



# Filgotinib for Treating Moderately to Severely Active Ulcerative Colitis: An Evidence Review Group Perspective of a NICE Single Technology Appraisal

Antoinette D. I. van Asselt<sup>1,2</sup> · Nigel Armstrong<sup>3</sup> · Merel Kimman<sup>4,5</sup> · Andrea Peeters<sup>4</sup> · Kevin McDermott<sup>3</sup> · Lisa Stirk<sup>3</sup> · Charlotte Ahmadu<sup>3</sup> · Tim M. Govers<sup>6</sup> · Frank Hoentjen<sup>7,8</sup> · Manuela A. Joore<sup>4,5</sup> · Sabine E. Grimm<sup>4</sup>

Accepted: 12 January 2023 / Published online: 1 February 2023

© The Author(s) 2023

## Abstract

The National Institute for Health and Care Excellence invited the manufacturer (Galapagos) of filgotinib (Jyseleca<sup>®</sup>), as part of the Single Technology Appraisal process, to submit evidence for the clinical effectiveness and cost effectiveness of filgotinib for treating moderately to severely active ulcerative colitis in adults who have had an inadequate response, loss of response or were intolerant to a previous biologic agent or conventional therapy. Kleijnen Systematic Reviews Ltd, in collaboration with Maastricht University Medical Centre+, was commissioned to act as the independent Evidence Review Group. This paper summarises the company submission, presents the Evidence Review Group's critical review on the clinical and cost-effectiveness evidence in the company submission, highlights the key methodological considerations and describes the development of the National Institute for Health and Care Excellence guidance by the Appraisal Committee. The company submission included one relevant study for the comparison of filgotinib versus placebo: the SELECTION trial. As there was no head-to-head evidence with any of the comparators, the company performed two separate network meta-analyses, one for the biologic-naïve population and one for the biologic-experienced population, and for both the induction and maintenance phases. The Evidence Review Group questioned the validity of the maintenance network meta-analysis because it assumed all active treatments to be comparators in this phase, which is not in line with clinical practice. The economic analysis used a number of assumptions that introduced substantial uncertainty, which could not be fully explored, for instance, the assumption that a risk of loss of response would be independent of health state and constant over time. Company and Evidence Review Group results indicate that at its current price, and disregarding confidential discounts for comparators and subsequent treatments, filgotinib dominates some comparators (golimumab and adalimumab in the company base case, all but intravenous and subcutaneous vedolizumab in the Evidence Review Group's base case) in the biologic-naïve population. In the biologic-experienced population, filgotinib dominates all comparators in both the company and the Evidence Review Group's base case. Results should be interpreted with caution as some important uncertainties were not included in the modelling. These uncertainties were mostly centred around the maintenance network meta-analysis, loss of response, health-related quality-of-life estimates and modelling of dose escalation. The National Institute for Health and Care Excellence recommended filgotinib within its marketing authorisation, as an option for treating moderately to severely active ulcerative colitis in adults when conventional or biological treatment cannot be tolerated, or if the disease has not responded well enough or has stopped responding to these treatments, and if the company provides filgotinib according to the commercial arrangement.

✉ Antoinette D. I. van Asselt  
a.d.i.van.asselt@umcg.nl

<sup>1</sup> Department of Health Sciences, University of Groningen, University Medical Center Groningen (UMCG), PO Box 30.001, Hanzeplein 1, 9700 RB Groningen, The Netherlands

<sup>2</sup> Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>3</sup> Kleijnen Systematic Reviews Ltd, York, UK

<sup>4</sup> Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+, Maastricht, The Netherlands

<sup>5</sup> Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

<sup>6</sup> Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands

<sup>7</sup> Division of Gastroenterology, University of Alberta, Edmonton, AB, Canada

<sup>8</sup> Department of Gastroenterology, Radboud University Medical Center, Nijmegen, The Netherlands

### Key Points for Decision Makers

At its current price with a commercial arrangement in place, and disregarding confidential discounts for comparators, filgotinib was the cheapest of the technologies compared and dominated some comparators in the biologic-naïve population. Filgotinib dominated all comparators in the biologic-experienced population. Results were fairly robust but important uncertainties were not included in the modelling.

The maintenance phase network meta-analysis implied that all active treatments are comparators in this phase while actually the only valid comparator, according to clinical practice, is no treatment or the curtailment of the intervention on which induction was achieved. This may be a relevant issue for all appraisals where clinical evidence consists of more than one treatment phase.

Ulcerative colitis appraisals are generally hampered by the absence of good quality real-world evidence on the long-term risk of loss of response, health-related quality of life, treatment sequences, dose escalations and resource use.

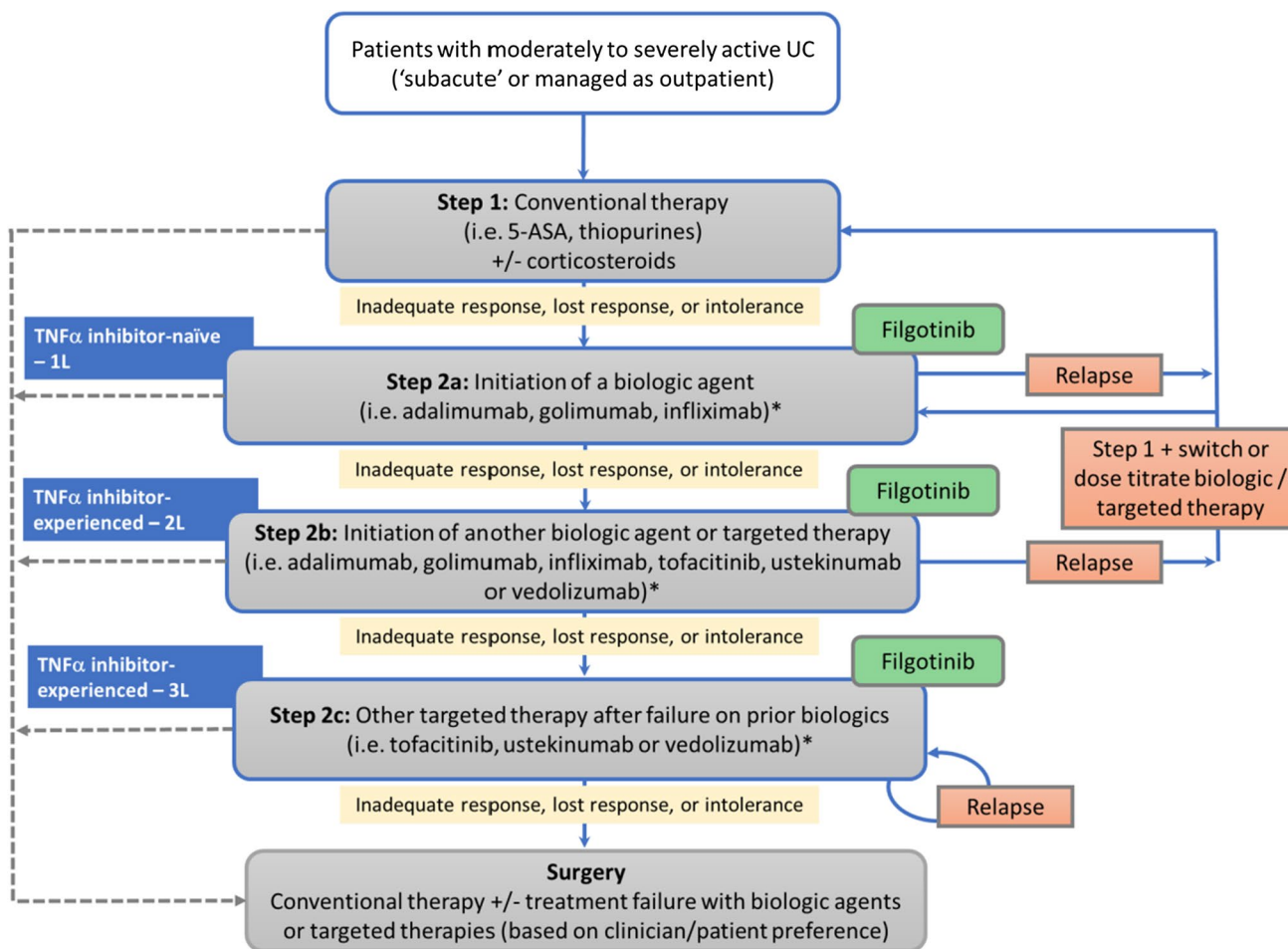
## 1 Introduction

Filgotinib, tradename Jyseleca<sup>®</sup>, was appraised within the National Institute for Health and Care Excellence (NICE) Single Technology Appraisal (STA) process. Health technologies must be shown to be clinically effective and to represent a cost-effective use of National Health Service (NHS) resources in order to be recommended by NICE. Within the STA process, the company (Galapagos) provided NICE with a written submission and a health economic model, summarising the company's estimates of the clinical effectiveness and cost effectiveness of filgotinib for the treatment of moderately to severely active ulcerative colitis (UC) in adults who have had an inadequate response, loss of response or were intolerant to a previous biologic agent or conventional therapy. This company submission (CS) was reviewed by an Evidence Review Group (ERG) independent of NICE. The ERG, Kleijnen Systematic Reviews in collaboration with Maastricht University Medical Centre+, produced an ERG report [1]. After consideration of the evidence submitted by the company and the ERG report, the NICE Appraisal Committee (AC) issued guidance on whether to recommend the technology by means of the Final Appraisal Determination, which is open for appeal [2]. This paper presents a

summary of the ERG report and the development of the NICE guidance. Furthermore, it highlights important methodological issues that may help in future decision making. Full details of all relevant appraisal documents (including the appraisal scope, CS, ERG report, consultee submissions, Appraisal Consultation Document, Final Appraisal Determination and comments from consultees) can be found on the NICE website [1, 2].

## 2 The Decision Problem

The NICE final scope defined the following population: people with moderately to severely active UC who have had an inadequate response, loss of response or were intolerant to a previous biologic agent or conventional therapy. [3] In the CS, the population was the same, with the addition that previous conventional therapies or biologic agents were described as: “conventional therapy (oral corticosteroids and/or immunomodulators), or a biologic agent (tumour necrosis factor [TNF]-alpha inhibitor or vedolizumab)” [1]. The company clarified that filgotinib could be included at any line in the biologic-experienced population (see Fig. 1). However, the biologic experienced network meta-analysis (NMA) and cost-effectiveness analysis were line agnostic. Additionally, dose escalation was applied to some comparators even though dose escalation was not recommended in the NICE guideline NG130 for UC. The company did, however, suggest that ‘second-line advanced’ therapy, which the ERG interpret as first line in the biologic experienced, was the most relevant population. However, this is inconsistent with the line implied by dose escalation, which the company indicated would occur as a last resort, i.e. pre-surgery [1]. The intervention (filgotinib, administered orally, 100 mg or 200 mg) was in line with the NICE scope. However, the modelled intervention was solely filgotinib 200 mg, administered orally once daily. Filgotinib 100 mg was not considered in the model, and this dosing is meant for patients with renal impairment only, which prompted the ERG to suggest limiting the decision problem to 200 mg only. Outcomes were in line with the NICE scope, although the scope listed mortality as an outcome, which was not addressed in the CS because no data on comparative mortality were available from the pivotal trial. The comparators in the CS were largely in line with the NICE scope, except for “conventional therapies, without biological treatments”, which were excluded from the NMA that was part of the CS. Instead, the placebo comparator in the NMA was used to inform the conventional treatment comparator in the economic model. However, the ERG questioned the relevance of conventional therapy as a comparator: in the NICE treatment pathway, it is proposed that filgotinib should be used for patients



\*Patients in response/remission remain on therapy with 12-month review

**Fig. 1** Proposed positioning of filgotinib within the National Institute for Health and Care Excellence treatment pathway. Courtesy of the company’s response to clarification questions by the Evidence

Review Group. *1L* first line, *2L* second line, *3L* third line, *5-ASA* 5-aminosalicylic acid, *UC* ulcerative colitis

who had an inadequate response, have lost response or are intolerant to conventional therapy. This is also the modelled population.

### 3 Independent ERG Review

The ERG reviewed the clinical effectiveness and cost-effectiveness evidence of filgotinib for this indication. As part of the STA process, the ERG and NICE had the opportunity to ask for clarification on specific issues in the CS, in response to which the company provided additional information [1]. Based on this information, the ERG produced an ERG base case by modifying the health economic model submitted by the company, and assessed the impact of alternative assumptions and parameter values on the model results.

Sections 3.1–3.6 summarise the evidence presented in the CS, as well as the review of the ERG.

#### 3.1 Clinical Effectiveness Evidence Submitted by the Company

The company’s clinical evidence came from the SELECTION trial in patients with moderately to severely active UC [4]. SELECTION is a phase IIb/III, randomised, double-blind, placebo-controlled trial comparing filgotinib 200 mg once daily, filgotinib 100 mg once daily and placebo during a 10-week induction study, followed by a maintenance study (weeks 10–58) in which the same interventions are compared to placebo after re-randomisation of those who responded to filgotinib during induction. The SELECTION trial was conducted under a single protocol but designed and analysed as three separate studies: two

induction studies and a maintenance study. The population of the induction period was stratified by biologic-naïve (cohort A) and biologic-experienced (cohort B) patients, resulting in the two induction studies.

The primary endpoint for the induction and maintenance studies was the proportion of patients achieving endoscopy/bleeding/stool frequency remission. Endoscopy/bleeding/stool frequency remission is defined as an endoscopic subscore of 0 or 1, a rectal bleeding subscore of 0 and at least a one-point decrease in stool frequency from baseline to achieve a subscore of 0 or 1. However, the primary outcome was not used in the economic model. The only outcomes used in the economic model (see Sect. 3.3) were Mayo Clinic Score (MCS) response (defined as: a MCS reduction of  $\geq 3$  points and at least 30% from the baseline score with an accompanying decrease in the rectal bleeding subscore of  $\geq 1$  point or an absolute rectal bleeding subscore of 0 or 1) and MCS remission (defined as: a MCS of 2 or less and no single subscore higher than 1).

In the induction phase, in cohort A, all efficacy outcomes in SELECTION showed statistically significant differences in favour of filgotinib 200 mg when compared with placebo [1]. In cohort B, a statistically significantly higher proportion of patients achieved endoscopy/bleeding/stool frequency remission at week 10 in the filgotinib 200-mg group compared with the placebo group, but MCS remission, endoscopic subscore of 0 and MCS remission (alternative definition) did not show statistically significant differences between groups. In the maintenance phase, all efficacy outcomes in SELECTION showed statistically significant differences in favour of filgotinib 200 mg when compared with placebo.

There were no trials identified comparing filgotinib versus comparators other than placebo, so the company undertook a systematic literature review and Bayesian NMA that aimed to provide a comparison of the efficacy of filgotinib with other comparators listed in the final NICE scope [1, 3]. The company separated their analysis into two populations:

- *biologic naïve* (cohort A population: patients without prior use of any biologic [TNF $\alpha$  inhibitor or vedolizumab], which aligns with the SELECTION cohort A); and
- *biologic experienced* (cohort B population: patients who have previously demonstrated an inadequate clinical response, loss of response to or intolerance to any biologic [TNF $\alpha$  inhibitor or vedolizumab], which aligns with the SELECTION cohort B).

The outcomes included in the NMA were: MCS remission, MCS response and mucosal healing (defined as an

endoscopic subscore of 0 or 1) [1]. These were assessed at two different timepoints, at the end of the Induction phase, and the end of the Maintenance phase, assumed to be, as in the SELECTION trial, 10 weeks and 48 weeks from re-randomisation, respectively. It is important to note that for the maintenance phase, outcomes at 58 weeks were conditional on response at 10 weeks. Results for the primary outcomes, which were those used in the cost-effectiveness analysis, depended on the phase and the population. The results of the NMAs are academic in confidence and therefore cannot be reported here [2].

### 3.2 Critique of Clinical Effectiveness Evidence and Interpretation

The CS and response to clarification provided sufficient details for the ERG to appraise the literature searches conducted as part of the systematic review to identify clinical effectiveness studies. A good range of databases and resources was searched.

Although re-randomisation of responders to the intervention permits an assessment of outcomes at the end of the maintenance phase conditional on having achieved a response, it does not inform the outcomes during the maintenance phase of those who did not achieve a response at the end of the induction period. There is no unbiased estimate (based on randomised trial data) of filgotinib versus placebo for the induction non-responders at the end of the maintenance phase because these patients were given the option to enter the long-term extension study where evidence as to how many patients were lost to follow-up or if they maintained the original treatment allocation was not reported [5]. Of course, if it is assumed that in clinical practice patients will switch treatment upon a lack of response at induction, then this might appear to be less of an issue. Although previous technology appraisals indicate that discontinuation should occur because of something resembling a lack of response (TA342 recommends that “Treatment should only continue if there is clear evidence of ongoing clinical benefit” and TA329 that patients “... should continue treatment only if there is clear evidence of response”), no time limit is expressed in terms of an induction period [6, 7]. Furthermore, in the cost-effectiveness analysis, because follow-up of non-responders is limited to the end of induction, the effectiveness of subsequent treatments is assumed to be the same regardless of the line of therapy in the biologic-experienced population. Therefore, re-randomisation also precludes an unbiased estimate of the long-term effectiveness of a sequence of biologic therapies. There were also issues identified in the risk of bias assessment of the SELECTION trial, particularly in terms of the balance of baseline characteristics, the effect of which is difficult to estimate.



The ERG considers that the NMAs were conducted using appropriate methods and that the induction phase NMA is appropriated to inform the question of the effectiveness of filgotinib 200 mg in comparison to the comparators in the decision problem in terms of response and remission. However, the ERG questions the validity of the maintenance phase NMA on a number of grounds. First, the population implied by the maintenance phase is not that of the decision problem, i.e. patients who have just finished the previous line of therapy. Second, it implies that all treatments are comparators in this phase when the only valid comparator, according to expected clinical practice, is no treatment or the curtailment of the intervention on which induction was achieved, i.e. treatment switching is not clinically relevant given successful induction. The NMA could also be considered to have questionable validity owing to the heterogeneity of the study populations included. This is because the population on entry to the maintenance phase in every trial is those patients who have responded only to the single intervention studied in that trial, which, of course, varies between the trials of different treatments. Therefore, the ERG replaced the maintenance phase NMA estimates of response that were inputs in the economic model with those that were trial based, i.e. estimates of induction response for each treatment from only the randomised controlled trials of that particular treatment (see Sect. 3.5).

### 3.3 Cost-Effectiveness Evidence Submitted by the Company

The company conducted searches for separate systematic literature reviews to identify cost-effectiveness outcomes, health-related quality-of-life (HRQoL) data and healthcare resource use for UC to address the decision problem and to inform the economic model structure. The company identified 12 UK-specific cost-effectiveness studies in UC, based on nine unique models, and used these to inform modelling choices.

The company constructed a Markov model in Microsoft Excel with nine health states and two transient states and a 10-weekly cycle length (Fig. 2). Distribution among health states in the first 10 weeks was based on the induction phase of the medication. After these first 10 weeks, the distribution among health states was based on the maintenance phase of the medication. Patients start in the model with ‘advanced therapy’ (Step 2 in Fig. 1) and have three options with regard to health states: Active UC, Response without remission and Remission. In the case of treatment failure (remaining in active UC), patients receive last-line conventional treatment and move to one of the following health states: active UC, response without remission or remission. Patients in active UC during last-line conventional treatment may undergo

surgery. Two types of surgery are included in the model: emergency surgery and elective surgery. These operations are modelled as transient states via which patients move to post-surgery states with or without complications. In these post-surgery states, all drugs are stopped. In the base-case analyses, no treatment sequences were modelled, i.e. there is no possibility to have different advanced therapies before moving to last-line conventional treatment. However, the model does give the opportunity to include up to four lines of advanced treatments, which was used for the scenario analyses. In response to a request by the ERG, the company implemented some treatment sequences, assuming no change in efficacy by line or order of therapy [1].

The model adopted the perspective of the NHS and Personal Social Services perspective. The model time horizon was lifetime. Costs and quality-adjusted life-years were discounted at a rate of 3.5% per year.

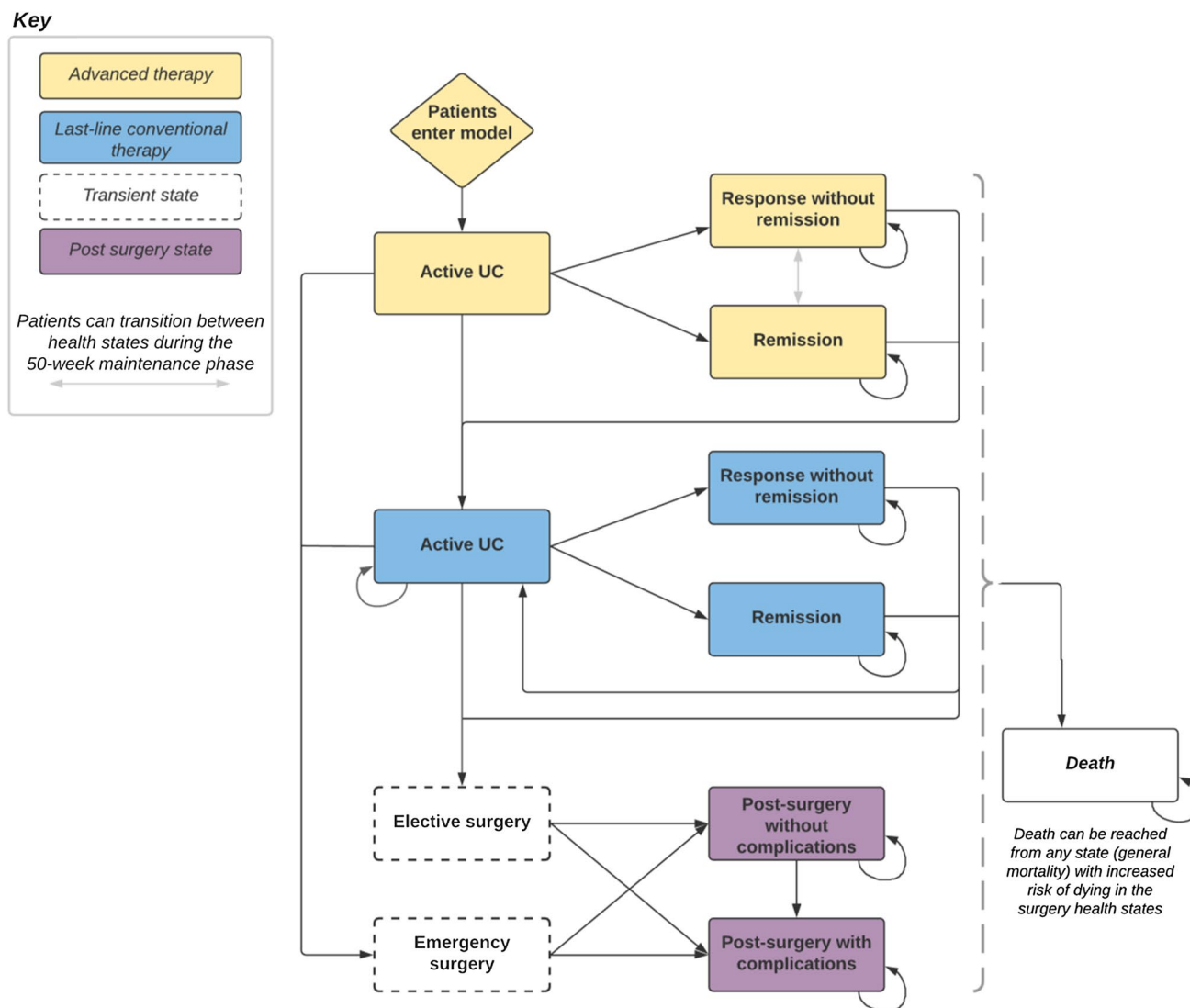
The patient population considered in the model was in line with the scope: adult patients with moderately to severely active UC who have had an inadequate response, loss of response or were intolerant to a previous biologic agent or conventional therapy, with biologic-naïve and biologic-experienced patients as subgroups. The baseline population characteristics applied in the model were based on the SELECTION trial induction study population, split into biologic-naïve and biologic-experienced patients.

The intervention considered in the model was filgotinib 200 mg, administered orally once daily. Filgotinib 100 mg was not considered in the model, as explained in Sect. 2. Comparators considered in the cost-effectiveness analysis were, in line with the scope of the appraisal, first-line biologics (TNF $\alpha$  inhibitors: infliximab, adalimumab, golimumab), advanced biologics (ustekinumab, vedolizumab) and a JAK inhibitor (tofacitinib). Conventional therapy was considered as a comparator and also modelled as a last-line therapy. It was however removed as a comparator later because of questionable relevance as filgotinib is intended for patients who had an inadequate response, have lost response or are intolerant to conventional therapy.

The main sources of evidence on treatment effectiveness used for the intervention and comparators were as follows:

- induction NMA estimates of MCS response and remission at induction;
- maintenance NMA estimates of MCS response to estimate loss of response by using the proportion with no response to estimate a constant rate of loss of response (see explanation below).

The distribution of patients at the end of the induction phase was informed by the NMA for the induction period alone. The NMA results included probability of overall response and remission. The proportion of patients achieving



**Fig. 2** Company's model structure for moderately to severely active ulcerative colitis (UC)

response (i.e. without remission) was estimated as the difference of patients receiving overall response (including remission), and patients achieving remission. The remainder of the population (1 minus overall response) would be in active UC at the end of the induction phase. Then, the output of the maintenance NMA was used to estimate long-term loss of response. As the maintenance NMA output was overall response (including remission), the probability of no response was calculated using the complement (i.e. 1 minus overall response). This probability of no response for the 50-week maintenance phase was recalculated into a 10-week probability. This 10-weekly loss of response was then applied to the total group of responders (both patients with response without remission and remission) to calculate transitions to the active UC health state. The relative proportions of patients in response without remission and

remission were assumed to remain constant, thus there would be no modelled transitions from remission to response without remission or vice versa. The risk of loss of response was extrapolated beyond the trial periods and assumed to be constant. A scenario with a 25% reduced risk of loss of response after the first year was provided later by the company. Additionally, the distribution over the response (without remission) and remission states, according to the situation in the maintenance phase, was assumed to remain constant in subsequent cycles. This would imply that if at the end of the maintenance phase, there are fewer patients in remission than in response, this cannot turn around anymore. After the 50-week maintenance period, patients are assumed to remain at the same level of response, and on the same treatment indefinitely, until they lose response and move to the active UC health state. However, because in

clinical practice patients in stable remission may discontinue treatment, a stopping rule was explored in a scenario analysis. A proportion of patients are expected to experience long-term complications after undergoing a colectomy. The rates of long-term complications post-surgery were obtained from Ferrante et al. [8], a study that reported a 46% rate of pouchitis in patients with UC undergoing a proctocolectomy over 6.5 years of follow-up, which resulted in an estimated 10-week probability of 1.81%. This probability in the model was thus related to the incidence of all pouchitis events, but the utility assigned to the post-surgery with complications health state was based on chronic pouchitis. Given that chronic pouchitis has a greater impact on HRQoL, the probability of chronic pouchitis (19%) is likely more appropriate. This change will favour treatments with higher numbers of patients in the active UC state.

The utility values for the health states in the economic model were based on EQ-5D-5L data collected alongside the SELECTION trial and a study by Arseneau et al. [9]. EQ-5D-5L health states reported at baseline, 10 and 58 weeks were mapped to EQ-5D-3L values using the crosswalk described by van Hout et al. [10]. The utility data were then analysed to predict the mean utility for each pre-surgical health state of the model, i.e. remission, response without remission and active UC. Baseline data were used for the active UC health state. For the remission and response without remission health states, the utility values calculated at the end of the induction phase (week 10) were used, as these estimates were based on a higher number of patients than the values at 58 weeks. Utilities were not treatment or population (biologic naive vs biologic experienced) specific. Utility values for the surgery with complications and post-surgery states were based on the study by Arseneau et al. [9], as the SELECTION trial did not collect appropriate data for these health states. The only adverse events relating to pharmaceutical treatments considered for the analysis were serious infections, defined as all serious adverse events in the Infections and Infestations system organ class. Experiencing an adverse reaction results in a fixed disutility. The disutility for pneumonia was applied for the estimated proportion of patients with a serious infection. Finally, an adjustment of the health state utility values by age and sex was applied to all patients in the model.

In line with NICE requirements, the model only considered direct medical costs. Cost and healthcare resource use inputs comprised drug acquisition, administration costs, costs associated with management of adverse events and background disease management costs. Costs were obtained from the published literature, 2018/19 NHS reference costs (published in 2020) [11] and the Monthly Index of Medical Specialties 2021 [12]. Costs were applied per cycle and

are estimated separately for induction treatment and maintenance treatment.

Drug acquisition costs of the intervention and advanced treatments are based on UK costs and dosing regimens from the Monthly Index of Medical Specialties 2021 [12]. Treatment costs per 10-weekly cycle are based on the recommended posology for each treatment. Where more than one posology was available, dose escalation of 30% was considered and a weighted average cost was applied based on the number of patients estimated to receive an escalated dose. This estimation was based on a systematic review of the literature in Crohn's disease [13]. It was assumed that the dose escalation was similar in UC. This estimate was varied in a scenario analysis. For drugs with weight-based dosing, doses for patients were computed based on a simulated baseline weight distribution, using a normal distribution with mean and standard deviation based on the SELECTION trial. Drug acquisition costs of conventional therapy provided in the model are based on UK costs obtained from the Monthly Index of Medical Specialties 2021 [12]. The usage of each treatment was sourced from TA547, both for conventional therapy alone, and as a concomitant therapy with advanced treatments [14].

Costs of administration were dependent on the mode of administration, i.e. intravenous, subcutaneous or oral. Orally administered drugs were assumed to have no administration cost. It was assumed that for subcutaneous injections, patients either self-inject their medication or acquire no administration costs otherwise due to homecare and support schemes offered by the manufacturers. The administration costs for intravenous drugs were assumed to be equal to the cost of an outpatient visit. Disease management costs comprised regular outpatient visits, blood tests, endoscopy and hospitalisations. Resource use inputs associated with each health state were based on a UK cost-effectiveness model, Tsai et al. [15]. The number of hospitalisation episodes in the model deviate from Tsai et al. 2008 and were increased, based on clinical expert opinion that hospitalisation rates increase as patient health worsens. The estimated annual hospitalisation episodes were increased from 0.30 reported in Tsai et al. to 1.20 for the response without remission health state, and to 1.50 for the active UC health state. Finally, the model included costs of adverse events in the form of serious infections.

The company's base-case results, without conventional care as a comparator, and with treatment sequences added upon request of the ERG, are presented in Table 1 for both populations. Filgotinib, at its current price, and disregarding confidential patient access schemes for comparators and subsequent treatments, is the cheapest of the technologies under comparison and dominates all comparators

**Table 1** Cost-effectiveness results, company base case

Technologies	NMB (20k)	NMB (30k)	Incremental QALYs FIL vs X	Incremental costs (£) FIL vs X	ICER (£/QALY) pairwise FIL vs X
<b>Company base-case (biologic-naïve), without conventional care, with company's treatment sequences</b>					
Filgotinib	£115,860	£252,927	–	–	–
Golimumab	£103,804	£240,308	0.056	– £10,931	FIL dominates
Tofacitinib	£105,216	£242,610	– 0.033	– 11,298	£345,631 SW
Adalimumab	£103,314	£239,788	0.059	– £11,361	FIL dominates
Infliximab	£100,385	£237,575	–0.012	– £15,722	£1,273,598 SW
SC vedolizumab	£101,019	£240,185	–0.210	– £19,040	£90,695 SW
IV vedolizumab	£96,228	£235,088	–0.179	–£23,220	£129,463 SW
<b>Company base-case (biologic-experienced), without conventional care, with company's treatment sequences</b>					
Filgotinib	£113,927	£247,120	–	–	–
Adalimumab	£110,610	£243,370	0.043	–£2452	FIL dominates
Tofacitinib	£109,527	£242,615	0.010	– £4191	FIL dominates
Vedolizumab SC	£109,411	£242,446	0.016	– £4200	FIL dominates
Ustekinumab	£108,259	£241,160	0.029	– £5085	FIL dominates
Vedolizumab IV	£108,110	£241,013	0.029	– £5239	FIL dominates

FIL filgotinib, ICER incremental cost-effectiveness ratio, IV intravenous, NMB net monetary benefit, QALY quality-adjusted life-year, SC subcutaneous, SW South-Western, X comparator

in the biologic-experienced population. In the biologic-naïve population, filgotinib dominates golimumab and adalimumab, while for all other comparators filgotinib saves costs but generates fewer quality-adjusted life-years. Notably, the incremental cost-effectiveness ratios reported here are not those used for decision making as confidential patient access schemes for the comparators and subsequent treatments are not included here.

In one-way sensitivity analyses, the economic model was found to be most sensitive to varying the health state specific costs, and treatment efficacy during the maintenance phase. In addition, several scenario analyses were undertaken to assess the impact of key variables on the model outcomes. For both subgroups of patients, the results were generally consistent with the base-case analysis. The model was most sensitive to treatment efficacy estimates using NMA sensitivity analyses results and various utility inputs. A lack of robust utility estimates and inconsistency in published sources is a key limitation in UC modelling.

The model was validated by the company by seeking early scientific advice with the purpose of validating the economic model structure, assumptions and clinical evidence used in the model. For example, dose escalation and health state resource use estimates and various other model inputs and assumptions were validated during interviews with five England-based gastroenterologists. Internal quality assurance measures were undertaken throughout the model development using extreme values and formula auditing to

ensure the consistency of model estimates. Furthermore, the model output was compared to previously published cost-effectiveness model costs and quality-adjusted life-years. The company concluded that filgotinib represents a cost-effective option in moderate and severe UC.

### 3.4 Critique of Cost-Effectiveness Evidence and Interpretation

#### 3.4.1 Modelling Treatment Sequences

The company did not include treatment sequences in the base case. At the request of the ERG, some were included, although clinical validation was lacking [1]. Crucially, the company did not demonstrate that efficacy remains the same between lines of therapy in the biologic-experienced population. In fact, at Technical Engagement, the ERG found an analysis of the company's own trial, SELECTION, that showed a reduction in efficacy [4]. This analysis shows that the proportion who achieve remission at 10 weeks for filgotinib 200 mg decreased from 16.3% in patients with one biologic failure to 7.4% in those with failure to at least two. There was also an increase in the placebo group from 2.0% to 3.7%, which meant that the treatment effect in the form of a relative risk, as calculated by the ERG, decreased from 8.2 to 2.0. Of course, it is unclear how this would compare to other biologics, although it might be reasonable to assume a similar reduction in effectiveness.



### 3.4.2 Modelling Loss of Response Over Time

The most crucial issue in this appraisal and UC appraisals in general is that there is insufficient evidence on loss of response to treatments. The company modelled loss of response to be equal for those in remission and those in only response (without remission), implying that if at the end of the maintenance phase there are fewer patients in remission than in response, this cannot turn around anymore in the sense that there will always be fewer patients in remission than in response over the full-time horizon of the model. The ERG questioned the clinical plausibility of this assumption. Remission may be more difficult to attain than response, but once in remission, patients may be more stable, and stay in remission longer before they lose response than patients in response without remission. This was supported by the ERG clinical expert, and this was also an assumption used in a previous technology appraisal, TA633, where this approach was accepted by the NICE committee [16]. Unfortunately, the model does not allow for this to happen, nor did the company provide a means to explore the impact of this assumption in a scenario analysis. Additionally, loss of response rates were assumed to be constant over time based on the proportion of non-responders at the end of the maintenance phase. In reality, loss of response will probably decrease over time, but there is no evidence to say exactly how (as stated by the company and confirmed by the ERG's clinical expert), and whether the rate of decrease would be similar between treatments. The impact of this uncertainty remained largely unexplored, as the company provided only a scenario with decreased loss of response after the first year, but it was still equal between health states and not substantiated with evidence. Altogether, the ERG considers loss of response to be a source of substantial uncertainty and likely a source of bias. The company's approach of not modelling loss of response differentially per health state probably induced bias, the likely direction of effect of which cannot be reported here for reasons of confidentiality given that the results of the NMA were marked commercially in confidence by the company.

### 3.4.3 Probability of Pouchitis Not Aligned with Utility

The probability of pouchitis used in the model was related to the incidence of all pouchitis events, but the utility related to chronic pouchitis. Given that chronic pouchitis has a greater impact on HRQoL, the probability of chronic pouchitis is likely more appropriate.

### 3.4.4 Uncertainty in Health State Utilities

The ERG believes that there is considerable uncertainty about the most appropriate utility estimates to be used in

the model. First, baseline values of the SELECTION trial for active UC were used, and 10-week values for response without remission, and remission. The company did not provide justification for this discrepancy. The baseline values do not capture potential improvements in HRQoL experienced by patients over the trial (induction) period. Furthermore, baseline estimates include those of potential responders (with or without remission) as well as non-responders. The baseline utility estimates are therefore likely biased and not reflective of the longer term non-responders they are applied to. The ERG believed that using the 10-week data from the SELECTION trial for all pre-surgery health states is most appropriate for the base-case analysis. Second, using biologic-naïve and biologic-experienced specific utility values for the respective models is likely to lead to more accurate results.

### 3.4.5 Questionable Application of Dose Escalation and Uncertainty in Resource Use Estimates

The application of dose escalation to all comparators but filgotinib was questioned by the ERG. While the company clarified that this was in line with the summary of product characteristics of filgotinib and comparators, dose escalation does not appear to be recommended in the NICE guideline for UC at any line including immediately prior to surgery [17]. A systematic review of dose escalation in UC showed that the percentage who undergo a dose escalation of anti-TNFs is quite uncertain, varying from 5.0 to 70.8% depending on the treatment [18]. Additionally, the time to dose escalation varied with one study showing an increase with time from 16% at 6 months to 44% at 36 months. What appears to be clear is that it is not the case that it only occurs immediately prior to surgery, as claimed by the company [16]. Indeed, the company assumes 30% of patients undergo escalation based on a study in Crohn's disease, but that study states: "In general, dose escalation was used when patients had a partial response, absence of response, or loss of response" [13]. Additionally, it is unclear why there is only an increase in cost with dose escalation given that one would expect that dose escalation occurs in order to either re-induce or prolong a response. Indeed, the study by Gemayel et al. indicates that dose escalation also occurs in those who have responded to initial treatment [18]. It also provides some evidence that dose escalation can be very effective: for example, although only from one small ( $n = 41$ ) study, percentage response and remission with infliximab from 5 to 10 mg/kg at week 8 was 87% and 67%, respectively. Overall, the ERG considered that the company's inclusion of a dose escalation in the model might induce bias and was a source of uncertainty.

Second, there is uncertainty surrounding the resource use estimates for the health states, especially the active UC and response without remission health states. Based on early

scientific advice, expert input from five UK clinicians was sought by the company. However, this input was not used in the base-case analysis. The clinical experts advised lower hospitalisation rates and outpatient visits in the more severe health states (i.e. active UC and response without remission). Applying higher health state costs for active UC and response without remission in the base case is favourable for those treatments that result in lower proportions of patients in those health states.

### 3.5 Additional Work Undertaken by the ERG

Based on all considerations highlighted in the ERG critique, the ERG defined a new base case in which various adjustments were made to the company's base case. These included:

- disabling dose escalations for all comparators;
- using trial results instead of the NMA for the maintenance phase to inform efficacy;
- using a 10-week utility value instead of a baseline value for the active UC health state;
- using a lower probability for pouchitis, aligning with the utility value that was related to chronic pouchitis.

Furthermore, the ERG explored decreasing the loss of response by 25% after the first year (as also implemented by the company), applying alternative treatment sequences, excluding treat-through trials and the use of 26-week utilities in scenario analyses. All cost-effectiveness analyses were presented in terms of net monetary/net health benefit because there were multiple comparators for each subgroup and this helped with the interpretation of results. No probabilistic sensitivity analysis was performed because effectiveness estimates would not have been included as the ERG changed from using the NMA for maintenance trial efficacy to using trial estimates, which could not easily be incorporated in the probabilistic sensitivity analysis.

The ERG base-case and scenarios indicate that at its current price, and disregarding confidential patient access schemes for comparators, filgotinib dominates some comparators (all but intravenous and subcutaneous vedolizumab in the ERG's base case, plus infliximab in the scenario using 26-week utilities) in the biologic-naïve population and dominates all comparators in the biologic-experienced population. These results were fairly robust, but in the opinion of the ERG should be interpreted with caution because important uncertainties were not included in the modelling, most notably: loss of response (e.g. diminishing over time and differential per health state), uncertainty about HRQoL estimates and appropriate modelling of dose escalation.

### 3.6 Conclusions of the ERG Report and Technical Engagement

The company's economic model met most of the NICE reference case criteria, except for a full incremental analysis, which the model did not accommodate. Uncertainty remained about the effectiveness of filgotinib in the maintenance phase, given questionable validity of the maintenance NMA. The approach taken by the company to model loss of response was also a source of remaining uncertainty. It also remained uncertain to what extent the health state utilities applied in the model were appropriate, and whether the dose escalations for all comparators were justified.

## 4 Key Methodological Issues

### 4.1 Appropriateness of Maintenance Phase NMA

The purpose of an NMA is to compare outcomes between treatments that might be used for a particular patient population and so it makes no sense to compare treatments where the population varies by treatment. Note that population might be defined in terms of not only disease, but stage and treatment experience. Patients who have had a response induced by one treatment might therefore not be considered part of the same population as those induced by another, thus raising the question as to the validity of a maintenance phase NMA, as conducted by the company in this appraisal, where all treatments have been administered to patients only induced on the same treatment. This is also no surprise because those trials are of a re-randomisation design in order to compare continuation versus curtailment of treatment at maintenance as opposed to comparing continuation versus switching to a new treatment. Therefore, the ERG recommends that a maintenance phase NMA not be performed to inform effectiveness at induction or at maintenance when the choice is between continuation and curtailment of induction therapy.

### 4.2 How to Model Treatment Sequences

It is reasonable to assume that treatment outcomes vary by treatment experience (line and type of therapy). Therefore, the onus should be on the company to demonstrate that efficacy remains the same regardless in what line of therapy filgotinib is administered, which they did not do in this appraisal. In fact, the ERG found evidence that efficacy reduces in further lines of therapy with possibly a change in relative treatment effect. Therefore, the ERG recommends that any analysis should be informed by empirical estimates where available, informed by a systematic review to identify

any for the comparators, supplemented by plausible assumptions where empirical estimates are not available.

### 4.3 How to Model Loss of Response

The company and the ERG both agreed that loss of response would wane over time, but disagree as to whether the rate differs depending on the starting point, i.e. response without remission or remission, and also as to whether loss of response is likely to be constant over time. Nevertheless, the company has not presented any evidence to inform their assumption of no difference and constant loss of response. In this appraisal, the ERG recommends that an analysis of trial data be performed to estimate the rates of loss of response for each of the two groups of patients. This would be facilitated by designing trials with no re-randomisation. The ERG also establishes that more evidence on patterns of loss of response over time is needed to avoid oversimplistic approaches that could lead to biased estimates of both costs and effects of treatments.

## 5 National Institute for Health and Care Excellence Guidance

On 29 April, 2022, NICE recommended filgotinib, within its marketing authorisation, as an option for the treatment of moderately to severely active UC in adults. Filgotinib is recommended, within its marketing authorisation, as an option for treating moderately to severely active UC in adults when conventional or biological treatment cannot be tolerated, or if the disease has not responded well enough or has stopped responding to these treatments, and if the company provides filgotinib according to the commercial arrangement.

### 5.1 Consideration of Clinical Effectiveness

The SELECTION trial was considered broadly generalisable to UK clinical practice. Filgotinib is assumed to be more effective than placebo at inducing and maintaining remission. It might also be as effective as most comparators in the induction phase in the biologic-naïve population. However, in the biologic-experienced population, this is not so clear. In the maintenance phase, we could continue to argue that its efficacy in relation to other comparators is irrelevant if this is contingent on re-randomisation following response at induction. There does, however, remain a need for estimates of the relative efficacy beyond the induction phase contingent on inadequate response, loss of response or intolerance to previous therapy, which might be obtained from randomised controlled trials without re-randomisation. The committee concluded that it had concerns about the methodology of the

maintenance NMAs and that the effectiveness of filgotinib in the maintenance phase was uncertain.

### 5.2 Consideration of Cost Effectiveness

The AC considered the model structure to be appropriate, but was concerned that people receiving filgotinib were likely to have an increased risk of cardiovascular events, therefore balancing the benefit and risks before starting filgotinib was essential. The AC argued that cardiovascular adverse events should have been included in the model. Additionally, the AC agreed with the ERG that a long-term loss of response in the model should have been different in people in the ‘response without remission’ and ‘remission’ states and that a loss of response is unlikely to be constant over time. Furthermore, the AC considered the health state utility values for people with active UC to be uncertain. Comparator treatment sequences in the NHS vary and some of the company’s modelled treatment sequences were less plausible (for instance, tofacitinib after vedolizumab). The AC concluded that a range of treatment sequences for moderately to severe UC is plausible but the company’s modelled treatment sequences do not fully reflect clinical practice. The ERG’s approach of a consistent probability and quality-of-life impact of chronic pouchitis was considered appropriate, as well as the ERG’s approach to including the benefits of dose escalations with the costs. The AC noted that most of the uncertainties had a minimal effect on cost-effectiveness results. It considered the biologic-naïve and biologic-experienced subgroups separately. It concluded that, considering the uncertainty, the cost-effectiveness estimates for filgotinib, taking into account the commercial arrangement, compared with other treatments for moderately to severely active UC were below what NICE normally considers an acceptable use of NHS resources.

## 6 Conclusions

This article describes the STA considering filgotinib for moderately to severely active UC. Despite the uncertainty introduced by how loss of response and dose escalations were modelled, plus the uncertainty on health state utilities and on the effectiveness of filgotinib in the maintenance phase, the committee ruled that filgotinib can be considered a cost-effective use of NHS resources [2]. The committee noted that most of the uncertainties had a minimal effect on cost-effectiveness results. It therefore recommended filgotinib within its marketing authorisation, as an option for treating moderately to severely active UC in adults when conventional or biological treatment cannot be tolerated, or if the disease has not responded well enough or has stopped

responding to these treatments, and if the company provides filgotinib according to the commercial arrangement that was offered by the company at the time of submission.

**Acknowledgements** This summary of the ERG report was compiled after NICE issued the Final Appraisal Determination. All authors have commented on the submitted manuscript and have given their approval for the final version to be published. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of NICE or the Department of Health. Any errors are the responsibility of the authors.

## Declarations

**Funding** This project was funded by the National Institute for Health Research Health Technology Assessment Programme. Please visit the Health Technology Assessment programme website for further project information (<https://www.nihr.ac.uk/explore-nihr/funding-programmes/health-technology-assessment.htm>).

**Conflicts of interest/competing interests** Antoinette D.I. van Asselt, Nigel Armstrong, Merel Kimman, Andrea Peeters, Kevin McDermott, Lisa Stirk, Charlotte Ahmadu, Tim Govers, Frank Hoentjen, Manuela A. Joore, and Sabine E. Grimm have no conflicts of interest that are directly relevant to the content of this article.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Availability of data and material** The data used and generated during the current study is not publicly available other than what is published by NICE on their website at <https://www.nice.org.uk/guidance/ta792>. The reason for the confidentiality is that all information in the reimbursement dossiers sent to NICE by pharmaceutical companies is reserved except for data already available from public sources and data not marked as confidential by the companies.

**Code availability** Not applicable.

**Authors' Contributions** All authors have commented on the submitted manuscript and have given their approval for the final version to be published. NA, KMD and CA critiqued the clinical effectiveness data reported by the company. LS critiqued the literature searches undertaken by the company. FH provided clinical expert opinion on loss of response and treatment sequences. AvA, NA, MK, AP, TG, MJ and SG critiqued the mathematical model provided and the cost-effectiveness analyses submitted by the company. AvA acts as the overall guarantor for the article. This article has not been externally peer reviewed by *Pharmacoeconomics*.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

1. National Institute for Health and Care Excellence. Filgotinib for treating moderately to severely active ulcerative colitis [ID3736]. Committee papers. London: NICE; 2022. <https://www.nice.org.uk/guidance/ta792/documents/committee-papers>. Accessed 7 Nov 2022.
2. National Institute for Health and Care Excellence. Filgotinib for treating moderately to severely active ulcerative colitis. Technology appraisal guidance [TA792]. London: NICE; 2022. <https://www.nice.org.uk/guidance/ta792/resources/filgotinib-for-treating-moderately-to-severely-active-ulcerative-colitis-pdf-82611613942981>. Accessed 7 Nov 2022.
3. National Institute for Health and Care Excellence. Filgotinib for treating moderate to severe rheumatoid arthritis: final scope. London: NICE; 2020. <https://www.nice.org.uk/guidance/ta676/documents/final-scope>. Accessed 1 Oct 2021.
4. Peyrin-Biroulet L, Dotan I, Hibi T, Taliadouros V, Oortwijn A, Zhao S, et al. OP25 Efficacy of filgotinib in patients with ulcerative colitis by line of therapy in the phase 2b/3 SELECTION trial J Crohns Colitis. 2021;15(Suppl.\_1):S024–6.
5. Gilead Sciences Inc. NCT02914535. Clinical study protocol: a long-term extension study to evaluate the safety of filgotinib in subjects with ulcerative colitis [PDF provided with the company's submission]; 2019. p. 76. Accessed 17 Aug 2021.
6. National Institute for Health and Care Excellence. Vedolizumab for treating moderately to severely active ulcerative colitis. Technology appraisal guidance TA342. London: NICE; 2015. <https://www.nice.org.uk/guidance/ta342>. Accessed 1 Nov 2021.
7. National Institute for Health and Care Excellence. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy. Technology appraisal guidance TA329. London: NICE; 2015. <http://www.nice.org.uk/guidance/ta329>. Accessed 20 Aug 2021.
8. Ferrante M, Declerck S, De Hertogh G, Van Assche G, Geboes K, Rutgeerts P, et al. Outcome after proctocolectomy with ileal pouch-anal anastomosis for ulcerative colitis. *Inflamm Bowel Dis*. 2008;14(1):20–8.
9. Arseneau KO, Sultan S, Provenzale DT, Onken J, Bickston SJ, Foley E, et al. Do patient preferences influence decisions on treatment for patients with steroid-refractory ulcerative colitis? *Clin Gastroenterol Hepatol*. 2006;4(9):1135–42.
10. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–15.
11. NHS. National cost collection 2019. London: NHS England and NHS improvement, 2020. [https://www.england.nhs.uk/wp-content/uploads/2020/08/1\\_-\\_NCC\\_Report\\_FINAL\\_002.pdf](https://www.england.nhs.uk/wp-content/uploads/2020/08/1_-_NCC_Report_FINAL_002.pdf). Accessed 19 Jan 2023.
12. Monthly Index of Medical Specialities (MIMS). 2021. <http://www.mims.co.uk/>. Accessed 19 Jan 2023.
13. Einarson TR, Bereza BG, Ying Lee X, Lelli F. Dose escalation of biologics in Crohn's disease: critical review of observational studies. *Curr Med Res Opin*. 2017;33(8):1433–49.
14. National Institute for Health and Care Excellence. Tofacitinib for moderately to severely active ulcerative colitis. Technology appraisal guidance TA547. London: NICE; 2018. <https://www.nice.org.uk/guidance/ta547>. Accessed 1 Nov 2021.
15. Tsai HH, Punekar YS, Morris J, Fortun P. A model of the long-term cost effectiveness of scheduled maintenance treatment with

- infliximab for moderate-to-severe ulcerative colitis. *Aliment Pharmacol Ther.* 2008;28(10):1230–9.
16. National Institute for Health and Care Excellence. Ustekinumab for treating moderately to severely active ulcerative colitis. Technology appraisal guidance TA633. London: NICE, 2020. <https://www.nice.org.uk/guidance/ta633>. Accessed 1 Nov 2021.
  17. National Institute for Health and Care Excellence. NICE Pathways: ulcerative colitis overview. London: NICE; 2021 <https://pathways.nice.org.uk/pathways/ulcerative-colitis>. Accessed 17 Aug 2021.
  18. Gemayel NC, Rizzello E, Atanasov P, Wirth D, Borsi A. Dose escalation and switching of biologics in ulcerative colitis: a systematic literature review in real-world evidence. *Curr Med Res Opin.* 2019;35(11):1911–23.