. *Ann Oncol*. 2017;28(1):34–43.
22. Drilon A, Laetsch TW, Kummar S, Dubois SG, Lassen UN, Demetri GD, et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *N Engl J Med*. 2018 Feb;378(8):731–9.
23. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N Engl J Med*. 2015 Aug;373(8):726–36.
24. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. Drazen JM, Harrington DP, McMurray JJV, Ware JH, Woodcock J, editors. Vol. 377, *New England Journal of Medicine*. Massachusetts Medical Society; 2017. p. 62–70.
25. Rosenzweig M, Lorenzon R, Cacoub P, Pham HP, Pitoiset F, El Soufi K, et al. Immunological and clinical effects of low-dose interleukin-2 across 11 autoimmune diseases in a single, open clinical trial. *Ann Rheum Dis*. 2019;78(2):209–17.

26. Zheng H, Wason JMS. Borrowing of information across patient subgroups in a basket trial based on distributional discrepancy. *Biostatistics*. 2020 Aug 14;
27. Chu Y, Yuan Y. A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clin Trials*. 2018 Apr;15(2):149–58.
28. Hobbs BP, Landin R. Bayesian basket trial design with exchangeability monitoring. *Stat Med*. 2018 Nov;37(25):3557–72.
29. Psioda MA, Xu J, Jiang QI, Yang Z, Ibrahim JG. Bayesian adaptive basket trial design using model averaging. *Biostatistics*. 2019;1–16.
30. Park JJH, Siden E, Zoratti MJ, Dron L, Harari O, Singer J, et al. Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols. *Trials*. 2019 Dec 18;20(1):572.
31. Gerber DE, Oxnard GR, Govindan R. ALCHEMIST: Bringing genomic discovery and targeted therapies to early-stage lung cancer. *Clin Pharmacol Ther*. 2015;97(5):447–50.
32. Umićević Mirkov M, Coenen MJH. Pharmacogenetics of disease-modifying antirheumatic drugs in rheumatoid arthritis: Towards personalized medicine. Vol. 14, *Pharmacogenomics*. Future Medicine Ltd London, UK ; 2013. p. 425–44.
33. Liu S, Lee JJ. An overview of the design and conduct of the BATTLE trials. *Chinese Clin Oncol*. 2015;4(3):1–13.
34. Cunanan KM, Gonen M, Shen R, Hyman DM, Riely GJ, Begg CB, et al. Basket trials in oncology: A trade-off between complexity and efficiency. Vol. 35, *Journal of Clinical Oncology*. American Society of Clinical Oncology; 2017. p. 271–3.
35. Zhong X, Cheng B, Qian M, Cheung YK. A gate-keeping test for selecting adaptive interventions under general designs of sequential multiple assignment randomized trials. *Contemp Clin Trials*. 2019 Oct 1;85:105830.
36. Zhao YQ, Laber EB. Estimation of optimal dynamic treatment regimes. In: *Clinical Trials*. SAGE Publications Ltd; 2014. p. 400–7.
37. Lavori PW, Dawson R. Adaptive Treatment Strategies in Chronic Disease. *Annu Rev Med*. 2008 Feb;59(1):443–53.
38. Nahum-Shani I, Ertefaie A, Lu XL, Lynch KG, McKay JR, Oslin DW, et al. A SMART data analysis method for constructing adaptive treatment strategies for substance use disorders. *Addiction*. 2017 May 1;112(5):901–9.
39. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A “SMART” Design for Building Individualized Treatment Sequences. *Annu Rev Clin Psychol*. 2012 Apr 27;8(1):21–48.
40. Stroup TS, McEvoy JP, Swartz MS, Byerly MJ, Glick ID, Canive JM, et al. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Project: Schizophrenia Trial Design and Protocol Development. *Schizophr Bull*. 2003 Jan 1;29(1):15–31.



41. Managing Alcoholism in People Who Do Not Respond to Naltrexone - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jul 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT00115037>
42. Sequential Two-agent Assessment in Renal Cell Carcinoma Therapy: The START Trial - Full Text View - ClinicalTrials.gov [Internet]. [cited 2020 Jul 28]. Available from: <https://clinicaltrials.gov/ct2/show/NCT01217931>
43. Schulte PJ, Tsiatis AA, Laber EB, Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Stat Sci*. 2014;29(4):640–61.
44. Moodie EEM, Richardson TS, Stephens DA. Demystifying Optimal Dynamic Treatment Regimes. *Biometrics*. 2007 Jun 1;63(2):447–55.
45. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int J Biostat*. 2010;6(2).
46. Sands BE, Sandborn WJ, Panaccione R, O'Brien CD, Zhang H, Johanns J, et al. Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. *N Engl J Med*. 2019 Sep 26;381(13):1201–14.
47. De Benedetti F, Gattorno M, Anton J, Ben-Chetrit E, Frenkel J, Hoffman HM, et al. Canakinumab for the Treatment of Autoinflammatory Recurrent Fever Syndromes. *N Engl J Med*. 2018 May 17;378(20):1908–19.
48. Reich K, Gooderham M, Thaçi D, Crowley JJ, Ryan C, Krueger JG, et al. Risankizumab compared with adalimumab in patients with moderate-to-severe plaque psoriasis (IMMvent): a randomised, double-blind, active-comparator-controlled phase 3 trial. *Lancet*. 2019 Aug 17;394(10198):576–86.
49. Mrowietz U, Bachelez H, Burden AD, Rissler M, Sieder C, Orsenigo R, et al. Secukinumab for moderate-to-severe palmoplantar pustular psoriasis: Results of the 2PRECISE study. *J Am Acad Dermatol*. 2019 May 1;80(5):1344–52.
50. Lebwohl M, Blauvelt A, Paul C, Sofen H, Węglowska J, Piguet V, et al. Certolizumab pegol for the treatment of chronic plaque psoriasis: Results through 48 weeks of a phase 3, multicenter, randomized, double-blind, etanercept- and placebo-controlled study (CIMPACT). *J Am Acad Dermatol*. 2018 Aug 1;79(2):266-276.e5.
51. Weinblatt ME, Baranauskaite A, Niebrzydowski J, Dokoupilova E, Zielinska A, Jaworski J, et al. Phase III Randomized Study of SB5, an Adalimumab Biosimilar, Versus Reference Adalimumab in Patients With Moderate-to-Severe Rheumatoid Arthritis. *Arthritis Rheumatol*. 2018 Jan 1;70(1):40–8.
52. Rothenberg ML, Carbone DP, Johnson DH. Improving the evaluation of new cancer treatments: Challenges and opportunities. Vol. 3, *Nature Reviews Cancer*. European Association for Cardio-Thoracic Surgery; 2003. p. 303–9.
53. Foster JC, Taylor JMG, Ruberg SJ. Subgroup identification from randomized clinical trial data. *Stat Med*. 2011 Oct 30;30(24):2867–80.

54. Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively selecting a target population for a future comparative study. *J Am Stat Assoc.* 2013;108(502):527–39.
55. Janes H, Brown MD, Crager MR, Miller DP, Barlow WE. Adjusting for covariates in evaluating markers for selecting treatment, with application to guiding chemotherapy for treating estrogen-receptor-positive, node-positive breast cancer. *Contemp Clin Trials.* 2017 Dec 1;63:30–9.
56. Van Cutsem E, Lenz HJ, Köhne CH, Heinemann V, Tejpar S, Melezínek I, et al. Fluorouracil, leucovorin, and irinotecan plus cetuximab treatment and RAS mutations in colorectal cancer. *J Clin Oncol.* 2015 Mar 1;33(7):692–700.
57. Brown PM, Pratt AG, Isaacs JD. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. Vol. 12, *Nature Reviews Rheumatology.* Nature Publishing Group; 2016. p. 731–42.
58. Robinson WH, Mao R. Biomarkers to guide clinical therapeutics in rheumatology? Vol. 28, *Current Opinion in Rheumatology.* Lippincott Williams and Wilkins; 2016. p. 168–75.
59. Pouw J, Leijten E, Radstake T, Boes M. Emerging molecular biomarkers for predicting therapy response in psoriatic arthritis: A review of literature. *Clin Immunol.* 2020 Feb 1;211:108318.
60. Aslibekyan S, Brown EE, Reynolds RJ, Redden DT, Morgan S, Baggott JE, et al. Genetic variants associated with methotrexate efficacy and toxicity in early rheumatoid arthritis: Results from the treatment of early aggressive rheumatoid arthritis trial. *Pharmacogenomics J.* 2014 Feb;14(1):48–53.
61. Senapati S, Singh S, Das M, Kumar A, Gupta R, Kumar U, et al. Genome-wide analysis of methotrexate pharmacogenomics in rheumatoid arthritis shows multiple novel risk variants and leads for TYMS regulation. *Pharmacogenet Genomics.* 2014;24(4):211–9.
62. Kung TN, Dennis J, Ma Y, Xie G, Bykerk V, Pope J, et al. RFC1 80G>A is a genetic determinant of methotrexate efficacy in rheumatoid arthritis: A human genome epidemiologic review and meta-analysis of observational studies. *Arthritis Rheumatol.* 2014;66(5):1111–20.
63. Morgan MD, Al-Shaarawy N, Martin S, Robinson JI, Twigg S, Magdy AA, et al. MTHFR functional genetic variation and methotrexate treatment response in rheumatoid arthritis: A meta-analysis. *Pharmacogenomics.* 2014 Mar 1;15(4):467–75.
64. Simon RM, Freidlin B. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res.* 2005 Nov;11(21):7872–8.
65. Freidlin B, Jiang W, Simon R. The cross-validated adaptive signature design. *Clin Cancer Res.* 2010 Jan 15;16(2):691–8.
66. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol.* 2002 Jul 5;9(3):505–11.
67. Matsui S, Simon R, Qu P, Shaughnessy JD, Barlogie B, Crowley J. Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine. *Clin Cancer Res.* 2012 Nov 1;18(21):6065–73.

68. Cherlin S, Wason JMS. Developing and testing high-efficacy patient subgroups within a clinical trial using risk scores. *Stat Med*. 2020 Jul 14;sim.8665.
69. Callegaro A, Spiessens B, Dizier B, Montoya FU, van Houwelingen HC. Testing interaction between treatment and high-dimensional covariates in randomized clinical trials. *Biometrical J*. 2017 Jul 1;59(4):672–84.
70. Cherlin S, Wason JMS. Developing a predictive signature for two trial endpoints using the cross-validated risk scores method. 2020 Jul 3;
71. Senn S. Disappointing dichotomies. *Pharm Stat*. 2003 Oct;2(4):239–40.
72. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.
73. Wason JMS, Mander AP, Eisen TG. Reducing sample sizes in two-stage phase II cancer trials by using continuous tumour shrinkage end-points. *Eur J Cancer*. 2011;
74. Suissa S. Binary methods for continuous outcomes: a parametric alternative. *J Clin Epidemiol*. 1991;44:241–8.
75. Suissa S, Blais L. Binary regression with continuous outcomes. *Stat Med*. 1995;14:247–55.
76. Wason JMS, Seaman SR. Using continuous data on tumour measurements to improve inference in phase II cancer studies. *Stat Med*. 2013;32(26):4639–50.
77. Wason JMS, Jenkins M. Improving the power of clinical trials of rheumatoid arthritis by using data on continuous scales when analysing response rates: An application of the augmented binary method. *Rheumatol (United Kingdom)*. 2016;
78. McMenamin M, Barrett JK, Berglind A, Wason JMS. Employing latent variable models to improve efficiency in composite endpoint analysis. 2019 Feb 19;arXiv:1902.07037.
79. Lin C-J, Wason JMS. Efficient analysis of time-to-event endpoints when the event involves a continuous variable crossing a threshold. *J Stat Plan Inference*. 2020 Sep;
80. Wason JMS, Seaman SR. Using continuous data on tumour measurements to improve inference in phase II cancer studies. *Stat Med*. 2013;
81. Lin C-J, Wason JMS. Improving phase II oncology trials using best observed RECIST response as an endpoint by modelling continuous tumour measurements. *Stat Med*. 2017;
82. McMenamin M, Barrett JK, Berglind A, Wason JMS. Sample Size Estimation using a Latent Variable Model for Mixed Outcome Co-Primary, Multiple Primary and Composite Endpoints. 2019 Dec 11;ArXiv:1912.05258.
83. Wason J, McMenamin M, Dodd S. Analysis of responder-based endpoints: improving power through utilising continuous components. 2020.
84. Wason JMS, Brocklehurst P, Yap C. When to keep it simple - adaptive designs are not always useful. *BMC Med*. 2019 Aug;
85. McMenamin M, Grayling MJ, Berglind A, Wason JM. Increasing power in the analysis of responder endpoints in rheumatology: a software tutorial. *medRxiv*. 2020 Jul 29;2020.07.28.20163378.

# Tables

Table 1  
An overview of various types of adaptive design and their benefits.

<b>Adaptive design</b>	<b>Description</b>	<b>Benefits</b>
Group-sequential	Allows a trial to be stopped early for efficacy, futility, or safety, when there is enough evidence to justify doing so.	On average, the sample size that would be required by a trial is reduced; particularly for those with strong treatment effects.
Response adaptive randomisation	Allows the treatment allocation ratio(s) to be altered as the trial progresses.	Allocation can be skewed in favour of the treatment arm that appears to have higher efficacy; meaning more patients are expected to respond in the trial.
Multi-arm multi-stage (MAMS)	Allows multiple treatments to be evaluated in a single trial. Interim analyses allow less promising treatments to be removed from the trial early.	Highly efficient for evaluating multiple treatments at once.
Sample size re-assessment	Allows the sample size to be modified in response to the outcome variation or treatment effect observed in the interim.	The trial is more likely to be powered at the desired level, especially when there is limited data to inform a sample size calculation.
Biomarker adaptive	Allows the trial's population to be adjusted to avoid enrolling patients who don't benefit from a treatment; typically this involves incorporating information from, or adapting on, a biomarker.	Patient subgroups who will benefit most from particular treatments can be identified and prioritized.
Platform trial	Allows treatments to be added in to an ongoing trial. Typically involves several treatments being evaluated under an overarching protocol.	Efficient for evaluating multiple treatments as new ones become available over time.

Table 2  
Examples of composite responder endpoints used in IMID trials.

IMID	Endpoint	Definition
Ankylosing spondylitis	ASAS20 response	<ul style="list-style-type: none"> <li>• 20% improvement and <math>\geq 10</math> units of change (on a 0–100 scale) in each of 3 domains</li> <li>• No worsening of a similar amount in the fourth domain</li> <li>• (Components are physical function, pain, inflammation and patient’s global assessment)</li> </ul>
Crohn’s disease	Clinical remission	<ul style="list-style-type: none"> <li>• Crohn’s Disease Activity Index below a threshold (e.g., 150)</li> <li>• No use of steroids or rescue treatment</li> </ul>
Idiopathic arthritis-associated uveitis	Best corrected visual acuity above threshold and no light perception	<ul style="list-style-type: none"> <li>• Best-corrected visual acuity, thresholds <math>\leq 20/50</math>, <math>\leq 20/200</math></li> <li>• No light perception</li> <li>• Contribution of amblyopia, yes/no</li> </ul>
Juvenile arthritis	Response	<p>Improvement by 30% in at least 3 of:</p> <ul style="list-style-type: none"> <li>• MD global assessment</li> <li>• parent or patient global assessment</li> <li>• functional ability</li> <li>• number of joints with active arthritis</li> <li>• number of joints with limited range of motion</li> <li>• Erythrocyte Sedimentation Rate</li> </ul>
Juvenile dermatomyositis	Responder index	<ul style="list-style-type: none"> <li>• <math>\geq 4</math> point reduction from baseline in safety of estrogen in lupus national assessment (SELENA) systemic lupus erythematosus disease activity index (SLEDAI) score</li> <li>• No worsening (increase of <math>&lt; 0.30</math> points from baseline) in physician’s global assessment (PGA)</li> <li>• No new British Isles Lupus Assessment Group of SLE clinics (BILAG) A organ domain score or 2 new BILAG B organ domain scores compared with baseline</li> </ul>
Nonalcoholic steatohepatitis	Resolution of steatohepatitis without fibrosis	<ul style="list-style-type: none"> <li>• Improvement in NAS of two points</li> <li>• No worsening of fibrosis</li> </ul>
Sjogren's syndrome	Response	<ul style="list-style-type: none"> <li>• <math>&gt; 30\%</math> reduction in analog scales evaluating dryness, pain and fatigue</li> </ul>

Table 3

Summary of extracted data for the 97 included articles. The denominator for computing percentages (given to 1 decimal place) is 97 unless stated otherwise.

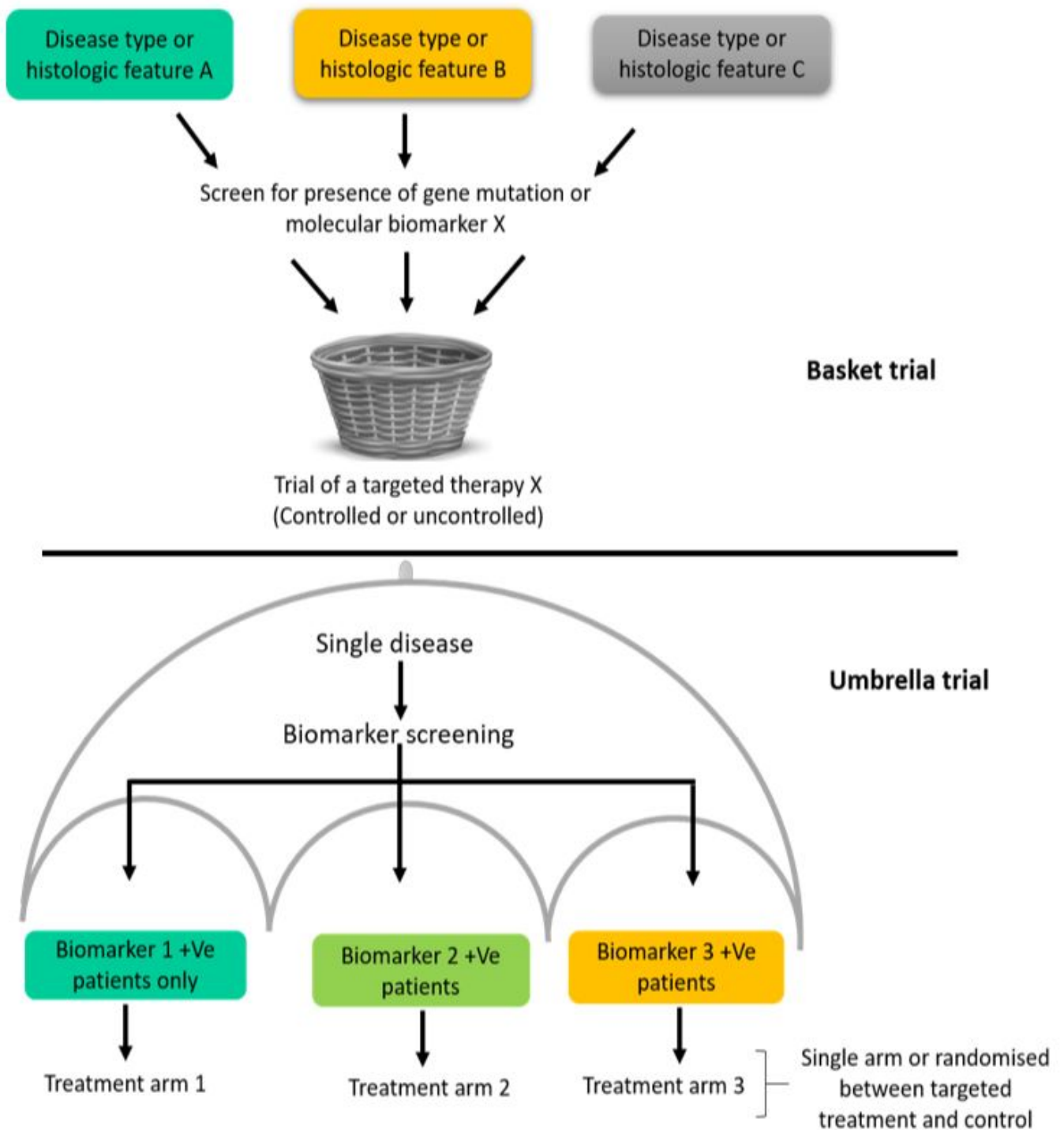
Question	n (%)
What immune-mediated inflammatory disease(s) was the trial conducted in?	30 (30.9)
Rheumatoid arthritis	10 (10.3)
Systemic lupus erythematosus	9 (9.3)
Psoriasis	9 (9.3)
Psoriatic arthritis	4 (4.1)
Juvenile idiopathic arthritis	4 (4.1)
Multiple sclerosis	3 (3.1)
Sjögren's syndrome	3 (3.1)
Systemic sclerosis	25 (25.8)
Other (see Supplementary Materials)	
How many treatment arms were in the trial?	5 (5.2)
1	63 (64.9)
2	13 (13.4)
3	11 (11.3)
4	2 (2.1)
5	3 (3.1)
6	
What was the total planned sample size according to the sample size calculation?	Median: 214 IQR: [113, 400] Range: [10, 5400]
How was the trial funded?	73 (75.3)
Industry	16 (16.5)
Academic	4 (4.1)
Mixed	3 (3.1)
Not reported	1 (1.0)
Charity	

Question	n (%)
Was any innovative design used? <sup>1</sup>	19 (19.6)
Yes	7 (7.2)
Group-sequential design/futility interim analysis	6 (6.2)
Sequential multiple assignment randomised trial design	4 (4.1)
Bayesian methods used	2 (2.1)
Sample size re-estimation	1 (1.0)
Basket trial design	
Did the trial design report the involvement of any evidence from routinely collected data, cohorts, or biobanks?	2 (2.1)
What was the length of patient recruitment (in weeks)? <sup>2</sup>	Median: 96 IQR: [55, 120] Range: [16, 296]
What was the primary endpoint timepoint (in weeks)? <sup>3</sup>	Median: 24 IQR: [12, 48] Range: [4, 240]
Did the exclusion criteria explicitly include the presence of another autoimmune disease? <sup>4</sup>	58 (59.8)
Were any endpoints based on dichotomizing continuous information?	66 (68.0)
Primary	81 (83.5)
Secondary	
How were dichotomized responder endpoints analysed? <sup>5</sup>	27 (27.8)
Cochrane-Mantel-Haenszel test	21 (21.6)
Logistic regression	9 (9.3)
Chi-square test	9 (9.3)
Cox model	9 (9.3)
Fisher's exact test	5 (5.2)
Log rank	16 (16.5)
Other	

Question	n (%)
Any high-dimensional data collected at baseline (gene expression, GWAS, synovial biopsies etc.) <sup>6</sup>	8 (8.2)
<sup>1</sup> One article used a sequential multiple assignment randomised trial design with an interim futility assessment.	
<sup>2</sup> One article did not report the recruitment period. To translate recruitment periods given in months to weeks, 4 weeks was taken to be equivalent to 1 month.	
<sup>3</sup> For two, one, and one article respectively the median, mean, and maximum follow-up times are used. To translate recruitment periods given in days and months to weeks, 30 days and 1 month were taken to be equivalent to 4 weeks.	
<sup>4</sup> One article reported “any serious illness” as an exclusion criteria and is considered as ‘No’ for the extraction.	
<sup>5</sup> Articles may have utilised more than one method.	
<sup>6</sup> Three articles that collected MRI imaging data, for which it was unclear as to whether this was high-dimensional, are listed as ‘No’ for the extraction.	

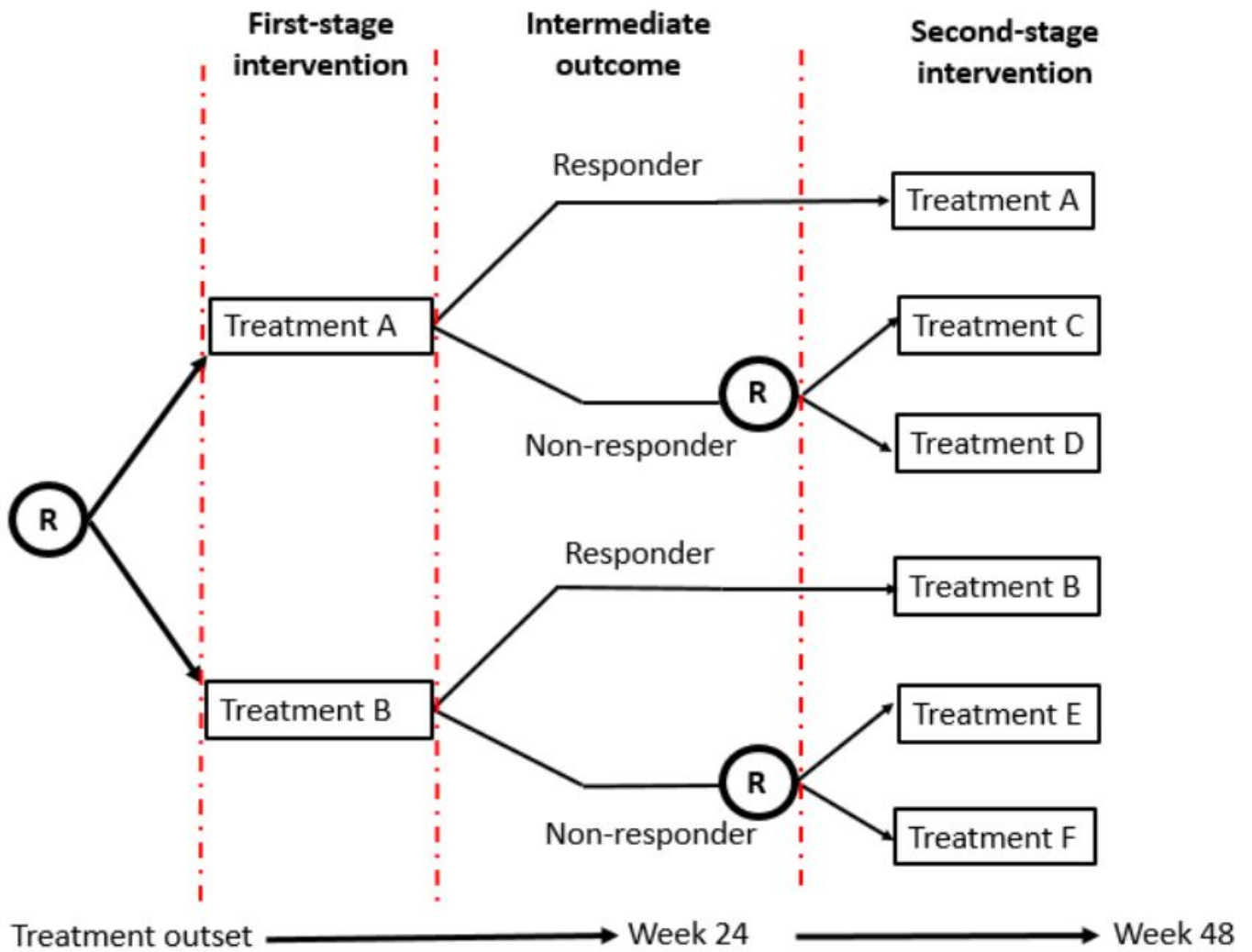
## Figures





**Figure 1**

Illustrations of umbrella and basket trial designs, with the sub-studies evaluating the new treatment(s) that are matched by the pre-defined biomarker(s) or genetic mutation(s).



**Figure 2**

An example SMART design. Only non-responders to the initial treatment are re-randomised in the second stage. R=randomisation.

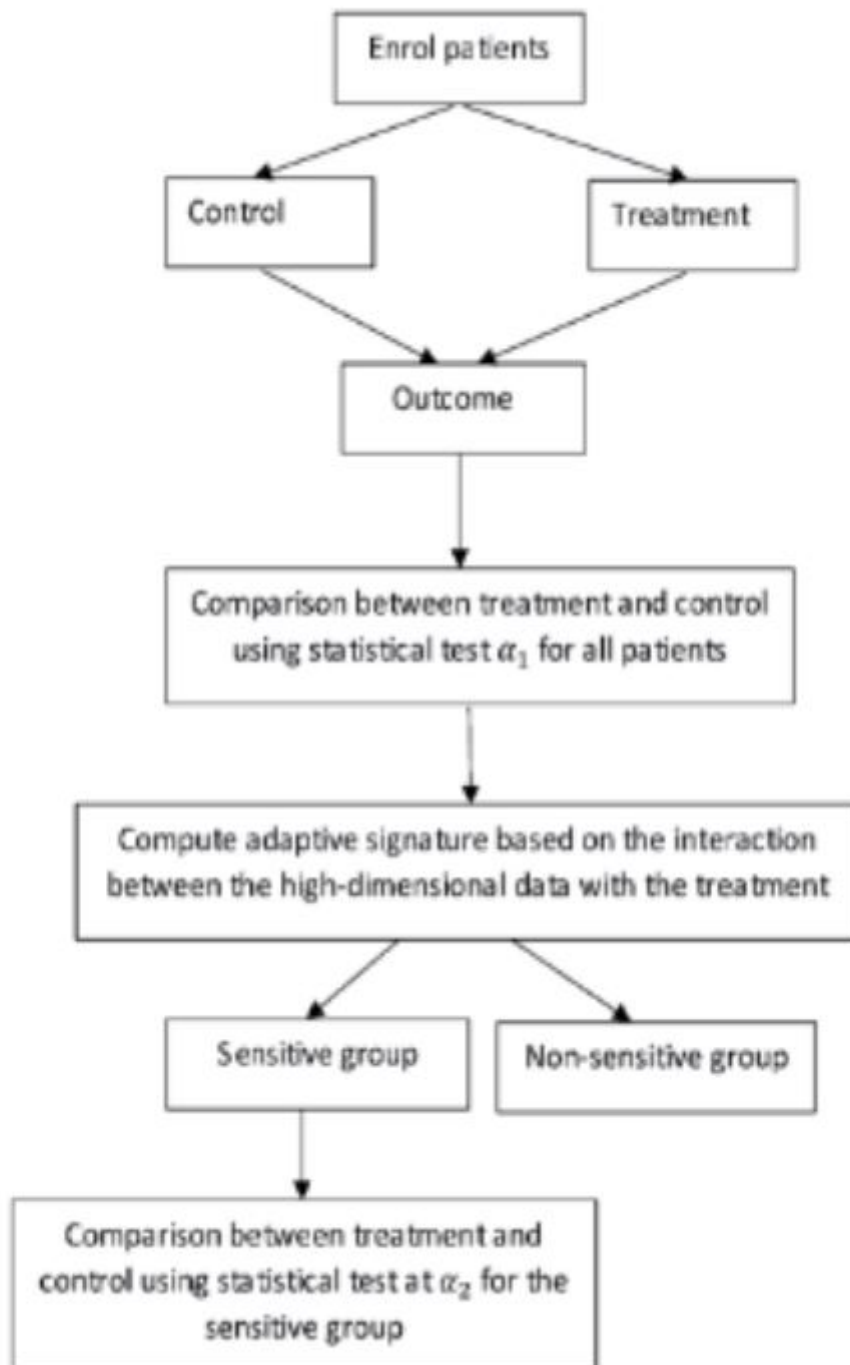


Figure 3

Schematic representation of the adaptive signature design.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraylingetalSupplementaryMaterials.xlsx](#)