Check for updates

Head-to-head trials in inflammatory bowel disease: past, present and future

Lieven Pouillon, Simon Travis, Peter Bossuyt, Silvio Danese and Laurent Peyrin-Biroulet

Abstract | With the increase in the number of novel drugs for inflammatory bowel disease (IBD), comparing therapeutic options or strategies has become a key challenge in IBD trials. Head-to-head trials designed and powered to enable formal comparisons are the gold standard in comparative research. Indeed, these trials are requested by some health authorities for evaluating the positioning of new treatments in IBD, as well as helping prescribing physicians to select the most appropriate treatment options for their patients. Despite head-to-head trials including aminosalicylate therapy in IBD having been performed decades ago, the first results of a randomized controlled trial directly comparing biologic agents with different modes of action have only now been published, mainly owing to important methodological issues. This Perspective provides an overview of the past, current and future concepts in IBD trial design, with a detailed focus on the role of comparative research and the challenges and pitfalls in undertaking and interpreting the results from such studies.

Effective drugs for inflammatory bowel disease (IBD) have been available for over 70 years[1]. Following the advent of anti-tumour necrosis factor (anti-TNF) therapy in the mid-1990s, treatment goals started to evolve and moved towards more robust targets[2], including mucosal healing, given its potential for disease modification, and probably also histological healing, which has been associated with a reduced risk of hospitalization, colectomy and colorectal cancer in patients with ulcerative colitis[3]. Multiple therapeutic options (biologic agents and small molecules) are currently available for IBD, and this armamentarium will increase further over the next decade[4] (FIG. 1). Given this increase in the number of novel treatments aiming for more stringent outcomes, comparative evidence is critical to help physicians and patients select the most appropriate therapeutic option, as well as to enable health authorities and payers to determine the optimal position of specific agents within treatment algorithms.

Until 2018, comparative data in IBD relied on indirect evidence provided by meta-analyses and comparative real-world evidence. Although the first head-to-head trials in IBD (sulfasalazine versus steroids and azathioprine, olsalazine versus mesalazine, or mesalazine Multi Matrix system (MMX, Cosmo) versus mesalazine) were performed decades ago[5–7], direct comparisons between biologic agents have been lacking until now (FIG. 1). Indeed, the modest effect size of all currently available drugs compared with placebo has always made competitive head-to-head trials in IBD particularly unattractive to industry[8], despite the appeal to clinicians, payers and patients. Methodological challenges concerning the study design and population, the choice of comparator and end point, the use of blinding, and different potential dosing and escape strategies further complicate head-to-head trials comparing biologic agents for IBD. In other immune-mediated inflammatory disorders (IMIDs), such as rheumatoid arthritis, multiple sclerosis and plaque psoriasis, however, such studies were performed much earlier[9–11], and direct comparative evidence remains a key request of health authorities and payers when considering the reimbursement of novel treatments in an era in which increasing health-care costs conflict with growing health-care budget constraints[12].

With the introduction of more robust outcomes such as endoscopic remission, which have been linked to reduced placebo response rates[13,14], direct comparisons between agents have become a more attractive option in IBD[15]. The VARSITY (the efficacy and safety study of vedolizumab intravenous compared to adalimumab subcutaneous in participants with ulcerative colitis) trial compared the efficacy and safety of intravenous vedolizumab (a monoclonal antibody against α4β7 integrin) and subcutaneous adalimumab (a monoclonal antibody against TNF) in moderate-to-severe active ulcerative colitis and was the first head-to-head trial between biologic agents with different modes of action in IBD[16]. Head-to-head trials comparing other treatment options such as etrolizumab (a monoclonal antibody against β7 integrin) and infliximab (an intravenously administered TNF blocker) or adalimumab are currently running and results are expected in the next 5 years.

In this Perspective, we provide an overview of the past, current and future concepts of trials in IBD, with a focus on head-to-head trials, together with a critical appraisal of comparative evidence in the treatment of IBD, emphasizing the methodological challenges and current available data.

Shifting priorities in IBD research

The main target of early IBD research was the development of drugs to control disease symptoms, improve quality of life and alter disease progression. This approach led to the current therapeutic armamentarium that consists of immunomodulators, anti-TNF blockers, anti-adhesion therapy, IL-12–IL-23 blockers and janus kinase (JAK) inhibitors[4]. Compared with other IMIDs, the number of therapeutic options with proven efficacy in clinical trials remains limited, and IBD studies continue to focus on how to optimize the use of existing drugs and treatment strategies. To choose the right drug for the right patient, the ideal scenario of personalized, predictive medicine is on the

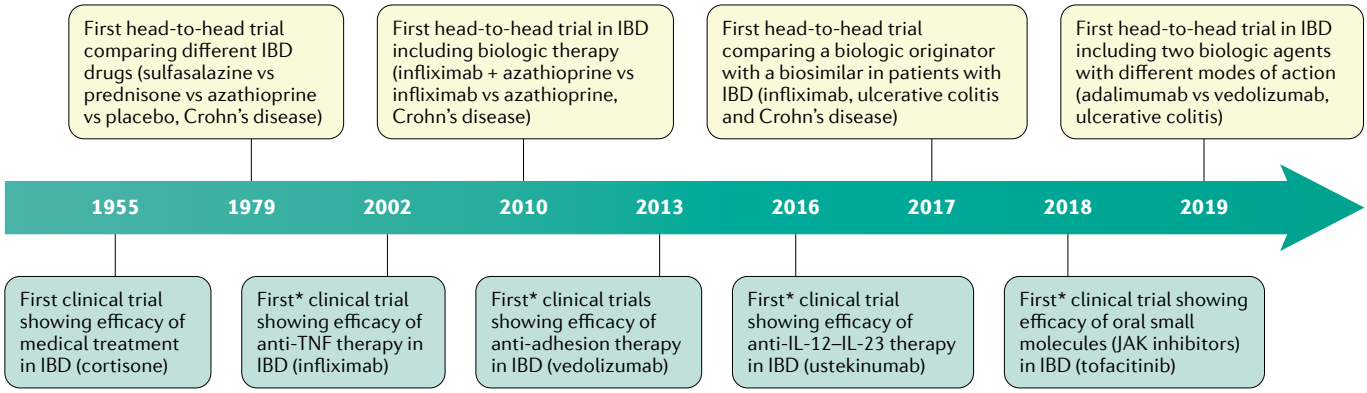

Fig. 1 | **Key events in IBD drug development.** Timeline showing key trials in IBD drug development (green panels) and head-to-head comparative research (yellow panels). IBD, inflammatory bowel disease; JAK, janus kinase; TNF, tumour necrosis factor. *Phase III, randomized controlled trial.

horizon[17], but practical insights into the clinical use of drugs in patients are still scarce. Thus, direct comparison of existing biologic agents and newer IBD drugs in head-to-head randomized controlled trials (RCTs) has become a key feature in current IBD research (BOX 1).

Comparing therapeutic options for IBD

Meta-analysis, real-world evidence and head-to-head trials are the tools for comparing different therapeutic options for IBD, each with its own strengths and weaknesses (TABLE 1). In this regard, it is important to distinguish a drug's efficacy (its effect in a controlled environment that can only be offered by a clinical trial) with its effectiveness (its effect in real-world circumstances).

Meta-analysis. Meta-analysis is the quantitative, scientific synthesis of research results from many individual studies, often RCTs. A meta-analysis should always be preceded by a systematic review, which aims to provide a robust overview of the efficacy of an intervention, a problem, or field of research. The systematic review process includes formal methodological guidelines for the literature search, study screening, data extraction and coding, along with detailed documentation of each step[18]. The systematic review aims to be transparent, reproducible and updatable and should address well-defined research questions. Only if the systematic review reveals sufficient and appropriate quantitative data can a meta-analysis be performed to assess the magnitude of the outcome across the selected relevant primary studies to analyse the causes of variation among study outcomes by using effect sizes[19].

In contrast to a narrative review, a systematic review with meta-analysis can accurately summarize results across studies in a comprehensive and quantitative manner. Nevertheless, owing to a lack of stringent methodological and reporting quality criteria, results of meta-analyses are often criticized and controversial[20]. Indeed, the use of statistically flawed approaches can lead to erroneous and misleading results. Thus, the term 'meta-analysis' should be applied only to studies that use well-established statistical procedures, such as appropriate effect-size calculation, weighting and heterogeneity analysis with statistical models that take into account the distinct hierarchical structure of meta-analytical data, or develop rigorously justifiable methodological advances of these methods[19]. Other caveats exist: a meta-analysis can highlight areas in which evidence is deficient, but cannot overcome these deficiencies. The over-representation or under-representation of populations, species or systems in the scientific literature, as well as selective or incomplete data reporting in primary publications (that is, publication bias), remain an important challenge to meta-analysis[21]. Furthermore, meta-analyses risk becoming rapidly outdated because results can be markedly different within a few years as more studies emerge.

Network meta-analysis is a technique for comparing multiple treatments simultaneously in a single analysis by combining direct and indirect evidence within a network of RCTs[22]. This approach has become particularly attractive in IBD research as it could help in assessing the comparative effectiveness of different treatments regularly used in clinical practice. Nevertheless, network meta-analysis is more complex and resource intensive than pair-wise meta-analysis. Assessment of transitivity — that is, ruling out systematic differences between the available comparisons other than the treatments being compared — is crucial for the validity of a network meta-analysis, and input from both clinicians and statisticians is needed when developing the clinical question[22].

In IBD, there is wide variety in the design of registration trials for new drugs. Key differences include patient populations, the choice of symptom-based measures (clinical remission or clinical response) versus objective outcomes (biomarker levels, endoscopy findings or histopathology analysis) as end points and variable analytical methods (for example, different methods to deal with missing data, such as last observation carried forward, in which the last observed score is used for all subsequent missing time points, or non-responder imputation, in which all participant dropouts are assumed to be non-responders)[23]. This variety makes meta-analysis particularly challenging and explains why comparable analyses often lead to differing conclusions. The primary goal of apparently similar meta-analyses often differs as well, affecting the number of included studies and, consequently, the results. By way of example, differences between clinical trials and meta-analyses for biologic agents approved for ulcerative colitis are shown in Supplementary Tables 1 and 2. In clinical trials, the exact definition and timing of the assessment of the primary end point vary, and the studied population can be more or less treatment refractory depending on the inclusion and exclusion criteria. Other differences are in the interpretation of clinical scoring systems (average versus worst-rank), corticosteroid tapering regimens and handling of missing data. For meta-analyses, comparisons can be made directly between different drugs, or indirectly versus placebo. Studied end points can be effectiveness, safety, or both.

Real-world evidence. Real-world evidence is defined as the clinical evidence regarding the usage and potential benefits or risks of a medicinal product derived from analysis of data acquired from clinical practice[24]. Information relates to patients' health status or delivery of health care, and is routinely collected from a variety of sources, such as electronic health records or administrative databases[24]. Real-world studies are observational and are able to provide large datasets from diverse patient populations over a long period of time[25]. The best source of real-world data remains debated, and to obtain the most granular information possible, big data analysis from electronic records, combining multiple data sources, might become the favoured tool[26].

Although evidence generated from practice-based observations seems to be of less scientific value than data from RCTs, real-world observations fulfil an important role in IBD research. Indeed, the strict protocol-specified criteria for enrolment into RCTs might not reflect clinical reality, in which health-care practitioners are faced with patients with wider ranges of disease severity and age, taking a broader range of concomitant medications and often with more and varying comorbidities[27]. Real-world evidence also offers the opportunity to compare the value of different health-care systems, as well as looking into 'hard' outcomes in the long-term, such as bowel damage and surgery in patients with IBD over a period of years, which is not possible in the shorter time frames of most RCTs. Nevertheless, comparing real-world evidence between different treatment strategies is fallible: data collection is primarily retrospective, and can have systematic bias in sampling and ascertainment, as well as being inherently incomplete compared with trial data collection, which has set criteria and protocols[24,25]. This limitation might in part explain why retrospective studies and RCTs sometimes lead to different conclusions. A typical example is therapeutic drug monitoring in patients with IBD, with some retrospective data supporting proactive infliximab dosing based on trough concentrations maintained at a certain level[28], whereas prospective trials have never confirmed this aspect[29,30].

Propensity score matching can counter systematic differences in baseline characteristics, by down-sizing known confounders and by reducing the effects of multiple covariates to a single score, the propensity score. This process enables the estimation of population-average treatment effects and helps to overcome selection bias by comparison of outcomes across treatment groups, pairs or pools of propensity score-matched patients[31]. However, matching ideally requires very large registries with thousands of patients to achieve groups with sufficient patient numbers, and propensity score matching can only be used if the patients had access to all treatment options being compared, with an equal probability of each treatment option. Furthermore, propensity score matching does not counter unobserved covariates, so residual confounding could still influence the results[32].

The ideal cohort for comparative real-world effectiveness research has a large (n >1,000) number of patients across multiple sites with varying practices, which provides granular data to enable deeper matching, multiple analytical approaches to assess consistency and instrumental variable analysis to estimate causal relationships. An example of such a cohort is the multicentre VICTORY (vedolizumab for health outcomes in inflammatory bowel disease) Consortium, that includes >2,500 patients from 12 different centres in the USA and Canada. This cohort has been used to compare the effectiveness of vedolizumab and anti-TNF therapy in patients with ulcerative colitis[33]. Preliminary results demonstrated that propensity score matching accounted for differences in age, sex, prior ulcerative colitis-related hospitalization, disease extent, disease severity, steroid-refractoriness or dependence, and prior anti-TNF therapy failure. Initial analysis showed a 54% versus 37% clinical remission rate favouring vedolizumab at 1 year (OR 1.54, 95% CI 1.08–2.18), implying a 54% benefit over anti-TNF agents, with fewer serious adverse events (OR 0.29, 95% CI 0.12–0.73)[33]. Caution is warranted when interpreting these data, as the retrospective collection of data from multiple centres might have affected uniformity.

Although RCTs often include patients with refractory disease, real-world effectiveness research has the additional benefit of taking into account a more varied set of patients with baseline characteristics often changing over time, which can gradually lead to improved effectiveness and disease outcomes in clinical practice. A large study from the USA showed that more patients who had not previously received biologic agents have been treated with vedolizumab since FDA approval in 2014 (from 6% to 12% among patients with Crohn's disease (2014–2015 versus 2015–2017) and from 29% to 36% among patients with ulcerative colitis (2014–2015 versus 2015–2017), with increasing remission and mucosal healing rates in patients with Crohn's disease and fewer hospitalizations and operations in patients with ulcerative colitis[34].

Head-to-head trials. Head-to-head trials are RCTs that compare different active interventions (for example, vedolizumab versus adalimumab in ulcerative colitis[16]) or treatment strategies (for example, early combined immunosuppression versus conventional management in Crohn's disease[35]) randomly assigned across groups or participants. The design of head-to-head trials differs from that of observational studies, in which a population is assigned to different interventions based on patient or provider factors and local guidelines (FIG. 2). Head-to-head trials are the current gold standard in comparative research; they are designed and powered to enable formal comparison between distinct therapies or therapeutic strategies[36].

Box 1 | **Past, present and future perspectives of IBD trials**

Past perspective

- Developing inflammatory bowel disease (IBD) drugs that induce and maintain clinical remission
- Developing disease-modifying drugs to reduce IBD-related complications

Present perspective

- Optimizing the use of existing treatment options (strategy trials)
- Enlarging the spectrum of disease-modifying IBD drugs by exploring modes of action different from that of anti-tumour necrosis factor (TNF) blockade
- Directly comparing biologic agents and new IBD drugs

Future perspective

- Identifying predictive biomarkers for disease evolution as well as response to therapy (personalized medicine)
- Revealing and therapeutically modifying gut microbiota–host interactions (towards definite cure of IBD?)

Table 1 | **Strengths and weaknesses of different comparative approaches**

Approach	Strengths	Weaknesses
Meta-analysis	Provides context that individual studies cannot provide Outcomes might include more precise estimate of treatment effects or risk factors for disease than individual studies Reduces the need for repeated research studies	Included studies should be similar enough to be pooled Potential research and publication bias Erroneous or poorly conducted studies can adversely affect results of entire meta-analysis Needs appropriate comparison methods to adjust for trial differences
Real-world evidence	Bridges the gap between clinical trials and practice Provides information on a population-based level from a wide variety of sources Captures long-term data about effectiveness and safety, including rare events, in heterogeneous populations Complements randomized controlled trials	Data completeness, accuracy and consistency may not be uniform (potential selection bias, information bias, recall bias and detection bias) Study populations are unselected, which limits treatment comparisons
Head-to-head trials	Gold standard: compare therapies in the same population and setting Increasingly required by regulatory authorities	Expensive Long timelines Eligible participants do not always reflect real-world patients owing to strict inclusion and exclusion criteria Require careful study design and selection of appropriate comparator and end points

There is speculation that pharmaceutical companies avoid head-to-head comparisons to avoid jeopardizing their market share by unfavourable results[37]. However, when the status of head-to-head trials across different indications, including IBD, was mapped, most (>55%) were industry sponsored, although fewer than 15% of these had two industry funders[38]. Industry-sponsored trials were larger, more commonly registered, more frequently used a non-inferiority design and had a higher citation impact than non-industry-sponsored trials. Interestingly, and some might suggest not surprisingly, they were more likely to show a favourable response towards the experimental treatment than non-industry-sponsored trials[38]. This finding might be a consequence of the non-inferiority design (discussed later). In our opinion, comparative trials carry greater credibility when driven by non-profit entities; although the latter face cost as a major barrier, the principal pitfalls of undertaking head-to-head trials remain in their design.

Head-to-head trial design

Type of comparison. Active controlled clinical trials methodology can be classified by type of comparison[7,15,16,39] (FIG. 3). The most convincing way of determining whether an investigational intervention is more efficacious than its comparator is by performing a superiority trial. In such comparison, the sample size calculation is conventionally based on achieving adequate power to demonstrate that the relevant confidence limit for a difference between the two treatments excludes zero, assuming that the experimental treatment is superior by a given amount ('delta'). It is, however, difficult for new agents to demonstrate superiority over established agents as the increase in efficacy might be small, so non-inferiority trials are often favoured.

Non-inferiority trials aim to demonstrate a minimum level of efficacy in a new drug compared with an established drug by a pre-specified degree, because the new drug might offer other advantages such as an improved safety profile, easier administration or reduced costs. The key challenge is to determine the non-inferiority margin, which is the clinically acceptable maximum difference between two treatments. This margin directly influences sample size estimates and study conclusions. To maintain the validity of non-inferiority trials, the margin should be narrow enough to preserve a clinically relevant amount of the active comparator's treatment effect. 'Clinical relevance' is a subjective term, open to debate and contention, but in IBD a difference in effect between an investigational drug and a comparator (be it placebo or active agent) of 10% is generally considered appropriate, because the efficacy of all agents is modest[40]. This margin depends, of course, on the outcome being compared (clinical remission or response, objectively confirmed or not). Although it is often thought that non-inferiority trials need to be much larger than superiority trials, there is no good reason why the size of the delta in superiority trials and the non-inferiority margin in non-inferiority trials, and by implication the sample size, should be different. Superiority trials are, however, sometimes smaller because the delta is chosen as the value that corresponds to the expected difference, with optimistic values selected to reduce the sample size[41].

Tsui et al.[42] evaluated whether published non-inferiority trials were adequately designed. Most trials (101 of 162) used an active comparator that had not even been shown to be effective itself, meaning that both treatments could be equally ineffective. In fewer than half the trials that used an effective comparator (25 of 61), the chosen margin was small enough to preserve more than 50% of the comparator's treatment effect[42]. In other words, only 25 of 162 trials compared active treatment with active treatment[42]. Remarkably, in almost 10% of non-inferiority trials, the design enabled the intervention to be declared non-inferior even if it was worse than placebo or another historical cohort[42]. In this regard, the NOR-SWITCH (switching from originator infliximab to biosimilar infliximab compared with maintained treatment with originator infliximab) trial is relevant. The NOR-SWITCH trial was a randomized, non-inferiority, double-blinded study, in which infliximab originator was compared with infliximab biosimilar (CT-P13) in patients with different IMIDs[39]. The primary end point was the percentage of participants experiencing disease worsening (variously defined). For IBD, disease worsening meant a change from a baseline Harvey-Bradshaw index of ≥4 points and a score of ≥7 points for Crohn's disease; for ulcerative colitis,

it meant a change from a baseline Partial Mayo score of >3 and a score of ≥5. The non-inferiority margin was set at 15%. Although the study reached its primary end point in the overall population (and also in the IBD population specifically), the range of risk differences between distinct types of IMIDs was notable, and the chosen margin has been criticized as being too wide to include clinically relevant differences[39] (Supplementary Table 3).

A favoured alternative to non-inferiority design is a placebo-controlled trial with a non-powered reference arm[43]. Here, the active comparator serves only to estimate the reference drug's efficacy, to confirm previous results and assist the regulators (or payers) in determining the drug's value. The advantages of this type of design include lower sample sizes, ranging between the active reference arm, the tested molecule and the placebo, and the opportunity to compare different dosage levels of the same drug. Nevertheless, interpretation remains complicated. Differences between the original (placebo-controlled) trial and the investigational drug, placebo-controlled with a non-powered reference arm, will cast doubt on the conclusions about the comparator. The lack of power in this type of study is, however, likely to explain discrepancies, unless outcome measures differ substantially[15]. In a pooled analysis of the two CORE (colonic release budesonide) I and II studies that analysed the outcomes of treatment with budesonide MMX at different doses versus placebo in patients with ulcerative colitis[44], the primary end point data were lower than those observed for other ulcerative colitis therapies at the same time point (18% versus 29–60%)[45–47]. Nevertheless, these latter trials used clinical remission alone as an end point, whereas the CORE studies combined clinical (no rectal bleeding and normal stool frequency) and endoscopic (no mucosal friability at full colonoscopy) outcome parameters[44].

Choice of comparator. Head-to-head trials comparing biologic agents in IBD have attracted serious interest because of ethical concerns raised by health authorities and industry around placebo treatment, the societal cost of therapy and repercussions of disease progression and adverse events[15]. These worries are shared by patients and clinicians, the latter guided by the principles of the 1964 Declaration of Helsinki, one of which states: "Physicians who combine medical research with medical care should involve their patients in research only to the extent that this is justified by its potential preventive, diagnostic or therapeutic value and if the physician has good reason to believe that participation in the research study will not adversely affect the health of the patients who serve as research subjects"[48]. Indeed, effective drugs for IBD have been available for a long time (>70 years)[1], so trials comparing two active treatments are overdue and reduce the chance of study participants receiving placebo. This approach should facilitate recruitment in IBD trials, now often disappointingly slow, because it is easier to convince a patient to participate in a clinical trial without a sham arm[15].

The choice of comparator needs to be carefully considered and depends on what the investigator or study sponsor would most like to know. There might not be such a thing as a 'good' or a 'bad' choice of comparator, as the key is the correct interpretation of the results, although fairness has to be taken into account. When performing head-to-head trials in IBD it is important to consider whether concomitant treatments should be allowed or not. For example, the use of infliximab in combination with azathioprine is superior to the use of infliximab monotherapy in both ulcerative colitis and Crohn's disease[49,50]; consequently, infliximab monotherapy cannot reasonably be used as an 'honest' comparator in a head-to-head trial with another biologic agent. Furthermore, concomitant steroids, and whether or not forced tapering of those should be imposed, is crucial to compare results in IBD regulatory trials. Put simply, investigator-decided steroid dose favours the placebo and reduces the absolute difference between placebo and study drug. To test its efficacy, a study drug is best tested in the context of mandatory steroid withdrawal. In the VARSITY head-to-head trial comparing adalimumab and vedolizumab for moderate-to-severe active ulcerative colitis, 36% patients in both groups were receiving concomitant steroids at baseline, but there was no forced steroid tapering[16]. At 52 weeks, the secondary outcome measure of corticosteroid-free remission rates was generally low across both treatment groups. There was a numerical but not statistically significant advantage for adalimumab over vedolizumab (26 of 119 (22%) versus 14 of 111 (13%); $P=0.08$). Whether the lack of forced steroid tapering also affected the primary outcome (clinical remission at week 52, which significantly favoured vedolizumab; $P=0.006$) of the VARSITY trial remains unclear, but the results for almost all the other secondary outcome measures (such as endoscopic mucosal healing) were consistent with the primary outcome, favouring vedolizumab. There was no multivariate analysis of data about steroid dosing during the study[16].

Use of blinding. Blinding refers to keeping trial participants, investigators (usually health-care providers) or assessors (those analysing the outcome data) unaware of the results of outcome assessments. Blinding can improve compliance and retention of

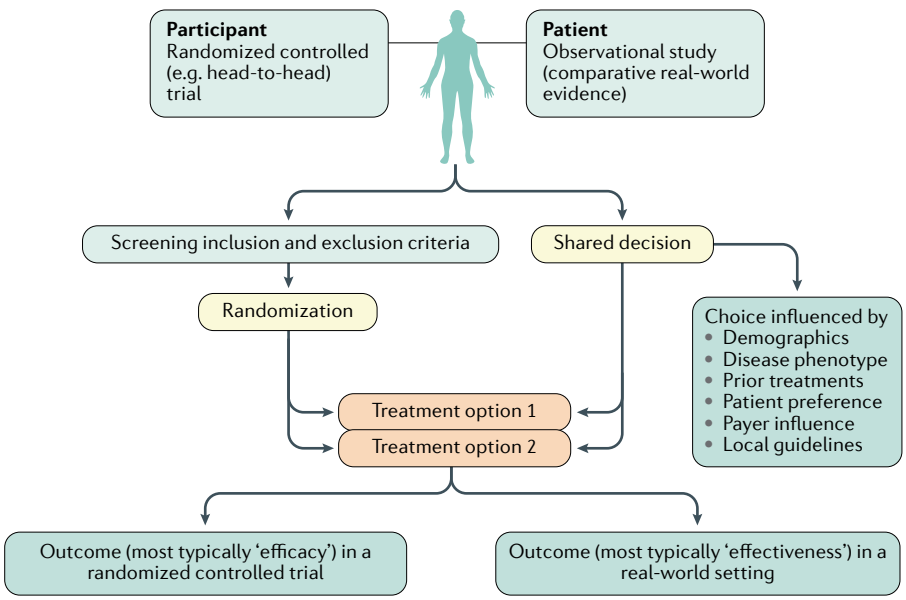

Fig. 2 | **Randomized controlled trials versus observational studies.** Differences in the design of randomized controlled trials and observational studies can affect comparative outcomes of distinct treatment options.

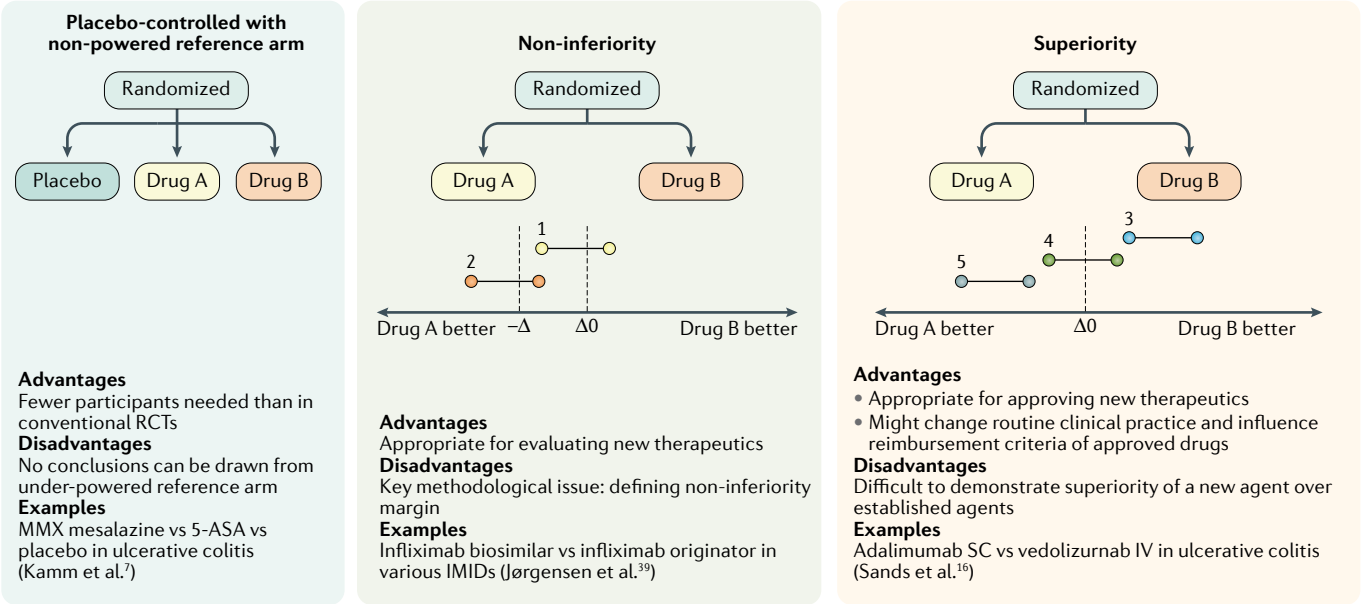

Fig. 3 | **Types of head-to-head trial and their advantages and disadvantages.** In a non-inferiority trial, the hypothesis that a new drug is non-inferior to the reference drug is tested, while a superiority trial is designed to test whether a new drug is more efficacious than the reference drug. In a placebo-controlled trial with a non-powered reference arm, the active comparator serves to estimate the reference drug's efficacy, to confirm previous results and assist the regulators (or payers) in determining the drug's value. 1, lower limit of confidence interval does not overlap with non-inferiority margin (hypothesis of non-inferiority accepted); 2, lower limit of confidence interval overlaps with non-inferiority margin (hypothesis of non-inferiority rejected); 3, confidence interval does not include zero treatment difference (drug B superior to drug A); 4, confidence interval does include zero treatment difference (no difference between drug A and drug B); 5, confidence interval does not include zero treatment difference (drug A superior to drug B). 5-ASA, 5-aminosalicylic acid; IMID, immune-mediated inflammatory disease; IV, intravenous; MMX, Multi Matrix system; RCT, randomized controlled trial; SC, subcutaneous.

trial participants while reducing biased supplemental care or treatment[51]. In head-to-head trials, the implications of double blinding, single blinding, non-blinding, or open-label treatment should be carefully considered. If a regulatory agency requests an active control arm in a phase III head-to-head trial, double blinding is mandatory, meaning that all participants, investigators and assessors have to be unaware of intervention assignments throughout the trial[32]. The main reason is to avoid expectation bias influencing treatment effect. When comparing drugs with different modes of administration (for example, oral versus subcutaneous or intravenous), blinding can be a logistical and expensive challenge, requiring a double-dummy blinding strategy. With this approach, the comparators do not need to look identical, because the placebo is given to match the appearance of each study drug[52].

Trials that are not double blinded should not automatically be deemed inferior. Single blinding means that only one of the three groups, normally the participants rather than investigators or assessors, remains unaware of intervention assignments. This approach can be sufficient if the purpose of the trial is to demonstrate a drug's effect in real-world situations[32]. In IBD research, the CONSTRUCT (comparison of infliximab and ciclosporin in steroid-resistant ulcerative colitis) trial compared the clinical effectiveness of infliximab with ciclosporin in patients with acute severe ulcerative colitis, through a single-blinded design[53]. A dynamic algorithm protected against investigator preference and randomly assigned patients in a one-to-one fashion, but local investigators and participants were aware of the treatment allocated, whereas the chief investigator and analysts were blinded[53]. There are examples in other fields of research, such as rheumatoid arthritis[9].

Non-blinding can be applied when testing outcome parameters that are unlikely to be influenced by participants' or investigators' expectations (for example, serum trough concentrations of the drugs being tested or immunogenicity in patients treated with a biosimilar versus biologic originator)[54,55]. An example in IBD research is the CALM trial, a multicentre, randomized, open-label, active-controlled, two-group, phase III, efficacy and safety study to evaluate two treatment algorithms, tight control and conventional management, in patients with moderate-to-severe active Crohn's disease[56]. The primary end point was endoscopic mucosal healing and absence of deep ulcers 48 weeks after randomization. Although participants and investigators were blinded to patient allocation, post-screening C-reactive protein (CRP) and faecal calprotectin levels, treatment in this study was non-blinded or open label. In our opinion, the design of the CALM trial potentially influenced the primary outcome, as interpretation of endoscopy was subject only to local reading, whereas a central reading process would have been less prone to expectation bias of the investigators. Nevertheless, objective (secondary) end points such as low CRP or faecal calprotectin levels differed between the two treatment strategies at the end of the trial, which provides reassurance that the primary outcome (tight control significantly increases the likelihood of endoscopic mucosal healing in early Crohn's disease) was a real effect[56].

Choice of end point. The design of head-to-head trials is also influenced and sometimes complicated by specific requirements of regulators. According to the European Medicines Agency (EMA) draft guidelines on the development of new drugs published in 2018, the primary end point in studies of ulcerative colitis or luminal Crohn's disease should concern the proportion of patients with symptomatic and endoscopic remission, defined and justified according to validated instruments used for evaluating signs, symptoms and

mucosal inflammation[57,58]. The EMA guidelines are complemented by those of the FDA, although the latter seems more aspirational by suggesting the incorporation of a not-yet-validated histology instrument in the definition of mucosal healing and the development of new patient-reported outcomes (PROs) in ulcerative colitis[59]. The timing of measuring the primary end point depends on the aim of the treatment course (for example, induction versus maintenance) and the pharmacodynamic properties of the tested drugs. A different time point for the assessment of symptomatic and endoscopic remission might be acceptable[57,58]. These are generic proposals that go some way to avoiding the artificial divide between induction and maintenance therapy in IBD, while also focusing on the need to control the biology of the disease, which is inflammation.

Unfortunately, validated cut-offs of clinical and endoscopic scoring systems are often lacking, leading to wide heterogeneity in applied definitions[60–62]. In the STRIDE (selecting therapeutic targets in inflammatory bowel disease) programme, an international specialist panel agreed that the target for clinical therapeutic trials in IBD was a PRO measure combined with objective evidence (imaging). For ulcerative colitis, the PRO was clinical remission, defined as complete resolution of rectal bleeding and altered bowel habit, and endoscopic remission, defined as a Mayo endoscopic subscore (or Ulcerative Colitis Endoscopic Index of Severity score) of 0 or 1 (REF.[2]). In Crohn's disease, PRO remission was defined as resolution of abdominal pain (using a visual analogue scale) and altered bowel habit (using the Bristol stool scale), and imaging remission as resolution of ulcerations at ileocolonoscopy or resolution of inflammatory findings on cross-sectional imaging[2]. A similar consortium defined a set of end points for the development of anti-fibrosis drugs in Crohn's disease[63]. These outcomes have yet to be introduced into regular practice as neither the existing PROs nor the anti-fibrosis drug's end points are prospectively validated[63,64], but the effect of such tight end points can be seen in the studies of budesonide MMX in ulcerative colitis: in the two registration trials, the remission rates with placebo were 5–7%, whereas those with the active agent were only in the region of 18%, although differences between the two remained statistically significant[44]. End points that act as a surrogate marker for future disease activity are elusive, although histopathology might yet prove of value, as concordance between clinical, endoscopic and histological remission has been associated with increased rates of steroid-free remission and a reduction in hospitalization for ulcerative colitis over a succeeding period of 6 years[65].

More stringent outcomes have reduced placebo response rates in IBD[66], but caution is warranted to avoid becoming overly restrictive. A post hoc analysis of data from the EXTEND (extending the safety and efficacy of adalimumab through endoscopic healing) trial assessed the effect of different definitions of clinical and endoscopic remission on treatment efficacy estimations in Crohn's disease[67]. By increasing the stringency of week-12 clinical end points, placebo response rates reduced by at least 12%, whereas absolute treatment effects increased by maximum of 10%. By contrast, when amending the endoscopic end point by lowering the target score, the treatment effect reduced from 24% to 8%, and composite end points further diminished response rates and effect sizes. These findings demonstrate that increasing the stringency of combined clinical endoscopic end point definitions in Crohn's disease trials reduces the ability to detect treatment-related changes in disease activity, so a focus on end points that reflect clinical benefit alone is warranted[67]. As a consequence, quality of life can be a primary outcome: it is the (often unstated) goal of medicine, but is rarely measured in practice. In the CONSTRUCT trial (see earlier), no differences between infliximab and ciclosporin for the treatment of acute severe colitis based on quality-adjusted survival, or the area under the curve of scores from the Crohn's and Ulcerative Colitis Questionnaire, were found[53]. Colectomy and mortality rates were likewise similar between the two treatment groups, but were only secondary outcomes[53]. The other IBD study that has used quality of life as a primary end point is the LIR!C (laparoscopic ileocolic resection versus infliximab treatment of distal ileitis in Crohn's disease) trial, an open-label RCT that demonstrated equal outcomes of infliximab treatment and ileocolic resection in patients with limited, non-stricturing, ileocaecal Crohn's disease[68].

Study populations. One of the drawbacks of head-to-head clinical trials is that they generally use exclusion criteria similar to phase II and phase III trials, which might not be relevant to the general IBD population[32]. Many trials are, at least in part, conducted at academic medical centres, for which the referral population might be more refractory to treatment, so most participants might have received treatment with immunosuppressive and/or biologic agents in the past and have a longer disease duration compared with the general IBD population[66]. Comparative trials should, therefore, critically appraise their inclusion and exclusion criteria, taking into account that in clinical practice, biologic therapy is often started for the cumulative effect of low-to-moderate disease activity and, depending on context, previous response to therapy, physician preference, patient preference and payer provision. One way to make clinical trials more relevant to everyday practice is to use a cluster randomization design, in which randomization is by centre, including all patients in one arm in a single centre, and not by individual patient, as a centre in reality will adopt a unified strategy to all their patients. The REACT (randomized evaluation of an algorithm for Crohn's disease treatment) trial used this approach: 41 centres were randomly assigned to either conventional management or early combined immunosuppression with a TNF antagonist and an antimetabolite. Although the primary end point (control of symptoms) was not different between the groups, the design demonstrated that the risk of major adverse outcomes (disease-related hospitalization, surgery or complications) was lower in the centres randomly assigned to early combined immunosuppression[35]. Almost 2,000 patients were recruited, and the response to early combined immunosuppression was broadly comparable to results in registration trials, suggesting that patients recruited individually to trials are more representative of clinical practice than commonly acknowledged[35].

Dosing and escape arms. Dosing in head-to-head trials comparing biologic treatments should mimic clinical practice. Escape arms including dose escalation and the possibility of switching to another biologic therapy should follow society treatment guidelines. Patients who 'escape' should not be included in the primary outcome, as the escape therapy might improve apparent outcomes.

Defining different dosing and escape arms can be a major hurdle in study design, potentially increasing the burden on participants. For example, in a current double-dummy blinded trial, the effectiveness and safety of infliximab versus etrolizumab is being compared in patients with ulcerative colitis who have not previously received biologic agents (NCT02136069)[69]. The study protocol does not offer the option of

Table 2 | **Principal published head-to-head trials with biologic agents in IBD**

Reference	Comparator	Treatment population	Number of patients	Primary end point	Period[a]	Type of comparison	Blinding	Main finding
Biologic agent vs conventional agent								
Colombel et al. (2010)[50]	Infliximab and azathioprine vs infliximab vs azathioprine	Moderate-to-severe active Crohn's disease, naive to IMM and biologic agents	508	Achieving CS-free clinical remission	26 weeks	Superiority	Double-blinded, double-dummy	Infliximab ± azathioprine superior to azathioprine alone
Laharie et al. (2012)[72]	Ciclosporin vs infliximab	Acute severe ulcerative colitis refractory to IV CS	115	Treatment failure[b]	98 days	Superiority	Non-blinded	Ciclosporin not more effective than infliximab
Panaccione et al. (2014)[49]	Infliximab and azathioprine vs infliximab vs azathioprine	Moderate-to-severe active ulcerative colitis, naive to biologic agents	239	Achieving CS-free clinical remission	16 weeks	Superiority	Double-blinded, double-dummy	Infliximab + azathioprine superior to either monotherapy infliximab or azathioprine
Feagan et al. (2014)[71]	Infliximab and methotrexate vs infliximab	Active Crohn's disease receiving CS, naive to biologic agents	126	Time to treatment failure[c]	50 weeks	Superiority	Double-blinded, double-dummy	Infliximab + methotrexate not superior to infliximab alone
Williams et al. (2016)[53]	Ciclosporin vs infliximab	Acute severe ulcerative colitis refractory to IV CS	270	Quality-adjusted survival	3 years	Non-inferiority	Single-blinded	No difference between ciclosporin and infliximab
Biologic agent vs biosimilar								
Jørgensen et al. (2017)[39]	Infliximab originator vs infliximab biosimilar	Crohn's disease, ulcerative colitis or other IMID, at least 6 months stable infliximab treatment	482	Disease worsening	52 weeks	Non-inferiority	Double-blinded	Switching from infliximab originator to infliximab biosimilar non-inferior to continuation of infliximab originator
Biologic agent vs biologic agent								
Sands et al. (2019)[16]	Vedolizumab IV vs adalimumab SC	Moderate-to-severe active ulcerative colitis	769	Achieving clinical remission	52 weeks	Superiority	Double-blinded, double-dummy	Vedolizumab superior to adalimumab

CS, corticosteroid; IBD, inflammatory bowel disease; IMID, immune-mediated inflammatory disease; IMM, immunomodulator; IV, intravenous; SC, subcutaneous. [a]Time point at which the primary outcome was assessed. [b]Absence of clinical response on day 7 or relapse between day 7 and day 98 or absence of CS-free remission on day 98 or severe adverse event leading to treatment interruption, colectomy or death. [c]Lack of CS-free clinical remission during week 14 or failure to maintain remission to week 50.

dose escalation in instances of treatment failure, although it is well established that therapeutic drug monitoring helps select infliximab-treated patients who could benefit from this approach[70]. Trial participants with non-response or loss of response to infliximab could, therefore, be switched to the etrolizumab arm, whereas they might potentially have regained response to infliximab dose escalation.

Comparing biologic agents for IBD

Existing head-to-head trials. Only a few trials have compared biologic agents with conventional treatment or with each other in patients with IBD (TABLE 2).

Several studies have aimed to show the superiority of combination therapy with immunomodulators and infliximab, compared with either treatment as monotherapy[49,50,71]. The superiority of the 'combotherapy' hypothesis was confirmed for infliximab and azathioprine in both Crohn's disease and ulcerative colitis[49,50], but negative for infliximab and methotrexate in Crohn's disease[71]. In this latter trial, patients receiving combination therapy less frequently developed antibodies to infliximab, hinting that a longer study duration could have altered primary outcomes, although the study duration was already longer than in trials with infliximab and azathioprine (50 weeks versus 16 to 30 weeks)[49,50,71]. Other landmark trials have compared the clinical effectiveness of ciclosporin and infliximab in patients with acute, severe steroid-refractory ulcerative colitis[53,72]. Laharie et al.[72] hypothesized that ciclosporin would less often lead to treatment failure; both participants and investigators were not blinded to treatment allocation and outcomes were assessed early after initiation of therapy. By contrast, Williams et al.[53] designed a non-inferiority trial with a different primary outcome, based on quality of life and with a longer follow-up; the chief investigator and analysts were blinded to treatment allocation.

Despite marked differences between the two study designs, the conclusions were identical: ciclosporin and infliximab are equivalent in the management of acute, severe ulcerative colitis[53,72].

The first comparison of biologic agents for the treatment of IBD confirmed that switching infliximab originator to its biosimilar was not inferior to continuing with the originator[39]. To date (2020), the only published RCT comparing biologic agents with two different modes of action is the VARSITY trial[16]. This trial demonstrated superior clinical remission with vedolizumab than with adalimumab in the treatment of moderate-to-severe active ulcerative colitis. Although the trial included both patients who had not and patients who had received anti-TNF agents, which could have impaired the outcomes in adalimumab-treated patients, the superiority of vedolizumab was more pronounced in those who had not received anti-TNF agents. By contrast, and as already remarked, corticosteroid-free remission rates numerically favoured adalimumab, although only a small percentage of patients treated with corticosteroids at baseline could stop these during the trial (40 of 230 in both arms), which lacked a mandated corticosteroid-tapering regimen. Although this aspect might be explained by residual disease activity, it might also reflect treatment habits in a steroid-dependent population. Notably, treatment optimization through dose adjustment in the VARSITY trial was not possible owing to the study design, although higher dosing of both adalimumab and vedolizumab might have improved clinical outcomes[73,74].

Other head-to-head trials involving biologic agents have focused on the optimized use of established treatment options, such as studies comparing an infliximab dosing strategy based on serum trough concentrations with an infliximab dosing strategy based on clinical features[29,30]. Although both studies failed to show superiority of concentration-based dosing, the results should be interpreted with caution. In the TAXIT (trough concentration adapted infliximab treatment) RCT only a highly selected group of patients treated with maintenance infliximab for at least 14 weeks and in stable clinical remission were included, and the optimization phase balanced trough levels of all participants between prespecified intervals even before randomization. This selection bias might have contributed to the lack of difference in primary outcome between the concentration-based and clinically based dosing arms[29]. In an RCT investigating tailored treatment with infliximab (TAILORIX; tailored treatment with infliximab for active luminal Crohn's disease) some methodological issues existed, as in the concentration-based arms the investigators could also decide to dose optimize infliximab based on clinical symptoms or biomarkers alone[30]. The TAXIT and TAILORIX trials highlight

Table 3 | **Selected ongoing head-to-head trials between biologic agents in IBD**

Study name and/or ClinicalTrial.gov number	Comparator	Type of study	Treatment population	Number of patients (±[a])	Primary end point	Period[b]
NCT02871635[77]	BI695501 (biosimilar) SC vs adalimumab SC	Phase III, randomized, double-blinded	Active luminal Crohn's disease	140	Clinical response	4 weeks
GARDENIA[69] NCT02136069	Etrolizumab SC vs infliximab IV	Phase III, randomized, double-blinded, double-dummy	Moderate-to-severe active ulcerative colitis, naive to biologic agents	390	Clinical response (week 10) and remission (week 54)	10 and 54 weeks
HIBISCUS 1 and 2 (REFS[78,79]) NCT02163759 NCT02171429	Etrolizumab SC vs adalimumab SC vs placebo	Phase III, randomized, double-blinded, double-dummy	Moderate-to-severe active ulcerative colitis, naive to biologic agents	350	Clinical remission	10 weeks
EXPEDITION[80] NCT03616821	Brazikumab IV/SC[c] vs vedolizumab IV vs placebo	Phase II, randomized, double-blinded, double-dummy	Moderate-to-severe active ulcerative colitis	375	Clinical remission	10 weeks
SEAVUE[81] NCT03464136	Ustekinumab IV/SC[c] vs adalimumab SC	Phase III, randomized, double-blinded, double-dummy	Moderate-to-severe active luminal Crohn's disease, naive to biologic agents	350	Clinical remission	52 weeks
VEGA[82] NCT03662542	Guselkumab IV/SC[c] + golimumab SC vs guselkumab IV/SC[c] vs golimumab SC	Phase II, randomized, double-blinded, double-dummy	Moderate-to-severe active ulcerative colitis	210	Clinical response	12 weeks
INTREPID[83] NCT03759288	Brazikumab IV/SC[c] vs adalimumab SC vs placebo	Phase II/III, randomized, double-blinded, double-dummy	Moderate-to-severe active Crohn's disease	1,140	Endoscopic response and clinical remission	12 and 52 weeks
VIVID-1 (REF.[84]) NCT03926130	Mirikizumab IV/SC[c] vs ustekinumab IV/SC[c] vs placebo	Phase III, randomized, double-blinded, double-dummy	Moderate-to-severe active Crohn's disease	1,100	Endoscopic response and clinical remission	52 weeks
GALAXI 1, 2 and 3 (REF.[85]) NCT03466411	Guselkumab IV/SC[c] vs ustekinumab IV/SC[c] vs placebo	Phase II/III, randomized, double-blinded, double-dummy	Moderate-to-severe active luminal Crohn's disease	2,000	Clinical response (phase II) and remission (phase III)	12 weeks

IBD, inflammatory bowel disease; IV, intravenous; SC, subcutaneous. [a]Final number likely to change. [b]Time point at which the primary outcomes were assessed. [c]IV during induction, and continued SC treatment during maintenance.

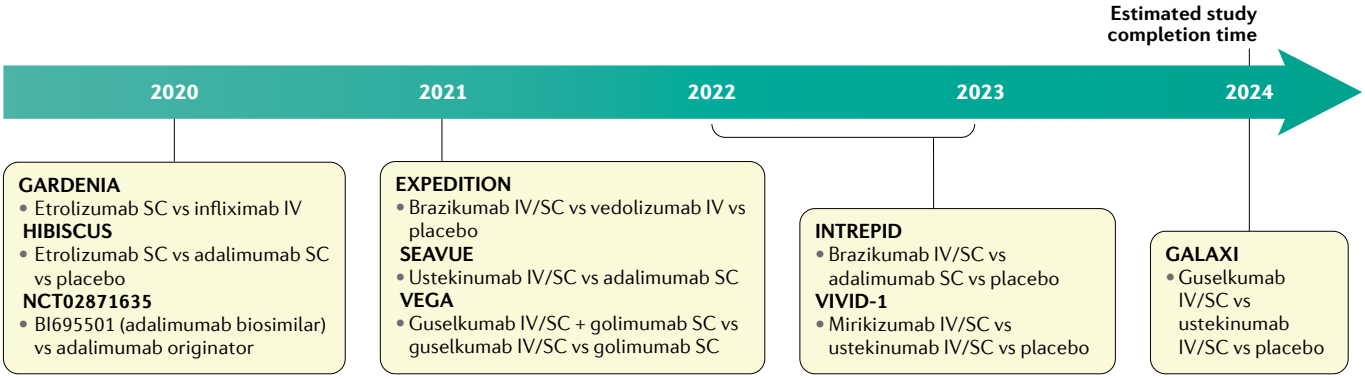

Fig. 4 | **Ongoing head-to-head trials in IBD.** Timeline showing estimated completion time of the main ongoing head-to-head trials between biologic agents in inflammatory bowel disease (IBD) (also see TABLE 3). IV, intravenous; IV/SC, IV during induction and continued SC treatment during maintenance; SC, subcutaneous.

the importance of performing a thorough analysis of a trial's design looking beyond the main outcomes. A detailed review of treatment strategy trials, including other examples comparing different routes of administration of the same drug[75], or top-down and step-up treatment strategies[35,76], is, however, outside the scope of this article, but a concise overview is given in Supplementary Table 4.

Ongoing head-to-head trials. Multiple head-to-head trials comparing biologic agents in the treatment of IBD are in progress. Most contrast biologic agents with different modes of action in a double-blinded, double-dummy setting. Primary end points in all trials are achieving clinical response or remission at pre-specified time points[69,77–85] (TABLE 3; FIG. 4).

Conclusions

Given the increase in the number of novel IBD treatments aiming for more stringent outcomes, comparative research has been prioritized in the current landscape of IBD trials. Largely owing to the request of health authorities and payers, several head-to-head trials between biologic agents for the treatment of IBD have been launched in the past few years, and this trend will continue over the next decade. Although the results of such trials will influence prescribers worldwide, it remains unlikely that the trials alone are going to determine the positioning of biologic therapy. Comparing active treatments in an RCT is expensive and complex and is mainly driven by industry. This approach enhances the likelihood of a favourable response to the experimental treatment, which should be taken into account when interpreting results[38]. Trial design will continue to face immense hurdles, given the evolving understanding of the sophisticated science behind biologic agents in the treatment of IBD. Strategy trials, belonging broadly to the range of head-to-head studies, have enlarged our knowledge about how to optimize the use of established therapies, illustrated by therapeutic drug monitoring or treat-to-target approaches[29,30,56]. Furthermore, biomarkers predicting response to therapy are opening the door to personalized medicine, further complicating concepts of a 'fair' head-to-head study design[86,87]. Head-to-head research might eventually evolve to head-to-head performance trials, comparing therapy allocation and optimization based on different (bio)markers, rather than simply comparing therapeutic agents. To maintain relevance, head-to-head trials between biologic agents should also select and stratify patients more effectively, based on disease and patient biology, while using the same objective measures and PROs as those applied in clinical practice. Several other questions remain unanswered. What is a clinically meaningful difference between two IBD drugs? What is the long-term effect on durability and disease complications of the differences observed between two IBD drugs? And should we favour non-inferiority trials, as efficacy of biologic agents and small molecules seems to have reached a plateau? Finally, the quest for optimization and comparison of established treatment options should not restrain the focus of future IBD research from improving the understanding of disease pathogenesis, as many mechanisms are poorly understood and need to be unravelled to improve long-term outcomes in patients with IBD.

Lieven Pouillon[1], Simon Travis[2], Peter Bossuyt[1], Silvio Danese[3,4] and Laurent Peyrin-Biroulet[5,6]✉

[1]Imelda GI Clinical Research Centre, Imeldaziekenhuis Bonheiden, Bonheiden, Belgium.

[2]Translational Gastroenterology Unit, NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, Oxford, UK.

[3]IBD Center, Department of Gastroenterology, Humanitas Clinical and Research Institute, Milan, Italy.

[4]Department of Biomedical Sciences, Humanitas University, Milan, Italy.

[5]Department of Gastroenterology, Nancy University Hospital, Vandœuvre-lès-Nancy, France.

[6]Inserm U1256 NGERE, Lorraine University, Vandoeuvre-lès-Nancy, France.

✉*e-mail: peyrinbiroulet@gmail.com*

https://doi.org/10.1038/s41575-020-0293-9

Published online: 17 April 2020

1. Truelove, S. C. & Witts, L. J. Cortisone in ulcerative colitis. *BMJ* **2**, 1041–1048 (1955).
2. Peyrin-Biroulet, L. et al. Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. *Am. J. Gastroenterol.* **110**, 1324–1338 (2015).
3. Chateau, T. et al. Histological remission in ulcerative colitis. *Am. J. Gastroenterol.* **115**, 179–189 (2020).
4. Sabino, J., Verstockt, B., Vermeire, S. & Ferrante, M. New biologics and small molecules in inflammatory bowel disease: an update. *Therap. Adv. Gastroenterol.* **12**, 175628481985320 (2019).
5. Summers, R. W. et al. National Cooperative Crohn's Disease Study: results of drug treatment. *Gastroenterology* **77**, 847–869 (1979).
6. Courtney, M. G. et al. Randomised comparison of olsalazine and mesalazine in prevention of relapses in ulcerative colitis. *Lancet* **339**, 1279–1281 (1992).
7. Kamm, M. A. et al. Once-daily, high-concentration MMX mesalamine in active ulcerative colitis. *Gastroenterology* **132**, 66–75 (2007).
8. Jeong, D. Y. et al. Induction and maintenance treatment of inflammatory bowel disease: a comprehensive review. *Autoimmun. Rev.* **18**, 439–454 (2019).
9. Weinblatt, M. E. et al. Head-to-head comparison of subcutaneous abatacept versus adalimumab for rheumatoid arthritis: findings of a phase IIIb, multinational, prospective, randomized study. *Arthritis Rheum.* **65**, 28–38 (2013).
10. Durelli, L. et al. Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). *Lancet* **359**, 1453–1460 (2002).
11. Gordon, K. B. et al. Efficacy and safety of risankizumab in moderate-to-severe plaque psoriasis (UltIMMa-1 and UltIMMa-2): results from two double-blind, randomised, placebo-controlled and ustekinumab-controlled phase 3 trials. *Lancet* **392**, 650–661 (2018).
12. Dulai, P. S., Singh, S., Ohno-Machado, L. & Sandborn, W. J. Population health management for inflammatory bowel disease. *Gastroenterology* **154**, 37–45 (2018).

13. Ma, C. et al. Systematic review with meta-analysis: endoscopic and histologic placebo rates in induction and maintenance trials of ulcerative colitis. *Aliment. Pharmacol. Ther.* **47**, 1578–1596 (2018).
14. Duijvestein, M. et al. Response to placebo, measured by endoscopic evaluation of Crohn's disease activity, in a pooled analysis of data from 5 randomized controlled induction trials. *Clin. Gastroenterol. Hepatol*. https://doi.org/10.1016/j.cgh.2019.08.025 (2019).
15. Peyrin-Biroulet, L., Lopez, A. & Sandborn, W. Head-to-head comparative studies: challenges and opportunities? *J. Crohns Colitis* **11**, S567–S575 (2017).
16. Sands, B. E. et al. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. *N. Engl. J. Med.* **381**, 1215–1226 (2019).
17. Digby-Bell, J. L., Atreya, R., Monteleone, G. & Powell, N. Interrogating host immunity to predict treatment response in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 9–20 (2020).
18. Liberati, A. et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* **339**, b2700 (2009).
19. Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. Meta-analysis and the science of research synthesis. *Nature* **555**, 175–182 (2018).
20. Ioannidis, J. P. A. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q.* **94**, 485–514 (2016).
21. Rogozin´ska, E., Marlin, N., Thangaratinam, S., Khan, K. S. & Zamora, J. Meta-analysis using individual participant data from randomised trials: opportunities and limitations created by access to raw data. *Evid. Based Med.* **22**, 157–162 (2017).
22. Rouse, B., Chaimani, A. & Li, T. Network meta-analysis: an introduction for clinicians. *Intern. Emerg. Med.* **12**, 103–111 (2017).
23. Ghosh, S. et al. Interpreting registrational clinical trials of biological therapies in adults with inflammatory bowel diseases. *Inflamm. Bowel Dis.* **22**, 2711–2723 (2016).
24. Corrigan-Curay, J., Sacks, L. & Woodcock, J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA* **320**, 867–868 (2018).
25. Blonde, L., Khunti, K., Harris, S. B., Meizinger, C. & Skolnik, N. S. Interpretation and impact of real-world clinical data for the practicing clinician. *Adv. Ther.* **35**, 1763–1774 (2018).
26. Olivera, P., Danese, S., Jay, N., Natoli, G. & Peyrin-Biroulet, L. Big data in IBD: a look into the future. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 312–321 (2019).
27. Ha, C., Ullman, T. A., Siegel, C. A. & Kornbluth, A. Patients enrolled in randomized controlled trials do not represent the inflammatory bowel disease patient population. *Clin. Gastroenterol. Hepatol.* **10**, 1002–1007 (2012).
28. Pouillon, L. et al. Mucosal healing and long-term outcomes of patients with inflammatory bowel diseases receiving clinic-based vs trough concentration-based dosing of infliximab. *Clin. Gastroenterol. Hepatol.* **16**, 1276–1283 (2018).
29. Vande Casteele, N. et al. Trough concentrations of infliximab guide dosing for patients with inflammatory bowel disease. *Gastroenterology* **148**, 1320–1329 (2015).
30. D'Haens, G. et al. Increasing infliximab dose based on symptoms, biomarkers, and serum drug concentrations does not increase clinical, endoscopic, and corticosteroid-free remission in patients with active luminal Crohn's disease. *Gastroenterology* **154**, 1343–1351 (2018).
31. Austin, P. C. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar. Behav. Res.* **46**, 399–424 (2011).
32. Fleischmann, R., Landewé, S. & Smolen, J. Review of head-to-head study designs in rheumatoid arthritis. *Semin. Arthritis Rheum.* **46**, 279–285 (2016).
33. Faleck, D. et al. Comparative effectiveness of vedolizumab and TNF-antagonist therapy in ulcerative colitis: a multicentre consortium propensity score-matched analysis [abstract OP026]. *J. Crohns Colitis* **12**, S019 (2018).
34. Koliani-Pace, J. L. et al. Changes in vedolizumab utilization across US academic centers and community practice are associated with improved effectiveness and disease outcomes. *Inflamm. Bowel Dis.* **25**, 1854–1861 (2019).
35. Khanna, R. et al. Early combined immunosuppression for the management of Crohn's disease (REACT): a cluster randomised controlled trial. *Lancet* **386**, 1825–1834 (2015).
36. Armstrong, K. Methods in comparative effectiveness research. *J. Clin. Oncol.* **30**, 4208–4214 (2012).
37. Lathyris, D. N., Patsopoulos, N. A., Salanti, G. & Ioannidis, J. P. A. Industry sponsorship and selection of comparators in randomized clinical trials. *Eur. J. Clin. Invest.* **40**, 172–182 (2010).
38. Flacco, M. E. et al. Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *J. Clin. Epidemiol.* **68**, 811–820 (2015).
39. Jørgensen, K. K. et al. Switching from originator infliximab to biosimilar CT-P13 compared with maintained treatment with originator infliximab (NOR-SWITCH): a 52-week, randomised, double-blind, non-inferiority trial. *Lancet* **389**, 2304–2316 (2017).
40. Olivera, P. et al. Physicians' perspective on the clinical meaningfulness of inflammatory bowel disease trial results: an International Organization for the Study of Inflammatory Bowel Disease (IOIBD) survey. *Aliment. Pharmacol. Ther.* **47**, 773–783 (2018).
41. Dunn, D. T., Copas, A. J. & Brocklehurst, P. Superiority and non-inferiority: two sides of the same coin? *Trials* **19**, 499 (2018).
42. Tsui, M., Rehal, S., Jairath, V. & Kahan, B. C. Most noninferiority trials were not designed to preserve active comparator treatment effects. *J. Clin. Epidemiol.* **110**, 82–89 (2019).
43. Wan, M. T. et al. Head-to-head trials of systemic psoriasis therapies: a systematic review of study design and maximum acceptable treatment differences. *J. Eur. Acad. Dermatol. Venereol.* **33**, 42–55 (2019).
44. Sandborn, W. J. et al. Induction of clinical and colonoscopic remission of mild-to-moderate ulcerative colitis with budesonide MMX 9 mg: pooled analysis of two phase 3 studies. *Aliment. Pharmacol. Ther.* **41**, 409–418 (2015).
45. Gross, V. et al. 3g mesalazine granules are superior to 9mg budesonide for achieving remission in active ulcerative colitis: a double-blind, double-dummy, randomised trial. *J. Crohns Colitis* **5**, 129–138 (2011).
46. Lichtenstein, G. R. et al. Effect of once- or twice-daily MMX mesalamine (SPD476) for the induction of remission of mild to moderately active ulcerative colitis. *Clin. Gastroenterol. Hepatol.* **5**, 95–102 (2007).
47. Kamm, M. A. et al. Effect of extended MMX mesalamine therapy for acute, mild-to-moderate ulcerative colitis. *Inflamm. Bowel Dis.* **15**, 1–8 (2009).
48. World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. *WMA* https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (2018)
49. Panaccione, R. et al. Combination therapy with infliximab and azathioprine is superior to monotherapy with either agent in ulcerative colitis. *Gastroenterology* **146**, 392–400 (2014).
50. Colombel, J. F. et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. *N. Engl. J. Med.* **362**, 1383–1395 (2010).
51. Schulz, K. F. & Grimes, D. A. Blinding in randomised trials: hiding who got what. *Lancet* **359**, 696–700 (2002).
52. Wan, M., Orlu-Gul, M., Legay, H. & Tuleu, C. Blinding in pharmacological trials: the devil is in the details. *Arch. Dis. Child.* **98**, 656–659 (2013).
53. Williams, J. G. et al. Infliximab versus ciclosporin for steroid-resistant acute severe ulcerative colitis (CONSTRUCT): a mixed methods, open-label, pragmatic randomised trial. *Lancet Gastroenterol. Hepatol.* **1**, 15–24 (2016).
54. Yoo, D. H. et al. Efficacy and safety of CT-P13 (biosimilar infliximab) in patients with rheumatoid arthritis: comparison between switching from reference infliximab to CT-P13 and continuing CT-P13 in the PLANETRA extension study. *Ann. Rheum. Dis.* **76**, 355–363 (2017).
55. Park, W. et al. Efficacy and safety of switching from reference infliximab to CT-P13 compared with maintenance of CT-P13 in ankylosing spondylitis: 102-week data from the PLANETAS extension study. *Ann. Rheum. Dis.* **76**, 346–354 (2017).
56. Colombel, J.-F. et al. Effect of tight control management on Crohn's disease (CALM): a multicentre, randomised, controlled phase 3 trial. *Lancet* **390**, 2779–2789 (2017).
57. European Medicines Agency. Guideline on the Development of New Medicinal Products for the Treatment of Ulcerative Colitis. *EMA* https://www.ema.europa.eu/en/documents/scientific- guideline/guideline-development-new-medicinal-products-treatment-ulcerative-colitis-revision-1_en.pdf (2018).
58. European Medicines Agency. Guideline on the Development of New Medicinal Products for the Treatment of Crohn's Disease. *EMA* https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-development-new-medicinal-products-treatment-crohns-disease-revision-2_en.pdf (2018).
59. Reinisch, W. et al. Comparison of the EMA and FDA guidelines on ulcerative colitis drug development. *Clin. Gastroenterol. Hepatol.* **17**, 1673–1679.e1 (2019).
60. Ma, C. et al. Heterogeneity in definitions of efficacy and safety endpoints for clinical trials of Crohn's disease: a systematic review. *Clin. Gastroenterol. Hepatol.* **16**, 1407–1419 (2018).
61. Vuitton, L. et al. Defining endoscopic response and remission in ulcerative colitis clinical trials: an international consensus. *Aliment. Pharmacol. Ther.* **45**, 801–813 (2017).
62. Vuitton, L. et al. IOIBD technical review on endoscopic indices for Crohn's disease clinical trials. *Gut* **65**, 1447–1455 (2016).
63. Danese, S. et al. Identification of endpoints for development of antifibrosis drugs for treatment of Crohn's disease. *Gastroenterology* **155**, 76–87 (2018).
64. Pittet, V. E. H. et al. Differences in outcomes reported by patients with inflammatory bowel diseases vs their health care professionals. *Clin. Gastroenterol. Hepatol.* **17**, 2050–2059 (2019).
65. Bryant, R. V. et al. Beyond endoscopic mucosal healing in UC: histological remission better predicts corticosteroid use and hospitalisation over 6 years of follow-up. *Gut* **65**, 408–414 (2016).
66. D'Haens, G. et al. Challenges to the design, execution, and analysis of randomized controlled trials for inflammatory bowel disease. *Gastroenterology* **143**, 1461–1469 (2012).
67. Feagan, B. et al. Performance of Crohn's disease clinical trial endpoints based upon different cutoffs for patient reported outcomes or endoscopic activity: analysis of EXTEND data. *Inflamm. Bowel Dis.* **24**, 932–942 (2018).
68. Ponsioen, C. Y. et al. Laparoscopic ileocaecal resection versus infliximab for terminal ileitis in Crohn's disease: a randomised controlled, open-label, multicentre trial. *Lancet Gastroenterol. Hepatol.* **2**, 785–792 (2017).
69. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT02136069 (2020).
70. Vande Casteele, N., Herfarth, H., Katz, J., Falck-Ytter, Y. & Singh, S. American Gastroenterological Association Institute technical review on the role of therapeutic drug monitoring in the management of inflammatory bowel diseases. *Gastroenterology* **153**, 835–857.e6 (2017).
71. Feagan, B. G. et al. Methotrexate in combination with infliximab is no more effective than infliximab alone in patients with Crohn's disease. *Gastroenterology* **146**, 681–688 (2014).
72. Laharie, D. et al. Ciclosporin versus infliximab in patients with severe ulcerative colitis refractory to intravenous steroids: a parallel, open-label randomised controlled trial. *Lancet* **380**, 1909–1915 (2012).
73. Paul, S. et al. Pharmacokinetics of adalimumab in inflammatory bowel diseases: a systematic review and meta-analysis. *Inflamm. Bowel Dis.* **20**, 1288–1295 (2014).
74. Peyrin-Biroulet, L. et al. Loss of response to vedolizumab and ability of dose intensification to restore response in patients with Crohn's disease or ulcerative colitis: a systematic review and meta-analysis. *Clin. Gastroenterol. Hepatol.* **17**, 838–846 (2019).
75. Sandborn, W. J. et al. Efficacy and safety of vedolizumab subcutaneous formulation in a randomized trial of patients with ulcerative colitis. *Gastroenterology* **158**, 562–572.e12 (2020).
76. D'Haens, G. et al. Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. *Lancet* **371**, 660–667 (2008).
77. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT02871635 (2019).
78. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT02163759 (2020).

79. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT02171429 (2020).
80. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03616821 (2020).
81. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03464136 (2020).
82. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03662542 (2020).
83. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03759288 (2020).
84. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/record/NCT03926130 (2020).
85. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT03466411 (2020).
86. Verstockt, B. et al. Low TREM1 expression in whole blood predicts anti-TNF response in inflammatory bowel disease. *EBioMedicine* **40**, 733–742 (2019).
87. Dulai, P. S. et al. Approaches to integrating biomarkers into clinical trials and care pathways as targets for treatment of inflammatory bowel diseases. *Gastroenterology* **157**, 1032–1043.e1 (2019).

Author contributions
L.P., S.T. and L.P.-B. researched data for and wrote the article. All authors made substantial contributions to discussion of content and reviewed/edited the manuscript before submission.

Competing interests
L.P. received personal fees from Abbvie, Ferring, Takeda; and advisory board fees from Takeda. S.T. received research support from AbbVie, Buhlmann, Celgene, International Organization for the Study of Inflammatory Bowel Disease, Janssen, Lilly, Takeda, UCB, Vifor and Norman Collisson Foundation; consulting fees from AbbVie, Allergan, Amgen, Arena, Asahi, Astellas, Biocare, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Buhlmann, Celgene, Chemocentryx, Cosmo, Enterome, Ferring, Giuliani SpA, GSK, Genentech, Gilead, Immunocore, Immunometabolism, Indigo, Janssen, Lexicon, Lilly, Merck, MSD, Neovacs, Novartis, NovoNordisk, NPS Pharmaceuticals, Pfizer, Proximagen, Receptos, Roche, Sandoz, Sensyne, Shire, Sigmoid Pharma, SynDermix, Takeda, Theravance, Tillotts, Topivert, UCB, VHsquared, Vifor and Zeria; speaker fees from AbbVie, Amgen, Biogen, Ferring, Janssen, Shire and Takeda; and declares no stock or share options. P.B. received educational grants from AbbVie, Mundipharma, Pfizer and Janssen; speaker fees from AbbVie, Takeda and Pfizer; and advisory board fees from Hospira, Janssen, MSD, Mundipharma, Roche, Pfizer, Takeda, Sandoz and Pentax. S.D. has served as a speaker, a consultant and an advisory board member for Abbvie, Ferring, Hospira, Johnson & Johnson, Merck, Millennium Takeda, Mundipharma, Pfizer, Tigenix, UCB Pharma and Vifor. L.P.-B. received personal fees from AbbVie, Janssen, Genentech, Ferring, Tillots, Pharmacosmos, Celltrion, Takeda, Boerhinger Ingelheim, Pfizer, Index Pharmaceuticals Sandoz, Celgene, Biogen, Samsung Bioepis, Alma, Sterna, Nestle, Enterome, Allergan, MSD, Roche, Arena, Gilead, Hikma, Amgen, BMS, Vifor, Norgine, Mylan, Lilly, Fresenius Kabi, Oppilan Pharma, Sublimity Therapeutics, Applied Molecular Transport, OSE Immunotherapeutics, Enthera and Theravance; grants from Abbvie, MSD and Takeda; and stock options from CTMA.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41575-020-0293-9.

RELATED LINKS
ClinicalTrials.gov: https://www.clinicaltrials.gov/

© Springer Nature Limited 2020