

Baricitinib for Previously Treated Moderate or Severe Rheumatoid Arthritis: An Evidence Review Group Perspective of a NICE Single Technology Appraisal

Shijie Ren¹  · Iñigo Bermejo¹ · Emma Simpson¹ · Ruth Wong¹ · David L. Scott² · Adam Young³ · Matt Stevenson¹

Published online: 3 March 2018

© The Author(s) 2018. This article is an open access publication

Abstract As part of its single technology appraisal process, the National Institute for Health and Care Excellence invited the manufacturer (Eli Lilly) of baricitinib (BARI; Olumiant[®]; a Janus kinase inhibitor that is taken orally) to submit evidence of its clinical and cost effectiveness for the treatment of moderate to severe rheumatoid arthritis (RA) after the failure of disease-modifying antirheumatic drugs (DMARDs). The School of Health and Related Research Technology Appraisal Group at the University of Sheffield was commissioned to act as the independent Evidence Review Group (ERG). The ERG produced a detailed review of the evidence for the clinical and cost effectiveness of the technology, based on the company's submission (CS) to NICE. The clinical-effectiveness evidence in the CS for BARI was based predominantly on three randomised controlled trials comparing the efficacy of BARI against adalimumab or placebo, as well as one long-term extension study. The clinical-effectiveness review identified no head-to-head evidence on the efficacy of BARI against all the comparators within the scope. Therefore, the company performed network meta-analyses (NMAs) in two different populations: one in patients who had experienced an inadequate response to conventional DMARDs (cDMARD-IR), and the other in patients who had

experienced an inadequate response to tumour necrosis factor inhibitors (TNFi-IR). The company's NMAs concluded BARI had comparable efficacy as the majority of its comparators in both populations. The company submitted a de novo discrete event simulation model that analysed the incremental cost-effectiveness of BARI versus its comparators for the treatment of RA from the perspective of the National Health Service (NHS) in four different populations: (1) cDMARD-IR patients with moderate RA, defined as a 28-Joint Disease Activity Score (DAS28) > 3.2 and no more than 5.1; (2) cDMARD-IR patients with severe RA (defined as a DAS28 > 5.1); (3) TNFi-IR patients with severe RA for whom rituximab (RTX) was eligible; and (4) TNFi-IR patients with severe RA for whom RTX in combination with methotrexate (MTX) is contraindicated or not tolerated. In the cDMARD-IR population with moderate RA, the deterministic incremental cost-effectiveness ratio (ICER) for BARI in combination with MTX compared with intensive cDMARDs was estimated to be £37,420 per quality-adjusted life-year (QALY) gained. In the cDMARD-IR population with severe RA, BARI in combination with MTX dominated all comparators except for certolizumab pegol (CTZ) in combination with MTX, with the ICER of CTZ in combination with MTX compared with BARI in combination with MTX estimated to be £18,400 per QALY gained. In the TNFi-IR population with severe RA, when RTX in combination with MTX was an option, BARI in combination with MTX was dominated by RTX in combination with MTX. In the TNFi-IR population with severe RA for whom RTX in combination with MTX is contraindicated or not tolerated, BARI in combination with MTX dominated golimumab in combination with MTX and was less effective and less expensive than the remaining comparators. Following a critique of the model, the ERG undertook exploratory analyses after

✉ Shijie Ren
s.ren@sheffield.ac.uk

¹ School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

² Department of Rheumatology, King's College Hospital NHS Foundation Trust, London, UK

³ Center for Lifespan and Chronic Illness Research, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK

applying corrections to the methods used in the NMAs and two programming errors in the economic model that affected the company's probabilistic sensitivity analysis (PSA) results. The ERG's NMA results were broadly comparable with the company's results. The programming error that affected the PSA of the severe cDMARD-IR population had only a minimal impact on the results, while the error affecting the severe TNFi-IR RTX-ineligible population resulted in markedly higher costs and QALYs gained for the affected comparators but did not substantially modify the conclusions of the analysis. The NICE Appraisal Committee concluded that BARI in combination with MTX or as monotherapy is a cost-effective use of NHS resources in patients with severe RA, except in TNFi-IR patients who are RTX-eligible.

Key Points for Decision Makers

Baricitinib (BARI) has shown comparable clinical efficacy to the majority of recommended biologic disease-modifying antirheumatic drugs (bDMARDs) in previously treated moderate to severe rheumatoid arthritis (RA).

A confidential Patient Access Scheme has been agreed with the Department of Health under which BARI will be available to the National Health Service (NHS) at a reduced cost.

Estimated incremental cost-effectiveness ratios for BARI, in combination with methotrexate (MTX) or as monotherapy, versus its comparators, are within the range usually considered as a cost-effective use of NHS resources in patients with severe RA. The exception is for patients who have had an inadequate response to a tumour necrosis factor inhibitor (TNFi) and who are eligible for rituximab (RTX) in combination with MTX as RTX is of similar clinical efficacy to BARI but has a significantly lower cost.

In patients with moderate RA and a 28-Joint Disease Activity Score (DAS28) between 4.0 and 5.1, the estimated ICER for BARI in combination with MTX versus intensive conventional DMARDs was estimated to be £37,420 per quality-adjusted life-year gained.

1 Introduction

The National Institute for Health and Care Excellence (NICE) is an independent organisation responsible for providing national guidance on promoting good health and preventing and treating ill health in priority areas with significant impact. Health technologies must be shown to be clinically effective and to represent a cost-effective use of National Health Service (NHS) resources in order for NICE to recommend their use within the NHS in England. The NICE single technology appraisal (STA) process usually covers new single health technologies within a single indication, soon after their UK market authorisation [1]. Within the STA process, the company provides NICE with a written submission, alongside a mathematical model that summarises the company's estimates of the clinical and cost effectiveness of the technology. This submission is reviewed by an external organisation independent of NICE [the Evidence Review Group (ERG)], which consults with clinical specialists and produces a report. After consideration of the company's submission (CS), the ERG report and testimony from experts and other stakeholders, the NICE Appraisal Committee (AC) formulates preliminary guidance, the Appraisal Consultation Document (ACD), which indicates the initial decision of the AC regarding the recommendation (or not) of the technology. Stakeholders are then invited to comment on the submitted evidence and the ACD, after which a further ACD may be produced or a Final Appraisal Determination (FAD) issued, which is open to appeal. An ACD is not typically produced when the technology is recommended within its full marketing authorisation; in this case, an FAD is produced directly. In this STA, while there was a restriction on the use of baricitinib (BARI), NICE directly produced an FAD having considered the probability of an appeal.

This paper presents a summary of the ERG report [2] for the STA of BARI for previously treated moderate to severe rheumatoid arthritis (RA), and a summary of the subsequent development of the NICE guidance for the use of this technology in England. Full details of all relevant appraisal documents (including the appraisal scope, ERG report, company and consultee submissions, FAD and comments from consultees) can be found on the NICE website [3].

2 The Decision Problem

RA is an autoimmune disease that causes chronic inflammation, progressive, irreversible joint damage, impaired joint function, and pain and tenderness caused by swelling of the synovial lining of joints. The condition is associated with increasing disability and reduced health-related

quality of life [4]. The primary symptoms are pain, morning stiffness, swelling, tenderness, loss of movement, fatigue, and redness of the peripheral joints [5, 6]. RA is associated with substantial costs, both directly (due to treatment acquisition and hospitalisation) and indirectly (due to reduced productivity) [7]. The condition has long been reported as being associated with increased mortality [8, 9], particularly due to cardiovascular events [10]. NICE estimates that there are 400,000 people in the UK with RA [11], with approximately 26,000 incident cases per year [12]. RA is more prevalent in females (3.6 per 100,000 per year) than in males (1.5 per 100,000 per year) [13]. For both sexes, the peak age of incidence in the UK is in the eighth decade of life, but all ages can develop the disease [13].

Two classifications have dominated the measurement of improvement in RA symptoms: American College of Rheumatology (ACR) responses [14] and European League Against Rheumatism (EULAR) responses [15]. In the UK, progression of RA is often monitored using the 28-Joint Disease Activity Score (DAS28). The EULAR response criteria use both the change in DAS28 and the absolute DAS28 score to classify a response as good, moderate or none [15]. Although EULAR response has been reported less frequently in RCTs than ACR responses [16], it is much more closely aligned to the treatment continuation rules stipulated by NICE, which require at least a moderate EULAR response or a DAS28 improvement of more than 1.2 points to continue treatment with biologic disease-modifying antirheumatic drugs (bDMARDs).

2.1 Current Treatment

NICE recommends a combination of conventional DMARDs (cDMARDs) as first-line treatment for people with newly diagnosed RA, including methotrexate (MTX) and at least one other cDMARD plus short-term glucocorticoids, ideally beginning within 3 months of the onset of persistent symptoms [11].

For patients who have severe active RA (defined as a DAS28 > 5.1), the NICE guidance recommends the use of the following bDMARDs: abatacept (ABT), adalimumab (ADA), certolizumab pegol (CTZ), etanercept (ETN), golimumab (GOL), infliximab (IFX), or tocilizumab (TCZ), each in combination with MTX after failure to respond to cDMARDs [17, 18]. For patients with severe RA for whom MTX is contraindicated or has been withdrawn, the guidance recommends the use of ADA, CTZ, ETN, or TCZ as monotherapy [17, 18]. Most of these drugs (all except ABT and TCZ) are tumour necrosis factor inhibitors (TNFi). After failure of the first TNFi, the guidance recommends rituximab (RTX) in combination with MTX for the treatment of severe active RA [19];

however, if RTX is contraindicated or withdrawn because of an adverse event (AE), the guidance recommends ABT, ADA, ETN, GOL, IFX, TCZ and CTZ in combination with MTX [18–22]. Additionally, if MTX is contraindicated or withdrawn because of an AE, the guidance recommends ADA, ETN or CTZ [18, 19] as monotherapy. The guidance also recommends TCZ in combination with MTX as a third-line biologic after inadequate response to RTX in combination with MTX [20], and recommends discontinuing treatment with bDMARDs unless a moderate EULAR response is achieved at 6 months or if the response is not maintained [18–22]. After treatment discontinuation, the next treatment in the sequence is initiated.

3 The Independent Evidence Review Group (ERG) Review

In accordance with the process for STAs, the ERG and NICE had the opportunity to seek clarification on specific points in the CS [23], in response to which the company provided additional information [23]. Given the possibility that BARI could be of similar efficacy and price to many of the comparators, the ERG had agreed in advance with NICE that if that were the case then the analyses undertaken would not be extensive. As such, the STA was an informal pilot for NICE's fast-track appraisal that was being introduced [24]. The ERG critiqued the network meta-analyses (NMAs) undertaken by the company and provided results from alternative NMAs. Errors identified in the coding of the probabilistic sensitivity analyses (PSA) performed by the company were corrected by the ERG. The evidence presented in the CS, as well as the ERG's review of that evidence, is summarised here.

3.1 Clinical Evidence Provided by the Company

Evidence was presented in the CS [23] for the efficacy of BARI in combination with MTX or as monotherapy in previously treated moderate to severe RA. The key clinical-effectiveness evidence was based on three randomised controlled trials (RCTs). Two RCTs recruited MTX or cDMARD intolerant or inadequate response (cDMARD-IR) patients with RA (RA-BEAM [25], RA-BUILD [26]); RA-BEAM [25] included placebo (PBO) and ADA as comparators, while RA-BUILD [26] included only PBO as a comparator. The remaining RCT recruited patients with RA who had experienced an inadequate response to TNFi (TNFi-IR; RA-BEACON [27]). This RCT used PBO as a comparator. Additionally, one long-term safety and tolerability study was included (RA-BEYOND [28]).

For the primary endpoint of an ACR20 response at 12 weeks follow-up, all three RCTs reported that BARI

4 mg was statistically significantly superior to PBO ($p \leq 0.001$). Furthermore, at 12 weeks, more patients reached an ACR20 in the BARI 4 mg treatment arm than the ADA treatment arm ($p = 0.01$). There was also an advantage over PBO for BARI 4 mg at 24 weeks and BARI 2 mg at 12 and 24 weeks follow-up. The most common AEs for BARI were increases in low-density lipoprotein cholesterol, upper respiratory tract infections, and nausea; other adverse drug reactions included herpes simplex, herpes zoster, acne, increased creatine phosphokinase, increased triglycerides, increased liver function tests (aspartate transaminase, alanine transaminase), neutropenia and thrombocytosis.

NMAs were performed to assess the relative efficacy of BARI compared with the comparators in the cDMARD-IR or TNFi-IR patients with moderate to severe RA. For the base-case analysis at week 24 in the cDMARD-IR population, BARI 4 mg in combination with cDMARDs was associated with a statistically significant higher odds of an ACR50 response compared with cDMARDs, ADA, PBO, ETN and sulfasalazine (SSZ). No statistically significant differences were found versus any other comparators for the ACR50 outcome, with the exception of CTZ in combination with cDMARDs, in which odds of ACR50 response were found to be significantly in favour of the comparator. A similar pattern of results was observed for BARI 2 mg.

For the base-case NMA at week 24 in the TNFi-IR population, BARI 4 mg in combination with cDMARDs demonstrated significantly higher ACR50 response rates than the cDMARD comparator. No statistically significant differences were found versus bDMARDs in combination with cDMARDs, with the exception of the comparison of BARI (both 4 and 2 mg in combination with cDMARDs) with TCZ in combination with cDMARDs, and the comparison of BARI 2 mg in combination with cDMARDs with RTX in combination with cDMARDs, in which statistically significant treatment effects in favour of the comparators were observed.

A treatment effect in combination with cDMARDs was assumed to be the same as a treatment effect in combination with MTX in the economic model.

3.1.1 Critique of the Clinical Evidence and Interpretation

The ERG found the searches for clinical-effectiveness evidence reported in the CS to be adequate, and believed that all published RCTs of BARI were included in the CS. The eligibility criteria applied in the selection of evidence for the clinical-effectiveness review were considered by the ERG to be reasonable and consistent with the decision problem outlined in the final NICE scope [3]. The quality

of the included RCTs was assessed using well-established and recognised criteria.

The ERG stated that the results presented in the company's NMAs should be treated with caution because several problems were identified with the methods used. A random-effects model was assumed for the study-specific baseline treatment effects (pooling non-active and active controls). Simultaneous models for baseline and treatment effects were used, which means that the relative treatment effects were also affected by the inappropriate pooling among the baselines. Studies that reported EULAR responses were synthesised along with EULAR response outcomes converted from studies that only reported ACR responses, which did not ensure that the relative rankings of treatments are maintained. A random-effects model was used for the cDMARD-IR population. In contrast, a fixed-effect model was used for the TNFi-IR population since the company stated that random-effects models were unstable and did not converge. The choice between the use of fixed-effect and random-effects models should depend on the objective of the analysis and the conduct of the included studies, rather than on model convergence.

3.2 Cost-Effectiveness Evidence Provided by the Company

The manufacturer supplied a de novo discrete event simulation model constructed in Microsoft Excel[®] (Microsoft Corporation, Redmond, WA, USA). The model simulated patients' disease progressions through the sequences of treatments being compared. For each treatment, patients may achieve good, moderate or no EULAR response at 24 weeks. The EULAR response rates for each treatment were based on the company's NMAs. Patients who achieved a moderate or good EULAR response were assumed to have an improvement in Health Assessment Questionnaire (HAQ) score and remained on treatment until loss of efficacy (as assessed by a clinician), AE or death. Patients who experienced no EULAR response discontinued treatment at 24 weeks and started the next treatment in the sequence until they were receiving palliative care alone. The HAQ score of a patient while receiving bDMARDs or BARI treatment was assumed to be constant following the initial response; in contrast, while a patient was receiving cDMARDs or palliative care, HAQ progression was assumed to be non-linear based on latent HAQ trajectory classes [23]. Time to treatment discontinuation for responders was assumed independent of treatment but was dependent on EULAR response category (moderate or good) and was modelled using Weibull curves fitted to British Society for Rheumatology Biologics Register data. At treatment discontinuation, patients were assumed to suffer a rebound in HAQ score equal to the

improvement achieved on treatment initiation and were started on the next treatment in the sequence. The mortality rate was assumed to be affected by the HAQ score of a patient at treatment initiation.

The model estimated the costs and quality-adjusted life-years (QALYs) over patients' remaining lifetimes. Euro-QoL-5 Dimensions (EQ-5D) values were calculated based on a mapping algorithm generated from HAQ scores and patient characteristics [29]. Costs were considered from an NHS perspective. The company's analysis included costs associated with drug acquisition, drug administration and monitoring, and hospitalisation. Administration and monitoring costs were based on Technology Appraisal 375 (TA375) [16] and NHS Reference Costs 2014/15, hospitalisation costs and resource use estimates were based on HAQ score bands as in previous NICE technology appraisals, and drug costs were taken from the British National Formulary [30]. An annual discount rate of 3.5% was used for costs and outcomes. Serious AEs (SAEs) were excluded from the base-case but were included in a scenario analysis.

The analyses presented in the CS relate to four different populations of RA patients: (1) cDMARD-IR patients with moderate RA; (2) cDMARD-IR patients with severe RA; (3) TNFi-IR patients with severe RA who were eligible for RTX in combination with MTX; and (4) TNFi-IR patients with severe RA for whom RTX in combination with MTX was contraindicated or not tolerated. The definition of severe RA was a DAS28 > 5.1, while moderate RA was defined as a DAS28 > 3.2 and ≤ 5.1. Baseline characteristics of patients were based on the relevant clinical BARI trials.

In the cDMARD-IR population with moderate RA, the deterministic incremental cost-effectiveness ratio (ICER) for BARI in combination with MTX compared with intensive cDMARDs was estimated to be £37,420 per QALY gained. In the cDMARD-IR population with severe RA, BARI in combination with MTX dominated all comparators except for CTZ in combination with MTX; the ICER of CTZ in combination with MTX compared with BARI in combination with MTX was estimated to be £18,400 per QALY gained.

In the TNFi-IR population with severe RA, when RTX in combination with MTX was an option, BARI in combination with MTX was dominated by RTX in combination with MTX. In the TNFi-IR population with severe RA for whom RTX in combination with MTX was contraindicated or not tolerated, BARI in combination with MTX dominated GOL in combination with MTX, and was less effective and less expensive than the remaining comparators. The ICERs for ETN biosimilars, CTZ and ADA, all in combination with MTX, compared with BARI in combination with MTX were lower than £30,000 per QALY

gained. The ICERs for intravenous TCZ and subcutaneous ABT, both in combination with MTX, compared with BARI in combination with MTX were estimated to be higher than £30,000 per QALY gained.

3.2.1 Critique of the Cost-Effectiveness Evidence and Interpretation

The company's model was based on the model developed by the Assessment Group (AG) in NICE TA375 [16], with some minor deviations. The ERG believed that the conceptual model was appropriate but suffered from a series of implementation errors and limitations.

The ERG noted that the company did not identify any evidence on the effectiveness of ADA, CTZ, ETN and IFX in combination with MTX in severe TNFi-IR patients. In the absence of such data, the company used the same efficacy estimates of these treatments in severe cDMARD-IR patients instead, which is a favourable assumption for these interventions; therefore, caution is advised when interpreting these results. The company rounded modified HAQ values to the nearest valid HAQ score rather than allowing the valid HAQ score to be sampled probabilistically. The ERG noted that this approach might lead to inaccurate estimations of HAQ scores as values might be rounded up more often than rounded down or vice versa.

The company intended to implement the trajectory of HAQ score while patients were receiving cDMARDs or palliative care based on the latent class approach used by the AG in TA375. However, the company assigned each patient to a single class based on the probability of class membership instead of using an average weighted by the probability of class membership.

The company assumed that patients who achieve a moderate or good EULAR response at 24 weeks experience a reduction in HAQ score instantaneously at treatment initiation. The ERG believed that the company's approach is likely to lead to an overestimation of treatment benefits in relation to savings in RA-related costs as the achievement of response would take at least a few weeks, and potentially up to 24 weeks for some patients.

In order to calculate the QALYs and costs produced in the time span between two events, the model used an area under the curve (AUC) approach for the HAQ score and then mapped this value to both an EQ-5D score and hospitalisation costs; however, since the relationships between HAQ score and EQ-5D, and between HAQ score and hospitalisation costs, are not linear, this approach may lead to inaccurate results.

The TCZ subcutaneous formulation was not included in the list of comparators despite intravenous TCZ being included. The company argued that it had excluded subcutaneous TCZ because (1) the available evidence for

subcutaneous TCZ was limited; (2) it provided a lower efficacy estimate than for intravenous TCZ; and (3) the cost difference between the two formulations was relatively small. The ERG noted that the difference in costs might be considerable, taking into account the administration costs and the confidential Patient Access Scheme (PAS). Intravenous ABT was included in the NMA but was excluded from the analyses. In response to a clarification request by the ERG, the company presented the results of intravenous ABT only for the cDMARD-IR population with severe RA, which led to similar results compared with subcutaneous ABT.

The company used one of the algorithms proposed by Hernández Alava et al. [29] to map HAQ scores to EQ-5D. The ERG noted that newer algorithms with a higher accuracy have since been published, such as that reported by Hernández Alava et al. [31] and used in TA375.

3.3 Additional Work Undertaken by the ERG

The ERG's critical appraisal identified a number of issues relating to the company's model and analysis. The ERG believed that the NMA is subject to potential limitations; some of the scenario analyses, as well as the PSA for the severe TNFi-IR, RTX-ineligible population, lacked face validity; the efficacy estimates for the cDMARD-IR population should only be used for the first-line of treatment; rounding to the nearest HAQ score may have introduced bias; the HAQ trajectory of a patient receiving cDMARDs or palliative care should have been calculated as a weighted average; assuming HAQ improvement upon treatment initiation overestimated treatment benefit in relation to savings in RA-related costs; averaging HAQ across large time periods led to inaccuracies in the calculation of costs and QALYs; subcutaneous TCZ should have been included in the list of comparators; newer mapping algorithms from HAQ scores to EQ-5D should have been used; BARI should have not been assumed to be provided before intensive cDMARDs for moderate patients; mortality rates differed between sequences; the distribution of weight for interventions where the dosage is weight-based should have been considered; and the dosage of IFX was inaccurate.

The ERG re-analysed both the ACR and EULAR outcomes at week 24 for both the cDMARD-IR and TNFi-IR populations in the NMAs; however, due to the similarity in efficacy between bDMARDs, the ERG undertook few exploratory analyses, except amending two programming errors that affected the company's PSA results.

3.3.1 Additional Network Meta-Analyses

In the ERG's NMAs, all cDMARDs were assumed to have equivalent efficacy and were grouped together. The company provided data in the format for NMA for the cDMARD-IR population EULAR outcomes. The ERG amended the EULAR data used for van de Putte et al. [32] so that the moderate EULAR responders did not include good EULAR responders.

For EULAR outcomes in the TNFi-IR population, and ACR outcome in the cDMARD-IR and TNFi-IR population, the ERG computed the number of responses in each category using the data provided in percentages reported in the CS and in response to clarification request. The ERG's ACR NMA used the same included studies as those used in the CS, while the ERG's EULAR NMA only included studies that reported EULAR outcomes, rather than introducing EULAR data converted from ACR data.

The model for the relative treatment effect used in the ERG's analyses was the same as in the NICE Decision Support Unit Technical Support Document [33] and did not assume a random-effects model for the baseline for each study. The baseline and relative treatment-effect models were run separately to make sure that the information in the baseline model did not propagate to the relative treatment-effect model.

A random-effects model was used for all ERG NMAs. For the TNFi-IR population, since data were sparse, an informative prior was assumed for the between-study standard deviation, as suggested by Ren et al. [34]. This was a lognormal distribution, with mean -2.56 and variance of 1.74^2 , which was truncated so that the odds ratio in one study would not be ≥ 50 times than in another. It represented the beliefs that heterogeneity being small was 15%, being moderate was 66%, and being high was 19%. The NMAs conducted by the ERG had total residual deviances, which indicated that the model used by the ERG provided a better fit than those conducted by the company.

For EULAR outcomes in the cDMARD-IR population, BARI 4 mg in combination with cDMARDs was associated with statistically significant beneficial treatment effects relative to PBO and cDMARDs. No statistically significant differences were found versus any other comparator, with the exception of TCZ in combination with cDMARDs, which was associated with statistically beneficial treatment effects relative to BARI 4 mg in combination with cDMARDs.

For ACR outcomes in the cDMARD-IR population, BARI 4 mg in combination with cDMARDs was associated with statistically significant beneficial treatment effects relative to PBO, cDMARDs and ADA monotherapy. No statistically significant differences were found versus any other comparator, with the exception of CTZ in

combination with cDMARDs, which was associated with a statistically significant beneficial treatment effect relative to BARI 4 mg in combination with cDMARDs.

For EULAR outcomes in the TNFi-IR population, BARI 4 mg in combination with cDMARDs was associated with statistically significant beneficial treatment effects relative to cDMARDs. No statistically significant differences were found versus RTX in combination with MTX, which was the only other comparator in the network.

For ACR outcomes in the TNFi-IR population, BARI 4 mg in combination with cDMARDs was associated with a statistically significant beneficial treatment effect relative to cDMARDs. No statistically significant differences were found versus any other comparator.

3.3.2 Exploratory Analyses for the Economic Model

The programming error that affected the PSA of the severe cDMARD-IR population had a minimal impact on the results; CTZ in combination with MTX had the greatest probability of being most cost effective, at a cost per QALY thresholds of £20,000 or more, followed by BARI in combination with MTX. In contrast, the programming error that affected the PSA of the severe TNFi-IR RTX-ineligible population resulted in markedly higher costs and QALYs gained for TCZ, ETN biosimilars, IFX biosimilars, GOL and ADA, all in combination with MTX. As a result, the affected comparators, with the exception of GOL, shifted from the southwestern to the northeastern quadrant in the cost-effectiveness plane compared with BARI in combination with MTX, with ICERs ranging from £10,197 to £37,063, thus aligning with the results of the deterministic analysis for this population. CTZ in combination with MTX had the greatest probability of being cost effective, at a cost per QALY thresholds of £20,000 or more. The ERG commented that the results in the TNFi-IR RTX-ineligible population are confounded by the assumption for some interventions that the EULAR responses obtained in the cDMARD-IR population was applicable to the TNFi-IR RTX-ineligible population.

3.4 Conclusions of the ERG Report

BARI in combination with MTX treatment was estimated by the company to have an ICER of £37,420 per QALY gained compared with intensive cDMARDs in the moderate RA cDMARD-IR population, and shortcomings in the analysis led the ERG to believe that the actual figure is markedly higher. In the severe RA cDMARD-IR population, BARI in combination with MTX dominated all its comparators, except for CTZ in combination with MTX; the ICER of CTZ in combination with MTX compared with BARI in combination with MTX was estimated to be

£18,400 per QALY gained. In the severe RA TNFi-IR population, when RTX in combination with MTX was an option, BARI in combination with MTX was dominated by RTX in combination with MTX. In patients with severe RA who have had inadequate response to a TNFi and for whom RTX in combination with MTX is contraindicated or not tolerated, BARI in combination with MTX dominated GOL in combination with MTX and was less effective and less expensive than the rest of its comparators. The ICERs for ETN biosimilars and ADA, all in combination with MTX, compared with BARI in combination with MTX, were estimated to be lower than £30,000 per QALY gained, while the ICERs for intravenous TCZ and subcutaneous ABT, both in combination with MTX, compared with BARI in combination with MTX, were estimated to be higher than £30,000 per QALY gained; however, the confidential PASs for ABT and TCZ were not included in these analyses and the results in the TNFi-IR RTX-ineligible population are confounded by the assumption for some interventions that the EULAR responses obtained in a cDMARD-IR population was applicable to the TNFi-IR RTX-ineligible population.

4 Key Methodological Issues

The company used a random-effects model for the study-specific baseline treatment effects that inappropriately pooled the non-active and active controls. In addition, studies that reported EULAR responses were synthesised along with converted EULAR response outcomes from studies that only reported ACR responses. A random-effects model was used for the cDMARD-IR population. In contrast, a fixed-effect model was used for the TNFi-IR population, given the justification that random-effects models were unstable and did not converge.

With the agreement of NICE, errors that were identified were not amended if they did not change the conclusions. This was the case where the efficacy of the bDMARDs was similar. As such, this was an informal pilot for NICE's fast-track appraisals.

5 National Institute for Health and Care Excellence Guidance

In June 2017, on the basis of the evidence available (including verbal testimony of invited clinical experts and patient representatives), the AC produced guidance that BARI in combination with MTX was recommended as an option following an inadequate response to intensive therapy with cDMARDs for treating severe RA. It was also recommended as an option, following an inadequate

response to other DMARDs, including at least one bDMARD, for treating severe RA if patients could not receive RTX in combination with MTX. The AC also produced guidance that BARI monotherapy was recommended if MTX was contraindicated or not tolerated, under the same criteria as for combination treatment. All recommendations were conditional on the company providing BARI with the agreed PAS.

5.1 Consideration of Clinical and Cost-Effectiveness Issues Included in the Final Appraisal Determination

This section summarises the key issues considered by the AC. The full list of the issues considered by the AC can be found in the FAD [35].

5.1.1 Current Clinical Management

In England, the AC considered the current clinical management for people with severe RA that has not responded to intensive treatment with combinations of conventional DMARDs, and noted that the NICE guidance recommends the following bDMARDs: ADA, ETN, IFX, CTZ, ABT, TCZ and GOL (each with MTX). For people who meet these criteria but cannot take MTX, the guidance recommends that ADA, CTZ, ETN or TCZ may be used as monotherapy. For people with severe RA who had had an inadequate response to at least one TNFi, the guidance recommends ADA, ETN, IFX, RTX and ABT (each with MTX) as options. If RTX in combination with MTX is contraindicated or withdrawn, the guidance recommends ABT, ADA, ETN or IFX (each with MTX); however, if MTX is contraindicated or withdrawn, the guidance recommends ADA, ETN, TCZ or CTZ. NICE also recommends CTZ in combination with MTX following either a TNFi or RTX.

5.1.2 Uncertainties in the Clinical Evidence

The AC considered the company's clinical evidence and accepted that the results showed that BARI was more clinically effective than cDMARDs, and was as effective as ADA for moderate to severe RA patients who had responded inadequately to cDMARDs. Furthermore, the AC considered that BARI is more clinically effective than cDMARDs alone for moderate to severe RA patients who had responded inadequately to bDMARDs, and that BARI has a similar safety profile to cDMARDs and ADA.

The AC heard from the ERG that there were problems with the methods used in the company's NMAs. These included the conversion of ACR data to EULAR data before synthesis, the use of simultaneous models for

baseline and treatment effects, the use of a random-effects model for one population and a fixed-effects model for the other, and poor model fit. In addition, the company had pooled the control data inappropriately. The ERG corrected the errors in the company's NMAs. Having reviewed both analyses, the AC concluded that the results of the corrected NMAs and the company's NMAs were broadly comparable.

5.1.3 Uncertainties in the Economic Modelling

The AC had some concerns with how costs were calculated. The company's model included costs associated with drug acquisition, drug administration and monitoring, and hospitalisation. The AC was aware that BARI and several of the bDMARDs have PASs. It noted that the company had incorporated the PAS prices for BARI, CTZ and GOL into the model, but, as advised by NICE, not the confidential PAS for ABT and TCZ. The company had also calculated the average cost of drug doses using the average weight, rather than the distribution of the weight of the modelled patient population. The AC was also aware that the company overestimated the number of doses, and therefore the costs, of IFX.

The AC had concerns that the company was likely to have overestimated the efficacy of ADA, CTZ, ETN and IFX in combination with MTX in severe TNFi-IR patients. The company assumed the same efficacy estimates of these treatments as those in severe cDMARD-IR patients because of the lack of evidence for these treatments. Where data on both were available, the EULAR responses for all treatments were higher in these patients than in those with an inadequate response to bDMARDs.

6 Conclusions

The evidence suggests that BARI in combination with MTX or as monotherapy has a comparable efficacy for treating moderate to severe RA as that of other bDMARDs already recommended by NICE. The economic analyses conducted by the company and the ERG estimated ICERs within the range usually considered by NICE as a cost-effective use of NHS resources for BARI in combination with MTX or as monotherapy versus some or all of its comparators in the considered populations. The exception was for TNFi-IR RTX-eligible patients and patients with moderate RA. Consequently, NICE recommended BARI in combination with MTX as an option for patients with severe RA who can tolerate MTX if (1) they have cDMARD-IR; (2) they have TNFi-IR, and RTX in combination with MTX is not an option; or (3) they have TNFi-IR and have already been treated with RTX in combination

with MTX. NICE recommended BARI monotherapy for cDMARD-IR and TNFi-IR patients with severe RA who cannot tolerate MTX.

Acknowledgements This summary of the ERG report was compiled after NICE issued the FAD. 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.

Author Contributions IB and MS critiqued the mathematical model provided and the cost-effectiveness analyses submitted by the company; ES critiqued the clinical-effectiveness data reported by the company; SR critiqued the NMAs performed by the company and conducted the additional NMAs; RW critiqued the literature searches undertaken by the company; and DS and AY provided clinical advice to the ERG throughout the project. All authors were involved in drafting and commenting on the final document. SR acts as the guarantor of the manuscript. This summary has not been externally reviewed by PharmacoEconomics.

Compliance with Ethical Standards

Funding This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme (project number 16/56/23). Visit the HTA programme website for further project information (<http://www.nihr.ac.uk>).

Conflicts of interest David L. Scott has no conflicts of interest relating to baricitinib; however, his department has received a peer-reviewed grant from Pfizer within the last 12 months to undertake academic research on polypharmacy in arthritis. The department has also received free ETN from Pfizer to use in an NIHR-funded programme grant in RA. David L. Scott and Matt Stevenson declare that they have written a commentary on Janus kinase inhibitors for *The Lancet* ([http://dx.doi.org/10.1016/S0140-6736\(17\)31659-8](http://dx.doi.org/10.1016/S0140-6736(17)31659-8)), which was written after the completion of this STA. Shijie Ren, Iñigo Bermejo, Emma Simpson, Ruth Wong, and Adam Young have no conflicts of interest to declare.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. <https://www.nice.org.uk/article/pmg9/>. Accessed 6 Apr 2017.
2. Stevenson M, Bermejo I, Simpson E, Ren S, Wong R, Scott D, et al. Baricitinib for treating moderate to severe rheumatoid arthritis: a single technology appraisal. NICE 2017. <https://www.nice.org.uk/guidance/ta466>. Accessed 6 Apr 2017.
3. National Institute for Health and Care Excellence (NICE). Baricitinib for moderate to severe rheumatoid arthritis. 2017. <https://www.nice.org.uk/guidance/ta466>. Accessed 6 Apr 2017.
4. Scott DL, Steer S. The course of established rheumatoid arthritis. *Best Pract Res Clin Rheumatol*. 2007;21(5):943–67.
5. Pincus T, Fuchs HA, Callahan LF, Nance EP, Kaye JJ. Early radiographic joint space narrowing and erosion and later malalignment in rheumatoid arthritis: a longitudinal analysis. *J Rheumatol*. 1998;25(4):636–40.
6. Drossaers-Bakker KW, Kroon HM, Zwinderman AH, Breedveld FC, Hazes JMW. Radiographic damage of large joints in long-term rheumatoid arthritis and its relation to function. *Rheumatology*. 2000;39(9):998–1003.
7. Allaire S, Wolfe F, Niu J, LaValley MP, Zhang B, Reisine S. Current risk factors for work disability associated with rheumatoid arthritis: recent data from a US national cohort. *Arthritis Care Res*. 2009;61(3):321–8.
8. Naz SM, Symmons DPM. Mortality in established rheumatoid arthritis. *Best Pract Res Clin Rheumatol*. 2007;21(5):871–83.
9. Dadoun S, Zeboulon-Ktorza N, Combescure C, Elhai M, Rozenberg S, Gossec L, et al. Mortality in rheumatoid arthritis over the last fifty years: systematic review and meta-analysis. *Joint Bone Spine*. 2013;80(1):29–33.
10. Meune C, Touzé E, Trinquart L, Allanore Y. Trends in cardiovascular mortality in patients with rheumatoid arthritis over 50 years: a systematic review and meta-analysis of cohort studies. *Rheumatology*. 2009;48(10):1309–13.
11. National Collaborating Centre for Chronic Conditions. Rheumatoid arthritis: national clinical guideline for management and treatment in adults. 2009. <https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0009576/>. Accessed 6 Apr 2017.
12. The National Audit Office. Services for people with rheumatoid arthritis. 2009. <https://www.nao.org.uk/wp-content/uploads/2009/07/0809823.pdf>. Accessed 24 Mar 2016.
13. Symmons DPM, Barrett EM, Bankhead CR, Scott DGL, Silman AJ. The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register. *Rheumatology*. 1994;33(8):735–9.
14. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American college of rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum*. 1995;38(6):727–35.
15. van Gestel AM, Prevoo MLL, van't Hof MA, van Rijswijk MH, van de Putte LBA, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: comparison with the preliminary american college of rheumatology and the world health organization/international league against rheumatism criteria. *Arthritis Rheum*. 1996;39(1):34–40.
16. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health Technol Assess*. 2016;20(35):1–610.
17. National Institute for Health and Clinical Excellence (NICE). Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for rheumatoid arthritis not previously treated with DMARDs or after conventional DMARDs only have failed. 2016. <https://www.nice.org.uk/guidance/ta375>. Accessed 6 Apr 2017.
18. National Institute for Health and Clinical Excellence (NICE). Certolizumab pegol for the treatment of rheumatoid arthritis. 2010. <https://www.nice.org.uk/guidance/ta186>. Accessed 6 Apr 2017.
19. National Institute for Health and Clinical Excellence (NICE). Adalimumab, etanercept, infliximab, rituximab and abatacept for

- the treatment of rheumatoid arthritis after the failure of a TNF inhibitor. 2010. <http://www.nice.org.uk/guidance/ta195>. Accessed 6 Apr 2017.
20. National Institute for Health and Clinical Excellence (NICE). Tocilizumab for the treatment of rheumatoid arthritis. 2012. <https://www.nice.org.uk/guidance/ta247>. Accessed 6 Apr 2017.
 21. National Institute for Health and Clinical Excellence (NICE). Golimumab for the treatment of rheumatoid arthritis after the failure of previous disease-modifying anti-rheumatic drugs. 2011. <http://www.nice.org.uk/guidance/ta225>. Accessed 6 Apr 2017.
 22. National Institute for Health and Clinical Excellence (NICE). Certolizumab pegol for treating rheumatoid arthritis after inadequate response to a TNF-alpha inhibitor. 2016. <https://www.nice.org.uk/guidance/ta415>. Accessed 6 Apr 2017.
 23. Eli Lilly. Baricitinib for treating moderate to severe rheumatoid arthritis [ID979]. 2017. <https://www.nice.org.uk/guidance/ta466>. Accessed 6 Apr 2017.
 24. National Institute for Health and Care Excellence (NICE). Fast track appraisal. 2017. <https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/process-guide-addendum-fast-track.pdf>. Accessed 6 Apr 2017.
 25. Taylor PC, Keystone EC, van der Heijde D, Weinblatt ME, del Carmen Morales L, Reyes Gonzaga J, et al. Baricitinib versus Placebo or Adalimumab in Rheumatoid Arthritis. *N Engl J Med*. 2017;376(7):652–62.
 26. Dougados M, van der Heijde D, Chen Y-C, Greenwald M, Drescher E, Liu J, et al. Baricitinib in patients with inadequate response or intolerance to conventional synthetic DMARDs: results from the RA-BUILD study. *Ann Rheum Dis*. 2017;76(1):88–95.
 27. Genovese MC, Kremer J, Zamani O, Ludivico C, Krogulec M, Xie L, et al. Baricitinib in Patients with Refractory Rheumatoid Arthritis. *N Engl J Med*. 2016;374(13):1243–52.
 28. Eli Lilly and Company. JADY Abbreviated Clinical Study Report. 2016. <https://www.nice.org.uk/guidance/ta466>. Accessed 6 Apr 2017.
 29. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Health*. 2012;15(3):550–61.
 30. British National Formulary. British National Formulary (BNF) 71. London: BMJ Group and Pharmaceutical Press; 2016.
 31. Hernández Alava M, Wailoo A, Wolfe F, Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. *Rheumatology*. 2013;52(5):944–50.
 32. van de Putte LBA, Atkins C, Malaise M, Sany J, Russell A, van Riel PLCM, et al. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. *Ann Rheum Dis*. 2004;63(5):508–16.
 33. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. 2013;33(5):607–17.
 34. Ren S, Oakley E, Stevens J. Incorporating genuine prior information about between-study heterogeneity in random effects pairwise and network meta-analyses. *Medical Decision Making (In press)*.
 35. National Institute for Health and Care Excellence (NICE). Baricitinib for moderate to severe rheumatoid arthritis—final appraisal determination. 2017. <https://www.nice.org.uk/guidance/ta466/documents/html-content>. Accessed 21 Nov 2017.