Vedolizumab for the treatment of adults with moderately to severely active ulcerative colitis: A Single Technology Appraisal

Errata

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Munira Essat, Research Associate, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Paul Tappenden, Reader in Health Economic Modelling, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Shijie Ren, Research Associate, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Alice Bessey, Research Associate, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Rachel Archer, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Ruth Wong, Information Specialist, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA Sami Hoque, Consultant Gastroenterologist, Barts Health NHS Trust, Whipps Cross University Hospital, Whipps Cross Road, London, E11 1NR Alan Lobo, Consultant Gastroenterologist, Sheffield Teaching Hospitals NHS Foundation Trust, Royal Hallamshire Hospital, Glossop Road, Sheffield, S10 2JF
Correspondence to	Munira Essat, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

| **Date completed** | October 2014 |

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 13/110/01.

The main differences noted between the studies in both the induction phase and maintenance phase relate to patient characteristics, study design (randomisation at baseline or re-randomisation of biologic induction-responders) and study duration. GEMINI1[8] and ULTRA2[37] both included patients with prior anti-TNF-α exposure and anti-TNF-α naïve patients, whilst ACT1,[23] ACT2,[23] PURSUIT-SC,[39] Suzuki 2014,[38] and ULTRA1[36] included only patients who were anti-TNF-α naïve. Within PURSUIT-M,[40] all recruited patients were golimumab induction-responders.[39] Patients with prior anti-TNF-α exposure may be a more difficult to treat population than those who are anti-TNF-α naïve. In GEMINI 1, failure to anti-TNFs was defined as inadequate response (i.e. primary non-responders to induction therapy with anti-TNF therapy), loss of response (i.e. secondary non-response/loss of response anti-TNF over time following initial response) or patients were intolerable to anti-TNFs. Whilst in ULTRA 2, previous use of anti-TNF agents other than adalimumab was permitted if the patient had discontinued its use due to a loss of response or intolerance to the agent for longer than 8 weeks (i.e. this study does not appear to have included primary non-responders to anti-TNFs). In two of the maintenance trials (GEMINI1[8] and PURSUIT-M[40]), only patients who responded to biologic induction therapy were included in the maintenance phase analysis; these patients were re-randomised to either active treatment or placebo at the start of the maintenance phase. In contrast, in ULTRA2,[37] ACT1/2,[23] and Suzuki 2014,[38] patients were randomised to induction and maintenance regimens at baseline. As noted in the MS[1] (page 124), these differences would have implications for the efficacy results. In addition, the duration of studies varied both in the induction phase (between 6 to 8 weeks) and the maintenance phase (between 52 to 54 weeks, further details are provided in Table 16 and Table 17). The MS[1] (page 125) notes that the difference in study duration in the maintenance phase would not have a great impact on the results; the ERG agrees with this statement.

Data for the study quality (validity) assessment of the RCT studies included in the NMA (see MS,[1] pages 116-122) appear to be derived from the published trial reports. Although a detailed evaluation of each of the included studies was not undertaken by the ERG, the studies appear to be reasonably well conducted (MS[1] pages 353-355). With the exception of GEMINI1, these trials have previously been reviewed as part of the multiple technology appraisal of infliximab, adalimumab and golimumab for the treatment of moderately to severely active UC after failure of conventional therapy.[41]

For the statistical analysis (MS,[1] pages 126-129), the manufacturer undertook separate NMAs for the anti-TNF-α naïve and anti-TNF-α experienced/failure subgroups and the ITT population. Induction phase data and maintenance phase data were synthesised separately. For the trials without re-randomisation at the end of the induction phase, the manufacturer's NMA assumes that patients that responded at the end of maintenance also all responded at end of induction. All outcome measures were modelled separately using a binomial likelihood and a logit link function. The models are reported on page 127 of the MS.[1]

4.4 Critique of the indirect comparison and/or multiple treatment comparison

An NMA was performed to compare treatment effects between vedolizumab, adalimumab, golimumab, infliximab and placebo for the outcomes of clinical response, clinical remission, discontinuation due to AEs and SAEs (Table 18) using data from the trials: GEMINI1,[8] ULTRA1,[36]

incorrect as the N for placebo and adalimumab should be 29 and 36, respectively. The Suzuki 2014[38] trial data for the maintenance phase clinical remission reported in Table 132 and 138 in the MS[1] report were incorrect as the n for placebo and adalimumab 40mg EOW should be 7 and 41, respectively. The PURSUIT-M[40] trial data for maintenance phase durable clinical response in the clarification response[4] page 59 were incorrect as the N for placebo, golimumab 50mg and golimumab 100mg should be 154, 151 and 151, respectively; the n for golimumab 50mg and golimumab 100mg should be 71 and 76, respectively. The ULTRA2[37] trial data for maintenance phase durable clinical response in the clarification response[4] page 59 were also incorrect as the n for adalimumab 40mg EOW should be 59. The ERG has not checked all the data presented by the manufacturer; hence it is unclear if data used for other outcome measures were all correct.

The manufacturer undertook separate NMAs of anti-TNF-α naïve and anti-TNF-α experienced/failure subgroups. However, the manufacturer did not provide a rationale for conducting such analysis on subgroups separately. The ERG considers that the disadvantage of conducting separate analyses is that the possibility of an interaction between treatment and subgroup cannot be explored. The ERG asked the manufacturer to conduct an additional meta-regression including type of population as a covariate to assess if there is an interaction. The manufacturer's response stated that *"when such a model is fitted to a small network, the model may pick up on variation which could be caused by any number of study differences (known or unknown) causing the result to be spuriously significant or not significant, e.g. due to a lack of data. At least 20 studies would be needed."*[4] and that because the maximum number of studies in any of the network was 7, no such analysis was performed. The ERG considers this point to be reasonable for conducting meta-regression in general. However, whether it is possible to undertake meta-regression analysis also depends on the number of treatments included and the assumption of the model coefficients. If conducting a meta-regression is indeed not possible, then the predictive distribution of treatment effects which incorporates extra variability should be presented.

Induction phase and maintenance phase data were synthesised separately by the manufacturer. The ERG considers this to be appropriate. The MS[1] acknowledges that the study designs of ULTRA2,[37] Suzuki 2014[38] and ACT1[23] are different from the designs employed within the GEMINI1 and PURSUIT-M[40] trials. In order to allow for comparison with adalimumab and infliximab, the manufacturer made the following adjustment to the trials without re-randomisation after the induction phase. When conducting the NMA for the maintenance phase, the manufacturer assumed that the responders at the end of induction were the same as the responders at the end of maintenance in calculating the probability of durable clinical response, clinical remission, mucosal healing, and CSF remission. However, this approach ignores the fact that non-responders at the end of induction could become responders at the end of the maintenance phase, and the number of events at the end of