OXFORD

Review Article

The Impact of Raising the Bar for Clinical Trials in Ulcerative Colitis

Bruce E. Sands,[a] Adam S. Cheifetz,[b] Chudy I. Nduaka,[c] Daniel Quirk,[c] Wenjin Wang,[c] Eric Maller,[c] Gary S. Friedman,[c] Chinyu Su,[c] Peter D. R. Higgins[d]

[a]Dr Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA [b]Department of Medicine and Division of Gastroenterology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA [c]Pfizer Inc, Collegeville, PA, USA [d]University of Michigan, Ann Arbor, MI, USA

Corresponding author: Bruce E. Sands, MD, Dr Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1069, New York, NY 10029, USA. Tel.: 212-241-6744; fax: 646-537-8647; Email: bruce.sands@mssm.edu

Abstract

In order to identify the practical implications for both health care practitioners and patients in understanding differences between the results of trials assessing therapies for ulcerative colitis [UC], we reviewed clinical trials of therapies for moderate to severe UC, with a focus on trial design. Over time, patient populations in UC trials have become more refractory, reflecting that patients are failing treatment with additional and different classes of drug, including conventional therapies, immunosuppressant drugs, and anti-tumour necrosis factor therapies. Outcomes used to measure efficacy have become increasingly stringent in order to meet the expectations of patients and physicians, and the requirements of regulatory bodies. Trial design has also evolved to integrate induction and maintenance therapy phases, so as to facilitate patient recruitment and to answer clinically important questions such as how efficacious therapies are in specific subpopulations of patients and during long-term use. As UC clinical trial design continues to evolve, and with limited head-to-head trials and real-world comparative effectiveness studies evaluating UC therapies, careful judgment is required to appreciate the differences and similarities in trial designs, and to understand how these variances may affect the observed efficacy and safety outcomes.

Key Words: Ulcerative colitis; inflammatory bowel disease; tofacitinib; tumour necrosis factor inhibitor therapy; anti-integrin therapy; small molecule; Janus kinase inhibitor; clinical trials

1. Introduction

Ulcerative colitis [UC] is a chronic condition characterised by mucosal inflammation of the colon.[1] Appropriate therapy depends on the activity, severity, and extent of disease.[2,3] For patients with moderate to severe UC, treatment options include traditional immunosuppressant therapies [e.g., thiopurines], biologic therapies,[3,4] and a Janus kinase [JAK] inhibitor. Biologic drugs approved for use in patients with moderate to severe disease include tumour necrosis factor [TNF]-alpha antagonists [infliximab, adalimumab, and golimumab][5–9] and

anti-integrin therapy [vedolizumab].[10,11] Tofacitinib is an oral, small-molecule JAK inhibitor approved for the treatment of UC.

As new therapies have become available, the design of clinical trials assessing their use for UC has evolved, reflecting changes in patient populations, patients' and physicians' treatment expectations, and regulatory body requirements.[12,13] Critically, there remains confusion surrounding these evolving aspects of UC trial design. Clinicians and patients may not take into account important differences between trials, which may lead to inappropriate comparisons

across trials of disparate design, and inappropriate conclusions being drawn. With limited head-to-head trials, comparing results among trials of differing design is challenging. Network meta-analyses comparing efficacy and safety of therapies for moderate to severe UC have been performed,[14,15] but such comparisons should also be interpreted cautiously due to acknowledged limitations of comparing trials conducted in different patient populations and with different designs, including the influence of assumptions made when adjusting data from different trials.[16] Furthermore, UC clinical trials enroll patients who may not accurately reflect the disease burden in the overall patient population, making it challenging for physicians to generalise results from clinical trials to the patients they see in clinical practice.[17]

We reviewed clinical trials of advanced therapies for the treatment of moderate to severe UC, including anti-TNF agents, anti-integrin therapies, and small-molecule therapies, focusing on trial design. We aimed to identify practical implications for health care practitioners and patients in understanding differences between trial results, and to identify goals for future trial design that would better inform clinicians and patients regarding appropriate UC therapy choice.

In the following sections of the manuscript, we review the characteristics of the patient populations in UC trials [including previous and concomitant medications], approaches taken to the overall design of UC trials, and efficacy endpoints used in UC trials [including the use of central and site-read endoscopy].

2. Design of Clinical Trials of Advanced UC Therapies

We reviewed clinical trials of therapies targeting patients with moderate to severe UC, to highlight aspects of modern UC trial design relevant to clinicians treating this patient population. Clinical trial programmes included in this review [Table 1] covered: anti-TNF agents (infliximab [ACT], adalimumab [ULTRA], and golimumab [PURSUIT]); anti-integrin therapies (vedolizumab [GEMINI] and etrolizumab [HICKORY]); and small-molecule therapies (tofacitinib [OCTAVE], a JAK inhibitor, and ozanimod [TOUCHSTONE], a sphingosine-1-phosphate receptor agonist).

2.1. Patient populations in UC trials

Patient populations in UC trials have become increasingly refractory, as selection criteria have evolved to include previous exposure to and/or failure with biologic therapies including anti-TNF agents. The trials included in this review differed in their inclusion of patients with previous exposure to anti-TNF therapies and had different requirements for previous failed therapies [Tables 2 and 3].

Whereas ULTRA 2 included patients who had secondary loss of response with anti-TNF [specifically infliximab, the only anti-TNF approved for UC at that time], it excluded patients with primary non-response; 40% of patients in ULTRA 2 had secondary loss of response. Approximately 39% of patients in GEMINI and 51–54% in OCTAVE had previous anti-TNF failure [primary or secondary loss of response].

Concomitant therapy with 5-aminosalicylates and corticosteroids was permitted in all completed trials. Across all trials, 31–55% of patients received concomitant thiopurines, except for OCTAVE and TOUCHSTONE where thiopurines were prohibited.

Among trials that included a maintenance phase, corticosteroid tapering at the beginning of the maintenance phase was mandatory in ACT, PURSUIT, OCTAVE Sustain, and GEMINI, and at the discretion of the investigator in ULTRA 2 and TOUCHSTONE. In GEMINI, patients who could not tolerate tapering were permitted to resume their steroid dose from the start of the induction phase, with tapering to resume thereafter—this was not considered rescue therapy and did not require patients to discontinue the study. In OCTAVE Sustain, patients who could not tolerate tapering were permitted to step back up to the earlier steroid dose, with tapering to resume thereafter. Unsuccessful steroid taper counted as treatment failure.

Average disease duration at baseline of the induction phase ranged from 5.4–8.4 years across trials. The proportion of patients with extensive colitis or pancolitis, vs left-sided colitis or other distributions, also varied between studies. A minimum disease duration was mandated for eligibility in ULTRA 2 [3 months] and OCTAVE [4 months]. No minimum disease duration was pre-specified in PURSUIT-SC. Most trials had selection criteria that excluded patients with isolated proctitis. Mean Mayo scores at induction baseline were similar across the trials, ranging from 8.3 to 9.1. Where reported, median C-reactive protein levels at induction baseline ranged from 3.2 to 10.0 mg/l across trials. GEMINI reported faecal calprotectin but not C-reactive protein.

In general, when interpreting efficacy results in clinical trials of patients with moderate to severe UC, the greater the number of therapy classes a patient population has been exposed to and/or failed treatment with, the more refractory that patient population. Accordingly, demonstration of efficacy in the most refractory patient populations may represent a higher bar. The level of inflammatory burden in a patient population—whether measured by endoscopic assessment, or by faecal calprotectin and/or C-reactive protein levels to the extent with which they correlate with inflammation—is also an important consideration. A patient population with low inflammatory burden may be unlikely to demonstrate a significant reduction in inflammatory burden, potentially complicated by a high

Table 1. Clinical trial programmes of advanced therapies for the treatment of ulcerative colitis.

Therapy	Clinical trial	ClinicalTrials.gov registration number	Years conducted
Infliximab[5]	ACT 1 and ACT 2 [phase 3 induction and maintenance]	NCT00036439 and NCT00096655	2002–2005
Adalimumab[6,7]	ULTRA 1 [phase 3 induction],	NCT00385736,	2007–2010
	ULTRA 2 [phase 3 induction and maintenance]	NCT00408629	2006–2010
Golimumab[8,9]	PURSUIT-SC [phase 2/3 induction],	NCT00487539,	2007–2010
	PURSUIT-M [phase 3 maintenance]	NCT00488631	2007–2011
Vedolizumab[10,11]	GEMINI 1 [phase 3 induction], GEMINI 1 [phase 3 maintenance]	NCT00783718	2008–2012
Tofacitinib[18]	OCTAVE Induction 1 and OCTAVE Induction 2 [phase 3 induction],	NCT01465763 and NCT01458951,	2012–2016
	OCTAVE Sustain [phase 3 maintenance]	NCT01458574	2012–2016
Ozanimod[19]	TOUCHSTONE [phase 2 induction and maintenance]	NCT01647516	2012–2015
Etrolizumab[20]	HICKORY [phase 3 induction and maintenance]	NCT02100696	2014–ongoing

Table 2. Baseline characteristics of patient populations in clinical trials of advanced therapies for the treatment of ulcerative colitis.

Clinical trial	Number of patients	Previous anti-TNF treatment, %	Previous anti-TNF failure, %	Concomitant thiopurines, %	Corticosteroid use at baseline, %	Disease duration, years	Disease location, %	Mean baseline Mayo score	Median baseline C-reactive protein, mg/l
Infliximab									
ACT 1 and ACT 2 [induction and maintenance]	ACT 1: 364 ACT 2: 364	0	0	42–55	49–65	Mean: 5.9–8.4	Left-sided: 53–63 Extensive: 38–47	8.3–8.5	6.0–10.0
Adalimumab									
ULTRA 1 [induction],	576	0	0	39–40	55–69	Median: 5.4–6.9	Left-sided: 32–47 Extensive: 46–56	8.7–9.0	3.2–6.4
ULTRA 2 [induction and maintenance]	494	40[a]	40[a]	35	59	Mean: 8.3	Left-sided: 39 Pancolitis: 49	8.9	4.1
Golimumab									
PURSUIT-SC [induction],	1065	0	0	31	45	Mean: 6.3	Left-sided: 58 Extensive: 42	8.5	4.8
PURSUIT-M [maintenance]	1228	0	0	31	48	Mean: 6.5	–	8.4	4.6
Vedolizumab									
GEMINI 1 [induction],	895	48	39[b]	34	54	Mean: 6.9	Left-sided: 38 Pancolitis: 37	8.6	N/A
GEMINI 1 [maintenance]	373	37–42	32[b]	35–40	57–58	Mean: 6.2–7.8	Left-sided: 36–42 Extensive: 11–13 Pancolitis: 32–42	8.3–8.4	N/A
Tofacitinib									
OCTAVE Induction 1 & 2 [induction],	1139	53–58	51–54	N/A	45–49	Median: 6.0–6.5	Left-sided: 30–35 Extensive/pancolitis: 49–54	8.9–9.1	4.4–5.0
OCTAVE Sustain [maintenance]	593	46–51	42–47	N/A	44–51	Median: 6.5–7.2	Left-sided: 31–34 Extensive/pancolitis: 52–55	3.3–3.4	0.7–1.0
Ozanimod									
TOUCHSTONE [induction and maintenance]	197	15–20	Not reported	N/A	34–40	Mean: 5.9–6.7	Left-sided: 61–63 Extensive: 37–39	8.3–8.6	3.9–4.9

Data reported are for the overall trial populations where available. Where these were not available, ranges are presented across the treatment groups included in the trial.

[a]ULTRA 2 included patients who had secondary loss of response with anti-TNF treatment, but excluded patients with primary non-response to anti-TNF treatment. The proportion of patients with previous anti-TNF failure in ULTRA 2 represents patients with secondary loss of response only.

[b]Data from vedolizumab prescribing information.[11] N/A, not applicable; TNF, tumour necrosis factor.

placebo response. Very high levels of inflammation may indicate a refractory patient population or a population in which response is likely to occur more slowly.

The implications of mandatory vs discretionary and/or variable tapering of corticosteroids is also of importance when comparing active treatment vs placebo. Where discretionary corticosteroid tapering is used, efficacy in patients on active treatment who are able to tolerate the corticosteroid taper may be masked by the propensity of non-responsive patients on placebo to remain on corticosteroids. Mandatory tapering of corticosteroids may allow for

Table 3. Permitted previous and concomitant therapies in clinical trials of advanced therapies for the treatment of ulcerative colitis.

Clinical trial	Prohibited concomitant therapies	Permitted concomitant therapies [minimum duration of previous treatment]	Tapering of concomitant corticosteroids
Infliximab ACT 1 and ACT 2 [induction and maintenance]	• Rectally administered corticosteroids or rectal 5-ASA [2 weeks]	• Corticosteroids [minimum not stated] • Thiopurines [minimum not stated] • 5-ASA [minimum not stated; ACT 2 only]	Mandatory attempt; after Week 8: 5 mg/week until a dose of 20 mg/day; thereafter, 2.5 mg/week until discontinuation
Adalimumab ULTRA 1 [induction], ULTRA 2 [induction and maintenance]	• Intravenously administered corticosteroids [2 weeks] • Cyclosporine, tacrolimus, mycophenolate mofetil, or methotrexate [30 days] • Therapeutic enema or suppository [14 days] • Investigational drugs [30 days or five half-lives]	• Corticosteroids ≥20 mg/day [14 days] • Corticosteroids <20 mg/day [40 days] • Azathioprine ≥1.5 mg/kg/day or 6-mercaptopurine ≥1 mg/kg/day [90 days; stable for 28 days] • 5-ASA [stable dose; minimum not stated]	Not mandatory; after Week 8: at the discretion of the investigator
Golimumab PURSUIT-SC [induction], PURSUIT-M [maintenance]	• Anti-TNF, B-, or T-cell-depleting agents [12 months] • Cyclosporine, tacrolimus, sirolimus, mycophenolate mofetil [8 weeks] • Investigational drugs [five half-lives]	• Corticosteroids ≤40 mg/day [stable for 2 weeks] • Thiopurines [stable for 4 weeks] • 5-ASA [stable for 2 weeks]	Mandatory attempt; from Week 1 of maintenance: 5 mg/week [for doses >20 mg/day] or 2.5 mg/week [for doses ≤20 mg/day]
Vedolizumab GEMINI 1 [induction], GEMINI 1 [maintenance]	• Anti-TNF [60 days] • Cyclosporine, thalidomide, or investigational drugs [30 days]	• Corticosteroids ≤30 mg/day [stable for 4 weeks; 2 weeks if being tapered] • Thiopurines [stable for 8 weeks] • 5-ASA [stable for 2 weeks] • Probiotics [stable for 2 weeks] • Anti-diarrhoeals [no minimum]	Mandatory attempt if clinical response achieved; from Week 6 or as soon as clinical response achieved: 5 mg/week [for doses >10 mg/day] or 2.5 mg/week [for doses ≤10 mg/day]; dose could be increased to the original dose with tapering to resume within 2 weeks
Tofacitinib OCTAVE Induction 1 & 2 [induction], OCTAVE Sustain [maintenance]	• Thiopurines or methotrexate [2 weeks] • Anti-TNF or interferon therapy [8 weeks] • Intravenously administered corticosteroids or rectally administered corticosteroids or 5-ASA [2 weeks] • Anti-adhesion molecule therapy, lymphocyte depleting agents, other immunosuppressants, or immunomodulatory biologics [1 year]	• Oral glucocorticoids ≤25 mg/day [stable throughout induction] • 5-ASA [stable throughout induction]	Mandatory attempt from Week 1 of maintenance: 5 mg/week [for doses >20 mg/day] or 2.5–5 mg/week [for doses 11–20 mg/day] or 2.5 mg/week [for doses ≤10 mg/day]; dose could be increased once during the study to the previous dose with tapering subsequently resumed to achieve steroid-free status. Inability to complete taper was counted as treatment failure per protocol
Ozanimod TOUCHSTONE [induction and maintenance]	• Immunomodulatory biologics [4 months] • Biologic agent or investigational drug [five half-lives] • Rectally administered steroids [2 weeks] • Live vaccine [4 weeks]	• Corticosteroids ≤20 mg/day [4 weeks, stable for 2 weeks] • 5-ASA [6 weeks, stable for 3 weeks]	Not mandatory; after Week 8: at the discretion of the investigator

Corticosteroid doses given are for prednisone or equivalent dose of other corticosteroid.

5-ASA, 5-aminosalicylates; TNF, tumour necrosis factor.

a more robust comparison of active therapy vs placebo, and greater weight may be assigned to efficacy achieved without corticosteroids. However, other relevant factors of trial design also need to be accounted for.

2.2. Integration of induction and maintenance phases in UC clinical trial programmes

The trials included in this review took varying approaches to the progression of patients from induction to maintenance phases. In the treat-through approach [Figure 1A], patients are randomised to active treatment or placebo for the duration of the study from induction through maintenance. In an integrated approach, patients are randomised into an induction randomised controlled trial [RCT], with eligible patients [typically those who are responders at the time of induction efficacy assessment] then re-randomised into a maintenance RCT [Figure 1B]. In the approach depicted in Figure 1C, an additional open-label active treatment arm is included in the induction phase to provide sufficient eligible patients for re-randomisation into the maintenance phase RCT.

ACT 1 and ACT 2 used a treat-through approach and assessed infliximab induction and maintenance efficacy, with patients being followed from randomisation through to Week 54 [ACT 1] and Week 30 [ACT 2]. ULTRA 1 was a stand-alone trial assessing adalimumab induction efficacy over 8 weeks. ULTRA 2 assessed adalimumab induction and maintenance efficacy with patients being followed from randomisation through to Week 52. TOUCHSTONE also had a treat-through design with patients followed for up to 32 weeks.

PURSUIT, OCTAVE, and GEMINI used integrated induction and maintenance RCTs, with eligible patients from the induction RCTs re-randomised into the maintenance RCT. The randomised portion of PURSUIT-M enrolled patients who had responded to golimumab induction therapy; whereas OCTAVE Sustain enrolled patients who responded to induction therapy with either tofacitinib or placebo. In GEMINI, to fulfil sample-size requirements for the maintenance phase, the induction phase included an additional cohort of patients who received open-label vedolizumab. Patients with clinical response to either blinded or open-label vedolizumab therapy were then re-randomised in the maintenance phase of the trial. Among patients enrolled in the maintenance phase, 252 received open-label vedolizumab induction therapy and 121 had received blinded induction therapy with vedolizumab.
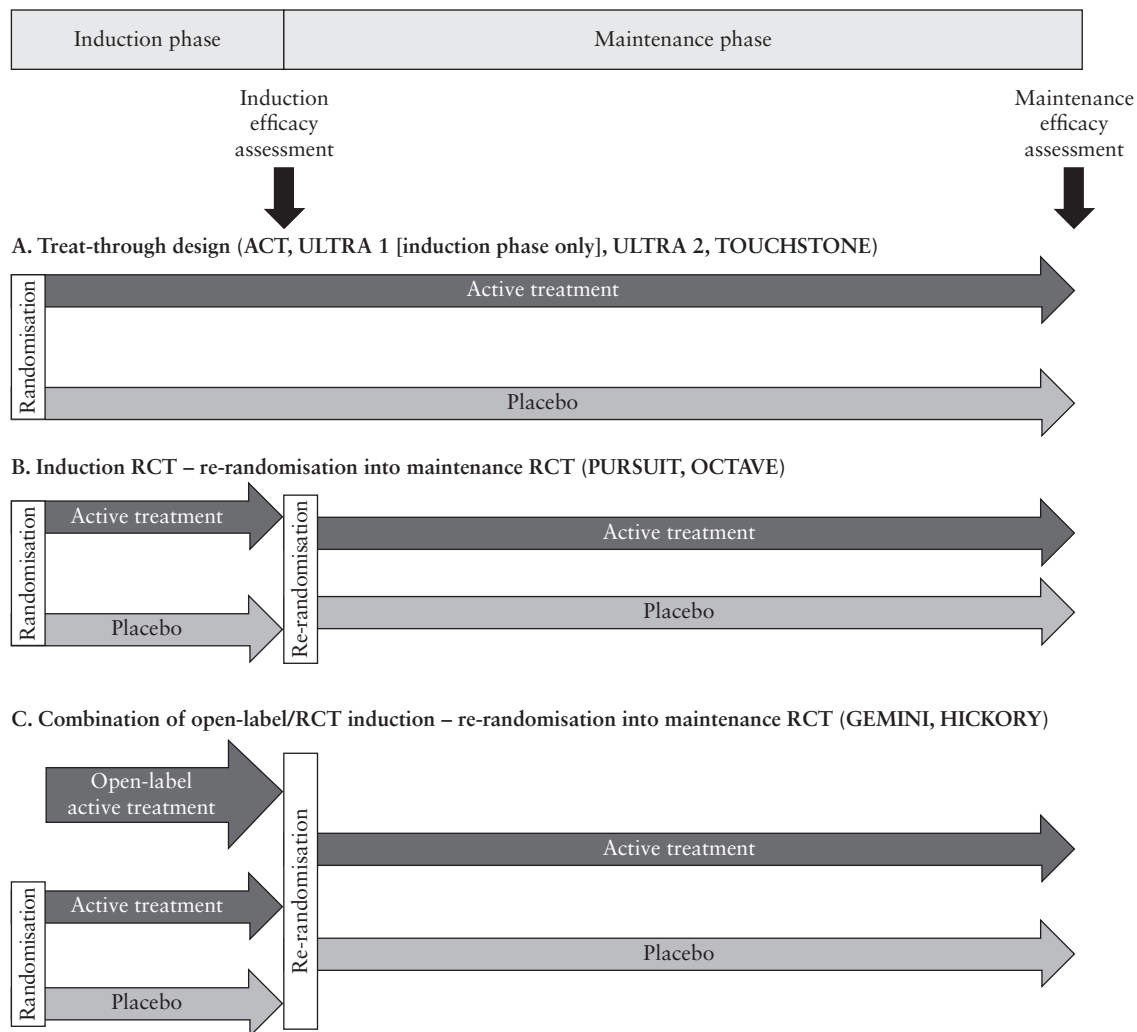

Figure 1. Different approaches to integration of induction and maintenance phases. [A] Treat-through design. [B] Integrated induction and maintenance RCTs. [C] Combination of open-label and RCT induction followed by maintenance RCT. RCT, randomised controlled trial.

The 66-week HICKORY trial also features integrated induction and maintenance phases, and open-label and blinded cohorts; however, full details of the study design have not yet been published.

The two fundamental approaches to UC trial design described above [i.e., the treat-through approach and the integration of induction and maintenance RCTs with separate randomisation for each phase] each have their own merits.

One benefit of the treat-through approach is in the ease of interpretation, given that the patient populations randomised to each treatment arm remain relatively constant throughout the duration of the study [with minor fluctuations due to discontinuations]. Conversely, interpretation of results from integrated induction and maintenance studies is slightly complicated by the need to account for the re-randomisation criteria applied at entry to the maintenance phase. Another situation in which the treat-through approach may be beneficial is when a delayed response to therapy might be expected. The continued assessment of patients beyond the induction phase—and while still under randomised controlled conditions—may allow the identification of specific subgroups of patients in whom delayed response to treatment occurs. This may also more accurately represent clinical practice where clinicians and patients may persist with a therapy if sufficient initial benefit is observed to justify extended induction treatment. Given that in the integrated approach it is usually only induction responders who are re-randomised into the maintenance phase, delayed response can typically only be assessed under open-label [i.e., non-randomised] conditions.

A benefit of the integrated approach is that it allows a separate assessment of induction efficacy and maintenance efficacy. This may be of relevance where regulatory bodies require discrete evidence of an agent's efficacy as induction and maintenance therapy.[21] A further benefit of integrated induction and maintenance RCTs is the potential to assess a number of clinical scenarios while under randomised controlled conditions. Patients who complete the induction phase as responders to active treatment may be re-randomised to a higher or lower dose of active treatment in the maintenance phase, which can be used as a surrogate to evaluate dose intensification and dose de-escalation, respectively. Patients may also be re-randomised to placebo maintenance therapy, which can be used as a surrogate to evaluate treatment interruption [a consideration for patients who wish to become pregnant or those undergoing transition of care].

Taking into consideration the above, it is important when comparing efficacy among UC maintenance trials to understand the randomisation criteria applied to patients' progression between the induction and maintenance phases of each trial, and the use of open-label arms to supply patients to the maintenance trials.

2.3. Efficacy endpoints

Several primary efficacy outcomes were used across the trials, all based on the Mayo score [Tables 4 and 5].[22] Clinical response was the primary efficacy endpoint in ACT 1 and ACT 2, in the PURSUIT trials, and for the induction phase of GEMINI:

- Clinical response: ≥3 points and ≥30% reduction from baseline total Mayo score plus decrease of ≥1 in rectal bleeding subscore or absolute rectal bleeding subscore ≤1.

Clinical remission was the primary efficacy endpoint in the ULTRA trials, for the maintenance phase of GEMINI, and in TOUCHSTONE:

- Clinical remission: total Mayo score ≤2, no subscore >1.

The OCTAVE and HICKORY trials used a more stringent definition of remission as the primary endpoint—equivalent to the definition of clinical remission used in the other trials, with the additional requirement of a rectal bleeding subscore of zero:

- Remission: total Mayo score ≤2, no subscore >1, rectal bleeding subscore = 0.

In addition to variation in the specific endpoint used for the primary evaluation of efficacy, the timing of its assessment also varied. For induction trials, primary efficacy assessment was at Week 6 for PURSUIT-SC and GEMINI; Week 8 for ACT, ULTRA, OCTAVE, and TOUCHSTONE; and Week 14 for HICKORY. For trials that assessed maintenance efficacy as a primary or a co-primary endpoint alongside an induction co-primary endpoint, assessments were made after at least 52 weeks of therapy. Not all trials included maintenance efficacy as a pre-specified primary endpoint [i.e., ACT and TOUCHSTONE].

Except for TOUCHSTONE, based on available publications/protocols, each of the trials that assessed maintenance efficacy included a measure of corticosteroid-free remission. ACT 1 and ACT 2 reported clinical remission and discontinued use of corticosteroids at Week 30 [ACT 1 and ACT 2] and Week 52 [ACT 2 only]; ULTRA 2 reported remission at Week 52 with discontinuation of corticosteroids for at least 90 days before Week 52; PURSUIT-M reported corticosteroid-free clinical remission at Week 52 among patients receiving corticosteroids at baseline. Due to the mandatory steroid tapering requirement, it is important to note that the Week 52 remission endpoint in OCTAVE Sustain in essence represents steroid-free remission, with the exception of one protocol deviation in each of the active treatment arms (one patient taking prednisone 2.5 mg/day and another taking prednisone 7.5 mg/day, in the tofacitinib 5 mg and 10 mg twice daily [BID] treatment arms, respectively). Additionally, OCTAVE Sustain reported sustained corticosteroid-free remission among patients in remission at baseline, specifying that the patient had to be in remission at both Week 24 and Week 52, and that corticosteroids must have been discontinued ≥4 weeks before those time points. GEMINI reported corticosteroid-free remission at Week 52.

Recently, mucosal healing has been identified as an important measure of disease activity in inflammatory bowel diseases.[23,24] Although there is discussion about what constitutes mucosal healing and how it should be measured,[25] it is most commonly assessed in UC clinical trials using the Mayo endoscopic subscore.[26] Each of the completed trials reviewed here included the same definition of mucosal healing: Mayo endoscopic subscore of 0 or 1, corresponding to normal or inactive disease [subscore = 0] or mild disease with evidence of erythema, decreased vascular pattern [subscore = 1], or mild friability [subscore = 1]; OCTAVE required any observed friability to be scored as subscore 2 or more; among the other published trials, assessment criteria permitted mild friability to be scored within the mild disease category [i.e., subscore = 1]. Whereas the latest modification to the Mayo endoscopic subscore in the OCTAVE trial makes the endpoint more stringent, such evolutions further complicate contextualisation of trial results, as even if the overall definitions of Mayo score-based endpoints appear identical between trials, differences in the assessment criteria of the endoscopic subscore may further hamper comparisons between trials.

A further consideration in interpreting the Mayo score is the approach taken to calculation of the stool frequency and rectal bleeding subscores. Typically these are calculated based on data from the most recent 3 consecutive days, with either the average result or the

Table 4. Primary efficacy endpoints in clinical trials of advanced therapies for ulcerative colitis.

Clinical trial	Induction primary efficacy endpoint	Maintenance primary efficacy endpoint
ACT 1 and ACT 2 [induction and maintenance]	Clinical response [Week 8]	No primary endpoint for maintenance phase
ULTRA 1 [induction],	Clinical remission [Week 8],	N/A,
ULTRA 2 [induction and maintenance]	Clinical remission [Week 8; co-primary with Week 52 endpoint]	Clinical remission [Week 52; co-primary with Week 8 endpoint]
PURSUIT-SC [induction],	Clinical response [Week 6],	N/A,
PURSUIT-M [maintenance]	N/A	Clinical response [Week 54]
GEMINI 1 [induction],	Clinical response [Week 6],	N/A,
GEMINI 1 [maintenance]	N/A	Clinical remission [Week 52]
OCTAVE Induction 1 and	Remission[a]	N/A,
Induction 2 [induction],	[Week 8],	
OCTAVE Sustain [maintenance]	N/A	Remission[a] [Week 52]
TOUCHSTONE [induction and maintenance]	Clinical remission [Week 8]	No primary endpoint for maintenance phase
HICKORY [induction and maintenance]	Remission[a] [Week 14; co-primary]	Remission[a] [Week 66; among randomised patients in clinical remission at Week 14; co-primary]

[a]The OCTAVE and HICKORY trials used a more stringent definition of remission as the primary endpoint—equivalent to the definition of clinical remission used in the other trials with the additional requirement of a rectal bleeding subscore = 0.

N/A, not available.

Table 5. Mayo score criteria for efficacy endpoints used in ulcerative colitis clinical trials.

	Clinical response	Clinical remission	Remission
Total Mayo score	≥3-point and ≥30% reduction from baseline	≤2	≤2
PGA subscore		≤1	≤1
Rectal bleeding subscore	≥1-point reduction from baseline or absolute subscore ≤1	≤1	0
Stool frequency subscore		≤1	≤1
Endoscopic subscore		≤1	≤1

The total Mayo score comprises four subscores [PGA; rectal bleeding; stool frequency; endoscopic], each scored from 0 to 3, with higher scores indicating more severe disease.

PGA, Physician's Global Assessment.

worst result used to generate the subscore.[25] For the trials that have reported the approach taken [PURSUIT, GEMINI, OCTAVE, and TOUCHSTONE], all reported the average result. In cases where the worst score is used to calculate stool frequency and rectal bleeding subscores, there may be a bias toward an overall lower estimate of absolute efficacy, but conversely, treatment effect size may be exaggerated since it is possible that placebo response rates may be lower. However, in the absence of studies comparing the two approaches, it is not possible to determine what the predominant effect would be.

2.4. Central vs local reading of endoscopy

The importance of central endoscopy readers has recently been noted as a means of reducing subjectivity and potential bias associated with local or site-read endoscopy.[27,28] Local reading of endoscopies may introduce differential bias, with a tendency to record higher scores during the screening period to qualify the patient for enrolment, and a tendency to record lower scores at subsequent assessment of outcomes to permit continued qualification for study treatment. Furthermore, site or local reading of endoscopy may

contribute to variation in placebo responses observed in UC clinical trials.[28,29] Previous studies have demonstrated the benefit of performing centralised reading of endoscopy, which reduced variability in the assessment of placebo efficacy.[28] Additionally, draft guidance from the U.S. Food and Drug Administration [FDA] on the use of imaging endpoints in clinical trials suggests that centralised reading may decrease variability in image interpretation.[30] Whereas additional training and/or further validation of local vs central reading of endoscopy may allow for trials to be performed using local reading of endoscopy,[27] it is also conceivable that differences between central and local reading observed in controlled conditions would be larger if local readers were not aware that they were being observed [i.e., the Hawthorne effect].

Among the completed trials reviewed here, only more recently conducted studies [OCTAVE and TOUCHSTONE] used central reading of endoscopy, both to determine patients' eligibility to participate in the trials and to evaluate efficacy outcomes; earlier trials used local reading of endoscopy for eligibility and efficacy assessments.

3. Impact of Patient Population and Central vs Local Reading of Endoscopy on Efficacy Outcomes

To illustrate the possible effects of changes in patient population and disparity in the assessment of efficacy endpoints, we explored data from the OCTAVE Sustain trial [for which efficacy data based on both central and local reading of endoscopy have been published].[18,31] We contrasted efficacy measured using central-read endoscopy in the overall population of the trial [of whom approximately 50% had previous exposure to anti-TNF] against efficacy measured using local-read endoscopy in the anti-TNF-naïve subgroup [Table 6]. For remission at Week 52, there was a 10.8 percentage point difference in treatment effect size for the 5 mg BID dose of tofacitinib when comparing the rate of remission in the overall population based on central reading vs the response rate in anti-TNF-naïve patients using local reading of endoscopy [i.e., a relative increase of 47%]. For the 10 mg BID dose, the difference was 7.3 percentage points [25% relative increase]. A similar pattern was observed for mucosal healing,

with a difference of 8.8 percentage points [36% relative increase] for the 5 mg BID dose and 7.6 percentage points [23%] for the 10 mg BID dose. As shown, patient inclusion characteristics and changes in trial design may have a significant impact on the reported efficacy based on point estimates of the data.

3.1. Practical implications for clinicians and patients, and goals for future trials of UC therapies

In this review of clinical trials assessing advanced therapies for the treatment of moderate to severe UC, we noted several important aspects of trial design that should be accounted for when evaluating trials, including: differences in disease characteristics of the trial populations; exposure to previous UC therapies and permitted concomitant therapies; progression of patients between induction and maintenance phases of therapy; and the endpoints used to determine efficacy. Although patients and clinicians often focus on absolute rates of response, our overview suggests that failing to account for these differences may be misleading. Accordingly when evaluating efficacy outcomes across trials, placebo-adjusted response rates, numbers needed to treat, and/or risk ratios may allow for more meaningful efficacy comparisons than comparing absolute efficacy responses. Though network meta-analyses comparing therapies for moderate to severe UC have been conducted, limitations of such comparisons have highlighted the need for randomised comparative efficacy trials.[14–16,32]

The ultimate goal for trials assessing novel therapies should be the design of efficient trials that generate scientifically rigorous data and answer clinically important questions. Comparative effectiveness and head-to-head trials of advanced therapies for the treatment of moderately to severely active UC would be welcomed, but require lengthy recruitment periods to enrol sufficient sample sizes to be adequately powered. Trial designs that facilitate and encourage recruitment of patients would be beneficial in ensuring that sample-size requirements are easily and quickly met, enabling safe and efficacious treatments to reach clinical practice more quickly, and investigation into ineffective therapies to be more rapidly concluded. Several trials reviewed here contained design elements that achieve this, including: preferential randomisation to active treatment over placebo [OCTAVE]; allocation to open-label active treatment [GEMINI and HICKORY]; and allowance for patients receiving placebo to transfer to open-label active therapy early in the case of loss of response or relapse [TOUCHSTONE].

Real-world data studies of therapies for UC are also needed to supplement data from clinical trials. This may allow evaluation of therapies in larger patient populations that may more accurately represent the UC patient population, given that many patients with moderate to severe inflammatory bowel disease would not qualify for participation in the RCTs reviewed here.[17] In the absence of head-to-head RCTs, real-world data may also assist with comparing advanced therapies for the treatment of UC.

3.2. Regulatory body recommendations for UC clinical trial endpoints

An additional consideration in the design of clinical trials and choice of endpoints is the requirements of regulatory bodies. Draft

Table 6. Impact of patient population and use of central-read vs local-read endoscopy on efficacy outcomes in OCTAVE Sustain.

	OCTAVE Sustain					
	Central-read endoscopy, anti-TNF-naïve and anti-TNF-experienced patients			Local-read endoscopy, anti-TNF-naïve population[a]		
	Placebo	Tofacitinib 5 mg BID	Difference [95% CI]	Placebo	Tofacitinib 5 mg BID	Difference [95% CI]
	[N = 198]	[N = 198]		[N = 106]	[N = 108]	
Remission at Week 52, n [%]	22 [11.1]	68 [34.3]	23.2*** [15.3–31.2]	14 [13.2]	51 [47.2]	34.0*** [22.6–45.4]
Clinical remission at Week 52, n [%]	22 [11.1]	68 [34.3]	23.2*** [15.3–31.2]	14 [13.2]	52 [48.1]	34.9*** [23.5–46.4]
Mucosal healing at Week 52, n [%]	26 [13.1]	74 [37.4]	24.2*** [16.0–32.5]	17 [16.0]	53 [49.1]	33.0*** [21.3–44.8]
Clinical response at Week 52, n [%]	40 [20.2]	102 [51.5]	31.3*** [22.4–40.2]	26 [24.5]	60 [55.6]	31.0*** [18.6–43.5]
	Placebo	Tofacitinib 10 mg BID	Difference [95% CI]	Placebo	Tofacitinib 10 mg BID	Difference [95% CI]
	[N = 198]	[N = 197]		[N = 106]	[N = 96]	
Remission at Week 52, n [%]	22 [11.1]	80 [40.6]	29.5*** [21.4–37.6]	14 [13.2]	48 [50.0]	36.8*** [24.9–48.7]
Clinical remission at Week 52, n [%]	22 [11.1]	81 [41.1]	30.0*** [21.9–38.2]	14 [13.2]	49 [51.0]	37.8*** [25.9–49.7]
Mucosal healing at Week 52, n [%]	26 [13.1]	90 [45.7]	32.6*** [24.2–41.0]	17 [16.0]	54 [56.3]	40.2*** [28.1–52.3]
Clinical response at Week 52, n [%]	40 [20.2]	122 [61.9]	41.7*** [32.9–50.5]	26 [24.5]	63 [65.6]	41.1*** [28.6–53.6]

***$p < 0.0001$ vs placebo. Data are full analysis set with non-responder imputation.

[a]Based on data from baseline of induction studies.

BID, twice daily; CI, confidence interval; N, number of evaluable patients; n, number of patients with efficacy response; TNF, tumour necrosis factor.

guidance from the FDA[25] and the European Medicines Agency [EMA][21] acknowledges that the total Mayo score has been a commonly used tool for registration trials, but cites the Physician's Global Assessment subscore as a limitation of the tool. Accordingly, the FDA's draft recommendation for the primary efficacy outcome of UC clinical trials is a definition of clinical remission based on the Mayo stool frequency, rectal bleeding, and endoscopic subscores only: stool frequency subscore = 0, rectal bleeding subscore = 0, and endoscopic subscore ≤1. An ongoing trial of the JAK inhibitor ABT-494 [upadacitinib] in patients with moderate to severe UC[33] has a primary outcome measure based on the adapted Mayo score [i.e., Mayo score excluding the Physician's Global Assessment subscore], with clinical remission defined as stool frequency subscore ≤1, rectal bleeding subscore = 0, and endoscopic subscore ≤1. The FDA's draft guidance does not support a definition of mucosal healing based on endoscopic findings without a validated histological assessment, and the EMA guidelines define mucosal inflammation assessed by endoscopy as 'endoscopic healing' rather than 'mucosal healing'.[21] Indeed, histological remission may be a goal for future trial designs, given that patients who demonstrate histological healing have better outcomes than those with ongoing histological activity.[34] However, there is a need for standardised assessments of histological disease activity as well as endpoint definitions. The draft FDA and EMA guidance recommends that disease signs and symptoms would be best measured by a patient-reported outcome rather than a clinician-reported outcome, and also advises standardisation of such measures across patients.[21,25] Finally, the FDA guidance recommends that the definition of corticosteroid-free remission used in UC trials should define a minimum period of time over which the patient is both steroid-free and in remission. For future trials that may incorporate this or further iterations of these guidelines, it will be important to acknowledge the potential impact on efficacy outcomes. In addition, trials should be able to report data based on previous definitions for context, e.g., as secondary or additional endpoints.

As UC trials have evolved, patient populations have become more refractory, and the expectations of treatment and stringency of efficacy measures have increased. Despite the well-acknowledged difficulty in comparing agents across trials for the same indication, it is inevitable that clinicians will need to make their best judgments of the data across studies. In the absence of head-to-head RCTs of multiple agents, it is critical that clinicians understand the ways in which studies are comparable or disparate in design and population, and how any differences may affect interpretation of the results.

Funding

Conflict of Interest

BES has received grant support, personal fees, and non-financial support from Pfizer Inc during the conduct of this research; consulting fees and research grants from AbbVie, Amgen, Bristol-Myers Squibb, Celgene, Janssen, and Takeda; and consulting fees from Akros Pharma, Allergan, Arena Pharmaceuticals, Boehringer Ingelheim, EnGene, Forward Pharma, Gilead, Immune Pharmaceuticals, Lilly, Lycer, Lyndra, MedImmune, Oppilan Pharma, Receptos, Synergy Pharmaceuticals, Target PharmaSolutions, Theravance, TiGenix, TopiVert Pharma, UCB Pharma, Vedanta Biosciences, and Vivelix Pharmaceuticals outside the submitted work. ASC has been on consulting/advisory boards for AbbVie, AMAG, Ferring, Janssen, Pfizer Inc, and Takeda; and has received consulting fees and research support from Miraca Laboratories.

CIN, DQ, WW, EM, GSF, and CS are employees and stockholders of Pfizer Inc. PDRH has received consulting fees from AbbVie, Amgen, Genentech, JBR Pharma, and Lycera.

Acknowledgments

Author Contributions

Review concept and design: BES, ASC, CIN, DQ, WW, EM, GSF, CS, and PDRH. Literature search and drafting of the article: BES, ASC, CIN, DQ, WW, EM, GSF, CS, and PDRH. Critical review of subsequent drafts: BES, ASC, CIN, DQ, WW, EM, GSF, CS, and PDRH. All authors read and approved the final draft for submission.

References

1. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *Lancet* 2017;**389**:1756–70.

2. Harbord M, Eliakim R, Bettenworth D, *et al*. Third European evidence-based consensus on diagnosis and management of ulcerative colitis. Part 2: current management. *J Crohns Colitis* 2017;**11**:769–84.

3. Kornbluth A, Sachar DB; Practice Parameters Committee of the American College of Gastroenterology. Ulcerative colitis practice guidelines in adults: American College of Gastroenterology, Practice Parameters Committee. *Am J Gastroenterol* 2010;**105**:501–23; quiz 524.

4. Dignass A, Lindsay JO, Sturm A, *et al*. Second European evidence-based consensus on the diagnosis and management of ulcerative colitis. Part 2: current management. *J Crohns Colitis* 2012;**6**:991–1030.

5. Rutgeerts P, Sandborn WJ, Feagan BG, *et al*. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005;**353**:2462–76.

6. Reinisch W, Sandborn WJ, Hommes DW, *et al*. Adalimumab for induction of clinical remission in moderately to severely active ulcerative colitis: results of a randomised controlled trial. *Gut* 2011;**60**:780–7.

7. Sandborn WJ, van Assche G, Reinisch W, *et al*. Adalimumab induces and maintains clinical remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2012;**142**:257–65.e1–3.

8. Sandborn WJ, Feagan BG, Marano C, *et al*.; PURSUIT-SC Study Group. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014;**146**:85–95; quiz e14–5.

9. Sandborn WJ, Feagan BG, Marano C, *et al*.; PURSUIT-Maintenance Study Group. Subcutaneous golimumab maintains clinical response in patients with moderate-to-severe ulcerative colitis. *Gastroenterology* 2014;**146**:96–109.e1.

10. Feagan BG, Rutgeerts P, Sands BE, *et al*.; GEMINI 1 Study Group. Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2013;**369**:699–710.

11. US Food and Drug Administration. *ENTYVIO® [Vedolizumab] Highlights of Prescribing Information*. 2014. https://www.accessdata.fda.gov/drugsatfda_docs/label/2014/125476s000lbl.pdf Accessed December 4, 2018.

12. D'Haens G, Feagan B, Colombel JF, *et al*.; International Organization for Inflammatory Bowel Diseases [IOIBD] and the Clinical Trial Committee Clincom of the European Crohn's and Colitis Organisation [ECCO]. Challenges to the design, execution, and analysis of randomised controlled trials for inflammatory bowel disease. *Gastroenterology* 2012;**143**:1461–9.

13. Hindryckx P, Baert F, Hart A, Magro F, Armuzzi A, Peyrin-Biroulet L; Clinical Trial Committee Clincom of the European Crohn's and Colitis

Organisation [ECCO]. Clinical trials in ulcerative colitis: a historical perspective. *J Crohns Colitis* 2015;**9**:580–8.

14. Singh S, Fumery M, Sandborn WJ, Murad MH. Systematic review with network meta-analysis: first- and second-line pharmacotherapy for moderate-severe ulcerative colitis. *Aliment Pharmacol Ther* 2018;**47**:162–75.

15. Bonovas S, Lytras T, Nikolopoulos G, Peyrin-Biroulet L, Danese S. Systematic review with network meta-analysis: comparative assessment of tofacitinib and biological therapies for moderate-to-severe ulcerative colitis. *Aliment Pharmacol Ther* 2018;**47**:454–65.

16. Cameron C, Ewara E, Wilson FR, *et al*. The importance of considering differences in study design in network meta-analysis: an application using anti-tumor necrosis factor drugs for ulcerative colitis. *Med Decis Making* 2017;**37**:894–904.

17. Ha C, Ullman TA, Siegel CA, Kornbluth A. Patients enrolled in randomised controlled trials do not represent the inflammatory bowel disease patient population. *Clin Gastroenterol Hepatol* 2012;**10**:1002–7; quiz e78.

18. Sandborn WJ, Su C, Sands BE, *et al*.; OCTAVE Induction 1, OCTAVE Induction 2, and OCTAVE Sustain Investigators. Tofacitinib as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2017;**376**:1723–36.

19. Sandborn WJ, Feagan BG, Wolf DC, *et al*.; TOUCHSTONE Study Group. Ozanimod induction and maintenance treatment for ulcerative colitis. *N Engl J Med* 2016;**374**:1754–62.

20. Peyrin-Biroulet L, Feagan BG, Mansfield J, *et al*. Etrolizumab treatment leads to early improvement in symptoms and inflammatory biomarkers in anti-TNF-refractory patients in the open-label induction cohort of the phase 3 HICKORY study. *J Crohns Colitis* 2017;**11**:S6–7. [Abstract OP011].

21. European Medicines Agency—Committee for Medicinal Products for Human Use. *Guideline on the Development of New Medicinal Products for the Treatment of Ulcerative Colitis*. 2018. https://www.ema.europa.eu/documents/scientific-guideline/guideline-development-new-medicinal-products-treatment-ulcerative-colitis-revision-1_en.pdf Accessed November 23, 2018.

22. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomised study. *N Engl J Med* 1987;**317**:1625–9.

23. Neurath MF, Travis SP. Mucosal healing in inflammatory bowel diseases: a systematic review. *Gut* 2012;**61**:1619–35.

24. Neurath MF. New targets for mucosal healing and therapy in inflammatory bowel diseases. *Mucosal Immunol* 2014;**7**:6–19.

25. U.S. Department of Health and Human Services FaDA. *Ulcerative Colitis: Clinical Trial Endpoints. Guidance for Industry*. 2016. http://www.fda.gov/downloads/Drugs/Guidances/UCM515143.pdf Accessed April 25, 2018.

26. Vaughn BP, Shah S, Cheifetz AS. The role of mucosal healing in the treatment of patients with inflammatory bowel disease. *Curr Treat Options Gastroenterol* 2014;**12**:103–17.

27. Gottlieb K, Travis S, Feagan B, Hussain F, Sandborn WJ, Rutgeerts P. Central reading of endoscopy endpoints in inflammatory bowel disease trials. *Inflamm Bowel Dis* 2015;**21**:2475–82.

28. Feagan BG, Sandborn WJ, D'Haens G, *et al*. The role of centralized reading of endoscopy in a randomised controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;**145**:149–57.e2.

29. Travis SP, Schnell D, Krzeski P, *et al*. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity [UCEIS]. *Gut* 2012;**61**:535–42.

30. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research [CDER], Center for Biologics Evaluation and Research [CBER]. *Clinical Trial Imaging Endpoint Process Standards. Guidance for Industry*. 2015. http://www.fda.gov/downloads/Drugs/Guidance_Compliance_Regulatory_Information/Guidances/UCM268555.pdf Accessed September 8, 2016.

31. Feagan BG, Vermeire S, Sandborn WJ, *et al*. Tofacitinib for maintenance therapy in patients with active ulcerative colitis in the phase 3 OCTAVE Sustain trial: results by local and central endoscopic assessments. *Am J Gastroenterol* 2017;**112**:S329.

32. Stidham RW, Lee TC, Higgins PD, *et al*. Systematic review with network meta-analysis: the efficacy of anti-tumour necrosis factor-alpha agents for the treatment of ulcerative colitis. *Aliment Pharmacol Ther* 2014;**39**:660–71.

33. ClinicalTrials.gov. *A Study to Evaluate the Safety and Efficacy of Upadacitinib [ABT-494] for Induction And Maintenance Therapy in Subjects With Moderately to Severely Active Ulcerative Colitis [UC]*. 2017. https://clinicaltrials.gov/ct2/show/NCT02819635 Accessed February 1, 2018.

34. Christensen B, Hanauer SB, Erlich J, *et al*. Histologic normalization occurs in ulcerative colitis and is associated with improved clinical outcomes. *Clin Gastroenterol Hepatol* 2017;**15**:1557–64.e1.