Early phase and adaptive design clinical trials in rheumatoid arthritis: a systematic review of early phase trials.

Rheumatology Advances in Practice 2 (2), rky045. 10.1093/rap/rky045

Please note: Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Abstract

Objective. Adaptive designs can enable highly sophisticated and efficient early phase trials, but the clinical inference from these trials is surrounded by complexity, and currently there is a paucity but steadily increasing amount of use of these designs in all fields of medicine. We aim to review early phase trials in RA to discover those that have used adaptive designs and benchmark trial characteristics.

Methods. From an OVID search for journal articles reporting the results of early phase trials in rheumatology, 35 studies were found, with 9 subsequently excluded; 11 were added from manual searches and 19 from searching the references. Study characteristics were extracted from the 56 papers (describing 62 trials), including the number of arms, number of patients, the primary outcome and when it was measured.

Result. One early phase trial using an adaptive design was found. The benchmark early phase trial in RA is a phase II double-blinded randomized trial, with four arms (one control and three intervention), each with 34 patients, and ACR20 measured at 16 weeks as the primary outcome.

Conclusion. The one adaptive design reviewed here, and a simulation study found in the search, both indicate that adaptive designs can be applied to early phase trials in RA. We have described the benchmark, which the efficiency of early phase trials using an adaptive design needs to exceed. These efficient designs could drive down numbers required, time for data collection and thus cost. Changes have been suggested, but more needs to be done.

Key words: rheumatoid arthritis, adaptive design, early phase trial, systematic review

Key messages
- A paucity of adaptive designs is found in early stage RA trials.
- Research into RA treatments must make use of adaptive designs.
- Research into relevant, early time-point, patient-reported, RA outcome measures is required to aid these designs.

Introduction

Randomized controlled trials are the gold standard in evidence-based medicine. However, many clinical trials struggle to meet the recruitment target and time line [1]. This issue is not unique to rheumatology: randomized controlled trials are resource intensive in terms of time, personnel, finance and the available patient pool. Some of these obstacles might be mitigated by the use of adaptive trial designs, which have been developed to improve clinical trial efficiency. The application of adaptive designs in early phase trials has become highly pragmatic, efficient and pertinent. Both the US Food
and Drug Administration and the European Medicines Agency accept adaptive designs but have issued guidance on aspects that require special consideration [2, 3]. However, oncology is the only area where adaptive designs are established [4, 5], and there is a relative paucity of their use across other fields of medicine, including rheumatology.

An adaptive design trial is one in which modifications are made at various time points, dependent on pre-specified outcomes collated from the data observed up to that point. The modifications, time points [or more often, a point after which n patients have been recruited or have sufficient outcome(s) data] and the outcomes are many and variable, even within specific disease areas. Modifications can include adaptations to the randomization schedule, sample size re-estimation (both blinded and unblinded) and changes to the inclusion/exclusion criteria and to the mix of drugs defined as the intervention. The various and multiple interim analyses in these designs can allow for early curtailment of the trial owing to futility, safety or non-inferiority, and certain designs allow for intervention arms (such as multi-arm, multi-stage [6]) to be dropped for these reasons. Other designs allow for an operationally seamless transition from phase II(a/b) to phase III [7]. These designs are far removed from trials with incorporated interim analyses, which are often less statistically rigorous [8].

Since their introduction, researchers have formulated increasingly inventive designs to bring about the desired purposes for undertaking these trials: flexibility, efficiency and value for money (ever more pertinent in the current political and economic global and academic climates). It is most important, however, to reduce any possible harm to patients by attempting to minimize the expected sample size required to prove (or otherwise) the hypothesis in question.

The methodology for these designs started with Gehan’s 1961 [9] single-arm design, Fleming’s [10] single-stage design of 1982, then Simon’s [11] two-stage design of 1989, with the notion of dual outcomes, for both efficiency and toxicity, introduced by Bryant and Day in 1995 [12]. Now we look to UK-based names, such as Royston et al. [13], Burnett et al. and Hills and Burnett [14, 15], Mander et al. and Wason and Mander [16, 17], and Magirr et al. [18]. In the USA, luminaries include Thall and Cook [19], who introduced the concept of Bayesian-adaptive designs, and Berry et al. [20], who have furthered that field, in addition to O’Quigley et al. [21].

Outcome Measures in Rheumatology (OMERACT) has established a virtual special interest group to examine how adaptive designs might be applied to rheumatology research, in particular addressing issues set out in the US Food and Drug Administration and European Medicines Agency guidance [2, 3]. We focus on early phase trials, because the US Food and Drug Administration and European Medicines Agency have accepted adaptive designs for early phase trials, but not for phase III trials. Therefore, the purpose of this systematic review was, first, to discover those early phase trials that have used adaptive designs in RA, and second, to describe the characteristics of early phase trials to determine the benchmark, which the efficiency of early phase trials using an adaptive design needs to exceed. The results from this study will form the evidence base that will underpin work of the OMERACT Adaptive Clinical Trial Design Special Interest Group, by providing the requirements for designing an early phase trial in RA on matters such as the number of arms, the number of patients, the primary outcome and when it is measured. To implement an adaptive design into any such trial, it will need to be shown that it can be superior in efficiency to the benchmark.

**Methods**

**Search and eligibility criteria**

The initial step for this systematic review was to search for journal articles reporting the results of early phase trials in rheumatology. This was undertaken in January 2015 using the OVID online database using the following search terms:

(rheu* OR arth*) AND (early phase or phase 1 or phase I or phase 2 or phase 2a or phase 2 b or phase II or phase lla or phase llb) AND trial.ti.

The search included no specific terms for adaptive designs, because the relevant articles should also be identifiable as early phase trials under the terms used. This search also included papers on PsA, OA and systemic JIA, but these were later removed because the objective of this systematic review focused on RA as a specific exemplar. Ethical approval was not obtained to undertake this systematic review, because it was not required.

**Data collection**

Data were collected on phase, randomization, blinding, whether the trial used an adaptive design (and, if so, what sort), time period for data collection, countries involved, amount of participation time, inclusion and exclusion criteria, the intervention(s) and the treatment schedules for control and intervention groups, the primary outcome, when the primary outcome was assessed, the total number, the number of arms and number in each arm, and any additional outcomes, plus any further notes. Results relating to the primary outcome were also collected and were summarized using forest plots created by RevMan5.3 [22] (with specific primary outcomes at specific time points grouped together).

When the point at which the primary outcome was assessed was not entirely clear, we applied the following strategy: where adverse events were the primary outcome, the amount of participation time was used; where multiple outcomes were listed, the worst case (longest duration) was taken.
Results

Based on abstracts, 35 papers were found in the OVID search, of which nine papers were excluded (supplementary Table S1, available at Rheumatology Advances in Practice online) [23–31]. Eleven papers were found through further manual searches, and 19 more came from searching the reference lists of papers already found for review, or to be used in the discussion. This gave a total of 56 papers (see supplementary material extracted data, available at Rheumatology Advances in Practice online) [32–87]. This information, and that relating to how, and which, papers came to be reviewed here is available is Fig. 1.

In 54 of these papers, a single trial was reported, except for Choy et al. [35], in which three separate trials are reported, and for Namour et al. [64], in which 5 separate trials are reported. Thus, there are 56 papers reporting on 62 trials. Therefore, for each analysis, it is noted whether it is referring to the number of papers or the number of trials.

There was one example of an early phase trial using an adaptive design in the 62 trials, which was part A of Choy et al. [35].

The trials took place in a variety of different geographical regions, as shown in Table 1; the most common was the USA. Phases were listed as I, Ib, II, Ila, Iib, I/II and Ila/b, with phase II being the most frequent (supplementary Table S2, available at Rheumatology Advances in Practice online). All but five (plus a protocol that was combined with another to form a trial [68]) used randomization. Of the 62 trials, 49 were double-blinded, one was single-blinded (within Choy et al. [35], seven were open-label and one used the term ‘masked’ [57]. The length of overall data collection time (from 20 of 56 papers) ranged from 3 to 61 months, with a mean of 19.9 (s.d.: 12.84) months and a median of 16.5 (interquartile range (IQR): 13.0; 23.5) months. The amount of trial participation time ranged from 0.86 to 76 weeks (from 61 of 62 trials), with a mean of 18.5 (s.d.: 13.77) weeks and a median of 16.0 (IQR: 9.0; 24.0) weeks.

Numerous inclusion and exclusion criteria were applied across these papers (see supplementary material, available at Rheumatology Advances in Practice online). The main areas were for inclusion were as follows: fulfilled ACR [88] criteria; DAS28 [89], CRP, ESR and/or morning stiffness above specified values; number of swollen and tender joints above a specified value; failed previous DMARD treatment; length and stability of MTX treatment; and, age, generally >18 years, and sometimes limited to a maximum of 75 years. For exclusion, these were as follows: recent DMARD treatment; and stability of NSAID treatment.

Of the 62 trials, 14 did not include a control arm. Forty-six used some form of placebo, although two [53, 54] used background MTX throughout all arms in combination with placebo or the study treatment.

The most common primary outcome was ACR20 [90] (Table 2), with many secondary outcomes measured (usually reflecting the RA core set) [91]. The full list is available in the supplementary material, available at Rheumatology Advances in Practice online, but the most frequent were the individual components of DAS28 [89] and the ACR 50 and 70 response criteria.

The time at which the primary outcome was measured ranged from 2 to 26 weeks (from 37 of 62 trials), with a mean of 13.6 (s.d.: 7.32) weeks and a median 12.0 (IQR: 9.0; 20.0) weeks.

The total number of patients ranged from 12 to 509 (from all 62 trials), with a mean of 133.8 (s.d.: 132.55) and a median of 73.0 (IQR: 55.3; 222.5). The number of arms ranged from 2 to 9 (from all 62 trials), with a mean of 4.3 (s.d.: 2.06) and a median of 4.0 (IQR: 2.8; 6.0). The number of patients per arm ranged from 3.13 to
252.00 (from all 62 trials), with a mean of 34.01 (S.D.: 39.041) and a median of 23.42 (IQR: 8.44; 51.46). The forest plots (Figs 2 and 3) show that the odds ratios and effect sizes found in these papers are highly varied, but these results are either positive or indicate null findings (i.e. no evidence of a difference). There are no published negative results available.

**Discussion**

There are many key issues that should be addressed for adaptive design trials before use for the rheumatic diseases (Table 3). There was only one example of an early phase trial using an adaptive design in the 62 trials, which was part A of Choy et al. [35]. From the papers reviewed here, the benchmark early phase trial in RA is a phase II double-blinded randomized trial, with four arms (one control and three intervention) containing ~134 patients (34 in each arm), and a primary outcome of ACR20 [90] measured at 16 weeks. (The mean points to 13.6 weeks, but given that measurements are generally taken every 4 weeks, this rounds up to 16 weeks). It should be noted, however, that this is a very general set of values taken from widely varying research, some of which was inherently biased to have certain criteria by the very nature of the interaction between the intervention and the population sampled.

It appears that the standard is to use composite measures, such as ACR20 [90] or DAS28 [89], measured at a time point of 16 weeks. For adaptive designs to be implemented properly in this field and to provide the improvements in efficiency desired, we require a highly discriminate outcome measure, which can be assessed at early time points for interim analyses implemented in these designs. Analysing the response characteristic of commonly used outcome measures over time will be vital. Given the small sample, missing data imputation might also have an important role.

A paper found in the search but not reviewed, in that it was a simulation study, was by Thygesen et al. [27]. Here, the authors undertook a set of simulations around

---

**Table 1** Countries in which early phase clinical trials in RA are taking place (data from 39 of 56 papers)

<table>
<thead>
<tr>
<th>Country</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>15</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>11</td>
</tr>
<tr>
<td>UK</td>
<td>9</td>
</tr>
<tr>
<td>Belgium</td>
<td>8</td>
</tr>
<tr>
<td>Germany</td>
<td>7</td>
</tr>
<tr>
<td>Russia</td>
<td>5</td>
</tr>
<tr>
<td>Canada, Norway, Poland, Serbia/(former state of) Serbia and Montenegro</td>
<td>4</td>
</tr>
<tr>
<td>Austria, Australia, Finland, Hungary, Ukraine</td>
<td>3</td>
</tr>
<tr>
<td>Brazil, China, Czech Republic, Denmark, Japan, Mexico, New Zealand, Romania, Spain, Sweden</td>
<td>2</td>
</tr>
<tr>
<td>Argentina, Belarus, Chile, Colombia, Egypt, Estonia, France, Georgia, Greece, India, Israel, Italy, Lithuania, Malaysia, Philippines, Portugal, Republic of Korea, Slovakia, South Africa, Switzerland, Taiwan, Thailand, Turkey</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 2** Primary outcomes (data from 37 of 62 trials)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACR20</td>
<td>18 (48.6)</td>
</tr>
<tr>
<td>DAS28</td>
<td>6 (16.2)</td>
</tr>
<tr>
<td>Adverse events</td>
<td>2 (5.4)</td>
</tr>
<tr>
<td>Area under curve formed from ACR20 at three time points</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>ACR20 + ACR50</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>CRP + ESR</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>Efficacy (based on clinical symptoms, signs and laboratory tests; &lt;30% is ineffective, ≥30% to &lt;50% is effective, and ≥50% is remarkable)</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>Modified Paulus approach</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>MRI erosion score</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>Paulus20</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>Radiological score (Van der Heijde modified Sharp score)</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>SJC</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>Time exhibiting Paulus20, weeks</td>
<td>1 (2.7)</td>
</tr>
<tr>
<td>TJC + SJC</td>
<td>1 (2.7)</td>
</tr>
</tbody>
</table>

SJC: swollen joint count; TJC: tender joint count.
a Bayesian dose-finding procedure applied to an adaptive seamless phase I/II trial, specifically in RA. This was a bivariate procedure to look at both safety and efficacy, and the set-up was thus that the hypothetical trial collected data on these when the patient received the treatment every 4 weeks. For safety, information on all adverse events was collected, and if any of these matched a pre-defined list, the event was considered a F;G.2

### Early phase trials in RA
https://academic.oup.com/rheumap
For efficacy, CRP and ACR20 were also recorded. The outcomes were thus:

(i) Occurrence of a dose limiting event within 4 weeks of treatment (yes or no)—a measure of safety;

(ii) Success with respect to ACR20 at 4 weeks and a 25% reduction in CRP at 4 weeks relative to baseline (yes or no)—an early indicator of efficacy;

(iii) Success with respect to ACR20 at 16 weeks (yes or no)—a more reliable indicator of efficacy.\[27\]

This suggests that it is possible to look at outcomes, such as CRP and ACR20, at a relatively early time point. Equally, in part A of Choy et al. [35], we see that the primary outcome was the change in DAS28 from baseline to day 14. At this point, and having been measured every week, the safety, tolerability, efficacy, pharmacokinetics and pharmacodynamics were assessed. Also, in a paper [40] using the Paulus20 outcome [92], efficacy was assessed at week 4. However, in the majority of the papers reviewed here, it seems that researchers are rarely keen to use ACR20 or DAS28 outcomes at a time point that would be suitable for an adaptive design. Vital research is required to investigate early, potentially patient-reported, outcomes that are predictive of eventual ACR20 or DAS28 results so that sample size and trial design may be adjusted.
and the various simulations of Thygesen et al. [27] required numbers as low as 74. Although there is not enough evidence here to say for certain, this is indicative of the notion that adaptive designs tend to require fewer patients than standard trial designs.

Another relevant factor here is the overly restrictive inclusion and exclusion criteria, which makes for highly stringent recruitment barriers. This then drives up the length of time required for data collection, which was almost 2 years on average. To collect data on only ~134 patients in this time shows that the throughput of patients into these trials is simply too slow.

It is also worth noting that all the papers reviewed here provided only positive or no-evidence-of-a-difference results (Figs 2 and 3), which, along with the fact that most of this research is done in the private sector, explains the small number of papers available. Publication bias based on the main trial outcome [93] has already been shown to be a disappointing barrier to furthering rheumatological research, which undoubtedly means that time, money and other resources are wasted on research that is already known to be fruitless.

In addition to the methodological papers laid out in the Introduction [9–21], there are numerous examples of adaptive designs in practice in the field of oncology, many of which are summarized in reviews [5, 94]. These have mostly made use of group sequential methods [95, 96]. Part A of Choy et al. [39], was a randomized, double-blind placebo-controlled Bayesian-adaptive dose-finding study. This used trial simulations of the Bayesian-adaptive pharmacokinetic/pharmacodynamic design to estimate the sample size and a Bayesian-adaptive dose-finding algorithm to identify subsequent doses after the starting dose [97].

In a 2015 editorial in The Journal of Rheumatology, Pope et al. [1] wrote:

> Over the last 2 decades, we have seen advances in the clinical management of rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, systemic sclerosis, vasculitis, and systemic lupus erythematosus. Yet trial designs and entry criteria required for drug development have not kept pace with medical care and thus no longer reflect patients seen in typical rheumatology practices in Canada, the United States and Western Europe [1].

In addition to this editorial, the Canadian Rheumatology Research Consortium has also suggested changes [98], which focus on both inclusion and exclusion criteria and study design.

We already know that updates can be made in the field of rheumatology. Five years before his paper defining the ACR20, 50 and 70 outcomes [90], Felson et al. [99] had published on the need for standardized outcomes and reporting. Also, a review in 2009 [100] showed how far trial design for RA had advanced in the previous decade, which noted the already existing debate [101, 102]. We are therefore hopeful that researchers in rheumatology and clinical trialists can start a change of the tide. The key objectives of the OMERACT Adaptive Clinical Trial Design Virtual Special Interest Group are to identify and address key barriers to adaptive design for clinical trials in RA and to improve clinical efficiency, which will benefit patients, researchers and funders.

**Conclusion**

Our search discovered a single example of an early phase trial that used an adaptive design in RA. We have described the benchmark, which the efficiency of early phase trials using an adaptive design needs to exceed. Research into treatments for RA should make use of adaptive designs if there is a desire to move forward in the world, or else it will begin to lag behind other clinical research efforts [5]. Beyond the academic elegance and need for statistical a priori definitions, these designs can reduce numbers in terms of participants and resources and can lead to trials that are operationally connected.

Whether we create, test and use objective early-time-point outcomes for use in adaptive designs may well be the important question. We must look for innovative ways of measuring outcomes in RA at much earlier time points. Also, we would need to look at which of these outcomes is more discriminatory.

**Funding:** The Cardiff Regional Experimental Arthritis Treatment and Evaluation (CREATE) Centre is joint funded by Health and Care Research Wales (HCRW) and Arthritis Research UK (ARUK). The Parker Institute, Bispebjerg and Frederiksberg Hospital, is supported by a core grant from the Oak Foundation (OCAY-13-309).

**Disclosure statement:** The authors have declared no conflicts of interest.

**Supplementary data**

Supplementary data are available at Rheumatology Advances in Practice online.

**References**


Burnett AK, Hills RK, Hunter AE et al. The addition of gemtuzumab ozogamicin to low-dose Ara-C improves remission rate but does not significantly prolong survival in older patients with acute myeloid leukaemia: results from the LRF AML14 and NCRI AML16 pick-a-winner comparison. Leukaemia 2013;27:75–81.


60 Moreland LW, McCabe DP, Caldwell JR et al. Phase I/II trial of recombinant metholin human tumor necrosis...


64 Namour F, Vanhoutte FP, Beetens J. Pharmacokinetics, safety, and tolerability of GLPG0259, a mitogen-activated protein kinase-activated protein kinase 5 (MAPKAPK5) inhibitor, given as single and multiple doses to healthy male subjects. Drugs R D 2012;12: 141–63.


83 Kremer JM, Bloom BJ, Breedveld FC et al. The safety and efficacy of a JAK inhibitor in patients with active...


