Use of tumour necrosis factor alpha (TNF α) inhibitors adalimumab and infliximab for Crohn's disease

Report commissioned by: NHS R&D HTA Programme

On behalf of: The National Institute for Health and Clinical Excellence

Produced by: West Midlands Health Technology Assessment Collaboration
Department of Public Health & Epidemiology
University of Birmingham
Edgbaston
Birmingham
B15 2TT

Authors: Ms Janine Dretzke, Department of Public Health & Epidemiology
Mr Jeff Round, Institute of Health Sciences, University of Leeds
Dr Richard Edlin, Institute of Health Sciences, University of Leeds
Dr Martin Connock, Department of Public Health & Epidemiology
Dr Claire Hulme, Institute of Health Sciences, University of Leeds
Ms Jadwiga Czeczot, Department of Public Health & Epidemiology
Ms Anne Fry Smith, Department of Public Health & Epidemiology
Prof Chris McCabe, Institute of Health Sciences, University of Leeds
Dr Catherine Meads, Department of Public Health & Epidemiology

Correspondence to: Dr Catherine Meads
West Midlands Health Technology Assessment Collaboration
Department of Public Health & Epidemiology
University of Birmingham
Edgbaston
Birmingham
B15 2TT

Date completed: July 2008

Expiry Date: tbc

Source of funding

This report was commissioned by the NHS R&D HTA programme as project number 06/60/01.

Declared competing interests of the authors:

None

Acknowledgements

Dr Tariq Iqbal and Dr Ralph Boulton (University Hospital Birmingham) for providing clinical advice, Natalie Rowles for administration, Denise Cann (NACC) for advice on the patient perspective and Dr Fujian Song for internal peer review.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NHS R&D HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Dretzke J, Edlin R, Round J, Connock M, Hulme C, Czeczot J, Fry-Smith A, McCabe C, Meads C. Use of tumour necrosis factor alpha (TNF α) inhibitors adalimumab and infliximab for Crohn's disease. Systematic review and economic evaluation. Health Technology Assessment # tbc.

Contributions of Authors

The report's authorship is as follows:

Protocol: J Dretzke, C Meads, M Connock with comments from all co-authors

Literature searching: A Fry-Smith

Study selection, data extraction and quality assessment clinical effectiveness: J Dretzke, M Connock, J Czeczot, C Meads

Clinical effectiveness analysis: M Connock with comments from all co-authors

Systematic review cost-effectiveness: R Edlin, C Hulme, C McCabe

Appraisal of industry models: R Edlin, C Hulme, C McCabe with comments from all co-authors

De novo economic model: C McCabe, R Edlin, J Round

Writing of the report: J Dretzke, R Edlin, J Round, M Connock, C Meads, C Hulme, C McCabe with contributions from all co-authors

Project management: C Meads

ABOUT "HOME UNIT"

The West Midlands Health Technology Assessment Collaboration (WMHTAC) is an organisation involving several universities and academic groups who collaboratively undertake research synthesis to produce health technology assessments. Most of our members are based in the Department of Public Health & Epidemiology, University of Birmingham, however other members are drawn from a wide field of expertise including economists and mathematical modellers from the Health Economics Facility, University of Birmingham.

WMHTAC produce systematic reviews, health technology assessments and economic evaluations for NHS R&D HTA programme (NCCHTA), the National Institute for Health and Clinical Excellence (NICE), and for the health service in the West Midlands. WMHTAC also undertakes methodological research on research synthesis, and provides training in systematic reviews and health technology assessment.

CONTENTS

1. DEFINITION OF TERMS AND LIST OF ABBREVIATIONS

AE	Adverse event
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
CDEIS	Crohn's Disease Endoscopic Index of Severity
CEAC	Cost-effectiveness acceptability curve
CI	Confidence interval
CUA	Cost-utility analysis
EMEA	European Agency for the Evaluation of Medicinal Products
FDA	US Food and Drug Administration
HBI	Harvey-Bradshaw Index
HrQoL	Health-related quality of life
IBD	Inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
IBS	Irritable bowel syndrome
ICD	Infliximab clinical discretion
ICER	Incremental cost-effectiveness ratio
IMT	Infliximab maintenance treatment
IQR	Inter-quartile range
ITT	Intention-to-treat analysis
LOCF	Last observation carried forward
LVCF	Last value carried forward
NACC	National Association for Colitis and Crohn's Disease
PCDAI	Paediatric Crohn's Disease Activity Index

PDAI	Perianal Disease Activity Index
QALY	Quality-adjusted life year
QOL	Quality of life
RCT	Randomised controlled trial
SD	Standard deviation
SPC	Summary of Product Characteristics
TAR	Technology Assessment Report
TNF	Tumour necrosis factor

There is some difficulty with using the term 'episodic treatment' because it has several possible definitions, depending on where it is being used. Possible definitions include the following:

1. Giving treatment when patient experiences a disease relapse (if signs and symptoms reoccur) (see previous NICE guidance). The relapse could occur once in several years or much more frequently, such as every 11 weeks (see Rutgeerts 2004 report[3] of ACCENT 1 – median time interval between episodic infusions)
2. Treatment given to the comparator arm (ie placebo arm) of the ACCENT 1 trial[3,4], (see diagram for the treatment given). This includes patients who were given placebo and patients who were given infliximab, ie crossovers. It also does not distinguish between responders and non-responders
3. Treatment 'as needed with infliximab' (see Rutgeerts 2004 report of ACCENT I)[3].
4. 'Intermittent therapy' or 'induction only/reinitiation therapy' (see Abbott's industry submission response to WMHTAC TAR, top of page 2)[5]
5. Three retreatments for those who initially respond but subsequently relapse (see economic model in previous TAR p34)[6]
6. Retreatment with a single dose of infliximab (see Marshall model)[7]
7. Retreatment when patients relapse or do not respond (see Jaisson-Hot model) [8]

There is also some difficulty with the term "maintenance treatment". Generally this is thought to mean keeping patients who have initially responded to treatment in continuing response or remission. However, the following definitions have also been used:

- Any scheduled maintenance treatment (see most RCT reports and Jaisson-Hot cost effectiveness analysis) [8]
- Any continuing treatment, (to distinguish between induction and maintenance therapy and this continuing treatment can be episodic or scheduled maintenance) (see ACCENT I trial) [3,4]
- Any treatment that includes an induction and a maintenance phase (see Schering Plough response to WMHTAC TAR p3)[6]

In this report, the term episodic treatment has been used in different places in the clinical effectiveness section, particularly with reference to the ACCENT I trial, but does not specify what was meant by the term. In the critical appraisal of the infliximab industry submission, the term 'infliximab clinical discretion' has been used for clarity because the precise definition of episodic treatment that was being used in the model could not be determined.

2. EXECUTIVE SUMMARY

2.1 Background

Crohn's disease (CD) is a severe, life-long disease characterised by inflammation of the gastrointestinal mucosa. Main symptoms include chronic diarrhoea, abdominal pain, rectal bleeding and weight loss, and growth failure in children. Common complications are strictures (narrowing of the bowel), fistulas (creation of abnormal passageways between the bowel and other structures) and perianal disease (comprised of fissures, fistulas and abscesses). The disease is characterised by recurring flares of variable duration alternating with periods of remission of variable duration. There is no cure and most patients will need to take medication for large periods of their life and many will require surgery. CD manifests itself mainly during late adolescence or early adulthood; prevalence estimates range from 50 to 375 per 100,000. The impact on patients and society is high as ill-health can be life-long and can negatively affect education and employment as well as patients' quality-of-life. Costs to the NHS are high, particularly for patients needing hospitalisation.

Conventional treatment pathways are complex and include a wide range of drugs (corticosteroids, aminosalicylates, immunosuppressants, antibiotics), nutritional therapy and surgery. More recently, a group of drugs called tumour necrosis factor inhibitors (anti-TNF-α agents) have been evaluated for their effectiveness in CD. One of these, infliximab, is currently recommended by NICE (2002) for patients with severe, active CD, where patients are refractory to or intolerant of conventional treatment.

2.2 Objectives

The objectives of this Technology Assessment Report (TAR) were:

- to update a previous TAR on the effectiveness and cost-effectiveness of infliximab in adults with moderate to severe CD or fistulising CD who are refractory to or intolerant of conventional treatment
- to review the evidence on the clinical and cost-effectiveness of infliximab in children with moderate to severe CD who are refractory to or intolerant of conventional treatment
- to review the evidence on the clinical effectiveness and cost-effectiveness of a further anti-TNF-α antibody, adalimumab, in adults with moderate to severe CD who are refractory to or intolerant of conventional treatment
- to investigate whether there is evidence for greater clinical or cost-effectiveness for either adalimumab or infliximab

2.3 Methods

Clinical effectiveness

Standard systematic review methods were used for study identification and selection, data extraction and quality assessment. Only RCTs comparing adalimumab or infliximab to standard treatment (placebo), RCTs comparing adalimumab to infliximab, or RCTs comparing different dosing regimens of either adalimumab or infliximab in patients with moderate to severe, active CD intolerant or resistant to conventional treatmentwere eligible for inclusion. Outcomes reported in the trials were mainly based around changes in the Crohn's Disease Activity Index (CDAI), a questionnaire measuring various parameters associated with CD. Results are reported for those trial arms where dosing regimens were consistent with the respective licence indications. Results were presented in Forest plots but not pooled due to the existence of either a single trial or clinical heterogeneity where there was two trials that potentially could have been pooled. Formal indirect comparisons were not undertaken due to clinical heterogeneity of trials. Results are reported for those trial arms where dosing regimens were consistent with the respective licence indications.

Cost-effectiveness

A systematic review of published studies on the cost and cost-effectiveness of adalimumab and infliximab was undertaken. The economic models of cost-effectiveness submitted by the manufacturers of both drugs were critically appraised and, where appropriate, rerun using parameter inputs based on the evidence identified by the authors of the TAR. A de novo Markov state transition model was constructed to calculate the incremental cost-effectiveness ratio for adalimumab and infliximab therapy respectively compared to standard care.

2.4 Results

Clinical effectiveness review

Based on 11 trials, there was evidence from both induction and maintenance trials that both adalimumab and infliximab therapy were beneficial compared to placebo (standard care) for adults with moderate to severe CD and, for infliximab, for adults with fistulising CD; results were statistically significant for some time-points. These results were based on changes to the CDAI and, for fistulising disease, on rates of fistula closure. Results from maintenance trials were almost exclusively based on sub-groups of 'responders'. There was no direct evidence to show that 'responders' were more likely to benefit from treatment than 'non-responders' in the longer term. The maintenance trials, in the main, did not inform on persistence of the response (remission) state where point prevalence was reported. There is likely to be a benefit of infliximab therapy for children, but these results are uncertain as the trials had no placebo (standard care) arm; rates of spontaneous improvement could therefore not be quantified but are likely to be high. There was no valid evidence

regarding the relative effectiveness of 'episodic' and 'scheduled' infliximab treatment regimens. Few differences were found between treatment and standard care arms for selected adverse events, though high proportions of scheduled crossovers resulted in a lack of a true placebo group in most of the maintenance trials.

Cost-effectiveness review

No published studies on the cost-effectiveness of adalimumab were identified. The four independently funded studies identified for infliximab suggested high cost-effectiveness ratios (all above £50,000/QALY for non-fistulising disease and all above £100,000/QALY for fistulising disease).

Appraisal of industry submissions

For adalimumab there was a lack of clarity over the source and interpretation of data used in the industry model and key elements of the model could not be verified. Corrected results for both severe CD, and moderate and severe (combined) CD were substantially higher than in the industry submitted model; in the severe sub-group of patients the corrected ICER approached cost-effectiveness (at a threshold £30,000). For infliximab, errors were identified in the industry model (active CD), some of which could not be corrected. The authors' revision of the model (active CD) suggested that infliximab was cost-effective for episodic (clinician discretion) treatment, although an exact description of this intervention was lacking. The revised model indicated that scheduled maintenance treatment with infliximab was unlikely to be cost-effective. The revised industry model for fistulising CD also suggested that infliximab was unlikely to be cost-effective. The model was provided for paediatric CD was non-functional.

De novo economic model

A Markov model was developed from the NHS/PSS perspective to estimate the incremental cost per QALY for both drugs compared to standard care in (a) episodic therapy (as it was defined for the denovo economic model) for moderate and severe disease; and (b) maintenance therapy for moderate and severe disease. The model had a one-year time horizon and was constructed and analysed in Data TreeAge Pro 2006. The findings were that for induction, both adalimumab and infliximab are cost effective (dominant relative to standard care) in the management of severe CD and that adalimumab (but not infliximab) is cost effective for moderate CD, according to limits usually accepted by NICE. Neither drug is cost effective as maintenance therapy for moderate or severe disease by these criteria.

A budget impact assessment suggested that total cost to the NHS in England and Wales for induction in severe disease only could range between £17 and £92 million and for maintenance for one year between £140 and £200 million. These totals would be less if treatment was directed towards only

those CD patients whose condition was refractory to other treatment or who were intolerant or experience toxicity from these treatments and where surgery was inappropriate. It is unclear how many people would be in this category so the precise budget impact if the current NICE guidance is maintained was unclear.

2.5 Discussion

Regarding clinical effectiveness, there were concerns about the trial design and lack of clarity particularly regarding the maintenance trials, which may have affected interpretation of results. These related to the division of patients into sub-groups (responders and non-responders) at different time-points; the high proportions of scheduled crossovers resulting in a lack of a true placebo group; and uncertainties regarding the handling of missing binary and continuous data. Overall, the trials showed a benefit of both adalimumab and infliximab therapy over standard care, as measured by CDAI related outcome measures (or fistula closure for patients with fistulising CD). Uncertainties remain over the size of the effect for both drugs, the duration of effect (after 1 year), the best type of treatment regimen (e.g. scheduled or as required) and the type of patient who would benefit most (e.g. in terms of disease severity or being an early 'responder'). There are also uncertainties over whether the CDAI derived measures are adequate for capturing clinically meaningful changes in disease severity. Whilst trial populations overall may appear homogenous based on similar CDAI scores, individual patients are likely to vary in their disease manifestations and severity. All of the trials were in patients with 'moderate to severe' CD (or fistulising CD) and therefore none exactly matched the licence indications or NICE guidance, which specify the use of these drugs in patients with 'severe' disease. All trials were multi-centre and applicability to UK populations, particularly in terms of standard care being provided, and in terms of patients having failed or having become intolerant to conventional treatment, was uncertain.

The uncertainties in the clinical data (as outlined above) complicated the economic analyses. The published economic models relied heavily on a small body of data and data from small samples. In such cases, the interpretation of economic models within the published papers was difficult. Assessments of the industry-submitted models were hampered by inconsistent use of data and by lack of clarity about the source and interpretation of data. Both manufacturers submitted Monte Carlo simulation Markov models but unfortunately some of the models had serious errors. Also Markov models assume zero memory; how long a patient has been in a health state and how they got there may impact on resources and could be important in a CD patient group. Both the published cost effectiveness studies and the industry submission models lacked input of long term data.

2.6 Conclusions

Anti-TNF therapy with adalimumab or infliximab may have a beneficial effect compared to standard care on CDAI related outcome measures for induction and maintenance. Formal comparisons between the two drugs were not possible due to clinical heterogeneity between trials. Uncertainty remains regarding the size and duration of the effect of the two drugs and over the type of patient that is likely to benefit more or less from treatment. The findings were that for induction, both adalimumab and infliximab are cost effective (dominant relative to standard care) in the management of severe CD and adalimumab (but not infliximab) is cost effective for moderate CD, according to limits generally accepted by NICE. Neither drug is cost effective as maintenance therapy for moderate or severe disease. Perhaps, most importantly, the analysis reflected the fact that a substantial number of patients would achieve remission under standard care and that the incidence of relapse amongst those in remission was such that maintenance therapy would have to show greater effectiveness than at present and/or be much less costly than it currently is in order to reach the levels of generally accepted cost-effectiveness.

3. BACKGROUND

3.1 Description of health problem

3.1.1 Description of Crohn's disease

Inflammatory bowel disease refers to a group of chronic intestinal diseases characterised by inflammation of the gastrointestinal mucosa. The most common types of inflammatory bowel disease are ulcerative colitis and Crohn's disease (CD). CD can affect any part of the gastrointestinal tract, from mouth to anus, but most commonly the terminal ileum (35%) or the ileocaecal region (40%) are affected.[9]

The main symptoms of CD are dependent on disease location and include chronic or nocturnal diarrhoea, abdominal pain, anal lesions, rectal bleeding and weight loss. Clinical signs include pallor, cachexia, abdominal mass or tenderness, or perianal fissures, fistulas or abscesses. Systemic symptoms include malaise, anorexia or fever.[9-11] Extraintestinal symptoms related to intestinal inflammation include spondylarthritis, cutaneous manifestations or ocular inflammation. [11] In children, growth failure may be the primary manifestation of CD.[12]

CD can be defined using the Vienna classification, i.e. by location (L1 = terminal ileal, L2 = colonic, L3 = ileocolic, L4 = upper gastrointestinal), by disease behaviour (B1 = inflammatory (non-stricturing, non-penetrating), B2 = stricturing, B3 = penetrating) and age at diagnosis (A1 less than 40 years old, A2 greater than 40 years).[13] Stricturing disease refers to the narrowing of the bowel, which can lead to bowel obstruction, whilst penetrating (or fistulising) disease refers to the creation of abnormal passageways (fistulas) between the bowel and other structures such as the skin. Inflammatory disease (non-stricturing, non-penetrating) causes inflammation without any strictures or fistulas.

Approximately 40-50% of patients present with ileocolonic disease at the time of diagnosis, approximately 30% have isolated small bowel disease and approximately another 30% have pure colonic disease. It is estimated that only 10-15% of patients have a change in disease localisation in the 10 years after diagnosis.[14] Disease behaviour at diagnosis is inflammatory (non-stricturing and non-penetrating) in 70% of patients, stricturing in 17% and penetrating (fistulas or abscesses or both) in 13% of patients.[15]

Where the ileum and colon are affected, this is usually complicated by intestinal obstruction, inflammatory mass or abscess. Where disease is limited to the colon, patients commonly present with

rectal bleeding, perianal complications and extraintestinal complications involving the skin or joints. Gastric and duodenal manifestations include nausea and vomiting, epigastric pain or gastric outlet obstruction.[16]

Common complications are strictures, fistulas and perianal disease. Fistulas can develop between loops of bowel adjacent to the bladder, vagina or the skin. Perianal disease comprises fissures, fistulas and abscesses, and perianal manifestations may precede the onset of bowel symptoms. [9,16] Symptomatic perianal disease requiring therapy occurs in around 35% of Crohn's disease patients.[17] CD may also be complicated by sequelae related to malabsorption such as anaemia or metabolic bone disease.[16] Rare complications include acute dilatation, perforation and massive haemorrhage, especially when the disease affects the colon.

CD is characterised by recurring flares alternating with periods of remission. Most patients take medication for a large period of their life because if they stop they might experience a disease flare but some drugs are tapered off during periods of remission, then if a patient experiences a flare they then return to therapy.[14]

3.1.2 Aetiology

The aetiology of CD remains unknown. It is generally accepted that the disease is a response to environmental triggers (infection, drugs or other agents) in genetically susceptible individuals.[10] Smoking has been shown to be a risk factor in CD, with suggestions that smokers are more than twice as likely to develop the disease. [18] Areas under investigation to identify pathogenic mechanisms include: epidemiology (e.g. diet, drugs, water supply), the gut/environmental interface (e.g. work on luminal bacteria), the inflammatory process (e.g. cell signalling pathways) and genetics (e.g. studies on gene expression).[10] Exacerbating factors include intercurrent infections, smoking and the use of nonsteroidal anti-inflammatory drugs, whilst the issue of stress initiating or exacerbating CD remains controversial.[16]

3.1.3 Diagnosis

No definitive diagnostic test exists for CD. Overlapping features with other inflammatory bowel diseases, a potentially insidious onset, and the heterogeneity of manifestations and/or presentation without gastrointestinal symptoms can make diagnosis difficult.[16] Diseases with symptoms in common with CD include infectious diarrhoea, small bowel lymphoma, ulcerative colitis, appendicitis, coeliac disease and irritable bowel syndrome (IBS). A detailed clinical history, physical examination, laboratory tests and endoscopic evaluation are necessary to make an accurate diagnosis. A diagnosis of inflammatory bowel disease should be contemplated in patients presenting with chronic (bloody or non-bloody) diarrhoea, particularly nocturnal diarrhoea and/or weight loss,

abdominal pain, fever or extraintestinal manifestations. Family history of the disease should be considered. Signs of volume depletion, ulceration of the oral mucosa, perianal lesions or abdominal tenderness may be observed on physical examination. Laboratory tests should rule out infection and look for markers of inflammatory bowel disease. such as low serum albumin level or Vitamin B_{12} deficiency. Imaging studies of the bowel may be helpful; abdominal radiography may reveal mucosal oedema or dilated loops of small bowel or colon consistent with either inflammation or obstruction. On endoscopy, CD is characterised by deep, linear ulcerations that can occur as segmental areas of mucosal involvement separated by areas of normal intervening mucosa ('skip lesions'). Biopsy findings usually demonstrate transmural inflammation.[19]

CD may be unsuspected and incorrectly diagnosed in the elderly, with as many as 60% of patients being misdiagnosed initially compared to a misdiagnosis rate of only 15% in younger people. The delay in diagnosis has been calculated as 6.4 years after onset of symptoms in older patients compared to 2.4 years in younger individuals.[20]

3.1.4 Natural history

The disease location of CD is fairly stable; however the behaviour of the disease can vary substantially during its course. The disease changes from non-stricturing to either stricturing (in 27%) or penetrating disease (in 29%).[15] After the first year of diagnosis, 10-30% of patients have an exacerbation, 15-25% have low activity and 55-65% are in remission; 13-20% have a chronic active course of disease activity, 67-73% have a chronic intermittent course and only 10-13% remain in remission for several years.[15] Most patients with CD will require surgery within 20 years.[15] The lifetime risk for developing fistulas has been reported to be between 20-40%. Perianal fistulas are most common, followed by entero-enteric, with many patients developing a fistula at or before diagnosis of CD.[21] CD is associated with an increased risk of colonic carcinoma and the overall mortality is slightly higher than that of the overall population. [10]

A Danish study[22] of an inception cohort of 373 CD patients found the following disease activity distributions: 80% of patients had high activity at diagnosis, decreasing to an almost stable value of 30% in the following 25 years; a constant 15% of patients overall had low activity and around 55% could expect to be in remission each year. Individual patients however changed from year to year between relapse and remission. The study further found that over a 10 year period 20-30% of patients could expect to go into remission each year. There was a slight indication of the disease 'burning out' as late in the disease course (more than 15 years post diagnosis) slightly more patients (29%) changed from activity to remission compared to 14% changing from remission to activity. A separate analysis of 171 patients followed for at least 7 years after diagnosis found that, between years 3 and 7, 25% of patients had active disease every year, 22% were in remission and 53% changed between years in

remission and years with relapse.[22] This disease course was independent of initial treatment, age, sex, localisation and symptoms at diagnosis or time from onset to diagnosis. With regard to hospital admissions, 83% were admitted during the year of diagnosis, this decreased during the following five years to a constant 20% each year.

A US modelling study[23] studied a retrospective cohort and estimated a future life expectancy of 46.4 years for a representative CD patient aged 28.1 years at time of diagnosis. The projected clinical course consisted of 11.1 years in remission (with no medication), 18.9 years in post-surgical remission (no medication), 12.7 years of receiving aminosalicylate or a similar medication and disease severe enough to require corticosteroids or immunosuppressives lasted 3.2 years. This was based on a sample of 174 patients and on treatment practices used between 1970 and 1993, which may have changed over the course of the study.

A Norwegian study[13], which followed up 221 CD patients prospectively for 5 years found that during the observation period 28% had undergone surgery. At the time of the 5-year visit 54% used sulfasalazin and 5-aminosalicylic acid, 25% used oral glucocorticosteroids and 13% used azathioprine. There were 16% who had symptoms that interfered with everyday activities and 72% had taken oral glucocorticosteroids at some point during the 5 years.

These cohort studies and the models based on them indicate that the clinical course estimates will vary depending on a variety of characteristics of the patients within the cohort.

3.1.5 Incidence and prevalence

CD can occur at any age, but manifests itself mainly during late adolescence or early adulthood. Peak onset is between 15 and 30 years of age.[24,25] The incidence in younger years is higher in women than in men.[24,26] There is some inconsistency regarding differences in prevalence between women and men overall, with some studies finding a higher prevalence in women, and some finding no difference.[26] There is an increased prevalence amongst first- and second-degree relatives suggesting the involvement of genetic factors.[25] CD may also present later in life (sixth and seventh decade) when there tends to be more colonic involvement and disease manifestations may be less severe.[20]

The extent of CD varies across the world and is most common in developed countries, with the UK having one of the highest rates. It was previously thought that IBD occurred less frequently amongst ethnic minorities. However, studies of migrant populations have shown that ethnic and racial differences are more likely to be attributable to lifestyle and environmental influences than true genetic differences. Similar rates of IBD have been found in African-Caribbean and white children and adults in the UK.[18] No association between CD and social class was found in a UK prevalence

study; it has been suggested that this is attributable to exposure to risk factors becoming more similar across social classes.[26]

In regions with a high prevalence of CD, the incidence increased between the 1950s and 1980s, and stabilised after that, which can be explained by an increased availability of gastroenterology units and increased awareness of the disease.[25,27] Some studies suggest that there is still an upward trend, which may be due to continued variations in environmental risk factors.[26] Increases in less developed countries have recently been noted, and it has been suggested that this is a result of changes in lifestyle (e.g. more exposure to smoking, changes in diet).[25]

Table 1 shows the incidence and prevalence of CD in the UK taken from studies published from 2000 onwards. The incidence ranges from 3.8 to 10 per 100,000 per year and the prevalence ranges from 50 to 375 per 100,000. For children, the British Paediatric Surveillance Unit (BPSU) found an estimated incidence of 5.3 per 100,000 per year.[12] Differences in incidence and prevalence estimates may result from the way data is gathered, changes in disease awareness and diagnosis over time, or changes in disease risk factors. There is no national CD database that could be used to determine numbers of CD patients.

Table 1. Incidence and prevalence of CD in the UK

Study	Population/sample	Incidence CD (adults)	Prevalence CD (adults)
Carter et al., 2004[10]	Review by the British Society of Gastroenterology (based on several studies, no details on sample size)	5-10/100,000 per year	50-100/100,000
Ehlin et al., 2003[26]	The 1970 British Cohort study and the 1958 National Child Development Study (one week national birth cohorts); total sample population of 22,680 (70% of target population)	NR	1970 cohort at age 30: 375/100,000 (95% CI 262, 488) 1958 cohort at age 30: 211/100,000 (95% CI 127, 295) 1958 cohort at age 42: 325/100,000 (95% CI 221, 430)
Rubin et al., 2000[27]	Systematic search of GP records in North England (based on population of 135,723)	8.3/100,000 per year (95% CI 7.5-20.3)	144.8/100,000 (95% CI 124.8-168.8)
NACC[28]	UK (no details on sample)	5-10/100,000 per year	100/100,000
Shivananda et al., 1996[29]	Multi-centre study of 20 centres across Europe during 1991-93, one of these in Leicester (total sample size unclear)	Non-immigrants: 3.8/100,000 per year (95% CI 0.7, 6.9) Immigrants: 5.6/100,000 per year (95% CI 0.0, 12.5) All aged 15-64	NR
Stone et al., 2003[30]	Fifteen general practices recruited through the Trent Focus Collaborative Research Network, UK (based on population of 86,801)	NR	130/100,000 (95% CI 107, 157)
Yapp et al., 2000[31]	Information from clinical records, the department of pathology database and a questionnaire sent to local family practitioners in the city of Cardiff (total sample size unclear)	5.6/ 100,000 per year (95% CI 4.4-6.8)	NR
NR=not reported			

3.1.6 Impact of health problem

3.1.6.1 Significance for patients in terms of ill-health

The impact on patients and society is high, as patients are often diagnosed at a young age and ill health may be life-long. Medical treatments can cause secondary health problems and surgery can result in complications such as impotence or intestinal failure. Patients can find symptoms embarrassing and humiliating, and may have difficulties in gaining employment or insurance. Younger people in particular may have psychological problems and growth failure or retarded sexual development. Approximately 75% of patients are fully capable of work one year after diagnosis and 15% of patients are unable to work after 5-10 years of disease.[10] Similarly, a Danish study[22] found that, except for the year of diagnosis, 75-80% of patients were fully capable of work each year, 9-16% were incapable and 11-9% only partly capable; after 15 years, 15% of patients obtained a disablement pension. The National Association of Colitis and Crohn's Disease (NACC) website[32] states that most sufferers can be maintained in remission for most of the time and are able to lead a full working life, however, some with severe disease do not achieve their educational and career potential.

Information sheets produced by NACC[32]relating to the most frequently asked questions to the NACC helpline cover the following issues: difficulties finding insurance companies who will provide life cover, travel, critical illness, mortgage protection or health insurance (when offered, insurance can be more expensive than if they did not have CD); managing bloating and wind; managing diarrhoea; concerns for young people (particularly focusing on emotional aspects such as embarrassment, body image, anxiety); and supporting someone with CD.

A prospective cohort study[33] of health-related quality of life in 231 patients with CD found that patients' main worries (in decreasing order of magnitude of concern) related to 'having an ostomy bag', 'uncertain nature of disease', 'energy level', 'having surgery', 'pain and suffering', 'eating normally', 'feelings about my body' and 'effects of medication'. Other concerns related to loss of bowel control, career/finances, sexual relationships, body/self-image, being a burden to others, developing cancer or dying early. Quality of life (QoL) as measured in this study by the Short Form-36 (SF-36) was lower for CD patients compared to the general population (the SF-36 measures various aspects of physical and mental functioning). Factors having a negative impact on QoL were active disease, hospitalisation, receiving steroids, having colonic disease and surgery.

A discussion with a patient representative, who has also worked for the NACC helpline highlighted the following issues of particular concern to patients who contact the helpline (personal communication, Denise Cann, NACC, 5th September 2007):

- difficulty in coping with unpredictability of disease (particularly where patients have been in remission) and a lack of control over it
- difficulty in gaining employment or staying employed, finding insurance
- impact on family and social life
- impact on relationships, sexual activity and pregnancy
- embarrassing nature of disease, e.g. flatulence, need to frequently use toilets due to diarrhoea, incontinence
- distressing symptoms such as rectovaginal fistulas where faeces can be passed through the vagina
- coping with the general tiredness, malaise and lack of energy
- coping with side effects of treatments
- fear that (new) treatment may not work
- coping with depression
- difficulty particularly for children and teenagers to cope emotionally
- costs: drug and continence prescription charges, cost of many sets of clothing/linen, trips to hospital, loss of earnings

3.1.6.2 Significance for NHS

A UK study from 2004[34] calculated the cost of Crohn's disease. The setting was an NHS university hospital with a target population of around 330,000. Table 2 lists the costs for different patient groups.

Table 2. Cost of Crohn's disease

Patient group	Mean cost for 6 months*
All CD patients (with complete 6 month follow-up- 'prevalent' cases)	£1652 (95% CI £1221, £2239)
Ambulatory group	£516 (95% CI £452, £618)
Patients hospitalised during study period	*£6923 (95% CI £5415, £8919)*+
Quiescent disease	£275 (95% CI £235, £319)
Ambulatory patients suffering disease exacerbation ('flare')	£578 (95% CI £431, £701)
Hospitalised patients	*£5444 (95% CI £3894, £9242)*+
New 'incident' cases	£2662 (95% CI £1006, £5866)

* to include costs of primary care visits, add approximately £30 per patient per 6 months
\+ we were unable to resolve the discrepancy between these two figures; a reply from the author was not received

Costs comprised all secondary care costs, including drugs, tests (e.g. endoscopy, laboratory tests), in- and outpatient services and surgery. Cost estimates also included all associated costs such as staff salaries, pharmacy services and other miscellaneous costs. Costs did not include visits to a GP but these were estimated separately and amounted to less than £30 per patient per 6 months. The median

number of days lost from household and recreational activities in six months were 20 (IQR 9 to 60). 50% of employed patients had some loss of employment days, with a median loss of earnings of £299 (IQR £119 to £597). Mean out of pocket expenses were £66 (range 0 to £750) and included travel and over the counter medication. No patient in this cohort received infliximab or another anti-TNF-α.

The contribution of different items and services to the overall cost of CD in all patients was as follows (estimated from Figure 1 in paper): 37% surgery, 24% in-patient costs, 11% out-patient costs, 11% tests (laboratory, x-ray, endoscopy) and 17% drugs.

Six-month resource use in ambulatory and hospitalised CD patients is shown in Table 3 (adapted from Table 2 in Bassi 2004[34]). There were a total of 260 bed days for CD within the 6 month period, 196 surgical bed days and 12 days of intensive care bed occupancy.

Table 3. Resource use in hospitalised and ambulatory CD patients

Parameter (per 6 months)	**Ambulatory CD patients (n=130) Mean (range)**	**Hospitalised CD patients (n=28) Mean (range)**
Outpatient services (visits)		
IBD related	2.2 (0-7)	2.9 (0-8)
Extraintestinal	1.25 (1-3)	-
Dietician	0.07 (0-3)	0.1 (0-1)
Stoma nurse	-	0.03 (0-1)
Laboratory	7.6 (0-28)	35.3 (9-66)
Radiology		
Plain x ray	0.07 (0-1)	1.4 (0-4)
Barium enema	0.01 (0-1)	0.07 (0-1)
Barium follow-through	0.1 (0-1)	0.30 (0-2)
Ultrasound abdomen	0.02 (0-1)	0.18 (0-1)
CT abdomen/pelvis	0.01 (0-1)	0.01 (0-1)
MRI abdomen/pelvis	-	0.07 (0-1)
White blood cell scan	0.01 (0-1)	0.07 (0-1)
DEXA scan	0.07 (0-1)	-
Fistulogram	0.01 (0-1)	-
Endoscopy		
OGD	0.15 (0-1)	0.11 (0-1)
Sigmoidoscopy	0.05 (0-2)	0.18 (0-1)
Colonoscopy	0.1 (0-1)	0.3 (0-3)
Hospital admission	N/A	
No. of admissions		1.1 (1-2)
Length of each admission (days)		14 (4-40)
CT=computed tomography, MRI=magnetic resonance imaging, OGD=oesophagogastroduodenoscopy, DEXA= dual energy X-ray absorptiometry (for measuring bone density)		

3.1.7 Measurement of disease severity in adults

Working definitions of disease severity have been developed by the Practice Parameters Committee of the American College of Gastroenterology (2001)[11], which are:

Mild- moderate disease:
" Mild-moderate Crohn's disease applies to ambulatory patients able to tolerate oral alimentation without manifestations of dehydration, toxicity (high fevers, rigors, prostration), abdominal tenderness, painful mass, obstruction, or >10% weight loss"
Moderate-severe disease:
"Moderate-severe disease applies to patients who have failed to respond to treatment for mild-moderate disease or those with more prominent symptoms of fever, significant weight loss, abdominal pain or tenderness, intermittent nausea or vomiting (without obstructive findings), or significant anaemia."
Severe-fulminant disease:
" Severe-fulminant disease refers to patients with persisting symptoms despite the introduction of steroids as outpatients, or individuals presenting with high fever, persistent vomiting, evidence of intestinal obstruction, rebound tenderness, cachexia, or evidence of an abscess."
Remission:
"Remission refers to patients who are asymptomatic or without inflammatory sequelae and includes patients who have responded to acute medical intervention or have undergone surgical resection without gross evidence of residual disease. Patients requiring steroids to maintain well-being are considered to be 'steroid-dependent' and are usually not considered to be 'in remission'."

The severity of Crohn's disease is difficult to assess, and a global measure encompassing clinical, endoscopic, biochemical and pathological features is not available.[35] The most widely used disease activity measures include the Crohn's Disease Activity Index (CDAI), the Harvey-Bradshaw Index (HBI) or Simple Index, a simplified version of the CDAI, and the Perianal Disease Activity Index (PDAI). A commonly used health related quality of life measure is the Inflammatory Bowel Disease questionnaire (IBDQ). Other measures include the Crohn's Disease Endoscopic Index of Severity (CDEIS).

The CDAI was developed in the 1970s as there was a need for a single index to assess disease severity. Variables measured include number of liquid stools, abdominal pain, general well being, extraintestinal complications, use of antidiarrhoeal drugs, abdominal mass, haematocrit and body weight. Scores range from 0 to approximately 600 (see Appendix 1 for full description of the index and the scoring system used). Values of below 150 are suggestive of quiescent disease (remission)

and values above 450 are associated with very severe disease. [36] Some investigators have arbitrarily labelled CDAI scores of 150-219 as mildly active disease and scores of 220 to 450 as moderately active disease.[35]

The CDAI has been criticised for having limitations. It does not cover aspects of quality of life, such as psychological, social, sexual and occupational functioning. A patient with a low CDAI score may still be severely limited by the disease in those areas.[37] Substantial variability exists when different observers review the same case histories and calculate the CDAI score, although this can be reduced after discussion and education about the terminology. The calculation is based in part on a daily diary kept by the patient for seven days before the evaluation. In practice some investigators and study coordinators assist the patient to retrospectively complete the diary at the time of an evaluation visit; there is no information on the prevalence of this practice. The CDAI score may be low in patients whose primary symptom is drainage of enterocutaneous fistulas, presumably because the presence of an actively draining fistula contributes only 20 points to the score. The CDAI is therefore not an appropriate instrument for assessing the activity of draining abdominal or perianal enterocutaneous fistulas. The CDAI has been criticised for giving too much weight to 'general well-being' and 'intensity of abdominal pain', as these are relatively subjective items. However these aspects of disease are important to patients.[38]

Clinical studies have variously defined a clinical response as a decrease in CDAI of 50, 60, 70 or 100 points. The FDA and EMEA suggested in 2000 that a meaningful decrease of in the CDAI score is a decrease of 100 points.[38]

The HBI is a modified/simplified version of the adult CDAI. It uses a single day's reading for diary entries and excludes three variables (body weight, haematocrit and use of drugs for diarrhoea). Code values are added together rather than summing the products of code values and coefficients (see Appendix 1). Scores range from 0 to 20. The CDAI can be predicted reasonably well from the HBI.[39] Other instruments derived from the CDAI are: the Cape Town Index (CTI), which includes parameters on subjective symptoms, physician clinical findings and laboratory data; the three-variable version of the CDAI used for survey research; and the Van Hees Index (VHI), which includes laboratory parameters, sex (male or female) and seven clinical features and excludes subjective, patient related items such as well-being and pain.[37]

The PDAI was developed to account for the morbidity and impairment of quality of life of patients with perianal disease, and to evaluate the effectiveness of perianal disease treatment. Variables

include discharge, pain/restriction of activities, restriction of sexual activity, type of perianal disease (including number of fistulas) and degree of induration. Scores range from 0 to 20.[17]

The reliance on traditional disease activity measures (such as the CDAI) to measure treatment effectiveness fails to take into account the impaired quality of life experienced by CD patients. The Irritable Bowel Disease Questionnaire (IBDQ) is a health related quality of life measure. It is a 32 item questionnaire and evaluates general activities of daily living, intestinal function, social performance, personal interactions and emotional status. Four-dimensional scores cluster items under: bowel function, emotional function, systemic function and social function. Scores range from 32 to 224.[40]

The CDEIS was developed to take into account endoscopic data, such as lesion severity, when assessing severity of the disease. Variables include the presence or absence of deep or superficial ulceration in various segments of the intestinal tract, the surface involved (in cm), surface ulcerated (in cm) and presence of ulcerated stenosis. Scores range from 0 to 30. [41]

3.1.8 Measurement of disease severity in children

The paediatric CDAI is a multi-item measure of severity that includes linear growth and places less emphasis on subjectively reported symptoms and more on laboratory parameters of intestinal inflammation compared to the adult CDAI. It includes 11 variables including weight, height, abdominal mass, perirectal disease, extraintestinal manifestation, haematocrit, erythrocyte sedimentation rate, albumin, abdominal pain, number of liquid stools and general well-being. Scores range from 0 to 100: ≤10 indicates inactive disease, 11-30 mild disease and >30 moderate to severe disease.[42,43]

3.2 Current service provision

CD treatment includes nutrition, drugs and surgery. Nutrition includes complete elemental diets and nutritional supplements. Drug treatments can include aminosalicylates (mesalazine, sulfasalazine), corticosteroids (prednisolone, budesonide, i/v hydrocortisone, methylprednisolone). Licensed drugs are being used in unlicenced indications for chronically active CD, including immunomodulators (azathioprine, mercaptopurine and methotrexate) and the antibiotic metronidazole.[44] Cytokine modulators (also known as biologics) such as adalimumab and infliximab are licensed for severe active CD. Use of infliximab is subject to NICE guidance (see below). Adalimumab is discussed in the next section (see Section 3.3, Description of technology under assessment). Surgery is not curative and is used to manage symptoms. In patients with fistulas, treatment can include seton use and surgery. At least 50% of CD patients require surgical treatment in the first 10 years of disease and around 70-80% require surgery within their lifetime.[10]

NICE guidance on the current use of infliximab in Crohn's disease is as follows (Technology Appraisal Guidance No. 40):

"1.1 Infliximab is recommended for the treatment of patients with severe Crohn's disease who fulfil all three of the following criteria:

- *Patients who have severe active Crohn's disease. These patients will already be in very poor general health with weight loss and sometimes fever, severe abdominal pain and usually frequent (3–4 or more) diarrhoeal stools daily. They may or may not be developing new fistulas or have extra-intestinal manifestations of the disease. This clinical definition normally corresponds to a Crohn's Disease Activity Index (CDAI) score of 300 or more and a Harvey-Bradshaw Index of 8/9 or above*

- *Patients whose condition has proved to be refractory to treatment with immunomodulating drugs (e.g. azathioprine or 6mercaptopurine, methotrexate) and corticosteroids, or who have been intolerant of, or experienced toxicity from, these treatments.*

- *Patients for whom surgery is inappropriate (e.g. because of diffuse disease and/or a risk of short bowel syndrome).*

1.2 Treatment can be repeated for those patients who match the above criteria and have responded to the initial treatment course, but then relapsed. A decision about whether or not to re-administer infliximab after the first course or subsequently should be made only after discussion with the patient who has been fully informed of the potential risks and benefits of repeated therapy (episodic treatment).

1.3 Infliximab should be prescribed by a gastroenterologist experienced in the management of Crohn's disease.

1.4 Infliximab is not recommended for patients with fistulising Crohn's disease who do not have the other criteria for severe active Crohn's disease as detailed in section 1.1."

For current conventional treatment, the recommendations below are taken from the UK guidelines for the management of inflammatory bowel disease in adults from 2004[10] (see

Appendix 2 for full details on medical management of CD). In this guideline, treatment options are complex and depend on the severity of disease, whether first line treatments have failed, side effects, stage/type of disease (active, in remission, chronic, fistulising). Also, some treatment may be adjunctive.

For patients with active, ileal/ileocolonic/colonic disease, options include aminosalicylates (e.g. mesalazine), corticosteroids (e.g.prednisolone), antibiotics (e.g. metronidazole), immunosuppressants (e.g. azathioprine), nutritional therapy and surgery. Patients with fistulising and perianal disease can be treated with antibiotics or immunosuppressants, infliximab where CD is severe and active and fistulas are refractory to other treatment, nutritional therapy and surgery.

The efficacy of treatment for maintenance of remission depends on how remission was achieved (medically or surgically), on risk of relapse and site of disease. In addition to smoking cessation (one of the most important factors in maintaining remission), aminosalicylates, immunosuppressants or antimetabolites (e.g. methotrexate) can be used. Infliximab can be used for up to 44 weeks as part of a treatment strategy including immunomodulation. Corticosteroids are not effective for the maintenance of remission, although some patients appear steroid dependent. Immunomodulation should be tried as first line treatment in steroid dependent patients; infliximab should be reserved for patients with moderate to severe CD who are refractory or intolerant of treatment with steroids, mesalazine, azathioprine/mercaptopurine and methotrexate and where surgery is considered inappropriate. It has been estimated that around 2% of patients have severe, drug refractory disease but this is based on a Markov model rather than cohort data.[45]

In children, enteral nutrition is used as primary therapy for active CD by the majority of paediatric gastroenterologists in the UK.[12]

An audit[46] carried out in collaboration between the British Society of Gastroenterology, the Royal College of Physicians, the Association of Coloproctology of Great Britain and Ireland and the National Association for Colitis and Crohn's Disease (NACC) found marked variation in the resources and quality of care: They found that:

- 44% of sites did not have an IBD nurse specialist;
- there was poor provision of dietetic services;
- there was a lack of adequate toilet provision in hospitals;
- fewer than one fifth of hospitals were able to refer patients directly for psychological support;
- 42% of patients with IBD had a stool sample sent for culture;
- 52% of CD patients were weighed
- 37% seen by a dietician;

- many patients with CD were receiving inappropriately prolonged course of steroids;
- there was inadequate prophylactic bone protection therapy for patients on systemic steroids and inadequate screening for osteoporosis;
- there was infrequent participation in clinical research into IBD in the UK.

3.3 Description of technology under assessment

Adalimumab and infliximab are tumour necrosis factor inhibitors (anti-TNF-α antibodies). TNF-α is a cytokine, a small protein molecule acting as a cell messenger and involved in inflammatory conditions. It is a key mediator of the inflammation associated with CD and can be detected in diseased areas of the bowel wall and in blood and faeces of patients with the disease.[47] Both adalimumab and infliximab are manufactured antibodies that bind to and inhibit TNF-α thus reducing the inflammatory response. They belong to the pharmacotherapeutic group of selective immunosuppressive agents.[48] The term 'biologics' is also applied to these drugs as their production depends on cells that have been genetically engineered to produce a specific protein.

Adalimumab (Humira ®, Abbott Laboratories) is a recombinant, fully human monoclonal antibody expressed in Chinese Hamster Ovary cells. It binds specifically to TNF and neutralises its biological function. Adalimumab is available as Humira 40mg solution; each 0.8ml single dose vial contains 40mg of adalimumab. It is administered by subcutaneous injection. Treatment with adalimumab should be initiated and supervised by specialist physicians experienced in the treatment of CD. After training, patients may self-inject with adalimumab, with medical follow-up as necessary. Adalimumab is also licensed for use in rheumatoid arthritis, psoriatic arthritis and ankylosing spondylitis.[49]

The licence indication for CD detailed in the SPC[49] is as follows:

"Humira is indicated for treatment of severe, active Crohn's disease, in patients who have not responded despite a full and adequate course of therapy with a corticosteroid and/or an immunosuppressant; or who are intolerant to or have medical contraindications for such therapies. For induction treatment, Humira should be given in combination with cortiocosteroids. Humira can be given as monotherapy in case of intolerance to corticosteroids or when continued treatment with corticosteroids is inappropriate.

The recommended Humira induction dose regimen for adult patients with severe Crohn's disease is 80 mg at week 0 followed by 40 mg at week 2. In case there is a need for a more rapid response to therapy, the regimen 160 mg at week 0 (dose can be administered as four injections in one day or as

two injections per day for two consecutive days), 80 mg at week 2, can be used with the awareness that the risk for adverse events is higher during induction.

After induction treatment, the recommended dose is 40 mg every other week via subcutaneous injection. Alternatively, if a patient has stopped Humira and signs and symptoms of disease recur, Humira may be re-administered. There is little experience from re-administration after more than 8 weeks since the previous dose. During maintenance treatment, corticosteroids may be tapered in accordance with clinical practice guidelines. Some patients who experience decrease in their response may benefit from an increase in dose intensity to 40 mg Humira every week."

Infliximab (Remicade®, Schering-Plough) is a chimaeric human-murine monoclonal antibody manufactured from a recombinant cell line. It binds with high affinity to soluble and transmembrane forms of TNF thus inhibiting the functional activity of TNF. Infliximab is available as Remicade® 100mg powder for concentrate for solution for infusion; each vial contains 100mg of infliximab. Treatment with infliximab should be initiated and supervised by specialist physicians experienced in the treatment of CD. Infliximab is administered intravenously over a 2-hour period. Infusions should be administered by qualified healthcare professionals trained to detect infusion related issues; patients should be observed for at least 1-2 hours post-infusion for acute infusion-related reactions and emergency equipment (such as adrenaline) must be available. Patients may be pre-treated in order to avoid infusion related reaction, particularly where these have occurred previously. Infliximab is also licensed for use in rheumatoid arthritis, ulcerative colitis, ankylosing spondylitis, psoriatic arthritis and psoriasis.[50]

The licence indication for CD detailed in the SPC[50] is as follows:
"Adult Crohn's disease:Remicade is indicated for:

- *treatment of severe, active Crohn's disease, in patients who have not responded despite a full and adequate course of therapy with a corticosteroid and/or an immunosuppressant; or who are intolerant to or have medical contraindications for such therapies.*
- *treatment of fistulising, active Crohn's disease, in patients who have not responded despite a full and adequate course of therapy with conventional treatment (including antibiotics, drainage and immunosuppressive therapy).*

Paediatric Crohn's disease:
Treatment of severe, active Crohn's disease, in paediatric patients aged 6 to 17 years, who have not responded to conventional therapy including a corticosteroid, an immunomodulator and primary

nutrition therapy; or who are intolerant to or have contraindications for such therapies. Remicade has been studied only in combination with conventional immunosuppressive therapy.

Severe, active Crohn's disease
5 mg/kg given as an intravenous infusion over a 2-hour period. Available data do not support further infliximab treatment, in patients not responding within 2 weeks to the initial infusion. In responding patients, the alternative strategies for continued treatment are:
• Maintenance: Additional infusions of 5 mg/kg at 2 and 6 weeks after the initial dose, followed by infusions every 8 weeks or
• Readministration: Infusion of 5 mg/kg if signs and symptoms of the disease recur

Fistulising, active Crohn's disease
An initial 5 mg/kg infusion given over a 2-hour period is to be followed with additional 5 mg/kg infusion doses at 2 and 6 weeks after the first infusion. If a patient does not respond after these 3 doses, no additional treatment with infliximab should be given.
In responding patients, the strategies for continued treatment are:
• Additional infusions of 5 mg/kg every 8 weeks or
• Readministration if signs and symptoms of the disease recur followed by infusions of 5 mg/kg every 8 weeks

In Crohn's disease, experience with readministration if signs and symptoms of disease recur is limited and comparative data on the benefit / risk of the alternative strategies for continued treatment are lacking.

Crohn's disease (6 to 17 years)
5 mg/kg given as an intravenous infusion over a 2-hour period followed by additional 5 mg/kg infusion doses at 2 and 6 weeks after the first infusion, then every 8 weeks thereafter. Some patients may require a shorter dosing interval to maintain clinical benefit, while for others a longer dosing interval may be sufficient. Available data do not support further infliximab treatment in paediatric patients not responding within the first 10 weeks of treatment."

3.3.1 Adverse events with anti-TNF treatment

A number of adverse events have been associated with anti-TNF therapy and have been reported for infliximab and adalimumab. As the immune response may be suppressed, infections may be more likely to occur. These include tuberculosis, other bacterial infections including sepsis and pneumonia, fungal infections and opportunistic infections such as pneumocystosis or cytomegalovirus infection. Cases of re-activation of hepatitis B infection have been observed, as have rare cases of jaundice and

hepatitis, optic neuritis and onset or exacerbation of demyelinating disorders including multiple sclerosis (MS). A deficiency of TNF may result in the initiation of an autoimmune process and the occurrence of lupus-like syndrome has been observed. There is the possibility of an increased risk of lymphoma or other malignancies, worsening of heart failure or of adverse events of the haematological system (e.g. cytopenias). Infliximab has been associated with acute, infusion related reactions (including anaphylactic shock) and delayed hypersensitivity reactions. Injection site reactions are common with adalimumab. Common adverse events for both infliximab and adalimumab are upper respiratory infections (such as sinus infections), headache, rash, nausea and stomach pains. The development of anti-TNF antibodies may be associated with a decrease in efficacy and predispose the patient to an additional risk of recurrent delayed or acute allergic reactions.[49-53]

As outlined in the licence indications, patients eligible for treatment with anti-TNF therapy are adults or children with *severe*, active (or fistulising CD) who have not responded to and/or are intolerant to conventional treatment. There is no standard definition for what constitutes *severe* Crohn's disease. NICE guidance defines *severe* as a score of >300 on the Crohn's Disease Activity Index (CDAI) or 8 to 9 on the Harvey-Bradshaw index. The group that developed the CDAI defines values of 150 and below as *quiescent* disease and values above 450 as *extremely severe* disease; no intermediate cut-off point is given for *severe* disease.[36] The NICE scope for the current appraisal states that the population of interest consists of patients with *moderate to severe* Crohn's disease; there is no standard definition of what constitutes *moderate to severe.* Trials have described patients with a CDAI of 220-400 (or 450) as having moderate to severe Crohn's disease.[54]

This report will consider the following patient groups (where information is available): adults with moderate to severe, active CD intolerant or resistant to conventional treatment, children with moderate to severe active CD intolerant or resistant to conventional treatment, and adults with fistulising CD intolerant or resistant to conventional treatment. Where possible, patients with severe (rather than moderate to severe) CD will be considered as this is in line with the licence indication.

3.3.2 Degree of diffusion

There is no up-to-date evidence available on the degree of diffusion of adalimumab and infliximab for CD treatment in the UK. The only evidence that is available from routinely collected data is for the total number of adalimumab and infliximab prescriptions for all conditions.

4. DEFINITION OF THE DECISION PROBLEM

The main aims of the report are:

- to update a previous TAR[6] on the effectiveness and cost-effectiveness of infliximab in adults with moderate to severe CD or fistulising CD who are refractory to or intolerant of conventional treatment
- to review the evidence on the clinical and cost-effectiveness of infliximab in children with moderate to severe CD who are refractory to or intolerant of conventional treatment
- to review the evidence on the clinical effectiveness and cost-effectiveness of a further anti-TNF-α antibody, adalimumab, in adults with moderate to severe CD who are refractory to or intolerant of conventional treatment
- to investigate whether there is evidence for greater clinical or cost-effectiveness for either adalimumab or infliximab

4.1 Decision Problem

4.1.1 Interventions

Adalimumab and infliximab. These drugs are for use in patients with severe active CD or fistulising active CD (infliximab), who have not responded to conventional treatment or who have experienced toxicity from these treatments. There has been a distinction made between induction treatment and maintenance treatment but it is unclear where the boundary lies between these for the interventional drugs. Similarly there has been a distinction between 'episodic' treatment, ie treatment when a disease flare starts (or at a clinician's discretion), and maintenance treatment, where patients are treated at regular (scheduled) intervals with the intention of keeping them in remission, but it is unclear where the boundary lies between these treatment strategies. It would further be useful to know the most effective dosing regimen for each of the drugs.

4.1.2 Comparators

Conventional treatment includes no treatment, dietary intervention, drug treatment with aminosalicylates, methotrexate, corticosteroids (prednisolone, budesonide and hydrocortisone), azathioprine, metronidazole or surgical intervention.

Given that licences for both drugs are for use only when conventional treatment has failed, it is unlikely that RCTs would compare the drugs to conventional treatment. Instead, the most likely comparator will be no treatment or placebo, but where patients in all trial arms continue to receive

elements of conventional therapy. Another relevant comparator may be a different dosing regimen of the same drug.

For comparisons between both drugs under review, head-to-head comparisons within the same trial would be the ideal scenario. It is important to note that, because of earlier licensing, infliximab could be viewed either as the intervention of interest in some of the RCTs or as part of conventional treatment in others. It would also be useful to establish the effectiveness of both drugs compared to non-drug treatments such as surgery or nutrition, particularly in children.

4.1.3 Population and relevant sub-groups

Infliximab is licensed for use in adults and children with *severe*, active Crohn's disease or in adults with fistulising disease who are intolerant or resistant to treatment. Adalimumab is licensed for *severe*, active Crohn's disease; current information does not indicate whether this is in adults only. There is no standard definition for what constitutes *severe* Crohn's disease. NICE guidance defines *severe* as a score of >300 on the Crohn's Disease Activity Index (CDAI) or 8 to 9 on the Harvey-Bradshaw index. The group that developed the CDAI defines values of 150 and below as quiescent disease and values above 450 as extremely severe disease; no intermediate cut-off point is given for *severe* disease.[36]

The NICE scope for the current appraisal stated that the population of interest consists of patients with '*moderate to severe'* CD. There is no standard definition of what constitutes '*moderate to severe'* but RCTs have described patients with a CDAI of 220-400 as having moderate to severe Crohn's disease.[54] Note that this assessment report is therefore investigating treatments outside their licence indications. The main thrust of the work should be to investigate the clinical effectiveness of treatments in patients with a CDAI score of 300 or more. However, it is unlikely that any RCTs have included only these CD patients. The options therefore are:

- To only look at subgroups of patients in RCTs with a CDAI score of 300 or more. This is unlikely to be a valid comparison unless the RCT stratified patients by being more or less than CDAI 300
- To widen the inclusion criteria of the assessment report to include RCTs where CD patients had lower CDAI scores[55]

It may be that there is a different effectiveness of the interventions in CD patients with CDAI scores of more than 220 compared to more than 300.

Most work on measurement of CD has been carried out in adult patients. Where a child has CD, it is unclear how this would be consistently categorised as severe CD or moderate to severe CD. Although

there is a children's version of CDAI – PCDAI it is unclear how well this measure is validated and how it relates to CDAI cut-off points.

It could be important to look at populations of patients who have failed either infliximab or adalimumab therapy to determine if unresponsiveness to a particular drug is a persistent state and whether unresponsiveness to one drug can be linked to similar unresponsiveness to the other. Finally, it is unclear exactly how resistance to treatment is measured or how long a treatment trial would go on for before a patient would be categorised as being resistant or responsive to treatment.

4.1.4 Outcomes

Key factors are the clinical effectiveness of both drugs particularly in terms of enhancing patient quality of life, maintenance of remission, delaying disease progression and prolonging survival. More specifically, outcomes could include overall survival, progression free survival, health-related quality-of-life, disease activity (remission, response, relapse, changes in disease activity indices, number of fistulas for fistulising disease), maintenance of response to treatment over time, need for surgery, need for an ostomy, hospitalisation rates, need for steroid treatment, dropout rates from TNFα treatment and adverse effects of treatment. It is unclear how outcomes such as mucosal healing would impact on clinical outcomes such as quality of life.

Where disease severity and effect of treatment is measured by CDAI or PCDAI scores it is uncertain how large a change in CDAI score constitutes a clinically significant change and whether this would be the same change for more severe CD compared to less severe CD.

Trials in patients with fistulising disease will measure fistula closure but it is uncertain whether this is a good measure of effectiveness as abscesses can form if the fistula is no longer patent so abscess occurrence may be a better outcome measure. Other clinical outcomes could include abscess formation rates and seton use (if reported).

4.2 Overall aims and objectives of assessment

The overall decision problem is 'What is the cost effectiveness of adalimumab and infliximab in the management of moderate to severe CD in the UK NHS?' Ideally, this analysis would be based on head-to-head comparisons. In the likely absence of these, this decision problem is operationalised as number of complementary cost effectiveness analyses (depending on availability of data):

- What is the expected incremental cost effectiveness ratio for infliximab therapy (induction or episodic/clinician discretion or scheduled maintenance) compared to standard care in the management of moderate to severe CD?

- What is the expected incremental cost effectiveness ratio for adalimumab therapy (induction or episodic/clinician discretion or scheduled maintenance) compared to standard care in the management of moderate to severe CD?
- What is the expected incremental cost effectiveness ratio for one dosing regimen of infliximab therapy compared to another dosing regimen of infliximab in the management of moderate to severe CD?
- What is the expected incremental cost effectiveness ratio for one dosing regimen of adalimumab therapy compared to another dosing regimen of adalimumab therapy in the management of moderate to severe CD?
- What is the expected incremental cost effectiveness ratio for (different dosing regimens of) infliximab therapy compared to (different dosing regimens of) adalimumab therapy in the management of moderate to severe CD?

5. ASSESSMENT OF CLINICAL EFFECTIVENESS

5.1 Methods for reviewing clinical effectiveness

5.1.1 Search strategy

A search was undertaken to identify existing good quality systematic reviews in order to document the evidence base to date. Searches for primary studies were restricted to RCTs. The following sources were searched for relevant primary studies:

- Bibliographic databases: Cochrane Library (CENTRAL) 2007 Issue 2; MEDLINE (Ovid) 2000 to May / June 2007; MEDLINE In-Process & Other Non-Indexed Citations (Ovid) 4 June and 26 June 2007; EMBASE (Ovid) 2000 to May / June 2007. Searches were based on index and text words that encompass the condition: Crohn's disease and the interventions: adalimumab, certolizumab pegol, infliximab and natalizumab.[a] Where it was appropriate, a methodological 'filter' was applied to identify randomised controlled trials.
- EMEA, FDA and other relevant web sites.
- Citations of relevant studies.
- Contact with experts.
- Research registries of ongoing trials including National Research Register 2007 Issue 2, Current Controlled Trials and ClinicalTrials.gov
- Submissions from industry.
- Hand search of conference abstracts in 2006 and 2007: British Society of Gastroenterology (BSG), Digestive Disease Week (DDW), United European Gastroenterology Meeting (UEGW), European Crohn's and Colitis Organisation, Federation of Clinical Immunology Societies.

Searches were not limited by language. Full search strategies can be found in Appendix 3.

[a] Natalizumab and certolizumab pegol were originally part of this technology appraisal so were included in the searches. They were subsequently dropped from the report after completion of searches.

5.1.2 Inclusion and exclusion criteria

Only studies meeting the following inclusion criteria were included:

Study Design: Randomised controlled trials (RCTs)

Population: Adults (≥ 18 years) and children (6-17 years) with moderate to severe, active Crohn's disease intolerant or resistant to conventional treatment; adults (≥ 18 years) with fistulising Crohn's disease resistant to conventional treatment.
'Moderate to severe' disease includes patients with an average CDAI score of 220 or above or those that are described by trial authors as having moderate to severe disease.

Intervention: Adalimumab *or* infliximab (any dosage/treatment regimen)

Comparator:

- Conventional treatment without TNF-α inhibitors including no treatment, placebo, dietary intervention, drug treatment with aminosalicylates, methotrexate, corticosteroids (prednisolone, budesonide and hydrocortisone), azathioprine, metronidazole or surgical intervention
- Adalimumab and infliximab compared to each other
- Different dosage or treatment regimens of the same drug

Outcomes:
At least one of the following: overall survival, progression free survival, health-related quality-of-life, disease activity (remission, response, relapse, changes in disease activity indices, number of fistulas for fistulising disease), need for surgery, hospitalisation rates, adverse effects of treatment.

Trials that looked at both induction and maintenance of remission were included.
Study designs other than RCTs were excluded.

Based on the above inclusion/exclusion criteria, study selection was made independently by two reviewers. Discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. All discrepancies were resolved in this way.

5.1.3 Data extraction strategy

Information on study characteristics, study quality and results for each trial was extracted by one reviewer and checked by a second reviewer. Four reviewers were involved in data extraction. A standardized data extraction form was used, based on the form designed for the previous TAR on infliximab.[6] The data extraction template can be found in Appendix 4. Where necessary the template was adapted to accommodate details relevant to a specific trial. Where required, information was extracted from graphs as follows (see Appendix 5): the graph was scanned into a Word document, overlayed with an appropriate template with graph gridlines, printed and enlarged to A3 size and information extracted using the gridline template. To reduce error in this procedure extracted information was checked by comparing graph readings with any available values in the report text and/or by redrawing the graph using the extracted data and comparing this with the original (see Appendix 5 for examples). A full set of completed data extraction forms is available on request. Data extraction discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. All discrepancies were resolved in this way.

5.1.4 Quality assessment strategy

Quality assessment was based on the published papers only and note was taken that absence of a quality criterion may be due to lack of reporting rather than actual poor methodological quality. Authors were not contacted for further information. Quality assessment was descriptive, a quality scoring system was not used. The quality criteria assessed were based on guidelines suggested by the Cochrane Collaboration, inviting consideration of threats arising from selection, performance, attrition and detection biases. Individual checklist items were: randomisation, concealment, blinding, comparability of groups, follow-up of trial participants, handling of missing data (intention-to-treat analysis), power calculation and selective reporting (see Appendix 4 for checklist). Study quality was assessed by one reviewer and checked by a second. Discrepancies were resolved by discussion, with involvement of a third reviewer when necessary. All discrepancies were resolved in this way.

5.1.5 Handling of manufacturer and other submissions

The main industry submissions (including appendices) were checked for additional relevant trials and additional clinical effectiveness data for included trials. Because editorial constraints meant the results available in published accounts of the trials were necessarily selective, information in the submitted Clinical Study Reports was sourced as required for purposes of balance and completeness. It was not possible to systematically review all such additional information submitted due to the volume of the submissions (e.g. more than 38,000 pages for the Clinical Study Report of ACCENT I[3,4], more than 5000 pages for the Clinical Study Report of Targan 1997[54], both included studies). No references to specific sections of the Clinical Study Reports were made in the main industry submissions. Please note that the clinical study reports for the CLASSIC, CHARM and GAIN RCTs that were received

from the manufacturers of adalimumab started on section 4 and had no page numbers or tables of contents. Also some of the appendices were missing, particularly ones referred to in the text as having all of the raw results in tables. Therefore it is unclear whether some pages are missing from the middle of these reports or not and potentially the most useful appendices were not supplied. For details on how the submitted economic models were assessed see sections 6.2 and 6.3.

5.1.6 Analysis strategy

The clinical effectiveness section of this report mainly focuses on the results from RCTs and/or RCT trial arms in which the drugs were administered within the limits their current respective licence indication (see Appendix 6). Results of trials are organised and reported in four categories:

- induction trials in adult populations predominantly or wholly constituted from non-fistulising patients;
- maintenance trials in adult populations predominantly or wholly constituted from non-fistulising CD patients;
- trials in paediatric patients; and
- trials in populations constituted wholly of patients with fistulising CD.

Results are reported within these four categories on a trial by trial basis except with regard to adverse events and side effects which were considered simultaneously across all included trials across both drugs. Most outcome results are presented in Forest plots so as to provide an overview of the quantitative spread of effect sizes. These are accompanied with brief narrative commentary. In some instances outcome results are tabulated. Both placebo and intervention rates and both rate difference and rate ratio effect sizes are presented for most outcomes in the results section of this report. The confidence intervals quoted were not adjusted for repeated measures.

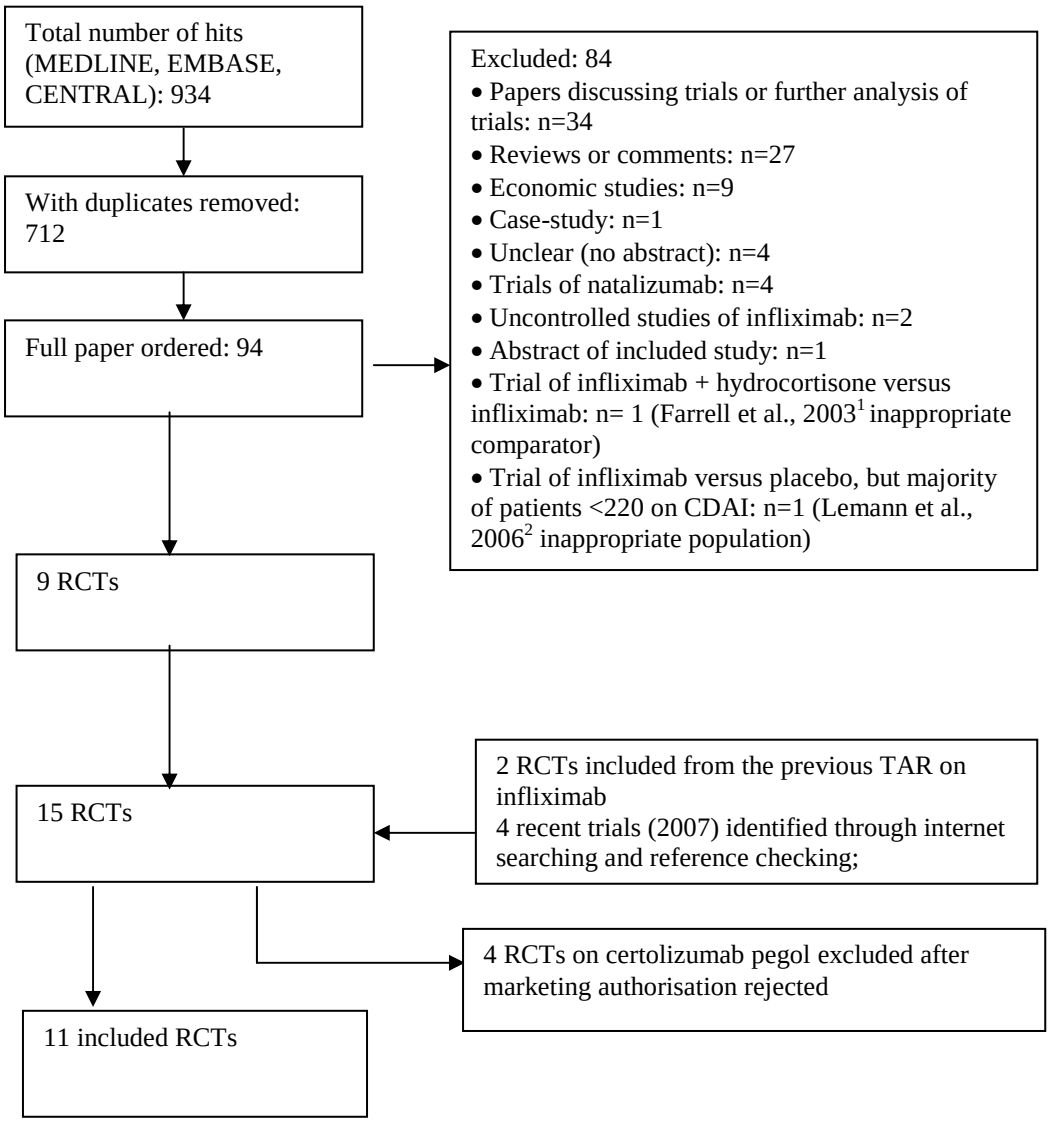
The clinical heterogeneity of trials, or the existence of only a single trial, precluded pooling of data in meta-analysis. The feasibility of undertaking indirect comparison analysis was considered in depth in order to assess the relative effectiveness of different drugs because there were no RCTs directly comparing both drugs included in this technology appraisal. However, indirect comparisons were not done because of the variation in placebo effect sizes in the RCTs (induction trials), the lack of identity in the apparently common comparator (i.e. placebo arm maintenance trials), and because of the reporting of subgroup results only at follow up (i.e. variously defined responders only) in many of the RCTs.

5.2 Results

5.2.1 Quantity of research available

Eleven relevant trials were identified, some supported by multiple publications. Figure 1 details the trial identification process.

Figure 1. Study identification process



At the time of writing of this report, 11 hardcopies of ordered publications were still outstanding or not available; none of these are likely to contain new trial data (see Appendix 7 for details of publications).

Eleven RCTs were included in total. Seven trials meeting the inclusion criteria were identified through the main database searches. Two additional studies from the previous TAR on infliximab were included,[6]as were two trials from 2007 which had been published after the search cut-off date.

Searching through the main industry submissions from both manufacturers did not yield any additional RCTs. The search for conference abstracts yielded no further relevant trials. An abstract of the study by Hommes 2005[56] was identified (Management of recent onset Crohn's Disease: a controlled, randomised trial comparing step-up and top-down therapy), which is referred to in the discussion section of this report. This study did not meet the criterion of a population of CD patients who are resistant or intolerant to conventional treatment.

The search for ongoing trials yielded four potentially relevant RCTs, all of adalimumab (see Appendix 8). All were at the recruitment stage (or not yet recruiting) at the time the information was verified by the respective manufacturers. Two were trials (induction and maintenance) of adalimumab in Japanese patients with moderate to severe CD. Two multi-centre trials of adalimumab were in patients with moderate to severe ileocolonic CD and in children with moderate to severe CD respectively. Two ongoing trials of infliximab were identified, but did not meet the inclusion criteria as they compared either infliximab with infliximab + methotrexate or infliximab with infliximab + azathioprine. No ongoing trials of head-to-head comparisons of adalimumab and infliximab were identified. No preliminary reports of any of these ongoing trials were identified in the manufacturer submissions.

5.2.2 Tabulation of included studies

All of the included RCTs recruited patients having 'moderate to severe CD' defined according to CDAI scores of between 220 and 450, or 220 and 400; it is therefore they do not reflect the intended licensed population of severe active CD (ie CDAI score of more than 300).

The included studies encompassed two trial designs, induction-therapy and maintenance-therapy, in any of three populations: adults predominantly or wholly non-fistulising, fistulising adults and children. Table 4 gives an overview of the included studies with reference to trial design and recruited patient population.

Table 4. Overview of the eleven included trials

TYPE OF TRIAL	DRUG	POPULATION		
		Wholly or predominantly non-fistulising adults	Fistulising adults	Children
Induction	Infliximab	**Targan** et al., 1997[54]†	**Present** et al., 1999[57]	**Baldassano** et al., 2003[43]
	Adalimumab	**CLASSIC I** Hanauer et al., 2006[58] **GAIN** Sandborn et al., 2007 [59]	No trials identified	No trials identified
Maintenance	Infliximab	**Rutgeerts** et al., 1999[55] **ACCENT I** Hanauer et al., 2002 [3,4]	**ACCENT II** Sands et al., 2004 [60]	**REACH** Hyams et al., 2007 [42]
	Adalimumab	**CLASSIC II** Sandborn et al., 2007 [61] **CHARM** Colombel et al., 2007[62]	No trials identified	No trials identified

† **D'Haens e**t al., 1999[63] described a subgroup of patients from **Targan** et al., 1997.[54]

Of the 11 included RCTs, nine compared infliximab or adalimumab to placebo. Two RCTs compared different doses of infliximab only and these were both in children.[42,43] Two RCTs of infliximab were in patients with fistulising disease.[57,60] Both induction and maintenance trials were identified for both drugs. All RCTs were multicentre studies conducted mainly in North America and Europe. No RCTs of head-to-head comparisons of adalimumab and infliximab were identified. No RCTs of adalimumab in children were identified. Based on the information in the published papers, all RCTs were either industry sponsored or in part industry sponsored, had participants from industry involved in study design or manuscript writing, or had one or more authors with industry involvement.

In the induction trials, patients not receiving anti-TNF treatment received short duration anti-TNF or placebo to see if a favourable clinical response was induced. In the maintenance trials, all patients received short term induction therapy with anti-TNF and then continued with longer term anti-TNF or placebo. In the maintenance trials most published results reported only the follow up of patients who initially responded to the induction therapy and results for "non-responders" were generally not provided.

The most widely reported outcomes were based on CDAI scores (see Appendix 1 for details). Although group mean or median CDAI scores were usualy recorded at various times of followup, the variance of these scores was incompletely reported and trials emphasised binary outcome measures derived by dichotomising CDAI scores. Three such binary measures were used:

- Response 70; defined as a reduction of 70 or more in CDAI score relative to baseline

- Response 100; defined as a reduction of 100 or more in CDAI score relative to baseline.
- Remission; defined as a CDAI score of less than 150

The definitions of the binary measures given above were often qualified by stipulation of additional criteria usually including no requirement for a change in concomitant medication because of worsening clinical condition and no requirement for surgery.

This section describes the results about the effectiveness of the anti-TNF interventions. The results reviewed were taken mainly from publications. When judged necessary for purposes of completeness and balance, information in the unpublished Industry Trial Reports was also sourced.

There are four sections in the clinical effectiveness results, induction treatment in adults (predominantly non-fistulising), maintenance in adults (predominantly non-fistulising), treatments in adult patients exclusively with fistulising CD and lastly, paediatric CD (18 years old or less). Within each section infliximab is reported before adalimumab and the earliest trial publication date first. Each of the four sections are organised for each trial as follows:

- Description of intervention used in the trial and other unusual points about the trial design
- Report of outcomes organised as A. Response 70, B. Response 100, C. Remission, D. Other outcomes, E. Other considerations, in the first two sections. Primary and secondary outcomes in the last two sections.
- Quality assessment
- Summary for that trial (in box)

Adverse events and side effects are considered simultaneously across all included trials for both drugs at the end of the clinical effectiveness section (see section 5.2.2.6), just before the discussion of clinical effectiveness (see section 5.2.3).

5.2.2.1 Induction trials in adult populations (wholly or predominantly non-fistulising)

Induction trials are patients who were not receiving anti-TNF therapy at the time of randomisation. Three trials were identified. One, Targan 1997,[54] compared infliximab with placebo. A further publication, D'Haens 1999[63], reported on a subgroup from Targan and so will not be further discussed. Two trials compared adalimumab to placebo (CLASSIC I[58], GAIN[59]). Apart from the subgroup study the trials recruited patients who had initial CDAI scores between 220 and 450. The outcomes reported are summarised in Table 5 and trial details are summarised in Table 6.

Table 5. Outcomes measured in induction trials with mainly non-fistulising adult populations

	% with REMISSION	% with RESPONSE 100	% with RESPONSE 70	CDAI score	IBDQ score	Other outcomes
Infliximab						
Targan1997[54]	√	X	√	√	√	CRPc
Adalimumab						
CLASSIC I[58]	√	√	√	√	√	CRPc
GAIN[59]	√	√	√	√	√	CRPc, Improvement in draining fistulas, fistula remission at week 4 (in sub-group)
CRPc = C-reactive protein concentration.						

Table 6. Main study and population characteristics: induction trials in predominantly or wholly non-fistulising adult populations

Study* Drug	Study wks N	Population: severity of CD (baseline CDAI and IBDQ if stated)	Intestinal areas affected	Main concomitant medication, % not on any medication	Previous/concomitant treatment with anti-TNF inhibitors	Intervention and comparator (dosing regimen)
Targan et al., 1997[54] Infliximab	4‡ 108	Moderate to severe, CDAI 220-450 Eligible if receiving mesalamine or oral corticosteroids or mercaptopurine or azathioprine Mean baseline CDAI (SD): 288 ± 54 placebo, 312 ± 56, 318 ± 59, 307 ± 50 infliximab groups Mean baseline IBDQ (SD): 128 ± 29 placebo, 122 ± 29, 116 ± 23, 118 ± 28 infliximab groups	Mainly ileum/colon, also colon only, some ileum only.	Aminosalicylates or corticosteroids, also mercaptopurine or azathioprine % not on medication (if any) not stated	Exclusion criterion: previous treatment with monoclonal antibodies	One 2-hour IV infusion of: 5mg/kg, or 10mg/kg, or 20mg/kg infliximab or of placebo
Hanauer et al., 2006 CLASSIC I[58] Adalimumab	4 299	Moderate to severe, CDAI 220-450 Mean baseline CDAI (SD): placebo, 296 (60); adalimumab groups 299 (57); 301 (61); 295 (52) Median baseline IBDQ (range): placebo, 131 (52-200); adalimumab groups 129 (81-218); 128 (63-200); 127 (37-192).	Mainly ileum & colon	Aminosalicylates, also corticosteroids, immunosuppressives , and few on antibiotics % not on medication (if any) not stated	Exclusion criterion: infliximab or other anti-TNF therapy	Subcutaneous infusion at weeks 0 and 2: 40mg/20mg or 80mg/40mg or 160mg/80mg adalimumab at week 0 and 2 respectively. Placebo at weeks 0 and 2
Sandborn et al., 2007 GAIN[59] Adalimumab	4 325	Moderate to severe, CDAI 220-450 Mean baseline CDAI (SD): placebo 313 (66); adalimumab 313 (58) Mean baseline IBDQ (SD): 124 (28) placebo, 120 (27) adalimumab	Mainly ileum or colon, some rectum, perianal or anus or gastro-duodenal	Corticosteroids or immunosuppressives, also oral aminosalicylates % not on medication (if any) not stated	Patients must have been treated with infliximab and either lost response or been intolerant; excluded patients with primary non-response to infliximab	Subcutaneous injections 160mg adalimumab at week 0 and 80mg at week 2 or placebo at weeks 0 & 2

* all were multi-centre studies conducted in the US, Canada and Europe and sponsored by industry. ‡ there was an open label extension beyond week 4.

Targan 1997[54] (Infliximab)

This RCT had four arms. Patients were randomised to a single intravenous infusion of placebo (N=25) or of infliximab at 5 mg/kg (N=27), at 10 mg/kg (N=28) or at 20 mg/kg (N=28). Disease status (remission, response 70 and CDAI score) was monitored at baseline and at weeks 2 and 4 after infusion. The 4 week blinded phase was followed by an open label phase with a further 12 weeks of follow up. The primary outcome measure was defined as a response 70 at week 4 with no change in any concomitant medication.

A. Response 70

Response 70 at week 4 was the primary outcome. Results for response 70 at weeks 2 and 4 are summarised in Figure 2. For response 70 at week 4 there was a statistically significant difference in favour of the infliximab groups (combined) compared to placebo ($P < 0.001$). The percentage of placebo patients achieving response 70 was 16% or less at both time points and for infliximab groups at week 4 and was between 50% and 81% depending on dose regimen. Point estimates of percentage response were associated with considerable uncertainty. The rate of response 70 at week 4 for the combined infliximab groups was 61% (95% CI: 51% to 71%). At week 4 the rate difference (infliximab – placebo) was between 0.34 and 0.65, and rate ratio (infliximab/placebo) was between 3.1 and 5.1 depending on dose. Both rate difference and rate ratio at week 4 reached statistical significance in favour of intervention.

Figure 2. Response 70 rates in Targan 1997

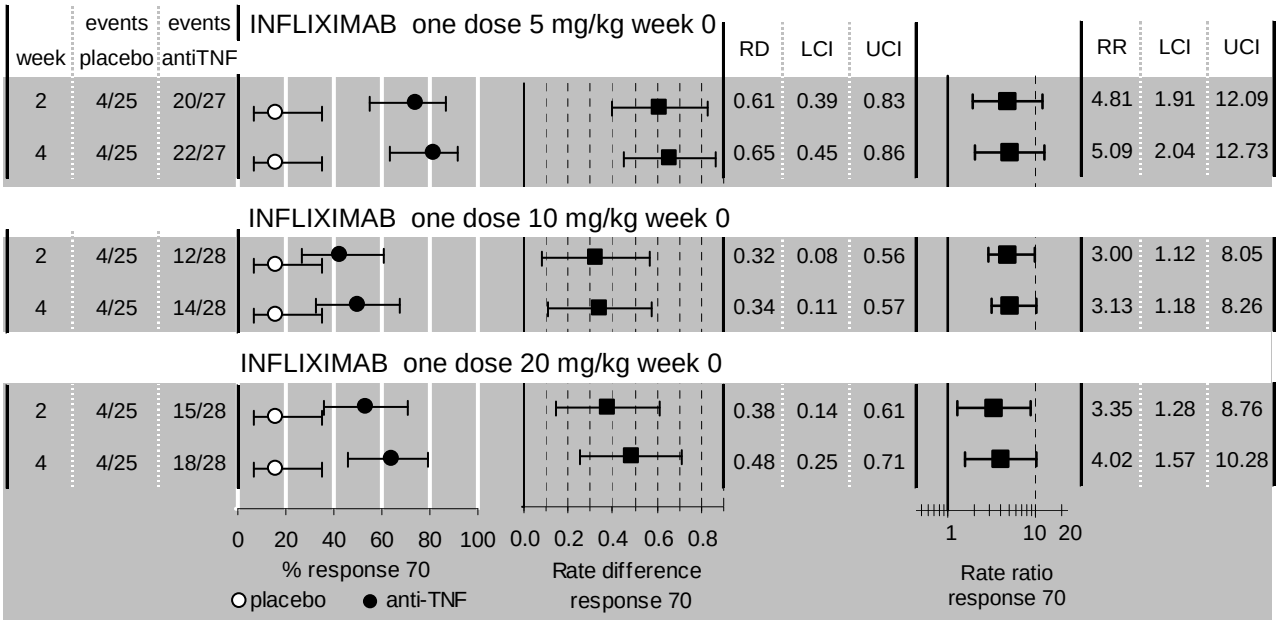


Table 7 summarises the comparison between different dose regimens for response 70 at week 4. The low dose regimen (5mg/kg) appeared more effective than the 10 mg/kg regimen (p = 0.009). The difference between dose regimens for other comparisons did not reach statistical significance.

Table 7. Rate difference between dose regimens in response 70 at week 4 in Targan

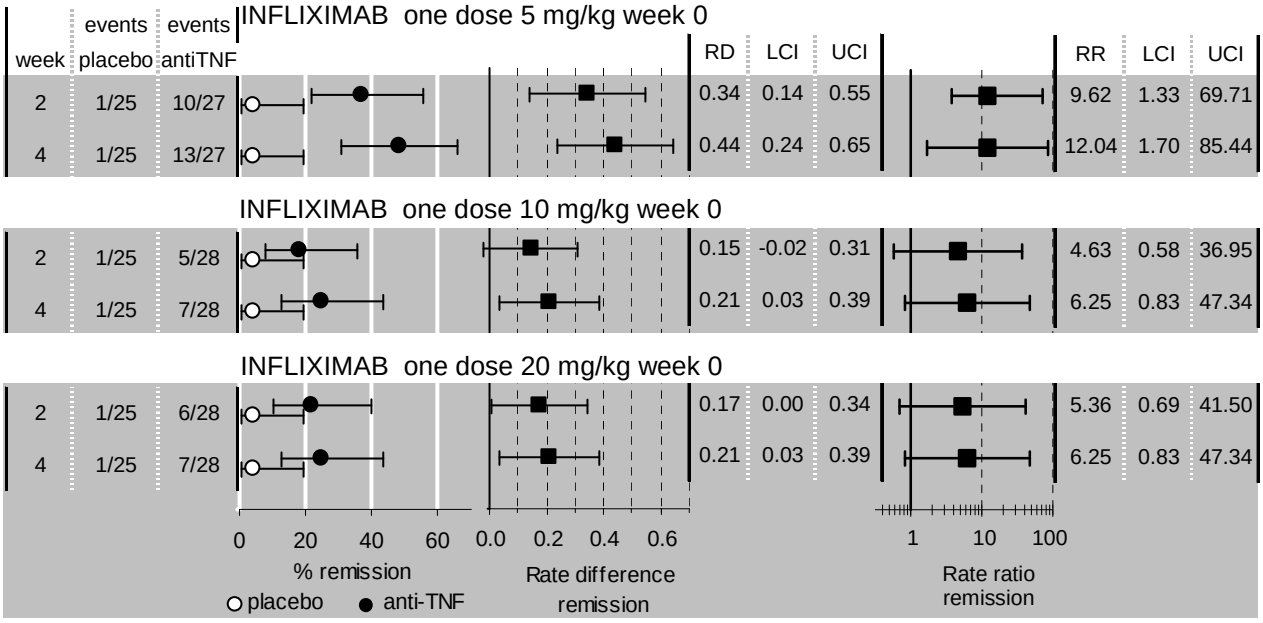
DOSE COMPARISON	RATE DIFFERENCE	*LCI*	*UCI*
5 mg/kg v 10 mg/kg	0.315	*0.079*	*0.551*
5 mg/kg v 20 mg/kg	0.172	*-0.058*	*0.402*
10 mg/kg v 20 mg/kg	-0.143	*-0.399*	*0.114*

B. Response 100 was not reported.

C. Remission

Figure 3 summarises remission rates. At four weeks between 25% and 48% of patients in the infliximab groups were in remission, depending on dose, but only one placebo patient achieved remission.

Figure 3. Remission rates in Targan 1997



At week 4 rate difference p = <0.001, 0.0206, 0.0206, for 5, 10 & 20 mg/kg dose regimens. At week 4 rate ratio p = 0.013, 0.076, 0.076 for 5, 10 & 20 mg/kg dose regimens

There was a discrepancy between remission rates published in Targan 1997[54] and rates presented in the manufacturer's submission. The latter for the 5 mg/kg group at week 4 were placebo rate 4% (1/24), infliximab rate 0% (0/24). These remission rates generate a negative risk difference

(infliximab – placebo) at week 4 (- 0.04). Confidence intervals[a] for rate ratios (infliximab/placebo) in the manufacturer's submission were described as "unadjusted", but were unexpectedly narrow compared to those calculated using standard software packages or using the standard error of *ln* (rate ratio) given by[64]: $([e_i]^{-1} + [e_p]^{-1} + [T_i]^{-1} + [T_p]^{-1})^{0.5}$ where e_i , and e_p , are the number of patients with the outcome in the intervention and placebo arms respectively, and $T_{i,}$, and T_p are total number of patients in the intervention and placebo arms respectively.

Maintenance of initial response to single infusion

At week four there were 54/83 (65%) responders (response 70) to infliximab (combined dose groups); by 12 weeks (see E. Open label phase below) there were 34 responders (41%). At week four 27/83 (33%) patients given infliximab had gained remission and at 12 weeks 20 patients (24%) were in remission.

D. Other outcomes

At week 4 favourable responses to treatment were reported for CDAI scores, for quality of life scores (IBDQ), and for CRP levels. The results reported are summarised in Table 8 .

Table 8. Mean (SD) values for CDAI, IBDQ and CRP concentrations at baseline and week 4

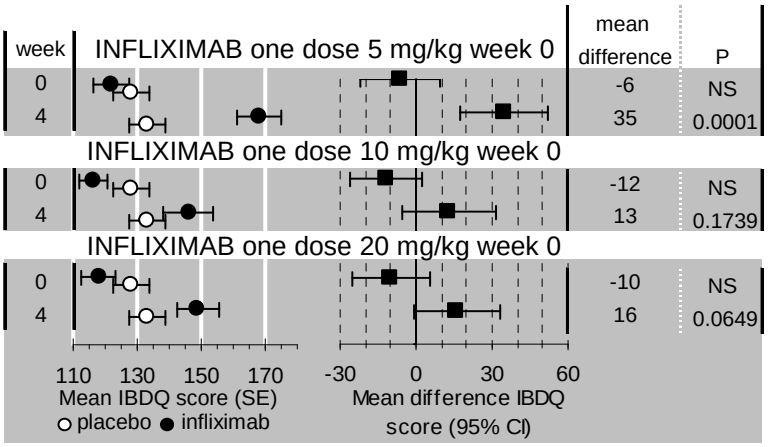
	Placebo N=25	5mg /kg N =27	10mg /kg N=28	20mg /kg N =28	All infliximab groups N =83
Score on CDAI					
Baseline	288 ± 54	312 ± 56	318 ± 59	307 ± 50	312 ± 55
4 weeks	211 ± 82	166 ± 76 [a]	226 ± 115 [b]	211 ± 107 [a]	201 ± 103 [a]
Score on IBDQ					
Baseline	128 ± 29	122 ± 29	116 ± 23	118 ± 28	118 ±27
4 weeks	133 ± 28	168 ± 36 [a]	146 ± 41 [c]	149 ± 35 [d]	154 ± 38 [e]
CRP (mg/litre) [f]					
Baseline	12.8+ 13.9	22.1 + 23.6	23.2 + 34.2	22.4 ± 23.9	22.5 ± 21.4
4 weeks	14.8 ± 18.6	5. 1 ± 9.3 [g]	12.1 ± 18.6	6.9 ± 11.6 [a]	8.3 ± 1.39 [a]

[a] P < 0.001; [b] P = 0.003; [c] P =0.02; [d] P=0.03; [e] P = 0.001; [f] Levels of CRP below 8mgIL are considered normal; [g] P= 0.004; Authors calculated P values for change from baseline comparing placebo with intervention using analysis of variance with the van der Waerden normal scores blocked according to centre. If the treatment effect was significant, the infliximab treatment groups were compared with the placebo group with linear contrasts.

Figure 4 shows the mean difference in IBDQ score (infliximab – placebo) at week 4. Mean difference reached statistical significance only for patients who received the low dose regimen.

[a] This discrepancy in confidence intervals applies to CDAI-based binary rate ratios for all trials in the infliximab industry submission.

Figure 4. Mean IBDQ scores and mean difference at baseline and week 4 of Targan



E. Other considerations – Open label phase

In the open label phase of the trial, extending by at least 12 weeks from week 4, non-responder patients at week 4 were eligible for a 10 mg/kg infusion of infliximab. The distribution of this second infusion amongst the patient groups is summarised in Table 9. Of the original 25 placebo group patients 19 non-responders received infliximab; 29 non-responder patients that had received a first dose of infliximab received the second dose. Table 9 lists the percentage of the patients (not responsive at week 4) in each group that subsequently achieved response 70 at follow up weeks 4, 8 and 12 after the second infusion.

Table 9. Numbers of patients receiving second infusion in open label phase of Targan

Original randomisation group (N)	Number receiving and not receiving second infusion (%)		Response 70 at times after second infusion (non-responders at week 4 after first infusion)		
	Did not receive	Received	Week 4	Week 8	Week 12
Placebo (25)	6 (24)	19 (76)	11/19 (58%)	13/19 (68%)	10/19 (53%)
5 mg/kg group (27)	21 (78)	6 (22)	2/6 (33%)	3/6 (50%)	1/6 (17%)
10 mg/kg group (28)	13 (46)	15 (54)	6/15 (40%)	5/15 (33%)	5/15 (33%)
20 mg/kg group (28)	20 (71)	8 (29)	2/8 (25%)	4/8 (50%)	2/8 (25%)
Combined infliximab groups					10/19 (53%)
ALL groups (108)	60 (56)	48 (44)	21/48 (44%)	25/48 (52%)	18/48 (37%)

Of patients unresponsive to the first dose of infliximab 28% (8/29) responded by week 12 following the second dose, compared to 53% (10/19) of patients whose second infusion was their first exposure to active intervention. During this open label phase there was a lack of a true placebo control group and the results therefore only suggest that some patients poorly responsive to an initial infusion may respond subsequently on receipt of further infusion. Whether a 10 mg/kg second dose represents the most appropriate dose regimen for this second-dose strategy is unknown.

Quality Assessment (based on published report)

Randomisation, allocation concealment, and blinding (up to week 4) were all adequate. Baseline characteristics were similar between groups except for CRP levels and for the proportion of patients with ileal involvement. Placebo CRP level (mean 12.8, SD 13.9) was substantially lower than that for the active intervention groups (mean (SD): 22.1 (23.6), 23.2 (34.2), and 22.4 (23.9) for 5 mg/kg, 10 mg/kg, and 20 mg/kg groups respectively). The potential impact on results of the imbalanced CRP levels is difficult to determine. Follow up appeared almost complete. The original study protocol did not specify the use of intention-to treat analysis, but the publication stated that patients were analysed according to assignment. A power calculation was conducted; this assumed a 30% response in the placebo group presumably reflecting the authors' assessment of placebo rates reported in other CD trials. The actual placebo response rate observed was less than half this value (16%) and was low compared to other similar trials. The low placebo rate and imbalance of placebo CRP level may indicate an atypical placebo population possibly stemming from the small sample size of the group (N=25).

Targan 1997. Summary of effectiveness evidence.

A single IV infusion of infliximab (5, 10 or 20 mg/kg) was more effective than placebo at delivering a clinical response (a reduction of ≥ 70 points in CDAI score) at week 4 of follow up ($p < 0.005$ for rate differences and $p < 0.022$ for rate ratios). Estimates of the percentage of patients responding to infliximab were associated with considerable uncertainty and at four weeks ranged between 50% and 80% depending on dose. Of the dose regimens used, the lowest appeared to be the most effective, suggesting the possibility that the most appropriate dose could be less than the lowest used in the trial (5 mg/kg). A proportion of patients (~30%) not responsive at week 4 did respond subsequently when given a second dose of infliximab (10 mg/kg); although it is likely this "second-dose" response required active intervention, this was not properly demonstrated because the trial lacked a true placebo comparator after week 4. The most effective dose regimen for a "second-dose" response was uncertain. After week 4 nearly all trial participants had received active intervention and inferences about the relation of outcomes to infliximab were obscured. The Targan trial was completed more than a decade ago and no further induction trial of infliximab in this population has been conducted so the uncertainties described above remain to be addressed.

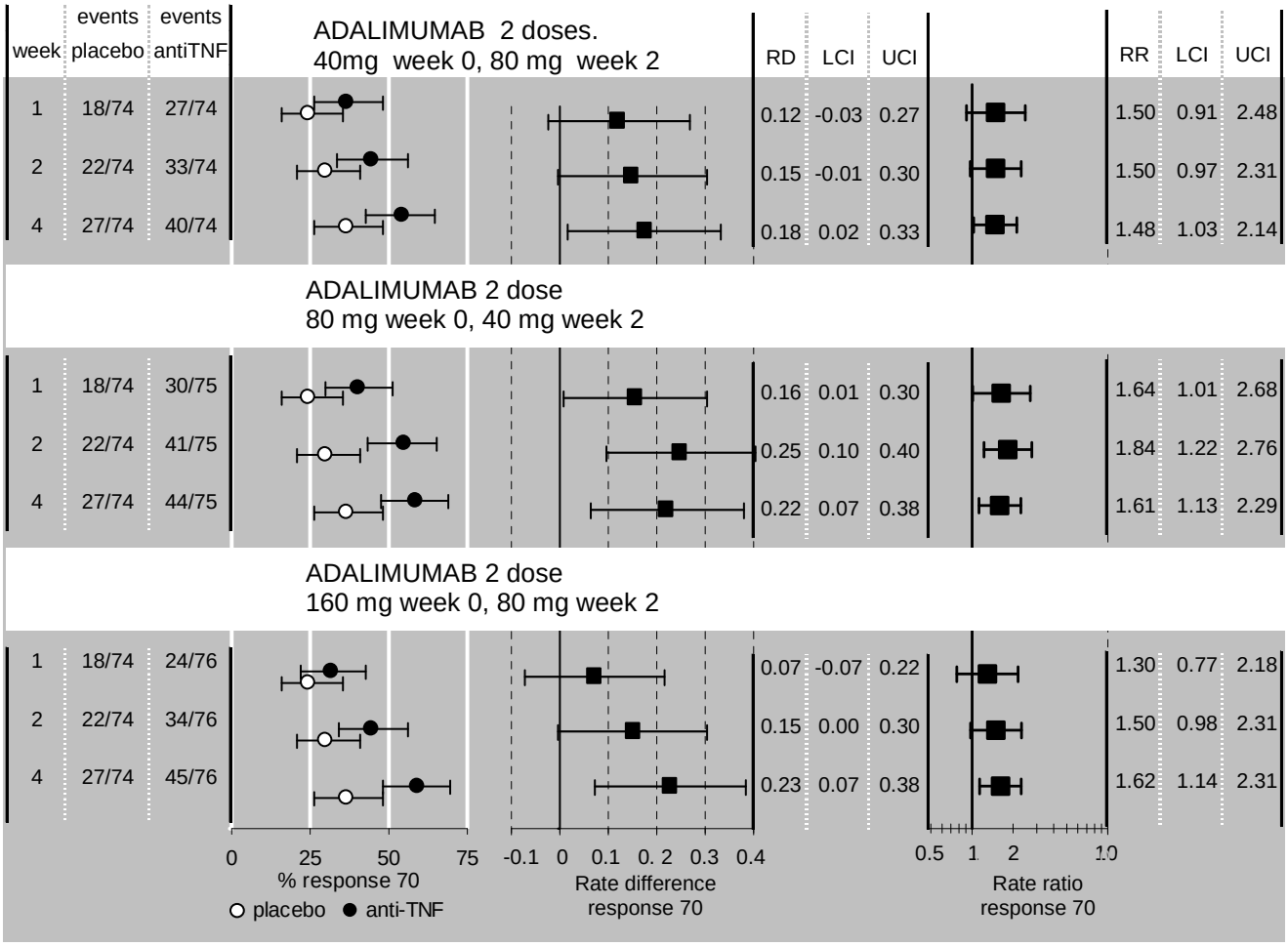
CLASSIC I[58] (adalimumab)

Patients (N=299) were randomised to two subcutaneous injections 2 weeks apart of either placebo (N=74), or of adalimumab at dose regimens of 40mg then 20mg (N=74), at 80mg then 40mg (N=75), or at 160mg then 80mg (N=76). Patients were excluded if they had previously received any anti-TNF treatment. At baseline 11% of patients had fistulas. Outcomes were monitored at weeks 1, 2 and 4 after the first injection. The primary outcome was defined as the proportion of patients in remission at week 4 in the two high dose adalimumab groups versus the placebo group (tested using chi squared test).

A. Response 70

At week 4 for the less robust measure of a clinical improvement by > 70 points in CDAI score from baseline (response 70) a statistically significant result was observed for both rate difference and rate ratio for all three dose regimens (results are summarised in Figure 5).

Figure 5. Rates of response 70 in CLASSIC I

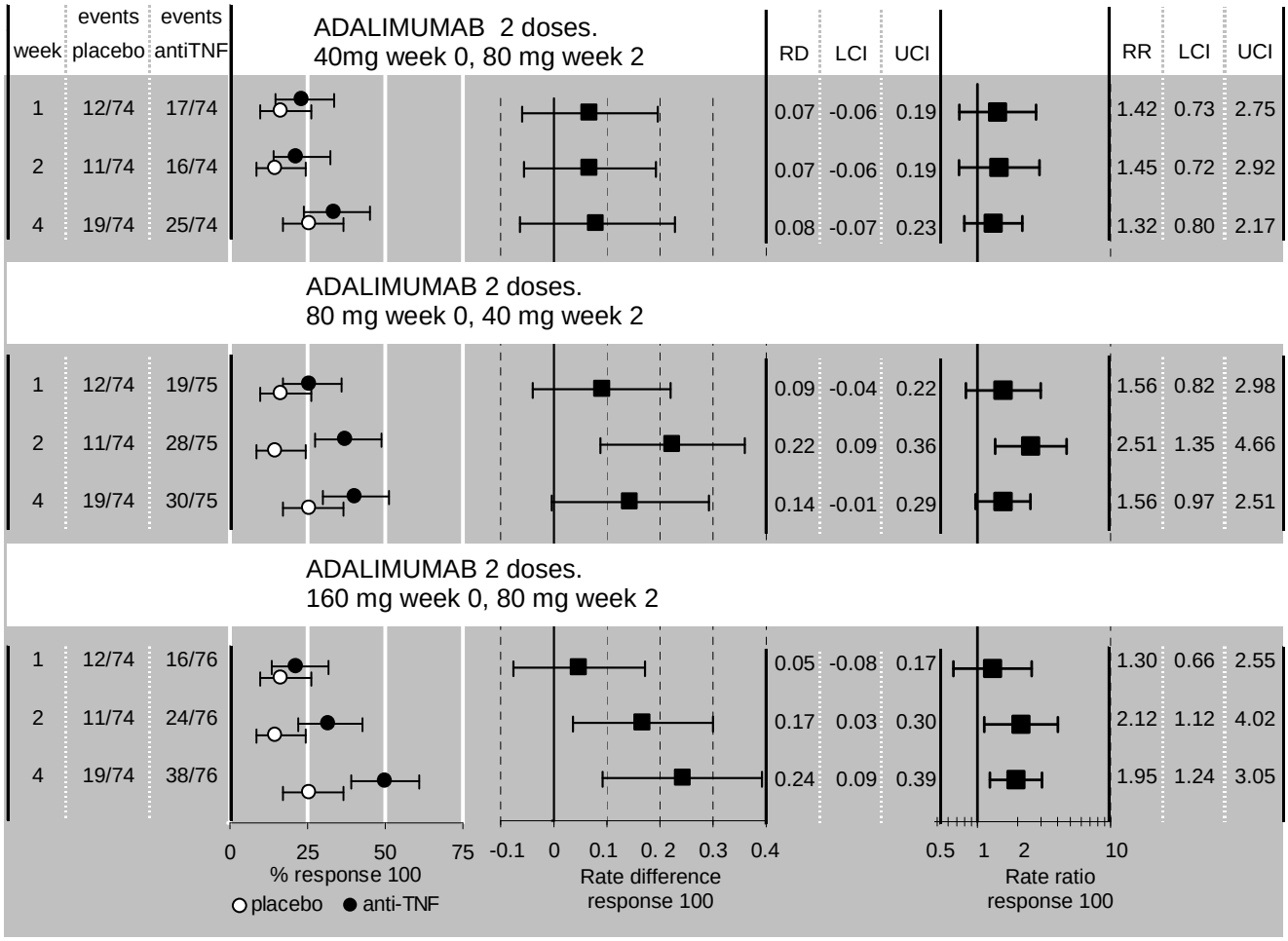


At week 4 for rate difference P = 0.029, 0.005, 0.004 for 40/20, 80/40 and 160/80 dose regimens, At week 4 for rate ratio P = 0.0357, 0.0088, 0.0073 for 40/20, 80/40 and 160/80 dose regimens

B. Response 100

At week 4 the rate difference for response 100 (intervention – placebo) only reached statistical significance for the highest dose regimen while rate ratio (intervention / placebo) reached statistical significance for the two higher dose regimen groups. The results for response 100 are summarised in Figure 6.

Figure 6. Rates of response 100 in CLASSIC I.

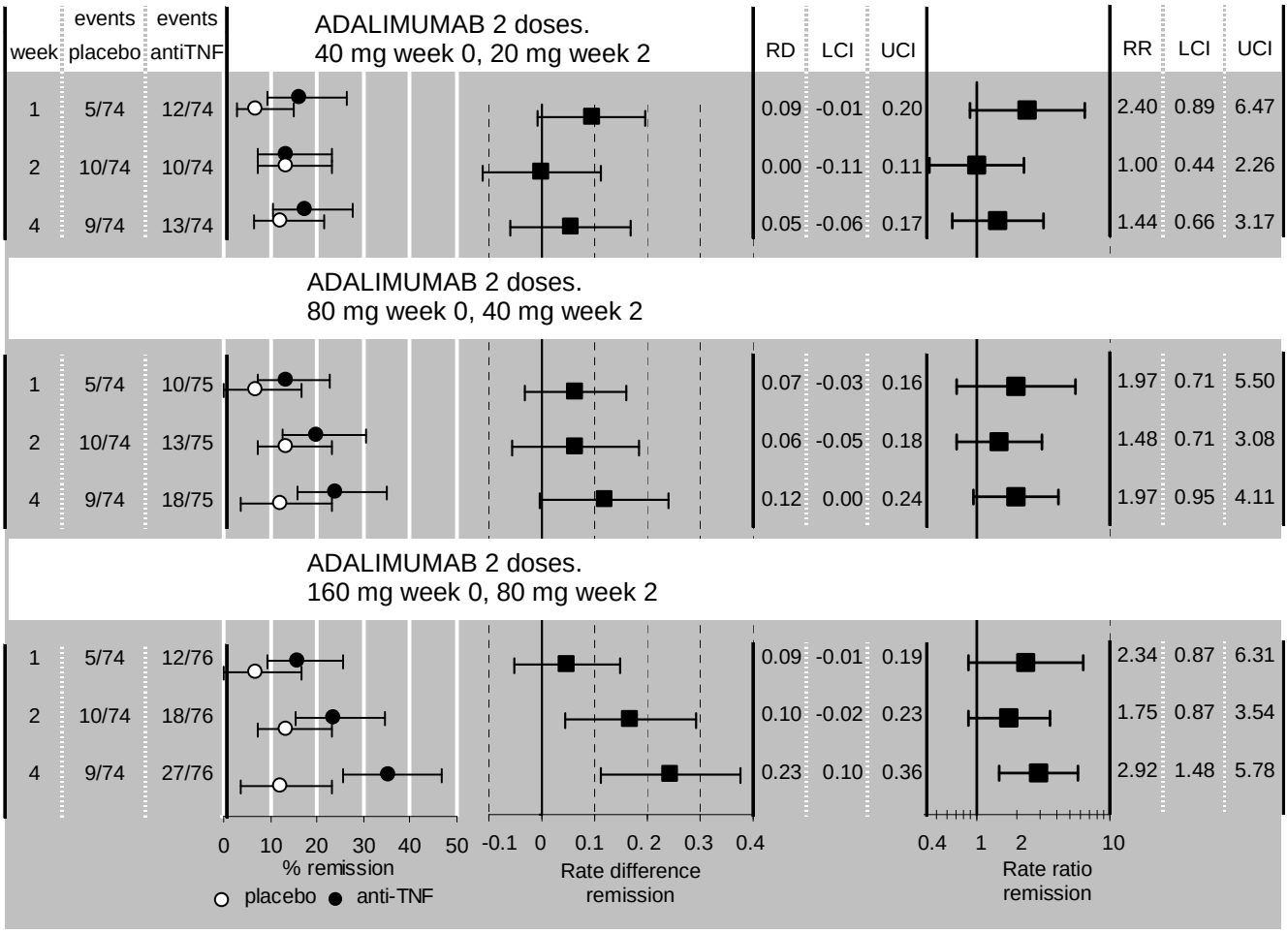


At week 4 for rate difference P = 0.279, 0.060, 0.0015 for 40/20, 80/40 and 160/80 dose regimens, At week 4 for rate ratio P = 0.284, 0.0682, 0.0036 for 40/20, 80/40 and 160/80 dose regimens.

C. Remission rates

Remission rates were the primary outcome in this RCT. For remission rates there was a statistically significant difference in favour of the two high dose adalimumab regimens relative to placebo for the proportion of patients in remission at (45/151 versus 9/74; P = 0.004). At week 4 the rate difference (intervention – placebo) and rate ratio (intervention / placebo) only reached statistical significance in the highest dose regimen group. Remission rates are summarised in Figure 7.

Figure 7. CLASSIC I remission rates



At week 4 rate difference P = 0.354, 0.057, 0.0005, for 40/20, 80/40 and 160/80 dose regimens. At week 4 rate ratio P = 0.359, 0.0691, 0.0021 for 40/20, 80/40 and 160/80 dose regimens

For each of the three CDAI-based binary outcome measures there was an apparent linear dose response trend with greater effectiveness for higher dose.

D. Other outcomes

At week 4 favourable responses to treatment were reported for CDAI scores, for Quality of life scores (IBDQ), and for CRP levels. The results reported are summarised in Table 10 .

Table 10. Mean (SD) values for CDAI, IBDQ and CRP concentrations at baseline and week 4

	Placebo N=74	40/20 N =74	80/40 N=75	160/80 N =76
Score on CDAI: mean (SD)				
Baseline	296 (60)	299 (57)	301 (61)	295 (52)
4 weeks	240 (NR)	228 (NR)	210 (NR) [a]	193 (NR) [b]
Score on IBDQ median and range				
Baseline	131 (52 – 200)	129 (81 – 218)	128 (63 – 200)	127 (37 – 192)
4 weeks	147 (NR)	147 (NR)	158 (NR) [c]	158 (NR) [c]
CRP (mg/litre) median (range)[d]				
Baseline	0.9 (0 – 17.3)	0.9 (0 – 11.3)	0.9 (0 – 14.9)	0.7 (0 – 9.3)
4 weeks	0.8 (0 – 9.3)	0.3 (0 – 8.6) [e]	0.4 (0 – 34.0) [f]	0.2 (0 – 4.6) [g]

Comparisons versus placebo: [a] P < 0.01; [b] P = < 0.001; [c] P = <0.05; [d] Levels of CRP below 8mgIL are considered normal; [e] P=0.032; [f] P = 0.0002 [g] P= 0.0001; NR = not reported.

E. Other considerations - Subgroup analyses

Logistic regression failed to show a relationship between baseline CRP levels or concomitant immunosuppressive therapy on the one hand and difference between placebo and adalimumab remission rates at week 4.

For the small subgroup of patients with fistulas (11%) no significant differences were observed between placebo and intervention with regard to fistula improvement or remission.

Quality Assessment (based on published report)

Randomisation, allocation concealment, and blinding were adequate. Baseline characteristics were reasonably well balanced between groups. There were no losses to follow up and withdrawals were limited to 5%. Efficacy estimates appear to have been calculated using ITT analysis but this was not stated explicitly. A power calculation was conducted; this assumed 20% and 45% remission rates in the placebo and intervention arms respectively (the observed placebo rate in the trial was about 12%). Last observation carried forward was used for analysis of IBDQ scores but the amount of missing data was not stated.

CLASSIC I. Summary of effectiveness evidence.

Two subcutaneous injections of adalimumab given two weeks apart at 40mg then 20mg, or at 80mg then 40mg, or at 160mg then 80mg, were more effective than placebo at achieving remission (CDAI score < 150) at week 4 after the first injection (p = 0.004 for the two high dose regimens combined versus placebo). The percentage of placebo treated patients gaining remission at week 4 was ~12% compared to between ~18% and ~36% for adalimumab treated patients depending on dose regimen received. Point estimates of response 70 rates, response 100 rates and remission rates were associated with considerable uncertainty but for all three outcome measures a trend was evident for higher doses to be more effective. At week 4 of follow up rate differences (intervention – placebo) and rate ratios (intervention / placebo) for the highest dose regimen reached statistical significance in favour of adalimumab for all three outcomes. Subgroup analyses failed to identify any baseline characteristics associated with a better response to active intervention relative to placebo.

GAIN[59] (adalimumab)

In this trial 325 patients were randomised to two subcutaneous injections 2 weeks apart of either placebo (N=166), or of adalimumab at a dose regimen of 160mg then 80mg (N=159). To be included patients had to have been previously exposed to infliximab treatment and found to be intolerant (N=190) or unresponsive (N=164) or intolerant and unresponsive (N=40). The primary response was defined as the proportion of patients in remission at week 4 after the first injection.

A. Response 70, B. Response 100 and C. Remission.

The primary outcome was remission rates. The remission rate at week 4 was 7% in the placebo group and 21% in the adalimumab group (P < 0.001). This result and those for the secondary outcomes as reported are summarised in Table 11. The CDAI-based binary response outcome measures reported are summarised graphically in Figure 8. At weeks two and four rate differences (adalimumab – placebo) and rate ratios (adalimumab/placebo) were in favour of the intervention and reached statistical significance.

D. Other outcomes

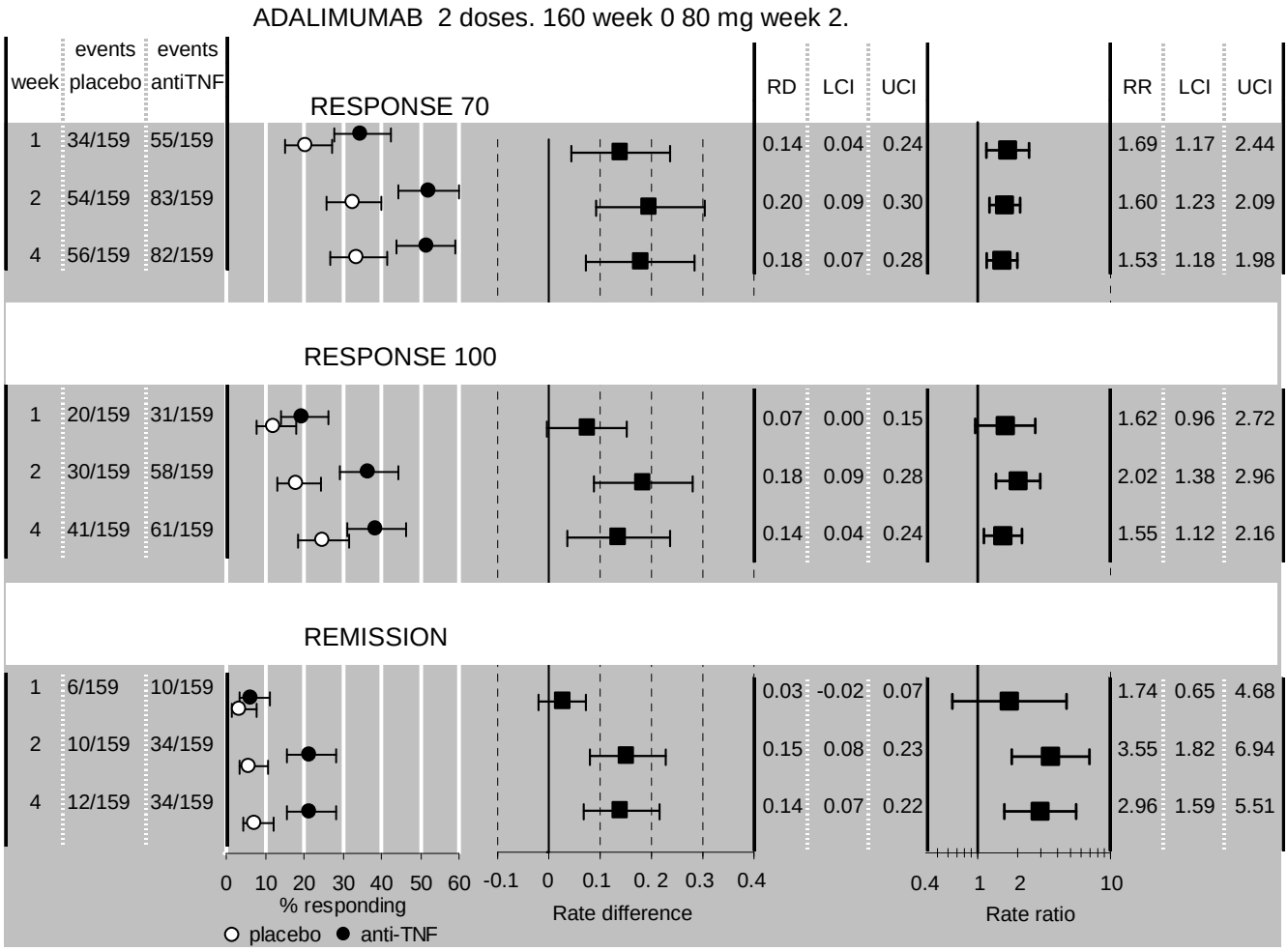
Results for these are also shown in Table 11. Mean CDAI scores reduced from baseline to a greater extent with adalimumab than with placebo (at week 4, P < 0.001 for mean change from baseline). At week 4 the improvements from baseline in IBDQ scores were 30 and 15 for the adalimumab the placebo groups respectively. CRP levels at week 4 relative to baseline were more normalised in the intervention than placebo group. The change from baseline comparing adalimumab to placebo reached statistical significance in favour of adalimumab.

Table 11. Outcome measures reported in the GAIN trial. [redacted]

	Placebo N=159	Adalimumab 160/80 N =164	Difference (95% CI) (adalimumab - placebo)	P [a]
Remission (rate; %) [b]				
Week 1	4%	6%	2.7% (-2.0 to 7.4)	[redacted]
Week 2	6%	21%	14.7% (7.2 to 22)	[redacted]
Week 4	7%	21%	14.2% (6.7 to 21.6)	[redacted] [c]
Response 70 (rate; %)				
Week 1	21%	35%	14.1% (4.5 to 23.7)	0.004
Week 2	33%	52%	19.7% (9.1 to 30.1)	[redacted]
Week 4	34%	52%	17.8% (7.3 to 28.4)	[redacted]
Response 100 (rate; %)				
Week 1	12%	20%	7.4% (-0.5 to 15.4)	[redacted]
Week 2	18%	37%	18.4% (8.9 to 27.9)	[redacted]
Week 4	25%	38%	13.7% (3.7 to 23.7)	[redacted]
CDAI: mean (SD)				
Baseline	313 (66)	313 (58)	0	
Week 1	287 (NR)	264 (NR)	-23	
Week 2	281 (NR)	232 (NR)	-49	
Week 4	264 (NR)	226 (NR)	-38	
IBDQ score: mean (SD)				
Baseline	124 (28)	120 (27)	+4	
Week 4	139 (NR)	150 (NR)	+11	<0.001
CRP: median (range) mg/L [d]				
Baseline	7.0 (0 – 235)	9.0 (0 – 115)	+2	
Week 4	7.0	5.0	-2	
Change from baseline	0	4	4	Significant

[a] Comparisons adalimumab versus placebo: [b] primary outcome % remission at week 4. [c] chi squared test. [d] Levels of CRP below 8mgIL are considered normal

Figure 8. Response 70, response 100 and remission rates reported in GAIN



At week 4 rate difference P = 0.001, 0.007, and 0.0002, for response 70, response 100 and remission respectively. At week 4 rate ratio P = 0.0014, 0.009, and 0.0006 for response 70, response 100 and remission respectively.

E. Other considerations - Subgroup analyses

The primary outcome (remission at week 4) was reported for subgroups of patients defined according to: previous response or intolerance to infliximab; receiving or not receiving immunosuppressive agents at baseline; receiving or not receiving corticosteroids at baseline; having a negative or positive test for antibodies to inflixmab. Rate difference was in favour of adalimumab relative to placebo for all subgroups.

A small proportion of patients (14%, N=45) had draining fistulas or perianal fistulas at baseline. Rates of fistula improvement and remission were similar between placebo and adalimumab groups.

Quality Assessment (based on published report)

Randomisation, allocation concealment, and blinding were adequate. Baseline characteristics were well balanced between groups. There were no losses to follow up and withdrawals were limited to

4%. Efficacy estimates appear to have been calculated using ITT analysis for remission and response outcomes. For continuous variables such as IBDQ last observation was carried forward; the amount of missing data for IBDQ was small (eight patients). A power calculation was conducted; this assumed 20% and 35% remission rates in the placebo and intervention arms respectively (the observed rates at week 4 in the trial were 7% and 21% respectively).

GAIN. Summary of effectiveness evidence.

Two subcutaneous injections of 160mg and then 80mg of adalimumab given two weeks apart were more effective than injections of placebo at achieving remission (CDAI score < 150) at week 4 after the first injection (P <0.001). The % of placebo treated patients gaining remission at week 4 was 7% (95% CI: 4% to 12%) compared to 21% (95% CI: 14% to 27%) for adalimumab treated patients. At weeks 2 and 4 of follow up rate differences (intervention – placebo) and rate ratios (intervention / placebo) reached statistical significance in favour of adalimumab for remission, response 70 and response 100. A statistically significant difference in favour of adalimumab versus placebo was observed for change in IBDQ score at week 4 relative to baseline.

Pooling and indirect comparison.

The two adalimumab trials differed with respect to their populations: CLASSIC I excluded patients if they had previously received any anti-TNF treatment while the GAIN trial recruited only patients who had previously experienced infliximab treatment but had proved intolerant or unresponsive; because of these clear population differences pooling results from the two trials was not undertaken. The existence of only a single induction trial for infliximab in this population precluded pooling.

No head to head induction trial of infliximab versus adalimumab has been conducted. A possible approach to compare effectiveness of the two drugs is by indirect comparison using trials with a "common" comparator (e.g. placebo). The Targan population, in contrast to that in GAIN was naïve to anti-TNF therapy and therefore indirect comparison between these trials was not judged productive. The placebo rates for remission and response 70 in Targan were low compared to those in the adalimumab trials and are indicative of likely differences between the potentially "common" comparator groups possibly stemming from the very small sample size of the placebo group in the Targan trial. Because of the likely difference in target placebo populations indirect comparison was judged more likely to be misleading than informative. It is relevant that neither industry submission undertook an indirect comparison between these induction trials. One way clinical heterogeneity may be expressed is in different response rates in placebo groups. Although CDAI scores at baseline may be similar between trials this could mask considerable clinical heterogeneity because CDAI is a summary score and patients can achieve the same score yet may have problems with quite different

aspects of their disease.

5.2.2.2 Maintenance trials in adults (wholly or predominantly non-fistulising)

These are trials in which all patients receive short term induction therapy with anti-TNF and then procede to longer term treatment with either placebo or anti-TNF. The predominant aim of these trials was to investigate whether anti-TNF was superior to placebo in maintaining any favourable clinical response observed from induction therapy. Since no true placebo comparator existed during the induction therapy it is not possible to determine how much of the favourable clinical response seen from induction was actually attributable to active intervention. This complicates interpretation of results.

Four trials were identified, two with infliximab (Rutgeerts et al., 1999 [55] and ACCENT I (Hanauer et al., 2002 & Rutgeerts 2004,[3,4]) and two with adalimumab (CLASSIC II (Sandborn et al., 2007[61]) and CHARM (Colombel et al., 2007[62]). The main study characteristics are summarised in Table 12. These studies were characterised by distinct differences in induction regimens.

The Rutgeerts 1999 trial was an extension of the Targan 1997 infliximab induction trial. Patients eligible had received variably 1 or 2 previous infusions of placebo or of infliximab at doses of 5, 10 or 20 mg/kg. Patients with a response 70 were then eligible for the trial. The induction regimen of participants in this trial was variable and not clearly defined, making it difficult to identify the precise target population involved.

Similarly to Rutgeerts 1999 the CLASSIC II trial was an extension of a previously conducted induction trial, namely the CLASSIC I study of adalimumab. Patients eligible for CLASSIC II were required to be in remission (CDAI < 150) at week 4 of CLASSIC I and also 4 weeks later. These patients may have received two subcutaneous injections two weeks apart of various doses of adalimumab (40 mg then 20 mg, or 80 mg then 40 mg or 160 mg then 80 mg) or of placebo.

The ACCENT I (infliximab) and CHARM (adalimumab) trials were free-standing maintenance trials with more straight forward induction regimens. In ACCENT I patients received a single induction infusion of 5 mg/kg of infliximab. In CHARM patients received subcutaneous induction injections of 160 mg of adalimumab and of 80 mg adalimumab 2 weeks apart.

The main study and population characteristics are shown in Table 12. The main outcome measures described in the published reports of the four trials are summarised in Table 13 .

Table 12. Main study and population characteristics: maintenance trials in adults predominantly or wholly non-fistulising

Study‡ Drug	weeks N	Population: severity of CD Baseline CDAI and IBDQ if stated	Areas affected	Main concomitant medication. *	Previous anti-TNF therapy	Intervention and comparator (dosing regimen)
Rutgeerts et al., 1999 [55] ** Infliximab	48 73	Moderate to severe, CDAI 220-400 "treatment resistant". Median CDAI: Placebo 305; infliximab 310 Median IBDQ: Placebo 121; infliximab 111	Mainly ileum and colon or colon only, some ileum only	Corticosteroids or immunosuppressive agents "allowed", Non-responders to aminosalicylates "eligible".	Excluded if had received monoclonal antibodies prior to Targan study	Variable treatment with infliximab or placebo in previous RCT then re-randomisation to placebo or infliximab (10mg/kg or 20 mg/kg IV) at 8 week intervals.
Hanauer et al., 2002 & Rutgeerts 2004 ACCENT I[3,4] Infliximab	54 573	Moderate to severe, CDAI 220-450 CDAI median *IQR*: Placebo 292 *256-341*; infliximab 303 *268-346* & 297 *256-346* IBDQ Median *IQR:* Placebo 126 *110-144*; infliximab 126 *109-146* & 131 *109-152*	Mainly ileum/colon, also colon only or ileum only; some gastro-duodenum	Corticosteroids, immunosuppressives, oral aminosalicylates	Excluded if previously treatmed with any anti-TNF agent.	All receive 5 mg/kg infliximab IV; then 7 additional infusions (week 2, 6, then every 8 weeks) of either placebo or infliximab (5mg/kg or 10mg/kg) (NB both infliximab groups received 5 mg/kg at weeks 2 & 6)
Sandborn et al., 2007[61] CLASSIC II[61] *** Adalimumab	56 55	All patients in remission week 0 (week 4 CLASSIC I[58]) Baseline corresponds to CLASSIC I[58] week 4 CDAI mean *SD*: Placebo 107 *62*; adalimumab 106 *33* & 88 *50* IBDQ median *range*: Placebo 191 *138-224*; adalimumab 188 *128-213* & 200 *138-216*	No further details As CLASSIC I[58]	Mainly oral aminosalicylates or corticosteroids, some immunosuppressive agents	Unclear if all previously received adalimumab or if patients in remission after placebo were included.	Subcutaneous infusion 40mg adalimumab from weeks 4-55, weekly or every other week (eow). Not stated if placebo weekly or eow.
Colombel et al., 2007[62] CHARM[62] Adalimumab	56 778	Moderately to severely active CD, CDAI 220-450 CDAI mean *SD* ¥: 313.1 *62.0* IBDQ median *range* ¥: 122.0 *44-205* (¥whole group, includes patients who withdrew before randomisation)	Mainly ileum or colon, few gastr-oduodenal or other (not stated)	Corticosteroids, immunosuppressive agents, oral aminosalicylates	424 (49.6%) previously exposed to anti-TNF (must <u>not</u> have exhibited an initial non-response)	All received adalimumab 80 mg subcutaneously, then 40 mg at week 2; randomisation at week 4, then 40 mg adalimumab, weekly or eow. Not stated if placebo weekly or eow.

‡ all were industry sponsored multi-centre studies mostly conducted in US, Canada and Europe; in CHARM centres in Australia and S. Africa also participated. * % 'not on any medication' was not stated in any study. ** An extension of the Targan 1997 trial. *** An extension of the CLASSIC I trial.

Table 13. Outcomes measured in maintenance trials with mainly non-fistulising adult populations

	% of patients in remission (CDAI score <150	% achieving 100-point response on CDAI	% achieving 70- point response on CDAI	CDAI score (mean or median)	IBDQ score (mean or median)	Additional outcomes
Infliximab						
Rutgeerts 1999[55]	√	X	√	√	√	Median CRPc. Time to loss of response.
ACCENT I[3,4]	√	X	√	√	√	Patients with CD related intra-abdominal surgery; CD related hospitalisations; patients discontinuing and remaining free from corticosteroids; mucosal healing (sub-group)
Adalimumab						
CLASSIC II[61]	√	√	√	X	√	Median CRPc, % of patients discontinuing steroids without loss of remission
CHARM[62]	√	√	√	√	√	% patients in remission at week who were also in remission at week 56; median t in remission; corticosteroid free remission; fistula response

Rutgeerts 1999 [55] (infliximab)

The Rutgeerts 1999 trial was an extension of the Targan 1997 infliximab induction trial and included 73 of the original 108 patients. Targan consisted of a 4-week comparison between placebo and one dose of infliximab in three arms (5mg/kg, 10 mg/kg or 20 mg/kg). This was followed after a maximum of two weeks by an open label phase with 12 weeks of follow up that started with the option of a 10 mg/kg dose of infliximab for week-4-non-responder patients. To be eligible to enrol in Rutgeerts 1999 the Targan week-4-responder patients needed to achieve a response 70 at week 8, and the week-4-non-responder patients needed to achieve a response 70 at week 8 after the open label option of a 10 mg/kg infusion of infliximab. Four weeks after qualifying (week 8 after induction infliximab or 8 weeks after open label infliximab the eligible patients were randomised to IV infusion of placebo or 10 mg/kg of infliximab (designated week 12 of maintenance phase) and a further three infusions at 8 week intervals (a total of 4 infusions after becoming eligible to participate; administered weeks 12, 20, 28, 36). Follow up continued to week 48.

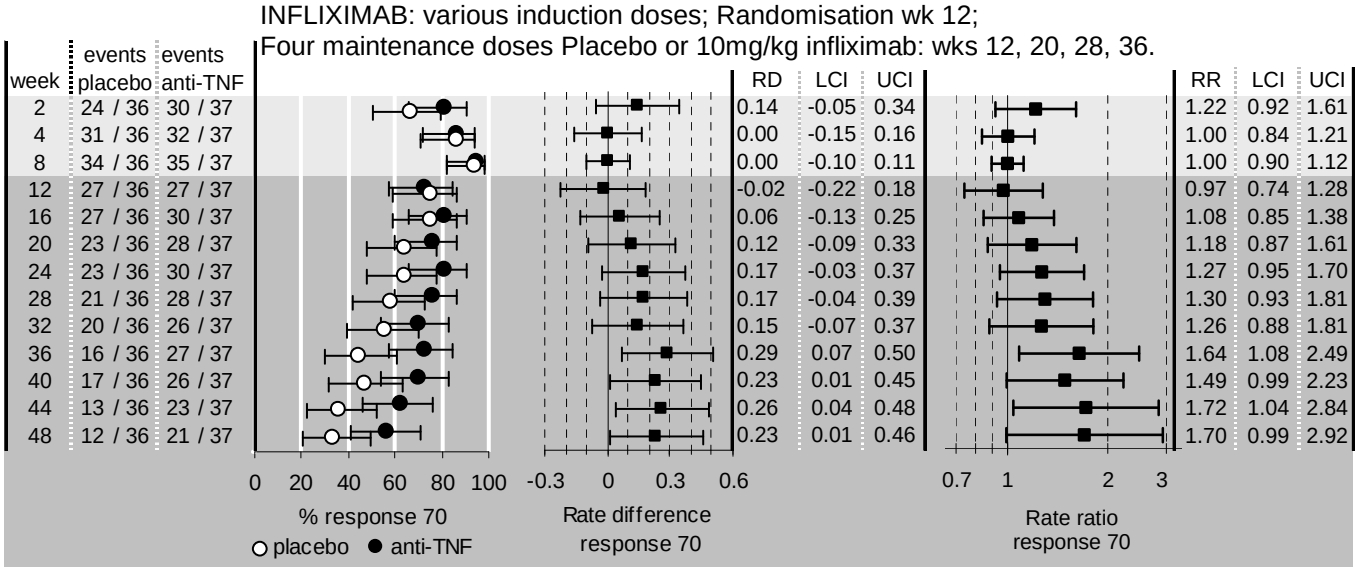
The induction regimen in this study was variable between patients in duration and in exposure to infliximab. In consequence, induction was ill-defined and the distinction between the

induction regimen and maintenance regimen was also unclear. The eligible patients could have received any of the following possible infusions of infliximab: one 5 mg/kg, one 10 mg/kg, one 20 mg/kg, one 5 mg/kg and one 10 mg/kg, two 10 mg/kg, one 20 mg/kg and one 10 mg/kg, or no infliximab if they had received an induction dose of placebo and achieved a response 70 at week 8 (N=4). How closely the trial induction phase corresponds to the licence indication is uncertain.

A. Response 70

No primary outcome measure was identified. The response 70 results presented (summarised in Figure 9) referred to point prevalence at assessment time points and do not necessarily indicate maintenance of individual patient response. At week 8 more than 90% of patients had a response 70 (CDAI reduced by > 70 points relative to baseline in Targan 1997). At week 12 (randomisation week) this had diminished to about 75% and by week 48 had further diminished to 33% in the placebo group and 57% in the infliximab group (p = 0.038 for rate difference and p = 0.054 for rate ratio). Point estimates were associated with considerable uncertainty. The authors stated that of patients with response 70 at the last infusion (week 36) 62% of the infliximab group and 37% of the placebo group maintained their response for the 8 weeks to week 44 (p = 0.16).

Figure 9. Response 70 rates in Rutgeerts 1999



At weeks 24 & 48 for rate difference P = 0.094 & 0.038. At weeks 24 & 48 for rate ratio P = 0.108 & 0.054.

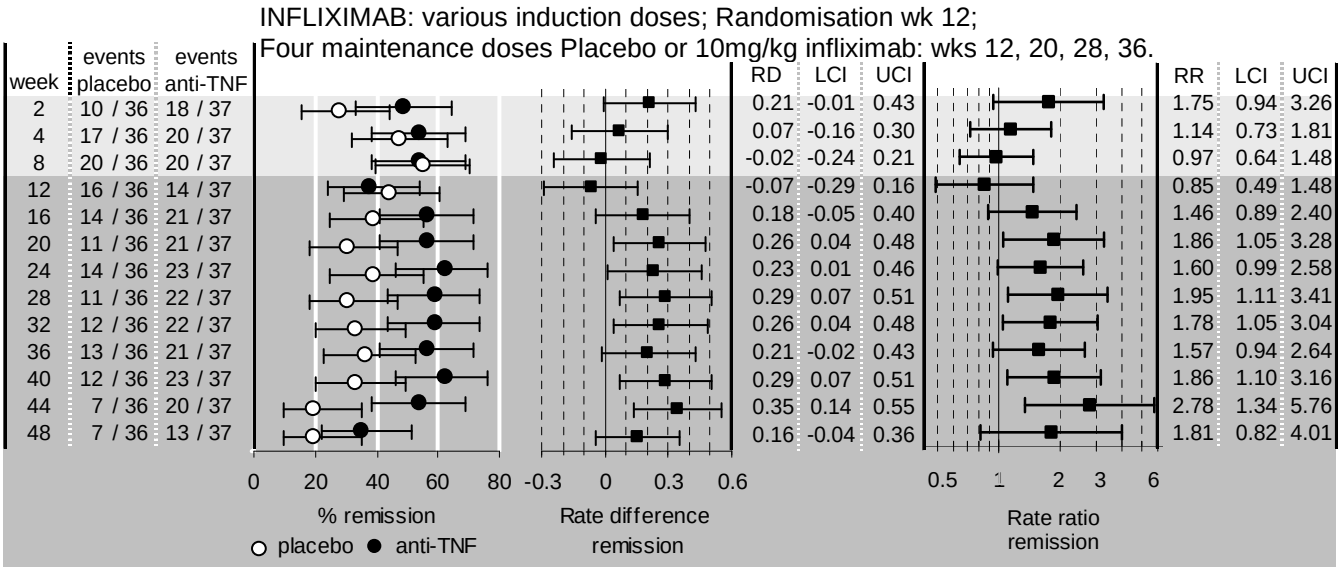
B. Response 100

This outcome was not reported

C. Remission

The point prevalence of remission at different follow up weeks was reported (results are summarised in Figure 10). Point estimates were associated with considerable uncertainty. At randomisation (week 12) ~38% of patients were in remission in the infliximab group, this increased to ~ 60% during weeks 16 to 40. The corresponding values for the placebo group were ~ 44% (week 12) and 35% (weeks 16 to 40). Rate difference (infliximab – placebo) and rate ratio (infliximab / placebo) just reached statistical significance ($P < 0.05$) at most time points for week 16 to week 40.

Figure 10. Remission rates in Rutgeerts 1999

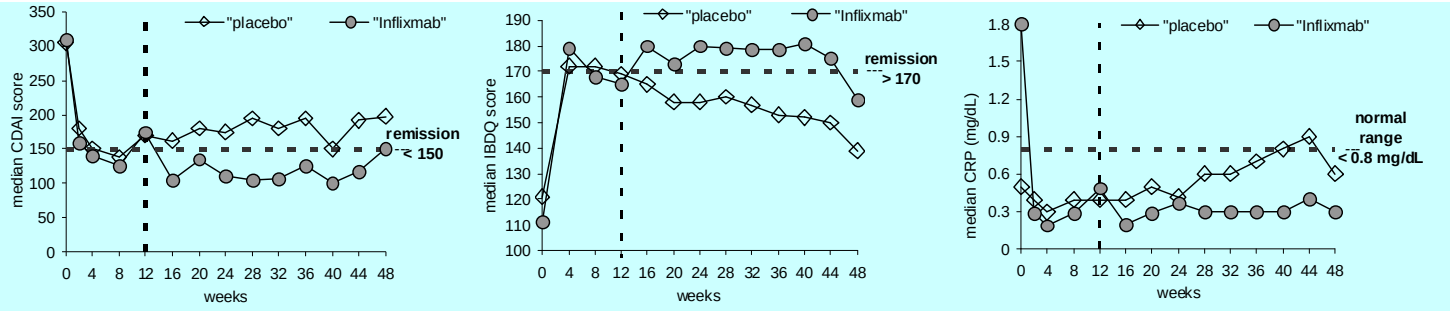


D. Other outcomes

Time to loss of response for patients achieving a response at "*any time*" during follow up after randomisation was reported. The criteria for loss of response were not explicit. Over 48 weeks it is possible for a patient to enter a response state on several occasions. The publication did not make clear which occasion(s) were used in the analysis, or how and if double counting was avoided. The log rank test for difference between placebo and infliximab groups just failed to reach statistical significance ($p = 0.057$).

Median CDAI score, median IBDQ score and median CRP concentrations were reported but range of values and statistical analyses for these outcomes were not presented. The results were in favour of infliximab relative to placebo with greater reduction in CDAI scores, larger increases in IBDQ scores and more "normalisation" of CRP concentrations. The results published are summarised in Figure 11.

Figure 11. Median CDAI, IBDQ and CRP levels reported in Rutgeerts 1999



Data taken from published graphs and redrawn. Where necessary the authors' carried last observation forward.

Quality assessment (based on published report)

Randomisation, allocation concealment and blinding were adequate. Baseline values for those characteristics reported were evenly balanced, but values for CRP, which was not balanced in the original Targan trial, were unclear. Analysis of response 70 and remission rates was by intention to treat; the results presented were point prevalence values at various follow up times, they therefore represent maintenance of response at the group level only and not maintenance by individual patients. For continuous outcomes last observation was carried forward where necessary but the amount of missing data was not reported. No primary outcome was identified and no power calculation was described; the combined trials appear to have been powered only for the induction analysis of Targan 1997 (at week 4 of that study). The maintenance part of the study was probably underpowered. About 33% of patients withdrew.

Rutgeerts 1999. Summary of effectiveness evidence.

The study recruited patients from amongst responders (CDAI score reduced by 70 points) following on from the Targan trial and the resulting induction phase varied between patients in both duration and dose regimen. Subsequent maintenance treatment with infliximab (4 infusions of 10 mg/kg at 8 week intervals) generated a greater proportion of patients with a response 70 and with remission than did treatment with placebo. Point prevalence estimates for these outcomes were associated with considerable uncertainty. The trial left unanswered how well a clinical response is sustained at the individual patient level.

ACCENT I (infliximab)

This was a free-standing maintenance trial (i.e. newly started). There were 580 eligible patients (CDAI range 220 to 400) of whom 573 received a single induction infusion of 5 mg/kg of infliximab. Two weeks later patients were randomised either to placebo, or to 5 mg/kg infliximab at weeks 2 and 6 and then every 8 weeks to week 54, or to 5 mg/kg at weeks 2 and 6 and then 10 mg/kg infliximab every 8 weeks to week 54 (these groups are here termed 5 mg/kg and 10 mg/kg groups respectively). At week 2 (randomisation week) patients were classified as responders (335/573, 58.5%) or non-responders (238/573, 41.5%) depending on whether they achieved a response 70 (a reduction of > 70 points in CDAI score at week 2 relative to baseline). At week 14 patients who initially responded but then worsened were eligible to cross over to treatment with increased dosage of infliximab; this cross over treatment for the placebo group was termed "episodic treatment". The results for responders were published in 2002 (Hanauer et al[4]) and patients who crossed over to increased dosage after week 14 for most of these analyses were considered as treatment failures.

Effectiveness results published for responders only in 2002 (Hanauer [4]) are reviewed below, and results for all patients, irrespective of responder status at week 2 and published in 2004 (Rutgeerts[3]), are considered in the following section.

ACCENT I: Results for responders

Of the 335 responders (58.5% of those who had received an induction dose of 5 mg/kg of infliximab) 110 were randomised to placebo, 113 to the 5 mg/kg infliximab group and 112 to the 10 mg/kg infliximab group.

A. Response 70

The published results for responders[4] included graphical presentation of point prevalence of response 70 at weeks 30 and 54. These results are summarised in Table 14. A statistically significant difference in rates in favour of infliximab versus placebo was reported for both infliximab groups at both weeks 30 and 54. The manufacturer's submission provided point prevalence rates for response 70 for all assessment visit weeks from 2 to 54. These results are summarised in Figure 12.

Point estimates were associated with appreciable uncertainty. Week 2 response rates of ~ 90% had diminished in all groups by week 54 to 15% in the placebo group and 38% and 47% in the 5 mgkg and 10 mg/kg infliximab groups respectively. Rate differences (infliximab – placebo) remained fairly constant from week 14 onward. Rate differences and rate ratios (infliximab / placebo) reached statistical significance in favour of infliximab at all visit times

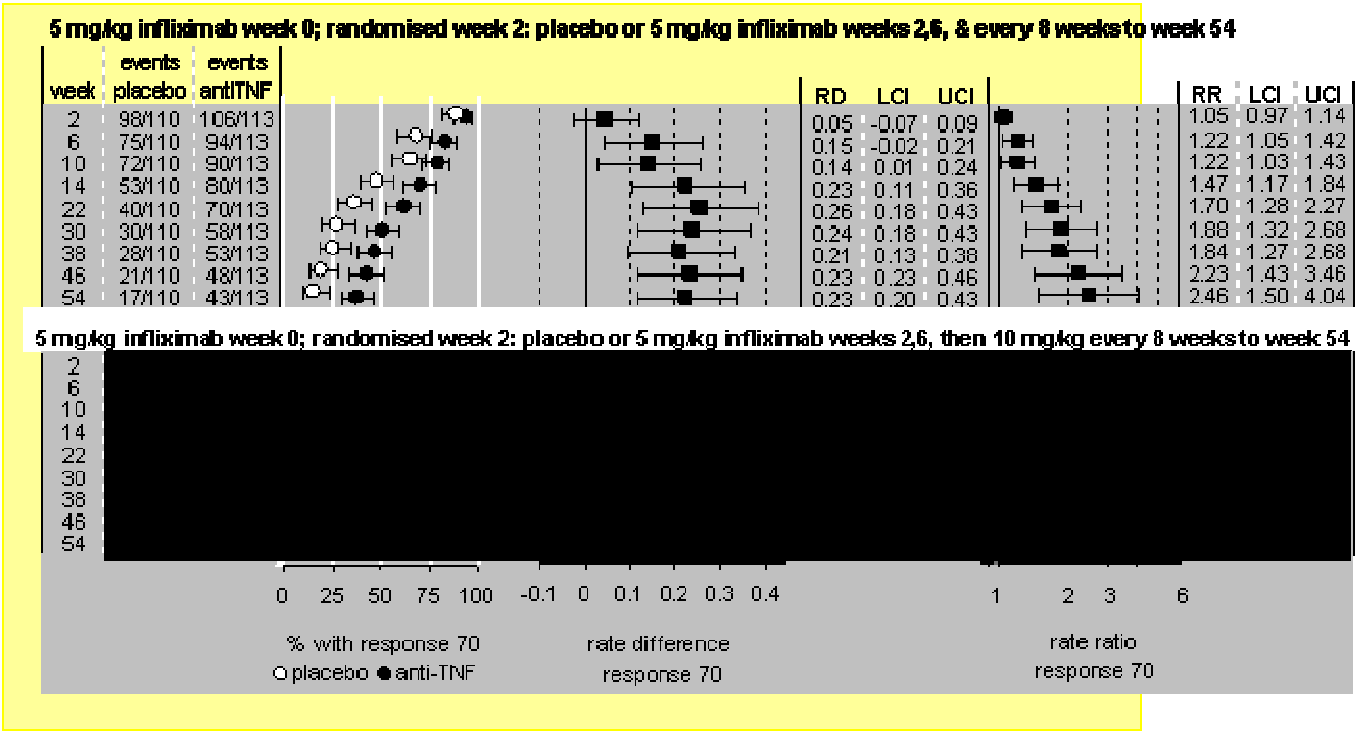
from week 10 to week 54. It is unclear why week 2 response rates were less than 100%; it is possible some patients with a 70 point CDAI reduction from baseline nevertheless required surgery or a change in concomitant medication for worsening of clinical condition. After week 2, decline of response occurred in both placebo and intervention groups, then after week 10 rate differences remained similar (for example for the 5 mg/kg arm rate differences (inflixmab – placebo) remained similar after week 14 as follows: at weeks 10, 14, 22, 30, 38, 46, and 54 rate differences were 0.14, 0.23, 0.26, 0.24, 0.21, 0.23, 0.23). This suggested that most benefit of infliximab was delivered in the first 10-12 weeks of the trial.

Table 14. Published response 70$^{\Theta}$ rates for responders at weeks 30 and 54 in ACCENT I

Dose regimen (N)	week 30*		week 54*	
	response 70 (%)*	P**	response 70 (%)*	P**
PLACEBO (110)	27%	NA	16%	NA
5 mg/kg group (113)	51%	0.0002	38%	0.0001
10 mg/kg group (112)	58%	<0.0001	47%	0.0001

$^{\Theta}$ Response 70 defined as reduction of ≥70 CDAI points from baseline and no requirement for medication change or for surgery. * Data read from published graph. ** intervention v. placebo.

Figure 12. Response 70 rates for responders throughout follow up in ACCENT I



B. Response 100

This outcome was not reported.

C. Remission rates

Remission was a co-primary outcome. The results published for remission at week 30 and week 54 are summarised in Table 15. For this outcome patients who worsened and crossed over to "episodic treatment" (allowed from week 14 onward) were counted as treatment failures (i.e. as no longer in remission). The results reported measured the point prevalence of remission for each group at week 30 and did not require maintenance of response from week 2 to 30 at the patient level. A statistically significant greater proportion of patients were in remission at weeks 30 and 54 in the infliximab groups than in the placebo group. At week 30 the rate differences (infliximab – placebo) were 18% and 25% for the 5 mg/kg and 10 mg/kg groups respectively and the corresponding numbers needed to treat (30 weeks) were 5.66 and 4. Note this NNT estimate does not include non-responders who had been administered induction infliximab. The point prevalence of remission had dimished somewhat by week 54.

Table 15. Remission$^{\Theta}$ rates for responders reported at weeks 30 and 54 in ACCENT I

Dose regimen (N)	Week 30			Week 54	
	Remission: % *(95% CI* ‡*)* (number)	Odds Ratio (95% CI) intervention / placebo	P*	Remission (%)**	P*
PLACEBO (110)	21% *(14% to 29%)* (23)	NR	NA	14%	NA
5 mg/kg group (113)	39% *(30% to 48%)* (44)	NR	0.003	28%	0.007
10 mg/kg group (112)	45% *(36% to 55%)* (50)	NR	0.002	38%	<0.0001
5 mg/kg & 10 mg/kg groups combined (225)	42% *(36% to 48%)* (94)	2.7 (1.6 to 4.6)	NR	33%	

$^{\Theta}$ Remission defined as a CDAI < 150 and no requirement for change in medication or for surgery. ‡ calculated from published values. *intervention v. placebo ** Data read from published graph.

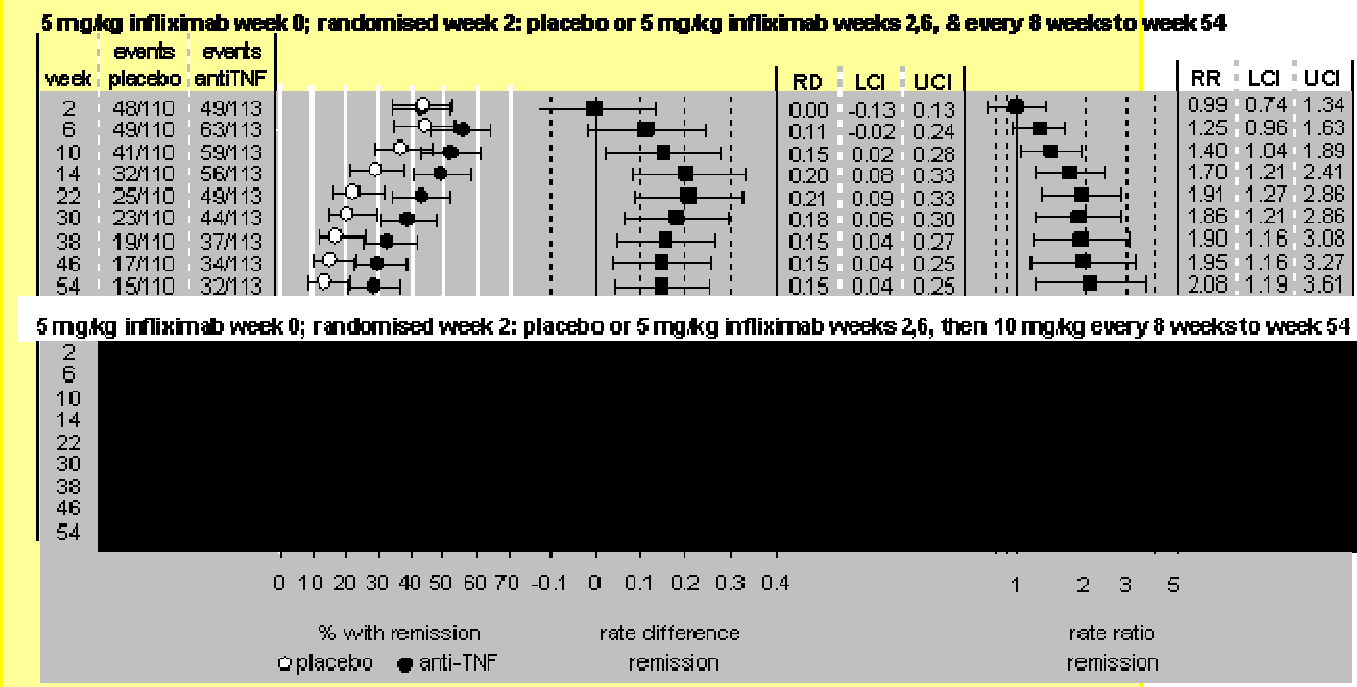
The unpublished Industry Trial Report for ACCENT I provided information regarding the maintenance of remission at the individual patient level for weeks 14 to 54. The percentages were slightly discrepant with those in the published report as indicated below:

Table 16. Patient level maintenance of remission reported in ACCENT I

	% in remission at all visits from week 14 to 54		
	PLACEBO	5 mg/kg group	10 mg/kg group
Published report	11%	25%	33%
Trial report	[redacted]	[redacted]	[redacted]
[redacted]			

The manufacturer's submission and the Industry Trial Report provided CIC point prevalence rates for remission for all assessment visits from week 2 to 54. These results are summarised in Figure 13.

Figure 13. Remission rates for responders throughout follow up in ACCENT I



At week 30 for rate difference p = 0.0027 and <0.0001 for 5 mg/kg and 10 mg/kg groups respectively. At week 30 for rate ratio p = 0.0047 and 0.00025 for 5 mg/kg and 10 mg/kg groups respectively.

Point estimates were associated with appreciable uncertainty. [redacted] From week 10, remission rates diminished in all groups and rate difference (infliximab – placebo) diminished or remained fairly constant; rate differences and rate ratios (infliximab/placebo) reached statistical significance at all visit times from week 10 onwards. It is evident that loss of remission was continuous after week 6 to 10 of follow up and that the advantage of intervention over placebo was mostly gained by about week 6 to 10, the phase of the study in which dose frequency was greatest. Thereafter decline of response was about the same for both placebo and intervention groups despite continued infliximab every 8 weeks in the treatment arms; for example for the 5 mg/kg arm rate differences (inflixmab – placebo) remained similar after week 14 as follows: at weeks 10, 14, 22, 30, 38, 46, and 54 rate differences were 0.15, 0.21, 0.20, 0.18, 0.15, 0.15, 0.15.

D. Other outcomes

The primary outcome in ACCENT I was identified as time to loss of response. (NB. A protocol amendment added the proportion of responder patients in remission at week 30 as a co-primary outcome, which has been reported above). Loss of response was defined as a CDAI of ≥ 175, a CDAI increased by ≥ 35% and a CDAI increased by ≥ 70 points relative to the qualifying value for a response on at least two consecutive assessments, or requirement for change in medication or requirement for surgery. Assessments were scheduled at weeks 0, 2, 6, 10, 14 and then every 8 weeks to week 54. With this definition of loss of response it is possible for an individual responder to no longer qualify as achieving a response 70 status but counter-intuitively nevertheless to not have lost response[a]. For this primary outcome patients in the active intervention arms had significantly longer time to loss of response than patients given placebo (p = 0.0002, log rank test). The median times to loss of response are summarised in Table 17 .

Table 17. Median time to loss of response in responders in ACCENT I

Dose regimen (N)	Median time (weeks) to loss of response	Interquartile range
PLACEBO (110)	19	10 to 45
5 mg/kg group (113)	38	15 to >54
10 mg/kg group (112)	> 54	21 to > 54
5 mg/kg and 10 mg/kg groups combined (225)	46	17 to > 54

Published effectiveness results for responders included median CDAI scores and median IBDQ scores. These are summarised Table 14. For missing values of CDAI and IBDQ the nearest observation was carried forward. CDAI scores and IBDQ scores diminished and increased respectively to a greater extent in the infliximab groups than in the placebo group. The interquartile ranges for median values during follow up were not reported.

Table 18. Median CDAI and IBDQ scores for responders during follow up in ACCENT I

	CDAI :median*				IBDQ median *			
week	placebo N=110	5 mg/kg N=113	10 mg/kg N=112	P	Placebo N=110	5 mg/kg N=113	10 mg/kg N=112	P
0	290	305	305	NS [a b]	129	128	130	NS [a b]
2	157	155	152	0.01[a] 0.04[b]	173	169	173	NR
6	159	138	140	<0.0001 [a] <0.002 [b]	165	174	161	NR
10	165	131	127	<0.0001 [a b]	160	170	169	NS [a b]
14	197	145	125	<0.0001 [a b]	155	167	172	0.05[a] 0.0076[b]
22	217	163	135	<0.0001 [a b]	142	164	169	0.013[a] <0.0001[b]
30	225	172	150	<0.0001 [a b]	144	162	167	0.015[a] 0.001[b]
38	238	214	140	<0.0001 [a b]	137	151	170	0.015[a] <0.0001[b]
46	235	200	142	<0.0001 [a b]	135	144	169	0.06[a] <0.0001[b]
54	238	192	152	<0.0001 [a b]	136	150	167	0.015[a] <0.0001[b]

* Data read from graph. [a] Comparison 5 mg/kg group versus placebo: [b] Comparison10 mg/kg group versus placebo Test for significance were done by ANOVA.

[a] For example an individual with CDAI of 221 at enrolment would qualify as a responder at week 2 with a CDAI score reduced by 71 points to 150. If this patient's CDAI subsequently rose to 170 they would no longer be in a response 70 but would nevertheless not have lost response because the increase in score from week 2 was less than 70 points, less than 35% of week 2 score and below a score of 175.

The manufacturer's submission provided information about quality of life measures (SF-36). The SF-36 scores were reported separately for mental and physical components for weeks 30 and 54 of the trial and mean improvement from baseline was reported. Standard deviations of values were provided. The results are summarised in Table 19 . Change from baseline for SF-36 physical component reached statistical signicance in favour of infliximab at both weeks 30 and 54

Table 19. SF-36 results reported for responders in ACCENT I

		SF-36 score			Mean difference from Baseline	
SF-36 component	GROUP	Baseline	Week 30	Week 54	Week 30	Week 54
Physical Component	Infliximab	33.9 ± 8.8	37.2 ± 11.3	36.5 ± 11.0	3.1 ± 9.5	2.5 ± 9.0
	Placebo	33.0 ± 8.5	40.4 ± 11.3	39.2 ± 11.9	7.3 ± 10.3	6.1 ± 10.8
					P=0.002	*P=0.014*
Mental Component	Infliximab	39.8 ± 11.3	42.8 ± 12.0	42.1 ± 12.0	2.9 ± 11.2	2.0 ± 10.9
	Placebo	38.8 ± 11.3	43.2 ± 11.4	43.9 ± 12.2	4.6 ± 12.7	5.1 ± 12.8
					P=0.348	*P=0.072*

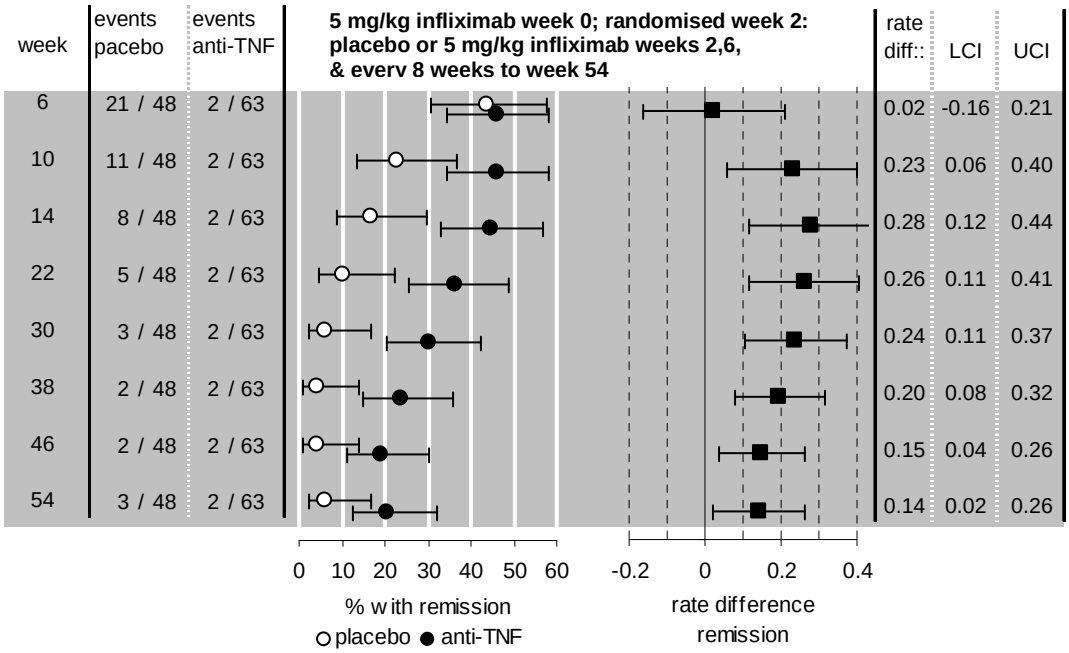
The results for infliximab refer to the 5mg/kg group only. P values refer to comparison between infliximab and placebo groups.

Median daily steroid dose was reduced by week 14 in all groups and then remained constant. The reduction in the infliximab groups was greater than that for the placebo group. The odds ratio for discontinuation of steroid use (infliximab / placebo) at week 54 was 4.2 (95% CI 1.5 to 11.5).

E. Other considerations - Subgroup analysis of remission rate in severe CD patients

The manufacturer's submission for infliximab provided CIC information about the proportion of responder patients who initially had severe disease (defined as a baseline CDAI score > 300) and who achieved remission status during follow up. Results presented referred to patients classified as having severe disease who were randomised to the 5 mg/kg infliximab group (n= 63/113 (56%)) and placebo group (n= 48/110 (44%)). No information was provided regarding patients with severe disease amongst non-responders. The remission rates in placebo and 5 mg/kg infliximab arms and the rate difference for this subgroup of patients are shown in Figure 14.. Remission rates were slightly poorer in this more severe CDAI group than for all responders, but a similar pattern was shown during follow up in that most of the advantage from the intervention was achieved with the first three doses (early phase). Thereafter, remission decayed away at approximately similar rates in the two arms even though patients in the intervention arm received further doses of infliximab and rate differences decreased from week 14 onward.

Figure 14. Remission rates and rate difference for responders with severe disease in ACCENT I



ACCENT I (responders): Quality assessment (based on published report of Hanauer 2002)
Randomisation, allocation concealment and blinding were adequate. Baseline characteristics were only reported for all patients, (ie for all responders and for all non-responders). It was therefore not possible to judge if baseline characteristics were evenly balanced between the three arms of responders that were analysed for effectiveness outcomes. Similarly the number of patients who withdrew was reported for all enrolled patients and it was not possible to determine how many responders discontinued their randomised treatment. Where necessary the nearest or last observation was carried forward for continuous outcomes but the amount of missing data was not reported. A power calculation was conducted and based on the primary outcome of loss of response. The definition of loss of response was complex, did not correspond to a failure to maintain a response 70 status, and its clinical meaning was difficult to gauge.

ACCENT I. Summary of effectiveness evidence for responders.

Of 573 patients (with baseline CDAI 220 to 400), 58.5% (335) achieved response 70 two weeks after a single induction infusion of 5 mg/kg of infliximab. These patients were designated "responders". It is unclear if the three trial arms of randomised responders were well balanced at baseline. Of responders, [redacted] were in remission (CDAI < 150) at week 2. This represented [redacted] of the original 573 patients. The proportion of responders with remission had declined by week 30 to 23% (95% CI: 14% to 29%) for those who only received placebo after induction and to 39% (95% CI: 30.% to 48%) for those who received four infusions of 5 mg/kg infliximab (at weeks 2, 6, 14 and 22) and to 42% (95% CI: 36% to 55%) for those received four infusions consisting of 5 mg/kg at weeks 2 and 6 and 10 mg/kg at weeks 14 and 22. Rate differences (infliximab – placebo) and rate ratios (infliximab/placebo) for remission at week 30 reached statistical significance in favour of infliximab for both infliximab groups. By week 54 the percentage of patients in remission had diminished further in all three groups. Most of the advantage of intervention relative to placebo was achieved by weeks 10 to 14; thereafter rate differences remained fairly stable. A similar pattern of results was observed for response 70. Published information regarding maintenance of remission at the patient level (as distinct from group level) was meagre. Between weeks 14 and 54, 11% of placebo patients retained remission at all six study visits; the corresponding values were 25% and 33% respectively for 5 mg/kg and 10 mg/kg infliximab groups. Somewhat lower values of [redacted] respectively were quoted in the Industry Trial Report. Results favouring infliximab over placebo were reported for several other outcomes including median CDAI scores and median IBDQ scores. These measures required last or nearest observation carried forward in order to allow for missing data.

ACCENT I: Results for all patients (Rutgeerts[3]) (Infliximab)

The results for all 573 patients that received an induction dose in ACCENT I were presented by Rutgeerts et al. 2004 in a paper published two years after that describing results for responders only. Separate results for non-responders have not been published. The 573 patients were 335 responders and 238 non-responders (defined according to whether a 70 point reduction in CDAI score was attained by week 2 after the induction infusion).

Randomisation at week 2 resulted in allocation of 188 patients to the placebo group, 192 to the 5mg/kg group and 193 to the 10 mg/kg group.

The authors stated *"the primary objective of the analysis was to examine the difference in efficacy between episodic and scheduled treatment strategies with infliximab under conditions that simulate clinical practice"*. For this purpose the patients in the original placebo group were designated as receiving "episodic strategy", and those in the infliximab groups as receiving a "5 mg/kg scheduled strategy" and a "10 mg/kg scheduled strategy" respectively. From week 14 onward patients who had shown a response to infliximab therapy at any time but then worsened were eligible to cross over to "*active episodic treatment as needed with infliximab 5, 10, and 15 mg/kg for patients originally assigned to episodic, 5 mg/kg scheduled, and 10 mg/kg scheduled treatment strategies respectively"*. This description is confusing since it clearly states that *active episodic treatment* is given in both episodic and scheduled strategies, which renders a comparison of episodic and scheduled strategies problematical. The publication designates the start of episodic treatment to be week 14. See Appendix 9 for patient flow through the trial.

The treatment regimens received before week 14 in each of the randomised groups were as follows:

Placebo / "episodic group":	5 mg/kg infliximab week 0, placebo weeks 2, and 6.
5 mg/kg group "scheduled strategy"	5 mg/kg infliximab week 0, week 2 and week 6.
10 mg/kg group "scheduled strategy"	5 mg/kg infliximab week 0, week 2 and week 6.

Treatment to week 14 was therefore similar for the two infliximab "scheduled strategy" groups and was determined according to randomisation. From week 14, cross over to an increase in infliximab dosage was allowed in all three trial arms for patients whose CD worsened. The criteria for worsening were "*an increase CDAI of ≥ 70 points from the qualifying score with a total score of at least 175, an increase in CDAI of 35% or more from baseline value, or the introduction of new treatment for active Crohn's disease"*. From week 14 onward it was possible for patients in different arms to be receiving identical infliximab

treatment; for example a placebo patient might cross over at week 14 to receive 5 mg/kg and this corresponds to treatment received by a 5 mg/kg "scheduled strategy" patient who did not cross over. This complicates the interpretation of any comparisons between groups.

A. Response 70

No primary outcome was identified. Analyses were according to randomised group irrespective of cross over after week 14 to different treatment regimen, and comparisons were drawn between the "episodic group" and the two "scheduled strategy" groups. The results for response 70 for all patients in ACCENT I are summarised in Figure 15. By week 14 statistically significant differences in CD status were evident between placebo group and intervention groups. P values for rate differences and rate ratios are shown in Table 20 .Rate differences and rate ratios for comparison between "episodic" and "scheduled" strategies after week 14 were in favour of "scheduled strategies" but failed to reach statistical significance at most time points. Interpretation of these differences is problematical.

Figure 15. Response 70 rates for all patients in ACCENT I

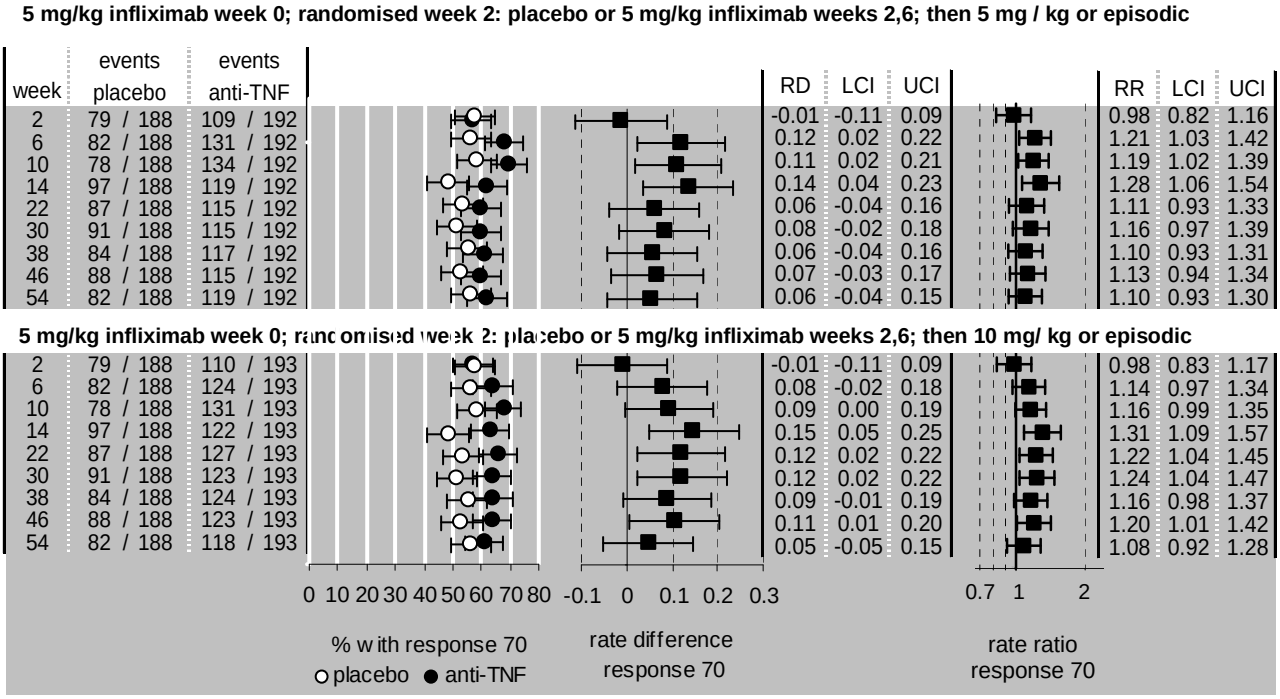


Table 20. P values for comparison of response 70 rates at week 14 for all patients in ACCENT I

	rate difference (placebo – active intervention)		rate ratio (placebo / active intervention)	
	vs. 5 mg/kg group	vs. 10 mg/kg group	vs. 5 mg/kg group	vs. 10 mg/kg group
P	0.00725	0.00326	0.00865	0.00418

B. Response 100

This outcome was not reported

C. Remission rates

Figure 16 summarises the published results for rates of remission at clinic visits to end of follow up (week 54). Week 14 remission rates were greater in the two "scheduled treatment" arms (37.5% in the 5 mg/kg group and 43% in the 10 mg/kg group) than in the "episodic" group (25.5%). P values for week 14 comparisons between placebo and intervention groups are shown in Table 21 .

Figure 16. Remission rates for all patients in ACCENT I.

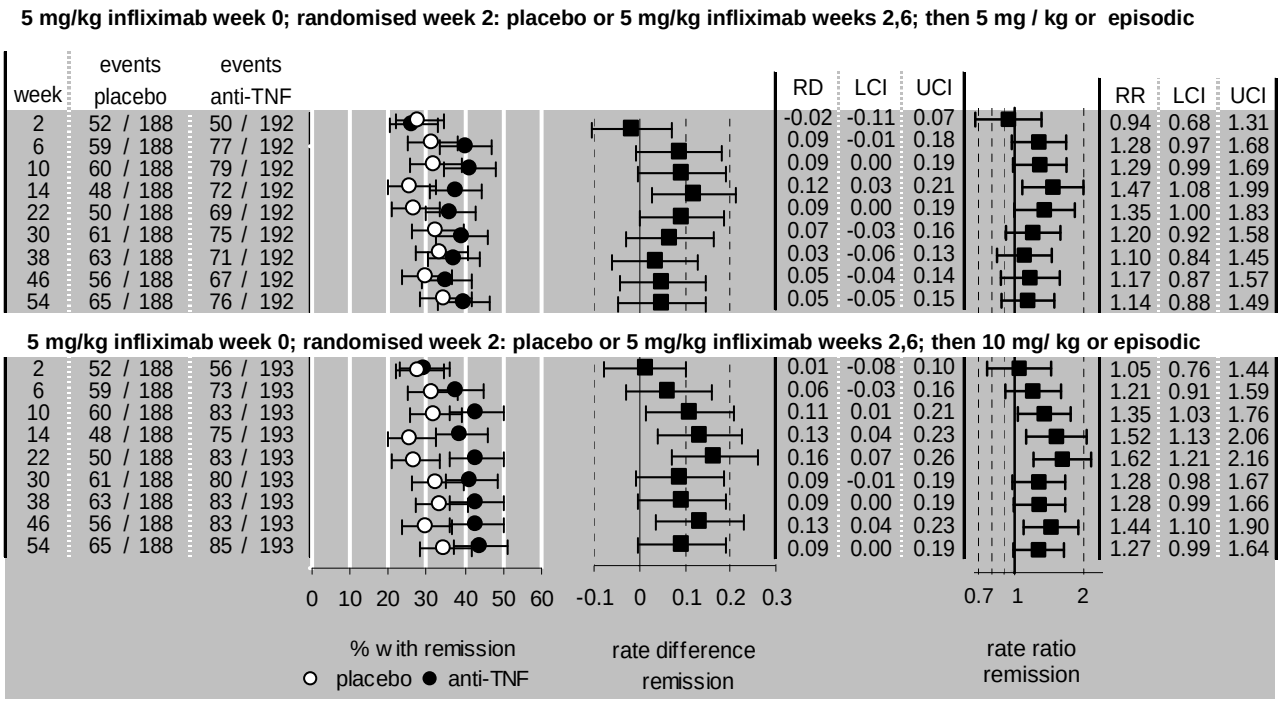


Table 21. P values for comparison of remission rates at week 14 for all patients in ACCENT I

	rate difference (placebo – active intervention)		rate ratio (placebo / active intervention)	
	vs. 5 mg/kg group	vs. 10 mg/kg group	vs. 5 mg/kg group	vs. 10 mg/kg group
P	0.0113	0.0049	0.0135	0.0063

Treatment regimens up to week 14 were strictly prespecified and designed to examine effectiveness for maintenance of the induced response. After week 14 treatment regimens became variable (termed "*episodic*" by the authors). It is clear that by week 14 the CD status of patients in the placebo/"episodic" arm had departed from that of patients in the two "scheduled strategy" arms; this means that at baseline (week 14) for the comparison of "episodic" with "scheduled strategies" the groups were imbalanced. Comparisons between "episodic" and "scheduled" strategies after week 14 are not randomised comparisons. For a randomised comparison of the two strategies patients should have been re-randomised at

week 14. Furthermore, to what extent remission after week 14 depended on active intervention is not known since after week 14 no inactive (placebo) arm existed.

Rate differences and rate ratios for comparison between "episodic" and "scheduled" strategies after week 14 were in favour of "scheduled strategies" but failed to reach statistical significance at nearly all time points. Interpretation of these differences is problematical because, as described above, the comparisons are not between properly randomised groups and because patients in all groups were allowed the option of "*episodic*" treatment.

D. Other outcomes

Median CDAI score and the proportion of patients with IBDQ score greater than 170 were reported and are summarised in Table 22 and presented graphically in Appendix 10. By week 14, statistically significant differences in CDAI median scores were evident between the placebo group and the intervention groups. Differences were less pronounced after week 14, especially for the placebo v. 5 mg/kg comparison. The percentage of patients with IBDQ score greater than 170 did not differ significantly between placebo and 5 mg/kg groups, but after week 14 favoured the 10 mg/kg group relative to placebo.

Table 22. CDAI and IBDQ results for all patients in ACCENT I

	CDAI :median[*]				% patients with IBDQ score > 170 [*]			
week	placebo N=188	5 mg/kg N=193	10 mg/kg N=192	P	Placebo N=188	5 mg/kg N=193	10 mg/kg N=192	P
0	292	303	297	NS [a b]	4.8	5.2	8.3	NS [a b]
2	197.5	205	195	NS [a b]	35.6	32.3	35.8	NS [a b]
6	205	180	180	NS [a b]	33.5	41.7	38.3	NS [a b]
10	187.5	170	167.5	< 0.05 [a b]	35.1	41.7	39.9	NS [a b]
14	225	185	182.5	< 0.05 [a b]	29.8	38.0	40.9	NS [a] < 0.05 [b]
22	212.5	185	167.5	< 0.05 [a b]	29.3	37.0	44.0	NS [a] < 0.05 [b]
30	212.5	180	177.5	NS [a] < 0.05 [b]	33.5	39.6	44.0	NS [a] < 0.05 [b]
38	200	187.5	170	NS [a] < 0.05 [b]	35.1	34.9	47.7	NS [a] < 0.05 [b]
46	205	190	175	NS [a] < 0.05 [b]	33.5	35.4	48.7	NS [a] < 0.05 [b]
54	205	185	170	NS [a] < 0.05 [b]	35.1	37.5	46.1	NS [a] < 0.05 [b]

[*] Data read from graph. Comparisons: [a] 5 mg/kg group versus placebo, [b] 10 mg/kg group versus placebo, no adjustment for repeated measures.

The manufacturer's submission provided information regarding CD-related hospitalisation rates and rates for intra-abdominal surgery. These rates and the relative risk for the 5 mg/kg "scheduled maintenance"group relative to the "episodic" are summarised in Table 23. The results for mucosal healing observed for a small subgroup of patients (N=58) at European study centres that underwent endoscopy examination are also tabulated. The interpretation of the comparisons is problematical for the reasons already described, in particular after week 14. The extent to which avoidance of hospitalisation and abdominal surgery might depend on the administration of active intervention is not measurable because no true control (placebo) group existed after that time.

Table 23. Endoscopy, hospitalisation and abdominal surgery results: all patients in ACCENT I.

	Endoscopy		Hospitalisations[Θ]		Abdominal surgery	
	mucosal healing at week 54	P[*]	n/N (%)	Relative risk *(95% CI)*	n/N	Relative risk *(95% CI)*
Placebo	4/22	NA	71/188 (38%)	NA	14/188	NA
5 mg/kg group	8/19	0.093			5/193	0.348[†] *(0.128 to 0.947)*
10 mg/kg group	8/17	0.053			6/192	
Combined 5 & 10 mg/kg groups	16/36	0.041	86/305 (23%)	0.591[*†] *(0.455 to 0.768)*	11/385	0.373 *(0.173 to 0.806)*

[*] Comparison for placebo v infliximab. [Θ] The hospitalisation rates were presented differently to other rates as number per 100 patients, rather than number per total at risk. We have calculated the number of hospitalisations based on the reported % and the known total numbers of patients. [†] Values presented in industry submission for hospitalisation were 0.6 *(0.5 to 0.7)* and for abdominal surgery were 0.3 *(0.2 to 0.6)*.

Quality assessment of ACCENT I (all patients): (based on published report of Rutgeerts 2004)[3]

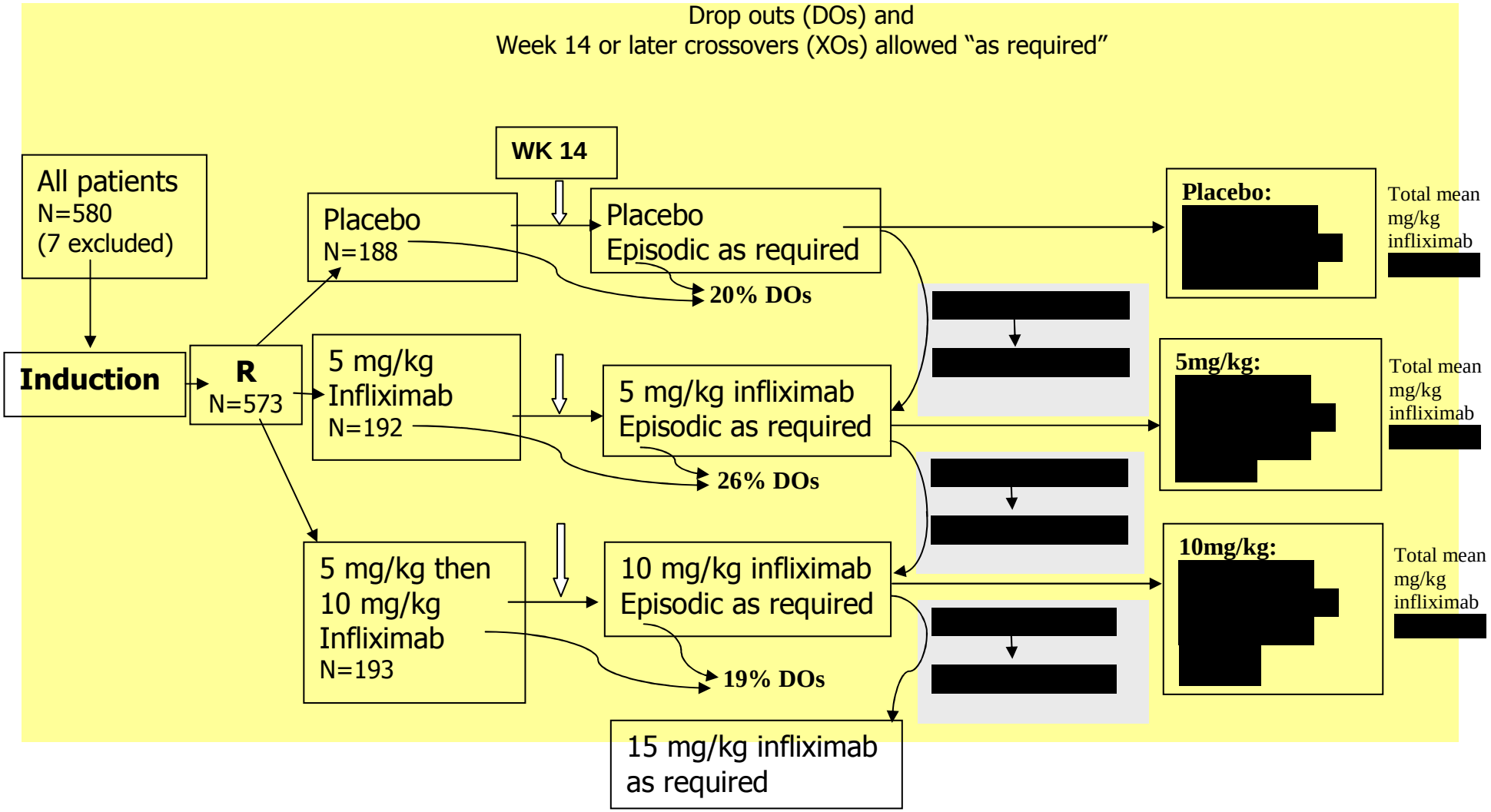
Randomisation, allocation concealment and blinding were adequate. Baseline characteristics at week 0 were well balanced. Where necessary, the nearest or last observation was carried forward for continuous outcomes but the amount of missing data was not reported. No power calculation was conducted for the analysis of all patients. The number of patients who withdrew was reported except for patients who crossed over to a 15 mg/kg dose regimen from a 10 mg/kg regimen. The proportion of patients who withdrew before the end of the trial was substantial.

Trial design, withdrawals, cross overs and validity of comparisons.

It must be questionable whether the "episodic" (placebo) arm did "*simulate clinical practice"* as stated to be an objective of the study. Patients in this arm of the study received one dose of 5 mg/kg infliximab at week 0, followed by an interim period of more than 3 months with no active infliximab therapy before the "*episodic*" use of infliximab according to worsening disease (for patients "*who had responded at any time to infliximab therapy*"). There is little evidence to support the idea that this resembles clinical practice. The scheduled strategy is difficult to define since it did not follow a prescribed programme of treatment as might be anticipated by the term "scheduled strategy" but encompassed "*episodic*" treatment in the same manner as the "episodic" arm.

Because of the large numbers of patients that withdrew from treatment and crossed over to dose escalations, the actual treatments received in the three different trial arms are difficult to define. Figure 17 summarises the progression of patients through the trial with respect to withdrawal from treatment and crossover to increased dose of infliximab.

Figure 17. Withdrawals and cross overs in ACCENT I



Over a period of one year, about a quarter of patients withdrew from treatment, and of those allocated active intervention at randomisation only about half completed the trial receiving the treatment regimen to which they had been allocated at randomisation.

The authors' stated primary objective *" .. was to examine the difference in efficacy between episodic and scheduled treatment strategies with infliximab"* [3]. They concluded that the scheduled treatment strategy was superior to episodic treatment. Unfortunately the comparisons were compromised by strong biases introduced as a result of the study design. These biases are explained below:

a] Cross-over to increased infliximab was allowed for patients *"who had responded a any time to infliximab therapy"* and subsequently worsened. In the placebo group ("*episodic strategy*") 78 of 188 patients (41%) were classified at week 2 as non-responders and received no further infliximab to week 14; these patients were unlikely to become responsive and therefore to qualify for cross-over to active intervention. In contrast to this group the week-2 non-responders in the "scheduled strategy" arms received additional doses of infliximab (5 mg/kg) at both weeks 2 and 6, boosting their opportunity to "*respond at any time"* to infliximab. The greater opportunity to respond at any time in the "scheduled strategy" arms represents a strong bias in their favour in any subsequent comparison with the episodic arm. Relative to the scheduled strategy this resulted in a substantial proportion of patients in the episodic arm being denied access to active therapy. This is reflected in the very large difference between arms in their exposure to infliximab stated to be 3 and 5 times greater in the two scheduled strategy arms compared to episodic.

b] Episodic treatment was introduced at week 14 of the trial, but by this time the CD status of patients in the placebo "episodic" arm was significantly inferior to that in the scheduled strategy arms in terms of several efficacy measures. This advantage for the scheduled strategy arms is reflected in increases from week 2 in the response 70 and remission rates at weeks 6 and 10 not seen in the placebo group. The result is a bias in favour of scheduled strategy for any comparison between strategies at times after week 14. Essentially the compared arms were unbalanced at the start of the compared strategies (week 14).

ACCENT I. Summary of effectiveness evidence for all patients.

Two infusions of 5 mg/kg of infliximab at weeks 2 and 6 after a single induction infusion of 5 mg/kg were better than placebo infusions at generating remission and response 70. At week 14, rate differences (infliximab – placebo) and rate ratios (infliximab/placebo) were in favour of infliximab and reached statistical significance ($p < 0.02$ for remission, $p < 0.01$ for response 70).

At week 14 "episodic" treatment was introduced and subsequent comparisons were made between the original placebo arm (designated "episodic reatment strategy") and original infliximab arms (termed "scheduled treatment strategies"). Because of bias strongly in favour of scheduled strategy groups the post-14 week comparisons were not valid estimates of the relative effectiveness of strategies. Biases identified arose from: [a] reduced opportunity for cross-over to active therapy for patients in the episodic group compared to the scheduled groups; [b] gross imbalance in disease status at the start of the strategies (week 14).

Difficulties in interpreting post-14 week comparisons between groups were compounded by the very high rate of withdrawal from treatment and the use of "episodic" treatment in all three arms of the trial so that the distinction between episodic and scheduled strategies was obscured except for the fact that the original infliximab groups were allowed larger dosages of active intervention.

CLASSIC II (adalimumab)

The CLASSIC II trial was an extension of the previously conducted adalimumab induction trial CLASSIC I which had enrolled 299 patients. To be eligible for CLASSIC II, patients were required to be in remission (CDAI < 150) at week 4 of CLASSIC I and also 4 weeks later (equivalent to week 8 of CLASSIC I and designated week 4 of CLASSIC II). These patients may have received two subcutaneous injections two weeks apart of various doses of adalimumab (40 mg then 20 mg, or 80 mg then 40 mg or 160 mg then 80 mg) or two injections of placebo. Fifty five eligible patients entered CLASSIC II, this means about 12 patients did not retain remission from week 4 to week 8 of CLASSIC I or declined to participate. The 55 patients were randomised at week 4 of CLASSIC II to receive placebo (N=18) or 40 mg of adalimumab every other week (eow) (N=19) or 40 mg of adalimumab weekly (N=18) from week 4 to 54. Thus CLASSIC II analysed only strong responders from the CLASSIC I trial.

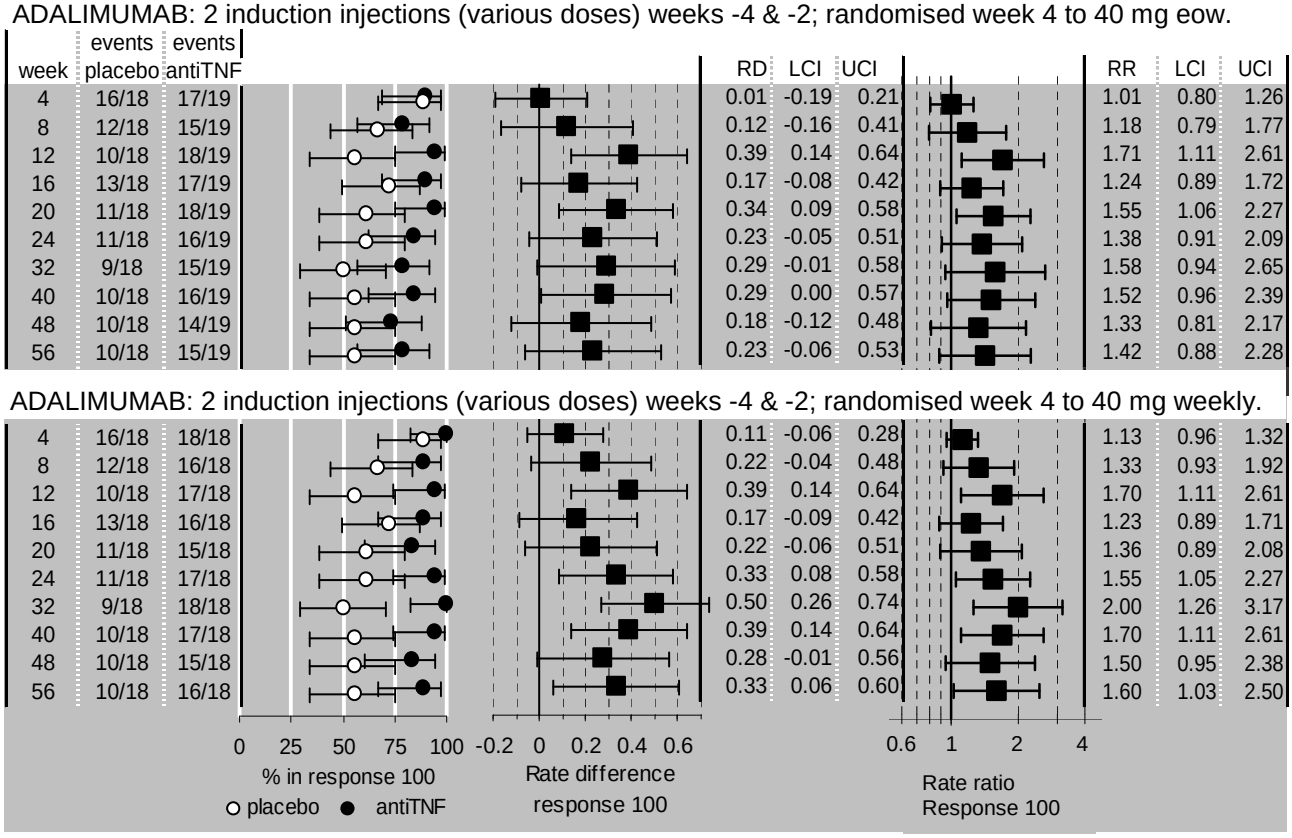
For the purposes of the "*primary efficacy analysis*" patients who "*had continued non-response defined as a decrease of ≤ 70 points v week 0 of CLASSIC I"* were considered treatment failures and became eligible for open label treatment. This means patients in remission at start of CLASSIC II became treatment failures if they ceased to qualify as response 70 responders relative to their baseline CDAI score in CLASSIC I. In addition, patients who flared during CLASSIC II follow up were also counted as treatment failures and were eligible for open label treatment. CD flare was defined as an increase of ≥ 70 points above the week 4 CLASSIC II value (which by definition was < 150) AND a CDAI score > than 150 (no longer in remission). Thus a patient in remission at week 4 (CLASSIC II) with a CDAI score of 149 would require to move to a CDAI of at least 219 to be classified as having experienced flare. For this patient a score of 218 would not count as a flare but could count as treatment failure if their week 0 CLASSIC I CDAI score had been less than 288 (for reference the mean baseline CDAI score at week 0 for 299 CLASSIC I patients was 298).

A. Response 70 and B. Response 100.

Response 100 and response 70 rates throughout follow up were among the secondary outcome measures of efficacy. Results reported for responses 100 and 70 and are summarised in Figure 18. The placebo rates were high for these less rigorous measures of effectiveness and the rate differences (adalimumab – placebo) and rate ratios (adalimumab / placebo) failed to reach statistical significance at most time points.

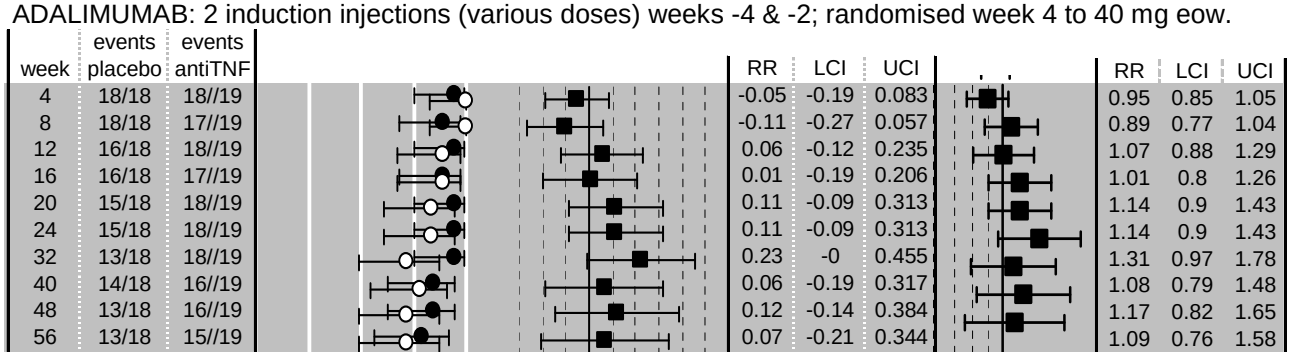
Figure 18. Response 100 (upper panel) and response 70 (lower panel) rates in CLASSIC II

ADALIMUMAB: 2 induction injections (various doses) weeks -4 & -2; randomised week 4 to 40 mg eow.

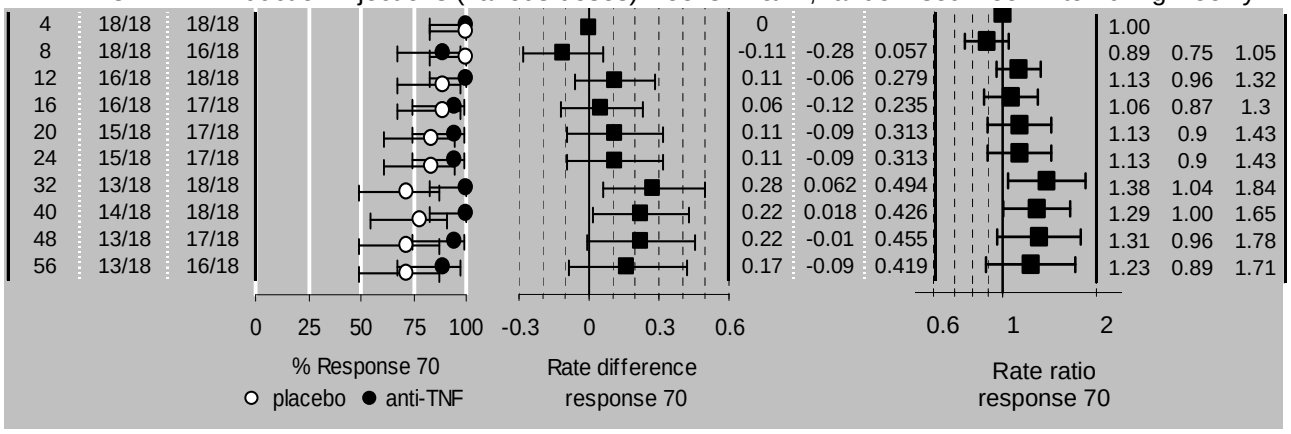


ADALIMUMAB: 2 induction injections (various doses) weeks -4 & -2; randomised week 4 to 40 mg weekly.

ADALIMUMAB: 2 induction injections (various doses) weeks -4 & -2; randomised week 4 to 40 mg eow.



ADALIMUMAB: 2 induction injections (various doses) weeks -4 & -2; randomised week 4 to 40 mg weekly.



C. Remission

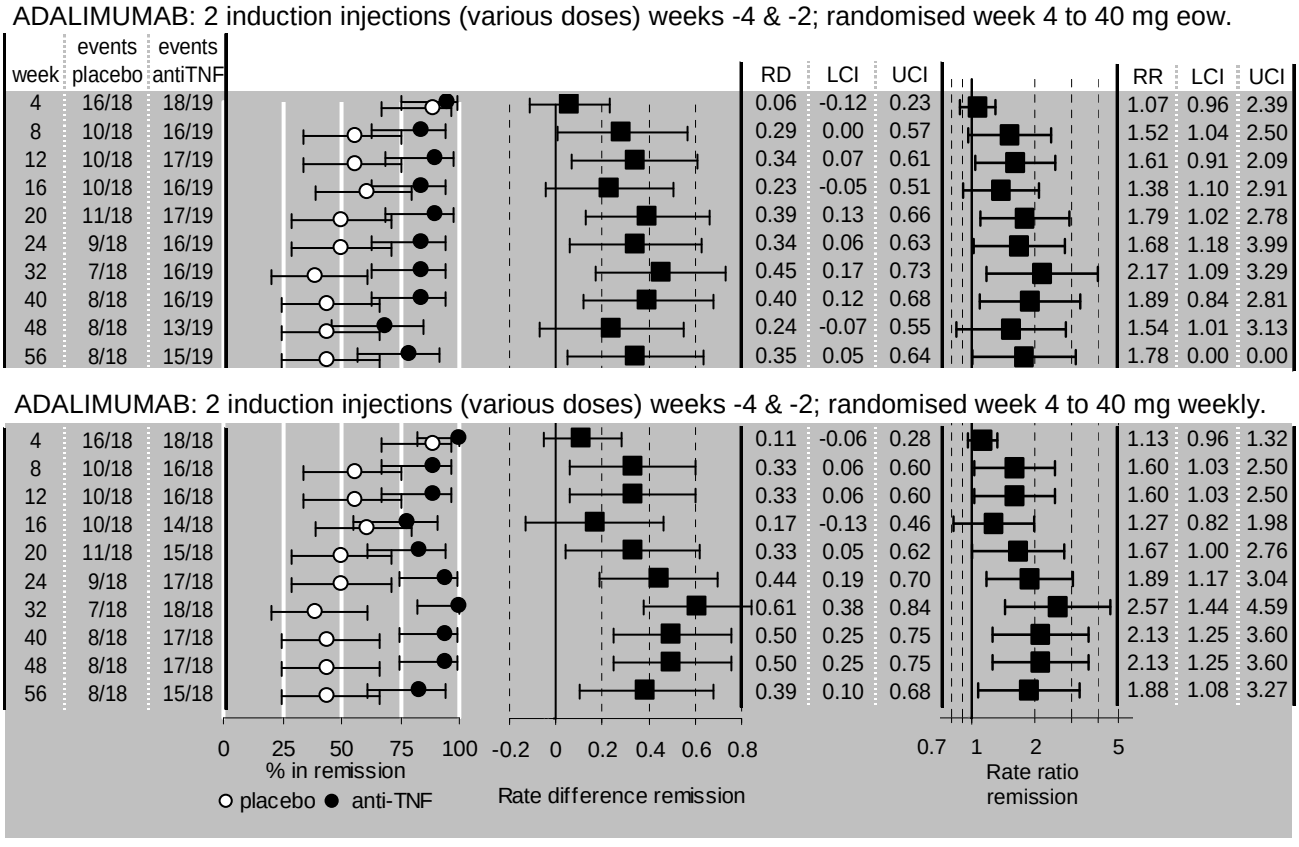
The primary outcome was the proportion of patients in remission at week 56 in each arm of the randomised cohort. Remission throughout follow up was among the secondary outcome measures of efficacy. For the primary outcome, ten patients (18%) withdrew before week 56 (5 from placebo and 5 from adalimumab). These were counted as remission failures for the primary analysis. Remission rates at week 56 are summarised in Table 24. Remission rates during the trial are summarised in Figure 19.

Table 24. Remission rates at week 56 in CLASSIC II (primary outcome)

Dose regimen (N)	Number in remission (%; *95% CI*)	P*
Placebo (18)	8 (44%; *25 to 66*)	NA
40 mg adalimumab eow (19)	15 (79%; *57 to 91*)	< 0.05
40 mg adalimumab weekly (18)	15 (83%; *61 to 94*)	< 0.05
[redacted]**	[redacted]***	

* Adalimumab versus placebo. ** 60 patients increased regimen from eow to weekly. *** Trial report data; publication states 94.

Figure 19. Remission rates in CLASSIC II



Point estimates of remission rate during the trial were associated with considerable uncertainty, reflecting the small number of patients in the trial. The fact that rates rose and fell during follow up indicated the values reported referred to point prevalence. Nearly half of patients in the placebo group were in remission at week 56 despite not receiving active

intervention from 2 weeks prior to randomisation onwards. Rate differences (intervention – placebo) and rate ratios (intervention/placebo) were in favour of intervention at all follow up times and reached statistical significance at several time points.

D. Other outcomes

The results published for continuous measures are summarised in Table 25. These measures involved last observation carried forward to allow for missing values. The amount of missing values was not published but was available (CIC) in the unpublished Industry Trial Report. For week 56 changes in favour of adalimumab relative to placebo were reported for mean IBDQ and CDAI scores.

Table 25. IBDQ scores, CDAI scores and CRP concentrations reported in CLASSIC II

	Placebo N=18	Adalimumab 40 mg eow N =19	Adalimumab 40 mg weekly N =18	P
mean IBDQ score‡				
Week 0†	187.5	181	191.5	
Week 4	188.5	187	191	
Week 8	178	181	187	
Week 12	170.5	182.5	189	
Week 16	172.5	181	182	
Week 20	170.5	177	186.5	
Week 24	167.6	176.3	192.2	< 0.005 [a b]
Week 32	166.5	182	192	<0.05[a] < 0.005[b]
Week 40	167	179	188	< 0.005 [a b]
Week 48	163.5	178	183.5	
Week 56	162.4	178.4	185.6	
CDAI: mean change (*95% CI*) from baseline in CLASSIC I ††				
Week 56	-119.6 (*-74 to -65.1*)	-158 (*-202 to -99.8*)	-197.7 (*-248 to -147*)	< 0.005 [a b]
CRP concentration mg/dl: median (*range*) [Levels of CRP below 0.88 mg/dl are considered normal]				
24	0.5 (*0 to 1.2*)	0.4 (*0 to 1.9*)	0.1 (*0 to 1.6*)	NR
56	0.4 (*0 to 0.9*)	0.3 (*0 to 2.8*)	0.3 (*0 to 1.2*)	NR

‡ Data read from graph except for weeks 24 and 56.; last observation carried forward; the number observations at weeks 24 and 56 [redacted] for the placebo, eow and weekly groups respectively. [a] Comparison eow adalimumab versus placebo. [b] Comparison weekly adalimumab versus placebo. † week 4 of CLASSIC I. †† Last observation carried forward number of observations [redacted] for placebo, eow, and weekly groups respectively; CLASSIC I baseline CDAI scores for these patients not reported.

At the start of CLASSIC II 49% of patients were receiving systemic steroids or budesonide; 7 of the placebo group, 7 of the eow adalimumab group, and 8 of the weekly adalimumab group. Using last observation carried forward it was reported that by week 56 the number that had discontinued steroids was 4 in both the placebo and eow adalimumab groups, and 7 in the weekly adalimumab group.

E. Other considerations - Open label study

Most patients from CLASSIC I that did not qualify for CLASSIC II participated in an open label study in parallel with CLASSIC II. The results reported were not randomised comparisons and are outwith the inclusion criteria for this report.

Quality assessment (based on published report)

Randomisation, allocation concealment and blinding were adequate. Baseline characteristics at week 0 were well balanced. The study was powered for the primary outcome (remission at week 4) of CLASSIC I and no further power calculation was conducted for CLASSIC II. The number of patients who withdrew was reported; 5 of 18 placebo patients withdrew and 5 of 37 patients given adalimumab withdrew. There were 32 patients (58%) that completed to 56 weeks of double blind follow up. The last observation was carried forward as necessary for continuous outcomes but the amount of missing data was not reported.

CLASSIC II. Summary of effectiveness evidence.

The trial population (N=55) was recruited from responders in the previous CLASSIC I adalimumab induction trial (N=299). Only responders with a strong response (remission for at least a month) were selected; they had received various induction dose regimens.

Maintenance injections of 40 mg of adalimumab administered weekly or every other week generated a statistically significant greater proportion of patients in remission at week 56 than did placebo (frequency of administration not reported). About half of the placebo group and 81% of those who received infliximab were in remission at week 56. Point estimates of response rates were associated with considerable uncertainty due to the small size of the trial. There were no statistically significant differences in effectiveness between every other week and weekly adalimumab regimens.

CHARM (adalimumab)

This was a free-standing maintenance trial (i.e. newly started).[62] There were 854 enrolled patients (CDAI range 220 to 400) of whom 130 (15.2%) had fistulas at screening and baseline. An induction regimen consisting of a 80 mg injection of adalimumab at week 0 and a 40 mg injection two weeks later was followed by randomisation of 778 patients at week 4 to one of three arms as follows: placebo to week 56 (N=261), 40 mg adalimumab eow to week 56 (N=260), and 40 mg adalimumab weekly to week 56 (N=257). There were 76 (8.9%) withdrawals prior to randomisation. Assessment visits were planned for weeks 0, 2, 4, 6, 8, 12, 16, 20, 26, 32, 40, 48, 56 and 60.

At week 4 patients were classified as responders or nonresponders. Responders had to have a reduction of ≥ 70 CDAI points relative to baseline. Of the 854 patients given the induction regimen, 499 (58%) were categorised as responders and were the focus of the published effectiveness results. This population was different to that followed up in the other adalimumab maintenance trial CLASSIC II in that the latter were on average better responders having achieved remission from induction. The numbers of responders randomised to the three trial arms of CHARM were: 170 to placebo; 172 to adalimumab every other week (eow); 157 to adalimumab weekly.

The coprimary outcome measures were designated: the percentage of week-4 responders who achieved remission at weeks 26 and 56. Pre-specified secondary outcomes included: 1) percentage achieving response 70 and response 100 at weeks 26 and 56; 2) Change in IBDQ score from baseline at weeks 26 and 56; 3) percentage achieving clinical remission at weeks 26 and 56 who were able to discontinue corticosteroid use; 4) percentage achieving clinical remission at weeks 26 and 56 who were able to discontinue steroids for ≥ 90 days; 5) percentage of patients with fistula remission (closure of all fistulas that were draining at screening and baseline visits); 6) Median time in clinical remission among randomised responders achieving remission. Post hoc analyses examined subgroup responses and sustainability of response.

At or after week 12 patients with disease flare[a] or sustained nonresponse[b] were eligible to cross over to 40 mg adalimumab eow which coud be escalated to 40 mg weekly for patients with continued nonresponse or recurrent flare. For the primary effectiveness outcome (responders) any patients who crossed over were counted as remission failures.

A. Response 70 and B. Response 100

The published response 70 and response 100 rates at weeks 26 and 56 are summarised in Table 26 . Rates reached statistical significance in favour of adalimimab for both dose regimens at both time points.

[a] Flare was defined as an increase of ≥ 70 CDAI points from that at week 4 and a CDAI score > 220.

[b] Nonresponse was defined as a CDAI score not reduced by ≥ 70 points from week 0.

Table 26. Reported response 100 and response 70 rates in CHARM

Dose regimen (N)	Number with response 100 (%; *95% CI*)		P*
	Week 26	Week 56	
Placebo (170)	45 (26.5%; *20 to 34*)	28 (16.5% *12 to23*)	NA
40 mg adalimumab eow (172)	89 (52%; *44 to59*)	71 (41% *34 to 49*)	< 0.001
40 mg adalimumab weekly (157)	82 (52%; *44 to 60*)	75 (48% *40 to 56*)	< 0.001
	Number with response 70 (%; *95% CI*)		
	Week 26	Week 56	
Placebo (170)	48 (28%; *22 to 35*)	30 (18% *13 to 24*)	NA
40 mg adalimumab eow (172)	93 (54%; *47 to 61*)	74 (43%36 *to 50*)	< 0.001
40 mg adalimumab weekly (157)	88 (56%; *48 to 63*)	77 (49% *41 to 57*)	< 0.001

* Adalimumab versus placebo; χ^2 test

The unpublished Industry Trial Report for CHARM provided (CIC) values for response 70 at time points for all assessment visits. These are summarised in Figure 20. [redacted]

Figure 20. Response 70 rates amongst responders in CHARM

[redacted]

Similar CIC results were observed for response 100 and are summarised in [redacted]

Figure 21. Rates of response 100 amongst responders in CHARM

[redacted]

C. Remission

The primary outcome was the proportion of patients in remission at weeks 26 and 56. The results are summarised in Table 27 . The difference between adalimumab groups and placebo reached statistical significance in favour of adalimumab for both dose regimens.

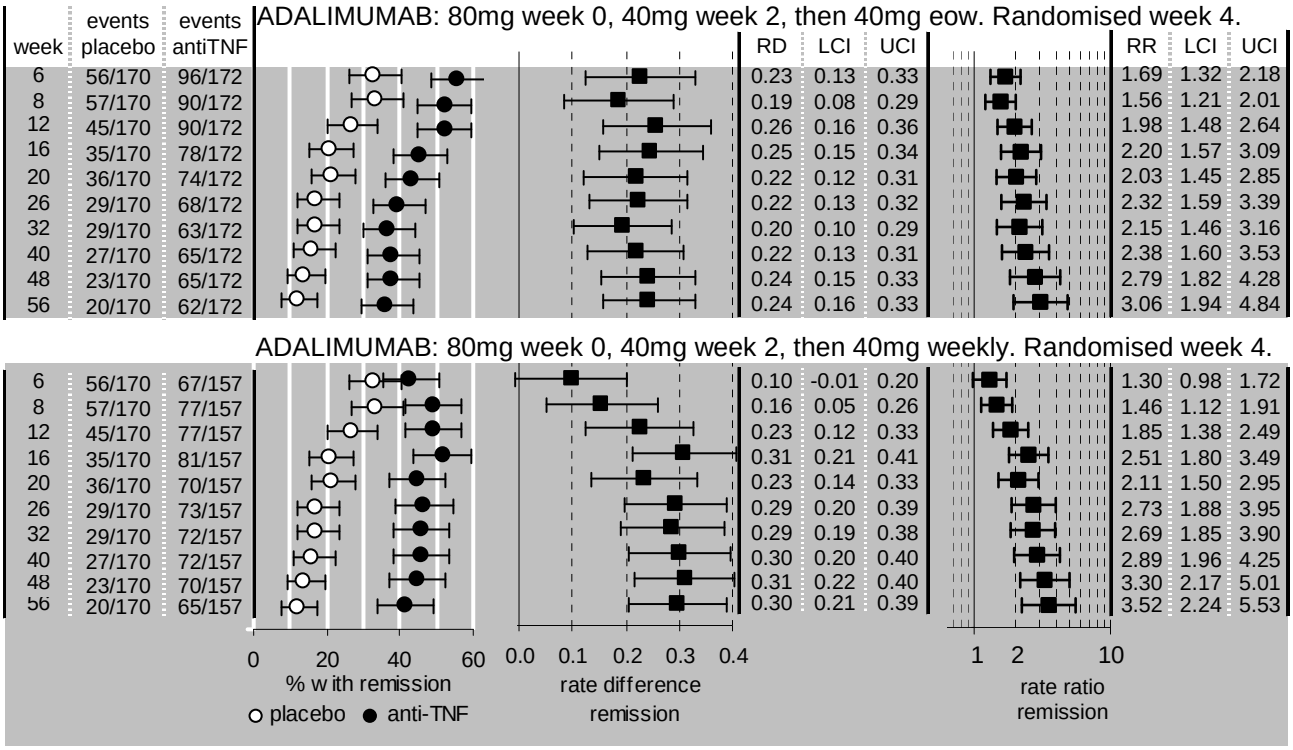
Table 27. Remission at weeks 26 and 56 in CHARM

Dose regimen (N)	Number in remission (%; *95% CI*)		P*
	Week 26	Week 56	
Placebo (170)	29 (17%; *12 to 23*)	20 (12% *8 to 13*)	NA
40 mg adalimumab eow (172)	86 (40%; *32 to 47*)	62 (36% *29 to 43*)	< 0.001
40 mg adalimumab weekly (157)	73 (47%; *39 to 54*)	65 (41% *34 to 49*)	< 0.001

* Adalimumab versus placebo; Cochran-Mantel-Haenszel χ^2 test adjusting for prebvious anti-TNF use.

The secondary outcomes of remission rates for each follow up visit to week 56 are summarised in Figure 22. Rate differences (adalimumab – placebo) and rate ratios (adalimumab/placebo) reached statistical significance in favour of adalimumab at all time points after week 6. Rates of remission in the adalimumab eow arm diminished through follow up. From week 12 to 16 onward, rate differences remained stable so that most benefit of the intervention appeared to be delivered in the first quarter of the trial. The rates reported were group point prevalence values and do not reflect maintenance of remission at the patient level. The difference in rates between the two adalimumab regimens at week 56 was not significant (rate difference p = 0.32, rate ratio p = 0.32).

Figure 22. Remission rates reported during follow up in CHARM



At week 56 rate difference and rate ratio for both regimens of adalimumab v. placebo p = <0.0001

Patient level maintenance of remission was published for weeks 26 to 56. In the adalimumab arms, 81% of patients in remission at week 26 sustained remission to week 56; this represented 114 patients and 27% of all those randomised to adalimumab. For patients

randomised to placebo 48% of those in remission at week 26 sustained remission to week 56. This represented 14 patients and 5% of all those randomised to placebo. The median time in clinical remission that started at any time was 127 days for placebo group, 378 days for adalimumab eow group and > 392 days for the adalimumab weekly group ($p = 0.002$ and $p< 0.001$ versus placebo respectively). Over 56 weeks it was possible for a patient to enter a remission state on several occasions. The publication did not make clear which occasion(s) were used in the analysis or how and if double counting was avoided.

D. Other outcomes

Published mean CDAI and IBDQ scores are summarised in Table 28 . No variance information was provided. After week 12, CDAI and IBDQ scores for patients who crossed over to increased adalimumab doses were included in the calculation of group mean scores although this was not made explicit. Mean CDAI scores decreased, and mean IBDQ scores increased, to a greater degree respectively in the adalimumab groups than the placebo group. Since a true placebo group did not exist after week 14 the results thereafter are difficult to interpret. Last observation was carried forward; the proportion of patients evaluated at week 56[redacted]

Table 28. Group mean CDAI and IBDQ scores reported for responders in CHARM

	CDAI :mean *				IBDQ median *			
week	placebo N=170	40 mg adalimumab eow N=172	40 mg adalimumab weekly N=157	P	placebo N=170	40 mg adalimumab eow N=172	40 mg adalimumab weekly N=157	P
0	318	317	310	NR	125	128	123	NR
2	215	200	203	NR	NR	NR	NR	NR
4	170	153	162	NR	166.6	174	165	NR
6	178	150	162	NR	NR	NR	NR	NR
8	183	147	155	NR	NR	NR	NR	NR
12	[redacted]			NR				NR
16				NR				NR
20				NR				NR
26				NR				NR
32				NR				NR
40				NR				NR
48				NR				NR
56				NR				NR

* Data read from graph. Values after week 12 were calculated including values for patients who crossed over to increased adalimumab.

From week 8 the responder patients who at baseline were receiving steroids could begin reducing steroide use (presumably at physician's discretion). This involved 66 placebo patients, 58 and 74 patients respectively in the adalimumab eow and adalimumab weekly groups. The percentage of these patients who were in remission at week 26 and who had discontinued steroids was 3% (2/66) in the placebo group, 35% (20/58) and 30% (22/74) and in the adalimumab eow and weekly groups. Corresponding percentages at week 56 were 6%, 29% and 23% respectively. The percentages who were in remission at week 26 and who were steroid free for at least 90 days were 3% in the placebo group and 19% and 15% in the

adalimumab eow and weekly groups. Corresponding percentages at week 56 were 5%, 29% and 20% respectively.

Hospitalisation rates

Details on hospitalisation rates from the CHARM trial[62] were reported in the industry submission, referenced to published abstracts by Wu 2007[65]and by Feagan 2007[66]. The latter abstract reports the hospitalisation rates in the placebo arm and the combined adalimumab arms, which were 22.4% and 14.0% respectively. The 56 week actuarial CD-related hospital admission rates for the placebo and for the combined adalimumab arms were 13.9% and 5.9% respectively. A difference in relative risk was apparent at two weeks after randomisation and placebo patients had 4.5 times the risk of hospitalisation at month 3 compared to adalimumab patients. Wu (2007)[65] used a Cox proportional hazard regression model and found that lower CDAI scores were associated with a decreased risk of hospitalisation and CD related hospitalisation. Simulated one-year rates indicated that a 70-point reduction on the CDAI throughout the follow-up period reduced all-cause hospitalisation risk by 28.3% and CD related hospitalisation by 36.5% at year-end. Further simulations indicated that remission was associated with a 43.7% decrease in the one-year risk of all-cause hospitalisation and a 60.3% decrease in CD related hospitalisation.

E. Other considerations - Subgroup analyses and cross-over issues

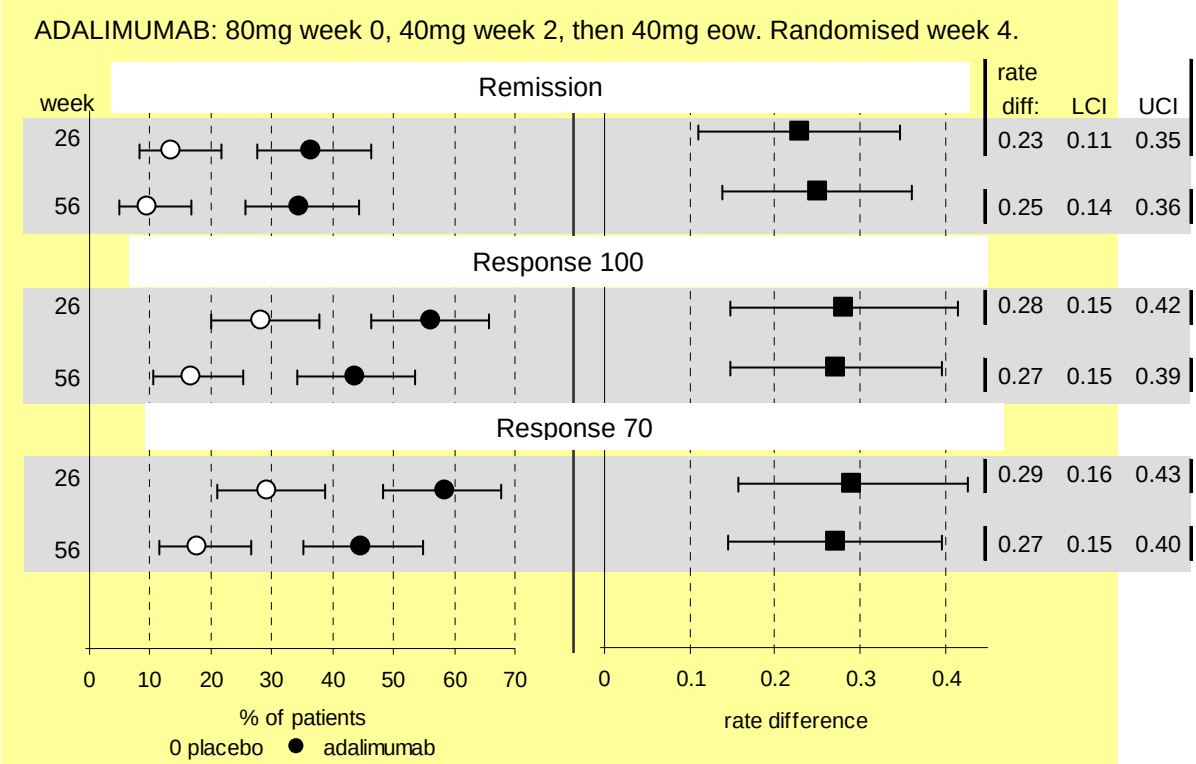
Outcomes for patients with draining fistulas are included in the next section.

The manufacturer's submission to NICE provided week 26 and 56 results for placebo and eow adalimumab group patients who had severe disease at baseline (CDAI > 300). Results for all severe patients and for severe week-4 responders were provided allowing calculation of results for nonresponders with severe CD (Table 29). There were 96 severe CD patients in both placebo and eow adalimumab groups. Rates of remission, response 70, and response 100 are summarised in Figure 23. Remission rates at week 56 in adalimumab and placebo arms were 35% and 10% respectively; higher rates were recorded for the less stringent response 70 and 100 outcomes. These rates were similar to those reported for all week 4 responders (within 5%; listed in Table 29). The rates in the eow arm for nonresponders with severe CD were about half those for week 4 responders CD

Table 29. Response rates for severe CD patients in CHARM

		All week 4 responders		Severe CD responders		All severe CD patients		Severe CD nonresponders
OUTCOME	week	placebo N=170	eow anti-TNF N=157	placebo N=96	eow anti-TNF N=96	placebo N=149	eow anti-TNF N=135	eow anti-TNF N=39
Remission	26	17%	**40%**	14%	**36%**	11%	**30%**	**15%**
	56	11.8%	**36%**	9%	**33%**	8%	**27%**	**8%**
Response 100	26	26.5%	**52%**	28%	**56%**	21%	**47%**	**26%**
	56	16.5%	**44%**	17%	**44%**	13%	**36%**	**18%**
Response 70	26	28%	**54%**	29%	**58%**	23%	**47%**	**28%**
	56	17.6%	**43%**	18%	**45%**	13%	**36%**	**18%**

Figure 23. Response and remission rates for severe disease responders in CHARM



Other post hoc subgroup analyses

Several post hoc analyses explored the effectiveness of adalimumab amongst subgroups of patients defined according to various criteria including: baseline C-reactive protein level greater or less than 1 mg/ml; concomitant treatment with or without immunosuppressant medication; previous experience of anti-TNF therapy or no previous experience. No statistically significant subgroup differences in adalimumab effectiveness were observed.

Premature withdrawal from treatment and cross-over due to worsening disease

The published information about withdrawal from treatment and cross over to open label therapy was difficult to disentangle. The Industry Trial Report provided fuller detail. Of 499 responders 29% (144) withdrew prematurely: [redacted] of the placebo group [redacted] of the

eow adalimumab group [redacted] and [redacted] of the weekly adalimumab group [redacted]. The Industry Trial Report stated the overall premature discontuation rate amongst all patients was [redacted], with 79 of these occurring before randomisation. Amongst all 788 randomised patients withdrawals during the randomised phase were [redacted] in placebo group, [redacted] in the adalimumab eow group, and [redacted] in the adalimimab weekly group, giving an overall rate of [redacted]

Cross over to open label treatment after week 12 involved [redacted] of patients randomised to placebo, [redacted] of those randomised to adalimumab eow, and [redacted] of those randomised to adalimumab weekly. These numbers represented patients experiencing worsening disease by flare or discontinued response. Transfer to open label for patients in the weekly adalimumab group involved continuation of the same dose regimen (since cross over was described as "*..switched to open label treatment with 40 mg adalimumab eow…. escalated to 40 mg weekly for those with continued nonresponse or recurrent flare"* After cross over *"..continued nonresponse with open-label 40mg weekly dosage resulted in withdrawal"* [62] however there was no published information about how long the state of flare or nonresponse was allowed to continue before withdrawal was implemented. The number of responder patients that crossed over to open label was not published. The Industry Trial Report allowed calculation of cross overs and withdrawals amongst all randomised patients; this information is summarised in Figure 24.

Figure 24. Withdrawals from treatment and cross overs for flare or nonresponse in CHARM

*There was a discrepancy concerning one patient in the values for the adalimumab weekly group

Quality assessment (based on published report of Colombel 2002)[62]

Randomisation, allocation concealment and blinding were adequate. Baseline characteristics were only reported for all patients, for all responders and for all non-responders, not for each of the trial arms. It was therefore not possible to judge if baseline characteristics were evenly balanced between the three arms of responders (placebo and eow or weekly adalimumab groups) that were analysed for effectiveness outcomes. The frequency of placebo injections was not documented. Information about patients who withdrew was reported. After week 12, patients with disease flare or nonresponse were allowed to cross over to open label treatment. It was difficult to determine how many responders and how many randomised patients in each group crossed over to open label treatment. There was no statement defining how long after cross over flare or nonresponse was allowed to continue before withdrawal was implemented. Where necessary, the nearest or last observation was carried forward for continuous outcomes but this was not stated explicitly and the amount of missing data was not reported. A power calculation was conducted and based on the primary analysis of 4-week responders achieving remission at weeks 26 and 56.

The published text stated "..*secondary efficacy analyses were conducted for all treated patients, including both randomised responder and randomised nonresponder groups (all randomised patients who failed to achieve a clinical response at week 4)*".[62] Although this might be technically correct, in the sense that analyses were conducted, it is misleading because the results of these analyses were not reported with the single exception of data on healing of fistulas for a subgroup of patients with fistulas at baseline and screening.

CHARM. Summary of effectiveness evidence.

778 patients given induction injections of 80mg and 40mg of adalimumab separated by 2 weeks were randomised at week 4 to maintenance therapy with placebo or 40mg adalimumab every other week or weekly. Only results for responders were published. Responders were defined as patients who at week 4 had a CDAI score reduced by ≥ 70 points from baseline. At weeks 26 and 56 there were significantly more responder patients in remission in the eow and weekly adalimumab groups than the placebo group, 40% and 47% respectively versus 29% at week 26, and 62% and 65% respectively versus 20% at week 56 ($p < 0.001$ for adalimumab v. placebo). The rate difference (adalimumab – placebo) for remission reached statistical significance in favour of adalimumab from week 8 onwards and remained stable from about week 12 or 16 to the end of follow up (week 56), indicating that most of the benefit from active intervention was delivered during the first quarter to third of the trial.

The proportion of responders (response 70) had diminished to less than 50% in all groups by week 56. Response 70 rates diminished [redacted] delivered during the first part of the trial. Premature withdrawal from randomised treatments (adalimumab and placebo) was [redacted]; withdrawal rate from active intervention (adalimumab) was [redacted] for responders [redacted] as nonresponders [redacted]. Amongst the whole trial population randomised to adalimumab maintenance therapy, [redacted] crossed over to open label treatment due to flare or nonresponse. The distribution of cross-overs between responders and nonresponders was unclear.

Pooling and indirect comparisons

The two adalimumab trials, CLASSIC II and CHARM, differed fundamentally with respect to populations analysed for outcome results. CLASSIC II reported results for responders who had achieved remission whereas the responders in CHARM had achieved only the less stringent response of a 70 point reduction in CDAI score. It would be inappropriate to combine the results from these two trials. It is relevant that the manufacturer's submission for adalimumab did not adopt a pooling approach. In a recent Cochrane review[67] the authors stated "*the two studies evaluating adalimumab were evaluated separately due to heterogeneity among the two trials* (i.e. CLASSIC II and CHARM)". Surprisingly the results section of the review provided pooled results for remission (random effects model) and a further different pooled result (which may have been fixed effects) was presented in the discussion. On contacting the authors regarding these inconsistencies, we have been informed that the review will be amended and the modified version made available in the Cochrane Library in July 2008.

The two infliximab trials Rutgeerts 1999 (extension from Targan 1997) and ACCENT I both employed a 10 mg/kg infliximab maintenance therapy arm and both reported results for responders based on a CDAI score reduced from baseline by ≥ 70 points. Therefore there is potential for pooling results. However the pre-maintenance "induction" phases of the two trials were very different so that the populations analysed for maintenance outcomes were likely to be quite different at the start of maintenance. Responders in ACCENT I were selected two weeks after a single exposure to a 5 mg/kg dose of infliximab. In contrast responders in Rutgeerts 1999 were selected between 8 and 12 weeks after their first exposure to infliximab and were required to have a response 70 lasting 4 weeks. A further considerable difference between the responders in the two trials was the degree of exposure to infliximab

prior to their selection as responders; in ACCENT I responders were defined after a single 5 mg/kg exposure, whereas Rutgeerts responders could have been exposed to any of the following: one 5 mg/kg, one 10 mg/kg, one 20 mg/kg, one 5 mg/kg and one 10 mg/kg, two 10 mg/kg, one 20 mg/kg and one 10 mg/kg, or no infliximab. The cumulative effect of these differences in responder population (up to 6 fold difference in exposure, different requirement in duration of response 70, and between 4 and 6 fold difference in duration of induction phase) is that the populations were unlikely to be sufficiently similar for the pooling of results to be informative. The current version of the recent Cochrane review reports pooled results for these trials but the modified version planned for July 2008 may well not.

Indirect comparisons between the placebo controlled maintenance trials so as to gain an estimate of relative effectiveness of the two anti-TNF agents was not undertaken for this non-fistulising adult population. Indirect comparison requires that trials for different interventions of interest share a common comparator arm. For the maintenance trials the differences between "placebo" groups were numerous and not easily quantifiable; different induction drugs were administered on differing numbers of occasions for different periods of time, followed by selection of responders by differing criteria representing different proportions of the randomised populations. The basis of indirect comparison depends on strict comparability of the trial arms common to the compared trials (in this case the placebo arms). In these circumstanes indirect comparison would be misleading and unjustified. It is noteworthy that neither of the manufacturers' submissions performed formal indirect comparison based on these trials.

5.2.2.3 Trials recruiting patients with fistulas

Two trials, Present 1999[57] an induction trial, and ACCENT II[60] a maintenance trial, compared infliximab to placebo for adults with fistulising CD. There were no trials of adalimumab that enrolled only from this patient group. In these two trials all patients had one or more fistulas at the time of randomisation and the main outcome measures focused on the status of fistulas during follow up. The outcomes measured are listed in Table 30 and the main trial characteristics summarised in Table 31. For reference purposes this section also includes fistula status results for the small subgroups of adult patients that had fistula in other trials

Table 30. Outcomes measured in trials of fistulising CD

	% achieving 70- point response on CDAI	CDAI score (mean or median)	IBDQ score (mean or median)	PDAI score (mean or median)	Response: Reduction of 50% or more of draining fistula	Complete response: absence of draining fistula	Additional outcomes
Infliximab							
Present 1999[57]	X	√	X	√	√ (at 2 or more consecutive visits)	√ (at 2 or more consecutive visits)	Time to beginning of response; duration of response
ACCENT II[60]	√ (sub-group only)	√	√	X	√ time until loss of response (at consecutive visits >4 weeks apart)	√	Subsequent response amongst previous non-responders, response rate in patients who lost response and crossed over

Table 31. Main study and population characteristics for trials in fistulising adult populations

Study ‡ *Drug*	Study wks N	Population: severity of CD (baseline PDAI*, CDAI and IBDQ if stated)	Intestinal areas affected; Fistula: N & location	Main concomitant medication.**	Previous/concomitant treatment with anti-TNF inhibitors	Intervention and comparator (dosing regimen)
Present et al., 1999[57] *Infliximab*	18 94	≥ 1 draining abdominal or perianal fistula of ≥ 3 months duration. Mean baseline PDAI (IQR): 9 (7-10.5) placebo; 8 (7-10), 10 (8-12) infliximab groups. Mean baseline CDAI (SD): 193 (92) placebo; 184 (98), 185 (97) infliximab groups. (IBDQ not stated)	Mainly ileum & colon, also ileum only, & colon only. Fistula: 1, 45%; > 1, 55%; mainly perianal fistula, a few abdominal.	Aminosalicylates (mainly), also mercaptopurine or azathioprine, corticosteroids & antibiotics	Exclusion criterion: Infliximab within 3 months of study; no further details.	Intravenous infusions of placebo, 5mg/kg infliximab or 10mg/kg infliximab at weeks 0, 2 and 6. Study visits at least every 21 days; total follow up not stated.
Sands et al., 2004[60] **ACCENT II** *Infliximab*	54 282	≥ 1 draining abdominal or perianal fistula of ≥ 3 months duration. PDAI scores not stated. CDAI at baseline: 60% ≥150, 33%≥ 220; Median baseline IBDQ (IQR) (responders): 168 (145-193) placebo, 155 (135-187) infliximab 161. *[136-176 (non-responders)]*	Mainly ileum and colon, also ileum only and colon only. Fistula: 1, 44%; > 1, 56%; mainly perianal fistula, some abdominal or recto-vaginal.	Aminosalicylates (mainly), also mercaptopurine or azathioprine, corticosteroids and antibiotics, few methotrexate. Previous medication: mercaptopurine / azathioprine (mainly) & antibiotics, some cyclosporine, tacrolimus or methotrexate	Exclusion criterion: previously treated with infliximab	Intravenous infusions of 5mg/kg infliximab at weeks 0, 2 and 6 for all patients; at week 14 responders and non-responders randomised to placebo or 5mg/kg infliximab at weeks 14, 22, 30, 38 and 46.

‡ Both studies were industry sponsored multicentre trials conducted in US, Canada and Europe. * PDAI = perianal disease activity index.** % not on any medication not stated.

Present 1999[57] (infliximab)

Present 1999[57] was a small study that randomised 31 patients to placebo, 31 and 32 patients respectively to 5 mg/kg or 10 mg/kg infliximab infused at weeks 0, 2 and 6. Follow up extended to at least week 18 with assessment visits every four weeks from week 2 onward.

The primary outcome was a greater than 50% reduction in the number of draining fistulas relative to baseline evaluated by physical evaluation and observed over at least two consecutive study visits at any time during the trial. Secondary outcomes included: complete absence of draining fistula observed over at least 4 weeks (i.e. across at least two consecutive study visits) at any time during the study, time to beginning of response, and duration of response. Changes in CDAI and PDAI scores were reported for some patients.

The results for the primary outcome (50% reduction in draining fistula occuring at any time over at least 2 consecutive clinic visits) and for complete absence of draining fistula over two consecutive clinic visits are summarised in Figure 25. For both these outcomes infliximab at both dose regimens was more effective than placebo (P = 0.002 and P = 0.02 for 5 and 10 mg/kg regimens respectively). The point estimates for response rates were associated with substantial uncertainty because of the small group size; for the combined infliximab groups the response rate was 62% (95% CI 50% to 73%) compared to 26% (95% CI 14% to 43%) for the placebo group (P < 0.001). For those with a response the median time to response was 6 weeks in the placebo group and 2 weeks in the infliximab groups (see Table 32)

Figure 25. Rates and rate differences for 50% reduction and absenceof draining fistulas

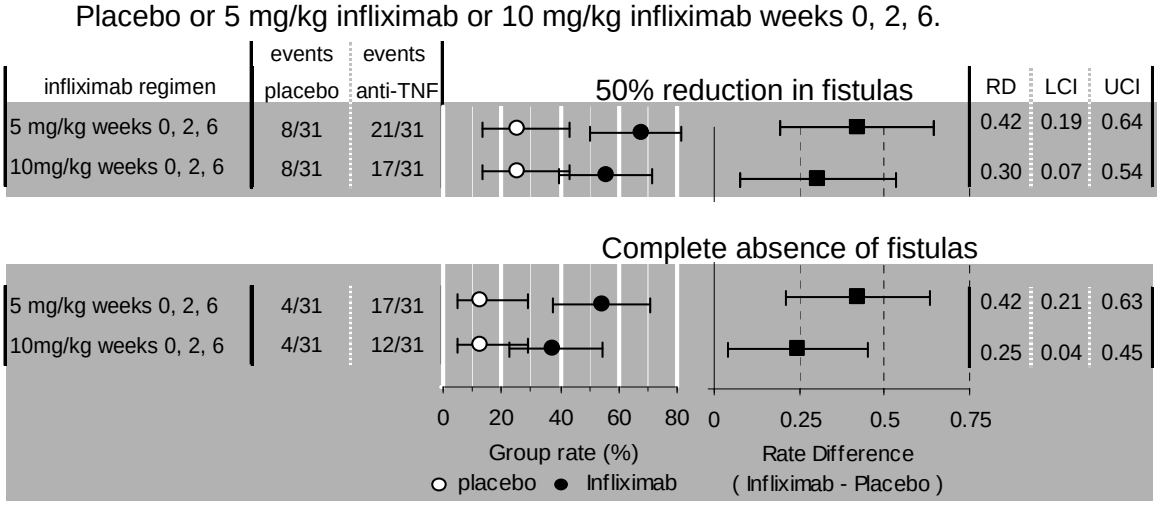


Table 32. Time to onset of primary outcome

Length of time to the beginning of a response‡ (days: MEDIAN & interquartile range)				
	Placebo n=8	5 mg/kg. n=21	10 mg/kg. n=18	5 or 10mg/kg n=39
MEDIAN [days] interquartile range	42 15–72	14 14–42	14 14–42	14 14–42
P vs placebo	NA	NR	NR	NR
‡ Only patients with primary response included. NR = not reportred. NA = not applicable				

The median duration of response (defined as the maximum period during which the patient experienced a 50% reduction in draining fistulas) was approximately three months. For infliximab patients, 29/63 (46%; 95% CI 34% to 58%) experienced complete absence of draining fistulas for at least two consecutive clinic visits compared with 4/31 (13%; 95% CI 5% to 29%) of patients in the placebo group (p <0.001).

The median CDAI and PDAI scores reported for baseline and weeks 2 and 18 of follow up are summarised in Table 33 . By week 2, statistically significantly better (i.e. lower) scores were found for the infliximab groups compared to the placebo group. The statistical significance of the difference between groups had weakened or disappeared by week 18. Not all patients contributed data for the analyses (i.e. this was not an intention to treat analysis).

Table 33. CDAI and PDAI scores reported in Present trial

	PLACEBO *N = 25*			Infliximab 5mg/kg weeks 0, 2, 6 *N = 27*			Infliximab 5mg/kg weeks 0, 2, 6 *N = 27*		
Outcome	week	Median	IQR	Median	IQR	P†	Median	IQR	P†
CDAI score‡	0	162	126-265	163	99-284	0.71	203	112–254	0.66
	2	171	114-252	108	83-203	0.04	111	89–164	0.06
	18	160	72-206	104	47-177	0.23	123	58–175	0.32
PDAI score‡	0	9	7-10.5	8	7-10	0.69	10.0	8.0–12.0	0.31
	2	8	6-10	6	3-7	0.02	6.0	4.0–8.0	0.04
	18	7	4-9	4	1-7	0.05	5.0	3.0–8.0	0.14
† anti-TNF v placebo using analysis of variance procedure. ‡ Last observation carried forward for missing values. CDAI= Crohn's disease activity index. PDAI = perianal disease activity index									

Quality assessment based on published report[57]

Randomisation and blinding were adequate and allocation concealment was likely to have been adequate. Baseline characteristics were generally well balanced although there was a greater proportion of patients in the infliximab groups that had had previous segmental resections compared to the placebo group. Draining fistulas of less than three months duration were excluded from the primary analysis. However the number or frequency of these fistulas was not reported, and it was unclear if these were also excluded from the secondary outcome of a complete absence of a draining fistula. Total follow up time for the primary outcome was unclear. No power calculation was performed. Last observation was carried forward for CDAI and PDAI analyses but the amount of

missing data was unclear. There were only six premature withdrawals from treatment, four from the placebo group and one patient from each of the infliximab groups.

PRESENT. Summary of effectiveness evidence.

Patients with one or more draining fistula of more than three months' duration, and an unreported number of fistulas of less than three months' duration, were randomised to placebo or 5 mg/kg infliximab or 10 mg/kg infliximab, by IV infusion at weeks 0, 2 and 6. More patients in the infliximab groups than in the placebo group achieved the primary outcome defined as: a reduction in the number of three month-duration draining fistulas present at baseline by at least 50% lasting for at least two consecutive clinic visits. The percentage of patients responding to infliximab was 62% (95% CI 50% to 73%) compared to 26% (95% CI 14% to 43%) for the placebo group (p <0.002). The median time to response was two weeks for infliximab groups and six weeks for placebo group. The duration of response was the same for both groups (median about 12 weeks).

More patients in the infliximab groups than in the placebo group achieved the secondary outcome of absence of draining fistula lasting for at least two consecutive clinic visits. The percentage of patients responding to infliximab for this outcome was 46% (95% CI 34% to 58%) compared to 13% (95% CI 5% to 29%) for the placebo group (p <0.001).

ACCENT II[60] (infliximab)

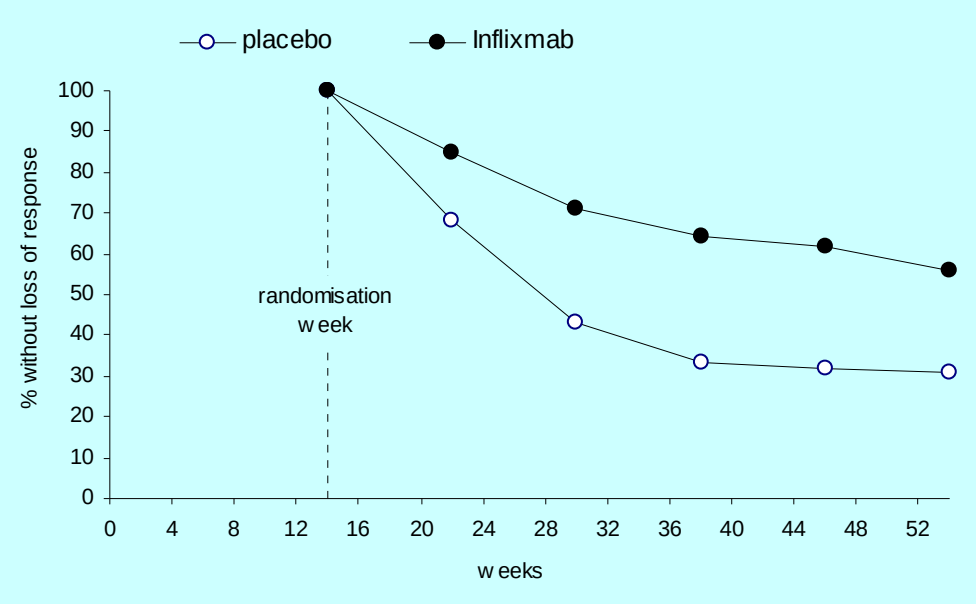
This was a maintenance trial that recruited 306 patients who had one or more fistulas of at least three months standing. Of the 306 enrolled patients 282 were assessed for "response" at week 14 after administration of infusions at weeks 0, 2 and 6 of 5 mg/kg infliximab. "Responders" were defined as those patients with at least 50% reduction in draining fistulas relative to baseline, observed at both weeks 10 and 14. Sixty nine percent (195 patients) were classified as responders. Both responders and non-responders were randomised to placebo (96 responders; 43 nonresponders) or to 5 mg/kg of infliximab (99 responders; 44 non-responders) which were administered at weeks 14, 22, 30, 38, and 46. Assessment visits were scheduled at weeks 0, 2, 6, 10, 14, 22, 30, 38, 46, and 54. After week 22 patients losing response could cross over to 5 mg/kg infliximab from placebo and from 5 mg/kg infliximab to 10 mg/kg infliximab. The fistula status outcome measures were:

- Loss of response; defined as a recrudescence of draining fistula or a change in therapy or a need for surgery or drop out due to lack of efficacy or a worsening of luminal disease activity.
- Response; defined as 50% reduction from baseline in draining fistula observed at consecutive visits 4 or more weeks apart.
- Complete absence of draining fistula.

Primary outcome

The primary outcome was designated as time to loss of response in responders. The results are summarised in Figure 26 .

Figure 26. Time to loss of response by responders in ACCENT II



Data taken from published graph and redrawn.

The median time to loss of response after randomisation was 14 weeks in the placebo group and more than 40 weeks in the infliximab group ($p < 0.001$ by log rank test). In the infliximab group, 42% of responders lost response, and in the placebo group, 62% lost response. The main reasons for loss of response in the primary outcome were: change in treatment (38% of placebo, 25% of infliximab) or recrudescence of fistula (22% placebo, 16% infliximab).

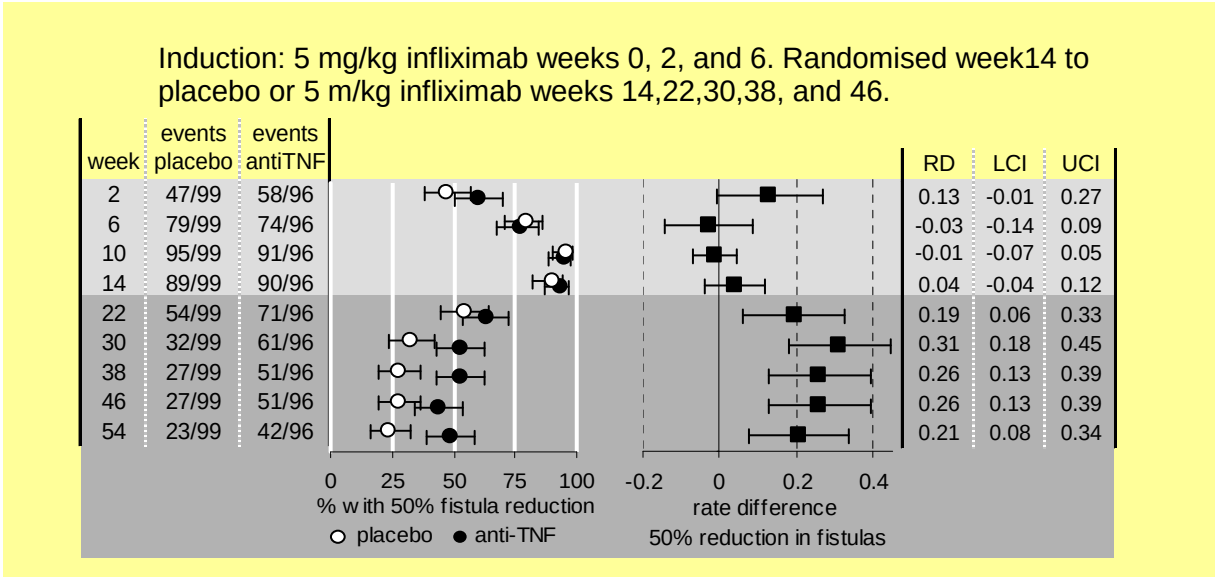
Response and complete response

At 30 weeks 33% and 64% of the placebo and infliximab groups respectively had a response (50% reduction in draining fistula from baseline for at least two consecutive visits), and at week 54 the corresponding percentages had diminished to 23% and 46% respectively ($p = 0.001$). The manufacturer's submission to NICE contained CIC information for additional weeks of follow up. These are summarised in Figure 27. Prior to randomisation, except at week 2, the rates were about equal as would be expected since all responder patients received identical induction therapy up to week 14 and baseline characteristics were well balanced. At week 2 a surprising difference between groups was observed with higher rate for the patients subsequently randomised to infliximab. At week 14 the placebo group did not receive infliximab. After randomisation at week 14, response rates diminished in both groups. From week 22 the rate difference (infliximab – placebo) reached statistical significance in favour of infliximab, after week 30 rate differences diminished indicating that most

benefit for maintenance of response from active intervention was delivered between week 14 and week 30. By week 30 the intervention group had received two extra infusions of infliximab compared to placebo patients.

Responders with loss of response during the post-randomisation phase were allowed to cross over after week 22 to an increased dose of infliximab. The renewed response rate in these cross over patients was reported as 25/41 (61%) in the placebo group (crossed over to 5 mg/kg dose) and 12/21 (57%) in the intervention group (crossed over to 10 mg/kg dose). However Figure 1 of the publication shows 50 crossovers from placebo and 28 from 5 mg/kg infliximab.[60]

Figure 27. Rates and rate differences for ≥ 50% reduction of draining fistulas in ACCENT II responders



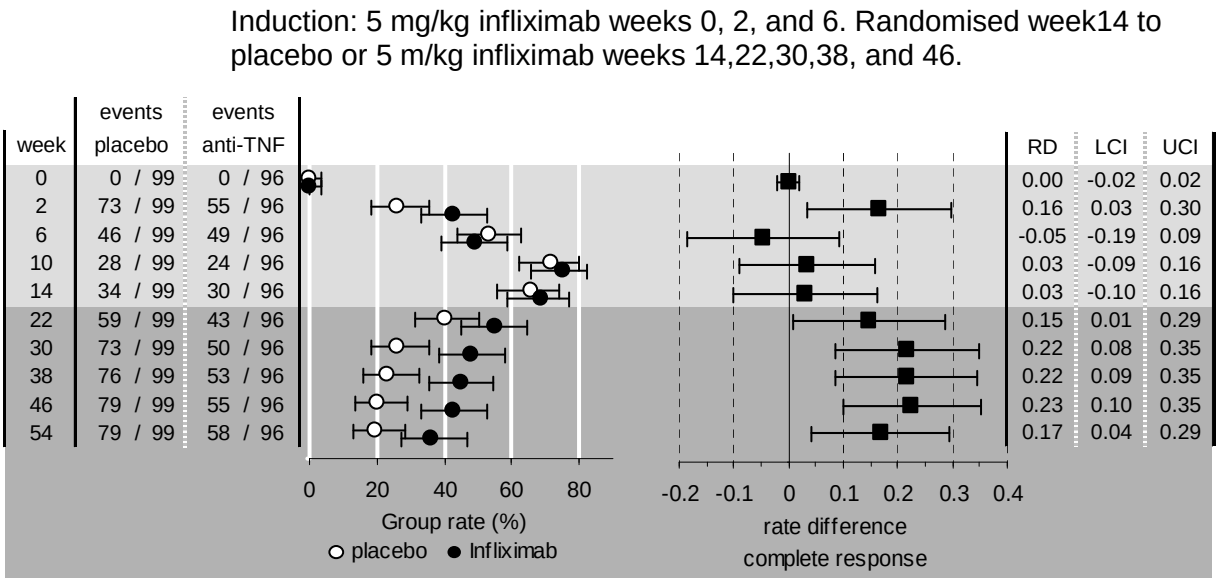
The published report provided information about the rates of "complete response" amongst responders.[60] A complete response was defined as a complete absence of draining fistulas. The definition for a response required ≥ 50% reduction in fistulas for at least four weeks, a "complete response" differed in that no minimum duration was specified. It was unclear, but likely, that this definition applied only to draining fistulas of at least three months' standing at baseline. The frequency of draining fistulas at baseline that were of less than three months standing was not reported. The results for a complete response are summarised in Figure 28.

At week 2, after only a single dose of infliximab 128/195 (66%) of patients already had a "complete response". Unexpectedly more patients who were subsequently randomised to infliximab had a complete response than those subsequently randomised to placebo (P = 0.014). By 14 weeks 66% and 69% of responder patients that were randomised to placebo and infliximab respectively had a complete response. The rate of complete response in responders diminished in both groups after week

14. From week 22 the rate differences (infliximab – placebo) reached statistical significance in favour of infliximab and from week 30 remained stable indicating that most benefit in maintenance of response from active intervention was delivered between week 14 and week 30.

The rate for a complete response amongst all enrolled patients at week 14 was reported to be 48% (147/306); this generated a 75% (147/[99+96]) complete response for responders at week 14 which according to published figure 2B corresponded to week 10 rather than to week 14.

Figure 28. Rates and rate differences for complete response amongst responders in ACCENT II



Hospitalisations and major surgery

The manufacturer's submission to NICE presented results for major surgery and for hospitalisation for all patients in ACCENT II whether they crossed over or did not cross over. For this purpose the placebo arm was termed "episodic treatment" and the infliximab arm "scheduled treatment". A 2.4 fold lower rate was reported for the scheduled treatment group. There were two important differences between these treatments. Firstly patients in the "episodic" arm experienced a four month mandatory withdrawal of active intervention (from weeks 6 to 22) not experienced by patients in scheduled treatment. Secondly after week 22 the "episodic" group patients were restricted to 5 mg/kg infliximab at episodes of worsening disease, where as the "scheduled treatment" group were able to receive 10 mg/kg. Restricted access to treatment (weeks 6 to 22) and restricted dosage represent biases likely to favour the "scheduled treatment" group for any comparisons after week 6. Furthermore the "episodic treatment" procedure was unlikely to reflect how an episodic strategy might be implemented in real world clinical practice, both with respect to the four month gap in active intervention and with regard to restriction of dose. Because of bias in the comparisons made and the probable dissimilarity between

the trial episodic treatment and likely clinical practice, it was considered here that the hospitalisation rate for the "episodic" treatment and the comparison with the scheduled treatment to be very approximate guides.

The considerations described above also apply to the values reported for the percentages of patients requiring major surgery (13% and 2% in "episodic" and scheduled treatment arms respectively).

Other outcomes reported for ACCENT II

The ACCENT II[60] published report presented the median decrease from baseline in CDAI score at weeks 30 and 54 for all patients. Improvements in median CDAI were statistically significantly greater for the infliximab group (p = 0.004). Median increases from baseline in IBDQ scores at weeks 30 and 54 were also significantly greater for the infliximab group than the placebo group. Baseline scores for all patients by group were not provided and baseline balance were therefore uncertain. In the case of missing values, the last observations were carried forward for the CDAI and IBDQ outcomes. The results are summarised in Table 34.

Table 34. Median CDAI and IBDQ changes ACCENT II[60]

	Median decrease in CDAI score from baseline			
WEEK	INFLIXIMAB *n=139*		PLACEBO *n= 143*	P infliximab v placebo
30	42		16	0.004
54	40		15	0.004
Median increase in IBDQ score from baseline				
30	14		4	0.002
54	10		5	0.03

Baseline scores were reported for responders by group, and for all nonreponders but not for all patients by group.

Further results for the ACCENT II[60] trial were presented in two separate papers.[68,69] One reported a *post hoc* analysis of the subgroup of responder patients with rectovaginal fistulas (11 received placebo and 14 received the 5mg/kg dose regimen of infliximab);[68] the other paper performed a *post hoc* analysis on incidence of abscess development in patients responding to infliximab with closure of fistulas.[69] The first of these papers was underpowered for firm conclusions to be drawn. In ACCENT II[60], cross-over to an increased dose of infliximab was allowed for all randomised groups (including placebo) from week 22 onward; this resulted in the mean dose of infliximab in the placebo group (quoted as 20 mg/kg) being approximately half that of the intervention groups (quoted as 40 mg/kg). The *post hoc* analysis for abscess development compared these two groups and reported no statistically significant difference in rates (15% vs 19%; p=0.526).

Quality assessment (based on published report Sands et. al. 2004[60])

Randomisation and blinding were adequate and allocation concealment likely to be so. Baseline characteristics for responders were well balanced between placebo and infliximab arms; however for the all-patient comparisons between infliximab and placebo arms (e.g. of change in IBDQ and CDAI scores relative to baseline) it was not possible to ascertain if groups were balanced at baseline. There was a lack of clarity in the methods section so that it was difficult to determine if the sentence ".. *Data for patients that crossed over from placebo to infliximab were censored before cross over occurred..*" referred to the survival analysis of loss of response. If it did, the reason for different handling of cross overs in the compared groups is difficult to interpret. The number of patients who withdrew prematurely was unclear except for discontinuation for adverse events. No power calculation was undertaken. The last observation was carried forward as necessary for continuous outcomes but the amount of missing data was not reported.

ACCENT II. Summary of effectiveness evidence.

After induction infusions of 5 mg/kg of infliximab at weeks 0, 2 and 6, 64% of enrolled patients were classified as responders. Responders were defined as patients experiencing at both weeks 10 and 14 a ≥ 50% reduction in the number of draining fistulas that were present at baseline of at least three months standing.

After week 14, the median time to loss of response by responder patients was greater for patients randomised to placebo than for those randomised to continued infliximab treatment of 5 mg/kg at eight-week intervals ($p < 0.001$). More responder patients randomised to inflixmab at week 14 experienced a response (closure of ≥ 50% of draining fistula for at least 4 weeks) than did responder patients randomised to placebo and from week 22 the rate difference (infliximab – placebo) reached statistical significance in favour of infliximab. After week 14, response rates diminished in both groups. From week 30, rate differences diminished indicating that most benefit from infliximab was delivered between week 14 and week 30.

Other trials reporting on subgroups of adults with fistulas

Two other trials reported on effectiveness of anti-TNF therapy for closure of fistulas - the GAIN induction trial of adalimumab[59] and the CHARM maintenance trial of adalimumab.[62] In the GAIN trial[59] at end of follow up (week 4) similar rates of fistula improvement were recorded for adalimumab and placebo groups (3/20 and 5/25 respectively). The CHARM trial[62] reported a measure termed "fistula remission" for the subgroup of trial patients that had fistula at screening and baseline. Fistula remission was defined as the percentage of patients with closure of all fistulas that were draining at screening and at baseline (separated by two weeks). Fistula remission was observed for 30% (21/70)

and 13% (6/47) of combined adalimumab groups and placebo group respectively at week 26 and for 33% (23/70) and 13% (6/47) respectively at week 56.

5.2.2.4 Paediatric CD trials

Patients in these trials were 18 years of age or less. Two trials, Baldassano 2003[43] and REACH (Hyams et al 2007[42]) looked at the effectiveness of different doses of infliximab in paediatric CD patients. There was no placebo arm in either trial. There were no trials of adalimumab in children. The outcomes measured are shown in Table 35 and the study characteristics are summarised in Table 36.

Table 35. Outcomes measured in trials of paediatric CD

	% of patients in remission (PCDAI score <10	% achieving response	PCDAI score (mean or median)	Additional outcomes
Infliximab				
Baldassano 2003[43]	√	√ decrease in CDAI of ≥ 70 points OR ≥ 10 points on PCDAI	√	Endoscopic lesion severity score (in consenting patients)
REACH[42]	√ (or: CDAI <150)	√ decrease in PCDAI of ≥ 15 points and total PCDAI score <30	√	IMPACT III score; % discontinuing corticosteroids; Change in height status (sub-group); Clinical response following crossover
PCDAI = paediatric CD activity index.				

Table 36. Main study and population characteristics: paediatric trials

Study & Sponsor	Country	Study length/size	Population: severity of CD (baseline CDAI and IBDQ if stated)	Intestinal areas affected	Main concomitant medication, % not on any medication	Previous/concomitant treatment with anti-TNF inhibitors	Intervention and comparator (dosing regimen)
Baldassano et al., 2003[43] Supported by Centocor	Multi-centre (US, Europe)	12 weeks n=21	Moderate to severe active disease despite previous treatment, PCDAI ≥30 or modified CDAI ≥200 Median PCDAI 56, 45, 41 infliximab groups (no placebo group), median modified CDAI score 455, 317, 312 (IBDQ not stated)	Mainly ileum and colon, also colon only and gastroduodenal	Mainly aminosalicylates and corticosteroids, also mercaptopurine or azathioprine, antibiotics, few methotrexate % not on medication (if any) not stated	No details	Single intravenous infusion of 1 mg/kg infliximab, 5 mg/kg infliximab or 10 mg/kg infliximab over at least 2 hours at week 0 (no placebo group)
Hyams et al., 2007 REACH[42] Supported by Centocor	Multi-centre (US, Canada, Europe)	54 weeks n=103	Moderate to severe CD PCDAI > 30 at baseline Mean baseline PCDAI (SD) 42.1 ± 9.2 and 40.1 ± 6.8 (infliximab groups, no placebo group) (no other baseline measures)	Mainly colon and/or ileum, also upper tract	Mainly mercaptopurine or azathioprine, also aminosalicylates and corticosteroids, few methotrexate % not on medication (if any) not stated	Exclusion criteria: Previously treated with infliximab or other anti-TNF agent	3 intravenous infusions as induction therapy with Infliximab 5mg/kg (weeks 0,2,6) followed by: 5 infusions of maintenance therapy with Infliximab 5mg/kg administered at weeks 14,22,30,38,46 or 3 infusions with Infliximab 5mg/kg at weeks 18,30,42

Baldassano 2003 (infliximab)

The small trial of Baldassano 2003[43] examined if a single dose of infliximab induced a response in paediatric patients. Patients were randomised to a 1 mg/kg (n=6), 5 mg/kg (n=7) or 10 mg/kg (n=8) infusion. Patients were followed up to week 12. The primary outcomes were improvements from baseline in PCDAI and modified CDAI score. Other outcomes were the percentage of patients responding and the percentage in remission.

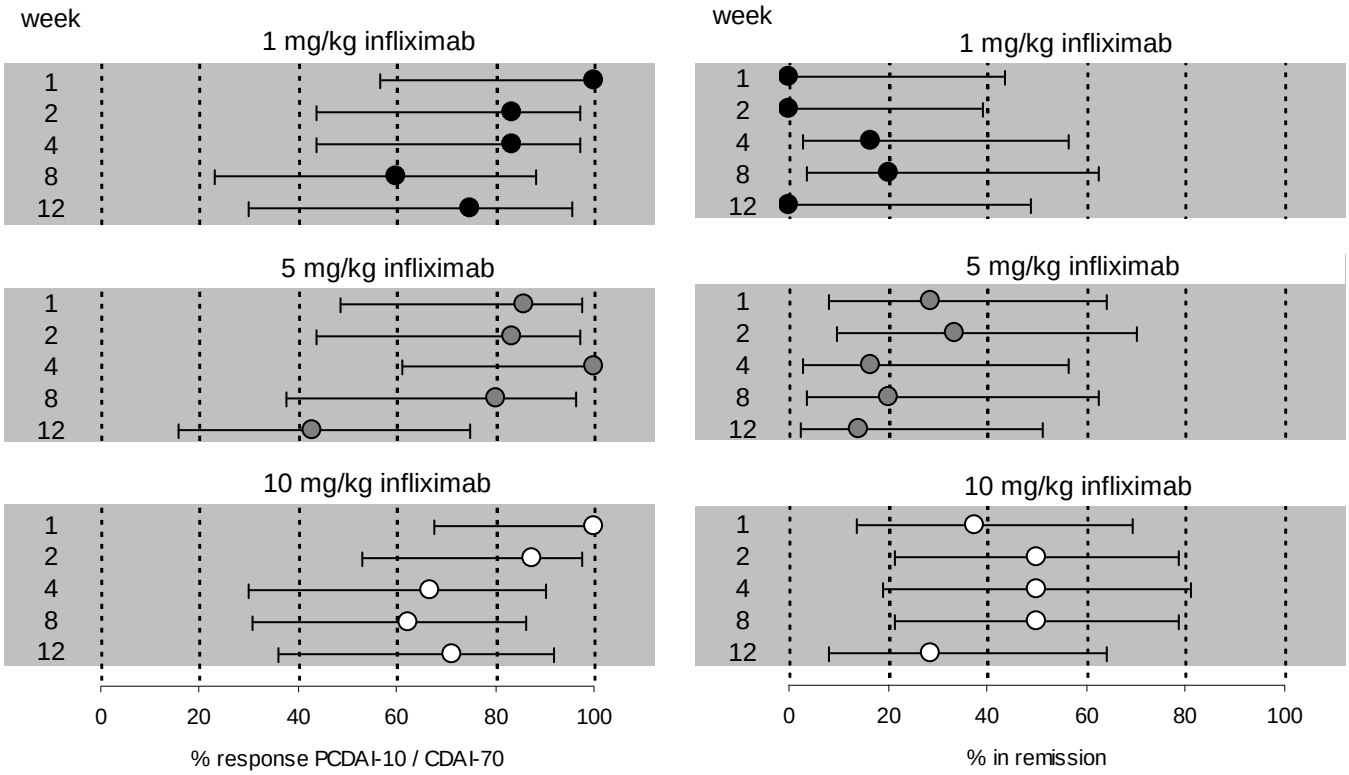
Table 37 shows the median percentage improvement in PCDAI score at various follow up times relative to baseline. No clear pattern relating to follow up time or dose regimen was apparent. To what extent improvement in scores resulted from infliximab treatment is impossible to determine because of lack of an appropriate placebo control group.

Table 37. Improvement in PCDAI score Baldassano 2003

Median % improvement from baseline in PCDAI score			
Week	Infliximab dose		
	1mg/kg	5mg/kg	10mg/kg
1	47	37	35
2	40	65	53
4	27	57	28
8	32	28	64
12	27	13	40

Response and remission results are summarised in Figure 29. All estimates were associated with great uncertainy due to the small number of participants. The proportion of patients in response approached 100% after one week in all groups and then tended to decline during follow up. There was little difference between the groups. How much of the response was intervention-dependent cannot be determined because of the lack of an appropriate placebo control group that did not receive infliximab. For remission, no clear pattern relating to dose or length of follow up was apparent. Again because of the lack of an inactive control it is impossible to determine the contribution of infliximab to the observed results.

Figure 29. Response and remission rates rates reported in Baldassano 2003 (results as reported, not ITT)



Response was defined as at least a 10 point reduction in PCDAI or at least a 70 point reduction in modified CDAI score; remission was defined as a PCDAI score less than 10 or a modified CDAI of less than 150

Quality assessment (based on published report)

Randomisation and blinding were adequate and allocation concealment likely to be so. With such small numbers in each group it is not surprising that some baseline characteristics were imbalanced, notably the 10 mg/kg group consisted almost exclusively of boys and the baseline CDAI score was substantially higher for the 1 mg/kg group than for the 5 mg/kg and 10 mg/kg groups. The number of patients completing the trial was reported to be 90%. No power calculation was done and analyses did not appear to be ITT.

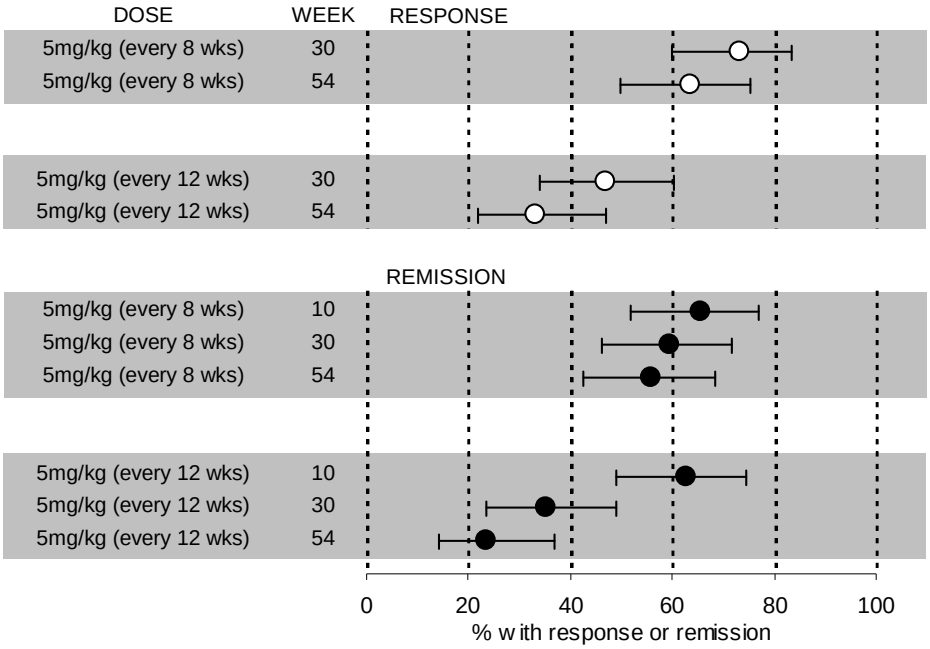
Baldassano 2003. Summary of effectiveness evidence.

An induction infusion of 1 or 5 or 10 mg/kg of infliximab improved PDAI scores relative to baseline. Induction increased the proportion of patients in response (40% to 100% depending on dose and follow up time) and in remission (0% to 50% depending on dose and follow up time). The study was underpowered so that these effectiveness estimates were associated with great uncertainty; no clear pattern was evident relating outcomes to dose regimen. The lack of a placebo control group renders interpretation of results problematical.

REACH (infliximab)

The REACH trial[42] was called an "induction and maintenance" study. Patients received induction doses of 5 mg/kg of infliximab at weeks 0, 2 and 6. Responders were defined as those who reduced baseline PCDAI by at least 15 points and had a score of 30 or less at week 10. Responders (only) at week 10 were randomised to either five further doses of 5 mg/kg every 8 weeks delivered at weeks 14, 22, 30, 38, and 46 or three further doses delivered every 12 weeks at weeks 18, 30, and 42. Of 112 patients entering the induction phase, 103 were classified as responders and 99 were analysed. The lack of a placebo control group not receiving infliximab means that it is difficult to determine to what extent maintenance of response after induction was attributable to infliximab intervention. No primary outcome was identified. Response and remission results were reported for weeks 30 and 54 and weeks 10, 30 and 54 respectively. These are summarised in Figure 30. The differences between the two dose regimens for both response and remission at both weeks 30 and 54 reached statistical significance ($p < 0.05$) in favour of the more frequent dose regimen.

Figure 30. Post induction response and remission rates for responders in the REACH trial



Response defined as decrease in PCDAI of ≥15 points from baseline and total no greater than 30. Remission defined as a PCDAI ≤ 10 points.

The REACH publication [42] also reported changes from baseline (mean and SD) in PCDAI score, IMPACT III score (a QoL measure), and daily corticosteroid use. Last observation was carried forward where values were missing. Information was provided for all "responders" (i.e. the two trial arms combined) or separately for the two different treatment groups, at weeks 10, 30 and 54. The results are summarised in Table 38. The "all responders results" do not represent a randomised comparison but rather a "before vs after-treatment" comparison for a subgroup of patients

(responders) that were selected because they exhibited a favourable response. Since a "no-treatment control group" was not included in this trial, the analyses do not provide robust quantitative information about the effectiveness of infliximab for paediatric patients and the favourable changes reported are difficult to interpret since an indeterminate proportion of the effects observed may have been infliximab-independent.

Table 38. Changes from baseline in outcome measures reported for the REACH trial

	PCDAI: mean *(SD)* decrease from baseline [improvement]			
GROUP	Week 10	Week 30	Week 54	N
Doses every 8 weeks	-33.2	NR	NR	52
Doses every 12 weeks	-29.4	NR	NR	51
Groups combined	-31 *(10)*	-25.5 *(16)*	-27 *(16)*	103
P < 0.001 vs baseline				
	IMPACT III: mean *(SD)* increase from baseline [improvement]			
GROUP	Week 10	Week 30	Week 54	N
Doses every 8 weeks	NR	24.7	26.5	NR
Doses every 12 weeks	NR	18.3	22.5	NR
Groups combined	23.9 *(16)*	21 *(18)*	24 *(17)*	76
P < 0.001 vs baseline				
	Corticosteroid dose (mg/kg): mean *(SD)* decrease from baseline [improvement]			N
GROUP	Week 10	Week 30	Week 54	
Doses every 8 weeks	NR	NR	NR	52
Doses every 12 weeks	NR	NR	NR	51
Groups combined	0.3 *(0.4)*	0.39 *(0.4)*	0.3 *(0.59)*	103
P < 0.001 vs baseline				
	Patients' corticosteroid use: n discontinued of N users *(%)*			
GROUP	Week 10	Week 30	Week 54	N
Doses every 8 weeks	12 of 24 at baseline *(50%)*	NR	10 of 12 at week 10 *(83%)*	NA
Doses every 12 weeks	3 of 12 at baseline *(25%)*	NR	5 of 9 at week 10 *(56%)*	NA
Groups combined	15 of 36 at baseline *(42%)*	NR	15 of 21 at week 10 *(71%)*	NA
Randomisation was at week 10. NR = not reported. NA = not applicable				

After week 10 responder patients were allowed to cross-over to increased infliximab for worsening disease state. The increases in infliximab allowed included transfer from infusions every 12 weeks to every 8 weeks and increase in infusion dose from 5 mg/kg to 10 mg/kg. The proportion of patients that crossed over was 40%. The number of responder patients that withdrew prematurely was reported as 22 (21%), but it was unclear if this included withdrawals of patients who had crossed over to increased infliximab.

Quality assessment (based on published report)

Randomisation was adequate and allocation concealment likely to be so. This was an open label study with no blinding. Baseline characteristics were well balanced. The number of patients withdrawing was reported but it was not clear if this also included cross-over patients who later withdrew. A power calculation was done and analyses were ITT.

REACH. Summary of effectiveness evidence.
A 10 week induction phase with infusions of 5 mg/kg of infliximab at weeks 0, 2 and 6 was followed by randomisation of responders at week 10 to further 5mg/kg infusions every 8 weeks or every 12 weeks. At week 10, 88% of enrolled patients were classified as responders. Response rates for responders diminished to less than 50% by week 54. The difference between dose regimens reached statistical significance in favour of the 8 weekly infliximab dose regimen. About 40% of patients crossed over to increased infliximab because of worsening disease status. About 20% of patients withdrew from treatment prematurely.

5.2.2.5 Results in non-responders

Published results for maintenance trials focused on early responders (determined at week 2, or 4 in the two large trials). It is important to attempt to determine if such a subgroup analysis can be justified.

The question of whether results were published for non-responders is summarised in Table 39. Out of all of the maintenance trials, only two trials (ACCENT I[3,4] and II[60]) published results including initial non-responders. Additional information was obtained from the industry submission for results in responders and non-responders from the CHARM[62] trial (see Appendix 11). Table 39 also details whether non-responders were randomised.

Table 39. Results reported for non-responders (maintenance trials)

Study	Were non-responders randomised to maintenance treatment?	Were results reported for both responders and non-responders separately (RCT only)?
CLASSIC II[61] (adalimumab)	No • Only those patients (from CLASSIC I[58]) in remission at week 0 and 4 eligible for randomisation; those not in remission at week 0 or no longer in remission at week 4 entered an open label cohort	No • Results reported for patients not eligible for randomisation who entered open label cohort
ACCENT I[3,4] (infliximab)	Yes • Responders and non-responders randomised	No • Results for ALL patients (responders + non-responders) reported in Rutgeerts 2004; results for responders only reported in Hanauer 2002 • Industry submission: sub-group analysis of mucosal healing and CDEIS scores in responders and non-responders together; hospitalisation and surgery reported for responders and non-responders together
Rutgeerts 1999[55] (infliximab)*	No • Responders from Targan 1997[54] RCT eligible • Initial non-responders in Targan 1997[54] given an additional 8 weeks of open label treatment during which they could respond and still be eligible for maintenance treatment • Unclear if any responders drawn from placebo group of RCT	No • No non-responders included in RCT • Proportion of non-responders at week 4 (Targan 1997[54] RCT) subsequently responding during open label treatment unclear
ACCENT II[60] (infliximab, fistulising)	Yes • Non-responders randomised for secondary analysis	Yes • Results reported for response
REACH[42] (infliximab, paediatric)	No • Only responders randomised (no placebo control)	No • No non-responders included in RCT (no placebo control)
CHARM[62] (adalimumab)	Yes • Non-responders randomised for secondary analysis (randomisation stratified by responder status)	No • Stated that secondary efficacy analyses were conducted for total population, but results presented only for fistula closure, which relates to a sub-group of patients (15% of patients have fistula) • Industry submission: present results (remission, CDAI decrease >70, CDAI decrease >100, IBDQ score) for responders and for patients with CDAI>300 . The trial report submitted to NICE contained information on non-responders.

ACCENT I[3,4] (infliximab)

Results for responders and non-responders were not published separately nor presented in the manufacturer's submission. However, by subtracting CIC information for responders from published information for all patients, it is possible in theory to calculate the response and remission rates for non-responders. Results for responders and for all patients were available in publications or CIC information in the industry submission for the following outcomes in the ACCENT I[3,4] trial:

a] Median CDAI scores at numerous follow up times. These were published in separate papers for responders only and for all patients.[4] No indication of variance was given so that robust analysis was not possible.

b] IBDQ scores. These were recorded but differently reported in the two publications (median scores for responders and a dichotomised outcome for "all" patients); this information cannot be used for estimation of non-responder results.[4]

c] Remission and response 70 for responder patients at multiple follow up times. The manufacturer's submission on infliximab provided CIC results for remission and response 70 for responder patients at multiple follow up times for ACCENT I[3,4]. Results for these outcomes for all patients were available in the public domain. It was possible to calculate the outcome for non-responder patients randomised to placebo or infliximab (5 mg/kg) by appropriate subtraction of responder rates from all-patient rates. Unfortunately, in practice this was meaningful for only the first 14 weeks of the trial because after week 14, patients who crossed over to increased infliximab dosage regimen on exacerbation of their disease contributed to the numbers achieving outcomes in the all-patient results but were discounted in the analyses for responders only.

The combined lack of complete long term results, and the introduction of crossover to different treatments at week 14 of the trial, made it difficult to determine the rates of response of 'non-responders' in the ACCENT I[3,4] trial, and renders problematical the interpretation of these rates when the limited available data allows their calculation. Appendix 11 provides the results calculated for non-responders in ACCENT I[4].[3]

ACCENT II (infliximab)

Limited results for responders and non-responders were reported separately for this trial[60] that investigated patients with fistulas. The response rate amongst initial non-responders was 7/44 (16%) in the placebo group and 9/43 (21%) in the infliximab group (p=0.6). A response was defined as a reduction of at least 50% from baseline in the number of draining fistulas at consecutive visits four or more weeks apart. The time point for this result was not stated and it is unclear whether these were patients who ever had a response during the 54 week trial. There are no details on whether these response rates were maintained. It is difficult to compare these results with those of the initial responders as the trial looked at the maintenance of response in initial responders rather than induction of response.

5.2.2.6 Adverse events

This section includes in-licence and non-licence trial results because all relevant evidence should be presented. All studies reported adverse events (AEs). In this section, the most serious AEs, and/or those thought potentially to be associated with anti-TNF therapy have been reported. Table 40 gives details on the number of patients with selected AEs, for both treatment and placebo groups. Where there were several treatment groups, these have been combined. AEs occurring during induction or open label periods of maintenance trials have been listed separately where details were given (CHARM[62] and CLASSIC II[61]). There were some differences in how papers reported or grouped together AEs (see notes at the foot of Table 41). Where details have not been reported (ND), it is possible that this was because the event did not occur. Excluding trials from the total count where the event did not occur, may lead to an overestimate the occurrence of this AE. Where patients experienced more than one type of AE within a category (e.g. infusion reactions), they will have been counted more than once.

AEs leading to withdrawal included worsening of CD, infection or obstruction. Serious infections included sepsis, colitis, abscess and pneumonia. Injection site reactions included burning, rash, pain, bruising or irritation, whilst IV infusion reactions included pruritus, chest pain, flushing, dizziness, dyspnoea, injection-site irritation and nausea. Very few deaths were reported.

Little difference was found between treatment and placebo groups for the selected AEs. The only cases of TB and lupus-like syndrome occurred in the treatment groups. AEs leading to withdrawal were slightly higher in the placebo groups and infusion reactions slightly higher in the treatment groups.

Table 40. Percentage of patients with selected AEs

	Treatment RCT data only	Placebo RCT data only	Induction or open label phase only (CHARM[62], CLASSIC II[61])
Deaths (both)	0.18% (3/1673)	0.11% (1/918)	N/A
Deaths (adalimumab)	0% (0/938)	0% (0/519)	0.09% (1/1075)
Deaths (infliximab)	0.41% (3/735)	0.25% (1/399)	N/A
Adverse events leading to withdrawal or discontinuation of treatment (both)	2.45% (43/1756)	6.36% (60/943)	N/A
Adverse events leading to withdrawal or discontinuation of treatment (adalimumab)	3.84% (36/938)	8.28% (43/519)	8.65% (93/1075)
Adverse events leading to withdrawal or discontinuation of treatment (infliximab)	0.85% (7/818)	4.0% (17/424)	N/A
Serious infections (both)	2.73% (47/1719)	3.42 (31/907)	N/A
Serious infections (adalimumab)	1.71% (16/938)	2.5% (13/519)	1.76% (19/1075)
Serious infections (infliximab)	3.97% (31/781)	4.64% (18/388)	N/A
TB (both)	0.23% (3/1323)	0% (0/707)	N/A
TB (adalimumab)	0.21% (2/938)	0% (0/519)	0% (0/1075)
TB (infliximab)	0.26% (1/385)	0% (0/188)	N/A
Cancer (both)	0.25% (4/1610)	0.56% (5/887)	N/A
Cancer (adalimumab)	0% (0/938)	0.38% (2/519)	0% (0/1075)
Cancer (infliximab)	0.60% (4/672)	0.81% (3/368)	N/A
Lupus (like syndrome) (both)	0.29% (3/1018)	0% (0/513)	N/A
Lupus (like syndrome) (adalimumab)	0% (0/421)	0% (0/258)	0% (0/221)
Lupus (like syndrome) (infliximab)	0.50% (3/597)	0% (0/255)	N/A
Demyelinating disorders (both)	0% (0/666)	0% (0/279)	N/A
Demyelinating disorders (adalimumab)	0% (0/554)	0% (0/279)	0.09% (1/1075)
Demyelinating disorders (infliximab)	0% (0/112)	ND	N/A
All infusion reactions (both)	16.4% (292/1777)	8.6% (81/943)	N/A
All infusion reactions (adalimumab)	17.5% (164/938)	7.7% (40/519)	12.7% (137/1075)
All infusion reactions (infliximab)	15.3% (128/839)	9.7% (41/424)	N/A
Anaphylactic reaction (both)	1.8% (2/112) NB possible reactions	ND	N/A
Anaphylactic reaction (adalimumab)	ND	ND	ND
Anaphylactic reaction (infliximab)	1.8% (2/112) NB possible reactions	ND	N/A

It appears that for reporting of AEs, the placebo groups of the maintenance trials also included patients who crossed over to a treatment group. For ACCENT I[3,4] and II[60], CHARM[62] and CLASSIC II[61], cross-over was specified as an option for those patients who had a non-response or experienced a disease flare. There were no details regarding potential crossovers from placebo to treatment in Rutgeerts 1999[55] (n=73). See section on quality for details on number of cross-overs from placebo groups (see Appendix 12).

Cross-over to treatment may have had an effect on the types and numbers of AEs reported in the placebo groups, for example an increase of those types of AEs associated with the treatment (e.g. infection) and/or an underestimate of AEs associated with no treatment (e.g. worsening of CD). It should be noted that in the maintenance, trials all patients (including those subsequently randomised

to placebo) initially received the study drug during the induction phase; the effects of this may have carried over into the placebo phase of the RCT.

None of the maintenance trials reported AEs for patients according to whether they had ever/never received the treatment during the RCT phase of the study. As some of the AEs reported are very rare, it is possible that any differences between treatment and placebo groups are due to chance.

Table 41. Number of patients with selected adverse events

	INFLIXIMAB							ADALIMUMAB					
STUDY	Baldassano et al., 2003[43]	Hyams et al., 2007[42] REACH	Present et al., 1999[57]	Sands et al., 2004[60] ACCENT II	Targan et al., 1997[54]	Hanauer et al., 2002[4] ACCENT I	Rutgeerts et al., 1999[55]	Hanauer et al., 2006[58] CLASSIC I	Sandborn et al., 2007[59] GAIN	Colombel et al., 2007[62]CHARM	Colombel et al., 2007[62]CHARM	Sandborn et al., 2007[61] CLASSIC II	Sandborn et al., 2007[61] CLASSIC II
Study duration (weeks) I=induction M=maintenance F=fistulising P=paediatric	12 weeks (I, P)	54 weeks (M, P)	18 weeks (I, F)	54 weeks (M, F)	4-16 weeks (I) (additional follow-up for open-label patients)	54 weeks (M)	48 weeks (M)	4 weeks (I)	4 weeks (I)	4 week induction period (n=854)	56 weeks (M) (n=778)	56 weeks (M)	Open label cohort
Deaths Rx	ND	0/ 112	0/63	0/138	ND	3/385	0/37	0/225	0/159	1/854	0/517	0/37	0/221
Deaths Placebo	N/A	N/A	0/31	0/144	ND	0/188	1/36	0/74	0/166	N/A	0/261	0/18	N/A
Adverse events leading to withdrawal or discontinuation of treatment Rx	ND	12/112	1/63	5/138	2/83 Unclear	45/385	6/37	2/225	2/159	54/854	30/517	2/37	39/221
Adverse events leading to withdrawal or discontinuation of treatment Placebo	N/A	N/A	0/31	12/144	0/25 Unclear	5/188	0/36	2/74	4/166	N/A	35/261	2/18	N/A
Serious infections Rx	ND	9/ 112	3/63	4/138	1/83	14/385	ND	2/225	0/159	10/854	14/517	0/37	9/221
Serious infections Placebo	N/A	N/A	0/31	9/144	1/25	8/188	ND	0/74	4/166	N/A	9/261	0/18	N/A
TB Rx	ND	ND	ND	ND	ND	1/385	ND	0/225	0/159	0/854	2/517	0/37	0/221
TB Placebo	N/A	N/A	ND	ND	ND	0/188	ND	0/74	0/166	N/A	0/261	0/18	N/A
Cancer Rx	ND	0/ 112	ND	0/138	ND	4/385	0/37	0/225 lymphoma	0/159	0/854	0/517	0/37 lymphoma	0/221 lymphoma
Cancer Placebo	N/A	N/A	ND	0/144	ND	2/188	1/36	0/74 lymphoma	0/166	N/A	1/261	1/18 lymphoma	N/A

	INFLIXIMAB							ADALIMUMAB					
STUDY	Baldassano et al., 2003[43]	Hyams et al., 2007[42] REACH	Present et al., 1999[57]	Sands et al., 2004[60] ACCENT II	Targan et al., 1997[54]	Hanauer et al., 2002[4] ACCENT I	Rutgeerts et al., 1999[55]	Hanauer et al., 2006[58] CLASSIC I	Sandborn et al., 2007[59] GAIN	Colombel et al., 2007[62]CHARM	Colombel et al., 2007[62]CHARM	Sandborn et al., 2007[61] CLASSIC II	Sandborn et al., 2007[61] CLASSIC II
Study duration (weeks) I=induction M=maintenance F=fistulising P=paediatric	12 weeks (I, P)	54 weeks (M, P)	18 weeks (I, F)	54 weeks (M, F)	4-16 weeks (I) (additional follow-up for open-label patients)	54 weeks (M)	48 weeks (M)	4 weeks (I)	4 weeks (I)	4 week induction period (n=854)	56 weeks (M) (n=778)	56 weeks (M)	Open label cohort
Lupus (like syndrome) Rx	ND	0/ 112	0/63	ND	ND	2/385	1/37	0/225	0/159	ND	ND	0/37	0/221
Lupus (like syndrome) Placebo	N/A	N/A	0/31	ND	ND	0/188	0/36	0/74	0/166	N/A	ND	0/18	N/A
Demyelinating disorders Rx	ND	0/ 112	ND	ND	ND	ND	ND	ND	ND	1/854	0/517	0/37	0/221
Demyelinating disorders Placebo	N/A	N/A	ND	ND	ND	ND	ND	ND	ND	N/A	0/261	0/18	N/A
All infusion reactions Rx	0/21	19/112	4/63	22/138	2/83	80/385	1/37	66/225	17/159	111/ 854 injection site reaction	80/517 injection site reaction	1/37	26/221
All infusion reactions Placebo	N/A	N/A	0/31	24/144	0/25	17/188	0/36	12/74	17/166	N/A	9/261 injection site reaction	2/18	N/A
Anaphylactic reaction Rx	ND	2 /112 possible reaction	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
Anaphylactic reaction Placebo	N/A	N/A	ND	ND	ND	ND	ND	ND	ND	N/A	ND	ND	NA

ND=no details (i.e. not stated in paper that this event did or did not occur); N/A= not applicable as no placebo arm

- Baldassano 2003[43]: did not use category of serious infection (1 pancreatitis, 1 sinusitis/appendicitis, 2 URTI)
- Present 1999[57]: did not use category of serious infection, but list infections under serious AE (pneumonia and abscesses)
- ACCENT II[60]: during long-term follow-up: 2 deaths, 2 cases of cancer and 1 of MS (all patients had received infliximab at some point)
- Targan 1997[54]: Two infusion reactions required discontinuation of infusions (unclear if led to discontinuation of study/treatment altogether); did not have category of serious infection but 2 infections required hospitalisation
- Rutgeerts 1999[55]: placebo patient who died was same one who had lymphoma (had induction treatment with infliximab); 1 patient withdrawn due to infusion reaction, unclear if further infusion reactions
- CHARM[62]: 2 cases of TB occurred during post-randomisation open label therapy

5.2.2.7 Development of antibodies

This section describes all included studies. Table 42 lists numbers of patients developing antibodies to the respective anti-TNF agents, anti-nuclear antibodies and antibodies to double-stranded DNA. Most (10/11) studies reported the development of antibodies to the respective anti-TNF agent; four studies reported anti-nuclear antibodies and eight studies reported anti-double-stranded DNA antibodies.

Five induction trials reported antibodies to an anti-TNF agent; these ranged from 0% to 6% (adalimumab: 0%, 1.3%; infliximab: 0%, 3.3%, 6%). All reported antibody development either for the intervention group only, or split by placebo and intervention group, except Present 1999[57], which reported antibodies for placebo and intervention group together. Targan 1997[54] included patients from the post RCT open label extension. This was also the longest logest follow study amongst induction trials (16 weeks) and had the highest level of antibodies (6%).

Five maintenance trials reported antibodies to an anti-TNF agent; these ranged from 2.6% to 17% (infliximab: 2.9%, 5.2%/9.2% (Rx/Placebo), 15%, 17%; adalimumab: 2.6%). Three studies reported antibodies for the intervention and placebo groups together (ACCENT II[60], Rutgeerts 1999[55] and CLASSIC II[61]). The majority of patients in CLASSIC II[61] came from the open label cohort component of the study. The lowest antibody levels occurred in CLASSIC II[61] (adalimumab); the other large adalimumab maintenance trial (CHARM[62]) did not measure antibodies.

Seven studies listed the proportion of inconclusive samples, which were generally high and ranged from 14% to 'most' patients. These samples had detectable concentrations of anti-TNF agent, which could compete for the detection of antibodies to the anti-TNF agent in the immunoassay used, and would therefore not give a valid result. It is unclear whether the overall percentages of antibodies to the anti-TNF agent would have been different if they could have been measured in all patients.

As with the adverse events described above, it should be noted that patients in the placebo groups of the maintenance trials would have all received the treatment as part of induction and may also have crossed over to a treatment group during later stages of the trials.

The proportions of anti-nuclear antibodies were variable: 25% in REACH[42] (infliximab), 18%/46% (Rx/Placebo) in ACCENT II[60] (infliximab), 35%/56% (Rx/Placebo) in ACCENT I[3,4] (infliximab) and 19% in CLASSIC II[61] (adalimumab).

Antibodies to double-stranded DNA were measured in three infliximab induction trials (range from 0% to 13%) and four maintenance trials (range 4% to 34%); only one adalimumab trial (CLASSIC II[61], 19%)measured this parameter.

Given the proportion of missing data (inconclusive samples), the varying numbers of patients receiving treatment in different trials (those who crossed over), and the relatively small number of trials it is not possible to conclude that one of the interventions is more or less likely to result in the development of antibodies to the anti-TNF agent. Whether different types of assays were used for the detection of antibodies, or whether there were differences in the number of frequency of assessments, which could have led to differences between studies or drugs was not investigated.

Based on the results from the ACCENT I trial,[3,4] it appeared that "episodic" treatment lead to the formation of fewer antibodies than scheduled treatment (28% in placebo/episodic treatment arm, 9% in 5 mg/kg scheduled arm and 6% in 10 mg/kg scheduled arm). It should be noted that the comparison between "episodic" and scheduled treatment is not a randomised one (see section on quality assessment). Given that the "episodic" group included patients who crossed over from the scheduled treatment groups and the fact that 46% of total samples were inconclusive, it is unclear how robust these results are.

Table 42. Antibodies to anti-TNF agent and DNA

	% evaluated for antibodies	Patients with antibodies to anti-TNF agent	Patients with anti-nuclear antibodies	Patients with antibodies to double-stranded DNA
ADALIMUMAB				
Hanauer et al., 2006CLASSIC I[58] INDUCTION	ND	Placebo: 1/74 (1.3%) positive (assumed N) Adalimumab groups: 1/225 (0.4%) positive (assumed N)	ND	ND
Sandborn et al., 2007 GAIN[59] INDUCTION	Appears to be measured in all	Placebo: 0/166 (0%) positive Adalimumab: 0/159 (0%) positive Presence of measurable adalimumab precluded determination of antibodies in most patients treated with adalimumab	ND	ND
Colombel et al., 2007CHARM[62] MAINTENANCE	Not measured in this study	Not measured in this study	Not measured in this study	Not measured in this study
Sandborn et al., 2007CLASSIC II[61] MAINTENANCE	269/276 (97%; includes 221 from open label cohort) for anti-adalimumab antibodies 185/276 (67%; includes 221 from open label cohort) for anti-nuclear and anti-DNA antibodies	7/269 (2.6%) positive All groups including open label cohort	36/185 (19%) positive At baseline and/or at final visit.	33/185 (19%) positive At baseline and/or at final visit.
INFLIXIMAB				
Baldassano et al., 2003[43] INDUCTION CHILDREN	21/21 (100%) NB no placebo group	0/21 (0%) positive	ND	0/21 (0%) positive
Hyams et al., 2007 REACH[42] MAINTENANCE CHILDREN	105/112 (94%) for antibodies to anti-TNF agent 91/112 (81%) for anti-nuclear antibodies 99/112 (88%) for anti-DNA antibodies NB no placebo group; includes patients who were not randomised to maintenance therapy	3/105 (2.9%) positive 21/105 (20%) negative 81/105 (77.1%) inconclusive sample (detectable infliximab concentration)	23/91 (25%) positive	7/99 (7%) positive

	% evaluated for antibodies	Patients with antibodies to anti-TNF agent	Patients with anti-nuclear antibodies	Patients with antibodies to double-stranded DNA
Present et al., 1999[57] INDUCTION FISTULISING	92/94 (98%)for antibodies to anti-TNF agent Unclear for anti-DNA antibodies (appears all in infliximab groups only)	3/92 (3.3%) positive 13/92 (14%) inconclusive sample Infliximab and placebo groups	ND	8/63 (13%) positive Infliximab groups only
Sands et al., 2004 ACCENT II[60] MAINTENANCE FISTULISING	258/282 (91%) for antibodies to anti-TNF agent 254/282 (90%) for anti-nuclear antibodies 243/282 (86%) for anti-DNA antibodies	44/258 (17%) positive 80/258 (31%) negative 134/258 (52%) inconclusive sample Not detailed by group	Infliximab: 56/122 (45.9%) positive Placebo: 24/132 (18.2%) positive	Infliximab: 27/116 (23.3%) positive Placebo: 8/127 (6.3%) positive
Targan et al., 1997[54] INDUCTION	101/108 (93%) for antibodies to anti-TNF agent 98/108 (91%) for anti-DNA antibodies NB samples include post RCT open label patients	6/101 (6%) positive who received infliximab blinded or as open label NB in 2/3 of patients infliximab was still detectable and may have interfered with assay	ND	3/98 (3%) positive who received infliximab blinded or as open label
Hanauer et al., 2002 & Rutgeerts 2004 ACCENT I[3,4] MAINTENANCE	442/573 (77%) for antibodies to anti-TNF agent NB number of antibodies reported according to actual treatment received bearing in mind that patients crossed over	Placebo: 41/442 (9.2%) positive Infliximab groups: 23/442 (5.2%) positive In 46% of patients infliximab still detectable therefore inconclusive	Placebo: 63 (35%) positive Infliximab groups: 363 (56%) positive	Placebo: 19 (11%) positive Infliximab groups: 123 (34%) positive
Rutgeerts et al., 1999 MAINTENANCE	71/73 (97%)	7/47 total (15%) positive Placebo: 5/ND Infliximab: 2/ND 24/71 (34%) inconclusive as measurable infliximab in sample	ND	2/47 total (4%) positive Placebo: 0/ND Infliximab: 2/ND
ND= no details				

5.2.3 Discussion of results and assessment of effectiveness

Patient heterogeneity

Patient heterogeneity may affect results across different trials. The inclusion criteria of the trials specified a CDAI score between 220 and 400 or 450. The inclusion of patients already at a CDAI level close to remission could have improved the remission rates found. However, if patients already had a low CDAI count, achieving a reduction of 70 points or 100 points would have been harder to achieve. The opposite would be more likely to be true for patients with very high initial CDAI scores. Therefore, it is unlikely that the initial wide spread of CDAI scores would have much impact on the results unless there were more patients at one end of the spread than the other. Mean CDAI scores at entry did not vary greatly between trials, so it appears unlikely that patient populations taken as a whole differed substantially between trials with respect to this parameter. Nevertheless populations probably did differ between trials since the placebo rates were heterogeneous. The corollary is that CDAI is not necessarily a reliable indicator of the seriousness of disease or of its likely progression. The CDAI score is a summary score and patients can achieve the same score yet have problems with very different aspects of their disease. Similarly, if a patient had a reduction of 70 points, that could be achieved in a variety of different ways. It is also uncertain whether a reduction of 70 or 100 points means the same in terms of reduction of disease severity for patients starting at different ends of the severity spectrum.

Cohort studies (e.g. Munkholme)[22] demonstrate that most CD patients, at some time in their disease history, experience "highly active disease" and that they cycle between highly active and quiescent periods of varying durations. Whether CD is severely debilitating for an individual depends to some extent on the frequency with which the episodes of highly active disease are repeated. Cohort studies show that this varies between patients. For these reasons a patient's CDAI score at a particular time, such as at recruitment into a trial, is not a good indicator for the likely duration of that level of disease activity or of the likely subsequent recurrence of highly active disease.

The licence indications for infliximab and adalimumab specify 'severe' CD but do not define how severe disease may be determined. It has been assumed that this is a CDAI score of 300+. Trials have recruited patients having 'moderate to severe CD' defined according to CDAI scores of between 220 and 450, or 220 and 400; it is therefore unclear to what extent these populations fully reflect the intended licensed population.

Induction trials - placebo rates

CD is a chronic relapsing and remitting disease. Induction trials selected patients in relapse. On average, irrespective of treatment, relapsed patients will tend to improve i.e. remit with time (their

CDAI scores will reduce as they regress to the mean). This tendency would be reflected in relatively high rates of improvement in placebo groups in placebo controlled trials and also in variation in these rates dependent on the relapse-remission cycling characteristics of each of the patients enrolled in the different trials.

The rates of response (reduction in CDAI of 70 or 100 points) and of remission in the placebo arms of the included induction trials varied from trial to trial and in some trials reached high levels (see Appendix 13 for details). Except in the Targan 1997 trial of infliximab,[54] by week 4 one third or more of placebo patients had already achieved the least stringent measure of improvement (response 70). Similarly, at least 20-25% achieved response 100 by week 4. Varied and high rates of placebo response have previously been documented for many CD intervention trials (Su et al., 2004)[70]. For dichotomous outcomes, variable placebo rates can profoundly influence effect size values such as rate difference and rate ratio. Thus placebo and intervention rates in two trials of 10% and 20% in one and 30% and 40% in the other generate identical outcome measures for rate difference (0.1, or 10%) but considerably different measures for rate ratio (2.0 and 1.3 respectively). For this reason both placebo and intervention rates and both rate difference and rate ratio effect sizes have been presented in this report for most outcomes in the results section. The confidence intervals quoted were not adjusted for repeated measures.

These high and varied placebo rates probably result from three influences: the tendency for CDAI scores to regress to the mean, from a placebo effect, and possibly from the effect of concomitant treatments allowed in the trials. The variation in placebo rates makes comparisons between trials problematical and indicates that CDAI scores alone are unlikely to be good prognostic indicators. Although recruited populations in the trials conformed to similar ranges or means or medians of CDAI score, they are likely to be clinically dissimilar.

Induction trials - effect sizes

By week 4 all induction trials, except for CLASSIC I[58] at the lower dose level for adalimumab (80 / 40 mg/kg weeks 0 and 2), exhibited statistically significant effect sizes for anti-TNF relative to placebo for remission and response, irrespective of whether these were measured in terms of rate difference or rate ratio. The trial of infliximab by Targan 1997[54] was remarkable in that the effect sizes observed were much greater than those seen in the other trials; placebo rates were notably lower in Targan 1997[54] than in any other trials. Targan 1997[54] was the earliest anti-TNF induction RCT and was a relatively small trial so that the point estimates of effectiveness were associated with more uncertainty than was the case for the larger induction trial of adalimumab (GAIN[59]). In the decade since the publication of Targan 1997[54], no infliximab induction trial has been reported that can

provide confirmatory evidence for the large effect size point estimates from the Targan 1997[54] trial. The response 70 rate at four weeks in the intervention arm of Targan 1997[54] was 81%. In the induction phase of the ACCENT I maintenance trial of infliximab,[3,4] the response 70 rate at week 4 was considerably less at 59%. ACCENT I[3,4] patients were administered the same dose at week 0 and patient baseline characteristics were similar to Targan 1997[54] e.g. very similar CDAI and IBDQ scores. The contribution of infliximab to the initial 59% response 70 rate in ACCENT I[3,4] cannot be gauged because of lack of an appropriate control group.

The follow up in the published adalimumab trial reports was to four weeks only, and there is no reliable evidence on effectiveness of induction with adalimumab beyond this time period. Targan provided data on infliximab for some patients up to 16 weeks (4 weeks induction + 12 weeks open label).

Maintenance trials - general comments on trial design

The maintenance trials conformed to what have been called "adaptive" trial designs. The main features of such designs have been reviewed recently by Chang and Chew[71]. ACCENT I, CHARM and CLASSIC II trials had adaptive trial design of the type described as "drop-the-loser" with in some cases "adaptive treatment switching".[71]. An inherent problem of "drop-the-loser" design is that groups that are dropped may contain valuable information regarding the response to treatment under study. A further problem concerns how such studies should be powered; whether for the interim analysis at the point when "losers" are dropped, or for the final analysis involving winners only. With treatment switching come problems of identifying the target population for the therapy of interest and a precise definition of the therapy provided. Treatment switching can lead to a change to a different hypothesis being tested. Chang and Chew state "From a statistical point of view adaptations to trial and or statisticall procedures could (i) introduce bias/ variation to data collection, (ii) result in a shift in location and scale of the target population, (iii) lead to inconsistency between the hypothesis being tested and the corresponding statistical tests".[71] In summary these trials are susceptible to difficulties of analysis and interpretation.

Maintenance trials in adult populations wholly or predominantly of nonfistulisng patients

For both drugs, one large maintenance trial has been published that employed within-licence treatment regimens: the CHARM trial [62](adalimumab) and the ACCENT I trial[3,4] (infliximab). The interpretation of results from the maintenance trials was hampered by the nature of the trial designs, most of which allowed for scheduled crossovers into other treatment arms (or to "open label treatment"). This led to a proportion of patients in the placebo arms of the trials receiving variable amounts of drug. In order to comply with an intention-to-treat analysis, these patients (and those who

withdrew) were mainly counted as treatment failures for binary outcomes such as remission or response. Not all trials clearly defined the handling of missing data or data for patients who crossed over. Where there was missing continuous data, the LOCF was used in ACCENT I but not in CHARM, the effect of which on results is unclear.

There were particular concerns over the ACCENT I trial (Rutgeerts 2004[3] publication), as its stated aim of comparing episodic with scheduled treatment is misleading as no patients were randomised to an episodic treatment arm. Furthermore, there were uncertainties regarding the impact of methods for handling of missing data in the analysis including both responders and non-responders.

Responder/non-responder subgroups

The interpretation of the maintenance trials was further complicated by the fact that a sub-group of patients (responders) were selected for analysis or randomisation at varying time-points after an induction period during which all patients received the study drug. For both of the large maintenance trials of within-licence treatment regimens (CHARM[62] - adalimumab and ACCENT I[3,4] - infliximab) the published effectiveness results all focused on the "responders" subgroup. Separate results for non-responders were not reported (see Table 39), although both CHARM[62] and ACCENT I[3,4] randomised both responder and non-responder patients. The definition of responders differed somewhat between the two trials. Furthermore, the induction phases used in both trials differed with respect to duration and number of induction doses administered. The consequence of these considerations is that attempting any comparisons of effectiveness between the trials is very problematical. The proportion of patients categorised as responders in each of these trials was CHARM[62] 64% and ACCENT I[3,4] 58%.

It is known from trials where results were also reported for (randomised) non-responders that initial non-responders can still respond later, so it is unclear to which patients this sub-group of responders actually represents in clinical practice. It is possible that a sub-group of responders chosen at a different time-point would have led to different results. There is no published evidence or information in the manufacturers' submissions to show that responders compared to non-responders benefit more from the treatment (compared to placebo). The selection of responders at different time-points in different trials also hampers any comparisons between the trials.

Reporting effectiveness results for a subgroup but not for all randomised patients (or not for all patients that commenced treatment) appears at odds with usual practice. For example in placebo controlled randomised trials of anti-TNF agents (infliximab, adalimumab, and etanercept) for the treatment of rheumatoid arthritis results for all patients have been analysed and presented[72]

Dichotomising patients into responders and non-responders only makes clinical sense if a "response" at the time of the dichotomisation is a good prognostic tool for identifying those patients most likely to benefit from maintenance of treatment. In order to find this out requires the comparison of results for responders and non-responders, which is the precise analysis that was not undertaken in these trials, unfortunately. Thus there is no evidence available to indicate that subgrouping patients in the ways described is a useful practice. The usefulness of the results reported for responders only is therefore questionable.

The ACCENT I[3,4] trial dichotomised patients according to their response at two weeks after the induction infusion of infliximab. The decision to do this may have been derived from previous research. The 1997 induction study of Targan[54] provided data up to four weeks after a single infusion of infliximab at 5mg/kg. This study reported that the mean CDAI score in the placebo group remained constant from week 2 to week 4, while the rate differences (infliximab vs placebo) for remission (score less than 150 points) and for a 70 point reduction in CDAI score increased from 0.37 to 0.44 and from 0.62 to 0.65 respectively. Placebo rates for these outcomes remained constant from weeks 2 to 4. Although the study was small and the point estimates were associated with considerable uncertainty these results imply that some patients not responding at two weeks in fact do go on to respond at a later time.

In the absence of appropriate analyses it appears that dichotomising patients as early as 2 weeks after a single infliximab infusion is probably premature and does not appear to be a clinically meaningful procedure.

In the CHARM trial the time chosen for categoriation into responders and non-responders (at week 4) was not based on efficacy data but was '*based on pharmacokinetic model estimates for when maximal drug concentrations should be present'*. CIC results were available for nonresponder patients at weeks 12, 26 and 56 so that it was possible to calculate the response rates amongst all randomised patients. The rate difference and rate ratio results for remission, response 100 and response 70 are summarised in Figure 31 and in Figure 32 respectively. [redacted]

Figure 31. Rate difference results for responders, nonresponders and all patients in CHARM

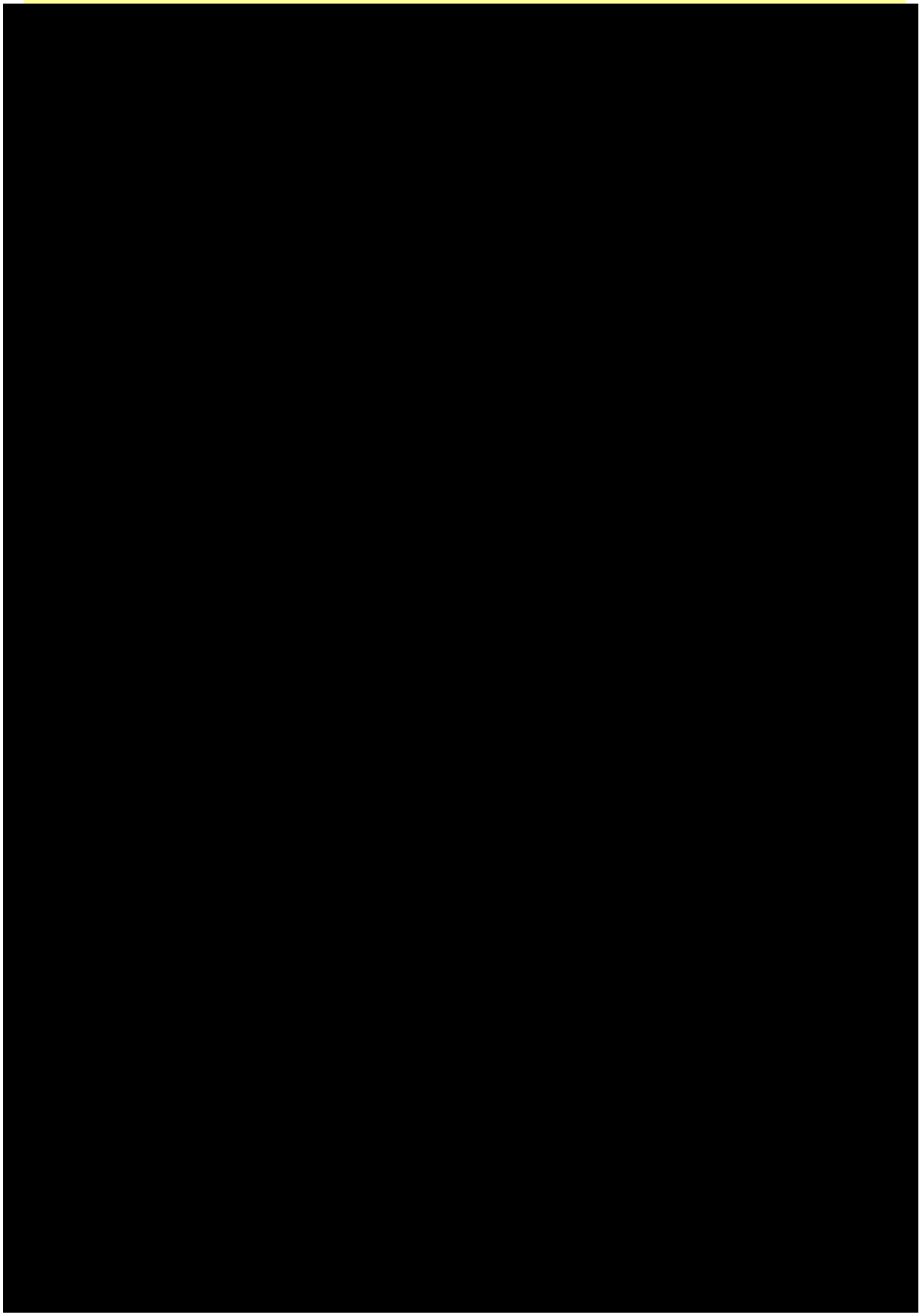
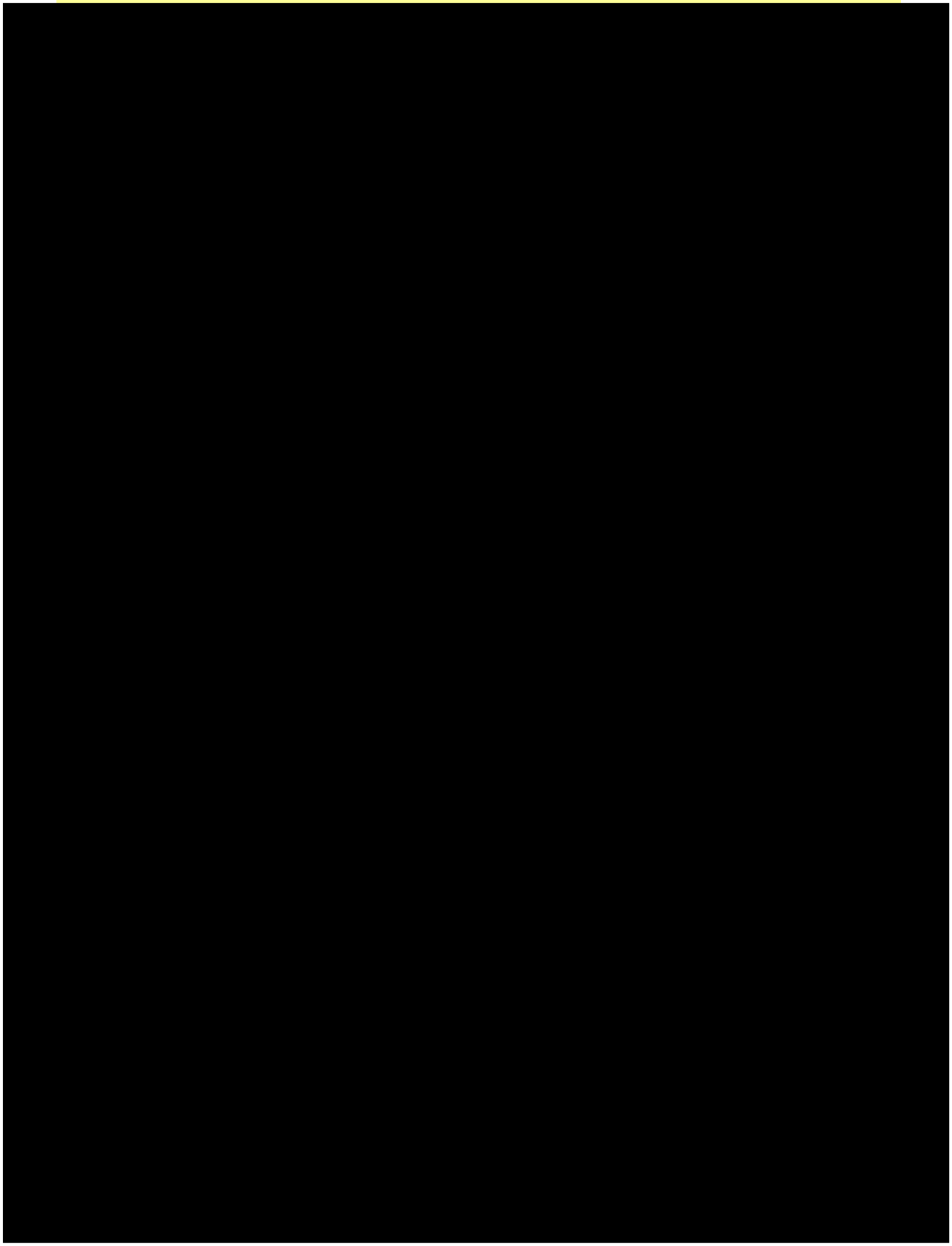


Figure 32. Rate ratio results for responders, nonresponders and all patients in CHARM



The two large maintenance trials (CHARM[62] and ACCENT I[3,4]) provided evidence that for the subgroups of patients defined as "responders" anti-TNF therapy was beneficial compared to placebo with respect to the proportions of patients exhibiting remission or response (70 or 100). Rates at multiple follow up times extending to week 56 for CHARM[62] (remission rates) were reported in the published paper. For the ACCENT I[3,4] trial, results for weeks 30 and 54 (response 70 and remission) only were published, but CIC information for multiple time points was provided. The higher rates of response for intervention versus placebo might lead to the conclusion that extended therapy over the prolonged follow up is beneficial and/or necessary for maintenance of response. However, examination of all the information available indicates that nearly all benefit observed for intervention over placebo was generated early on and that rate differences thereafter remained relatively stable or decreased. These results imply that prolonged treatment after the initial benefit has been attained is uneconomical, and since anti-TNF agents are associated with significant health risks, may be clinically ill-advised. The dose regimens required to attain this early benefit are likely to be different for adalimumab and infliximab.

The published results for the CHARM[62] trial graph "*% patients maintaining remission*" (Figure 2B in CHARM[62] publication) versus follow up time depicted increased rates of remission following after decreased remission rates, demonstrating that patients that achieved remission at late follow up times are counted as "*maintaining remission"* and that in fact, the point prevalence of remission is represented in the graph rather than maintenance of individual patients in remission. If this is the case, a late time point (e.g. 30 or 54 weeks) value reported does not necessarily inform about maintenance of response during follow up since it is possible that those registered as "*in response*" may only have achieved this status just prior to the time point reported. It was unclear, but appeared possible, that point prevalence of response was the statistic reported in the ACCENT I[3,4] published reports [redacted]

The most appropriate way to determine the ability of anti-TNF agents to maintain response in patients who were defined as responders is time-to-event analysis with statistical comparison using a log rank test. For ACCENT I,[3,4] median time to treatment failure was 19 weeks and 38 weeks for placebo and 5 mg/kg infliximab groups respectively ($p = 0.002$), however the definition of treatment failure used in this analysis was complex, did not correspond to a loss of response 70 status and its clinical impact difficult to gauge. The CHARM[62] trial (adalimumab) reported median duration of remission for those responders who achieved remission starting at any time during follow up. The median times were 127

days for the placebo group, 378 days for the 40 mg/kg adalimumab eow group and greater than 392 days for the 40 mg/kg weekly dosage regimen group.

Trials recruiting patients with draining fistula

One induction trial provided evidence that infliximab promotes fistula closure to a greater extent than placebo. The ACCENT II[60] trial of infliximab maintenance treatment focused on responders (69% of patients receiving induction doses). Infliximab maintenance treatment promoted closure of fistulas to a statistically significantly greater extent than did placebo. There was evidence that a reduction of dose frequency from every four weeks to every eight weeks was associated with poorer maintenance of fistula closure. Limited evidence from the CHARM[62] trial suggested that adalimumab may also promote fistula closure.

It is possible that fistula closure may not necessarily be a desirable outcome since it may result in increased development of abscesses. A *post hoc* analysis of patients participating in the ACCENT II[60] trial found no significant difference in abscess incidence between two groups receiving different mean dosages of infliximab. Interpretation of these results is problematical because results for the most appropriate comparison (placebo vs infliximab) were not available.

Trials recruiting paediatric patients

Two trials of infliximab, one induction and the other maintenance, reported on the treatment of paediatric patients. Unfortunately, in these trials all patients received infliximab and no reliable inferences regarding the effectiveness of the intervention were possible because the spontaneous response rates in the population were unknown. The more frequent of the two dosage regimens used in the REACH[42] trial (5 mg/kg every 8 weeks or every 12 weeks) resulted in statistically significant greater rates of response and remission, a dose response relationship that likely implies beneficial effect of infliximab relative to placebo or standard treatment but a placebo controlled trial would have provided far stronger evidence of effectiveness.

Differences in effectiveness of anti-TNF agents; indirect comparisons.

No head to head trials were found that compared the effectiveness of adalimumab and infliximab. However the existence of placebo controlled induction and maintenance trials for both drugs means that adjusted indirect comparisons of effectiveness were theoretically feasible using methods described by Glenny et al., 2005.[73]

The indirect comparison of trials was hampered in this case by a number of factors. One of these was differing placebo rates found for induction trials and unknown or uncertain placebo rates in

maintenance trials (because all patients receive active intervention early in the trial). Patients with CD can experience spontaneous clinical improvement without treatment. Su et al. (2004)[70] conducted a meta-analysis of CD trials looking at placebo rates for remission and response (based on the CDAI). The authors found substantial heterogeneity between placebo rates and found that these were in the main attributable to follow-up duration, number of follow-up visits and CDAI score at entry to the trial (see Appendix 13). Because of the variation in placebo rates in the induction trials indirect comparison was not done.

Indirect comparison of effectiveness using maintenance trials was judged unlikely to deliver valid results. For reponders the placebo arms of compared trials were not truly comparable because the groups had received different anti-TNF induction drugs on differing numbers of occasions and for different periods of time; furthermore the "responder" groups were constituted from different proportions of the randomised populations according to differing criteria. For all patients' results again placebo groups were not truly similar between trials and additionally availability of results for all patients in the adalimumab trials was limited (see Appendix 11); furthermore the permitted cross-over of variable proportions of placebo group patients to active intervention at weeks 12 and 14 of the CHARM and ACCENT I trials would render analyses unreliable.

Adverse events

The large amount of cross-overs in the trials made the comparison of adverse event rates between treatment and placebo arms difficult, as many patients in the placebo arms will have received some study drug. In addition, the maintenance trials either gave an induction bolus of the drug at the start of the trial then randomized to treatment or placebo, or enrolled patients from a previous induction trial. Similarly, it is difficult to tell what the true rates for the development of antibodies are for each of the drugs, again due to cross-overs and induction doses. It was not within the remit of this project to examine the test accuracy of antibody determination used in the trials.

Summary of effectiveness results

There were no included RCTs with severe CD patients only. They all included moderate to severe CD.

1] The general pattern of results is similar for the two drugs.

2] There is a good initial, clinically significant, improvement for the majority of patients when given induction treatment with infliximab or adalimumab. The short duration induction trials demonstrated that the majority of CD patients with moderate to severe disease gained clinical benefit from a single IV infusion of infliximab (5 or 10 mg/kg) or two subcutaneous injections of adalimumab (80mg and 40mg or 160mg and 80mg) separated by two weeks. Published estimates for the proportion benefiting

depended on the measure of clinical response employed and were associated with considerable uncertainty (e.g. 95% LCI to UCI ranged from 13% to 66% and 16% to 47% for remission at week 4 from infliximab and adalimumab respectively depending on dose and trial). Obtaining a valid estimate of effectiveness for the two drugs and for their relative efficacy was plagued with difficulties contingent on the small number of trials, their small size, differences between the populations examined, and uncertainties concerning the most appropriate induction regimen to be used and the imprecision of trial results.

3] Although there exists a core of responders of indeterminate size that maintain an anti-TNF-dependent response, in general the initial good response is not well maintained with extended treatment. This is evidenced in three ways:

a) The percentage of patients in response (or with remission) fades away after the first weeks or so of maintenance therapy.

b) Large numbers of patients drop out of treatment. In ACCENT I 34% of patients dropped out, some before dose escalation, some after; in the active treatment arms 25% withdrew in ACCENT I and [redacted] in CHARM. Amongst responders in CHARM, about [redacted] withdrew from active treatment.

c) Large numbers of patients required dose escalation and/or transfer to open label (CHARM) because of worsening disease. In ACCENT I, 68% in the 5mg/kg arm required dose escalation and 49% of those in the higher dose arm. In CHARM about 30% in the adalimumab arms crossed over to escalated dose or open label therapy.

These results indicate that during extended treatment an appreciable proportion of patients decide there is an unsatisfactory balance between the actual benefit of anti-TNF and its perceived benefits. The withdrawal rates in these trials are not similar to other monoclonal antibody interventions and contrast with the > 90% compliance over 52 weeks observed for IV weekly infused eculizumab.[74] The high requirement for dose escalation reflects efforts to resuscitate a fading response; the continuing drop out rate after escalation shows that these efforts meet with limited success.

Conclusion

Evidence from at least one induction and one maintenance trial for each drug administered within the licensed dose regimen demonstrates that for selected patients, relative to standard care, these anti-TNF agents (infliximab and adalimumab) deliver statistically significant benefits of disease remission and improvement based on CDAI binary measures. Remission, response 70 and response 100 rates measured in maintenance trials indicate that for "responders" nearly all benefit is achieved in about the first 12 weeks of treatment. Thereafter rate differences (anti-TNF minus placebo) remain relatively stable. These results imply that a short burst of treatment is likely to be more clinically and cost effective than prolonged treatment and that after about 12 weeks the likelihood the intervention will

be clinically and cost effective will steadily diminish as treatment is extended unless other favourable outcomes additional to those based on CDAI measures are delivered later than 10 to 12 weeks.

The recruitment of patients who may not have failed alternative treatments together with the selective reporting of outcomes for early responders in the maintenance trials means it is difficult to gauge the effectiveness of these drugs in maintaining favourable outcome amongst the whole patient population with moderate to severe CD who are resistant to other treatments. Because of inappropriate study designs, heterogeneity of patients, incomplete and/or selective reporting of outcomes and lack of head to head trials no convincing objective evidence was available to indicate whether one drug was superior to another either in respect to effectiveness or to safety.

6. ASSESSMENT OF COST-EFFECTIVENESS

6.1 Systematic review of existing cost-effectiveness evidence

6.1.1 Introduction

This chapter explores the published cost-effectiveness literature on the costs and benefits of TNF α inhibitors. Within the UK, the licensed anti-TNF treatments are adalimumab and infliximab. The following section goes on to describe the results of a systematic literature review of these treatments for CD.

When assessing the economic impact of CD, costs can be divided into direct costs and indirect costs. Direct costs refer to the costs of an intervention itself, and include the value of all resources consumed in the provision of an intervention, including all side-effects occurring as a result of treatment, and all future health care expenditures contingent on either the intervention or side-effects. Direct costs include all goods, services and other resources used, both within and outside of the healthcare sector. Healthcare costs include all medication, diagnostic tests, supplies, staff and medical facility costs. Costs outside healthcare can include the costs to the patients and to other public agencies. Indirect costs include those resources consumed that are not directly paid for by any party. Since the ability of patients to work is related to general health and the time spent in treatment, indirect costs include productivity gains and losses. Indirect costs also include the productivity costs of unpaid carers including family members.

In CD the perspective taken may have a significant impact on the costs associated with the disease and the overall conclusions drawn from the evidence. Several perspectives can be adopted and the NICE reference case recommends concentrating only on the direct health care and public social service (PSS) costs. A societal perspective that includes direct and indirect costs to all parties may be considered in sensitivity analyses.

6.1.2 Methods for reviewing cost effectiveness

6.1.2.1 Search strategy

A comprehensive literature search on the cost and cost-effectiveness of adalimumab, infliximab, certolizumab pegol and natalizumab for the treatment of CD from a UK perspective was conducted. [a]

Studies on costs, quality of life, cost effectiveness and modelling were identified from the following sources:

- Bibliographic databases: MEDLINE (Ovid) 2000 to May / June 2007, EMBASE (Ovid) 2000 to May / June 2007, Cochrane Library (NHS EED and DARE) 2007 Issue 2, and HEED database (June 2007).
- Industry submissions
- Internet sites of national economic units

Searches were not limited by language. Full search strategies can be found in Appendix 14.

In addition, searches for cohort studies of infliximab for CD and also clinical guidelines for CD were undertaken for background information for the decision analytic model. Full details can be found in Appendix 14.

6.1.2.2 Inclusion and exclusion criteria

Only studies meeting the following criteria were included.

Study design: Fully published economic evaluations only (abstracts without full publication were excluded)

Population: Patients with CD (adults or children)

Intervention: Adalimumab or infliximab (any dosage/treatment regimen)

Comparator: Conventional treatment without TNF-α inhibitors including no treatment, placebo, dietary intervention, drug treatment with aminosalicylates, methotrexate, corticosteroids (prednisolone, budesonide and hydrocortisone), azathioprine, metronidazole or surgical intervention

Outcomes: cost utility, cost effectiveness

6.1.2.3 Inclusion, quality assessment and data extraction strategies

Two reviewers independently reviewed studies for inclusion using title and abstract. Disagreements were resolved by discussion. After this initial sift, full papers were obtained and assessed for

[a] Natalizumab and certolizumab pegol were originally part of this technology appraisal so were included in the searches;they were subsequently dropped from the appraisal after completion of searches

inclusion. All studies were quality assessed using a standard checklist (Drummond and Jefferson, 1996[75]) by two independent reviewers. If a substantial part of the economic evaluation was missing due to the material being commercially in confidence, formal quality assessment methods were not used. Data extraction of included studies was perfomed by one reviewer, extracted data was then checked by a second reviewer.

6.1.3 Results of systematic review of existing cost effectiveness studies

6.1.3.1 Inclusion and exclusion of studies

Using the search strategy and previous knowledge of the literature, an initial 814 papers (805 + 9) were identified. Initial sifting identified 64 articles for further investigation. These articles identified seven further papers that could have provided relevant information for economic modelling but were not included in the systematic review of cost effectiveness studies.

6.1.3.2 Quantity and quality of included studies

Only four full papers met the review criteria and were subsequently reviewed. These were Arseneau 2001[76], from USA, Jaisson-Hot 2004[8] from France, Marshall 2002[7] from Canada and Clarke 2003[6] from UK. A further two papers were available in abstract form only. Given the difficulty of extracting reliable information from this format, these were not formally reviewed. Several excluded papers provided either the costs or benefits of treatment but not both.

None of the four papers declared any conflicts of interest. Two of the four papers peer-reviewed published works by independent researchers (Arseneau, 2001[76]; Jaisson-Hot, 2004[8]) and one was commissioned by the Canadian Collaborating Centre for Health Technology Assessment (CCHTA) (Marshall, 2002[7]), and one was an HTA report (Clark, 2003[6]) from the UK, commissioned by NICE. Given the restrictions on commercial in confidence information in HTA reports, the HTA was not quality assessed using the checklist.

Of the three quality-assessed studies, the CCHTA report[7] scored highly in comparison to the remaining two papers, and was both clearly written and transparent. These remaining two papers[8,76] (JH; A) omitted several key features (including an incremental analysis of all comparators) and resource usage was outlined in cost terms only.

6.1.3.3 Characteristics of economic studies

All four studies conducted cost utility analyses (CUA) of infliximab and were reported in a total of five papers. One study[6] considered non-fistulising and fistulising disease, two considered non-fistulising disease only[7,8] and one[76] considered fistulising disease only. No published economic studies were found for adalimumab in CD.

Within economic analyses of infliximab in CD, a lack of relevant observational data has lead to the widespread use of information from a relevant study conducted in Olmstead County, and in particular the model constructed in Silverstein (1999)[23]. Using 24 years of data, Silverstein (1999) [23] constructed a Markov model of the course of CD to calculate the excess lifetime costs of the disease. The model considered seven states (remission, mild disease, drug responsive severe disease, drug dependent severe disease, drug refractory severe disease, surgery and postsurgical remission) plus death. This model was not an economic evaluation but has been highly influential in the modelling carried out in non-fistulising CD.

Gregor et al (1997)[77] elicited health-related quality of life values from 180 consecutive Canadian CD patients. This evaluation provided both standard gamble and time trade-off data for hypothetical chronically active CD, acute disease and remitted states, and also by the patient's own health and CDAI status. This information was also used in a number of the economic models reviewed here.

Non-fistulising disease

Within the published models, the comparator treatment strategies comprised surgery and medical treatment (Jaisson-Hot)[8], placebo (Clark[6]) or usual care (Marshall[7]) in populations that were resistant/non-responsive to standard therapy. Only one model (Clarke[6]) was UK-based, with the others based in France (Jaisson-Hot[8]) and Canada (Marshall[7]). The French model[8] was lifetime-based, with the Canadian model[7] taking a one-year timeframe. The timeframe in the UK model[6] was unclear, with the treatment considered up to three re-treatments within a single year but stated that the timeframe of treatment was "1 or more years".

The French (Jaisson Hot et al, 2004[8]), Canadian (Marshall et al, 2002[7]) and UK (Clark et al, 2003[6]) models used the Olmstead County data when estimating transition probabilities. The French[8] and Canadian[7] models used this data to model states where infliximab was not used (in the French case, following surgery and in the Canadian model, from baseline). In the modified industry model in the UK HTA report[6], this Olmstead data was used to define post-remission health states for the infliximab arm. Clark et al[6] noted that the use of this information was likely to lead to bias if used to populate a Markov model that moved CD patients responding to treatment into a remitted state. They noted that the prognosis of those in a remitted state following disease flare and infliximab treatment was likely to differ from those who did not experienced a disease flare in the observational cohort.

Both the French[8] and Canadian[7] models used a third-party payer perspective and, whilst not described, the UK HTA report[6] likely used an NHS/PSS perspective in line with the NICE Reference Case.

Fistulising disease

In fistulising disease, the comparator treatment strategies comprised placebo (Clark 2003[6]) or the combination of 6-mercaptopurine/metronidazole and/or infliximab in different regimens (Arseneau 2001[76]). Both studies (one UK, Clark 2003[6], one USA, Arseneau 2001[76]) used a one-year timeframe. The USA model used a third-party payer perspective but the UK model was unclear on this point but again probably took an NHS and PSS perspective. A lack of existing clinical data beyond 18 weeks required Arseneau et al (2001) [76] to make strong assumptions about both the effectiveness of infliximab as second-line and reinfusion therapies, and regarding the longer-term chances of fistula recurrence.

The four models considered the cost-effectiveness of infliximab treatment for 70kg, adult CD patients (Clark 2003[6] fistulising and non-fistulising, Arseneau 2001[76], Marshall 2002[7]). In the remaining study the assumed weight of the Markov cohort was unclear (Jaisson-Hot 2004[8]).

Calculation of cost data

With the exception of the Marshall (non-fistulising Canadian model) [7] and Arseneau et al. 2001 (fistulising USA model) [76], the assessed models reported costs and resource usage poorly. None of the remaining models reported resource use separately from costs and in many cases the costs of individual items was not given. In the UK HTA model[6], the source of the cost data was not given. Expert opinion was used to estimate resource use items in two models (Jaisson-Hot 2004[8], Marshall 2002[7]).

As the models were typically of one year's duration, there was no discounting. In the single French analysis[8] of a longer duration (lifetime), a discount rate of 5% was used. (Whilst the USA analysis [76]is only one year in duration, it claimed to use a discount rate of 3%. It is not clear how this discounting was calculated.)

6.1.3.4 Health outcomes and data sources

Effectiveness in non-fistulising disease

In the UK model [6] much of the clinical data was removed due to confidentiality reasons. Effectiveness in the model was based on two scenarios where the effectiveness of infliximab was either aggregated across doses (Scenario 1) or based on the 5mg/kg dose (Scenario 2). Scenario 1 gave lower effectiveness estimates and was used in the Company submission. Values for both scenarios were given in summary tables. The French model (Jaisson-Hot 2004[8]) calculated effectiveness data from published evidence (Targan 1997[54]) and expert opinion but details were unclear. The Canadian model[7] used effectiveness data from the Targan 1997[54] and Rutgeerts 1999[55] trials.

Effectiveness in fistulising disease

The USA model by Arseneau 2001[76] converted pooled data from twelve studies to calculate transition probabilities. The model assessed benefits through fistula improvement rather than closure, so that it the improved state included both complete closure and symptomatic improvement. Whilst acknowledging that clinically relevant endpoints were a subject of debate, they acknowledge that this choice of definition was likely to increase the effectiveness of treatment and may have biased estimates.

The UK HTA model[6] used data from the Present 1999[57] study for fistulising disease but no details of precise estimates were given.

Utility estimates

Utility estimates were based on Gregor 1997[77] in the three non-fistulising models. As the Gregor estimates did not include fistulising states, separate estimates were used in the models for fistulising disease. The USA fistulising model by Arseneau 2001[76] used standard gamble utilities from 32 CD patients and 20 healthy volunteers, whilst Clark's (UK) modified industry model[6] used an unpublished algorithm based on CDAI and Perianal disease activity index (PDAI) scores from the Gregor data.

6.1.3.5 Cost-effectiveness results

The comparison of cost-effectiveness results across studies is always problematical. For comparison purposes, the methods used were to transform cost estimates based on purchasing power parities (as appropriate) and reflate according to all-item UK retail prices index figures[78] to provide estimates in £2006 where possible. Where the base year for costs was not given, figures could not be reflated and the original stated values are used here.

Differences in comparators, methods, data, and the non-disclosure or removal of pertinent information prevents reliable interpretation of the results of such comparisons. In these results, caution should be taken in the interpretation, since incremental cost-effectiveness ratios relate to the cost of increasingly more effective treatments whilst cost-effectiveness ratios may be compared to a common comparator. The former is preferred as it allows assessment of the marginal costs and effectiveness of treatment.

In only two of the models (Marshall, Arseneau) [7,76] were total costs and effectiveness data given for all the compared strategies. In only one of the models (Marshall, 2002) [7] was it possible to calculate incremental cost-effectiveness ratios. Back-calculation of figures was avoided as this may have introduced errors, whilst transforming provided figures is hazardous given that they are not displayed to sufficient precision.

In non-fistulising disease, the UK model[6] compared single and episodic treatment versus placebo. Against placebo, episodic treatment (a single 5mg/kg dose plus up to three 5mg/kg re-treatments within a single year) was estimated to have an ICER of £62,016 per QALY when using effectiveness data from the 5mg/kg dose group (base year not given). Treatment with a single dose of infliximab (no episodic re-infusions) was found to be less cost-effective.

The French model[8] estimated the cost-effectiveness against usual care only. As neither total nor incremental QALY figures were given (and back-calculating is not reliable) incremental figures could not be calculated. Against usual care, episodic treatment and maintenance treatments of infliximab were estimated to have cost-effectiveness ratios of €63,700 and €784,057 per QALY (base year not given).

The converted results from the Canadian model[7] suggested an incremental cost-effectiveness of £105,900 per QALY for single dose infliximab vs usual care, £280,600 per QALY for episodic vs single dose infliximab, and £407,000 per QALY for maintenance vs episodic treatment with infliximab.

In fistulising disease, the UK model[6]suggested a cost-effectiveness ratio of £102,000 (base year unclear) for initial infliximab treatment versus placebo. In the US model[76](converted to 2005 pounds sterling), only cost-effectiveness ratios could be calculated as the outcomes figures were not given with sufficient precision. Against the comparator treatment of 6-mercaptopurine and metronidazole (6MP/met), the interventions had cost-effectiveness ratios of £274,100 per QALY (infliximab, with 6MP/met as second line treatment), £278,000 per QALY (infliximab, with infliximab reinfusions as second line treatment), and £290,770 per QALY (6MP/met + episodic infliximab reinfusion).

Sensitivity analysis

The reporting of sensitivity analyses was variable, with probabilistic sensitivity analysis conducted in only the Canadian model[7]. This analysis suggested that usual care was favoured up to a threshold of approximately £105,000 per QALY, with a single dose of infliximab favoured between this figure and approximately £251,000 per QALY in non-fistulising disease (converted figures). Whilst this study suggested that the rate of surgical admissions for drug-refractory treatment had little effect on cost-effectiveness, it was sensitive to the variations in the cost of infliximab. With a sufficiently large price reduction it suggested that infliximab treatment may have become cost-effective.

In the UK study[6] for non-fistulising disease, the one-way sensitivity analyses conducted on utility, duration of response and the rate of averted surgeries did not result in any ICER below £40,000 per QALY (base year not given).

In fistulising disease, the UK model[6] varied the rate of success in re-treatment and re-closure of fistulas alongside the level of cost offset due to averted surgeries. In no case did this produce a ratio below £80,000 per QALY. In the other fistulising model (USA)[76], all ICERs remained above £79,000 per QALY (converted figures) regardless of the changes made in one-way sensitivity analyses other than in the price of infliximab. Only where the price of infliximab was reduced to £160 per dose (a reduction of 90% from the modelled price) did the ICER for episodic reinfusion fall marginally below £30,000 per QALY (converted figures).

6.1.3.6 Author conclusions

All of the studies considered were conducted by non-industry authors, with the Canadian[7] and UK[6] studies commissioned by national HTA bodies. The remaining studies[8,76] did not disclose any industry affiliation.

The results of all non-fistulising studies suggested that infliximab was not necessarily cost-effective over the usual range for thresholds. The study by Jaisson-Hot 2004 (France) [8] suggested that episodic infliximab treatment could possibly have been cost-effective but that maintenance treatment may not have justified the increased costs required. The study by Clark (UK) [6] suggested that the cost-effectiveness was relatively insensitive to changes in the key assumptions in their model but that the key criterion for the cost-effectiveness of episodic treatment would have been the duration of benefit.

The study by Marshall (Canada) [7] was limited by the use of non-Canadian data, the need to convert utility data to populate estimated states, the use of expert opinion to inform resource usage and the lack of longer term clinical data. They noted that whilst CD may severely impact morbidity and affect productivity, there was no detailed information available on productivity losses to make allowances for this. They justified the relatively short timeframe in their model with the lack of clinical data to populate a longer term model.

In fistulising disease, Clark (UK) [6] stated that the cost effectiveness ratios were high under even the most favourable assumptions for re-treatment and closure in the industry model. The study by Arseneau 2001 (USA) [76] suggested that the high cost-effectiveness ratio for infliximab was due to both the high incremental cost of infliximab and a similar effectiveness to 6MP/metronidazole treatment. They acknowledged the difficulties with the "fistula improvement" state and noted that

infliximab may have been more effective if it had promoted closure rather than merely symptomatic improvement. The availability of only 18 week data was also acknowledged as a difficulty.

6.1.3.7 Conclusions

There have been no published studies on the cost-effectiveness of adalimumab alone or in comparison to infliximab. Given the lack of comparison between both alternative treatments considered here, it is not possible to infer the relative cost-effectiveness of treatments from existing evidence. Also, whilst the indirect productivity costs of non-treatment may be appreciable in CD these costs were not included in the cost-effectiveness studies due to a lack of evidence as to their magnitude.

All four studies reviewed were conducted by non-industry authors, with the Canadian[7] and UK[6] studies commissioned by national HTA bodies. The remaining studies[8,76] did not disclose any industry affiliation. Taken together, the papers suggested that single use or episodic treatment with infliximab has a relatively high cost-effectiveness ratio for both non-fistulising and fistulising disease. Maintenance therapy was only considered for non-fistulising disease and this is partly due to its potentially prophylactic role in this disease group. Both models considering maintenance infliximab therapy suggested that it would have a particularly high cost-effectiveness ratio relative to both standard care and episodic infliximab treatment.

Full details for included studies of study characteristics, models used, costs and resources, efficacy data, total costs and outcomes, sensitivity analyses and author conclusions can be found in Appendix 15.

6.2 Critique of the submission on infliximab by Schering Plough

6.2.1 Model structure and inputs

The economic component of the Schering Plough submission[79] took the form of a cost utility analysis from the perspective of the health service provider. The model structure was informed by the structure of ACCENT I[3,4]and included the 'episodic' treatment over which concern has been previously expressed (see glossary). The term 'infliximab clinical discretion' (ICD) has been used in place of 'infliximab episodic treatment' in this analysis.

The stated aim of the industry model was to determine the cost effectiveness of infliximab scheduled maintenance treatment (IMT) compared to ICD and standard care (SC) without infliximab among patients with severe active CD (CDAI scores 220-400) in England and Wales. The scheduled maintenance treatment consisted of 5mg/kg infliximab at wk 0, 5mg/kg infliximab at wks 2 and 6 and every 8 wks thereafter. Those receiving ICD received an induction dose of 5mg/kg infliximab at wk 0 and thereafter 5mg/kg infliximab according to clinical discretion. Those receiving standard care received a placebo infusion at wks 2 and 6 and every 8 wks thereafter.

The same basic model, albeit using slightly different data sets, was used to compare IMT vs. standard care without infliximab for patients suffering from fistulising CD and for paediatric CD patients. Note that published reports of ACCENT I[3,4] do not inform the effectiveness of infliximab in these groups.

The analyses were primarily based on data from two recent trials, ACCENT I[3,4] and ACCENT II.[60] Further trial data came from Targan (Targan 1997,[54]), Present (Present et al, 1999[57]) and REACH (Hyams et al, 2007[42]). ACCENT I[3,4] was designed to compare a single 5mg/kg infusion of infliximab followed by maintenance or a placebo for patients with CD. Participants were recruited from across North America, Europe and Israel. Participants must have had CD for more than three months and a CDAI score of between 220 and 400. All participants were given 5mg infliximab at week 0. At week 2, whether participants were responders or not, they were randomly assigned to one of the following three groups:

Group I: Placebo infusion at wk 2, 6, and every 8 weeks thereafter to wk 46 (n=188)
Group II: 5mg infliximab at wk 2, 6, and every 8 weeks thereafter to wk 46 (n=192)
Group III: 5mg infliximab at wk 2, 6, followed by 10mg every 8 weeks thereafter to wk 46 (n=193).

Response was defined as a decrease in CDAI score of ≥70 points and a minimum 25% reduction in total CDAI score. After week 14 patients who initially responded but experienced exacerbation of symptoms could cross over to 5, 10, or 15mg of infliximab on an 'as needed' or 'episodic' basis.

The ACCENT II[60] trial compared long term treatment regimens for patients with fistulising CD. Participants all had CD with single or multiple draining fistulae and were recruited from across North America, Europe, and Israel. All participants were given 5mg infliximab at week 0, 2, and 6. At week 14 all patients, regardless of whether they were responders, were randomly assigned to one of the following two groups:
Placebo infusion at wk 14, 22, 30, 38, 46 and follow up at wk 54 (n=99)
5mg infliximab at wk 14, 22, 30, 38, 46 and follow up at wk 54 (n=96)

In this trial, response was defined as a reduction of at least 50% from baseline in the number of draining fistulae at consecutive visits 4 weeks apart. After week 22, patients in the placebo group who experienced a loss of response could crossover to IMT of 5mg of infliximab.

The economic analysis used a Markov model to simulate the progression of patients and to calculate the cost per QALY of the infliximab treatment over a five year period. For severe active CD the model states were active, remission, death, non-responding active (patients who failed to respond either initially by week 2 or discontinued treatment and the second week due to loss of response), surgery, post surgery remission and post surgery complications. The fistulising model replicated this but expanded the active state to: active + fistula closure, active + fistula, remission + fistula closure, remission + fistula. The severe active and fistulising models also differed in as much as in the severe active patients stayed in the active state (on treatment) for the first two weeks (assessment at week 2) whilst for the fistulising model they stayed in the active state for the first 14 weeks (reflecting the first assessment period at week 14). Transition probabilities for the active state were based on ACCENT II[60] and Present 1999[57] trial results. The transition probabilities for the 'on treatment' health states were estimated from the Targan 1997[54] and ACCENT I[3,4] studies; whilst the transition probabilities for the 'off treatment' health states were estimated from literature. The paediatric model state mirrored those of the severe active model with transition probabilities based on data from the Targan1997[54], ACCENT I[3,4] and REACH[42] studies.

The probability of surgery and post surgical states were obtained from a variety of sources (Marshall et al 2004[7] Wolters et al 2007 [80] and Jess et al 2007 [81]). The authors assumed an equivalent surgery rate (64%) in all three groups (severe active, fistulising and paediatric). Post surgery complications were estimated from Marshall et al 2004[7] which showed no significant differences between groups

with and without infliximab prior to surgery so a weighted average was used. Recurrence rates were based on those from Wolters et al 2007[80]. Whilst the study contained data from nine European countries this did not include the UK; expert opinion was sought to confirm the similarity of the estimates with the UK.

The methods for the estimation of quantities and unit costs were, in the main, comprehensively described. The cost of hospitalisation and assessments used data taken from Jewell et al 1998[9] (a UK study). This retrospective observational study (n=205) compared resource use six months pre- and six months post-infliximab. The pre infliximab figures were used for standard care. Data on post surgery health states (post surgery remission and post surgery complications) were not available so resource use was estimated by an expert panel (consisting of UK gastroenterologists) and the average estimates taken. The cost of immunomodulators were excluded from the analysis on the basis that the efficacy of the treatment would not have been affected (the authors did a post-hoc analysis of ACCENT I[3,4] and II[60] trials that indicated that there was no significant difference in infliximab treatment effect with or without immunomodulators). Adverse events were assumed to be included in the infusion and hospitalisation costs. The cost of infliximab infusions was estimated using an average adult body weight of 60kg which the authors stated was based on previous guidance from NICE. The costs of administering infliximab infusions (£96 per infusion) were said to have been taken from a previous study (referenced as: NICE. Administration costs estimated from CRD/CHE Technology Assessment group Psoriatic Arthritis HTA.[82] However the costs given in that study were considerably higher than the figure given (ie £257.50 per infusion).

Health state utilities were based on a number of different data sources. For severe active CD, health state preferences for pre surgery were taken from a Spanish study (n=201 CD patient responses to EQ-5D and converted into utilities using UK tariffs).[83,84] Surgery and post surgery preferences were based on data from a secondary care database of patients in Cardiff and Vale of Glamorgan.[31] No information was available for complications post surgery and a utility value of 0.4 was assigned (however, the table presented reported this as 0.5 so clarification is needed on this point). The authors justified the value given in terms of complications that would lead to significant hospitalisation.

The transitional probabilities were subject to sensitivity analyses with the exception of post surgery health states because no treatment effect was assumed beyond surgery. One way sensitivity analyses was conducted on patient weight, time horizon, discount rate, baseline age and infliximab administration cost and the resultant cost per QALY reported.

6.2.2 Model results

The results of the analysis are shown in Table 43 below.

Table 43. Cost-effectiveness of infliximab (Schering-Plough submission)

Time	Treatment	Mean Costs	Difference	Mean QALYs	Difference	ICERs
Severe Active CD						
5year	Maintenance	31040		2.145		
	Standard care	26209	4831	1.959	0.186	**25903**
5 year	ICD	25501		2.133		
	Standard care	26209	-708	1.959	0.174	**Dominant**
5 year	Maintenance	31040		2.145		
	ICD	25501	5539	2.133	0.01	**457386**
Fistulising						
5 year	Maintenance	36626		2.449		
	Standard care	30577	6049	2.247	0.202	**30005**
Paediatric						
5 years	Maintenance	33504		2.566		
	Standard care	27672	5833	2.146	0.42	**13891**

For infliximab maintenance therapy vs. standard care, changes in the discount rate for costs and QALYs resulted in a range of cost per QALY between £24,588 and £27,296. For a time horizon of two years and lifetime the cost per QALY was £26,462 and £28,432 respectively. Little change was seen as a result of changes to baseline age whilst an increase in the cost of administering infliximab from £96 to £124 increased the cost per QALY to £26,751.The largest increase was seen as a result of changes to patient weight, when 80kg was used the cost per QALY was £38,623 and for 70kg, £32,263. The latter figure assumed vial sharing. Similar results were shown in fistulising CD and paediatric CD where again the largest increase to cost per QALY resulted from changes in weight (80kg, £44,459; 70kg £37,232 for fistulising CD and 30kg, £8,942; 50kg, £18,841; 60kg, £23,791 for paediatric CD). For ICD treatment vs. standard care, the ICD treatment remained dominant in all scenarios with the exception of when patient weight was assumed to be 80kg when the cost per QALY was £445.

The authors concluded that the analyses demonstrated the cost effectiveness of infliximab maintenance in severe active CD and sub groups. The only caveat given was with regard to the resource estimates used for patients with infliximab. This information was taken from the study of 'episodic' care but was applied to both ICD AND infliximab maintenance therapy within this submission which, the authors suggested, meant that resource reductions from IMT were not captured.

6.2.3 Evaluating the industry submission for infliximab maintenance versus infliximab episodic and versus standard care

The design of the cost utility analysis is in line with what would be expected for this type of submission but the results are limited by a number of factors. Whilst the comparators appear to be justified, the analysis comparing both standard care and infliximab maintenance to ICD care is hampered by the definition of 'episodic care' used for the ICD comparison. Although episodic care was described as treatment 'as required', no further details were given.

The analysis relied heavily, but not exclusively, on the ACCENT trials. The information used in the economic analysis was based on responders only. Thus responders only inform the clinical effectiveness which is likely to overestimate the clinical results. In addition, the costs of those who did *not* respond do not appear to have been included, giving lower estimated costs. Both scenarios produce a lower cost per QALY.

In line with the other industry submissions, the primary comparison is with standard care. The rationale for this comparator is that the majority of patients eligible for biological treatments in England and Wales still receive standard care. Whilst the authors cite market research as evidence of this, unfortunately the information cited is not in the public domain. Also, the ACCENT trials were conducted outside the UK so it is not possible to determine how 'standard care' in the trials compares to that in the UK.

Throughout the submission a CDAI score of 220-400 was used to indicate severe active CD. Whilst there is no formal quantification of the level at which moderate CD becomes severe active, 220 is lower than has been used in a number of other studies, which makes comparison difficult.

The analysis used a Markov model. Markov models assume zero memory. How long a patient has been in a health state and how they got there may impact on resources and memory could be important in this patient group.

6.2.3.1 Evaluating the industry submission: Active CD (220 < CDAI < 400)

Clinical effectiveness of the comparators in the economic model

The economic model compared maintenance (IMT), ICD and standard care. As the details of ICD treatment are unclear, it is not possible to satisfactorily verify or interpret the model. In particular, the description of ICD treatment neither guarantees episodic care nor precludes the use of maintenance treatment. It is thus both extremely broad in definition and does not guarantee that clinically identical individuals would receive the same treatment. ICD is limited in the degree to which it represents an identifiable treatment strategy.

Much of the model was based on the ACCENT I[3,4] trial. This trial distinguished between those who did and did not respond by Week 2 on both the placebo maintenance and infliximab maintenance arms (of which the economic analysis considers only the 5mg/kg dosage). The placebo arm in the ACCENT I[3,4] trial included treatment with 1a) placebo maintenance treatment and 1b) ICD at 5mg/kg for those not responding to 1a. The infliximab arm in the ACCENT I[3,4] trial included treatment with 2a) maintenance at 5mg/kg and 2b) ICD at 10mg/kg treatment for those not responding to 2a.

Hanauer (2002) [4] compared the outcomes for Week 2 responders 1a) with treatment 2a), with those crossing to 1b) and 2b) considered to be treatment failures. Rutgeerts (2004)[3] attempted to compare outcomes for both Week 2 responders and non-responders in both their initial and crossover treatment, attempting to compare 1a) plus 1b) with 2a) plus 2b). As above, this comparison is inappropriate.

The economic model attempted to compare 1a) plus 1b) (as ICD) against only 2a) (as infliximab maintenance). Confidential clinical information in the industry submission suggested that, thus constituted, ICD would provide very similar and potentially superior clinical outcomes to maintenance therapy.

The clinical study report included data on how many patients retained a response to treatment at Week 30. Amongst Week 2 responders, 51% of those receiving IMT retained a response, as against [redacted] of those receiving standard care (placebo) at Week 30 (Week 54 Clinical Study Report). The ICD arm is based on those failing on placebo treatment receiving infliximab at 5mg/kg. All those receiving infliximab on ICD would have been considered failures and on this basis, we would expect that ICD would have the same effectiveness in retaining a response as placebo treatment.

A clearer comparison is available using the Week 30 Clinical Study Report includes data for both Week 2 responders and non-responders, those who crossed over, and those who received protocol-

prohibited medication changes or surgery. At 30 weeks, [redacted] of patients receiving ICD maintained a clinical response to treatment. As it would be expected that non-responders would have poorer outcomes than responders these results may have underestimated the effectiveness of ICD for Week 2 Responders. Given this, it would not be surprising for ICD to outperform maintenance therapy in health outcomes achieved, and any advantage for maintenance over ICD is likely to be very minor. As maintenance therapy patients received infliximab regularly, whereas ICD patients received infliximab according to clinician discretion (at the same dosage), ICD would be less expensive than maintenance. It appears that infliximab maintenance therapy is very unlikely to be cost-effective against ICD.

Use of Trial Data

Although the treatment scenarios were well presented, there were limitations, which were primarily associated with the sample characteristics of the studies used. The active CD treatment strategies estimated were based on ACCENT I[3,4] data together with data from another smaller study -Targan 1997[54]. The Targan 1997[54] study was used for transitions between Week 0 and Week 2, with transitions following Week 2 estimated using ACCENT I[3,4]. The use of Targan 1997[54] data is questionable as it comes from a smaller trial preferred in place of a larger trial (ACCENT I[3,4]) which provided less positive results.

The economic model also appeared to use two different populations in active CD, with both the standard care (placebo) and ICD arms using both Week 2 responders and non-responders and the maintenance arm using Week 2 responders only. If Week 2 responders do indeed have a better response than Week 2 non-responders, then this is likely to bias the comparisons in the economic model. Given the data above, this bias may account for any positive effect found for infliximab maintenance against ICD.

Inputs

The cost of drug infusions was estimated using an average adult body weight of 60kg which the authors' state was based on previous guidance from NICE. This is likely to have underestimated the cost per QALY. The Targan 1997[54] trial recorded mean body weights of between 68-74kg (for different treatment groups) whilst for ACCENT I[3,4] these are only recorded in the (confidential) clinical study reports [redacted].
There were four larger trials in the clinical effectiveness section (where n>100) that gave mean weight of included patients (Targan 1997[54], CLASSIC, GAIN[59] and CHARM[62]) and the mean weight was approximately 71.5 Kg. One way sensitivity analysis is carried out at the end of the industry

submission using weights of 70kg and 80kg increased the cost per QALY to £37,232 and £44,459 respectively.

A weight of 60kg exactly corresponds to the use of three 100ml vials of infliximab, and the model therefore assumed no wastage. The revised model uses a weight of 70kg, which remains a conservative estimate. The price of infliximab within the model was also increased in line with the figures cited in the industry; previously a slightly lower figure had been used. Any cost of wastage was not incorporated in the model.

Administration costs were taken to be half a day case - H26 (Day Case Rheumatology) in line with a recent HTA report (HTA, Psoriatic Arthritis[82] states psoriatic arthritis in submission references), which was £293.67 in 2006 Pounds Sterling.

Calculation of cost-effectiveness acceptability curves

The cost-effectiveness acceptability curves (CEACs) for active CD (220<CDAI<400) were calculated for ICD and maintenance therapy versus standard care but not together. All three comparator treatments should have appeared on the same CEAC as the piecewise comparisons are misleading. The CEACs provided in the industry submission are critically flawed as they did not distinguish between:

1) the case where an option is dominated or dominant and

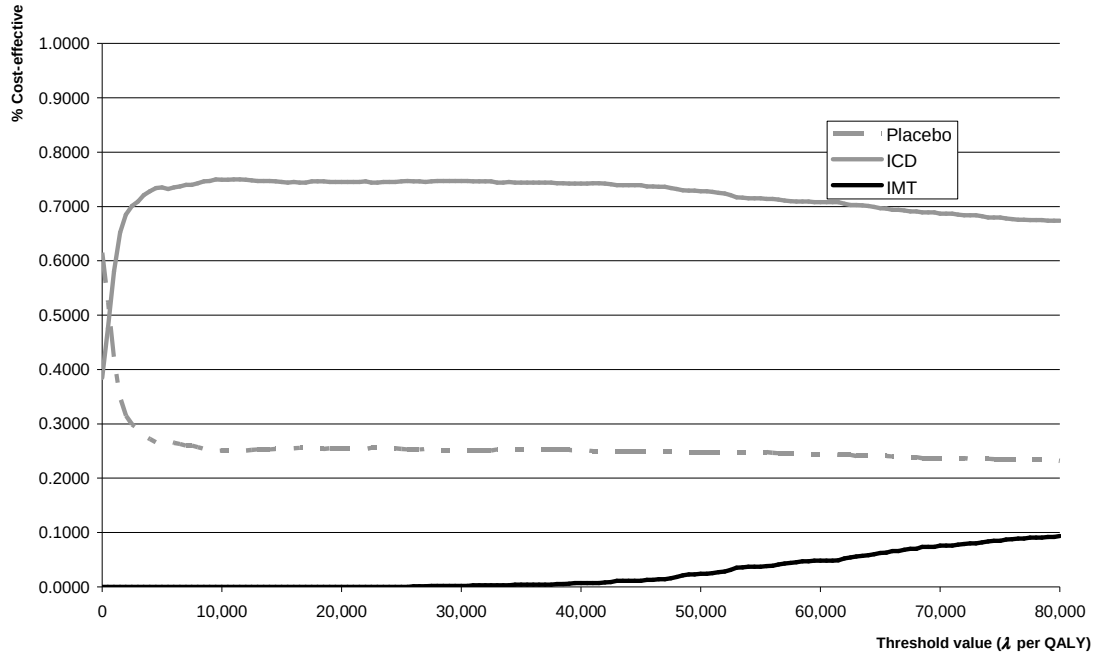
2) the case where an option is cost-effective above or below a critical value.

The probabilistic sensitivity analyses were re-calculated using appropriate methods and the changed input for weight, as above. One thousand iterations were used to construct this estimate.

Results from recalculation

Prior to re-calculation, the CEACs provided in the industry submission suggested that both ICD and maintenance care were increasingly likely to be cost-effective against standard care as the threshold value increased. Following re-calculation, the conclusions of the economic model do not correspond to those suggested by the industry submission. For threshold levels between £0 per QALY and £2,466 per QALY, placebo had the highest chance of being cost-effective. For threshold levels between £2,466 per QALY and approximately £481,000 per QALY, ICD treatment had the highest chance of being cost-effective. Infliximab treatment according to clinical discretion appears to be cost-effective, although this is contingent on a series of caveats, including the ill-defined nature of the "episodic" intervention itself.

Figure 33. CEAC for Active CD



6.2.3.2 Evaluating the industry submission: Fistulising CD

Use of Trial Data

For fistulising CD, the industry submission stated that evidence from the ACCENT I[3,4] trial suggested that maintenance infliximab therapy may bring significant QALY gains related to improved quality of life (as opposed to improved life expectancy) in adults with fistulising CD. However, it is not possible to ascertain from the submission or from published papers of the trial whether the sample included adults with fistulising CD. The submission did however report evidence presented in the ACCENT II[60] trial (fistulising CD patients) that showed a significantly longer time to loss of response for infliximab maintenance vs placebo maintenance and significantly improved CDAI scores for the infliximab group vs. placebo group.

The fistulising CD strategies were based on responders only. The ACCENT II[60] trial showed that 69% of the sample were responders after the induction period. There was, however, no placebo comparison during this period so it is not possible to determine the proportion of patients who would have gone into remission without infliximab. Whilst the proportion of responders was higher than the ACCENT I[3,4] trial, the definition of a responder differs and the time at which they were assessed as a responder/non-responder was much later (14 weeks rather than two weeks). This highlights the arbitrary nature of the time point chosen to identify responders and the impact it may have had on the results.

The treatment strategy was modelled on the Present 1999 trial[57](0-14 weeks) and the ACCENT II trial[60](14-54 weeks). The Present 1999 trial[57], like the Targan 1997 trial[54], had relatively small numbers in each arm (31-32). The ACCENT II[60] scheduled maintenance arm excluded those who switched to 10mg episodic treatment after week 22. Again it is not clear how standard care in the UK compares with that in the Present 1999 trial[57] (recruitment United States and Europe) and the ACCENT II trial.[60]

Inputs to fistulising model

As with severe active CD, the cost of drug infusions was estimated using an average adult body weight of 60kg which the authors' state was based on previous guidance from NICE. Questions remain about the suitability of this figure given that it is lower than the values found in clinical trials on which the analyses are based. As above, a figure of 70kg was used when considering this and changed both the cost of infliximab and administration costs to a more accurate figure.

The health state utilities were based on two different sources.

- A Spanish study (n=201 CD patients) measuring EQ-5D then converted into utilities using UK tariffs) which assumed that the preferences measured are concordant with UK patient preferences [83,84]
- A secondary care database of patients in Cardiff and Vale of Glamorgan measuring surgery and post surgery preferences. This was based on a small sample size and looked at surgery (less than two months after surgery, n=17) and remission post surgery (ie more than two months with no recurrence/complication n=21).[31]

Despite specific utility estimates being available for the fistulising CD model, they were not used because they are generally higher than the utilities found in the Spanish study, which was used to provide values in the Severe Active CD model. The authors stated that the estimates were not in accordance with the NICE reference case because they were taken from CD patients and healthy individuals (n=32 and n=20 respectively). The utilities allocated assumed that patients with fistula closure had identical utilities to patients without fistulas. For all other fistula states 0.15 was deducted from utility estimates, the authors gave no explanation for this figure but did include the variable in the sensitivity analysis.

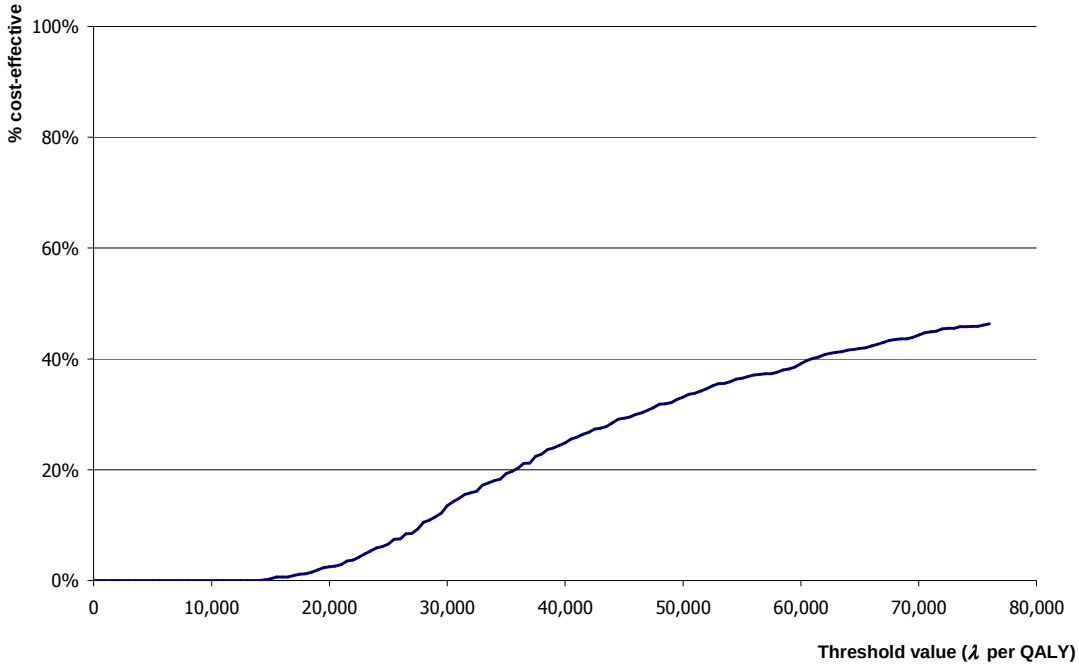
Calculation of cost-effectiveness acceptability curves

Similar problems were found in the calculation of CEACs in the fistulising CD model as above in Severe Active CD.

Results from recalculation of the model

Prior to re-calculation, the CEACs provided in the industry submission suggested that maintenance care was increasing likely to be cost-effective against standard care as the threshold value increased (Figure 34). At £20,000 and £30,000 per QALY, infliximab treatment was found to be cost-effective 32.5% and 48.1% of the time. Following the weight adjustment and recalculation of the CEAC, the curve shifted downwards. Now, at £20,000 and £30,000 per QALY infliximab treatment was found to be cost-effective only 2.5% and 13.5% of the time.

Figure 34. CEAC for Fistulising CD (Infliximab maintenance vs Placebo)



6.2.3.3 Evaluating the model: Paediatric CD

The economic model provided with the industry submission contained both circular references and broken links. Circular references mean that Excel must select on possible values for cells as it is not possible to calculate it. Whilst broken links may be repaired with the provision of the files containing the information on which the model is based, the file was not available. Any analysis conducted on this model would necessarily be based on such computational guesswork and would not withstand scrutiny. Given that there was no access to a functioning model we were able to neither verify nor respond to it beyond making the following comments. Note that the model and any cost-effectiveness figures based on it may have contained further errors of either the same or a different form to those identified above. Unless a functioning version of the model is provided it is not possible to consider the specifics of the Excel model.

Use of Trial Data

For paediatric CD, the submission states that evidence from the ACCENT I[3,4] trial 'suggested that maintenance infliximab therapy may bring significant QALY gains, related to improved quality of life (as opposed to improved life expectancy) in adult and paediatric patients with CD'. Whilst the sample characteristics provided in the trial summary do not show whether paediatric CD patients were included in ACCENT I[3,4], a paper in the Lancet (Hanauer et al, 2002[4] reporting on the ACCENT I[3,4] trial gave the patient age range as 18-76 suggesting no paediatric patients were included. It is difficult to tell whether the authors' statement is based on data from these subgroups or previous evidence.

Paediatric CD strategies were based on data from Targan 1997[54], REACH[42] and ACCENT I[3,4]. For both the scheduled maintenance and standard care arms, the model was based on the Targan 1997[54] trial response rates at Week 2. The Targan 1997[54] study was not a paediatric study and no age range was given (the mean ages that were given at baseline were 36 and 39.3 years). As there was a small sample in Targan 1997[54] it is unlikely that a sub sample of paediatric patients was used. The transitions of these Targan 1997[54] study responders were estimated using data from the REACH[42] trial (at between 2-54 wks). The REACH[42] trial was a paediatric study that compared scheduled maintenance every eight weeks vs every twelve weeks. Whilst the study only assessed for response at week 10, the response rate was particularly high (99/112) compared to the ACCENT trials. However, the REACH[42] study did not have a placebo arm so it is not possible to determine the proportion of patients who would have been classified as responders without infliximab.

The comparison of the two treatment strategies used in the model is inappropriate. The standard care arm was based on ACCENT I[3,4] data from week 2 onwards. Like the Targan 1997[54] study, ACCENT I[3,4] is not a paediatric study; thus the paediatric standard care treatment strategy is based only on adult study data; paediatric data is only used in the infliximab scheduled maintenance arm.

As with the adult comparisons, the paediatrics model was based on an optimistic assumption of 40kg weight. Of the paediatric studies used in the analysis, the mean weights recorded were between 45-55kg and 42-48kg.

6.2.3.4 Discussion of the Schering Plough submission

The Schering Plough submission included three sub-models considering 1) active CD for the CDAI range covered in the ACCENT I[3,4] trial including both moderate and more severe forms of CD, 2) fistulising CD, and 3)paediatric CD. The industry submissions contained errors, some of which were addressed in the revisions above. Others could not be corrected, such as the selective use of responders only in the infliximab maintenance arm in the Active CD model.

For active CD, the corrected models suggested that infliximab treatment (ICD) could be cost-effective, whilst maintenance care was unlikely to be cost-effective even at low multiples of the normal threshold values. The lack of detail on what constitutes ICD or "episodic" treatment is unhelpful.

To the degree that they can be investigated, the models provided by Schering Plough mostly meet the NICE reference case. There remain issues regarding the selection of studies, the use of data within the selected studies and some inputs used in the modelling. The utility data used in one model relies on a small sample but is broadly in line with the reference case.

Table 44. Compatibility of the industry model with the NICE reference case

Element of health technology assessment	Principles	
Defining the decision problem	The scope developed by the Institute	Yes
Comparator	Alternative therapies routinely used in the NHS	Yes
Perspective on costs	NHS and PSS	NHS only
Perspective on outcomes	All health effects on individuals	Yes
Type of economic evaluation	Cost-effectiveness analysis	Yes
Time horizon.	Sufficient to reflect any differences in costs or outcomes between the technologies compared.	Yes
Synthesis of evidence on outcomes	Based on a systematic review	Partial; doubt remains on selection of studies
Measure of health benefits	Quality-adjusted life years (QALYs)	Yes
Description of health states for calculation of QALYs	Health states described using a standardised and validated generic instrument	Yes
Method of preference elicitation for health state valuation	Choice-based method, for example, time trade-off, standard gamble (not rating scale)	Yes
Source of preference data	Representative sample of the public	Partial; one source of particularly small sample size
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes
Equity position	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
Modelling methods	Structural assumptions and data inputs clearly documented and justified.	Partial; assumptions made not justified
	Probabilistic sensitivity analysis should be conducted.	Yes

6.3 Critique of Abbott submission (adalimumab)

6.3.1 Introduction to the evaluation

An economic analysis was conducted for Abbott for their submission[5] to the National Institute for Health and Clinical Excellence by Analysis Group. The submission comprised two economic models – one comparing the cost-effectiveness of adalimumab as a maintenance therapy against standard care and one comparing the cost-effectiveness of adalimumab and infliximab as maintenance therapies. This latter model will be relevant only where both adalimumab and infliximab have been first justified as maintenance therapies versus standard care. Where one or both maintenance therapies are not cost-effective versus standard care, this comparison provides no information to decision makers.

This evaluation therefore begins by concentrating on the former model assessing the cost-effectiveness of adalimumab as a maintenance therapy. The model considers both fistulising and non-fistulising forms of CD together, and comprises a printed economic submission and accompanying working model in Excel. The economic model contained several assumptions that were not fully explained or justified at its initial submission. Abbott took the opportunity in responding to the draft assessment report to clarify some of these issues raised and their model incorporates elements of the health economic critique to their earlier version. At neither stage was a full and working probabilistic sensitivity analysis provided to identify all sources of uncertainty. However, a sufficient quantity of detail was provided in Abbot's revised version to allow replication of a probabilistic sensitivity analysis that incorporated only some sources of uncertainty.

This interpretation appears to be consistent with the Excel model provided but the interpretation of the model contained in the printed economic submission does not appear to be consistent with the Excel model provided.

6.3.1.1 Model inputs and structure

The stated aim of the company's primary submission was to produce a comparison of lifetime maintenance on adalimumab versus standard care. The adalimumab arm of the models was based on data up to Week 56 in the CHARM trial[62], which was then extrapolated to also produce a lifetime analysis by assuming that all those responding at Week 56 would continue to respond for the remainder of their lives. A regression based on the CLASSIC I trial[58] was used to provide standard care outcomes for the CHARM[62] arm. The company's comments received to our modified model are acknowledged below.

All patients enrolled in the CHARM[62] trial had baseline CDAI scores between 220 and 450. In this trial, all patients were given open label 80mg at week 0 and 40 mg at week 2 and then randomised blind at week 4 to a placebo, adalimumab 40mg every other week or adalimumab 40mg weekly. After Week 12, those who did not respond to randomised treatment (defined as a drop of less than 70 points in CDAI) were switched to open-label adalimumab 40mg every other week, as were those "responders" who experienced a treatment flare after Week 12. Those not responding to open-label adalimumab 40mg every other week were switched to adalimumab 40mg weekly. Those not responding to 40mg weekly were returned to standard care.

In the CLASSIC I[58] trial, patients had a baseline CDAI between 220 and 450 and had had no previous exposure to any anti-TNF therapy. There were 299 individuals who were randomised to either placebo (n=74) or adalimumab induction regimens in Weeks 0 and 2 of 40mg and 20mg (n=74), 80mg and 40mg (n=75), or 160mg and 80mg (n=76). All patients were followed for four weeks and the primary endpoint was the proportion with a CDAI score less than 150 in Week 4.

The industry submission compared the cost-utility of the 40mg adalimumab every other week (eow) strategy versus standard care. As the standard arm of the CHARM[62] trial began with adalimumab induction at 80mg in Week 0 and 40mg in Week 2, this did not provide suitable estimates for either the cost or the effectiveness of standard care. As the placebo arm in the CLASSIC I[58] trial received no adalimumab, the economic submission used this data to predict health states in the standard care arm.

The models in the industry submission were based on both the 56 weeks of the CHARM[62] trial and an extrapolation to give a lifetime model. The 56 week model included no discounting, the longer model using a discount rate of 3.5% for both costs and benefits. The lifetime model assumed that health remains constant across the group (in terms of the profile of health states) from Week 56 to death. The lifetime model assumed a baseline age of 37 (in line with the average age for CHARM[62]), with life expectancy of 66 years. There was no mortality between years 37 and 66 due to treatment, CD or from other causes.

The model structure was based around four health states defined as remission (CDAI < 150), moderate (150 ≤ CDAI < 300), severe (300 ≤ CDAI < 450) and very severe (450 ≤ CDAI). Patients enrolled in the CHARM[62] trial had baseline CDAI scores between 220 and 450, and fell in only the moderate and severe categories at baseline. Utility data was based on these health states. Costs were calculated based on both trial arms (for anti-TNF medication costs) and on the time spent in each of these disease states (for hospitalisation costs and all other costs). Overall costs and QALY benefits for

the CHARM trial[62] were calculated for both the baseline moderate and severe groups (150 ≤ CDAI < 450) and the baseline severe subgroup (300 ≤ CDAI < 450).

Univariate sensitivity analyses were conducted that modified the method of imputing states for those leaving the trial, that incorporated indirect costs and made several other changes to the cost assumptions. Using details in the revised industry model for adalimumab versus standard care, we were able to replicate the multivariate sensitivity analysis provided in the revised model. This analysis related only to the costs of hospitalisation and the cost and utility values associated the four disease states. There does not appear to be any allowance for uncertainty in the clinical effectiveness of adalimumab.

Estimates of standard care outcomes

The relationship between CDAI-based health states and prognostic factors was estimated using the CLASSIC I trial[58] using an ordered probit regression that predicted the chances of an individual falling into each of the four states (remission, moderate, severe and very severe). Variables were included for baseline CDAI, previous anti-TNF exposure, corticosteroid use, fistulising disease, and included time and treatment dummy variable. Health states for the first four weeks in the standard care arm were found by applying this regression to the clinical factors observed in the CHARM[62] 40mg eow arm. It was assumed that the proportion of people in each health state would remain constant from Week 4 onwards. Although patients previously receiving such treatment were excluded from the CLASSIC I[58] trial, the previous use of anti-TNF treatment appeared as a predictor in the ordered probit regression. It is unclear how this effect was estimated.

Estimates of adalimumab maintenance outcomes

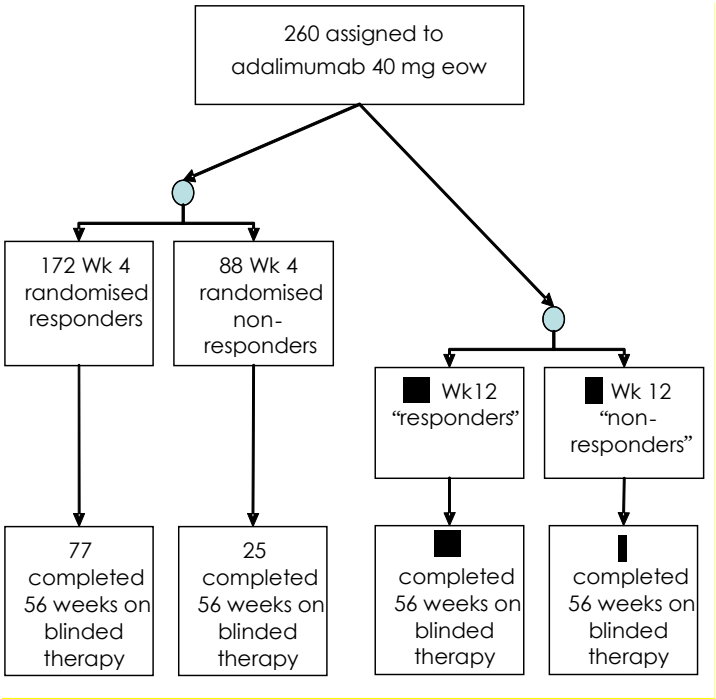
The adalimumab outcomes were estimated using data derived from CHARM trial[62] data. Within the CHARM trial[62], 778 patients were randomised but 854 patients were enrolled at Week 0 in order to achieve this sample. The 76 patients ([redacted] with CDAI ≥ 300) who withdrew prior to Week 4 did so for a variety of reasons, including adverse events (45), lack of efficacy (13), and in one case, death. (This death was judged not to be related to the use of adalimumab by the CHARM trial investigators[62].)

The revised industry model incorporated the costs of these non-randomised individuals by including them within the adalimumab modelling arm. Since the 260 individuals receiving adalimumab comprised approximately one third of the randomised CHARM, cohort it was assumed that one third of the non-randomised individuals would have been randomised to this arm.

The individuals who were not expected to receive a standard adalimumab course were modelled as if they were standard care patients but had incurred an additional £974 each in medication costs. The adverse events specific to these individuals due to adalimumab exposure were not modelled. Whilst this may bias results in favour of adalimumab against standard care, the magnitude of this bias is likely to be relatively small.

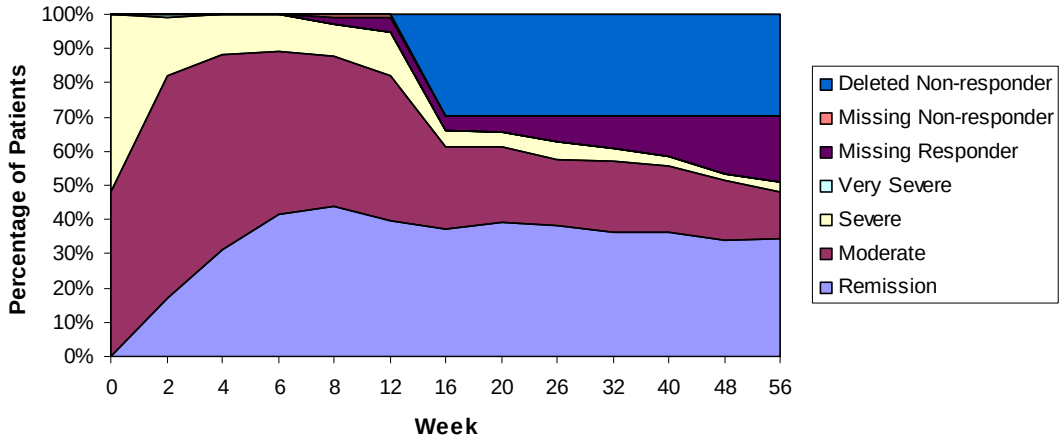
Those expected to be able to receive a standard adalimumab course (i.e. those randomised at four weeks) were modelled as the 40mg EOW arm from the CHARM tria[62]l, and was based on randomisation at four weeks. CHARM randomisation[62] was stratified by 4-week response (reduction in CDAI of 70 points from baseline). At 12 weeks, those not responding (reduction in CDAI of 70 points from baseline) could be shifted to open-label treatment and leave the randomised study. Figure 35 below shows the comparison between the randomised data at 4 weeks and the observational groups defined at 12-weeks. Of the 260 individuals, only [redacted] received their scheduled treatment at Week 56.

Figure 35. Clinical data: CHARM[62] evidence versus that used in the economic model.



Those patients removed from the trial at 12 weeks are referred to in the industry submission as "deleted non-responders" in the Figure 36 (reproduced as non-commercial in confidence) from the economic submission. Missing individuals were those who discontinued from the trial for other reasons, including disease flares and protocol violations.

Figure 36. Adalimumab outcomes from CHARM[62] trial



The economic model included a mixture of those responding and not responding at four weeks. The results in the economic model therefore differ from the co-primary clinical endpoints of the trial, which concerned Week 4 randomised responders only.

In the main analysis, the economic model used the last value carried forward (LVCF) to impute missing values. As a sensitivity test, the model included results where the course of the patients' disease reverted back to the state the patient was in prior to adalimumab therapy ("Simulated Placebo").

Health state cost and utility estimates

Each health state was linked to an expected number of hospitalisations using a Poisson regression model based on a variety of clinical and background characteristics. This was used to construct patient-level predictions of hospitalisation events per year. The unit cost was estimated using published UK data (Bassi et al, 2004[34]) and inflated using PSSRU figures to produce a cost per hospitalisation of £7,441 in 2006 pounds sterling. The CHARM trial[62] did not record CD-related surgeries, and so this hospitalisation factor incorporated the cost of surgery. (Note that the submission was inconsistent whether hospitalisation figure apply per year or over the 56 weeks of CHARM[62].)

Other non-hospitalisation disease costs (excluding anti-TNF medication) were estimated using Bassi et al (2004) [34]. Bassi et al used a seven state classification for CD states similar to Silverstein et al (1999)[23]. The model assumed that "very severe" corresponded to "indicated for surgery" in Bassi et al[34], with "severe", "moderate" and "remission" corresponding to "severe, drug-refractory", "mild disease" and "remitted" states. Estimated non-hospitalisation disease costs are given in

Table 45.

Table 45. Health state-based parameters in the industry submission

Health state	CDAI score	Non-hospitalisation disease costs[a]	Utility
Remission	< 150	8.45	0.859
Moderate	150 ≤ CDAI < 300	23.66	0.795
Severe	300 ≤ CDAI < 450	43.11	0.693
Very severe	450 ≤ CDAI	78.55	0.433

[a] UK£, 2006 prices

Table 45 also displays utility estimates for the four health states that were based on a reanalysis of previously published primary standard gamble data (Gregor et al, 1997[77]). These estimates were based on 180 consecutive Canadian patients presenting with CD between December 1995 and December 1996.

Adalimumab cost estimates

The cost per 40mg adalimumab dose was assumed to be £357.50, with one dose necessary every two weeks per patient after an initial three-dose induction in the first four weeks. No administration costs were included.

6.3.1.2 Results of the adalimumab industry submission

The results reported here refer to the adalimumab industry submission produced in response to comments in the draft assessment report. The industry submission 56-week model suggested that for baseline moderate and severe patients (150 ≤ CDAI < 450), the incremental costs of adalimumab 40mg eow treatment was £2,496 for an incremental increase of 0.0823 QALYs (see Table 46). The estimated incremental cost-effectiveness ratio was £30,319. For patients with severe disease (300 ≤ CDAI < 450), the incremental costs and benefits were estimated at £1,254 and 0.1045 giving an ICER of £10,959 per QALY.

Table 46. Results from the industry 56 week model

	Moderate and Severe			**Severe Only**		
	Adalimumab	**Standard care**	**Difference**	**Adalimumab**	**Standard care**	**Difference**
QALYs	0.8566	0.7743	0.0823	0.8384	0.7339	0.1045
Drug costs	£6,533	£0		£7,119	£0	
Health state related costs	£1,249	£2,049		£1,427	£2,407	
Hospitalisation	£2,028	£5,265		£2,598	£7,485	
NHS costs	£9.810	£7,315	£2,496	£11,146	£9,892	£1,254
ICER			**£30,319**			**£11,998**

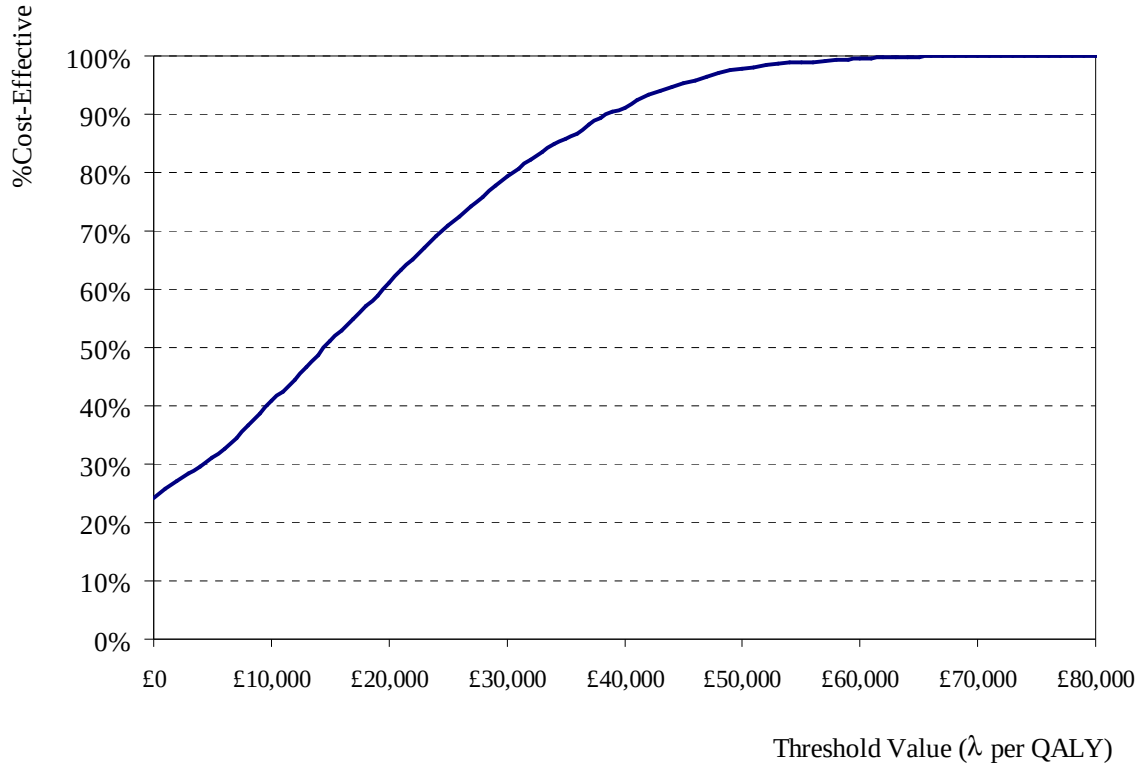
In the original submission, the univariate sensitivity analysis suggested that for patients with severe CD, adalimumab was close to or below £20,000 per QALY for a variety of different assumptions. When considering both moderate and severe CD together, the baseline assumption was close to £30,000 per QALY and typically exceeded this whenever any adverse change was made to the model assumptions. Whilst the industry submission included an induction regimen at 160mg and 80mg at Weeks 0 and 2, it should be noted that this was not used in the CHARM trial[62] and the results will differ if it is associated with higher adverse events. It is not expected that this would change significantly in the second model. The second industry submission model assumed that parameters for hospital and health-state related costs were distributed according to gamma distributions. Utilities for the remission state were assumed to be distributed according to a beta distribution, with constant ratios between all four utility values (see Table 47)

Table 47. Parameter distributions for distributions

Type		Type	A	B
Hospitalisation costs		Gamma	6.25	1,190.56
Health state related costs	Remission	Gamma	6.25	1.35
	Moderate	Gamma	6.25	3.79
	Severe	Gamma	6.25	6.90
	Very severe	Gamma	6.25	12.57
Utilities	Remission	Beta	3.5280	0.5791

In the second industry model (based on 1000 samples), adalimumab treatment has an estimated cost-effective below £20,000 per QALY in 60% of samples, and below £30,000 per QALY in 80.7% of samples. The diagram below is based on 5000 samples and gives similar results (61%, 79%).

Figure 37. CEAC for severe disease (last value carried forward method)



6.3.2 Evaluating the industry submission for adalimumab maintenance versus standard care

The inputs to the industry model of adalimumab maintenance were modified to investigate the robustness of the model. The revised model used the "Simulated Placebo" method of imputing missing values. Those leaving the CHARM[62] trial did so because of disease flare or other issues requiring protocol-violating treatments, and so their health may have been poorer than an "equivalent" simulated standard care outcome (which represented expected health at four weeks). The "simulated placebo" assumption is more neutral with respect to the prognosis of those leaving blinded CHARM[62] treatment than the LVCF used in the industry model.

Aside from a preference for "the simulated placebo", the major differences between the second industry model and the revised industry model are the assumptions made regarding the use of adalimumab beyond the study period. In both the first and second industry models it is assumed that all those receiving adalimumab at 56 weeks will continue to do so for their entire lives. In the revised model the rate of withdrawal from adalimumab maintenance post 56 weeks was increased from zero to that of the CHARM 40mg eow arm. Outcomes for patients with moderate CD were also inferred as a separate subgroup where this was possible. Unless otherwise indicated, the model description here is kept as in the industry submission.

6.3.2.1 The industry model's use of CHARM[62] data

The clinical endpoints of the CHARM[62] trial related to Week 4 responders (a reduction in CDAI of 70 points from baseline) and all published data referred to this group. This causes difficulties in interpreting the data, since terms are duplicated with few caveats. Where published data and the industry clinical submission referred to responders and non-responders, they did so based on a comparison of baseline and four week data (randomisation); the economic submission appeared to define this split using baseline versus Week 12 data.

The CHARM[62] trial randomised patients at four weeks to one of three (blinded) arms – placebo care, adalimumab 40mg eow and adalimumab 40mg weekly. This blinded treatment stage in the trial was maintained for 12 weeks for all randomised patients. Those who did not achieve a sufficient improvement in health at 12 weeks were termed "non-responders" in the economic model, which appeared to define this as a reduction of less than 70 points in CDAI from baseline. It appears that Week 12 non-responders were moved to open-label 40mg eow treatment, as were Week 12 responders who experienced disease flares after 12 weeks. Those receiving open-label eow treatment could subsequently move to weekly treatment as required, and then subsequently to standard care following persistent non-response.

Outcomes beyond 56 weeks

The initial industry submission did not adequately describe or justify the assumptions used in constructing its economic model. This was particularly problematic when considering lifetime costs and effects, since this extrapolated data at the end of CHARM trial[62] for an additional 37 years. The industry models suggested an average expected adalimumab 40mg eow use of 13.3 vials per year after Year 1, which was consistent with the numbers receiving eow treatment at 56 Weeks. However, with an approximately constant number of people leaving the trial's adalimumab arm from Week 7 onwards within CHARM, it could be predicted when the last individual would cease to receive adalimumab on this until-flare maintenance regimen. With the limited data made available from the economic model, it was predicted that the last dose of adalimumab corresponding to the blinded (and costed) treatment on CHARM[62] would have occurred in Week 189. A lifetime model was not necessary here as - under the assumptions of the placebo method of imputing lost values – the standard care and adalimumab model arms would have been identical after four years and so a 4-year timeframe would have sufficed.

In Year 1, predicted adalimumab usage was 19.13 vials when averaging across both Week 12 responders and non-responders. Note that doses in Weeks 52 and 54 were counted in Year 2. In Year 2, (Weeks 52-103) a total of 11.20 vials would have been used per randomised adalimumab 40mg eow patient. This is lower than 26.0 vials that would have been used if all patients had responded, and lower than the predicted usage in the economic model, as nearly 30% of responders were expected to cease using adalimumab in the second year. In Years 3 and 4, 6.00 and 1.11 vials were used on average.

In light of the lack of transparency in the industry model, this analysis was based on a 'best-guess' interpretation of the industry model that was consistent with the limited data presented. In particular, it appeared that the economic model was based on data considering only the blinded portion of the CHARM trial[62]. The revised model estimated resource use for Week 12 responders and non-responders up to Week 12, and non-responders after Week 12 (until missing or Week 56) for blinded eow treatment only. This figure, based on a constant loss of patients from blinded treatment from Week 6 (zero loss) to Week 56 ([redacted] patients lost) in the trial was similar to the estimated resource use in the industry model.

Other changes to model structure and interpretation.

Hospitalisation costs and disease state costs for Year 2 onward were taken from the industry model. Hospitalisation costs for Year 1 were as for the 56-week industry model, with four weeks of

"lifetime" costs removed (since this occurred in Year 2). These costs were weighted by the proportion of individuals expected to use adalimumab relative to those receiving standard care.

At randomisation (Week 4), the CHARM 40mg eow arm[62] included 125 patients with moderate CD (CDAI between 150 and 300) and 135 with severe CD (CDAI between 300 and 450). The industry submission provided the expected frequency of health states, adalimumab use and hospitalisations for both moderate and severe, and severe only groups. This allowed separate outcomes to be inferred for those with moderate disease within the 56 weeks of the CHARM model.

6.3.2.2 Results of revised adalimumab model

56 week results for severe and moderate subgroups

The industry submission predicted an incremental cost-effectiveness of £11,998 per QALY at 56 weeks for those with a baseline CDAI at or above 300. With the preferred "placebo method" of imputing missing data, this rises to £30,964 per QALY. The industry submission did not predict incremental cost-effectiveness for the moderate subgroup. In the estimates presented here, it was found that treatment was far less cost-effective than for the severe subgroup, and above £100,000 per QALY using the "placebo method" of imputing missing data.

In the 56 week model it appears that treatment for severe patients is likely to approach £30,000 per QALY under more conservative assumptions but will fall below £20,000 per QALY under optimistic assumptions. For moderate patients, even optimistic assumptions appear to give relatively large ICERs for adalimumab treatment.

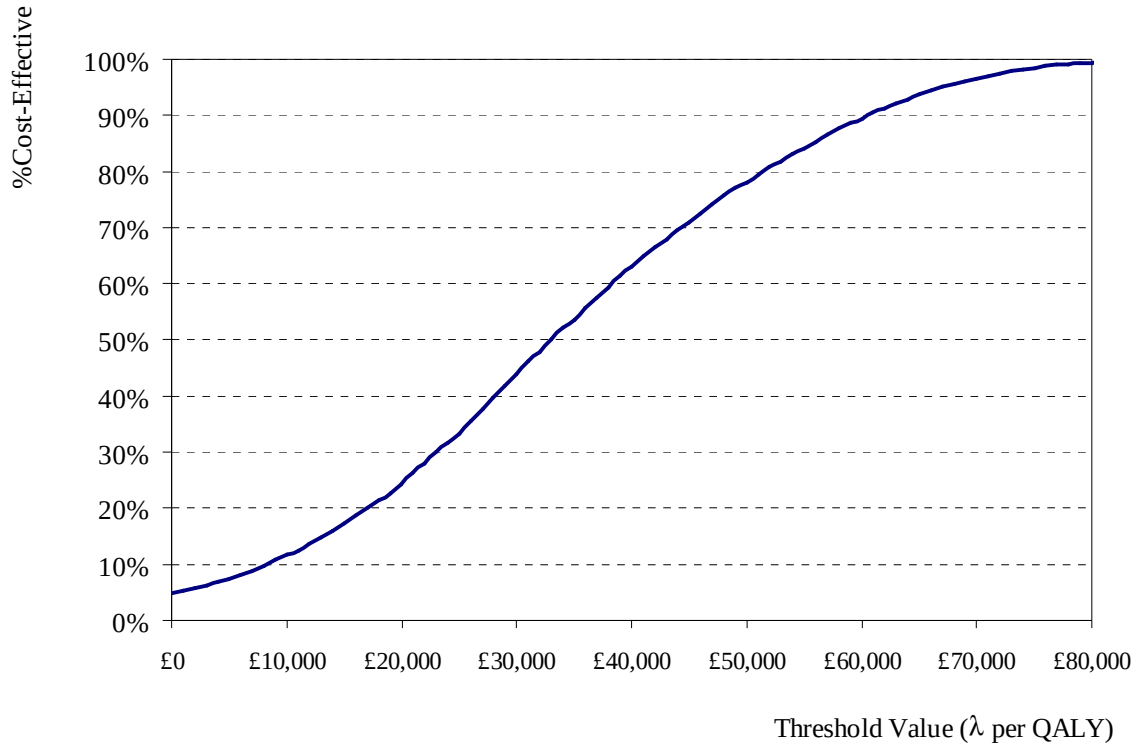
The numbers in Table 48 suggest that treatment of those with severe disease ($300 \leq CDAI < 450$) will be cost-effective under optimistic (LVCF) assumptions and marginally over £30,000 per QALY with the preferred and more conservative assumptions. There is less ambiguity surrounding the treatment of those with moderate disease ($150 \leq CDAI < 300$). Even under optimistic assumptions the smaller additional health benefit 0.0589 comes at a higher incremental cost, leading to an ICER above £60,000 per QALY.

Table 48. Cost-effectiveness of adalimumab in second industry models: severe and moderate sub-groups. and imputation method

	Severe Patients only			**Moderate Patients only**		
Values imputed using last value carried forwards – optimistic estimate						
	Adalimumab	**Standard care**	**Difference**	**Adalimumab**	**Standard care**	**Difference**
QALYs	0.8384	0.7339	0.1045	0.8769	0.8180	0.0589
Drug costs	£7,119	£0		£6,029	£0	
Health state related costs	£1,427	£2,407		£1,046	£1,663	
Hospitalisation	£2,598	£7,485		£1,465	£2,868	
NHS costs	£11,146	£9,892	£1,254	£8,540	£4,531	£4,009
ICER			**£11,998**			**£68,065**
Values imputed using simulated placebo – pessimistic estimate						
	Adalimumab	**Standard care**	**Difference**	**Adalimumab**	**Standard care**	**Difference**
QALYs	0.8225	0.7339	0.0886	0.8605	0.8180	0.0425
Drug	£7,119	£0		£6,029	£0	
Health state related costs	£1,565	£2,407		£1,205	£1,663	
Hospitalisation	£3,592	£7,485		£2,099	£2,868	
NHS costs	£12,636	£9,892	£1,254	£9,333	£4,531	£4,802
ICER			**£30,964**			**£113,008**

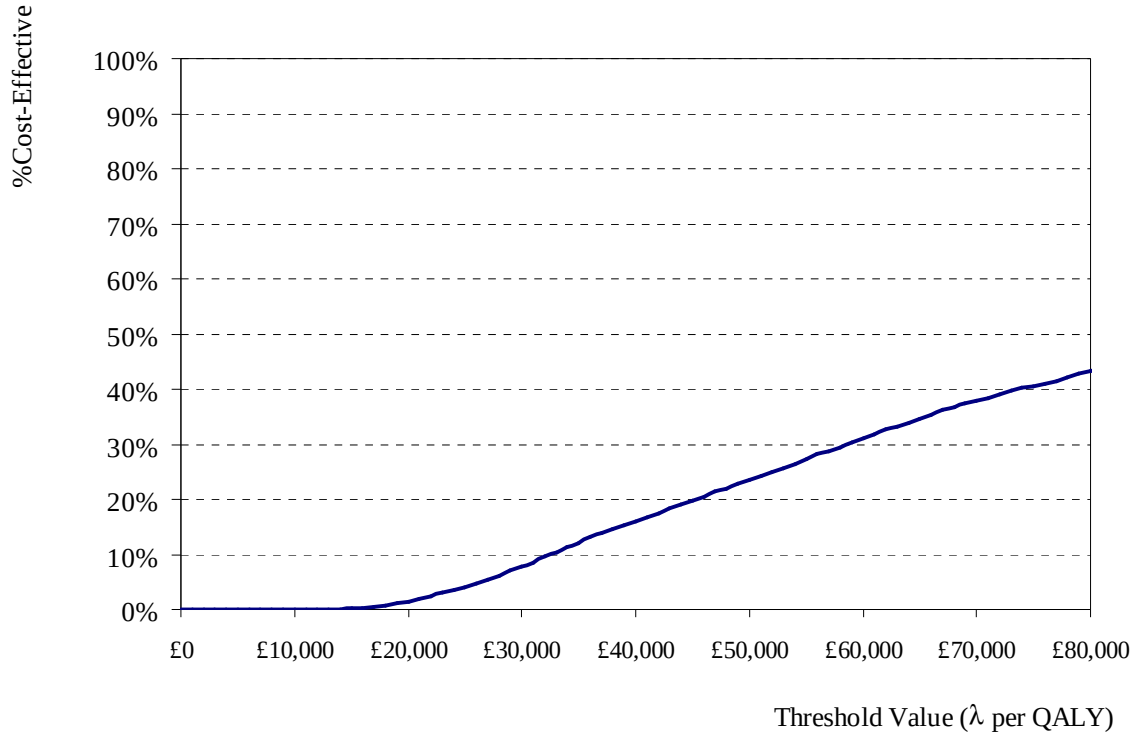
Figure 38 shows the CEAC for the treatment of severe disease where the (conservative) simulated placebo method is instead. Across 5000 samples, 24% fall of ICERs below £20,000 per QALY and 44% fall below £30,000 per QALY. These figures compare to 61% and 79% using LVCF.

Figure 38. CEAC for severe disease (placebo method)



For the moderate case, Figure 7 shows the CEAC for the optimistic case (LVCF). Here, only 1.5% of samples fall below £20,000 per QALY and only 7.8% fall below £30,000 per QALY. Under the pessimistic assumption (simulated placebo) these figures fall to 0.002% and 2%, respectively.

Figure 39. CEAC for moderate disease (last value carried forward method)



The effect of an increased "lifetime" model

Given the marginal cost-effectiveness of treating those with severe disease with adalimumab and the very large cost-effectiveness ratio for moderate disease, the issue of time horizon may prove critical.

Outcomes beyond 56 weeks were estimated by using calculated predictions of how many patients continued to receive treatment at any point. The industry economic submission gave costs for the adalimumab arm where the rate of response was fixed at the 56 week level.

Using the placebo method, the 56 week model for CHARM[62] suggested an ICER of £56,621 per QALY for both moderate and severe disease. The analysis suggested that continuing usage would improve overall cost-effectiveness but not do so drastically. The ICER in the first year was estimated to be £57,739 per QALY and this decreased to £43,933 per QALY in subsequent years. Over the four years in which the one-shot maintenance therapy is expected to affect outcomes, the overall cost-effectiveness of treatment was estimated to be £52,713. As such, it is not believed that extending the timeframe of the economic model for either the moderate or severe group would drastically improve the cost-effectiveness of treatment.

Table 49. Cost-effectiveness of adalimumab revised model: 56 weeks and 4 years

MODERATE AND SEVERE CROHN'S	**Lifetime model (industry)**			**4 year revised model (revised)**		
	Adalimumab	**Standard care**	**Difference**	**Adalimumab**	**Standard care**	**Difference**
QALYs	14.579	13.474	1.104	2.845	2.733	0.108
Drug costs	£90,919	£0		£13,149	£0	
Health state related costs	£24,559	£35,676		£5,778	£7,235	
Hospitalisation	£39,305	£92,162		£12,644	£18,667	
NHS costs	£154,783	£127,838	£26,945	£31,572	£25,902	£5,669
ICER			£24,385			**£52,713**

The 4 year revised model hinges on two assumptions: that the treatment referred to 40mg eow treatment, and that the patient loss continued at a constant rate. Comments provided alongside the second 56-week industry model suggested that both of these assumptions are inaccurate. In particular, the Abbott model includes observed drug usage including usage following dose-escalation to 40mg weekly. The total usage figures did not increase to the levels that would otherwise be expected because of a significant number of patients within CHARM[62] who did not receive their indicated treatment.

The industry figures for observed drug use did not allow a full breakdown of who received what treatment. They indicated that on a period-by-period basis (e.g. Weeks 12 to 16, Weeks 16 to 20 etc), those on a 40mg weekly dosage receive only [redacted] of the total dosage received by those on a 40mg eow dosage, rather than the 200% that would be expected. Assuming that all patients who received treatment had the indicated dose, then on a period by period based after Week 12 it appears that only [redacted] of patients prescribed 40mg eow, and [redacted] of patients prescribed 40mg weekly went on to receive it. This suggested that many individuals appeared to miss scheduled treatment on 40mg eow, and that those shifted on to 40mg weekly were unlikely to have received it. This may indicate issues in the tolerability of adalimumab.

The timing assumption used in the model was also conservative, given that those individuals who tolerate adalimumab well were more likely to continue using it than those who did not. Data from the company suggested that the rate of loss may slow but it is difficult to confirm this given that the data appeared to take a different baseline than that used in the economic analysis as it combined both

adalimumab arms from CHARM (remission rates for week 4 responders: [redacted]/517 at 52 weeks and [redacted]/517 at 18 months, with a smaller fall between 18 and 24 months). This suggested a continuing decline in adalimumab usage at a decreasing rate. Unfortunately, neither this nor a hypothesised relationship between timing and continued use (based on 56 week data only) was incorporated into the second company model and so it was not possible to report a more accurate "middle" case. Overall, both the lifetime industry model (reported below) and our modifications appear to be biased in opposite directions. It is reasonable to surmise that the underlying cost-effectiveness of adalimumab lies somewhere between the industry and revised model. On the basis of the hypothesised relationship between time and adalimumab usage the industry model overestimates usage by more than the revised model underestimates it.

6.3.2.3 Discussion of adalimumab industry submission

Neither the analysis of the 56-week CHARM trial[62]data nor the lifetime adalimumab economic model was based on the modified intention-to-treat analysis on which the major clinical findings of the published CHARM trial[62] were based. Both the published clinical data and the confidential clinical study report submitted to NICE were based on a division between responders (defined as those who had a reduction in CDAI of at least 70 points from baseline) at Week 4. In contrast, the economic model was based on response/non-response (again at 70 points from baseline) at Week 12. As such, the model is not immediately compatible with the main clinical findings concerning the proportion of Week 4 responders retaining response at 26 and 56 weeks.

In reviewing the economic evidence, there are concerns over the comparators used in the adalimumab model. Given the structure of the CHARM trial[62], standard care could have been compared with an induction only dose of adalimumab, an until-flare maintenance regimen (based on blinded treatment in CHARM[62]) and a lifetime regimen (based on blinded and open treatment in CHARM[62]). As the results for an induction only regime appear as the "standard care" arm of the CHARM trial[62], it should have been included in the economic submission.

Previous use of other anti-TNF therapy is an important predictor of response to adalimumab that is not addressed within the industry submission. In the CHARM[62] trial, patients with previous experience of anti-TNF therapy were only excluded if they had no clinical response to the therapy, or had used it in the past 12 weeks. Fifty percent of all CHARM[62] patients had had prior exposure to anti-TNF therapy. As the clinical response was superior in those who had not previously received anti-TNF treatment, it is highly likely that cost-effectiveness would be superior in this group. Given that the 56-week revised model suggested a relatively high cost-effectiveness ratio even for the severe subgroup this may be an important consideration.

A lack of clarity over the source and interpretation of data has hampered the analysis of the economic submission. Overall, the economic model met most of the requirements of the NICE reference case but crucial elements of the model could not be verified. The analysis here has attempted to address concerns over the methodology and interpretation of the economic model. It appears that the cost-effectiveness ratio for moderate CD patients is particularly high at 56 weeks and the analysis further suggests that this figure will not necessarily fall appreciably at a longer timeframe and particularly in light of the size of fall necessary to approach cost-effectiveness for moderate disease. The cost-effectiveness of adalimumab treatment is far more favourable for patients with severe CD.

Table 50. Compatibility of the model with the NICE reference case

Element of health technology assessment	**Principles**	
Defining the decision problem	The scope developed by the Institute	Yes.
Comparator	Alternative therapies routinely used in the NHS	Partial. Not all relevant comparators are used for the adalimumab.
Perspective on costs	NHS and PSS	NHS Only.
Perspective on outcomes	All health effects on individuals	Yes.
Type of economic evaluation	Cost-effectiveness analysis	Yes.
Time horizon.	Sufficient to reflect any differences in costs or outcomes between the technologies compared.	Yes. (Lifetime model)
Synthesis of evidence on outcomes	Based on a systematic review	Partial. Details unclear and not necessarily reproducible.
Measure of health benefits	Quality-adjusted life years (QALYs)	Yes.
Description of health states for calculation of QALYs	Health states described using a standardised and validated generic instrument	Yes.
Method of preference elicitation for health state valuation	Choice-based method, for example, time trade-off, standard gamble (not rating scale)	Yes.
Source of preference data	Representative sample of the public	No. Patient values used.
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes.
Equity position	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes.
Modelling methods	Structural assumptions and data inputs clearly documented and justified.	No.
	Probabilistic sensitivity analysis should be conducted.	Partial. No clinical undertainty is considered.

6.3.3 Industry submission comparison of adalimumab and infliximab maintenance therapies

The Abbott submission[5] also included a cost model comparing adalimumab and infliximab maintenance regimens. The stated aim of the maintenance comparison was to compare adalimumab against infliximab on the basis that infliximab is the alternative most likely to be displaced by the prescription of adalimumab. However, neither adalimumab maintenance nor infliximab maintenance would be the most appropriate comparators in such an analysis. Due to a lack of trial results comparing these treatments directly, the comparison is secondary in nature.

The infliximab comparator used appears to combine those who were judged to be responders and non-responders on the 5mg/kg arm of ACCENT I[4,3] using the Rutgeerts et al data including both 5mg/kg standard dosage and 10mg/kg "as needed" dosage. The adalimumab comparator used adalimumab maintenance at 80mg/40mg induction with 40mg dosage every other week. The adalimumab outcomes were found using a weighted sample from the CHARM[62] trial for those with CDAI between 220 and 400 (in line with ACCENT I[3,4]) and with weights derived so that the same gender distribution, median age and CDAI quartile figures (LQ, median, UQ) held across the infliximab and modified adalimumab groups. Those with CDAI above 400 were excluded from the analysis for comparability with the ACCENT I[3,4] trial. Those who had previously used anti-TNF treatments were not excluded from the adalimumab group, although this was an exclusion criterion in ACCENT I[3,4]. Missing data for both comparators was inferred using LVCF.

The adalimumab arm costs were found by assuming that all those receiving adalimumab and not responding at Week 12 would have continued to receive it, which lead to higher costs than in the main model. The infliximab arm drug costs were assumed to include an average wastage of 0.5 vials per infusion. Hospitalisation costs were estimated using the Poisson regression for the adalimumab group and observed hospitalisations (plus inferred data from the adalimumab group) in the infliximab group. The model used an excess hospitalisation of 0.098 per infliximab patient per 56 weeks.

As the infliximab data used remission/non-remission rather than the four health states of the main adalimumab model, the health-status based costs were estimated using the frequencies reported in Bassi et al[34]. A cost of £38.48 per week was calculated for non-remission costs, and the cost of remitted patients per week was assumed to be £8.45. Overall, the model suggested an excess cost for infliximab patients of £4,414 over 56 weeks, of which the majority was due to medication costs (£3,526).

Whilst this model also attempted to compare health outcomes, no summary quantitative figures were provided on which to base a cost-consequences analysis, cost-effectiveness or cost-utility analysis. Using the proportion of patients in remission (partially inferred using LVCF assumptions), it was claimed that adalimumab lead to a higher proportion of patients in remission from Week 6 onwards. In conjunction with the cost findings, the model claimed that adalimumab maintenance dominated infliximab maintenance.

This model is reported but not analysed it in depth. The model compared adalimumab maintenance against infliximab maintenance without comparing either against a standard treatment. As a comparison of non-standard treatments, it fell outside the scope of the assessment. Furthermore, since both adalimumab maintenance and infliximab maintenance appeared to have incremental cost-effectiveness ratios far outside the suggested ranges, the results of this model are of little practical relevance to the decision problem faced.

It was also noted that the infliximab regimen modelled here included the 10mg/kg dosage only. Given the uncertainty relating to which treatments were actually received by patients in ACCENT I[3,4], on what basis these treatments were received, and to what degree the treatments received would be legitimate in a NHS context, it would be difficult to place any confidence in this model.

6.4 Independent economic assessment

6.4.1 Introduction

The overall decision problem for this appraisal is 'What is the cost effectiveness of anti-TNFs in the management of moderate to severe CD in the UK NHS?' In order to undertake cost effectiveness analyses to address this decision problem it was necessary to (a) define moderate and severe CD; and (b) specify the specific roles for anti-TNFs in the management of CD that are to be evaluated; and (c) specify the patient groups for whom cost effectiveness will be assessed..

Disease severity can expressed in terms of current status or life course. It can be measured using a wide range of indices including frequency of symptoms, severity of symptoms, biochemical activity levels and intensity of treatment.

Available evidence does not provide a strong basis for differentiating CD severity in terms of life course. Munkholm et al[22] reported that 'The clinical course of CD differs markedly over time, with ever-relapsing cases, to a quiescent course with remission for several years, interrupted by years with relapse. No predictive factors have been found for the subsequent course with regard to age, sex, extent of disease at diagnosis and treatment in the year of diagnosis'.

The current severity of CD is difficult to assess, and a global measure encompassing clinical, endoscopic, biochemical and pathological features is not available.[36] The most widely used disease activity measures include the CD Activity Index (CDAI), the Harvey-Bradshaw Index (HBI) or Simple Index, a simplified version of the CDAI, and the Perianal Disease Activity Index (PDAI). A commonly used health related quality of life measure is the Inflammatory Bowel Disease questionnaire (IBDQ). Other measures include the CD Endoscopic Index of Severity (CDEIS).

The CDAI is the measure used in the anti-TNF clinical trials. It measures current disease severity using a recall period of the last 7 days. Variables captured in the measure include number of liquid stools, abdominal pain, general well being, extraintestinal complications, use of antidiarrhoeal drugs, abdominal mass, haematocrit and body weight. Scores range from 0 to approximately 600. Values of below 150 are suggestive of quiescent disease (remission) and values above 450 are associated with very severe disease.[37] Some investigators have arbitrarily labelled CDAI scores of 150-219 as mildly active disease and scores of 220 to 450 as moderately active disease.[36]

Given that the anti-TNF trials use CDAI to measure disease severity, the cost effectiveness analysis uses the following definitions of disease severity:

- Severe Disease CDAI > 300
- Moderate Disease 220<CDAI<300.

It should be noted that in line with the decision problem and the use of CDAI in the trials, this definition says nothing about the frequency of relapse. A patient who has been in remission for five years and relapses with moderate disease refractory to standard therapy is as much a target for treatment with anti-TNFs as a patient who has had two relapses in the last 12 months, with moderate disease that is refractory to standard therapy.

The scope for this appraisal identified three categories of use of anti-TNFs in the management of CD: Induction, Episodic and Maintenance. There is some uncertainty as to the precise definition of each of these categories.

Maintenance therapy is perhaps the most straightforward to define. It can be described as the chronic use of anti-TNF therapy to maintain remission in patients who have responded to anti-TNF therapy when in relapse. In maintenance therapy, the key challenges in arriving at a working definition are
(a) what is the criteria for defining a patient as a responder? Is it the achievement of remission or a specific improvement in their CDAI score? and
(b) how many doses of anti-TNF therapy can an individual receive before being a confirmed non-responder?

Defining episodic treatment is less straightforward (see glossary). In the literature and submissions to this appraisal we have identified seven different working definitions of episodic treatment. In the previous appraisal of anti-TNFs in the management of CD, episodic treatment was defined as giving up to three additional courses of treatment when a patient experienced a disease relapse if that patient initially responded to anti-TNF therapy. The relapse could have occurred once in several years or much more frequently. As with maintenance therapy, the key uncertainties in this definition are
(a) what is the criteria for defining a patient as a responder? Is it the achievement of remission or a specific improvement in their CDAI score? and

(b) how many doses of anti-TNF therapy can an individual receive before being a confirmed non-responder?

Induction treatment is the use of anti-TNF therapy with the aim to achieve remission. It is not straightforward to draw a distinction between repeated use of anti-TNF as induction and episodic therapy as described above. Induction therapy may merely be the initial application of anti-TNF to a patient in relapse which establishes their responder status prior to the subsequent provision of episodic or maintenance therapy. To consider the cost effectiveness of induction therapy divorced from its value in informing future decisions on episodic or maintenance therapy would be clinically unrealistic and would produce an inaccurate estimate of its cost effectiveness. As with episodic and maintenance treatment, there is a question regarding how many doses of anti-TNF an individual can receive before being confirmed as a non-responder.

Given the problems with assessing the cost effectiveness of induction therapy in isolation, we have not modelled induction therapy alone. Instead we examined the cost effectiveness of anti-TNF therapy in episodic and maintenance therapy, where episodic was defined as the patient having the opportunity to undergo a second course of treatment within the time frame of the model if they initially responded to the induction treatment but then subsequently relapsed. That is to say, all patients received induction therapy but only those who were responders were eligible to go on to further treatment if they subsequently relapsed.

There are a number of alternatives to defining responder status. Within the trials responders were defined in two distinct ways: (1) patients whose CDAI improved by a pre-specified amount following administration of anti-TNF; and (2) patients who achieved remission following administration of anti-TNF.

For the purpose of economic evaluation the first approach to defining a responder was problematic as it said nothing about the relative improvement in health in an individual for any given reduction in CDAI score. Both the health gain associated with any given improvement in the CDAI and the value attached to the health gain would depend upon the pre-treatment CDAI. Defining responders using a pre-specified improvement in CDAI does not differentiate between patients for whom treatment controls the disease and patients for

whom treatment merely reduces the severity of the symptoms. Thus it is not possible to ascribe a robust utility value for the health of responders defined in this way. By contrast, patients in remission are effectively free of symptoms and this is a health state for which it would be possible to ascribe a robust utility value. For this reason, response was defined as achieving remission following anti-TNF therapy.

There are a number of alternative approaches to determining how many doses, or cycles of treatment, individuals could receive before be confirmed as a non-responder including clinical opinion on actual practice and use as per licensed indication. The clinical advisor to the study team recommended that we allowed patients to have either two or three cycles of treatment before establishing a non-responder status. By contrast, the infliximab licence required patients to respond to the first cycle of treatment for subsequent maintenance or episodic treatment. Within the model, responder status was defined after the second cycle of treatment as this is in line with actual clinical practice.

The scope for this appraisal identified a number of patient groups for whom the committee would be interested in obtaining specific estimates of the cost effectiveness of anti-TNF therapy: Adults; Children; Severe; Moderate; Fistulising and Non-Fistulising.

The placebo randomised controlled trial evidence for the anti-TNFs did not include paediatric trials; even though infliximab does have a licence for use in the paediatric population. In the absence of estimates of effectiveness that can be used to model the magnitude of effect for anti-TNFs compared to standard care, robust modelling of the cost effectiveness of anti-TNF therapy in a paediatric population was not possible. However, to assist the committee in its deliberations; a scenario analysis using the adult models is presented, where paediatric costs have been substituted for the adult costs. This is equivalent to assuming that treatment is equally effective in paediatric populations.

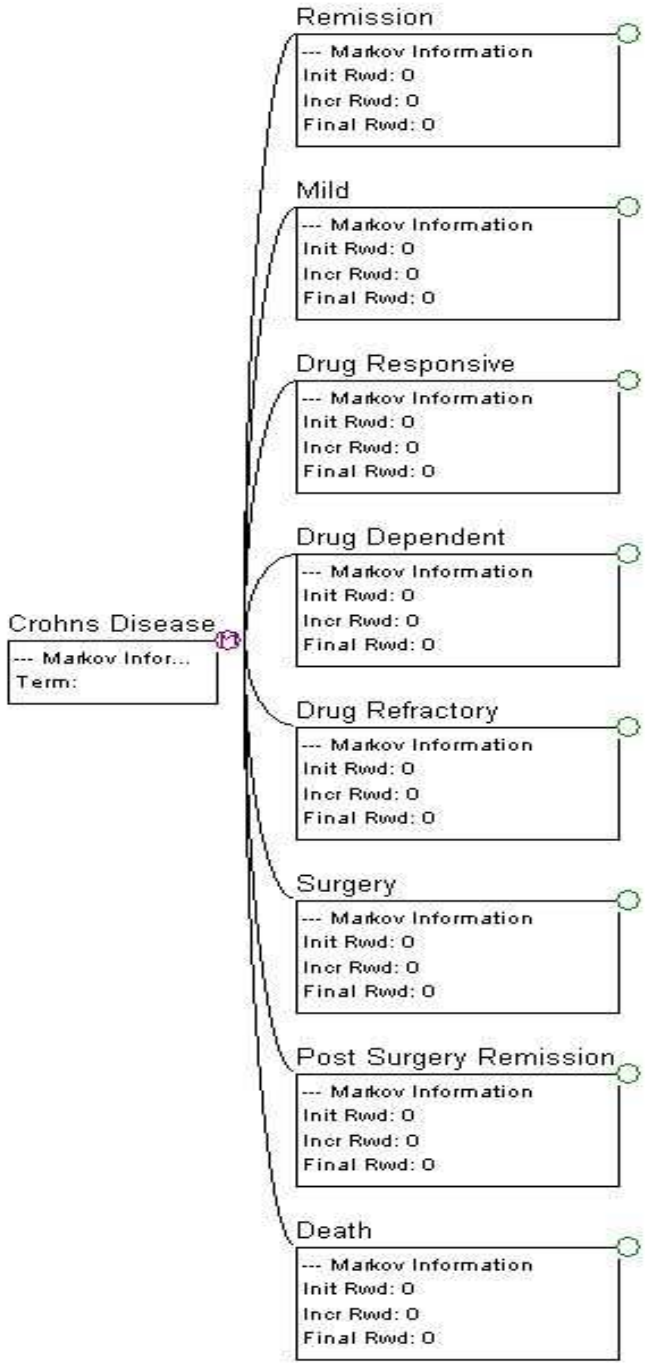
Separate models are presented for patients with moderate and severe disease as the value of the health gain associated with remission will be systematically different for these two patient groups, as are the likely costs of managing relapse.

Whilst the trials of anti-TNFs did differentiate between fistulising and non-fistulising disease it has not been possible to identify a long term usual care cohort study for fistulising patients. In the absence of this evidence, the trial-based evaluations submitted by the manufacturers provided the best estimates of the likely cost effectiveness of treatment in the fistulising population. However, it is important to emphasise the difference in the characteristics of the trial populations and the characteristics of the population that is the focus of the decision problem for this appraisal – i.e. *all* patients with moderate to severe disease which is refractory to standard treatment.

The trial populations were, for good reasons, frequently relapsing patients. Frequency of relapse is not an inclusion criterion for treatment in the decision problem for this appraisal. Patients who relapse more frequently have more capacity to benefit from an effective treatment and therefore, assuming effectiveness is not lower in patients who relapse frequently, treatment would be more cost effective in those patients than in the population defined for this appraisal.

In summary, de novo cost effectiveness analyses for adults with moderate to severe CD are presented, where response is defined as remission after one or two cycles of treatment.

Figure 40. Schematic of Silverstein et al.'s clinical classification[23]



The objective of the cost effectiveness analysis was to estimate the incremental cost per quality adjusted life year (QALY) for each drug compared to standard care in (a) induction therapy for moderate and severe disease; and (b) maintenance therapy for moderate and severe disease.

6.4.2 Methods

The cost effectiveness model is a simple Markov model. It consists of four primary states – remission, relapse, surgery and post surgery remission. Table 51 gives the costs and utilities associated with each state in the standard care arm. The utilities are derived from a widely cited study of health related quality of life in CD. Whilst the study does not meet the reference case specification in the NICE methods guide, in the absence of an alternative study that does meet these criteria, it has the desirable characteristics of providing values deriving a choice-based method (Time-Trade Off), being a well conducted study and providing utility values for differing severities of disease – the type of data that are required for this analysis. The exception to this was the utility weight for major surgery. In the absence of a published estimate it was assumed that the average utility for individuals in the major surgery state would be equivalent to being on level 2 of all dimensions of the EQ-5D – this gives a utility weight of 0.516.

The NICE Methods guide recommends that models of chronic diseases normally adopt a lifetime horizon. The reason for this recommendation being that disease exacerbations in chronic diseases are usually related to reductions in life expectancy. Wolters et al.(2007)[80] analysing a cohort followed from the years 1993/4 to 2003/4 reported that only age at diagnosis was associated with an increased mortality rate and that, whilst the evidence was weak, the Standardised Mortality Ratio for CD approached unity. As none of the trials provided evidence of an impact upon mortality, it is reasonable to assume there is not a differential mortality rate and therefore, a lifetime horizon would not add meaningfully to the precision of the cost effectiveness estimate. Therefore the time horizon for the model is one year and the cycle duration is four weeks; i.e. the model has 13 cycles. Both the inducation and maintenance model start with a cohort of 1000 patients in the relapse state.

Table 51. Parameters in the CD Cost Effectiveness Model

	Cost per cycle severe disease	Cost per cycle moderate disease	Utility per cycle severe disease	Utility per cycle moderate disease
Remission	52	52	0.073*	0.073*
Relapse	1489	474	0.056*	0.068*
Surgery	4592	4592	0.039*	0.039*
Post Surgery Remission	72	72	0.073*	0.073*
Adalimumab Induction *Adalimumab Maintenance*	1072.50 (see Table 67)			
Infliximab Induction *Infliximab Maintenance*	5809.26# (including loading dose) 1094.50 † (excluding loading dose)			

* from Gregor et al[77] , #see text in paragraph below Table 67, † "The cost of the maintenance therapy dose of infliximab was calculated in a similar fashion to that of the induction dose. Four vials of infliximab were required per treatment and each vial of infliximab costs £419.73. The cost of administration is £257.50 per individual treatment. Over the total course of treatment an average of 6.5 maintenance doses will be required over the course of 11.5 months. This gives (((4*£419.73)+257)*6.5)/11.5 for an average cost per dose of £1094.50."

As the model does not include a differential mortality rate, a one year time horizon is appropriate and thus there is no need to discount costs or benefits.

Only direct NHS costs are considered within the model. Costs are taken from the NHS Reference cost database 2005/6.[85] The exception to this being the cost for remission which was taken from the work by Bassi et al.[34] These were indexed using the PSSRU NHS Pay and Prices Index.[86] The Bassi et al cost for quiescent CD is used as the cost for the remission state. The reference costs for in-patient and out-patient major and intermediate interventions for Inflammatory Bowel Disease are used as estimates of the costs of severe, moderate relapses and major surgery.All costs were indexed to 2005/6 using the PSSRU NHS Pay and Prices Index.[86]

To construct the transition matrix from Silverstein et al's published matrix it was assumed that the Drug Responsive and Drug Dependent states are effectively managed with the standard care interventions. Thus, for the purposes of this analysis, the two categories are combined with the remission state into a single Remission on Standard Care state. This might mean that the cost in remission is an underestimate of the actual cost. However, if this is the case, the effect will be to make the interventions appear more cost effective than they actually are; i.e. it is an assumption that operates in favour of anti-TNF treatment.

The Drug Refractory state is assumed to be in relapse on the standard care package. The Surgery and Post Surgery Remission states are included in the model directly. This gives four primary states for the model:

- Remission;
- Relapse;
- Surgery; and
- Post Surgery Remission

It is worth noting that, in common with the Silverstein et al analysis, the matrix includes transitions from post-surgery remission from relapse and remission states. These transitions are most likely an artefact of the maximum likelihood method used to estimate the Silverstein transition matrix. Silverstein et al did not report complication rates from surgery and thus it is not included as a state in our model. As the number of this type of transition is small, it has not been considered here to have substantially weakened the Silverstein et al study as the preferred basis for modelling standard care. Figure 41 and Figure 42 are schematic diagrams of the standard care and anti-TNF pathways in the cost effectiveness model.

6.4.2.1 Modelling the disease course under standard care

The population of interest for this appraisal is an inclusive one, rather than the tightly defined populations often found in clinical trials. Therefore, it was important to identify evidence from a population cohort that identified patients who were resistant to standard therapy. It was also important the cohort reported data from the time before the advent of biologic therapy.

A frequently cited study by Silverstein and colleagues met these criteria. This study reported a two-monthly transition matrix estimated from 20 years follow-up of an inception cohort of 174 patients. Patients were characterised as being in one of eight states: remission, mild, drug responsive, drug dependent, drug refractory, surgery, post surgery remission and death.

Given the focus of the analysis on the cost effectiveness of anti-TNFs in moderate to severe disease and the lack of evidence for a differential mortality rate between standard and anti-TNF treatment, It was important to derive a transition matrix that did not include death or mild disease.

6.4.2.2 Derivation of the standard care transition matrix

To construct the transition matrix from Silverstein et al's published matrix it was assumed that the Drug Responsive and Drug Dependent states were effectively managed with the standard care interventions. Thus, for the purposes of this analysis, the two categories are combined with the remission state into a single Remission on Standard Care state.

A four-state matrix: remission, relapse, surgery and post-surgery remission was derived by the following steps:

Step 1; removing death: It was supposed that death from all states was equally likely. The chance of death in each (t0) state was divided by six and this was added to the six non-mild, non-death states.

Step 2: removing the mild state: A more complex process was used for the mild state. Here, the issue was not now where one entered the state from, but where one would exit to after leaving mild.

2.1) each of the non-death transition probabilities out of mild were deflated by the total chance of leaving the mild state (0.090). These probabilities for the exit state for mild were: remission, 0.636; drug responsive, 0.092; drug dependent 0.068; drug refractory 0.107; surgery 0.065; and post surgical recovery, 0.031.

2.2) These exit probabilities were multiplied by the chance of entering mild from each of the other initial health states (t0) and used to distribute the probabilities. Here, the chance of a person in remission remaining in remission increased by 0.070 (the chance of leaving for mild) x 0.636 (the chance of moving from mild to remission).

2.3) The initial Silverstein transition probabilities were increased by the probabilities in (2.1) and (2.2).

Step 3) The drug responsive state has Markov probabilities summing to 1.00001 due to rounding error in the original paper. These have now been corrected.

Step 4) Steps 1 to 3 produce a matrix in six states. The states remission, drug responsive, and drug dependent were then combined into a single remission state.

4.1) The chance of being in any one of these states was assessed using figures in Silverstein et al.[23] Of the three states, there was an 89.1% chance of being in remission, a 2.1% chance of being in a drug responsive state, and an 8.8% chance of being in a drug dependent state.

4.2) The chance of remaining in the (broader) remission state was calculated as the average of the chance of moving to any of the three earlier states from the three earlier states, weighted by the 89.1%, 2.1% and 8.8%.

4.3) The chance of moving from this (broader) remission state to a relapse, surgical or post-surgical state were also taken as a similar weighted average.

4.4) The chance of transiting to the (broader) remission state was calculated as the sum of the probabilities of the earlier states comprising the (broader) remission state.

5) This gave a matrix in four states (remission, relapse, surgery, and post surgical remission) for two monthly cycles. This is modified to form a one month transition matrix by halving the figures off the main diagonal and setting the diagonal entries to one minus the remaining values in each row. This process creates transition matrix in Table YY, which was used in the cost effectiveness model

The on-therapy cohort needs to be able to switch to standard care if they do not respond to anti-TNF treatment after two cycles of treatment. To facilitate this two additional states are included in the anti-TNF arm of the model. The first is 'Relapse 2'. Patients who remain in relapse after the first cycle of anti-TNF therapy transit to 'Relapse 2' for a second cycle of treatment. Patients who do not respond to the second cycle of anti-TNF treatment then transit to the Standard Care Relapse state.

Note that there is a small possibility (less than 1%) that patients in the standard care relapse state could then cycle into remission and subsequently re-enter the ant-TNF treatment pathway. In the interests of parsimony of the model, it was chosen not to complicate the model structure in order to capture the costs and effects of this small number of patients.

For the maintenance model, non-responders to two cycles of treatment are assumed transit to standard care relapse. Those who subsequently enter remission *do not* transit to remission with maintenance treatment. Rather they transit to standard care remission, with no possibility of restarting maintenance therapy.

Separate analyses for severe disease and moderate disease were undertaken. This is because the clinical course framework described above did not differentiate between these two states and it is not clear how a mild/moderate division could be placed upon the active disease patients reported in

Silverstein et al. [23] The implicit assumption is that the treatments are equally likely to achieve remission in moderate and severe disease.

The model was constructed and analysed in Data TreeAge Pro 2006 Healthcare.[87]

Figure 41. Schematic of Standard Care model structure

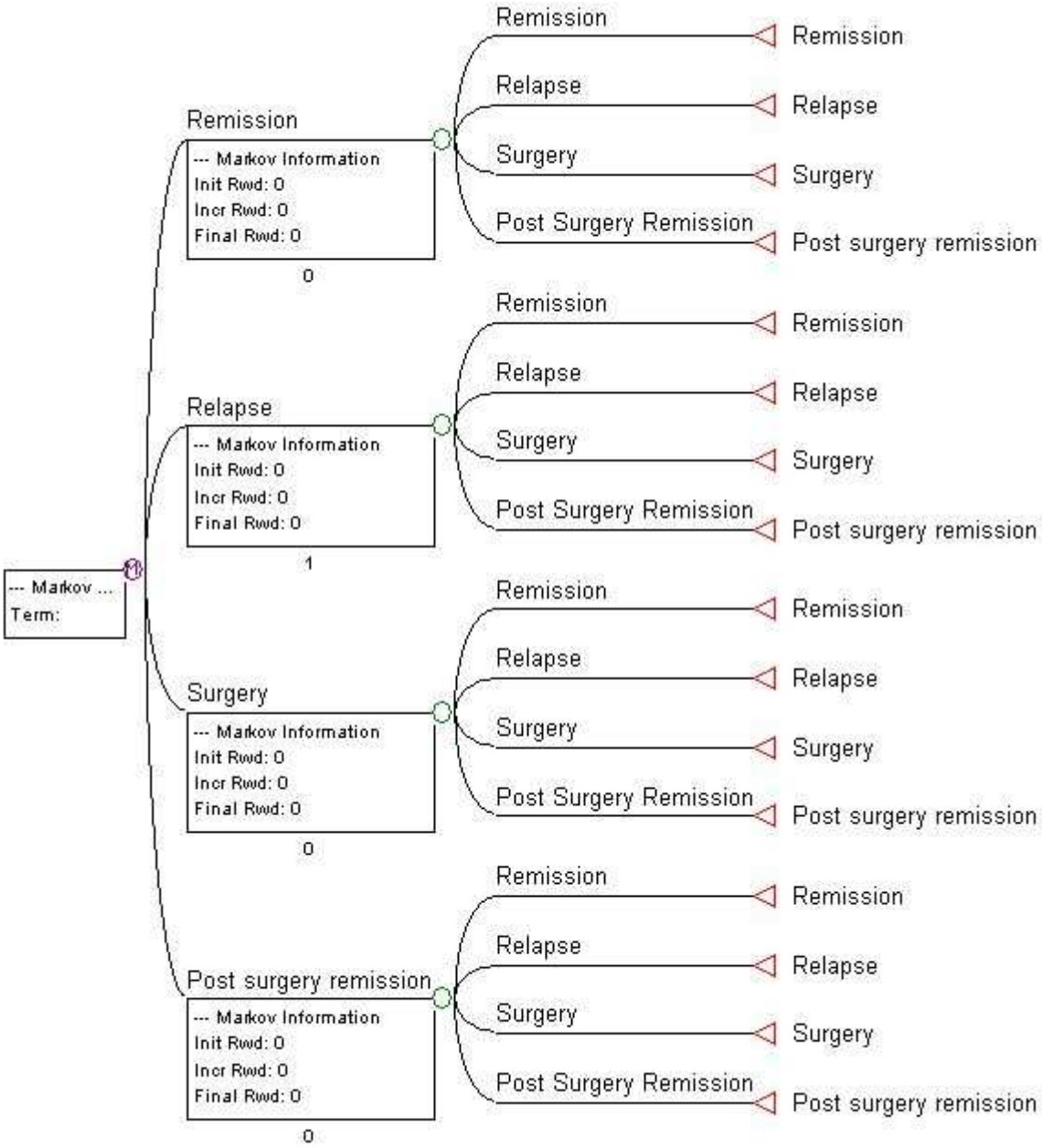
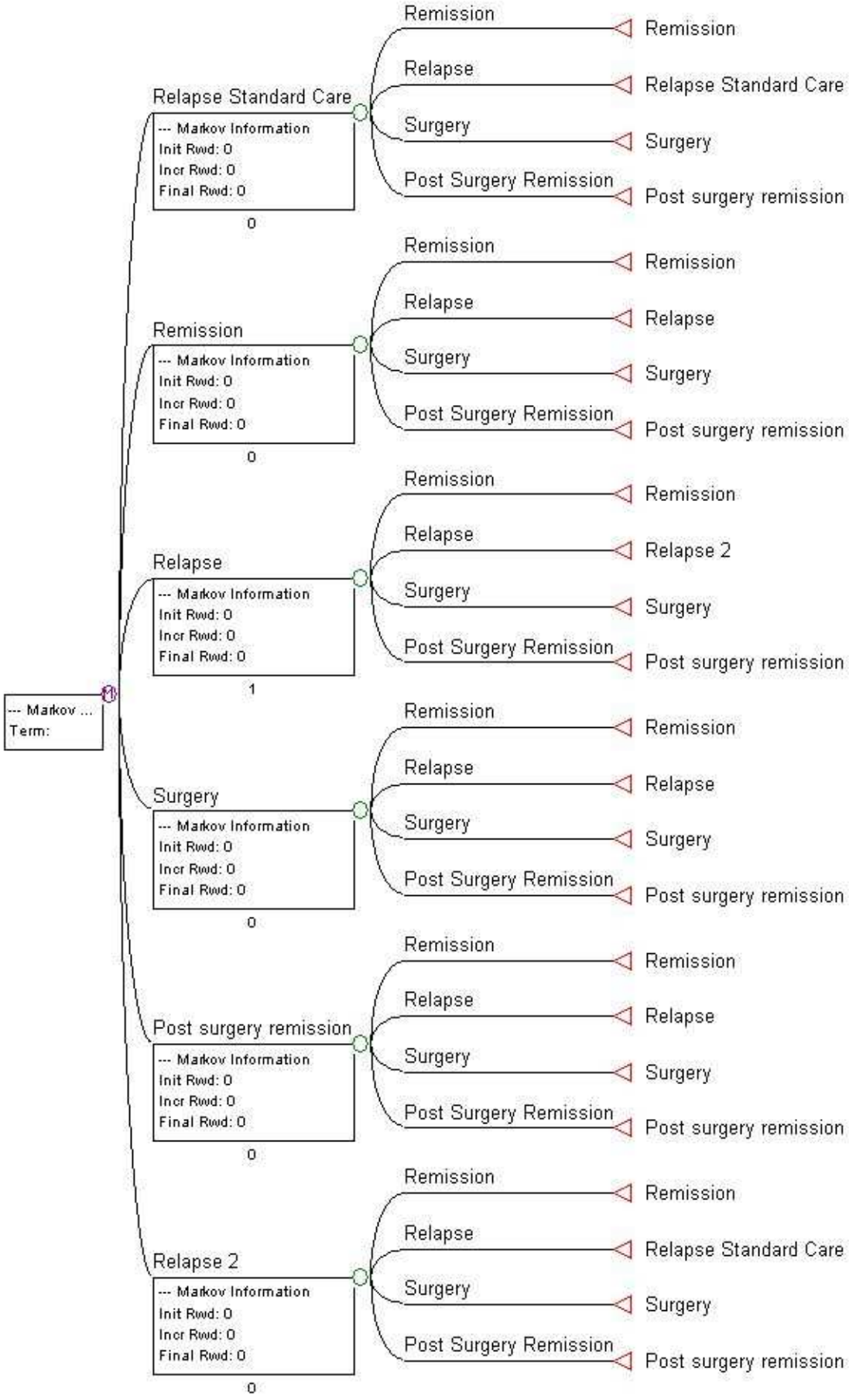


Figure 42. Schematic of anti-TNF model structure


--- Markov ...
Term:
Relapse Standard Care
--- Markov Information
Init Rwd: 0
Incr Rwd: 0
Final Rwd: 0
0
Remission
Relapse
Surgery
Post Surgery Remission
Remission
--- Markov Information
Init Rwd: 0
Incr Rwd: 0
Final Rwd: 0
0
Remission
Relapse
Surgery
Post Surgery Remission
Relapse
--- Markov Information
Init Rwd: 0
Incr Rwd: 0
Final Rwd: 0
1
Remission
Relapse 2
Surgery
Post Surgery Remission
Surgery
--- Markov Information
Init Rwd: 0
Incr Rwd: 0
Final Rwd: 0
0
Remission
Relapse
Surgery
Post Surgery Remission
Post surgery remission
--- Markov Information
Init Rwd: 0
Incr Rwd: 0
Final Rwd: 0
0
Remission
Relapse
Surgery
Post Surgery Remission
Relapse 2
--- Markov Information
Init Rwd: 0
Incr Rwd: 0
Final Rwd: 0
0
Remission
Relapse
Surgery
Post Surgery Remission

It was assumed in the model that maintenance therapy was equivalently effective in reducing the risk of patients in remission to those entering relapse. Ideally the trials would have reported separate remission to relapse and relapse to remission rates. In the absence of these data the inverse of the relapse to remission rate was applied to the remission relapse standard care transition probability. Thus, if anti-TNF maintenance therapy doubled the probability of entering remission from relapse, the model assumes it also halves the risk of relapsing amongst patients in remission.

Estimating the effectiveness for infliximab is less straightforward. The induction trial for infliximab was extremely small, (n=54)[54] In addition, the remission rate observed in the standard care arm of this study was substantially lower than that observed in the Silverstein data[23] and the much larger induction trials of adalimumab. As a result, there is substantial uncertainty around both the absolute and relative magnitude of effect for infliximab induction therapy as reported by Targan et al [54] In discussion with the Technical Lead at NICE, it was agreed that the effectiveness should be estimated by comparing the absolute remission rate for infliximab reported by Targan et al and the predicted transition rate from the standard care transition matrix. Table 52 shows the standard care transition matrix used in the model

Table 52. Transition matrix for the cost effectiveness model

	REM	REL	SUR	PSR
REM	0.9837	0.0059	0.0069	0.0035
REL	0.0713	0.8749	0.0348	0.0189
SUR	0.0521	0.0158	0.6709	0.2613
PSR	0.0054	0.0011	0.0026	0.9909

6.4.2.3 Sensitivity analysis

Monte Carlo simulation was used to estimate the expected mean costs and effects for standard care and each intervention. Each analysis used 10,000 simulations. The utilities were assumed to have a normal distribution with the mean and standard error as described by Gregor et al (1997).[77] The exception to this being the utility for the major surgery, where a value was assumed as described above. For the PSA it was assumed that the standard error around this estimate was comparable to the standard error for the Gregor et al figures and specified a normal distribution with a standard error of 0.001.

The costs, except for the drug costs, were taken from the NHS Reference.[85] Again a normal distribution was specified with the standard error being derived from the interquartile range reported in the database.

As proportions, the remission rates were specified as beta distributions. The standard errors were calculated, on the assumption of normality using the standard equation; ***a*** and ***b*** were then estimated using the following equations:

a = mean^2*(1-mean)/(se^2)

b = mean*(1-mean)/(se^2)-a

The distributions are reported in Table 53. From each analysis, the expected (mean) costs and outcomes for standard care and anti-TNF treatment, the associated expected incremental cost effectiveness ratio and the cost effectiveness acceptability curve are reported.

Table 53. Probability distributions

1	Utility_moderate_relapse	Utility distribution for moderate relapse states	Normal	0.068	0.0012
2	Remission_Utility	Remission in utility CD	Normal	0.073	0.00801
3	Adalim_Effectiveness	Adalimumab induction relative relapse rate	Beta	18	57
4	Utility_Severe_Relapse	Utility in Severe Relapse State	Normal	0.056	0.0012
5	Utility_Major_Surgery	Utility for major surgery state	Normal	0.039	0.001
6	Cost_Moderate_Relapse	Cost in moderate relapse	Normal	474	11.09
7	Major_Surgery_Cost	Cost of Major Surgery	Normal	4592	130.8
8	Cost_severe_relapse	Cost of Major Relapse	Normal	1489	39.71
9	Remission_state_cost	Cost in remission	Normal	52	3.85
10	Moderate_Surgery_Cost	Moderate Surgery Cost - Outpatient HRG F52	Normal	866	46.5
11	Moderate_surgery_utility	EQ-5D 22212 - assumed moderate surgery state utility	Normal	0.0546	0.001
12	Inflximab_Induction	Remission rate from Targan (see Figure 3)	Beta	13	14
13	Charm_placebo	Charm placebo remission proportion (see Figure 32)	Beta	36	224
14	Charm_active	Charm active maintenance (see Figure 32)	Beta	88	172
15	Accent_active	Infliximab maintenance remission rate all patients (see Figure 13)	Beta	63	113
16	Accent_Placebo	Accent placebo arm remission rate all patients (see Figure 13)	Beta	60	128

For each cost effectiveness analysis the following variables were included in the multivariate probabilistic sensitivity analysis (PSA):

- Utilities in remission, relapse, surgery, and post-surgical remission;

- Direct health care costs in remission, relapse, surgery and surgical remission;
- Effectiveness of anti-TNF therapy.

6.4.3 Results

Table 54 gives the mean costs and QALYs and expected ICERs for each intervention in induction and maintenance therapy.

Table 54. Cost Effectiveness of Anti-TNFs in CD

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	6,687.01	0.9637		6,405.44	0.9774		Anti-TNF dominates
Adalimumab Severe Disease	13,444.74	0.8866		11,215.42	0.9230		Anti-TNF dominates
Infliximab Moderate Disease	6,858.66	0.9646		10,010.62	0.9938		107,943.80
Infliximab Severe Disease	14,441.47	0.8862		12,593.69	0.9936		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	6,858.85	0.9649		14,724.78	0.9434		SC dominates
Adalimumab Severe Disease	13,447.52	0.8863		22,177.09	0.8270		SC dominates
Infliximab moderate disease	6,862.36	0.9636		30,397.34	0.9440		SC dominates
Infliximab severe disease	13,448.82	0.8876		39,980.18	0.8314		SC dominates

Figure 43 and Figure 44 present the CEACs for adalimumab and infliximab as induction therapy in patients with severe disease. Figure 45 and Figure 46 present the CEACs for adalimumab and infliximab as induction therapy in patients with moderate disease.

Figure 43. CEAC for adalimumab induction in severe disease

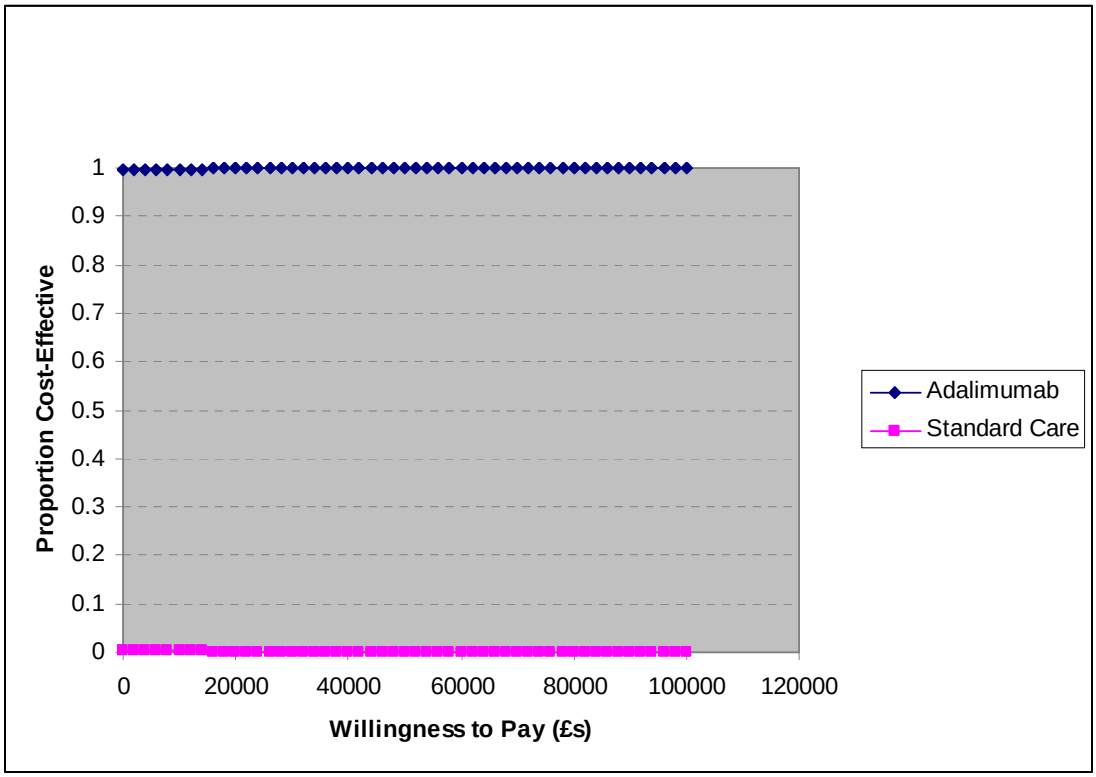


Figure 44. CEAC for infliximab induction in severe disease

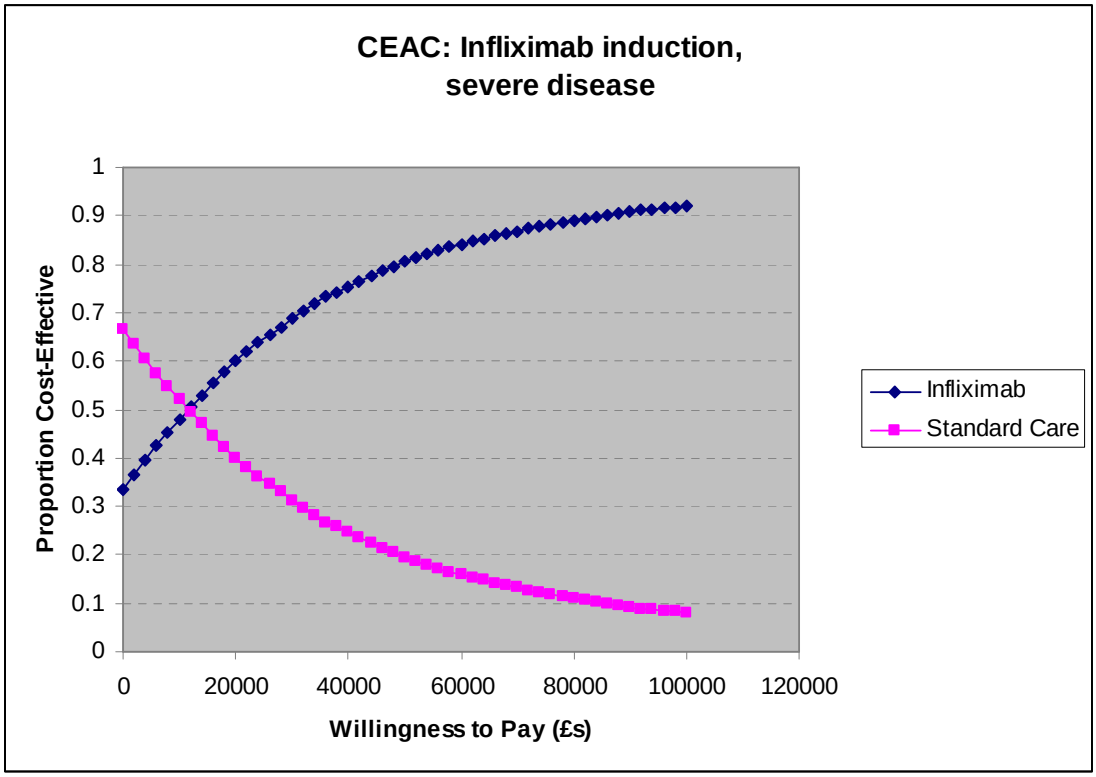


Figure 45. CEAC for adalimumab induction in moderate disease

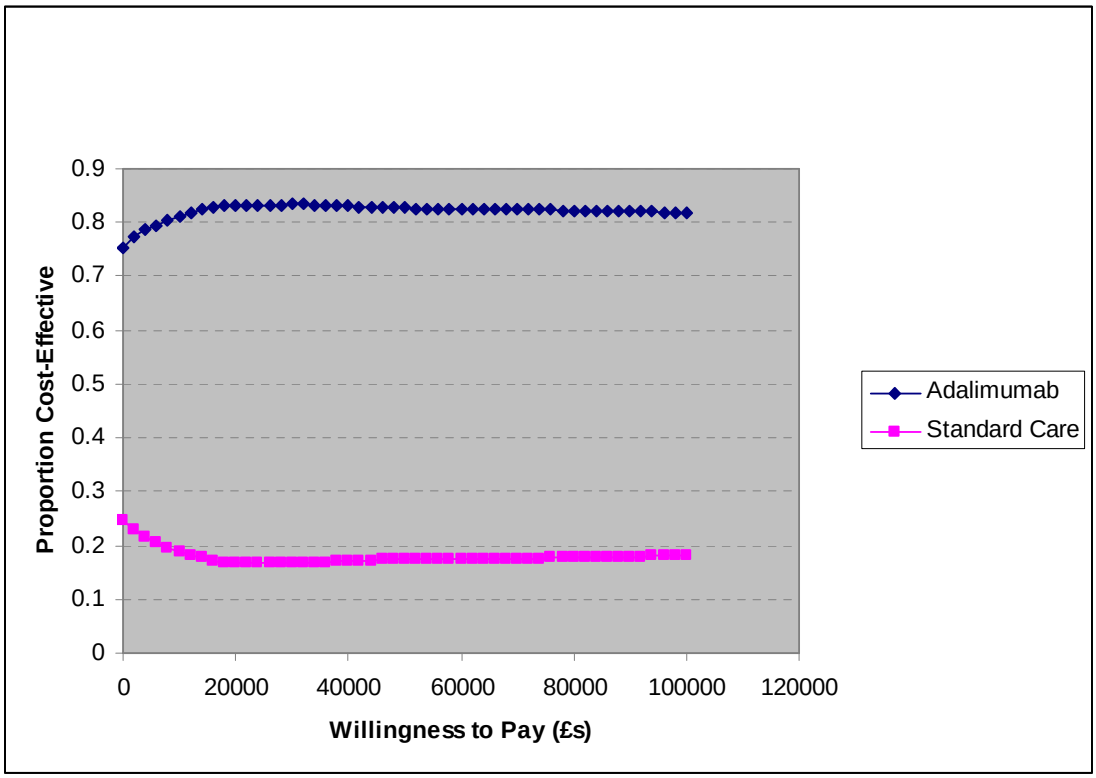
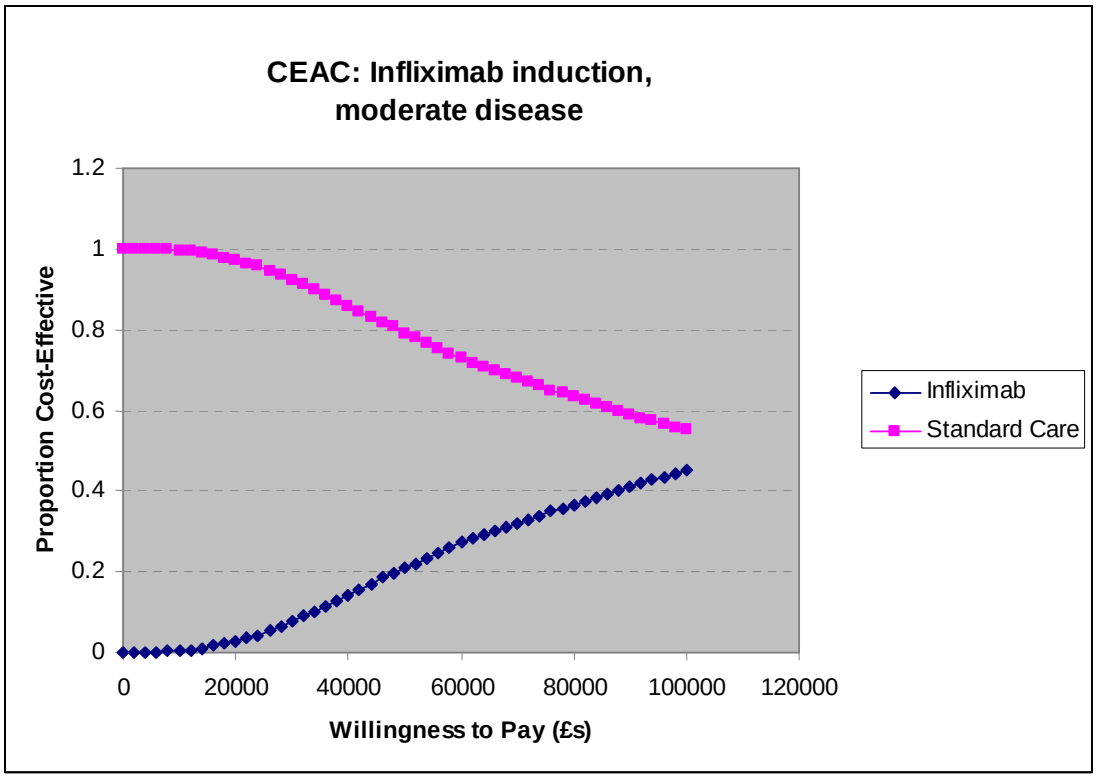


Figure 46. CEAC for infliximab induction in moderate disease



6.4.4 Discussion

The analysis described in this chapter indicates that one anti-TNF is not cost effective, according to the criteria laid out in the NICE Guide to Methods of Technology Appraisal, in the management of moderate CD, and whilst the expectation is that the other one is, there is significant uncertainty regarding its value. Neither are they cost effective as maintenance therapy for moderate or severe disease. Both treatments are highly cost effective, with no meaningful uncertainty, as induction therapy in severe disease.

The estimates of cost effectiveness in maintenance therapy must be viewed as exploratory. This is because of the shortness of the randomised placebo controlled period of the maintenance trials for these drugs. Essentially these do not provide evidence of the magnitude of effect compared to usual care due to allowing patients to cross over to 'episodic' treatment relatively quickly after trial commencement. The evidence required to model the cost effectiveness was the proportion of patients transiting between remission and relapse; and between relapse and remission, with and without treatment at regular time points. Given the absence of this evidence, it has therefore been necessary to postulate a maintenance effect based upon what has been reported in the trials. The implicit assumption for estimating the effect is that anti-TNFs interfere with the underlying biochemical process that causes relapses and that the effectiveness is equivalent whether the process has led to symptomatic relapse or not and whether the patients is in remission or relapse. Given that 81% of patients in remission were expected to be maintained in remission each year on standard care, the capacity for additional benefit from anti-TNF maintenance therapy is small; approximately 0.045 QALYs per annum for patients with severe disease. Against this background, it is unlikely that maintenance therapy has sufficient scope for generating health gain to justify its use at current prices, and by extension, there would be little value in undertaking further research into the question of the effectiveness of maintenance therapy.

Another key decision in estimating the cost effectiveness was to use the Silverstein et al[23] to model usual care for all three interventions and use the Targan et al trial[54] to provide an estimate of absolute effect, but not relative effect. The relative effect seen in the Targan et al trial[54] is an outlier due to the very low rate of remission at four weeks in the control arm. However, the absolute magnitude of effect is consistent with the remission rate seen in the pre-randomisation phase of the ACCENT I trial of infliximab.[3,4]

An important difference between this model and others is the use of a one year time horizon and the exclusion of death from the model. Silverstein et al reported a small risk of death in each state. [23] The mean risk of death varied between states, ie between 0.00015 and 0.00839. The greatest risks were

for the 'Drug Responsive' and Drug Refractory States (0.00626 and 0.00839 respectively).If these results had been used as the basis for incorporating mortality into the model and therefore adopting a lifetime horizon, the effectiveness of the drug in inducing remission would have produced an apparent mortality gain for treatment of approximately 0.00213 per additional remissions created – i.e slightly over two lives in a cohort of 1000 patients. As the total QALYs produced by induction therapy is in the region of seven QALYs per year, this would completely swamp the direct effectiveness, and make the treatments appear highly cost effective, even though the evidence does not support a *causal* link between status in the Silverstein et al [23] framework and mortality, and there is no direct evidence of a mortality benefit from anti-TNF treatment.

A further important consideration is the focus of this analysis on the cost effectiveness of these treatments in the induction of remission. The trials reported response rates for remission, CDAI response 70 and response 100. CDAI response rates cannot be converted into improvements in health without knowing the baseline CDAI for each patient. Given the non-linear relationship between utility and CDAI indicated in the Gregor et al study[77], it was not possible to attach a utility gain to a 70 or 100 point gain on the CDAI without knowledge of the pre-treatment CDAI status..

As discussed above, it was chosen to construct a model based upon health state (remission, relapse, requiring surgery and remission following surgery). This decision was guided by the desire to quantify the cost effectiveness of treatments in producing health rather than their cost effectiveness in shifting the clinical pathway. The results from Gregor et al study[77] suggested that when patients are grouped as per the Silverstein framework[23], the differences in mean health related quality of life are extremely small, and much smaller than those used in the remission, relapse etc… structure adopted in this analysis.

A simple model of CD was constructed which focused on the cost effectiveness of anti-TNF therapies in achieving or maintaining remission. The assumptions made here regarding cost of care and utilities gains from treatment have favoured the anti-TNFs over usual care. The analyses draw out the much larger health benefit for patients with severe disease compared to moderate disease and how this feeds through to ICERs that are likely to be acceptable for severe disease but not moderate disease. The analysis also highlights the important variations in effectiveness and cost between the therapies. Perhaps most importantly the analysis reflects the fact that a substantial number of patients will achieve remission under standard care and that the incidence of relapse amongst those in remission is such that maintenance therapy would have to be much less costly for it to be a cost effective option.

6.4.5 Sensitivity Analysis

The de novo cost-effectiveness model uses probabilistic sensitivity analysis to characterise the uncertainty in the clinical and cost data. This is a widely accepted method for addressing uncertainty in decision analysis modelling and is the preferred method according to the NICE reference case. However, in response to comments on the draft report from NICE and other stakeholders we have also undertaken a series of scenario analyses in order to explore the consequences for the estimates of cost-effectiveness if the necessary information was available.

In addition, stakeholders noted that there were aspects of the model that they wished to see corrected and where appropriate these changes have been made. The results of all scenarios analyses and other suggested changes are given in the text and the following set of tables:

Table 55: Original, uncorrected, base case analysis from the draft report (for reference)
Table 56: Original, uncorrected base-case analysis from the draft report, but with changes to the length of time the model runs. The model is now set to run for 13 cycles rather than 100.
Table 57: Includes changes made as suggested by the stakeholders except for the estimates of effectiveness which in this analysis still contain the 'unexplained' multipliers.
Table 58: New base case analysis including changes suggested by manufacturers and using observed events as estimate of effectiveness, not risk-ratios.
Table 59: Base case + 5 year time horizon
Table 60: Base case + 10 year time horizon
Table 61: Base case + 20 year time horizon
Table 62: Paediatric analysis: three vials
Table 63: Paediatric analysis: two vials
Table 64: 10mg/kg dosing

Table 55 shows the base-case results from the draft report that was sent for consultation. Stakeholders noticed an error in the programming for the model that led to the results being estimated over 100 cycles of four weeks instead of the intended 13 cycles. This has now been corrected and the results of this change are shown in Table 56. It is clear however that regardless of this change the overall conclusions on cost-effectiveness of the anti-TNF drugs would not have differed despite substantial differences in the ICER between the two analyses

Table 55. Original results (old base-case analysis)

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	20,141	7.247		20,302	7.2658		17,523
Adalimumab Severe Disease	30,337	7.124		27,867	7.173		Dominates
Infliximab Moderate Disease	20,134	7.241		25,455	7.282		129,781
Infliximab Severe Disease	30,339	7.131		28,990	7.24		Dominates
Maintenance							
Adalimumab Moderate Disease	20,132	7.2389		118,044	7.2595		4,753,010
Adalimumab Severe Disease	30,339	7.132		125,353	7.188		1,696,679
Infliximab moderate disease	20,137	7.2319		371,214	7.2376		61,592,456
Infliximab severe disease	30,350	7.1206		380615	7.1358		23,043,750

Table 56. Original model with correction to the time the models run

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	6,810.48	0.9644		6,900.03	0.9733		10,061.80
Adalimumab Severe Disease	13,394.60	0.8877		11,969.10	0.9221		Anti-TNF dominates
Infliximab Moderate Disease	6,806.29	0.9646		10,267.90	0.9926		123,628.93
Infliximab Severe Disease	13,92.30	0.8847		13,020.38	0.9607		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	6,807.67	0.9637		15,978.62	0.9724		1,054,132.18
Adalimumab Severe Disease	13,394.77	0.8872		21,345.61	0.9106		339,779.49
Infliximab moderate disease	6,807.51	0.9653		40,011.57	0.9670		19,531,800.00
Infliximab severe disease	13,395.30	0.8896		46,373.48	0.8912		20,611,362.50

The difference between Table 57 and Table 58 is in how we have estimated the effectiveness of each drug. In the original model, we estimated effectiveness as the relative risk of remission for adalimumab/infliximab relative to the placebo arm of the CHARM/ACCENT trials. In order for this to work in the model using Treeage software we included a multiplier to ensure that the total probability of the four events in that branch of the decision tree did not sum to greater than 1.0. The results of this analysis are included in Table 57. This method was queried and so an alternative approach was used, the results of which are in Table 58. Table 58 represents the new base-case analysis and includes those changes suggested by the stakeholders that we believed were appropriate to make.

Table 57. With RR and multipliers and all other changes

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	6,683.36	0.9651		6,516.75	0.9782		Anti-TNF dominates
Adalimumab Severe Disease	13,440.70	0.8867		11,464.96	0.9207		Anti-TNF dominates
Infliximab Moderate Disease	6,858.86	0.9644		10,395.42	0.9922		123,976.98
Infliximab Severe Disease	13,442.77	0.8852		13,086.78	0.9572		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	6,858.41	0.9640		12,849.43	0.9728		680,797.70
Adalimumab Severe Disease	13,442.17	0.8862		21,389.38	0.9093		344.035.06
Infliximab moderate disease	6,860.41	0.9653		40,071.93	0.9670		19,536,188.24
Infliximab severe disease	13,443.98	0.8871		46,419.49	0.8914		7,668,723.26

Table 58. Without RR and modifiers and with all other changes (new base-case analysis)

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	6,687.01	0.9637		6,405.44	0.9774		Anti-TNF dominates
Adalimumab Severe Disease	13,444.74	0.8866		11,215.42	0.9230		Anti-TNF dominates
Infliximab Moderate Disease	6,858.66	0.9646		10,010.62	0.9938		107,943.80
Infliximab Severe Disease	14,441.47	0.8862		12,593.69	0.9936		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	6,858.85	0.9649		14,724.78	0.9434		SC dominates
Adalimumab Severe Disease	13,447.52	0.8863		22,177.09	0.8270		SC dominates
Infliximab moderate disease	6,862.36	0.9636		30,397.34	0.9440		SC dominates
Infliximab severe disease	13,448.82	0.8876		39,980.18	0.8314		SC dominates

Following consulation on the draft report we were asked to consider an analysis of the cost-effectiveness of the anti-TNF agents based on the calculation of dosage using body surface area (BSA) instead of mass. However, advice from a clinical expert suggested that there is little evidence to suggest that dose-scaling based on BSA is likely to have an impact on the effectiveness of the treatment (Personal correspondence, Professor C Twelves, Cancer Research UK Leeds, April 2008). Morever, because the clinical evidence that is available was based on doses calculated based on mass, there is no suggestion as to what the differential effectiveness would be, making such an analysis speculative at best and highly misleading at worst. Finally, as the cost-effectiveness is based on a per-vial basis, minor adjustements to the dose required would not shift those catgories of treatments that were not cost-effective to being cost-effective, for the reasons discussed previously in the report.

We were asked to consider the cost-effectiveness of the anti-TNF treatments in the longer term and have included estimates of cost-effectiveness at 5, 10 and 20 years in Table 59, Table 60 and Table 61 respectively. We have not changed estimates of effectiveness in these scenarios as no reliable evidence is available to show the effectiveness of either drug at any of the longer-term time horizons. As a result, these results must be treated with caution. It should also be remembered that we do not have any evidence to suggest, if it were decided to alter the estimates of effectiveness, either the direction of change or the magnitude. These results are illustrative only and should not be considered to be reliable estimates of cost-effectiveness over the time frames modelled.

Table 59. Cost-effectiveness ratios at five years

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	14,072.71	4.7150		13,718.27	4.7339		Anti-TNF dominates
Adalimumab Severe Disease	25,167.34	4.6013		21,956.72	4.610		Anti-TNF dominates
Infliximab Moderate Disease	15,907.78	4.7158		19,825.54	4.7562		96,974.26
Infliximab Severe Disease	25,152.35	4.6112		23,502.55	4.7179		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	15,908.47	4.7129		68,590.94	4.4232		SC dominates
Adalimumab Severe Disease	25,145.82	4.6041		126.415.05	3.8737		SC dominates
Infliximab moderate disease	15,911.24	4.7162		131,791.85	4.4504		SC dominates
Infliximab severe disease	25,154.82	4.6033		174,536.89	3.9429		SC dominates

Table 60. Cost-effectiveness ratios at 10 years

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	20,936.27	9.4045		20,762.17	9.4252		Anti-TNF dominates
Adalimumab Severe Disease	36,324.17	9.3019		32,717.26	9.3594		Anti-TNF dominates
Infliximab Moderate Disease	25354.35	9.4601		30722.31	9.4508		120,008.60
Infliximab Severe Disease	36337.61	9.2641		35290.66	9.3844		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	25,352.24	9.4039		135,914.93	8.7702		SC dominates
Adalimumab Severe Disease	36,327.35	9.2638		230,521.75	7.6746		SC dominates
Infliximab moderate disease	25,350.12	9.4031		258,653.12	8.8204		SC dominates
Infliximab severe disease	36323.09	9.2913		343833.42	7.8221		SC dominates

Table 61. Cost-effectiveness at 20 years

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Adalimumab Moderate Disease	33,842.10	18.8296		33,775.57	18.8651		Anti-TNF dominates
Adalimumab Severe Disease	58,019.62	18.6816		53,533.09	18.7542		Anti-TNF dominates
Infliximab Moderate Disease	43,794.54	18.8253		51,632.49	18.8819		138,479.68
Infliximab Severe Disease	57,994.87	18.6323		57,488.00	18.7868		Anti-TNF dominates
Maintenance							
Adalimumab Moderate Disease	43,801.66	18.8144		270,492.17	17.4699		SC dominates
Adalimumab Severe Disease	57,969.90	18.6926		498,716.73	15.3051		SC dominates
Infliximab moderate disease	43,806.73	18.8307		512,576.53	17.5751		SC dominates
Infliximab severe disease	58,003.05	18.6571		682,431.89	15.5658		SC dominates

6.4.5.1 . Paediatric CD threshold analysis

The review of clinical effectiveness evidence reported found no good quality placebo-controlled evidence on the effectiveness of infliximab in paediatric CD. As a consequence, a threshold analysis based on the adult population effectiveness estimates to determine the estimated required effectiveness of infliximab in paediatric patients has been undertaken. This analysis was undertaken under the proviso that it must be interpreted with caution as it is often neither straight forward nor advisable to extrapolate the results of research in adults to a paediatric population. Children, it hardly needs saying, are not adults and there is no reason to uncritically accept the notion that research results that apply to adults are applicable to children. In the case of anti-TNF therapy in particular, it is important to consider how the effectiveness of the drugs might differ in a paediatric population, whether or not the same or similar adverse events can be expected, the differences in costs of the treatments including

both drug costs and the requirement for specialist paediatric services and finally the potentially different value attached to health in children when compared to adults.

Children undergo a period of rapid physiological development that is unique to that period of life. These changes affect the way in which in they respond to treatments relative to adults and in rare cases may have serious consequences (Wooltorton, 2003).[88] With respect to pharmacological treatment for various illnesses it is not always simply a case of prescribing a lower dose of the same drug, based on size or age, as it is often unclear if the drug acts in the same way in children as it does with adults. It must also be remembered that the side effect profile of a drug may differ in the paediatric population, as the oftenly cited example of paroxetine shows.[88] In that case, the drug was shown to be less effective in children than it was believed to be in adults and it also led to increases in suicidal behaviours and suicidal thoughts in children, with no evidence of such a side effect in adults.

It is also the case that the costs associated with treating children may differ from the costs of treating adults. This may be due to the different costs associated with the drug itself or related to some other factor such as the setting in which care takes place. In this case, since infliximab is dosed according to weight, the costs of treatment may be expected to be lower if a linear relationship between dose and effect is assumed (leaving aside the issue of whether the dose response relationship holds for the paediatric population as it assumed to do for the adult population). In any case, the cost of the drug is only a single factor in establishing the total cost of care for the paediatric population. It is often the case that children may need to be seen in specialist paediatric settings, which will attract different costs than those seen in adult clinics. On the whole, it should be expected that paediatric patients will cost a different amount to treat than adult patients. We have made an estimate of this cost for the threshold analysis though it is unclear how accurate this cost is given the paucity of evidence available on the required dose of the drug and the model and location of care for paediatric patients.

Owing to a lack of specific evidence, the assumption in the threshold analysis has been made that the utility weights assigned to children for all states in the model are the same as for adults in the same state. But research is clear that it is not necessarily appropriate to make this assumption. When assessing the health related quality of life of children it is necessary to consider issues related to understanding which domains of life children consider to be

important (Rosenbaum and Saigal, 1996),[89] the physiological and mental development of children from birth to adulthood (Harris and Butterworth, 2002)[90] and the social context in which children find themselves (Matza et al, 2004)[91] in relation to age-centred social roles, including aspects of life related to dependence and autonomy (Fox-Rushby and Parker, 1994).[92]

One of the most important considerations for researchers when developing or applying a measure for use in children is whether the domains of life that are being assessed are relevant and acceptable to the population being studied. Although this is true when developing a measure for use in any population, it has been argued that researchers should take particular note of this when developing measures for use with children (Petrou, 2003).[93] Those domains of life that are considered important in an adult population may not necessarily be appropriate indicators of quality of life in children (Petrou and Henderson, 2003).[94] It should be clear then that to apply adult utility values to an analysis of paediatric patients is a sub-optimal approach to the problem, though the information available permits no other course of action. .

After taking into consideration the above arguments, it is difficult to reach a reliable conclusion about the effectiveness of Infliximab for the treatment of paediatric CD (see Table 62 for results). Table 62 and Table 63 show the results from an analysis of a paediatric population where the mass of the patient is assumed to be between 40kg and < 60kg and between 20kg and < 40kg respectively. The average mass of the children in the evidence supplied by the manufacturer varied, with a mean of 49.1kg in Baldassano[43] and 43.8kg in REACH.[42] This analysis was conducted on a per vial basis, as once a vial is opened it must be used or discarded and this represents the true cost to the NHS.

It is clear in the analysis presented in Table 62 that induction therapy with infliximab for patients with severe disease is the only option that may be cost-effective, taking into the caveats of extrapolating from adult populations for effectiveness estimates as discussed above. . A threshold analysis of induction therapy in patients with moderate disease shows that even if treatment returned all patients to full health (ie, the average QALY was equal to 1.0) it would not be cost-effective, with an ICER equal to £51,071.39 – still well above the generally accepted threshold for cost-effectiveness. We undertook a similar threshold analysis for patients on maintenance therapy with both moderate and severe disease. If all

patients with moderate disease were returned to full health, the ICER for maintenance therapy would be 539,333.43. If all patients with sever disease were returned to full health the ICER for maintenance therapy would be 193,328.00.

In Table 63 there are similar results, but with one important difference. For patients with moderate disease, infliximab may be considered cost-effective for induction therapy, with an ICER of £13,573.75. For the other three scenarios there is no difference when compared with the results in Table 62. Threshold analysis of the maintenance models again shows that even where all patients return to full health, the ICER for either moderate or sever disease does not fall within generally accepted limits of cost-effectiveness. The ICER for moderate disease in this case is £410,378.80 and for severe disease £147,760.04.

It is simply the case here, as with treatment in adults, that for maintenance therapy, the potential for improvement in health compared with standard care is small, while the relative increased differences in costs are much larger and as a result it is unlikely that infliximab has the potential to be cost-effective for maintenance therapy.

Table 62. Infliximab paediatric analysis – 3 vials (children with mass between 40kg and < 60kg)

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Infliximab Moderate Disease	6,859.97	0.9654		8,627.04	0.9949		59,900.68
Infliximab Severe Disease	13,446.35	0.8871		11,220.84	0.9636		Anti-TNF dominates
Maintenance							
Infliximab moderate disease	6,859.44	0.9641		26,221.51	0.9486		SC dominates
Infliximab severe disease	13,447.09	0.8875		35,196.49	0.8435		SC dominates

Table 63. Infliximab paediatric analysis – 2 vials (children with mass between 20kg and < 40kg)

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Induction							
Infliximab Moderate Disease	6,858.17	0.9634		7,266.74	0.9935		13,573.75
Infliximab Severe Disease	13,444.03	0.8861		9,851.12	0.9622		Anti-TNF dominates
Maintenance							
Infliximab moderate disease	6,860.66	0.9651		21,182.88	0.9493		SC dominates
Infliximab severe disease	13,443.26	0.8869		30,154.92	0.8431		SC dominates

For induction, the initial dose of Infliximab costs more and is less effective than at the 5mg/kg dose so it is perhaps unsurprising that in moderate disease it compares unfavourably with standard care. The additional cost and the decreased effectiveness also suggests that 10mg/kg is not cost-effective for induction in patients with severe disease

Table 64. 10mg/kg dose. Infliximab only

	Standard Care			**Anti-TNF**			
	Mean Cost	Mean QALY		Mean Cost	Mean QALYs		ICER
Episodic							
Infliximab Moderate Disease	6860.83	0.9646		17616.22	0.9790		746,902.08
Infliximab Severe Disease	13,443.33	0.8860		22,198.30	0.9239		231,001.85
Maintenance							
Infliximab moderate disease	6,859.34	0.9652		42,326.46	0.9413		SC dominates
Infliximab severe disease	13,447.52	0.8859		52,424.52	0.8214		SC dominates

ASSESSMENT OF FACTORS RELEVANT TO THE NHS AND OTHER PARTIES

6.5 Budget impact assessment

The NICE guidance on infliximab from 2002 estimated that 31,000 patients in England and 1,800 in Wales had CD, that 2% had very severe disease and between 1050 and 4200 patients would have been eligible for treatment. These estimates were made in the absence of good quality CD prevalence studies. There is now more information on the UK prevalence of CD but not as much on the typical spread of severity.

It is estimated from the incidence/prevalence section in this report that the prevalence of CD in the UK is approximately 150 per 100,000 but could be between 50-400 per 100,000. The incidence of new cases of CD has been estimated to be approximately 5 per 100,000 per year but could be between 3.8 and 10 per 100,000 per year. The incidence and prevalence estimates from both industry submissions are shown in Table 65.

Table 65. Incidence and prevalence estimates of CD in industry submissions

	Incidence	Prevalence
Adalimumab submission[5]	10/100,000 per year used in budget impact section (derived from NICE guidance)	50-100/100,000 'however this is likely to be an underestimation' 62.5/100,000 used in budget impact section (derived from NICE guidance)
Infliximab submission[79]	14/100,000 per year	50-100/100,000, 145/100,000

These incidence and prevalence estimates are for all CD patients rather than those with moderate to severe CD or severe CD. A large cross-sectional survey of CD patients with CDAI scores used to indicate percentage with mild, moderate and severe CD was not found.

In a UK study of 172 CD patients attending a university hospital during a six month period, 7% were in remission, 33% had mild disease, 42% had severe disease, 8% had surgery and 10% were in post-surgery remission. Severity was judged by treatments being used rather than on CDAI score so severe CD patients were being treated with corticosteroids or immunosuppressive regimens.[34]

In a Canadian quality of life study, 180 consecutive CD patients referred to a tertiary care hospital had CDAI scores measured.[77] The overall mean (95%CI) CDAI score was 182 (166 to 199). There were 52 patients classified as 'chronically active therapy resistant' with mean CDAI (95%CI) of 246 (220 to 272), 34 patients classified as 'chronically active therapy responsive' with CDAI 72 (60 to 84), 45

patients classified as 'acute disease exacerbation' with CDAI 249 (217 to 281) and 49 patients in remission with CDAI 129 (110 to 148). This equates to 54% with severe disease (CDAI score greater than 220) and 46% with mild disease (CDAI score less than 220). These 46% of patients with mild disease would also be categorised as in remission (CDAI score less than 150).

In a regional cohort of 373 CD patients from Denmark, in the first year 80% had highly active disease (defined as more than four stools daily, blood or pus daily, severe or daily abdominal pains and systemic symptoms such as fever or weight loss).[22] In the second year 40% had high activity, 22% had low activity and 38% were in remission. In subsequent years the proportions were approximately 30%, 20% and 50% respectively.

From the three studies mentioned above it can be estimated that approximately 40% of CD patients will have moderate to severe disease and may be considered eligible for treatment according to the inclusion criteria for the RCTs.

In the Olmstead county cohort study of 174 CD patients, follow up information for up to 10 years was used in a Markov model to estimate the probability of future clinical course. From this it was estimated that 1.77% of CD patients might be in severe, drug refractory disease state. As this was based on a model, it may be much less reliable than actual cohort study results.[23]

With a prevalence of 150/100,000 and a total population of approximately 50 million in England and 3 million in Wales, there would be approximately 79,500 CD patients - 75,000 in England and 4,500 in Wales. If 40% had moderate to severe disease, this would be 31,800 CD patients. There is no information on the proportion of patients with severe CD as defined by a CDAI more than 300 within the moderate to severe category. However, it is noticeable that the mean CDAI score for all of the induction trials included in the clinical effectiveness review was approximately 300. These RCTs included patients described as having moderate to severe CD. From this it can be estimated that if there is a roughly normal distribution, approximately 50% of patients with moderate to severe CD will have a CDAI of more than 300.

Table 66. Estimated prevalence of CD severity

	Number in England and Wales	Percentage
All CD	79,500	100%
Moderate to severe CD	31,800	40%
Severe CD	15,900	20%
Severe and drug resistant CD (estimate from a Markov model only)	1,590	2%

The cost of treatment with the new interventions (induction and maintenance) for adults (non-fistulising CD) with both drugs is shown in Table 67. This includes the cost of administration in hospital or clinic in the case of infliximab. The administration cost would include the presence of a health professional during the two hours of the infusion and for a period of time afterwards. As there is a (small) risk of acute allergic reactions, emergency equipment should be available. No administration costs were given for adalimumab on the grounds that it can be given subcutaneously. However, training must be given before this can occur which will incur a cost.

Table 67. Estimated costs of new intervention from industry submissions

	From industry submission	Induction	Maintenance for one year
Adalimumab	Cost per 40mg vial - £357.50, No administration cost given	80mg at week 0 then 40mg at week 2 (2 doses) = £1072.50 (+admin)	40mg every other week (26 doses) = £9295 (+admin)
Infliximab	Cost per 100mg vial £419.73. Total cost per infusion £1,355.19 (assumes 60kg person so at 5mg/kg would need 3 vials, plus administration cost of £96)	One dose at weeks 0, 2 and 6 (3 doses) = £4065.57	5 mg/kg every 8 weeks (6.5 doses) = £8808.74

For infliximab the estimated three vials per person is likely to be an underestimate as the mean weight of patients from the four large trials included in the clinical effectiveness review that gave this information (CHARM[62], CLASSIC I[58], GAIN[59], Targan 1997[54]) suggested the mean weight of CD patients was approximately 71.5Kg so a dose of 5mg/Kg would require four vials per person. Also, it is unclear how the administration cost of £96 taken from an HTA report on psoriatic arthritis (Woolacott 2006[82]) was actually derived. In that HTA report, the annual administration cost for treatment (every 8 weeks so - 6.5 treatments) was estimated to be £1673.75 which equates to 257.50 per treatment. Taking these revised costs into account would give the induction dose estimate as £5,809.26 (((4 x 419.73) + 257.50) x 3) and maintenance for one year as £12,586.73 (((4 x 419.73) + 257.50) x 6.5) per person.

If 31,800 CD patients in England and Wales with moderate to severe CD receive treatment this equates to a total budget impact for both drugs that can be seen in Table 68. If only CD patients with a CDAI score more than 300 are treated, (a much more likely scenario) this equates to a total budget impact for both drugs that can be seen in Table 69. The current NICE guidance on infliximab states that it should be used in patients with severe active CD whose condition is refractory to other treatment or who are intolerant or experience toxicity from these treatments and where surgery is inappropriate. It is unclear how many people would be in this category so the precise budget impact if the current NICE guidance is maintained is unclear. The estimates below in Table 68 and Table 69 will be an overestimation.

Fistulising disease occurs in 17% to 43% of people with CD (ACCENT II[60]). In two trials in moderate to severe CD that also gave details on fistulising patients, the proportions were 14% (GAIN[59]) and 15% (CHARM[62]). Therefore it is possible that more people with fistulas have mild CD as measured by CDAI scores. If approximately 30% of all CD patients (23,850 in England and Wales) have fistulas then the estimated budget impact is shown in Table 71. Note that the prevalence used in these estimates does not include children. It is estimated that the incidence of CD in children is 5.3 per 100,000 per year (Jenkins 2001[12]) and that 20-30% of all new cases of CD are in people aged less than 20 years (infliximab industry submission[79]).

Table 68. Budget impact of new intervention for moderate to severe CD

	Induction	Maintenance for one year
Adalimumab from industry submission	£34,105,500	£295,581,000
Infliximab from industry submission	£129,285,126	£280,117,932
Infliximab from recalculation	£184,734,468	£400,258,014

Table 69. Budget impact of new intervention for severe CD

	Induction	Maintenance for one year
Adalimumab from industry submission	£17,052,750	£147,790,500
Infliximab from industry submission	£64,642,563	£140,058,966
Infliximab from recalculation	£92,367,234	£200,129,007

Table 70. Budget impact of new intervention for severe, drug resistant CD if Markov model accurate

	Induction	Maintenance for one year
Adalimumab from industry submission	£1,705,275	£14,779,050
Infliximab from industry submission	£6,464,256	£14,005,897
Infliximab from recalculation	£9,236,723	£20,012,900

Table 71. Budget impact of new intervention for fistulising CD

	Induction	Maintenance for one year
Adalimumab from industry submission	£25,579,125	£221,685,750
Infliximab from industry submission	£96,963,844	£210,088,449
Infliximab from recalculation	£138,550,851	£300,193,510

To put the above calculations into perspective, the total NHS drug bill for 2004-5 was £9,965,000,000 (Hansard 21st June 2006[95]). The mean annual cost of treating CD per patient (data collection in 2000, when infliximab was not being widely used) was approximately £3,300 (see Table 2) so if 31,800 patients were treated at that time this would have amounted to a cost of approximately £105,067,000.

As a comparison, the industry submission for adalimumab used the 2002 NICE guidance on infliximab to estimate that there would be a prevalence of 27,811, of whom 1,112 would be eligible for treatment with adalimumab. Combined with the incidence estimates for CD they estimated that 1287 CD patients would be eligible for adalimumab treatment in 2007, rising to 2000 patients in 2011. This would cost £11,971,784 in 2007, rising to £18,604,165 in 2011. They compare this to the budget impact of treating these patients with infliximab of £19,211,660 in 2007 to £29,854,950 in 2011.

The industry submission for infliximab estimated that the total cost of infliximab to the NHS per year would be £24,165,283 in the first year, rising to £38,916,321 in the fifth year. This assumed that 2,744 people would be eligible for treatment in the first year, rising to 4,419 people in the fifth year. They estimated that 28% of all patients with CD would be eligible for treatment with infliximab.

6.6 Mortality rates

No excess mortality rates with adalimumab or infliximab were found in any of the RCTs included in the clinical effectiveness review. However, there are reports in the medical press of relatively high rates of serious adverse events with disease modifying antirheumatic drugs. In a report of the serious adverse drug events reported to the US Food and Drug Administration between 1998 and 2005, infliximab was the 7th most frequently suspected drug for deaths and the 3rd most frequently suspected drug for disability and other serious outcomes. Adalimumab was also listed as having 2389 serious adverse drug events (Moore 2007[96]). It is not known how many people were taking these drugs.

In the UK, the drug analysis prints compiled from suspected adverse drug reactions are reported through the Yellow Card scheme. Fatal reactions reported up to 26 May 2006 are summarised in Table 72 below. The highest number of deaths was due to infections but it is surprising that the category of diseases of the circulatory system, particularly including myocardial infarctions, was relatively high for both adalimumab and infliximab. TB was not linked to many deaths. It is known

that the Yellow Card scheme tends to have an underreporting of adverse events. It has been calculated that £50,390,200 was spent on infliximab in 2006 (for all indications) (Hospital Prescribing England 2006[97]). Since infliximab costs £419.73 per vial, this would suggest that the NHS used 120,052 vials in 2006. If three vials are used per person, 40,000 people will have received infliximab, suggesting an overall mortality rate of very approximately 0.5%. It is unclear from this information whether there is an excess mortality in patients receiving infliximab. There is no information on the numbers of people taking infliximab for CD or the mortality rate in this patient group.

Table 72. Yellow card scheme reported deaths for adalimumab and infliximab

	Adalimumab	Infliximab
Infections (not TB)	31	70
TB	2	6
Neoplasm's	9	28
Mental and behavioural disorders	0	1
Diseases of the circulatory system	31	40
Diseases of the respiratory system	9	32
Diseases of the digestive system	0	4
Death/sudden death	11	23
Other	3	10
Total fatal outcome	96	214
Total number of reports	693	1949

7. DISCUSSION

7.1 Statement of principal findings

Clinical effectiveness review

- 11 RCTS were identified that had at least one study arm that included some participants within the UK licensed indication for adalimumab or infliximab. The results from these are summarised below. One further RCT[55] employed infliximab outside the licensed dose regimen; results from this trial and for other trial arms that used outside licence dose regimens are presented in Appendix 10.
- For adalimumab, two induction trials (CLASSIC I[58] and GAIN[59]) and two maintenance trials (CLASSIC II[61] and CHARM[62]) in adults with moderate to severe CD were identified
- For infliximab, one induction trials (Targan 1997[54]) and one maintenance trial in adults with moderate to severe CD (ACCENT I[3,4]), one induction (Present 1999[57]) and one maintenance trial (ACCENT II[60]) in adults with fistulising CD and one induction (Baldassano 2003[43]) and one maintenance trial (REACH[42]) in children with moderate to severe CD were identified
- All were placebo controlled trials, with the exception of the paediatric trials which compared different doses of infliximab, and there were no head-to head comparisons of the two drugs
- There were concerns regarding the trial design and study quality, particularly for the maintenance trials. These concerns related to the division of patients into sub-groups (responders and non-responders) at different time-points, the high proportions of scheduled cross-overs resulting in a lack of a true placebo group and uncertainties regarding the handling of missing binary and continuous data
- Particular concerns related to the ACCENT I[3,4] trial. The comparison between 'episodic' and 'scheduled' treatment described in the publication by Rutgeerts 2004 is not a valid comparison. The 'placebo' arm changed to 'episodic treatment' after 14 weeks and the scheduled maintenance arm participants could switch to episodic increased treatment. There was no randomisation to episodic and scheduled maintenance arms at the beginning of the trial.
- Statistically significant effect sizes in favour of anti-TNF therapy compared to placebo were found in all induction trials (except CLASSIC I[58], not statistically significant in favour of adalimumab) by week 4 for both CDAI response rates and remission; effect sizes in Targan 1997[54] (infliximab) were greater than those for adalimumab but were associated with greater uncertainty
- High and varied placebo response rates observed in the induction trials are thought to result from a tendency of CDAI scores to regress to the mean, from a placebo effect and possibly from differences in concomitant treatment in the trials
- There was statistically significant evidence from both large maintenance trials (CHARM[62], adalimumab and ACCENT I[4], infliximab[3]) that for the sub-groups defined as "responders" anti-TNF

therapy was beneficial compared to placebo with respect to remission or response rates at reported follow-up times. However, it appeared that point prevalence rather than sustained response (remission) was reported and so the results represented group rather than individual response (remission) and did not inform on persistence of the response (remission) state in the individual

- Indirect comparisons between adalimumab and infliximab were not done because they were judged unlikely to be valid due the heterogeneity between the trials caused by variation in placebo rates, the apparently arbitrary selection of responders only in the maintenance trials and the varied definition of responder status
- The practice of dichotomising patients into responders and non-responders was considered to only be clinically useful if 'responders' are more likely to benefit from maintenance of treatment. There was no evidence available from the identified trials to confirm or refute this
- There was evidence from both the induction and maintenance trial that infliximab promotes fistula closure to a greater extent than placebo (which was statistically significant for maintenance treatment). However, it is possible than fistula closure may not always be the most desirable outcome as it may result in increased development of abscesses
- In the paediatric infliximab trials, no reliable conclusions regarding the effectiveness of infliximab can be drawn as the spontaneous (placebo) response rates are not known; the dose response relationship observed in REACH[42] implied a beneficial effect of infliximab relative to standard care or placebo only
- Patient related quality-of-life was measured by the IBDQ in five trials (induction and maintenance). Overall there was a beneficial effect (statistically significant at some time-points) of anti-TNF therapy, shown by greater improvement (or less deterioration over time) in IBDQ scores in the treatment arms

Cost effectiveness review

- A review and quality assessment of existing published literature on the cost-effectiveness identified four papers for inclusion into the review. All concerned infliximab; no published studies on the cost-effectiveness of adalimumab were identified
- The four published infliximab cost effectiveness studies were all independently funded and the results suggested that single use or episodic treatment (various definitions) with infliximab had a relatively high cost-effectiveness ratio for both non-fistulising and fistulising disease (all above £50,000/QALY for non-fistulising disease and all above £100,000/QALY for fistulising disease).
- The results of both industry submissions (adalimumab and infliximab) typically showed ICERs of under £30,000 for both anti-TNFs versus standard care.
- For the adalimumab industry submission model there was a lack of clarity over the source and interpretation of data used in the industry model and key elements of the model could not be verified. Corrected results for both severe CD, and moderate and severe (combined) CD were substantially

higher than in the industry submitted model; in the severe sub-group of patients the corrected ICER approached cost-effectiveness (at a threshold £30,000).

- For infliximab, errors were identified in the industry model (active CD), some of which could not be corrected. The revised model was suggestive of infliximab being cost-effective for 'episodic' (clinician discretion) treatment, though the exact nature of this intervention remains unclear. Scheduled maintenance treatment with infliximab is unlikely to be cost-effective. The industry model for fistulising CD revised here also suggested that infliximab is unlikely to be cost-effective. No functioning model was provided for paediatric CD so no conclusions could be made from the reported findings.

De novo economic model

- A simple Markov model was developed from the NHS/PSS perspective to estimate the incremental cost per QALY for both drugs compared to standard care in (a) episodic therapy (as defined for the purposes of the economic model) for moderate and severe disease; and (b) maintenance therapy for moderate and severe disease. The model had a one-year time horizon and was constructed and analysed in Data TreeAge Pro 2006.
- The findings were that for 'episodic' treatment, both adalimumab and infliximab are cost effective (dominant relative to standard care) in the management of severe CD and that adalimumab (but not infliximab) is cost effective for moderate CD, according to the criteria laid out in the NICE Guide to Methods of Technology Appraisal. Neither drug is cost effective as maintenance therapy for moderate or severe disease.

Budget impact assessment

- A simple budget impact assessment was conducted using information from prevalence data and the industry submissions. It suggested that total cost to the NHS in England and Wales for induction in severe disease only could range between £17 and £92 million and for maintenance for one year between -£140 and £200 million. These totals would be less if only those CD patients whose condition is refractory to other treatment or who are intolerant or experience toxicity from these treatments and where surgery is inappropriate are treated. It is unclear how many people would be in this category so the precise budget impact if the current NICE guidance is maintained is unclear.

7.2 Strengths and limitations of the assessment

- Well established systematic review techniques were used for this technology assessment, which lends considerable strength to its validity and reliability.
- Searches for RCTs were conducted systematically. Using a sensitive search strategy is likely to have identified all of the relevant evidence; checking industry submissions did not yield additional RCTs.

- Both the licence indications (for adalimumab and infliximab) and current NICE guidance on infliximab specify the use of the drugs in 'severe' CD but the NICE scope for this work specified 'moderate to severe' CD. The identified induction RCTs (or induction phases of maintenance RCTs) included patients with moderate to severe CD or a CDAI score between 220 and 400 or 450. This means that none of the included trials matched the NICE guidance or licence indications with reference to the severity of CD. Subgroup results for patients with an initial CDAI score of 300 or more have been presented here if they were available from the trials. However, none of the trials planned for this specific subgroup so did not stratify by whether patients were above or below the 300 CDAI threshold. Furthermore, there are no consensus guidelines in the literature on what CDAI score constitutes 'severe' CD. Both the licence indications and the NICE guidance specify that adalimumab and infliximab should be used in patients who are resistant and/or intolerant to conventional treatment. Whilst many or most of the patients in the included studies were likely to meet this criterion, some may not have done. Only one study (Rutgeerts 1999[55] in infliximab) had as an inclusion criterion that patients should be treatment resistant.
- Considerable efforts were made to try to understand the flow of patients through the trials. Several of the included trials had very complicated structures where patients could take several different pathways with different tratments and these have been diagrammed to illustrate patient flow as clearly as possible.
- The assessment of relative effectiveness of adalimumab and infliximab was limited by the fact that no head-to-head comparisons were available. A formal indirect comparison was inappropriate due to clinical heterogeneity between trials, indicated by variation in placebo rates, and the variable subgroup selection of responders and non-responders.
- For dichotomous outcomes, variable placebo rates can influence the effect size values depending on the outcome measure used. In order to gain accurate estimates of effect sizes, both placebo and intervention rates and both rate differences and rate ratios are presented in the clinical effectiveness section.
- Trial designs for the maintenance trials were unusual so trial quality and any potential impact on the validity of results were investigated in detail.
- The systematic appraisal of both the published papers and industry models facilitated a comprehensive review of the cost effectiveness evidence in this area. However the evidence is limited; only four published economic studies met the review inclusion criteria, of which all considered infliximab, and none considered adalimumab. One paper was not quality assessed due to a lack of detail and the remaining three papers were of variable quality.

- The assessments of the industry models were hampered by inconsistent use of data, lack of clarity over the source and interpretation of data and, in one case, unclear details of treatment, which meant that it was not possible to satisfactorily verify or interpret the model.
- The strength of the new economic model presented here is its simple and transparent structure and inputs. However, CD is a very complex disease so it could be argued that the simple model presented here does not take account of all of the nuances of the disease. On the other hand, the more complicated a model becomes, the harder it is to establish accurate inputs to populate the model. Given that there was much uncertainty around a number of model parameters, not least the effectiveness estimates, on balance it was felt to be more appropriate to have a simpler model.

7.3 Uncertainties

- All of the included trials in the clinical effectiveness review were funded by the relevant drug companies. It is uncertain whether independently funded research in this area would yield different results. They may, however, have much more simple designs, which would aid interpretation of the results considerably.
- CD is a life-long condition with sometimes relatively long cycles of relapse and remission. The trials were mostly of one year's duration or less. It is uncertain whether the effect of the drugs would gradually wear off over time, and whether this might be associated with an increase in antibodies to the drug.
- The way the included trials were conducted and reported has provided considerable uncertainty as to the effectiveness of the drugs. Aspects to this are discussed in detail in the discussion of clinical effectiveness section (5.2.3) and, for the maintenance trials, include:
    - How the relatively large proportions who crossed over or were lost to follow up were counted
    - The use of point prevalence, rather than number of patients remaining in remission or as responders
    - Different or unclear handling of missing binary and contiuous data
    - The division of patients into sub-groups of responders and non-responders at different time-points
- One considerable uncertainty regards the division of patients as 'responders' and 'non-responders' on the basis of initial response to a single dose or up to three or four doses only. Where trials did give maintenance treatment to 'non-responders', the results have not been published. It may be that in the 'non-responder' group are CD patients who will respond to treatment but take longer to respond. The finding regarding the division of patients into responders and non-responders at specific time-points has implications for the licence indications.

The current licence indication for infliximab mentions that if patients have not responded to induction treatment within two weeks, there is no evidence to support further treatment. No evidence was identified to support this statement so it is unclear whether this part of the licence indication is evidence-based. It may be that the some of the so-called non-responders are taking longer to respond because of drug-drug interactions that have not been evaluated yet.

- The patients included in most of the trials had varying levels of severity of CD. They were mostly described as having moderate to severe CD or a CDAI score between 220 and 400 or 450. The trials were all multi-centre and there is no indication whether patients from different countries had different mean levels of severity. Patients in USA may have been enrolled at lesser severity level than UK or European patients because of the different health systems in different countries. Also there is no information on the ethnic group of participants. There is no information on whether the drugs were found to be more effective in one country compared to another or in one ethnic group compared to another. Therefore it is unclear how generalisable the results of these trials are to the UK.
- Applicability to individual patients is also uncertain. Although patients within the categories of 'moderate to severe' CD and fistulising CD may appear to be fairly homogeneous populations, this is unlikely to be the case in practice. Due the variable nature of the disease, these are actually likely to be very heterogeneous populations in terms of manifestation of disease, severity of disease, treatment (including surgical) history or concomitant medications and impact of disease on patients' lives. Therefore the effect of a drug on a specific type of patient is also unclear, and it is not known if there are sub-groups of patients who would benefit more or less from these drugs.
- The main outcome measures used in the trials are based on the CDAI, which may not be an adequate measure for capturing clinically meaningful changes in disease severity (see section 5.2.3 for further detail) or capturing aspects of quality of life such as psychological, social and occupational functioning. The disease specific quality of life measure IBDQ was reported in a number of trials but a generic quality of life measure such as EQ-5D may have been more useful.
- There was very little information from any of the included trials about hospitalization rates. This is a key cost driver in the economic models from industry and the economic model presented here. Also, some hospitalizations are for relatively minor procedures such as fistula drainage in someone who is relatively well and can just be an overnight stay whereas others are because patients are seriously ill and have to stay in hospital for weeks. Therefore simple counts of hospitalizations will not take into account all relevant information.
- The comparison of adverse events rates was affected by the design of the maintenance trials, as all patients initially received the study drug before being randomised to drug or placebo and additionally, patients in most maintenance trials had the opportunity to cross-over from placebo to

drug treatment if specified criteria were met. Therefore there is uncertainty around adverse events due to the study drug.

- The uncertainties in the clinical data (as outlined above) have complicated the economic analysis. It is difficult to define comparators where the details of treatment are uncertain. In such cases, the interpretation of economic models within the published papers becomes problematical.
- The published economic models relied heavily on a small body of data, primarily 24 years of data from the Olmstead County, USA. A Markov analysis of this data has been used widely. Similarly, in part, the industry models rely on data from small samples.
- Both the published cost effectiveness studies and the industry submission models lacked long-term data
- The analyses within all of the economic models typically used a Markov model. Markov models assume zero memory; how long a patient has been in a health state and how they got there may impact on resources. This could be important in a CD patient group.

7.4 Other relevant factors

- It was outside the remit of this assessment to look at the effectiveness of adalimumab or infliximab as first line or 'top-down' therapy. It has been suggested that there may be advantages to this approach in terms of avoiding complications such as surgery and hospitalizations. (Hommes 2005[56], Hanauer 2007[98])

8. CONCLUSIONS

8.1 Implications for service provision

Adalimumab and infliximab gave statistically significant effect sizes in favour of anti-TNF therapy compared to placebo in all induction trials for moderate to severe CD patients. There was statistically significant evidence from one large maintenance trial for adalimumab and one large maintenance trials for infliximab that for the sub-groups defined as "responders" anti-TNF therapy was beneficial compared to placebo with respect to remission or response rates at reported follow-up times. The findings of the economic model were that for induction, both adalimumab and infliximab were cost effective (dominant relative to standard care) in the management of severe CD and adalimumab (but not infliximab) was cost effective for moderate CD, according to the criteria laid out in the NICE Guide to Methods of Technology Appraisal. Neither drug was cost-effective as maintenance therapy for moderate or severe disease.

The cost effectiveness analysis highlights important variations in effectiveness and cost between the two therapies. Perhaps most importantly, the analysis reflects the fact that a substantial number of patients will achieve remission under standard care and that the incidence of relapse amongst those in remission is such that maintenance therapy with the anti-TNF drugs assessed here would have to be much less costly for it to be a cost effective option.

8.2 Suggested research priorities

- Independently funded RCT research on effectiveness of treatment
  - If the licence indication for both drugs remains for patients with severe active CD who are resistant and/or intolerant to other CD treatments, trials for anti TNF drugs should be conducted in these patients.
  - In order to take into account natural fluctuations in relapse and remission in CD, any future trials should be conducted for a period of at least one year
  - Any future trials in children should include a placebo (standard care) arm, as there is currently no evidence of the benefit of anti-TNF therapy compared to standard care
  - As there is currently no evidence that the sub-group of 'responders' is more likely to benefit than the whole group of eligible CD patients ('responders' and 'non-responders'), any future maintenance trials should be undertaken in the whole patient group; sub-group analysis of 'responders and 'non-responders' can be undertaken as part of the analysis providing the trial has sufficiently high patient numbers

- The potential benefit of 'episodic' treatment (treatment as required/deemed clinically necessary) compared to scheduled treatment should be investigated in an appropriately designed RCT, which has three treatment arms (placebo or standard care, 'episodic' treatment and scheduled treatment)
- CD is a relapsing and remitting condition. Each individual will have episodes of varying length and severity and periods of remission of varying length and mildness. Some patients will go into remission without the use of additional drug treatment. Therefore it is vital that this is taken into account when planning RCTs to assess accurately the added benefit of a particular drug treatment
- There should be no scheduled cross-overs in RCTs as this means that there is no true placebo arm and results become difficult to interpret, particularly where high proportions of patients cross over. Where patients need to use alternative treatment during the course of a trial, they should be considered as withdrawals
- Any future trials should measure quality-of-life (using a generic quality of life measure, for example EQ-5D) and should also record number and type of hospitalisations (including length of stay in hospital), as this information is important when considering the cost-effectiveness of treatments. If CDAI continues to be used as the main outcome measure there needs to be much more work on how this translates to the impact of the disease on the person. Does a change of 50 points from 150 to 200 have a similar magnitude of impact as a change from 350 to 400?
- Reporting of trial results needs to be clear, with results reported for all patients, and responders and non-responders separately if appropriate, and numbers of withdrawals at each time-point clearly stated

- Research into the natural history of CD
  - There is currently little information on the natural history of the disease in individual patients; a cohort study following individual patients over several years would provide information on the length of time patients have mild, moderate or severe disease or are in remission; this information in turn would facilitate the interpretation of trial results

.

9. APPENDICES

Appendix 1. Calculation of Crohn's Disease Activity Index (adapted from Best[36])

Variable	Description	Scoring	Multiplier
No. of liquid stools	Sum of 7 days		x 2
Abdominal pain	Sum of 7 days' ratings	0=none 1=mild 2=moderate 3=severe	x 5
General well-being	Sum of 7 days' ratings	0=generally well 1=slightly under par 2=poor 3=very poor 4=terrible	x 7
Extraintestinal complications	Number of complications listed	Arthritis/arthralgia, iritis/uveitis, erythema nodosum, pyoderma gangrenosum, aphtous stomatitis, anal fissure/fistula/abscess, fever >37.8 °C	x 20
Anti-diarrhoeal drugs	Use in the previous 7 days	0=no 1=yes	x 30
Abdominal mass		0= no 2=questionable 5=definite	x 10
Haematocrit	Expected-observed Hct	Men: 47-observed Women: 42-observed	x 6
Body weight	Ideal/observed ratio	(1-(ideal/observed)) x 100	x 1 (NOT< -10)

Appendix 2. Guidelines on the medical management of Crohn's disease

From: Carter et al., 2004[10] on behalf of the British Society of Gastroenterology

The severity of CD is more difficult to assess than UC. The general principles are to consider the site (ileal, ileocolic, colonic, other), pattern (inflammatory, stricturing, fistulising) and activity of the disease before treatment decisions are made in conjunction with the patient.

An alternative explanation for symptoms other than active disease should be considered (such as bacterial overgrowth, bile salt malabsorption, fibrotic strictures, dysmotility, gall stones) and disease activity confirmed (usually by CRP or ESR) before starting steroids. Individuals with CD have many investigations over their lifetime and imaging (colonoscopy, small bowel radiology) should not be repeated unless it will alter management or a surgical decision depends on the result.

1.1 Active ileal/ileocolonic/colonic disease

Patients should be encouraged to participate actively in the decision to treat with high dose aminosalicylates, different corticosteroids, nutritional therapy, antibiotics, new biological agents, or surgery. Infliximab is considered in section 1.5.

In mild ileocolonic CD, high dose mesalazine (4 g/daily) may be sufficient initial therapy (*grade A*).

For patients with moderate to severe disease, or those with mild to moderate ileocolonic CD that has failed to respond to oral mesalazine, oral corticosteroids such as prednisolone 40 mg daily is appropriate (*grade A*).

Prednisolone should be reduced gradually according to severity and patient response, generally over 8 weeks. More rapid reduction is associated with early relapse (*grade C*).

Budesonide 9 mg daily is appropriate for patients with isolated ileo-caecal disease with moderate disease activity, but marginally less effective than prednisolone (*grade A*).

Intravenous steroids (hydrocortisone 400 mg/day or methylprednisolone 60 mg/day) are appropriate for patients with severe disease (*grade B*). Concomitant intravenous metronidazole is often advisable, because it may be difficult to distinguish between active disease and a septic complication.

Elemental or polymeric diets are less effective than corticosteroids, but may be used to induce remission in selected patients with active CD who have a contraindication to corticosteroid therapy, or who would themselves prefer to avoid such therapy (*grade A*).

Elemental or polymeric diets are appropriate adjunctive therapy (*grade C*).

Total parenteral nutrition is appropriate adjunctive therapy in complex, fistulising disease (*grade B*).

Sulphasalazine 4 g daily is effective for active colonic disease, but cannot be recommended as first line therapy in view of a high incidence of side effects. It may be appropriate in selected patients (*grade A*).

Metronidazole 10–20 mg/kg/day, although effective, is not usually recommended as first line therapy for CD in view of the potential for side effects (*grade A*). It has a role in selected patients with colonic or treatment resistant disease, or those who wish to avoid steroids.
Topical mesalazine may be effective in left sided colonic CD of mild to moderate activity (*grade B*).
Azathioprine 1.5–2.5 mg/kg/day or mercaptopurine 0.75–1.5 mg/kg/day may be used in active CD as adjunctive therapy and as a steroid sparing agent. However, its slow onset of action precludes its use as a sole therapy (*grade A*).
Infliximab 5 mg/kg is effective (*grade A*), but is best avoided in patients with obstructive symptoms (see section 1.5).
Surgery should be considered for those who have failed medical therapy and may be appropriate as primary therapy in patients with limited ileal or ileo-caecal disease (*grade C*).

Recommendations
1.1.1 Initial treatment of active ileal or ileocolonic Crohn's disease with high dose mesalazine, corticosteroids, nutritional therapy, or surgery should be tailored to the severity of disease and take the views of the patient into account.
1.1.2 There is insufficient evidence to recommend the use of other agents outside trials or specialist centres.

1.2 Fistulising and perianal disease
Active perianal disease or fistulae are often associated with active CD elsewhere in the gastrointestinal tract. The initial aim should be to treat active disease and sepsis. For more complex, fistulising disease, the approach involves defining the anatomy, supporting nutrition, and potential surgery. For perianal disease, MRI and examination under anaesthetic are particularly helpful.
Metronidazole 400 mg tds (*grade A*) and/or ciprofloxacin 500 mg bd (*grade B*) are appropriate first line treatments for simple perianal fistulae.
Azathioprine 1.5–2.5 mg/kg/day or mercaptopurine 0.75–1.5 mg/kg/day are potentially effective for simple perianal fistulae or enterocutaneous fistulae where distal obstruction and abscess have been excluded (grade A).
Infliximab (three infusions of 5 mg/kg at 0, 2, and 6 weeks) should be reserved for patients whose perianal or enterocutaneous fistulae are refractory to other treatment and should be used as part of a strategy that includes immunomodulation and surgery (*grade A*).
Surgery (section 7), including Seton drainage, fistulectomy, and the use of advancement flaps is appropriate for persistent or complex fistulae in combination with medical treatment (*grade C*).

Elemental diets or parenteral nutrition have a role as adjunctive therapy, but not as sole therapy (*grade B*). There is insufficient evidence to recommend other agents outside clinical trials or specialist centres.

Recommendation

1.2.1 Controlled therapeutic trials combining medical and surgical therapy in perianal Crohn's disease should be conducted.

1.3 Other sites

The same general principles apply, although there are no randomised controlled trials in the treatment of gastroduodenal or diffuse small bowel disease.

Oral Crohn's disease. This is best managed in conjunction with a specialist in oral medicine. Topical steroids, topical tacrolimus, intra-lesional steroid injections, enteral nutrition, and infliximab may have a role in management but there are no randomised controlled trials.

Gastroduodenal disease. Symptoms are often relieved by proton pump inhibitors. Surgery is difficult and may be complicated by fistulation.

Diffuse small bowel disease. Stricture dilatation or strictureplasty with or without triamcinolone injection should be considered. Nutritional support before and after surgery is usually essential. Other approaches, including the combination of infliximab with surgery for residual strictures, are evolving.

1.4 Maintenance of remission

The efficacy of drug therapy appears to depend on whether remission was achieved with medical or surgical therapy, on the risk of relapse, and site of disease. Smoking cessation is probably the most important factor in maintaining remission.

To reduce the risk of relapse in CD:

All smokers should be strongly advised to stop (*grade A*), with help (counselling, nicotine patches, or substitutes) offered to achieve this.

Mesalazine has limited benefit and is ineffective at doses <2 g/day, or for those who have needed steroids to induce remission (*grade A*).

Azathioprine 1.5–2.5 mg/kg/day or mercaptopurine 0.75–1.5 mg/kg are effective, but reserved as second line therapy because of potential toxicity (*grade A*).

Methotrexate (15–25 mg IM weekly) is effective for patients whose active disease has responded to IM methotrexate (*grade A*). It is appropriate for those intolerant of, or who have failed, azathioprine/mercaptopurine therapy (*grade B*) once potential toxicity and other options, including

surgery, have been discussed with the patient. Folic acid 5 mg once a week, taken 3 days after methotrexate, may reduce side effects. Subcutaneous or oral therapy may be effective (*grade B*).
Infliximab is effective at a dose of 5–10 mg/kg every 8 weeks in patients who have responded to an initial infusion 12 weeks earlier, for up to 44 weeks (*grade A*). It is best used as part of treatment strategy including immunomodulation once other options, including surgery, have been discussed with the patient (*grade B*).
Sulphasalazine cannot be recommended (*grade A*).
Corticosteroids, including budesonide, are not effective (*grade A*), although some patients have chronic active disease who appear steroid dependent (below).

Recommendations
1.4.1 Patients with Crohn's disease who smoke should be offered help to stop.
1.4.2 Immunomodulation with azathioprine, mercaptopurine, or methotrexate is usually appropriate if patients relapse more than once per year as steroids are withdrawn.

1.5 Chronic active and steroid dependent disease
Long term treatment with steroids is undesirable. Patients who have a poor response to steroids can be divided into steroid refractory and steroid dependent. *Steroid-refractory disease* may be defined as active disease in spite of an adequate dose and duration of prednisolone (20 mg/d for 2 weeks) and *steroid dependence* as a relapse when the steroid dose is reduced below 20 mg/day, or within 6 weeks of stopping steroids. Such patients should be considered for treatment with immunomodulators if surgery is not an immediate consideration.
Azathioprine 1.5–2.5 mg/kg/day, or mercaptopurine 0.75–1.25 mg/kg/day are the first line agents of choice for steroid dependent disease (*grade A*).
Monitoring the FBC to detect neutropenia is advisable, although there is no evidence that this is effective because profound neutropenia and sepsis can develop rapidly. The FBC is best checked within 4 weeks of starting therapy and every 6–12 weeks thereafter, although may be done more frequently. Routine measurement of thiopurine methyltransferase activity before treatment, which may identify some (but not all) patients at risk of neutropenia, cannot yet be recommended but is debated. Large published series report safe use of azathioprine without TPMT assay.
Methotrexate IM 25 mg weekly for up to 16 weeks followed by 15 mg weekly is effective for chronic active disease (*grade A*). Oral dosing is effective for many patients (*grade B*).

Infliximab (5 mg/kg) should be reserved for patients with moderate to severe CD, who are refractory to or intolerant of treatment with steroids, mesalazine, azathioprine/mercaptopurine, and methotrexate, and where surgery is considered inappropriate (*grade A*).

Recommendation

1.5.1 Immunomodulation with azathioprine, mercaptopurine, or methotrexate should be tried if steroids cannot be withdrawn without deterioration in disease activity.

Appendix 3. Search strategy clinical effectiveness

Clinical Effectiveness Searches
Note: certolizumab pegol and natalizumab were originally part of this appraisal; they were subsequently excluded after searching had been completed

Source – MEDLINE (Ovid) 1950 to May Week 4 2007

1 (adalimumab or humira).mp., (540)
2 (certolizumab or cimzia).mp. (19)
3 (infliximab or remicade).mp. (3096)
4 (natalizumab or tysabri).mp.(208)
5 or/1-4 (3473)
6 Crohn Disease/ (21624)
7 crohn$.mp. (25626)
8 or/6-7 (25626)
9 5 and 8 (1046)
10 randomized controlled trial.pt. (235561)
11 controlled clinical trial.pt. (74973)
12 randomized controlled trials.sh. (48808)
13 random allocation.sh. (57966)
14 double blind method.sh. (91410)
15 single blind method.sh. (10959)
16 or/10-15 (399453)
17 (animals not human).sh. (4090275)
18 16 not 17 (365869)
19 clinical trial.pt. (436028)
20 exp clinical trials/ (191534)
21 (clin$ adj25 trial$).ti,ab. (130375)
22 ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab. (90759)
23 placebo$.ti,ab. (102414)
24 random$.ti,ab. (372182)
25 placebos.sh. (26175)
26 research design.sh. (47543)
27 or/19-26 (846379)
28 27 not 17 (743134)
29 28 not 18 (394326)
30 18 or 29 (760195)
31 9 and 30 (276)
32 limit 31 to yr="2000 - 2007" (258)

Source – MEDLINE (Ovid) 1950 to June Week 2 2007 *

1 ca2.mp. (105839)
2 d2e7.mp. (23)
3 cdp870.mp. (26)
4 pha-738144.mp. (0)
5 pha 738144.mp. (0)
6 (anti adj2 4 integrin).mp. (45)
7 anti alpha4 integrin.mp. (49)
8 anti alpha 4 integrin.mp. (32)
9 or/1-8 (105978)
10 crohn disease/ (21691)
11 crohn$.mp. (25715)
12 or/10-11 (25715)

13 9 and 12 (66)
14 randomized controlled trial.pt. (236980)
15 controlled clinical trial.pt. (75195)
16 randomized controlled trials.sh. (49205)
17 random allocation.sh. (58180)
18 double blind method.sh. (91776)
19 single blind method.sh. (11028)
20 or/14-19 (401708)
21 (animals not human).sh. (4106179)
22 20 not 21 (367711)
23 clinical trial.pt. (436884)
24 exp clinical trials/ (192444)
25 (clin$ adj25 trial$).ti,ab. (131452)
26 ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab. (91157)
27 placebo$.ti,ab. (102972)
28 random$.ti,ab. (374725)
29 placebos.sh. (26255)
30 research design.sh. (47827)
31 or/23-30 (851045)
32 31 not 21 (747002)
33 32 not 22 (396637)
34 22 or 33 (764348)
35 13 and 34 (26)
36 limit 35 to yr="2000 - 2007" (22)

* Additional search to account for alternative terminology used for the drugs.

Source – EMBASE (Ovid) 1980 to 2007 Week 22

1 (adalimumab or humira).mp. (2036)
2 (certolizumab or cimzia).mp., (230)
3 (infliximab or remicade).mp., (7811)
4 (natalizumab or tysabri).mp. (843)
5 or/1-4 (8685)
6 Crohn Disease/ (20817)
7 crohn$.mp. (23756)
8 or/6-7 (23756)
9 5 and 8 (2554)
10 limit 9 to "treatment (2 or more terms min difference)" (506)
11 limit 10 to yr="2000 - 2007" (492)

Source – EMBASE (Ovid) 1980 to 2007 Week 25*

1 ca2.mp. (115879)
2 d2e7.mp. (65)
3 cdp870.mp. (16)
4 pha-738144.mp. (1)
5 pha 738144.mp. (1)
6 (anti adj2 4 integrin).mp. (9)
7 anti alpha4 integrin.mp. (37)
8 anti alpha 4 integrin.mp. (2)
9 or/1-8 (116001)
10 crohn$.mp. (23876)
11 crohn disease/ (20928)
12 or/10-11 (23876)

13 9 and 12 (72)
14 limit 13 to ("treatment (2 or more terms min difference)" and yr="2000 - 2007") (17)

* Additional search to account for alternative terminology used for the drugs.

Cochrane Library (CENTRAL) 2007 Issue 2

#1 adalimumab OR humira
#2 certolizumab OR cimzia
#3 infliximab OR remicade
#4 natalizumab OR tysabri
#5 (#1 OR #2 OR #3 OR #4)
#6 crohn*
#7 MeSH descriptor Crohn Disease explode all trees
#8 (#6 OR #7)
#9 (#5 AND #8)

Cochrane Library (CENTRAL) 2007 Issue 2*

#1 ca2
#2 d2e7
#3 cdp870
#4 pha-738144
#5 pha next 738144
#6 antegren
#7 integrin
#8 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7)
#9 crohn*
#10 MeSH descriptor Crohn Disease explode all trees
#11 (#9 OR #10)
#12 (#8 AND #11)

* Additional search to account for alternative terminology used for the drugs.

Source – MEDLINE (Ovid) In-Process & Other Non-Indexed Citations June 04, 2007

1 (adalimumab or humira).mp (71)
2 (certolizumab or cimzia).mp (10)
3 (infliximab or remicade).mp (209)
4 (natalizumab or tysabri).mp. (14)
5 or/1-4 (249)
6 crohn$.mp. (549)
7 5 and 6 (77)
8 limit 7 to yr="2000 - 2007" (76)

Source – MEDLINE (Ovid) In-Process & Other Non-Indexed Citations June 26, 2007*

1 ca2.mp. (1007)
2 d2e7.mp. (0)
3 cdp870.mp. (0)
4 pha-738144.mp. (0)
5 pha 738144.mp. (0)
6 (anti adj2 4 integrin).mp. (0)
7 anti alpha4 integrin.mp. (2)
8 anti alpha 4 integrin.mp. (0)

9 or/1-8 (1009)
10 crohn$.mp. (602)
11 9 and 10 (0)

* Additional search to account for alternative terminology used for the drugs.

Ongoing studies

Source – National Research Register (2007 Issue 2)

See above Cochrane Library clinical effectiveness search strategy

Sources – Current Controlled Trials and ClinicalTrials.gov

Search terms: adalimumab OR humira; certolizumab OR cimzia; infliximab OR remicade; natalizumab OR tysabri; ca2 OR d2e7; cdp870 OR pha-738144; pha 738144 OR anti 4 integrin; anti alpha4 integrin OR anti alpha 4 integrin. References were selected where they also included Crohns disease.

Appendix 4. Data extraction form

Reviewer: Date:
Study author, year: Reference: Geographical location of the study:

Baseline Characteristics	Placebo n=	Drug1 n=	Drug2 n=	Drug3 n=
Mean Age ±SD				
Sex				
Ethnicity				
Mean weight (kg) ± SD				
Mean height ± SD				
Number smokers				
Mean duration of Crohn's disease (years) ± SD				
Intestinal area involved Ileum only Colon only Ileum/colon Jejunal only Perianal only other				
% with fistulas				
Where <u>all</u> with fistulising disease: Number of (draining) fistulas *Location of fistulas* Mean PDAI score				
Previous surgery for Crohn's				
Mean baseline CDAI ± SD				
Mean baseline IBDQ median (range)				
Other disease activity index or measure of disease severity (e.g. Harvey Bradshaw)				
Mean C-reactive protein (CRP) +/- SD				
Previous or concurrent biologic agent (state which) (% of patients previously received/receiving agent, % naïve)				
Other concurrent medication				
Corticosteroids (e.g. prednisone or budesonide) state which				
Immunosuppressive agent (e.g. mercaptopurine, methotrexate, azathioprine) state which				
Oral aminosalicylate				
Antibiotic				
Other: Specify				

Notes: (Identify any statistically significant differences)

Study design/methodology - See flow chart

List all outcomes:
Do not extract data on laboratory parameters

Outcomes: state which type of analysis (e.g. efficacy, ITT, safety etc.)

Outcome 1)

	Placebo n=	Drug1 n=	Drug2 n=	Drug3 n=
Baseline				
1st timepoint				
P vs placebo				
2nd timepoint				
P vs placebo				
3rd timepoint				
P vs placebo				

List number of patients for each study arm at each time-point
Repeat table for all relevant outcomes

Sub-group analyses (if applicable):

Safety

Adverse event		Placebo n=	Drug1 n=	Drug2 n=	Drug3 n=	
Average follow up						
Any Adverse Event (%)						
DEATH						
Adverse event leading to withdrawal						
GI	Nausea					
	Vomiting					
	Abdo. Pain					
CNS	Headache					
	Pain					
	Fatigue					
Infection	URTI					
	Other infection					
	Serious infection					
	TB					
Haematological						
Cardiovascular	Chest pain					
	Hypotension					
	Hypertension					
	Heart failure					
Skin	Pruritus					
	Injection site reaction (give details)					
Hypersensitivity	Acute					
	Delayed					
Respiratory	Dyspnoea					
MS	MS or symptoms of MS (e.g. demyelination)					
Bone marrow						
Other	Myalgia					
	Fever					
	Abscess					

	Antibodies to DNA					
	Human anti-TNF agent					
	Lupus arthritis					
	AE during or within 2 hrs of infusion					
	Other					
	Other					

Inclusion/exclusion criteria

Inclusion criteria:

Age/sex	
Duration of CD	
Severity of CD	
Surgical history	
Concurrent treatment (non biologics)	
Concurrent treatment (biologics)	
Previous treatment (non biologics)	
Previous treatment (biologics)	
Concurrent disease	
Female patients of child bearing potential included?	

Exclusion criteria:

Concurrent treatment (non biologics)	
Concurrent treatment (biologics)	
Previous treatment (non biologics)	
Previous treatment (biologics)	
Previous/imminent surgery	
Concurrent disease	

Are patients within UK licence in terms of severity of disease and resistance/intolerance to conventional treatment?

Follow-up of patients through trial

Number of patients enrolled:					
Number of patients excluded (state main reasons)					
Number of patients randomised:					
Number of patients at each time point and reasons for withdrawal					
	Placebo	Drug 1	Drug 2	Drug 3	Drug 4
Time point 1					
Time point 2					
Time point 3					
Time point 4					
Number completed					

Duration of study:

Number of infusions:
(how administered/where administered)

Number of assessments:

Additional notes on trial design (if applicable):

Funding source:

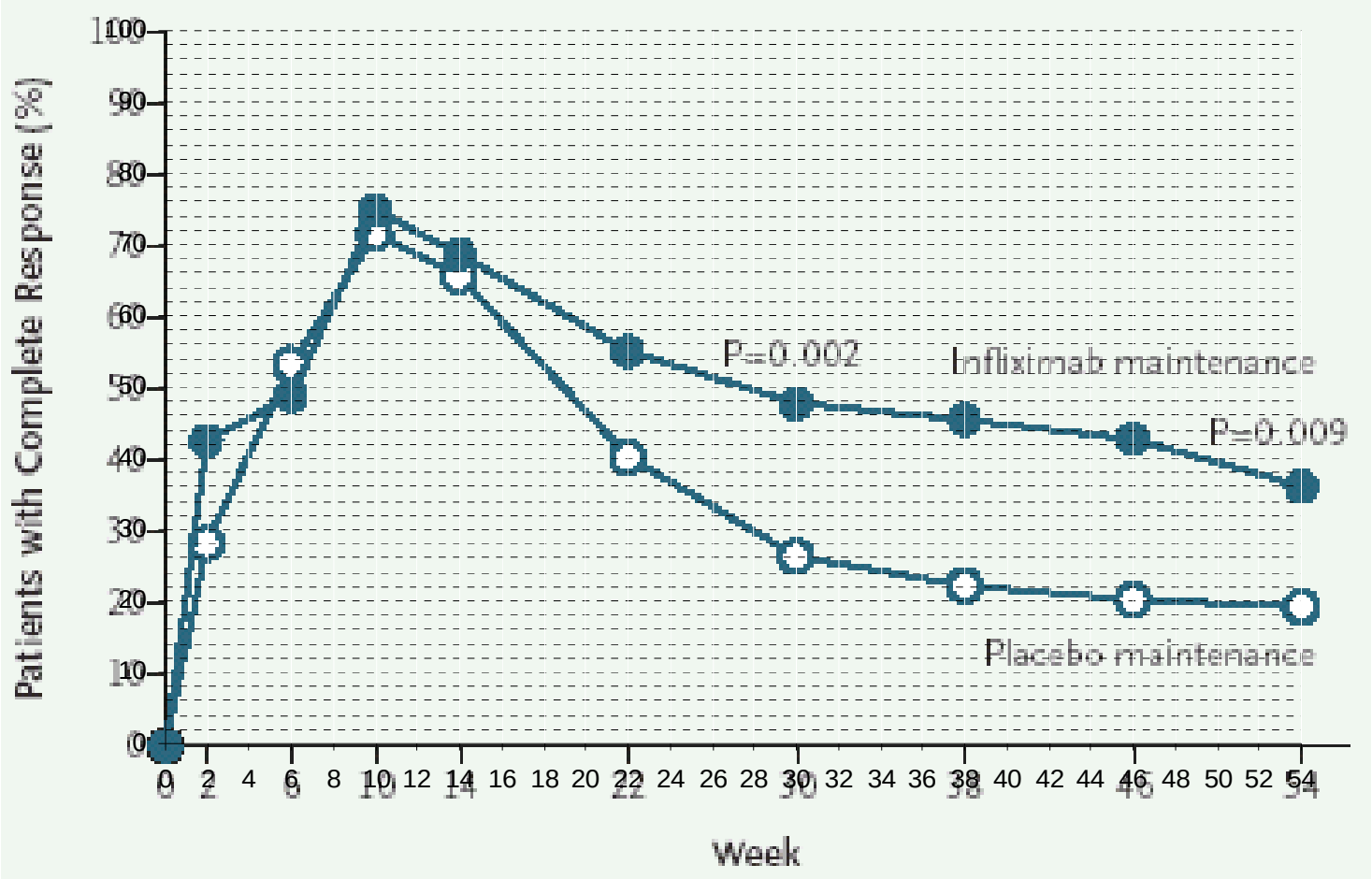
Quality Assessment

Randomisation	Details on method of randomisation	
	If described, was the method adequate?	
Concealment	Details of method of allocation concealment	
	If described, was the method adequate?	
Blinding	Details on placebo (indistinguishable from intervention?)	
	Details of blinding: patients	
	Details of blinding: study investigators	
	Details of blinding: study coordinators	
	Details of blinding: data analysts	
	Details of blinding: other	
Comparability of groups	Were groups comparable at baseline? For a) baseline scores For b) demographics	
	Were groups treated the same throughout the trial, with the exception of the intervention? For a) assessments For b) other care	
Analysis	Were all trial participants accounted for throughout trial?	
	Was loss to follow-up >20%? (state actual loss to follow-up for each time point)	
	Was it stated that an intention-to-treat analysis was performed?	
-ITT: data from all assessments used regardless of compliance with allocated treatment -Sensitivity analysis should be performed where assessment data missing	Was an ITT analysis performed for all relevant outcomes (according to the reported data), or was a sensitivity analysis performed? If other analysis (e.g. including open label patients, describe)	
	Was a sample size calculation performed?	
	Was there any selective reporting of outcome measures?	

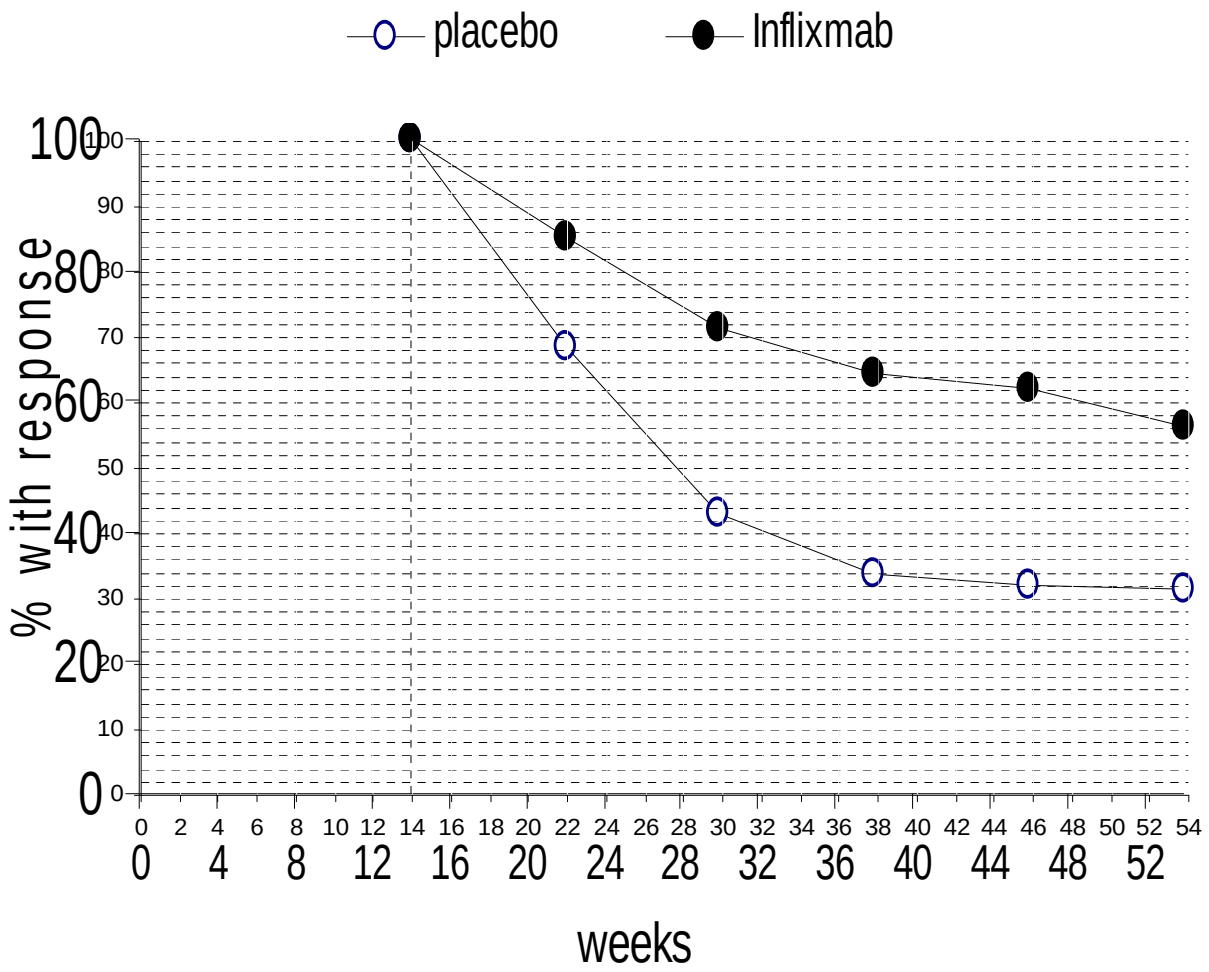
Description of which patients were included in which analysis: (primary, secondary, efficacy, ITT, open label, safety etc.)

Appendix 5. Extraction of data from published graphs

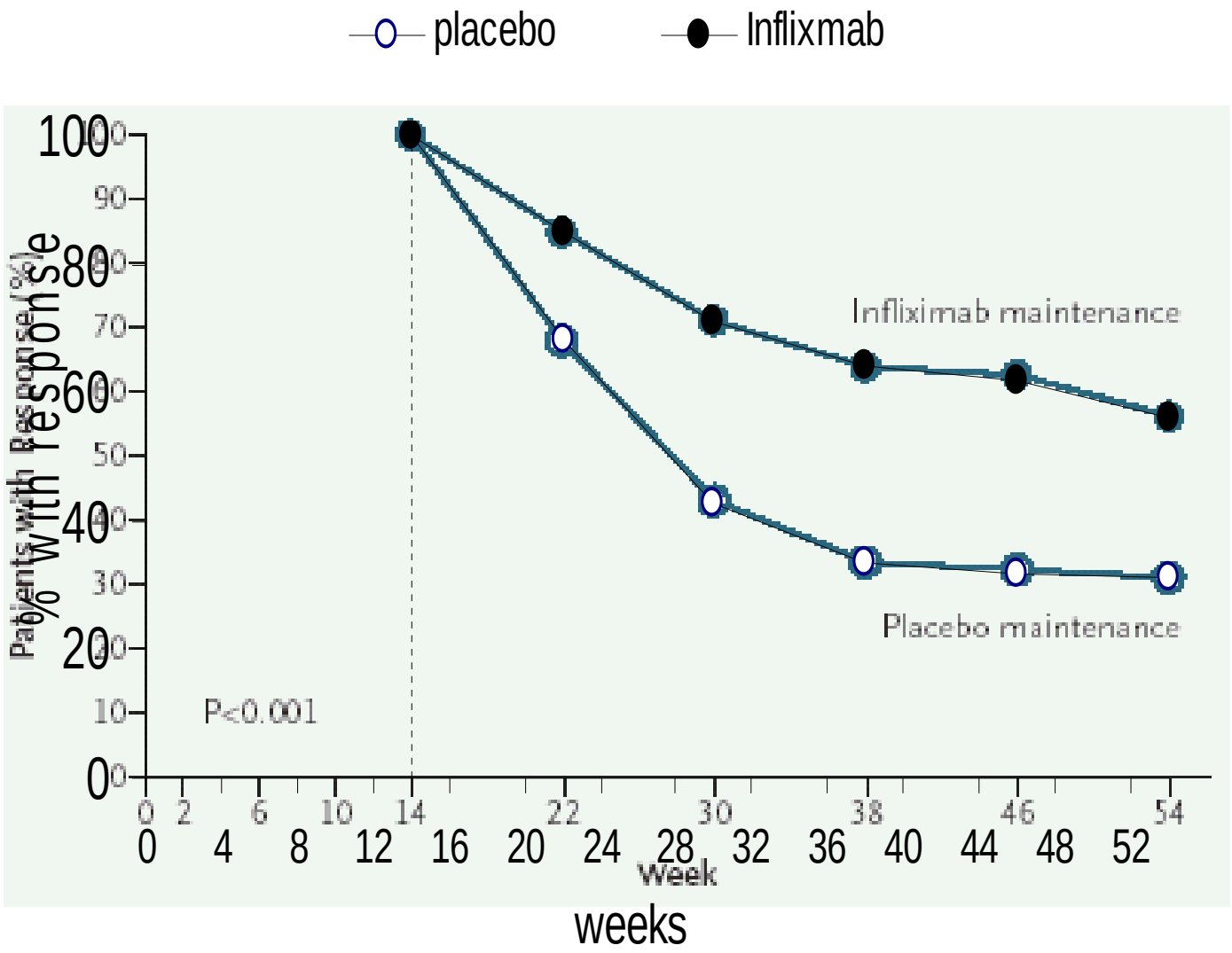
Scans of published graphs were overlayed with a grid, printed, enlarged to A3 and then used to extract data. The data was used to redraw the graph and compare with the original. Examples are shown below.
Scan of published graph with grid overlay

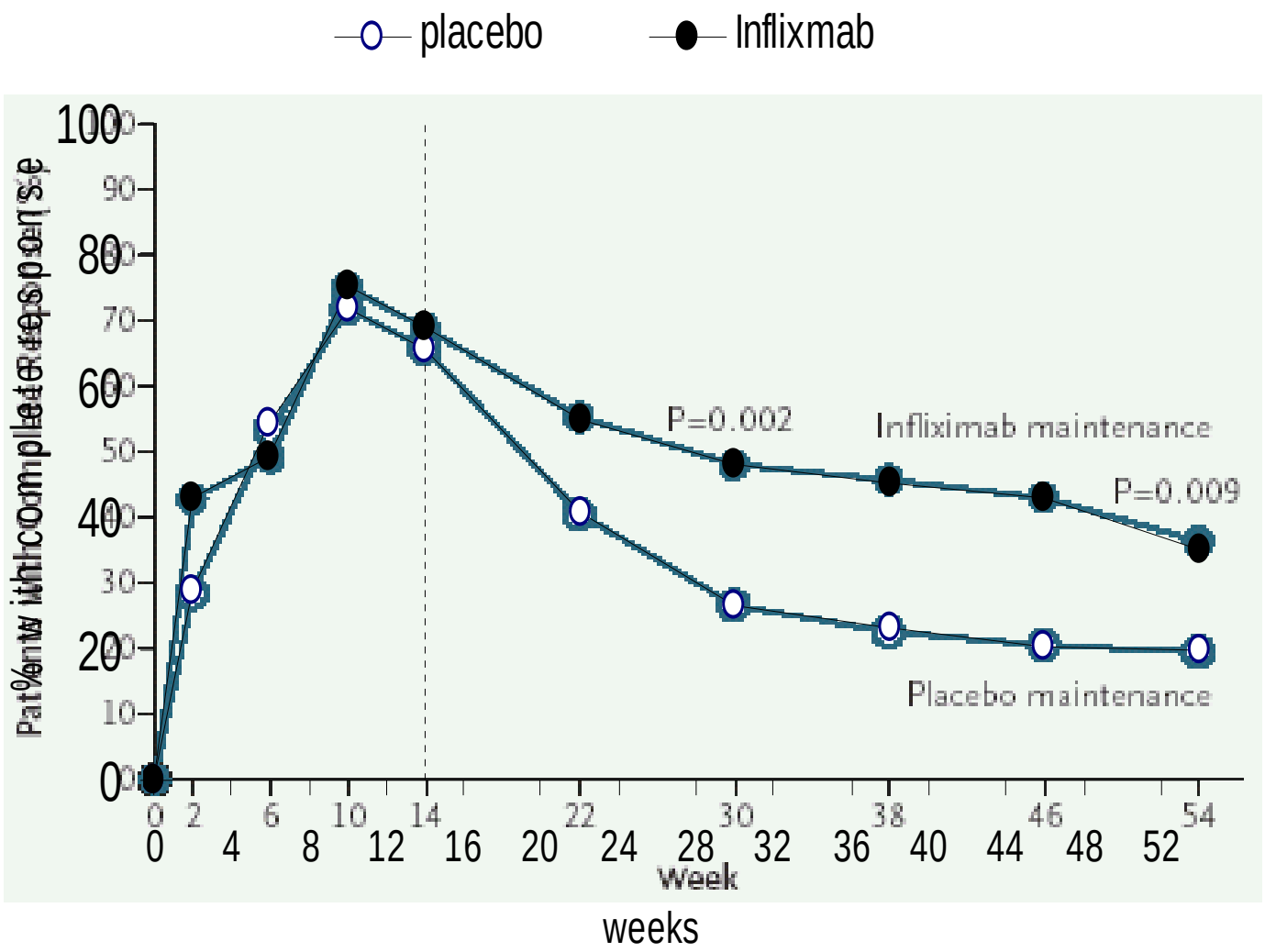


Scan of published graph with grid overlay



Scans of published graphs overlayed with graphs redrawn using data extracted from grid-overlayed originals.




placebo
Inflixmab
Patients with complete response
100
90
80
70
60
50
40
30
20
10
0
P=0.002
Infliximab maintenance
P=0.009
Placebo maintenance
0 2 6 10 14 22 30 38 46 54
0 4 8 12 16 20 24 28 32 36 40 44 48 52
Week
weeks

Appendix 6. Consistency of trials with licence indications

Table 73. Consistency of trials with licence indications

Licence indication	Study	Population/study characteristics
ADALIMUMAB		
• for treatment of severe, active Crohn's disease (NB *severe* is not further defined) • in patients who have not responded despite a full and adequate course of therapy with a corticosteroid and/or an immunosuppressant; or who are intolerant to or have medical contraindications for such therapies • recommended induction dose regimen is 80 mg at week 0 followed by 40 mg at week 2; in case there is a need for a more rapid response to therapy, the regimen 160 mg at week 0, 80 mg at week 2 can be used • after induction treatment, the recommended dose is 40 mg every other week via subcutaneous injection; patients who experience decrease in their response may benefit from an increase in dose intensity to 40 mg every week.	Hanauer et al., 2006[58] CLASSIC I INDUCTION	• moderate to severe CD (CDAI 220-450) 'despite conventional therapy' • concomitant steroids and immunosuppressants permitted (unclear if all patients resistant or intolerant) • 3 dose regimens used: 40mg/20mg, 80mg/40mg, 160mg/80mg at week 0 and 2 respectively
	Sandborn et al., 2007[59] GAIN INDUCTION	• moderate to severe CD (CDAI 220-450) • concomitant steroids and immunosuppressants permitted; unclear if all patients resistant or intolerant • all patients resistant/intolerant to infliximab • higher induction dose regimen used (160 mg at week 0, 80 mg at week 2)
	Colombel et al., 2007[62] CHARM[62] MAINTENANCE	• moderate to severe CD (CDAI 220-450) • concomitant corticosteroids and immunosuppressants permitted; unclear if all patients resistant or intolerant • 40 mg weekly or every other week compared to placebo
	Sandborn et al., 2007 CLASSIC II [61] MAINTENANCE	• see CLASSIC I[58] for patient characteristics • 40 mg weekly or every other week compared to placebo
INFLIXIMAB-ADULTS		

Licence indication	Study	Population/study characteristics
• for treatment of severe, active Crohn's disease (NB *severe* is not further defined) • in patients who have not responded despite a full and adequate course of therapy with a corticosteroid and/or an immunosuppressant; or who are intolerant to or have medical contraindications for such therapies • 5 mg/kg given as an intravenous infusion over a 2-hour period; available data do not support further treatment in patients not responding within 2 weeks • Maintenance: additional infusions of 5 mg/kg at 2 and 6 weeks after the initial dose, followed by infusion every 8 weeks • Readministration: infusion of 5 mg/kg (within 16 weeks following the last infusion) if signs and symptoms of the disease recur	Targan et al., 1997[54] INDUCTION	• moderate to severe CD, CDAI score between 220-450 • patients eligible if receiving mesalamine or oral corticosteroids or mercaptopurine or azathioprine (unclear if all patients resistant or intolerant) • single intravenous infusion over 2 hours of 5 mg/kg, 10 mg/kg or 20 mg/kg of infliximab
	Hanauer et al., 2002[4] & Rutgeerts et al., 2004 ACCENT I[3,4] MAINTENANCE	• moderate to severe CD, CDAI score between 220-450 • patients receiving corticosteroids, immunosuppressive agents, aminosalicylates or antibiotics eligible (unclear if all patients resistant or intolerant) • infusions at week 2 and 6 after the initial dose, then every 8 weeks of 5mg/kg or 10 mg/kg infliximab
	Rutgeerts et al., 1999[55] (follow-on from Targan 1997[54] trial) MAINTENANCE	• moderate to severe CD, CDAI score between 220-400 • concomitant corticosteroids or immunosuppressive agents allowed, patients who had not responded to aminosalicylates eligible • states that all patients treatment resistant (not specified which treatment (s) specifically) • 10mg/kg infliximab every 8 weeks
INFLLIXIMAB-FISTULISING CD		

Licence indication	Study	Population/study characteristics
• treatment of fistulising, active Crohn's disease in patients who have not responded despite a full and adequate course of therapy with conventional treatment (including antibiotics, drainage and immunosuppressive therapy) • an initial 5 mg/kg infusion given over a 2-hour period is to be followed with additional 5 mg/kg infusion doses at 2 and 6 weeks after the first infusion; if a patients does not respond after these 3 doses, no additional treatment with infliximab should be given • in responding patients, the strategies for continued treatment are: additional infusions of 5 mg/kg every 8 weeks or readministration if signs or symptoms of the disease recur followed by infusions of 5 mg/kg every 8 weeks	Present et al., 1999[57] INDUCTION	• single or multiple draining fistulas • concomitant aminosalicylates, corticosteroids, mercaptopurine, azathioprine or antibiotics permitted (unclear if all patients resistant or intolerant) • 5 mg/kg or 10 mg/kg infliximab at weeks 0, 2 and 6
	Sands et al., 2004[60] ACCENT II[60] MAINTENANCE	• single or multiple draining fistulas • concomitant aminosalicylates, corticosteroids, mercaptopurine, azathioprine, methotrexate or antibiotics permitted (unclear if all patients resistant or intolerant) • 5 mg/kg or 10 mg/kg infliximab at weeks 0, 2 and 6 (all patients), then 5mg/kg infliximab every 8 weeks
INFLIXIMAB-CHILDREN		
• for treatment of severe, active Crohn's disease (NB *severe* is not further defined) • in paediatric patients aged 6 to 17 years • who have not responded to conventional therapy including a corticosteroid, an immunomodulator and primary nutrition therapy; or who are intolerant or have contraindications to such therapies; Remicade has been studied only in combination with conventional immunosuppressive therapy • 5 mg/kg given as an intravenous infusion over a 2-hour period followed by additional 5 mg/kg infusion doses at 2 and 6 weeks after the first infusion, then every 8 weeks thereafter; some patients may require a shorter dosing interval to maintain clinical benefit, while for others a longer dosing interval may be sufficient Available data do not support further infliximab treatment in paediatric patients not responding within the first 10 weeks of treatment	Baldassano et al., 2003[43] INDUCTION	• moderate to severe, PCDAI ≥30 or modified CDAI ≥ 200 • active disease despite prior treatment with one or more of: corticosteroids, mercaptopurine or azathioprine, methotrexate, cyclosporine, tacrolimus • single 2-hour infusion of 1 mg/kg, 5 mg/kg or 10 mg/kg of infliximab
	Hyams et al., 2007[42] REACH MAINTENANCE	• moderate to severe, PCDAI ≥30 • Required concomitant treatment with azathioprine, mercaptopurine or methotrexate; permitted: aminosalicylates, oral corticosteroids, antibiotics or enteral nutrition (unclear if all patients resistant or intolerant) • 5 mg/kg infliximab at weeks 0, 2 and 6; followed by 5 mg/kg infliximab every 8 weeks or every 12 weeks

Appendix 7. Publications not obtained

Infliximab. A last resort for Crohn's disease after failure of steroids and azathioprine. Prescrire International 2000; 9(50):163-165.

Abramowitz L. Treatments of anoperineal localized Crohn's disease. [French]. Acta Endoscopica 2005; 35(5):748-750.

Bayes M, Rabasseda X, Prous JR. Gateways to clinical trials: November 2006. Methods & Findings in Experimental & Clinical Pharmacology 2006; 28(9):657-678.

Bayes M, Rabasseda X, Prous JR. Gateways to clinical trials: January/February 2007. Methods & Findings in Experimental & Clinical Pharmacology 2007; 29(1):53-71.

Dotan I, Yeshurun D, Hallak A, Horowitz N, Tiomny E, Reif S et al. [Treatment of Crohn's disease with anti TNF alpha antibodies--the experience in the Tel Aviv Medical Center]. [Hebrew]. Harefuah 368; 140(4):289-293.

Escher JC, van-den BG, Kate FT, te VA, van DS. Mucosal healing and down-regulation of inflammation with anti-tumor necrosis factor a (INFLIXIMAB) in children with refractory Crohn's disease. Journal of Pediatric Gastroenterology & Nutrition 2000; 31:S19.

Isaacs KL. Adalimumab induction therapy in Crohn disease. Evidence-Based Gastroenterology 2006; 7(3):67-68.

Koltun WA. A Paradigm for the Management of Complex Perineal Crohn's Disease in the Anti-TNF Era. Seminars in Colon & Rectal Surgery 2006; 17(2):61-67.

Mahadevan U. TNF-alpha antagonists: Benefits beyond remission. Reviews in Gastroenterological Disorders 2007; 7(SUPPL. 1):S13-S19.

Mealy NE, Bayes M. Treatment of gastrointestinal disorders: Certolizumab pegol. Drugs of the Future 2005; 30(6):600-601.

Schreiber S, Rutgeerts P, Fedorak RN, Khaliq-Kareemi M, Kamm MA, Boivin M et al. Erratum: A randomized, placebo-controlled trial of certolizumab pegol (CDP870) for treatment of Crohn's disease (Gastroenterology (2005) 129 (807-818)). Gastroenterology 2005; 129(5):1808.

Appendix 8. Ongoing trials

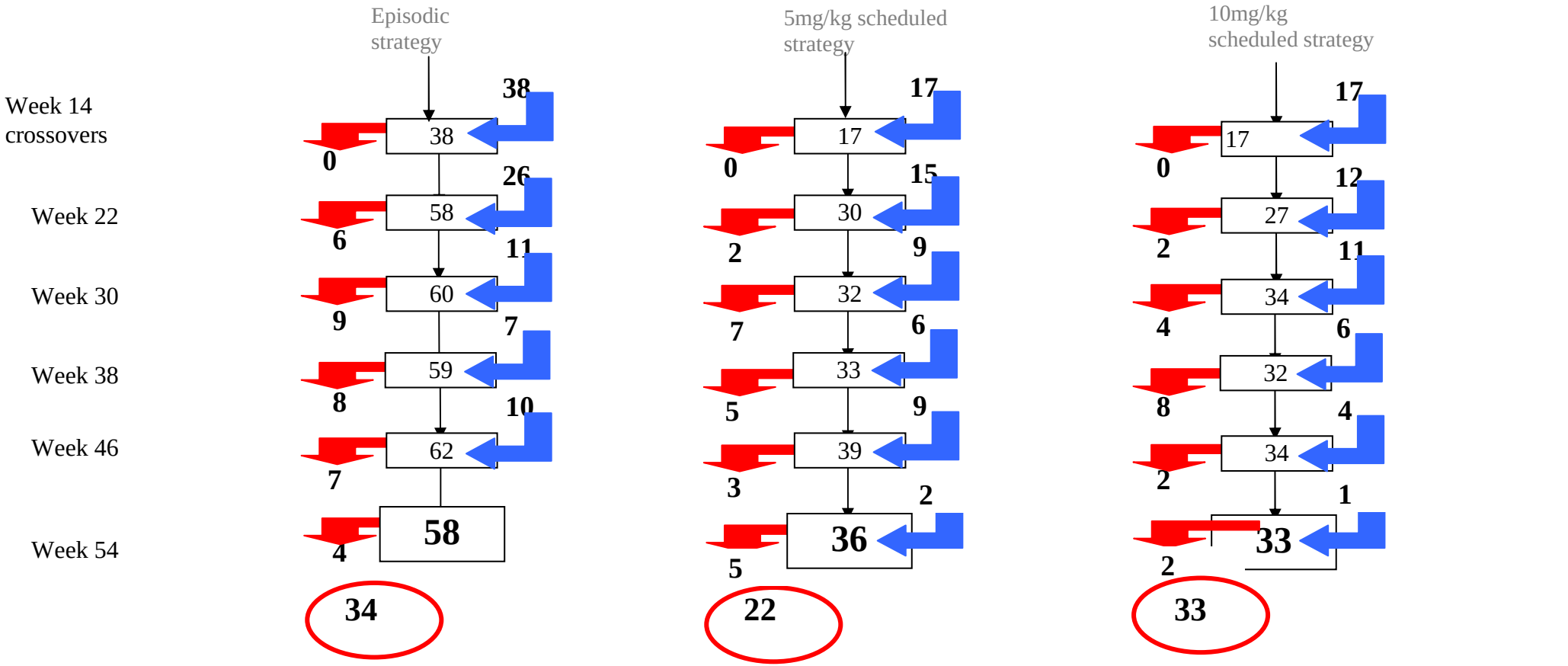
Table 74. Ongoing trials likely to meet inclusion criteria

Study/source	Country	Study design	Population	Treatment	Trial start/likely completion
Study M04-729 NCT00445939 Clinicaltrials.gov Information provided by Abbott	Japan	Multi-centre, randomised, double-blind, placebo-controlled study of adalimumab for the induction of clinical remission in Japanese subjects with Crohn's disease	Japanese subjects with Crohn's disease, CDAI score of ≥ 220 and ≤ 450; if previously received infliximab, subjects who discontinued due to a loss of response or intolerance	Adalimumab	Study start March 2007; recruitment stage (information verified March 2007)
Study M06-837 NCT00445432 Clinicaltrials.gov Information provided by Abbott	Japan	Multi-centre, randomised, double-blind, placebo-controlled study of adalimumab for the maintenance of clinical remission in Japanese subjects with Crohn's disease	Japanese subjects with Crohn's disease enrolled in and completed study M04-729	Adalimumab	Study start March 2007; recruitment stage (information verified March 2007)
Study M05-769 NCT00348283 Clinicaltrials.gov Information provided by Abbott	Multi-centre (US, Canada, Europe)	Multi-centre, randomised, double-blind, placebo-controlled study of the human anti-TNF monoclonal antibody adalimumab endoscopy trial to evaluate the effects on mucosal healing in subjects with Crohn's disease involving the colon	Patients with moderate to severe ileocolonic Crohn's disease	Adalimumab	Study start August 2006; recruitment stage (information verified April 2007)
Study M06-806 NCT00409682 Clinicaltrials.gov Information provided by Abbott	Multi-centre (US, Canada, Europe)	Multi-centre, randomised, double-blind, placebo-controlled study to evaluate the safety, efficacy and pharmacokinetics of the human anti-TNF monoclonal antibody adalimumab in paediatric subjects with moderate to severe Crohn's disease	Children aged 6-17 with moderate to severe Crohn's disease	Adalimumab	Study start March 2007; recruitment stage or not yet recruiting (information verified April 2007)

Table 75. Ongoing trials not meeting inclusion criteria

Study/source	Country	Study design	Population	Treatment	Trial start/likely completion
RP0401 NCT00132899 Information provided by Robarts Research Institute, Schering-Plough	Canada	A phase-III randomised, placebo-controlled, double-blind, parallel group, multi-centre study to evaluate the safety and efficacy of infliximab with methotrexate for the long-term treatment of Crohn's disease	Patients with symptoms that are persistent enough to require corticosteroid therapy	Infliximab (versus infliximab + methotrexate)	Study start December 2005, expected completion Dec 2007 (information verified December 2005)
CR004804 NCT00094458 Information provided by Centocor, Inc., Schering-Plough	Multi-centre (US, Canada, Europe	Multi-centre, randomised, double-blind, active controlled trial comparing Remicade® (Infliximab) and Remicade plus azathioprine in the treatment of patients with Crohn's disease naïve to both immunomodulators and biologic therapy (SONIC trial)	Patients with CDAI score of >220-<450	Infliximab (versus infliximab plus azathioprine)	Study start March 2005; recruitment stage or no longer recruiting (information verified May 2007)

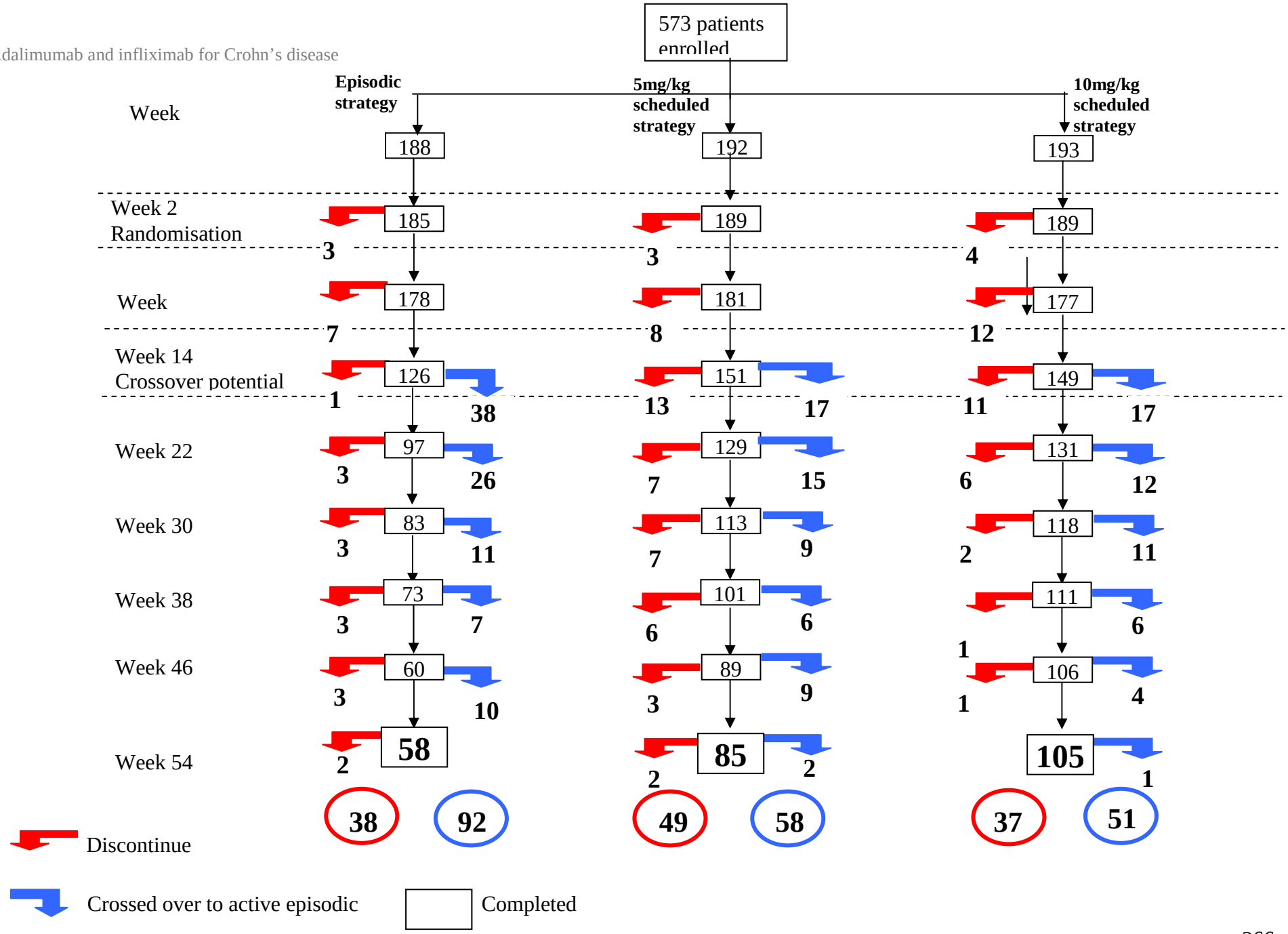
Appendix 9. Flow of patients through ACCENT I[3,4] trial



Flow-chart for patients who crossed over after week 14.

- Patients who crossed over to active episodic retreatment

- Patients who discontinued active episodic retreatment


573 patients enrolled
Episodic strategy
5mg/kg scheduled strategy
10mg/kg scheduled strategy
Week
188
192
193
Week 2 Randomisation
185
189
189
3
3
4
Week
178
181
177
7
8
12
Week 14 Crossover potential
126
151
149
1
38
13
17
11
17
Week 22
97
129
131
3
26
7
15
6
12
Week 30
83
113
118
3
11
7
9
2
11
Week 38
73
101
111
3
7
6
6
6
Week 46
60
89
106
3
10
3
9
1
4
Week 54
58
85
105
2
2
2
1
38
92
49
58
37
51
Discontinue
Crossed over to active episodic
Completed

Appendix 10. Results for all included studies irrespective of licence indication

This appendix presents the results from the included trials by outcome measure in the form of Forest plots.

INDUCTION TRIALS

Figure 47. Induction trials.— rate ratio of remission

STUDY	N	week	Drug	DOSE mg / kg	DOSE (weeks)	RR
Targan 1997	52	1	Inflix	5	0	4.63
Targan 1997	52	2	Inflix	5	0	10.19
Targan 1997	52	3	Inflix	5	0	11.11
Targan 1997	52	4	Inflix	5	0	12.04
Targan 1997	53	1	Inflix	10	0	2.679
Targan 1997	53	2	Inflix	10	0	5.357
Targan 1997	53	3	Inflix	10	0	5.357
Targan 1997	53	4	Inflix	10	0	6.25
Targan 1997	53	1	Inflix	20	0	2.679
Targan 1997	53	2	Inflix	20	0	5.357
Targan 1997	53	3	Inflix	20	0	5.357
Targan 1997	53	4	Inflix	20	0	6.25
CLASSIC I	148	1	Ada	40 / 20	0 & 2	2.4
CLASSIC I	148	2	Ada	40 / 20	0 & 2	1
CLASSIC I	148	4	Ada	40 / 20	0 & 2	1.444
CLASSIC I	149	1	Ada	80 /40	0 & 2	1.973
CLASSIC I	149	2	Ada	80 /40	0 & 2	1.48
CLASSIC I	149	4	Ada	80 /40	0 & 2	1.973
CLASSIC I	150	1	Ada	160 / 80	0 & 2	2.337
CLASSIC I	150	2	Ada	160 / 80	0 & 2	1.753
CLASSIC I	150	4	Ada	160 / 80	0 & 2	2.921
GAIN	325	1	Ada	160 / 80	0 & 2	1.74
GAIN	325	2	Ada	160 / 80	0 & 2	3.55
GAIN	325	4	Ada	160 / 80	0 & 2	2.958

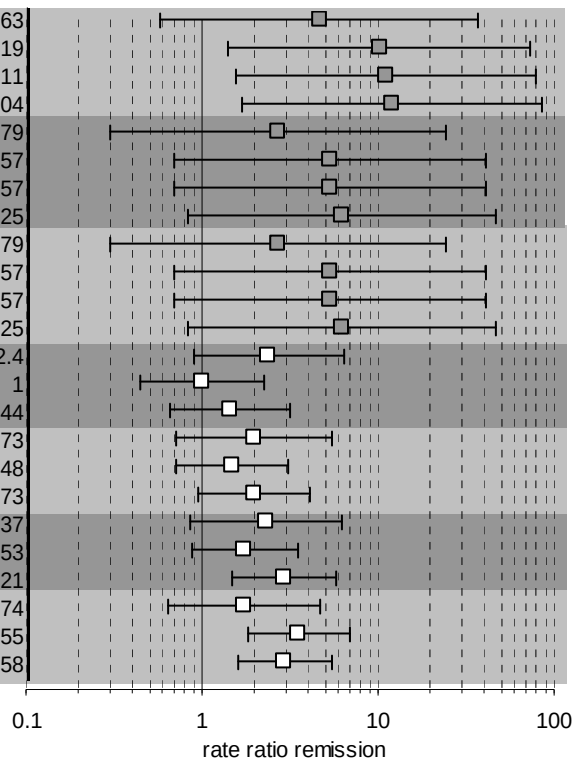


Figure 48. Induction trials.— rate difference of remission

STUDY	N	week	Drug	DOSE mg / kg	DOSE (weeks)	RD
Targan 1997	52	1	Inflix	5	0	0.15
Targan 1997	52	2	Inflix	5	0	0.37
Targan 1997	52	3	Inflix	5	0	0.40
Targan 1997	52	4	Inflix	5	0	0.44
Targan 1997	53	1	Inflix	10	0	0.07
Targan 1997	53	2	Inflix	10	0	0.17
Targan 1997	53	3	Inflix	10	0	0.17
Targan 1997	53	4	Inflix	10	0	0.21
Targan 1997	53	1	Inflix	20	0	0.07
Targan 1997	53	2	Inflix	20	0	0.17
Targan 1997	53	3	Inflix	20	0	0.17
Targan 1997	53	4	Inflix	20	0	0.21
CLASSIC I	148	1	Ada	40 / 20	0 & 2	0.09
CLASSIC I	148	2	Ada	40 / 20	0 & 2	0.00
CLASSIC I	148	4	Ada	40 / 20	0 & 2	0.05
CLASSIC I	149	1	Ada	80 /40	0 & 2	0.07
CLASSIC I	149	2	Ada	80 /40	0 & 2	0.06
CLASSIC I	149	4	Ada	80 /40	0 & 2	0.12
CLASSIC I	150	1	Ada	160 / 80	0 & 2	0.09
CLASSIC I	150	2	Ada	160 / 80	0 & 2	0.10
CLASSIC I	150	4	Ada	160 / 80	0 & 2	0.23
GAIN	325	1	Ada	160 / 80	0 & 2	0.03
GAIN	325	2	Ada	160 / 80	0 & 2	0.15
GAIN	325	4	Ada	160 / 80	0 & 2	0.14

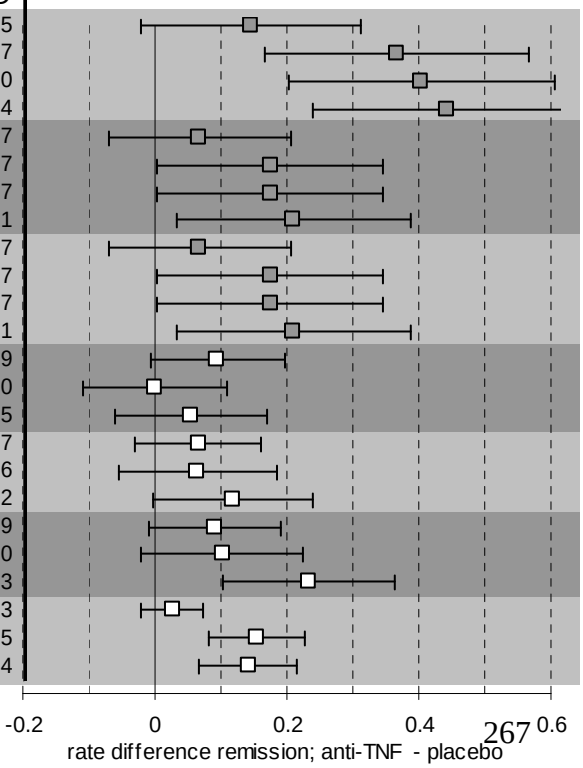


Figure 49. Induction trials.— rate ratio (RR) of response 100

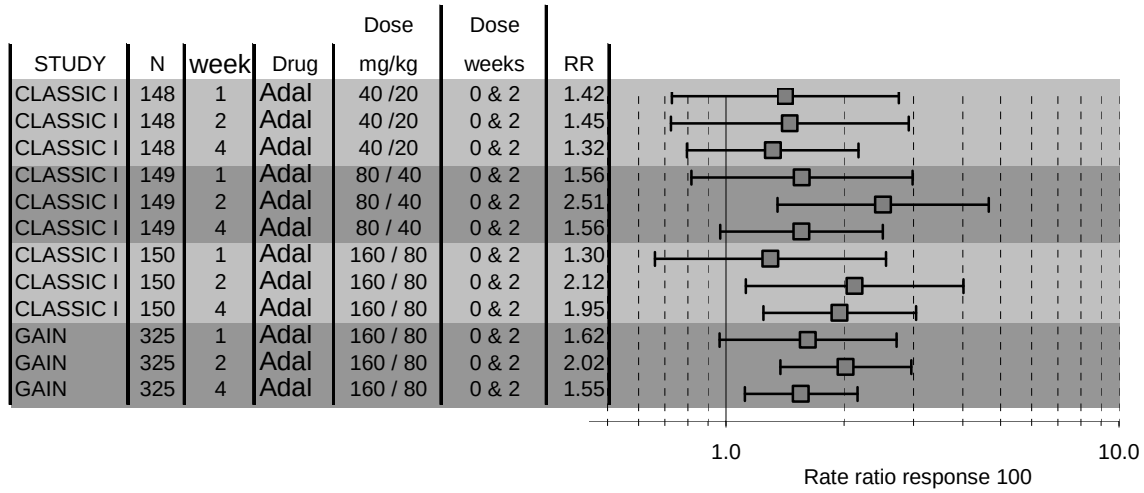


Figure 50. Induction trials.— rate difference (RD) of response 100

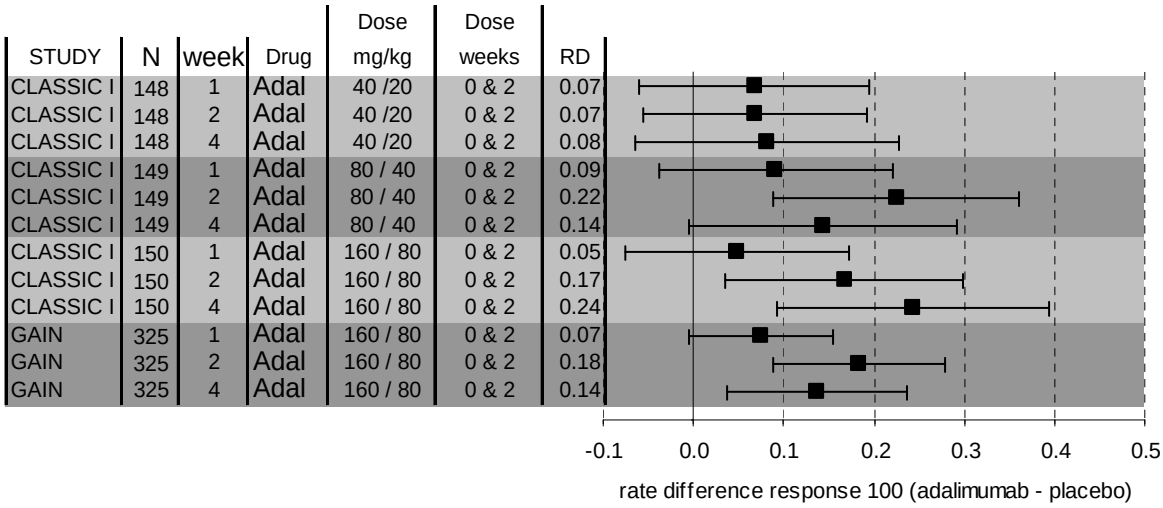


Figure 51. Induction trials .— rate ratio (RR) response 70

STUDY	N	week	Drug	Dose mg / kg	Dose week	RR
Targan 1997	52	1	Infl	5	0	5.09
Targan 1997	52	2	Infl	5	0	4.86
Targan 1997	52	3	Infl	5	0	5.09
Targan 1997	52	4	Infl	5	0	5.09
Targan 1997	53	1	Infl	10	0	3.57
Targan 1997	53	2	Infl	10	0	3.35
Targan 1997	53	3	Infl	10	0	3.13
Targan 1997	53	4	Infl	10	0	3.13
Targan 1997	53	1	Infl	20	0	3.57
Targan 1997	53	2	Infl	20	0	3.35
Targan 1997	53	3	Infl	20	0	3.57
Targan 1997	53	4	Infl	20	0	4.02
CLASSIC I	148	1	Ada	40 / 20	0 & 2	1.5
CLASSIC I 2006	148	2	Ada	40 / 20	0 & 2	1.5
CLASSIC I 2006	148	4	Ada	40 / 20	0 & 2	1.48
CLASSIC I 2006	149	1	Ada	80 / 40	0 & 2	1.64
CLASSIC I 2006	149	2	Ada	80 / 40	0 & 2	1.84
CLASSIC I 2006	149	4	Ada	80 / 40	0 & 2	1.61
CLASSIC I 2006	150	1	Ada	160 / 80	0 & 2	1.3
CLASSIC I 2006	150	2	Ada	160 / 80	0 & 2	1.5
CLASSIC I 2006	150	4	Ada	160 / 80	0 & 2	1.62
GAIN	325	1	Ada	160 / 80	0 & 2	1.69
GAIN	325	2	Ada	160 / 80	0 & 2	1.6
GAIN	325	4	Ada	160 / 80	0 & 2	1.53

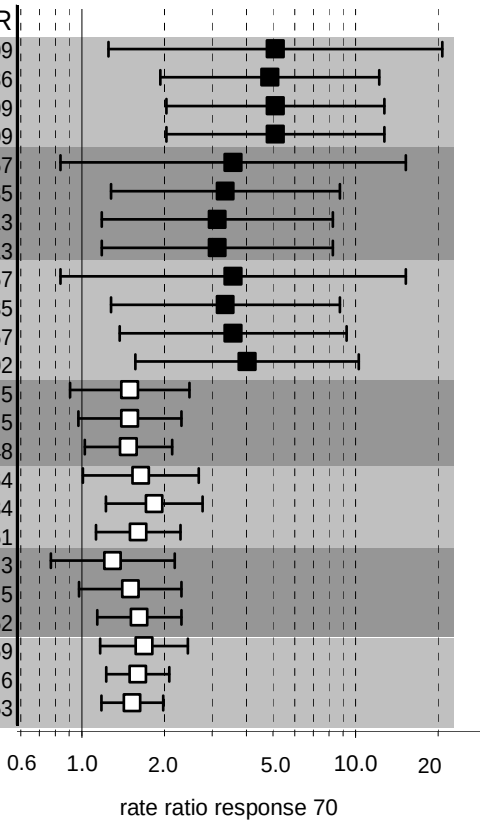


Figure 52. Induction trials .— rate difference response 70

STUDY	N	week	Drug	Dose mg / kg	Dose week	RD
Targan 1997	52	1	Infl	5	0	0.33
Targan 1997	52	2	Infl	5	0	0.62
Targan 1997	52	3	Infl	5	0	0.65
Targan 1997	52	4	Infl	5	0	0.65
Targan 1997	53	1	Infl	10	0	0.21
Targan 1997	53	2	Infl	10	0	0.38
Targan 1997	53	3	Infl	10	0	0.34
Targan 1997	53	4	Infl	10	0	0.34
Targan 1997	53	1	Infl	20	0	0.21
Targan 1997	53	2	Infl	20	0	0.38
Targan 1997	53	3	Infl	20	0	0.41
Targan 1997	53	4	Infl	20	0	0.48
CLASSIC I	148	1	Ada	40 / 20	0 & 2	0.12
CLASSIC I	148	2	Ada	40 / 20	0 & 2	0.15
CLASSIC I	148	4	Ada	40 / 20	0 & 2	0.18
CLASSIC I	149	1	Ada	80 / 40	0 & 2	0.16
CLASSIC I	149	2	Ada	80 / 40	0 & 2	0.25
CLASSIC I	149	4	Ada	80 / 40	0 & 2	0.22
CLASSIC I	150	1	Ada	160 / 80	0 & 2	0.07
CLASSIC I	150	2	Ada	160 / 80	0 & 2	0.15
CLASSIC I	150	4	Ada	160 / 80	0 & 2	0.23
GAIN	325	1	Ada	160 / 80	0 & 2	0.14
GAIN	325	2	Ada	160 / 80	0 & 2	0.2
GAIN	325	4	Ada	160 / 80	0 & 2	0.18

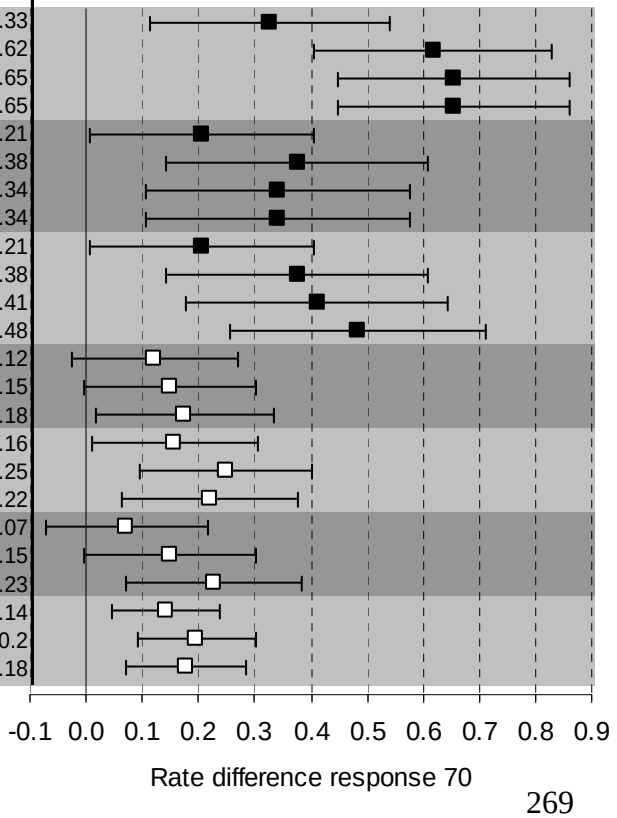


Figure 53. Induction trials.— CDAI scores (mean scores).

STUDY	drug	N	DOSE mg/kg	DOSE wks	WK	MEAN DIFF
CLASSIC I	Ada	148	40/20	0, 2	0	-3
CLASSIC I	Ada	148	40/20	0, 2	1	9
CLASSIC I	Ada	148	40/20	0, 2	2	6
CLASSIC I	Ada	148	40/20	0, 2	4	12
CLASSIC I	Ada	149	80/40	0, 2	0	-5
CLASSIC I	Ada	149	80/40	0, 2	1	17
CLASSIC I	Ada	149	80/40	0, 2	2	26
CLASSIC I	Ada	149	80/40	0, 2	4	30
CLASSIC I	Ada	150	160/80	0, 2	0	1
CLASSIC I	*Ada*	150	160/80	0, 2	1	17
CLASSIC I	Ada	150	160/80	0, 2	2	30
CLASSIC I	Ada	150	160/80	0, 2	4	47
GAIN	Ada	325	160/80	0, 2	0	0
GAIN	Ada	325	160/80	0, 2	1	23
GAIN	Ada	325	160/80	0, 2	2	49
GAIN	Ada	325	160/80	0, 2	4	38
D'Haens 1999	Inflix	15	5	0	0	-37
D'Haens 1999	Inflix	15	5	0	4	139
D'Haens 1999	Inflix	15	10	0	0	-60
D'Haens 1999	Inflix	15	10	0	4	41
D'Haens 1999	Inflix	15	20	0	0	-24
D'Haens 1999	Inflix	15	20	0	4	99
Targan 1997	Inflix	52	5	0	0	-24
Targan 1997	Inflix	52	5	0	2	90
Targan 1997	Inflix	52	5	0	4	105
Targan 1997	Inflix	53	10	0	0	-30
Targan 1997	Inflix	53	10	0	2	34
Targan 1997	Inflix	53	10	0	4	45
Targan 1997	Inflix	53	20	0	0	-19
Targan 1997	Inflix	53	20	0	2	55
Targan 1997	Inflix	53	20	0	4	60

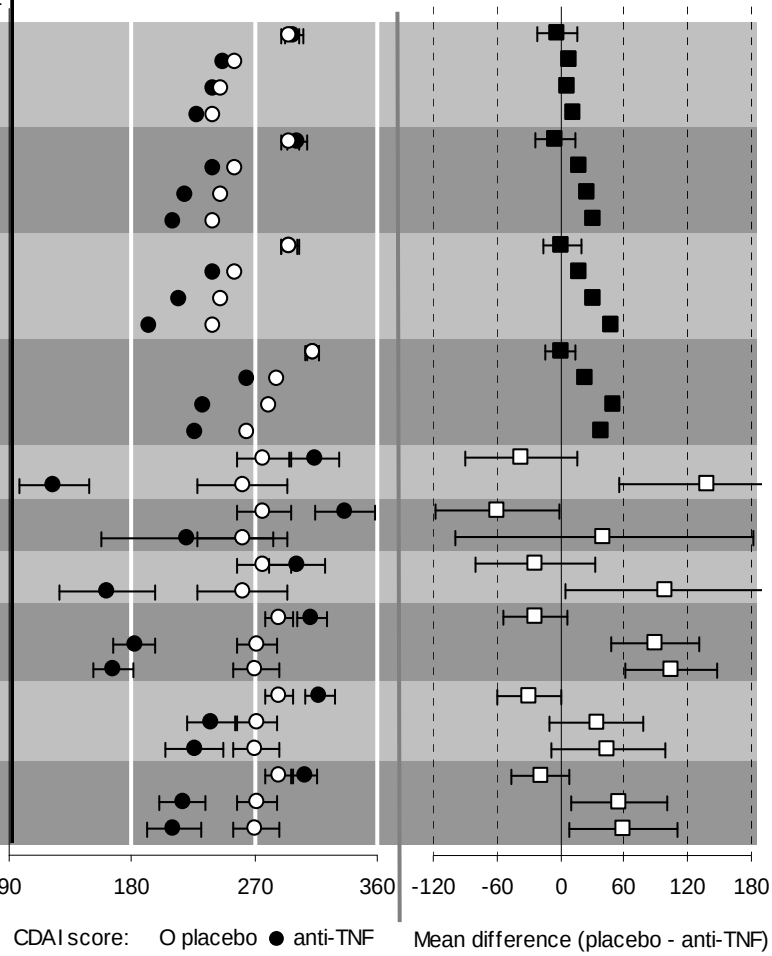
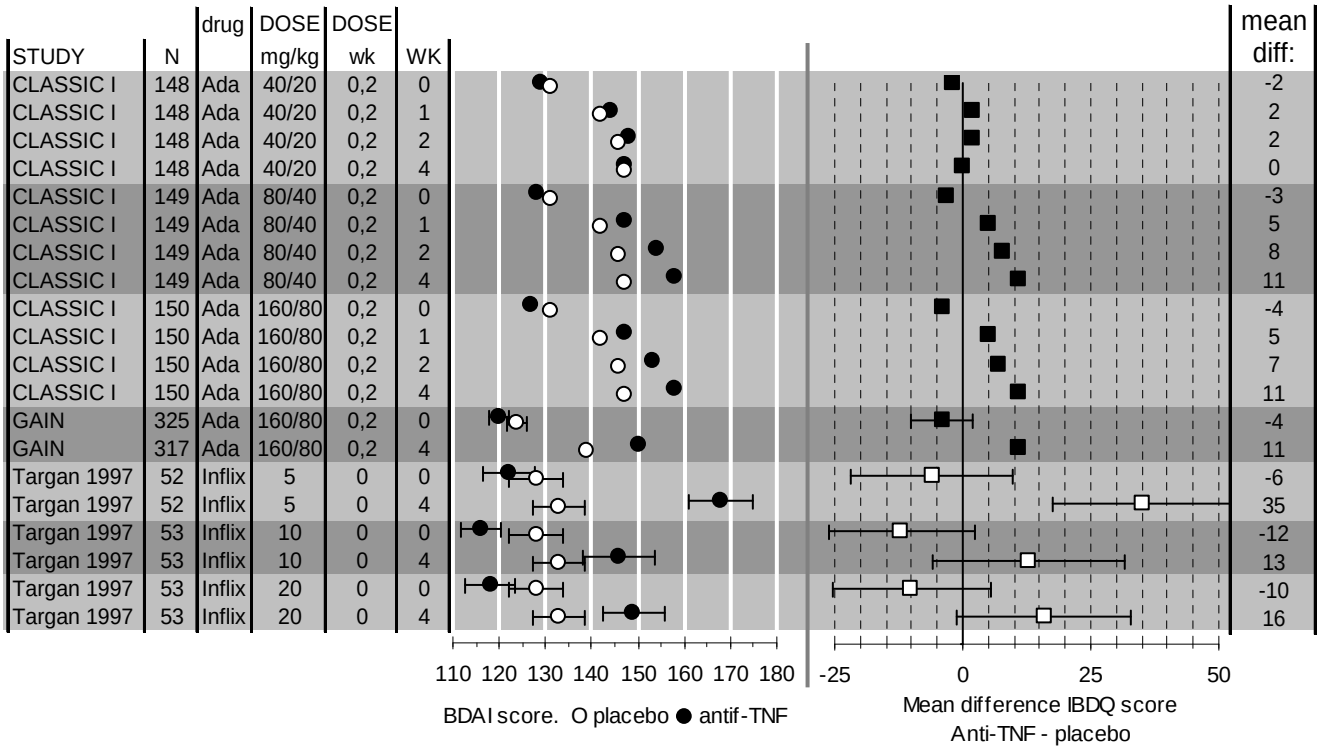


Figure 54. Induction trials.— IBDQ scores



MAINTENANCE TRIALS (unless stated 'dose weeks' refers to post randomisation doses).

Figure 55. Maintenance trials.— rate ratio (RR) remission. (responders)

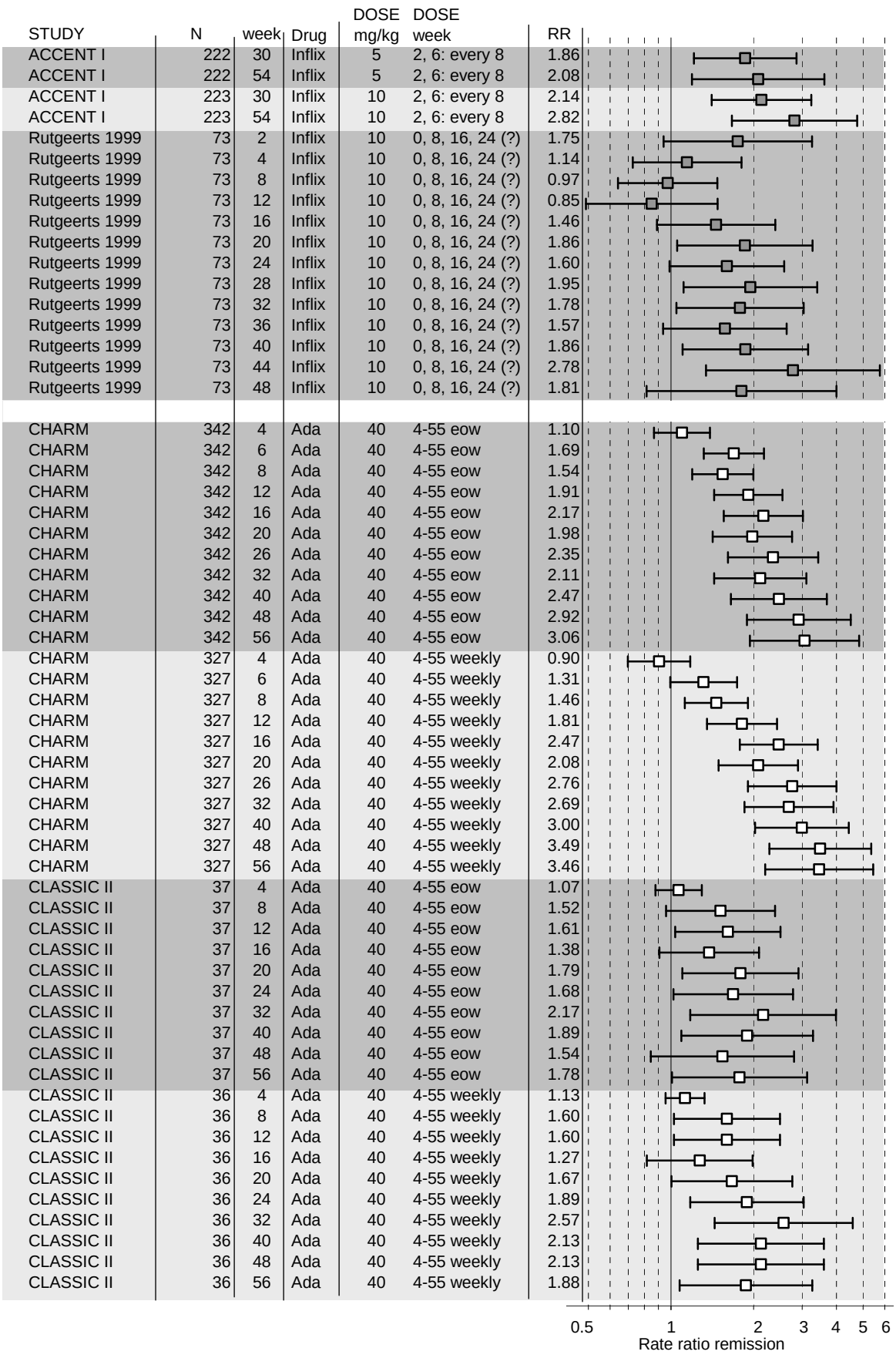


Figure 56. Maintenance trials.— rate ratio (RR) remission. (responders)

STUDY	N	week	Drug	DOSE mg/kg	DOSE week	RD
ACCENT I	222	30	Inflix	5	2, 6: every 8	0.18
ACCENT I	222	54	Inflix	5	2, 6: every 8	0.15
ACCENT I	223	30	Inflix	10	2, 6: every 8	0.24
ACCENT I	223	54	Inflix	10	2, 6: every 8	0.25
Rutgeerts 1999	73	2	Inflix	10	0, 8, 16, 24 (?)	0.21
Rutgeerts 1999	73	4	Inflix	10	0, 8, 16, 24 (?)	0.07
Rutgeerts 1999	73	8	Inflix	10	0, 8, 16, 24 (?)	-0.02
Rutgeerts 1999	73	12	Inflix	10	0, 8, 16, 24 (?)	-0.07
Rutgeerts 1999	73	16	Inflix	10	0, 8, 16, 24 (?)	0.18
Rutgeerts 1999	73	20	Inflix	10	0, 8, 16, 24 (?)	0.26
Rutgeerts 1999	73	24	Inflix	10	0, 8, 16, 24 (?)	0.23
Rutgeerts 1999	73	28	Inflix	10	0, 8, 16, 24 (?)	0.29
Rutgeerts 1999	73	32	Inflix	10	0, 8, 16, 24 (?)	0.26
Rutgeerts 1999	73	36	Inflix	10	0, 8, 16, 24 (?)	0.21
Rutgeerts 1999	73	40	Inflix	10	0, 8, 16, 24 (?)	0.29
Rutgeerts 1999	73	44	Inflix	10	0, 8, 16, 24 (?)	0.35
Rutgeerts 1999	73	48	Inflix	10	0, 8, 16, 24 (?)	0.16
CHARM	327	4	Ada	40	4-55 weekly	-0.04
CHARM	327	6	Ada	40	4-55 weekly	0.10
CHARM	327	8	Ada	40	4-55 weekly	0.16
CHARM	327	12	Ada	40	4-55 weekly	0.22
CHARM	327	16	Ada	40	4-55 weekly	0.31
CHARM	327	20	Ada	40	4-55 weekly	0.23
CHARM	327	26	Ada	40	4-55 weekly	0.30
CHARM	327	32	Ada	40	4-55 weekly	0.29
CHARM	327	40	Ada	40	4-55 weekly	0.31
CHARM	327	48	Ada	40	4-55 weekly	0.32
CHARM	327	56	Ada	40	4-55 weekly	0.29
CHARM	342	4	Ada	40	4-55 eow	0.04
CHARM	342	6	Ada	40	4-55 eow	0.23
CHARM	342	8	Ada	40	4-55 eow	0.18
CHARM	342	12	Ada	40	4-55 eow	0.25
CHARM	342	16	Ada	40	4-55 eow	0.25
CHARM	342	20	Ada	40	4-55 eow	0.21
CHARM	342	26	Ada	40	4-55 eow	0.23
CHARM	342	32	Ada	40	4-55 eow	0.19
CHARM	342	40	Ada	40	4-55 eow	0.22
CHARM	342	48	Ada	40	4-55 eow	0.25
CHARM	342	56	Ada	40	4-55 eow	0.24
CLASSIC II	36	4	Ada	40	4-55 weekly	0.11
CLASSIC II	36	8	Ada	40	4-55 weekly	0.33
CLASSIC II	36	12	Ada	40	4-55 weekly	0.33
CLASSIC II	36	16	Ada	40	4-55 weekly	0.17
CLASSIC II	36	20	Ada	40	4-55 weekly	0.33
CLASSIC II	36	24	Ada	40	4-55 weekly	0.44
CLASSIC II	36	32	Ada	40	4-55 weekly	0.61
CLASSIC II	36	40	Ada	40	4-55 weekly	0.50
CLASSIC II	36	48	Ada	40	4-55 weekly	0.50
CLASSIC II	36	56	Ada	40	4-55 weekly	0.39
CLASSIC II	37	4	Ada	40	4-55 eow	0.06
CLASSIC II	37	8	Ada	40	4-55 eow	0.29
CLASSIC II	37	12	Ada	40	4-55 eow	0.34
CLASSIC II	37	16	Ada	40	4-55 eow	0.23
CLASSIC II	37	20	Ada	40	4-55 eow	0.39
CLASSIC II	37	24	Ada	40	4-55 eow	0.34
CLASSIC II	37	32	Ada	40	4-55 eow	0.45
CLASSIC II	37	40	Ada	40	4-55 eow	0.40
CLASSIC II	37	48	Ada	40	4-55 eow	0.24
CLASSIC II	37	56	Ada	40	4-55 eow	0.35

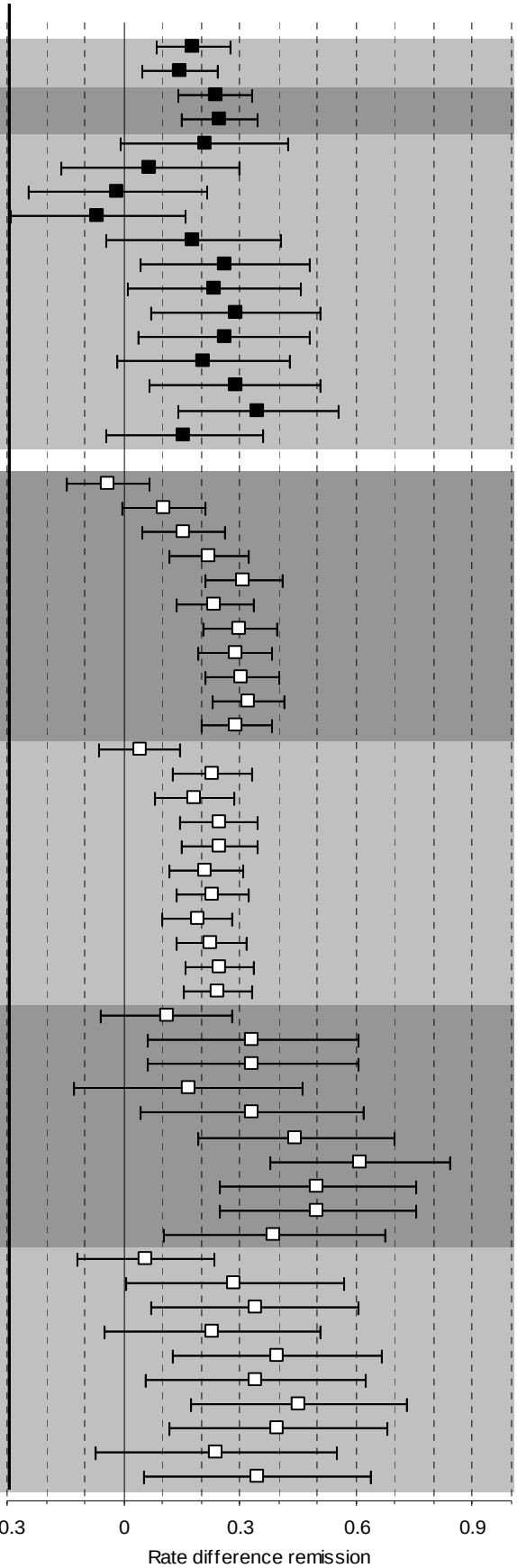


Figure 57. Maintenance trials .— rate ratio (RR) remission, all patients.

STUDY	N	week	Drug	DOSE mg/kg	DOSE week	RR
ACCENT I	380	2	Inflix	5	2, 6: every 8	0.94
ACCENT I	380	6	Inflix	5	2, 6: every 8	1.28
ACCENT I	380	10	Inflix	5	2, 6: every 8	1.29
ACCENT I	380	14	Inflix	5	2, 6: every 8	1.47
ACCENT I	380	22	Inflix	5	2, 6: every 8	1.35
ACCENT I	380	30	Inflix	5	2, 6: every 8	1.20
ACCENT I	380	38	Inflix	5	2, 6: every 8	1.10
ACCENT I	380	46	Inflix	5	2, 6: every 8	1.17
ACCENT I	380	54	Inflix	5	2, 6: every 8	1.14
ACCENT I	381	2	Inflix	10	2, 6: every 8	1.05
ACCENT I	381	6	Inflix	10	2, 6: every 8	1.21
ACCENT I	381	10	Inflix	10	2, 6: every 8	1.35
ACCENT I	381	14	Inflix	10	2, 6: every 8	1.52
ACCENT I	381	22	Inflix	10	2, 6: every 8	1.62
ACCENT I	381	30	Inflix	10	2, 6: every 8	1.28
ACCENT I	381	38	Inflix	10	2, 6: every 8	1.28
ACCENT I	381	46	Inflix	10	2, 6: every 8	1.44
ACCENT I	381	54	Inflix	10	2, 6: every 8	1.27

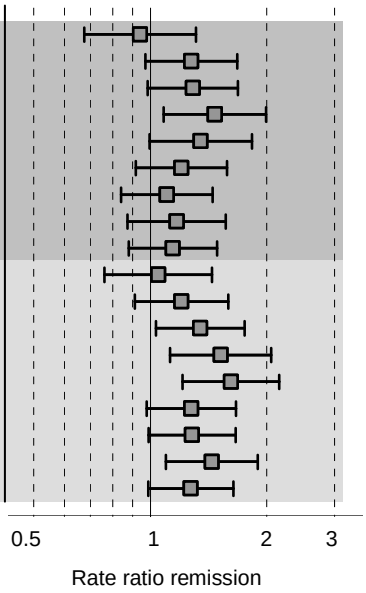


Figure 58. Maintenance trials .— rate difference (RD) remission, all patients.

STUDY	N	week	Drug	DOSE mg/kg	DOSE week	RD
ACCENT I	380	2	Inflix	5	2, 6: every 8	-0.02
ACCENT I	380	6	Inflix	5	2, 6: every 8	0.09
ACCENT I	380	10	Inflix	5	2, 6: every 8	0.09
ACCENT I	380	14	Inflix	5	2, 6: every 8	0.12
ACCENT I	380	22	Inflix	5	2, 6: every 8	0.09
ACCENT I	380	30	Inflix	5	2, 6: every 8	0.07
ACCENT I	380	38	Inflix	5	2, 6: every 8	0.03
ACCENT I	380	46	Inflix	5	2, 6: every 8	0.05
ACCENT I	380	54	Inflix	5	2, 6: every 8	0.05
ACCENT I	381	2	Inflix	10	2, 6: every 8	0.01
ACCENT I	381	6	Inflix	10	2, 6: every 8	0.06
ACCENT I	381	10	Inflix	10	2, 6: every 8	0.11
ACCENT I	381	14	Inflix	10	2, 6: every 8	0.13
ACCENT I	381	22	Inflix	10	2, 6: every 8	0.16
ACCENT I	381	30	Inflix	10	2, 6: every 8	0.09
ACCENT I	381	38	Inflix	10	2, 6: every 8	0.09
ACCENT I	381	46	Inflix	10	2, 6: every 8	0.13
ACCENT I	381	54	Inflix	10	2, 6: every 8	0.09

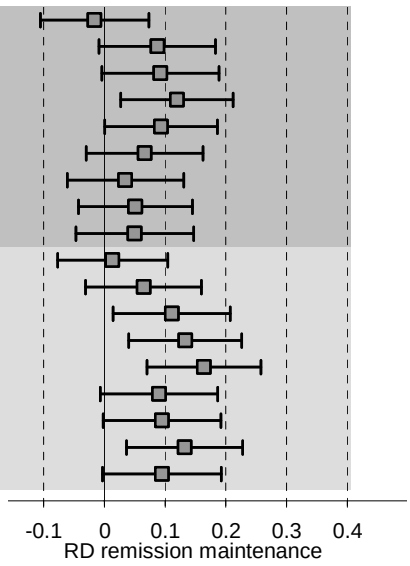


Figure 59. Maintenance trials.— rate ratio (RR) response 100

STUDY	N	week	Drug	DOSE mg/kg	DOSE week	RR
CHARM	327	26	Ada	40	4-55 weekly	1.97
CHARM	327	56	Ada	40	4-55 weekly	2.90
CHARM	342	26	Ada	40	4-55 eow	1.95
CHARM	342	56	Ada	40	4-55 eow	2.51
CLASSIC II	36	4	Ada	40	4-55 weekly	1.13
CLASSIC II	36	8	Ada	40	4-55 weekly	1.33
CLASSIC II	36	12	Ada	40	4-55 weekly	1.70
CLASSIC II	36	16	Ada	40	4-55 weekly	1.23
CLASSIC II	36	20	Ada	40	4-55 weekly	1.36
CLASSIC II	36	24	Ada	40	4-55 weekly	1.55
CLASSIC II	36	32	Ada	40	4-55 weekly	2.00
CLASSIC II	36	40	Ada	40	4-55 weekly	1.70
CLASSIC II	36	48	Ada	40	4-55 weekly	1.50
CLASSIC II	36	56	Ada	40	4-55 weekly	1.60
CLASSIC II	37	4	Ada	40	4-55 eow	1.01
CLASSIC II	37	8	Ada	40	4-55 eow	1.18
CLASSIC II	37	12	Ada	40	4-55 eow	1.71
CLASSIC II	37	16	Ada	40	4-55 eow	1.24
CLASSIC II	37	20	Ada	40	4-55 eow	1.55
CLASSIC II	37	24	Ada	40	4-55 eow	1.38
CLASSIC II	37	32	Ada	40	4-55 eow	1.58
CLASSIC II	37	40	Ada	40	4-55 eow	1.52
CLASSIC II	37	48	Ada	40	4-55 eow	1.33
CLASSIC II	37	56	Ada	40	4-55 eow	1.42

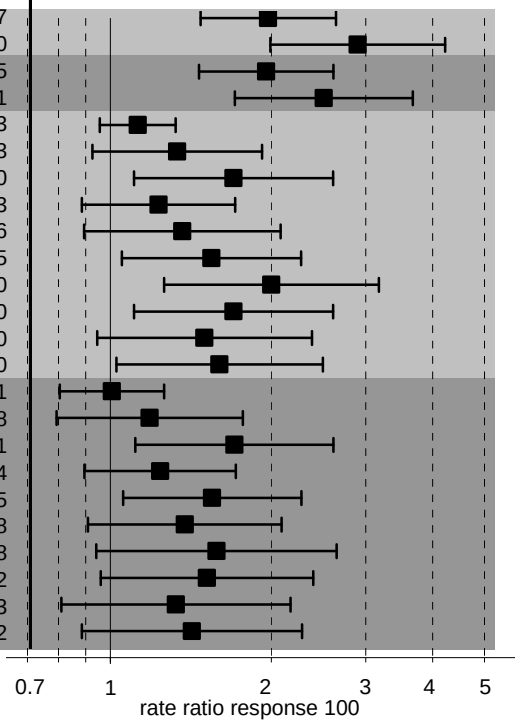


Figure 60. Maintenance trials.— rate difference response 100

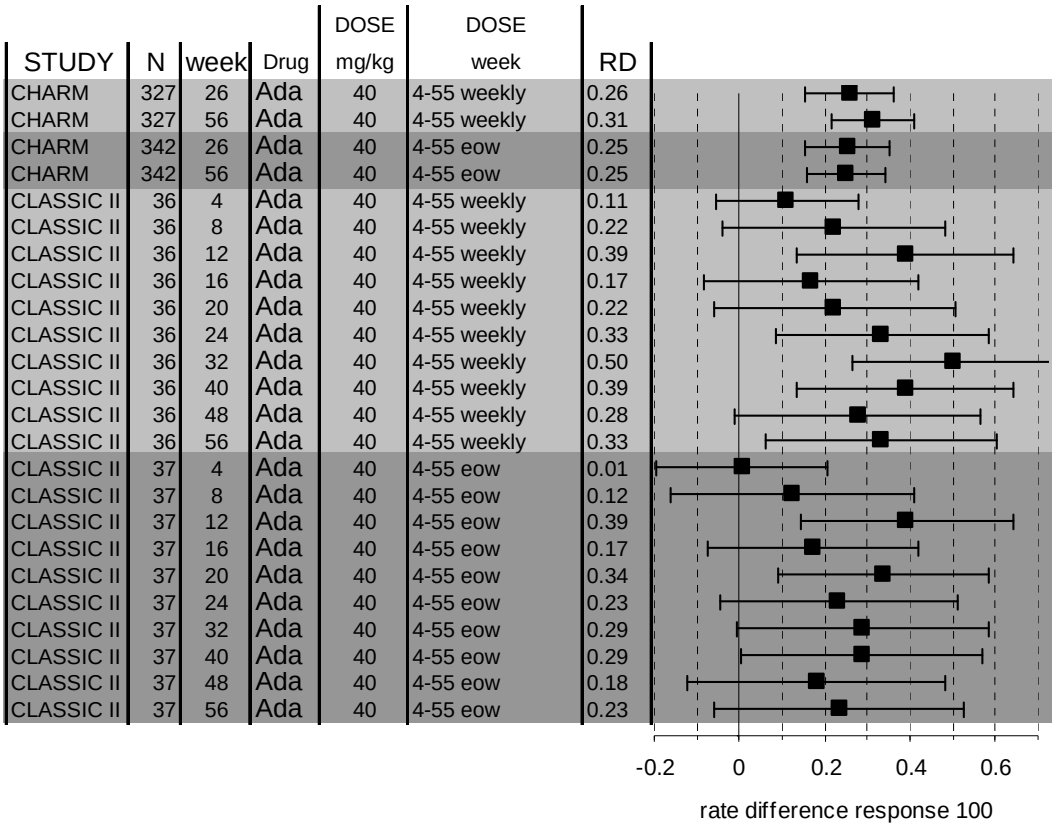


Figure 61. Maintenance trials.— rate ratio (RR) response 70 [responders].

("dose weeks" for Rutgeerts 1999 refers to all scheduled dose weeks including those prior to randomisation).

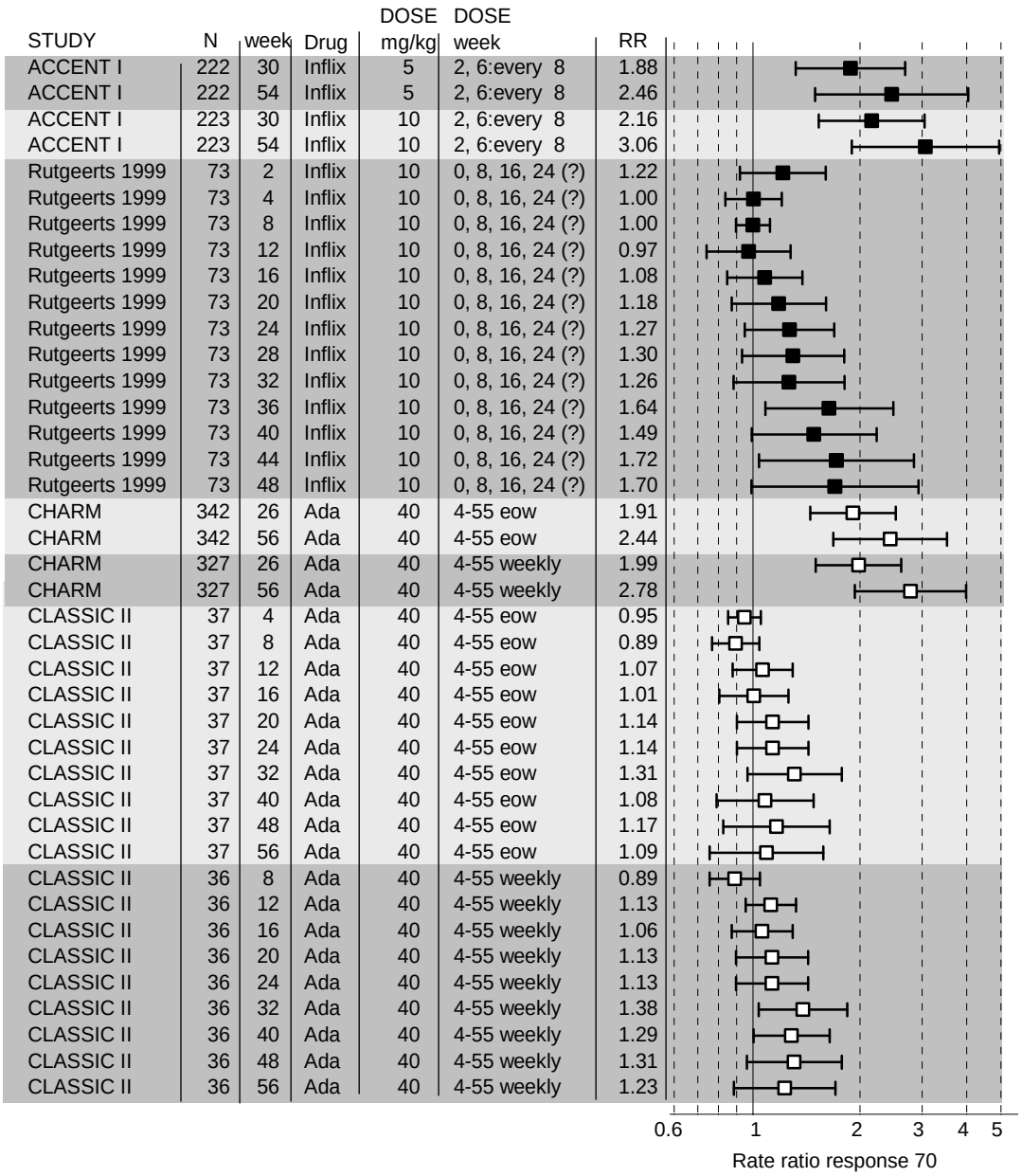


Figure 62. Maintenance trials.— rate difference (RD) response 70 [responders].

("dose weeks" for Rutgeerts 1999 refers to all scheduled dose weeks including those prior to randomisation).

STUDY	N	week	Drug	DOSE mg/kg	DOSE week	RD
ACCENT I	222	30	Inflix	5	2, 6:every 8	0.24
ACCENT I	222	54	Inflix	5	2, 6:every 8	0.23
ACCENT I	223	30	Inflix	10	2, 6:every 8	0.32
ACCENT I	223	54	Inflix	10	2, 6:every 8	0.32
Rutgeerts 1999	73	2	Inflix	10	0, 8, 16, 24 (?)	0.14
Rutgeerts 1999	73	4	Inflix	10	0, 8, 16, 24 (?)	0.00
Rutgeerts 1999	73	8	Inflix	10	0, 8, 16, 24 (?)	0.00
Rutgeerts 1999	73	12	Inflix	10	0, 8, 16, 24 (?)	-0.02
Rutgeerts 1999	73	16	Inflix	10	0, 8, 16, 24 (?)	0.06
Rutgeerts 1999	73	20	Inflix	10	0, 8, 16, 24 (?)	0.12
Rutgeerts 1999	73	24	Inflix	10	0, 8, 16, 24 (?)	0.17
Rutgeerts 1999	73	28	Inflix	10	0, 8, 16, 24 (?)	0.17
Rutgeerts 1999	73	32	Inflix	10	0, 8, 16, 24 (?)	0.15
Rutgeerts 1999	73	36	Inflix	10	0, 8, 16, 24 (?)	0.29
Rutgeerts 1999	73	40	Inflix	10	0, 8, 16, 24 (?)	0.23
Rutgeerts 1999	73	44	Inflix	10	0, 8, 16, 24 (?)	0.26
Rutgeerts 1999	73	48	Inflix	10	0, 8, 16, 24 (?)	0.23
CHARM	342	26	Ada	40	4-55 eow	0.26
CHARM	342	56	Ada	40	4-55 eow	0.25
CHARM	327	26	Ada	40	4-55 weekly	0.28
CHARM	327	56	Ada	40	4-55 weekly	0.31
CLASSIC II	37	4	Ada	40	4-55 eow	-0.05
CLASSIC II	37	8	Ada	40	4-55 eow	-0.11
CLASSIC II	37	12	Ada	40	4-55 eow	0.06
CLASSIC II	37	16	Ada	40	4-55 eow	0.01
CLASSIC II	37	20	Ada	40	4-55 eow	0.11
CLASSIC II	37	24	Ada	40	4-55 eow	0.11
CLASSIC II	37	32	Ada	40	4-55 eow	0.23
CLASSIC II	37	40	Ada	40	4-55 eow	0.06
CLASSIC II	37	48	Ada	40	4-55 eow	0.12
CLASSIC II	37	56	Ada	40	4-55 eow	0.07
CLASSIC II	36	4	Ada	40	4-55 weekly	0.00
CLASSIC II	36	8	Ada	40	4-55 weekly	-0.11
CLASSIC II	36	12	Ada	40	4-55 weekly	0.11
CLASSIC II	36	16	Ada	40	4-55 weekly	0.06
CLASSIC II	36	20	Ada	40	4-55 weekly	0.11
CLASSIC II	36	24	Ada	40	4-55 weekly	0.11
CLASSIC II	36	32	Ada	40	4-55 weekly	0.28
CLASSIC II	36	40	Ada	40	4-55 weekly	0.22
CLASSIC II	36	48	Ada	40	4-55 weekly	0.22
CLASSIC II	36	56	Ada	40	4-55 weekly	0.17

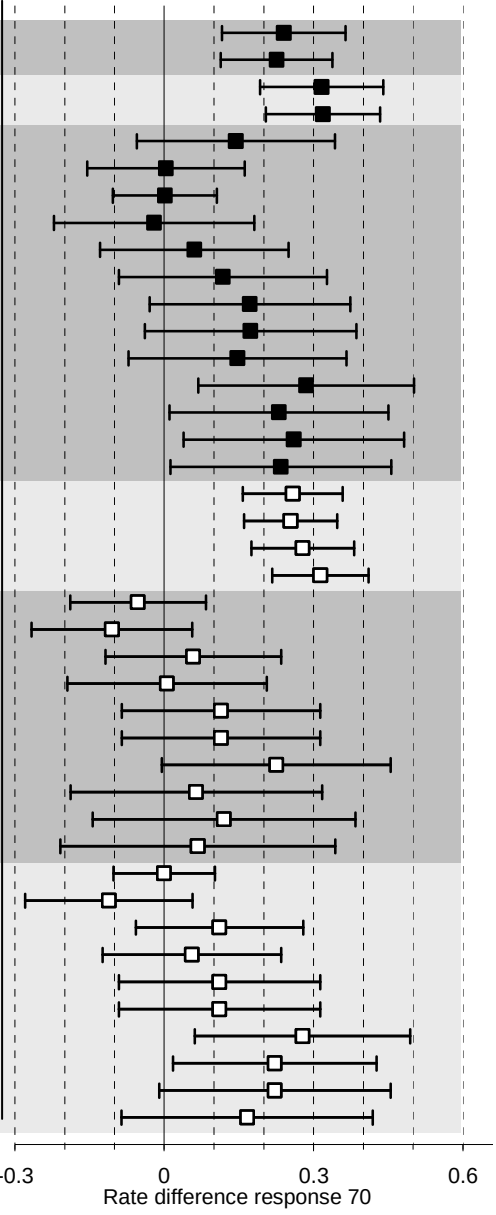


Figure 63. Maintenance trials.— rate ratio (RR) response 70 all patients

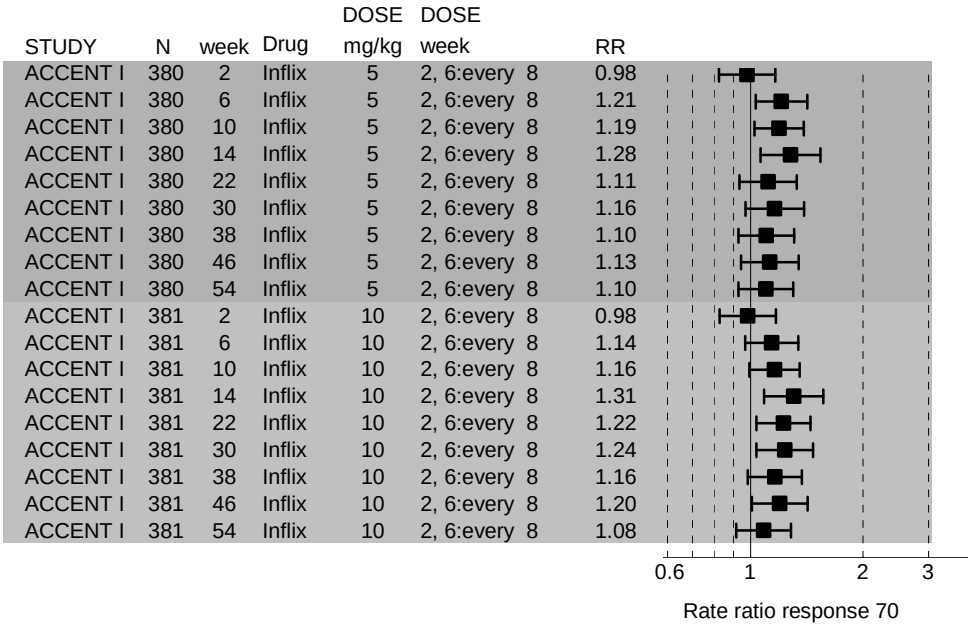


Figure 64. Maintenance trials.— rate difference (RD) response 70 all patients

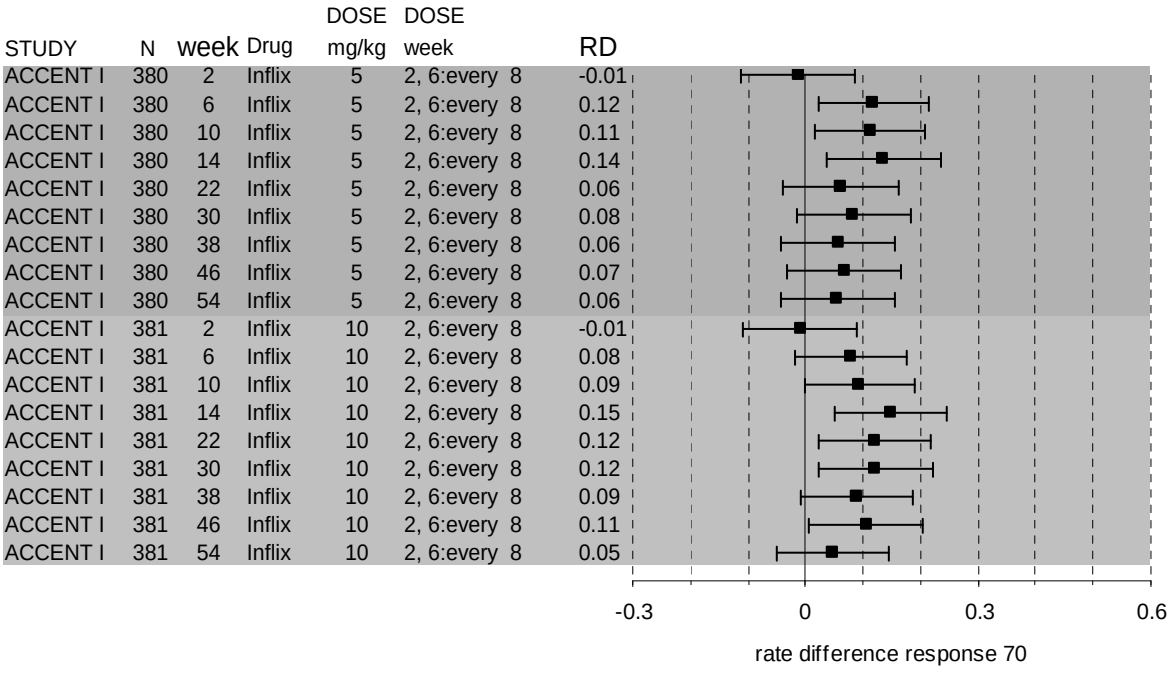


Figure 65. Maintenance trials.— CDAI scores for trials reporting median scores (Results for responders).

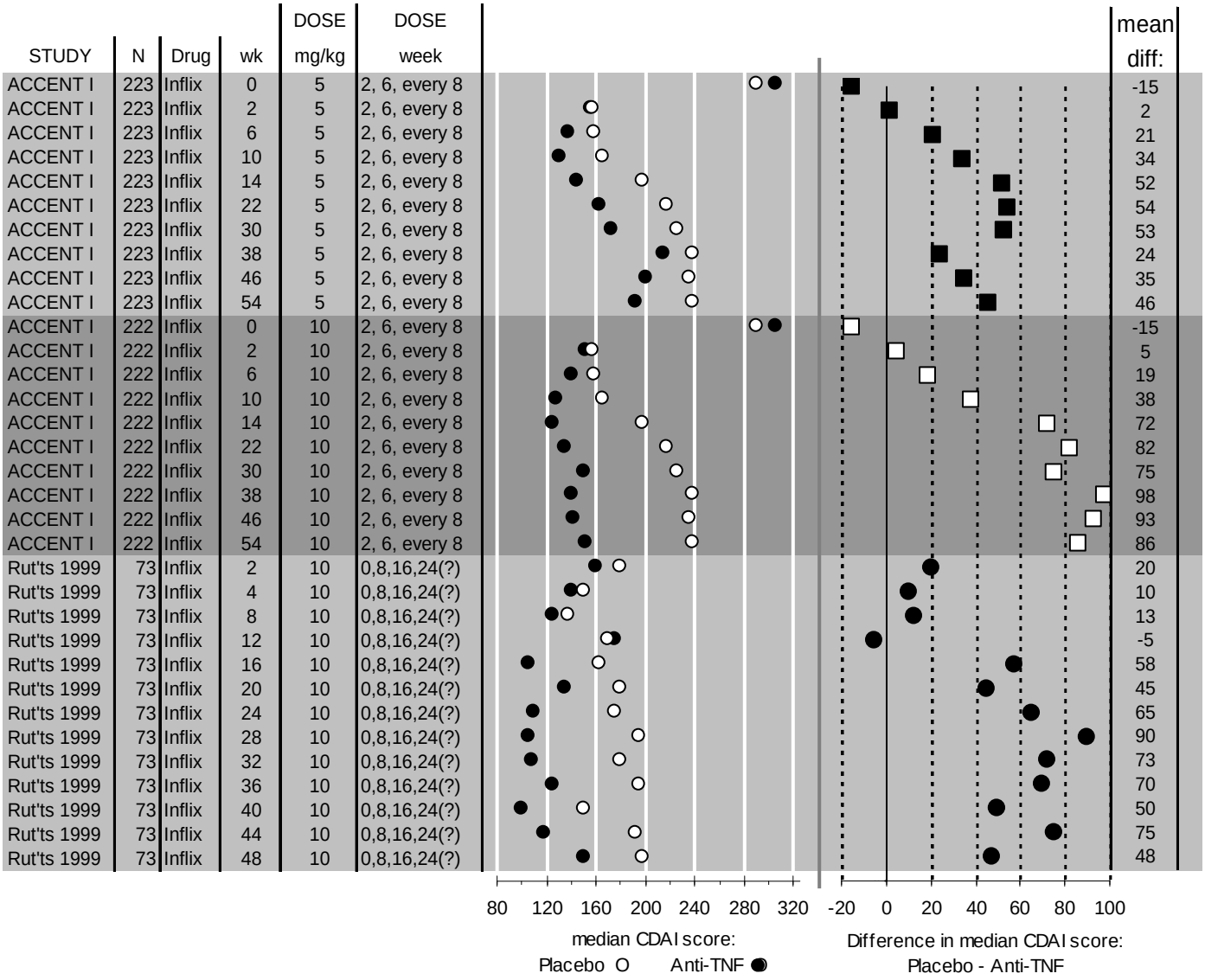


Figure 66. Maintenance trials.— CDAI scores for trials reporting median scores (Results for all patients).

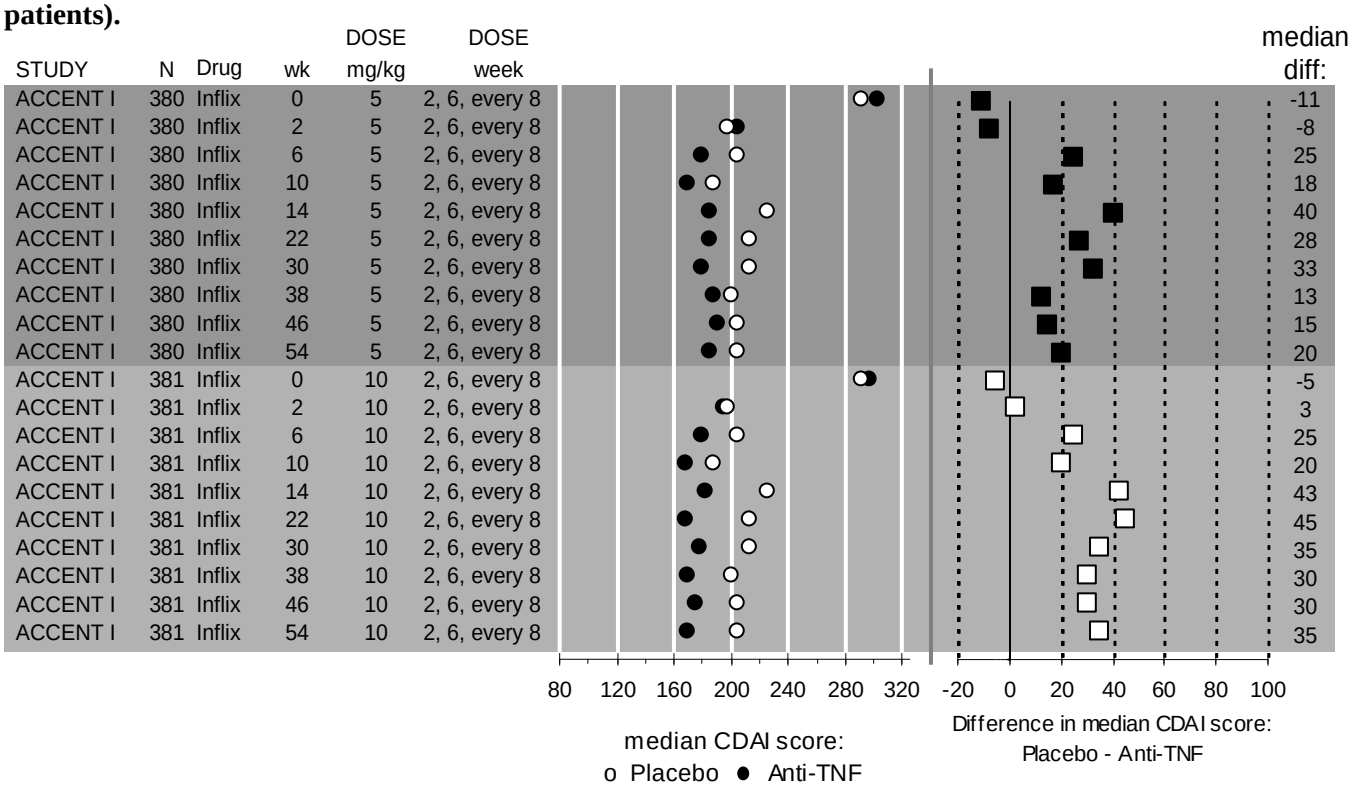


Figure 67. Maintenance trials.— CDAI scores for trials reporting mean scores

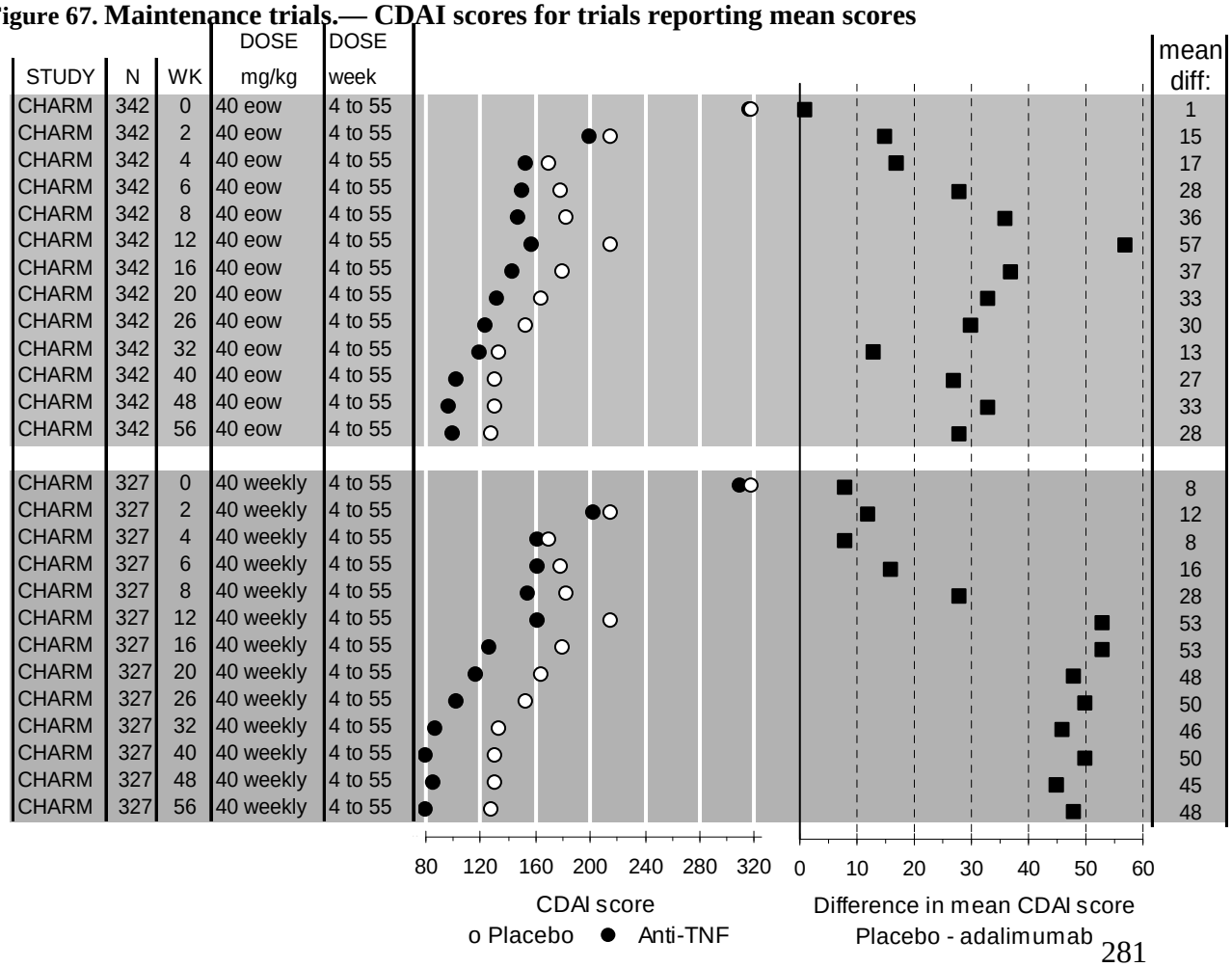


Figure 68. Maintenance trials.— IBDQ scores for trials reporting mean scores

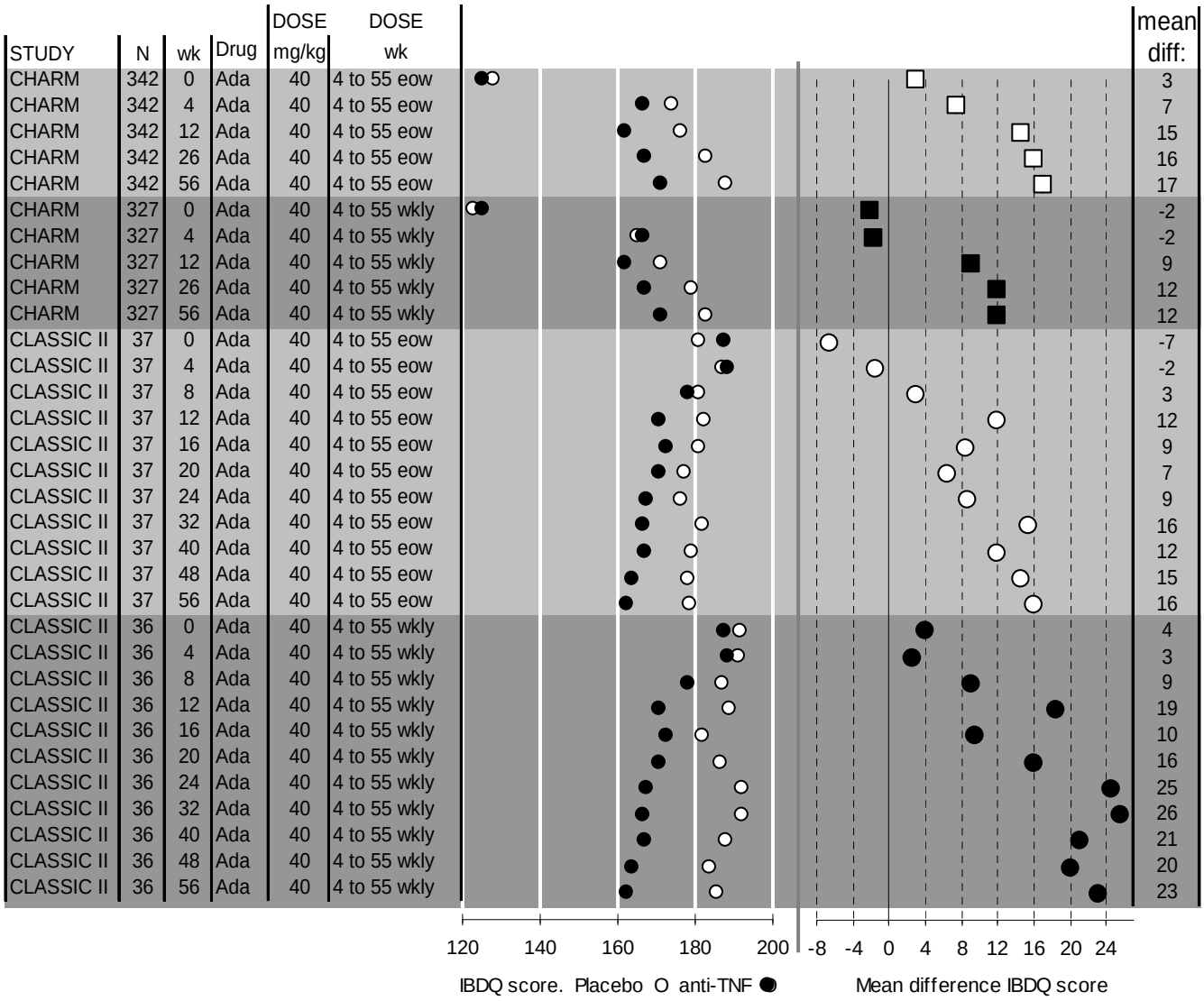


Figure 69. Maintenance trials.— IBDQ scores for trials reporting median scores.

(dose weeks refers to all scheduled doses including those prior to randomisation).

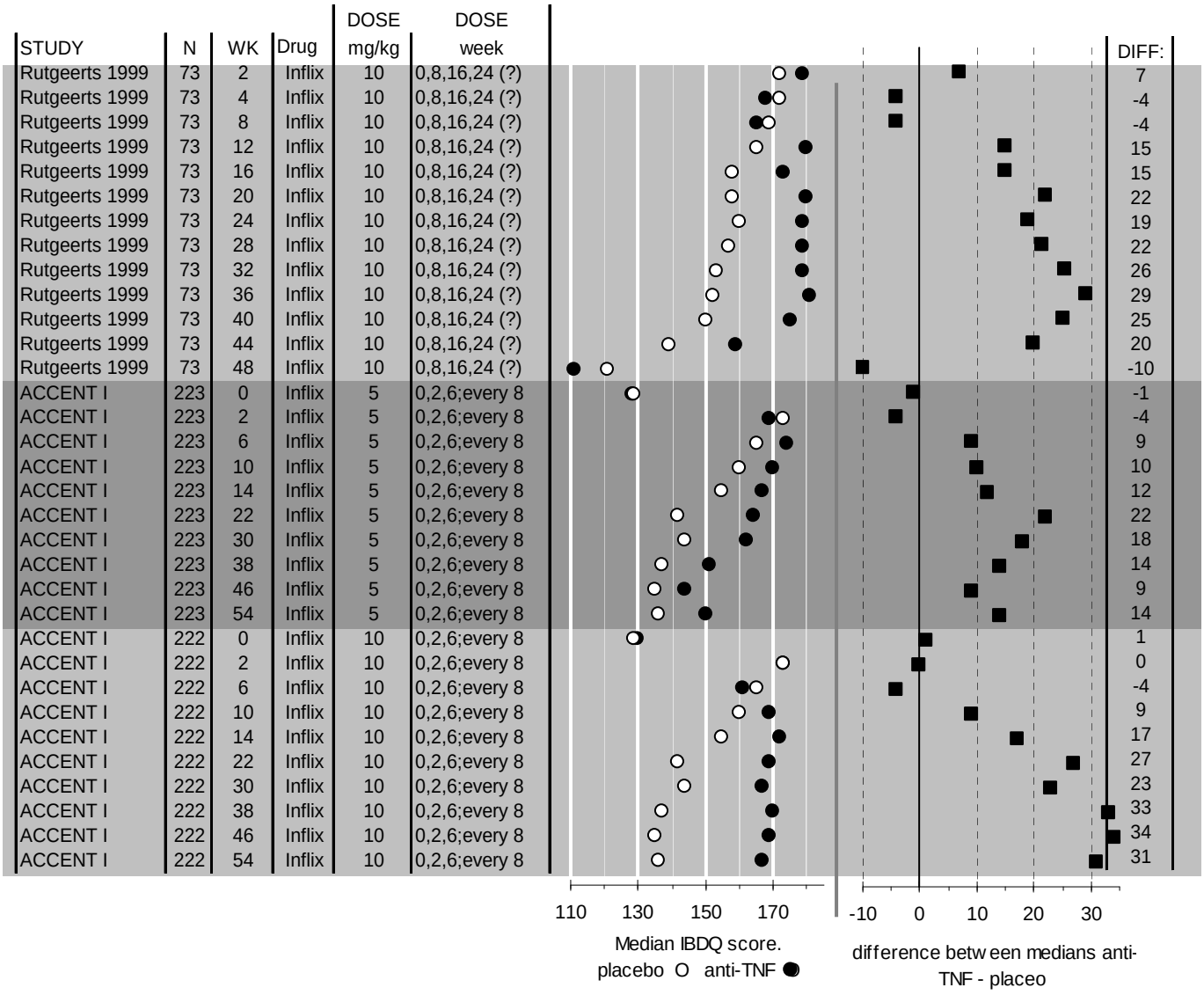
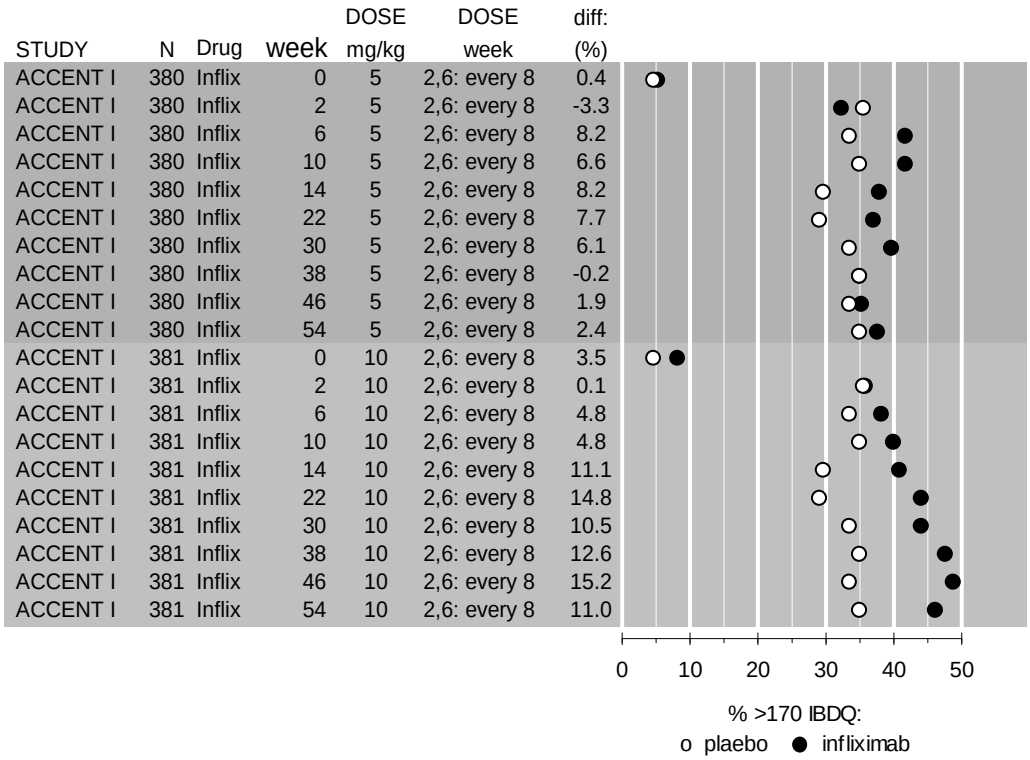


Figure 70. Maintenance trials.— IBDQ scores for trials reporting % of patients with IBDQ score more than 170

(taken as an indicator of remission. The results below refer to **all** the patients in the ACCENT I trial (not just "responder" patients).



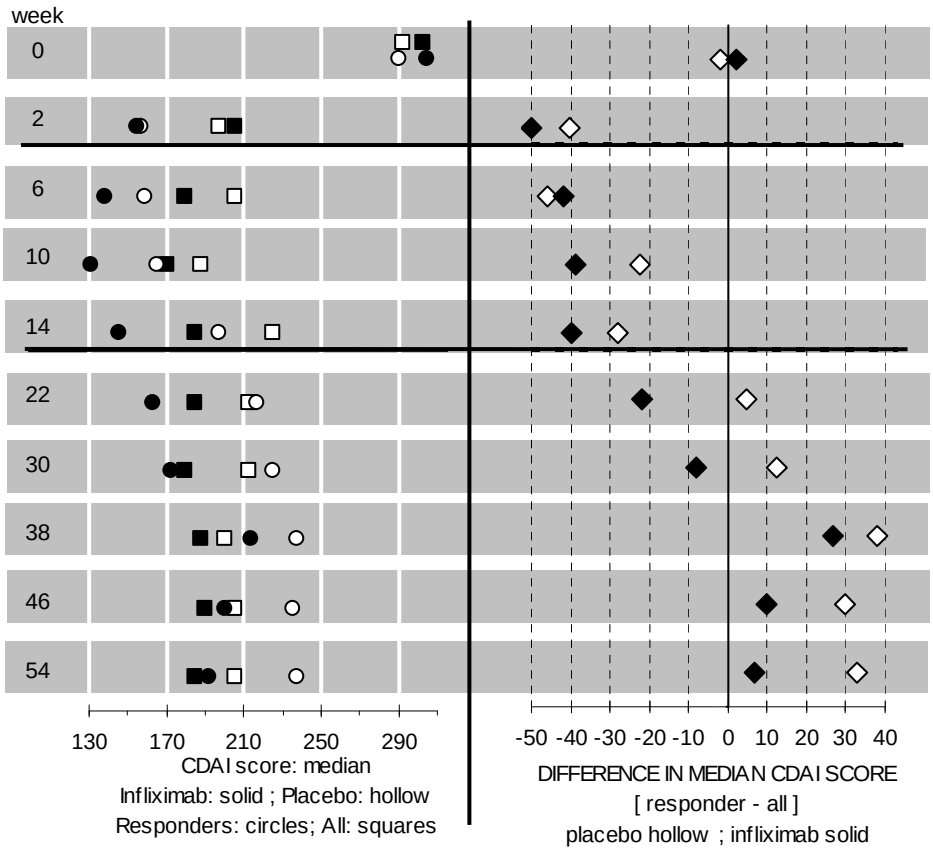
Appendix 11. Response rates amongst non-responders in maintenance trials

In this appendix, results for non-responders in the ACCENT I[3,4] trial (infliximab) are presented, followed by results for non-responders and for all patients in the CHARM trial (adalimumab).

ACCENT I[3,4] trial CDAI scores.

Examination of median CDAI scores in the two separate publications allowed an approximation of the response to treatment in "non-responders" at least to week 14, after which cross over to increased infliximab dosage was allowed for relapsing patients. The pertinent results for median CDAI scores are summarised in Figure 71.

Figure 71. Median CDAI scores in ACCENT I[3,4] trial and score difference between placebo and intervention



Placebo scores are hollow symbols, infliximab scores solid. Responders are represented by circles all patients by squares.

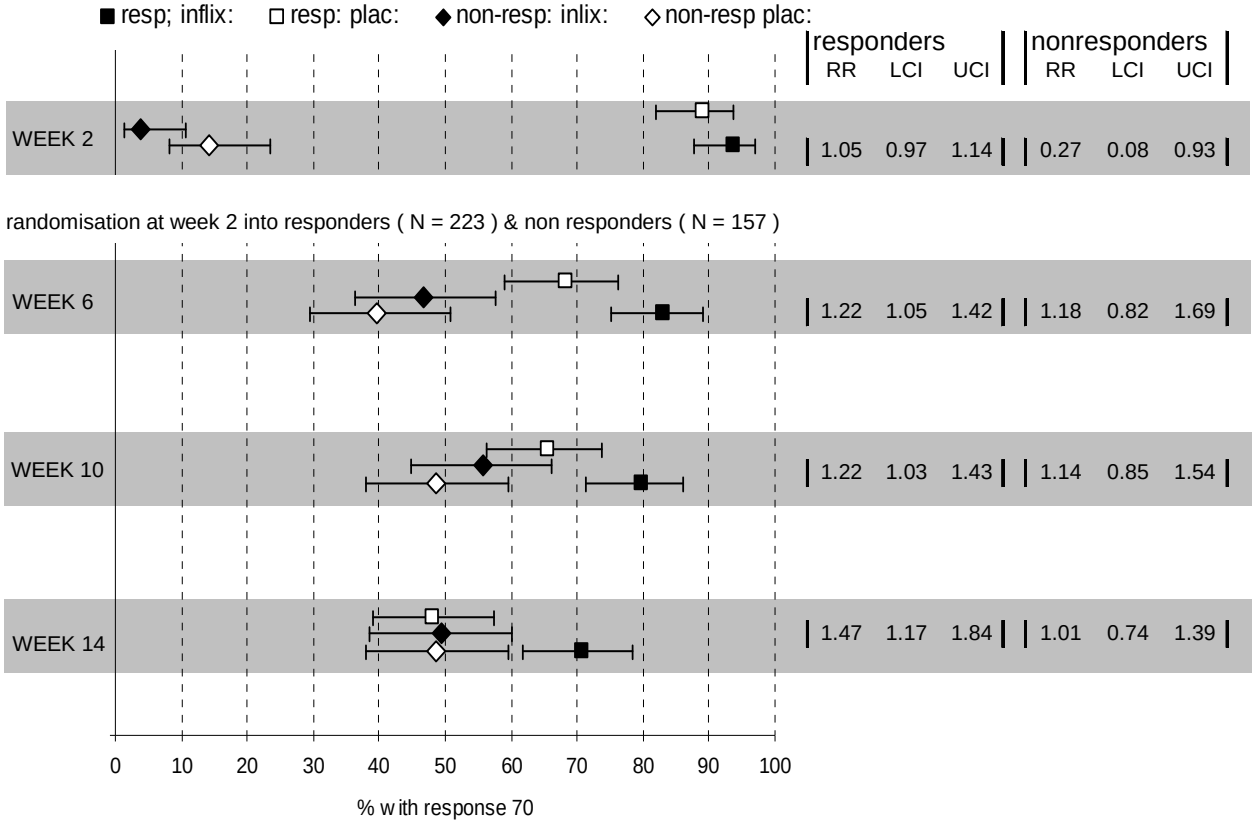
At randomisation (week 2) the difference in CDAI score of "responders" minus "all" was at its maximum, 50 points for infliximab and 40 for placebo respectively, and was determined by patient selection. After randomisation up to week 14 the difference for the infliximab treated patients (responders – all) remained fairly stable (at approximately 40 points) implying that during this phase of the trial "responders" and "non-responders" fare about equally well with respect to their CDAI

score at randomisation. After the introduction of permitted cross over for the "all patient" analysis at week 14 both infliximab and placebo treated groups exhibited striking increases in the score difference "responders" – "all". Since increase CDAI implies worse disease state, this trend implies non-responders were able to respond to treatment better than responders during this phase of the trial.

ACCENT I[3,4] remission and responder 70 rates.

Figure 72 shows the placebo and intervention response 70 rates for responders and calculated for non-responders. At week two a very large difference was evident as would be expected from the act of dichotomising patients into subgroups. Thereafter to week 14 response rates in both the intervention arms and the placebo arms gradually approached more closely. For non-responders there was only weak evidence that intervention was better than placebo. At week 14 both placebo groups and the non-responder intervention arm had a response rate of about 50%. The major difference between the responder and non-responder subgroups appeared to be the much larger proportion of placebo responders in the non-responder group; or conversely amongst the responder group there was a greater proportion of patients that required early doses of infliximab to achieve response. Unfortunately information beyond 14 weeks was not available except for responders.

Figure 72. Response 70 rates and rate ratios for responders and non-responders in ACCENT I[3,4] trial

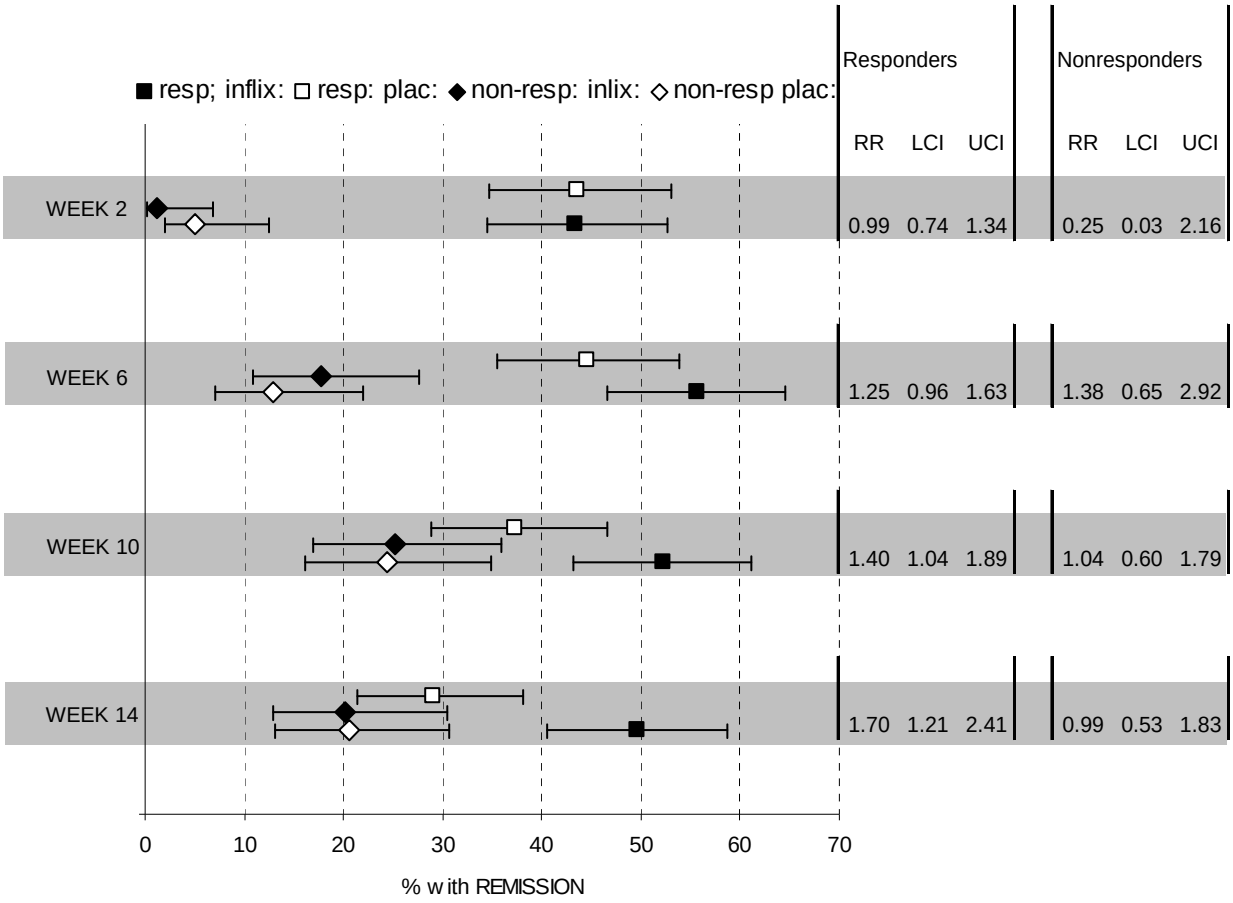


The 95% CI for rate ratio corrected from industry submission.

Similar results were seen for remission (Figure 73). At week 14, 20% of non-responders had attained remission irrespective of treatment. At week 14, the responder intervention arm exhibited 20% more patients in remission than the responder placebo group; thereafter this difference diminished. Unfortunately no information for non-responders was available beyond 14 weeks.

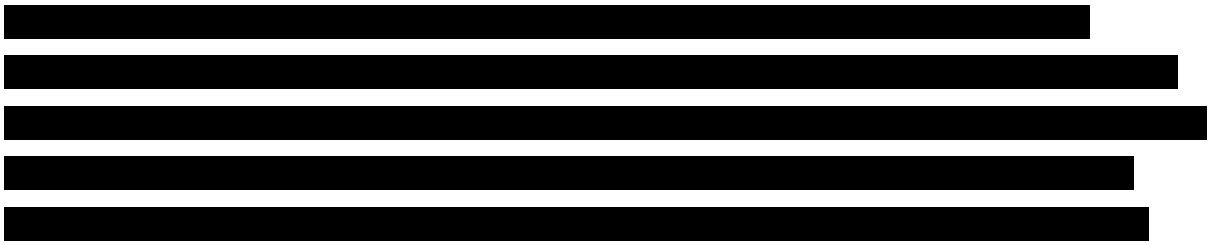
At week 14 the yield in percentage of patients with response 70 per dose, and of percentage of patients with remission per dose for strategies in which all patients received one dose, or all patients received three doses, or responders received three doses and non-responders a single dose was 49%, 21% and 25% for response 70, and 25%, 13%, and 17% for remission.

Figure 73. Remission rates and rate ratios for responders and non-responders in ACCENT I[3,4] trial



The 95% CI for rate ratio corrected from industry submission.

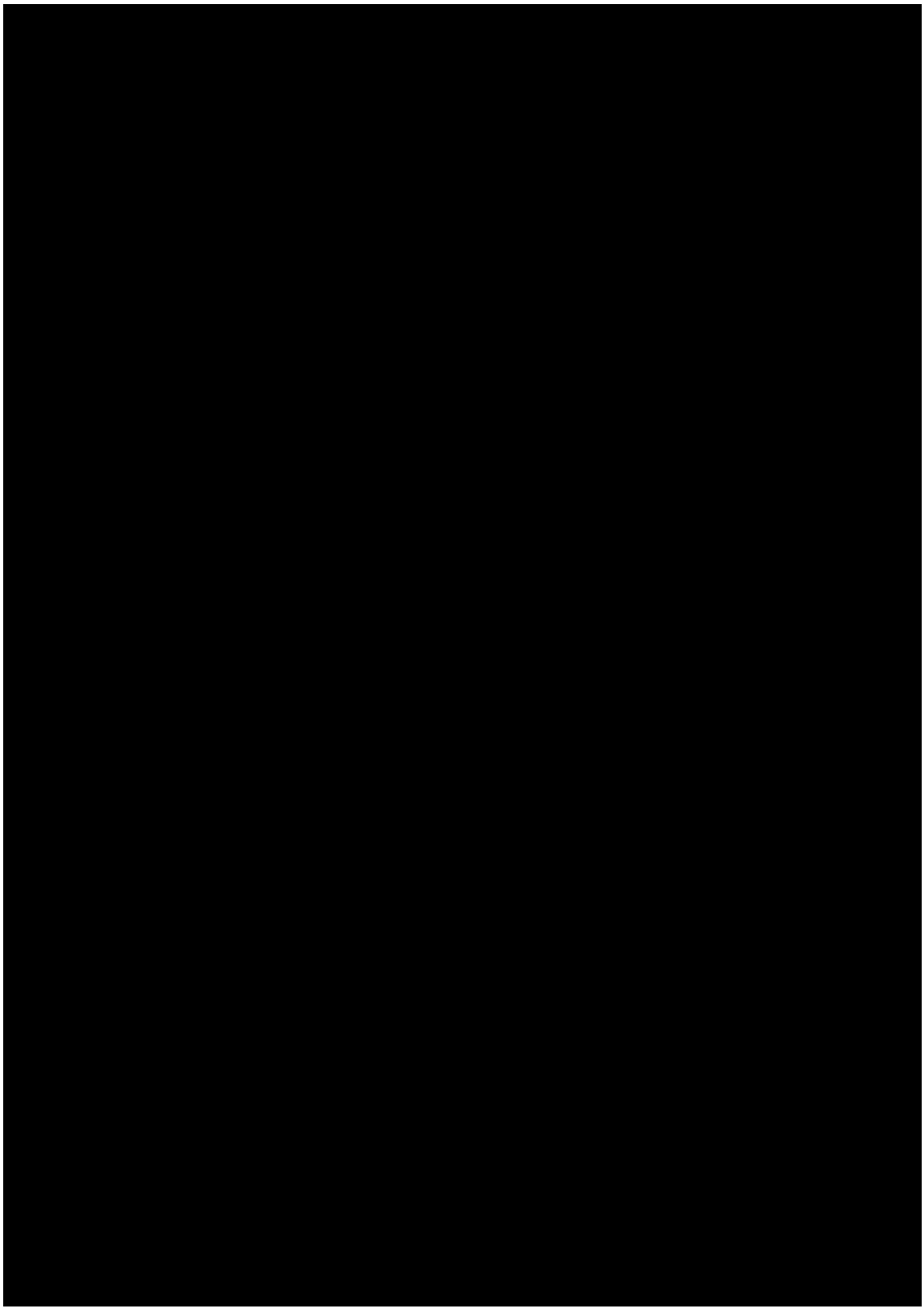
CHARM trial remission and response 70 and 100 rates.



[redacted]

Figure 74. Remission rate ratios in CHARM by patient group, dose regimen and trial week

Figure 75. Response 70 & 100 rate ratios; CHARM trial by patient group, dose regimen and trial week



Appendix 12. Quality assessment of trials

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
INDUCTION				
Hanauer 2006[58] CLASSIC I (Adalimumab)	4 week multi-centre, randomised, double-blind, placebo-controlled trial; 299 patients randomised to placebo, or 40/20mg (week 0 and 2), 80/40mg or 160/80mg adalimumab	• Placebo identical in appearance to adalimumab • Pharmacist preparing injections blinded • Patients blinded • Study investigators/coordinators blinded	All 299 patients included in efficacy analyses. Those with missing data at week 4 classified as remission failures (assume that also counted as response failures but not explicitly stated). Also states that "all analyses were as observed with the exception of the IBDQ data that assessed the last observation carried forward". (unclear what "as observed" means)	284/299 (95%) patients completed the trial; remaining patients withdrew. No loss to follow-up. Unclear how many patients contributing to each analysis.
Sandborn 2007[59] GAIN (Adalimumab)	4 week multi-centre, randomised, double-blind, placebo-controlled trial; 325 patients randomised to placebo or 160/80mg adalimumab	• No details on placebo • Patients blinded • Study investigators and data analysts blinded • Study site and Abbott Laboratories personnel blinded	For clinical remission and response measures, all patients included: considered patients with missing data to be non-responders. For continuous variables included only those patients with complete data.	No loss to follow-up; 14/325 (4%) discontinued intervention or placebo; unclear how many patients were counted as non-responders due to missing data or how many did not contribute to continuous outcome data
D'Haens 1999[63] (Infliximab)	4 week multi-centre, randomised, double-blind, placebo-controlled trial; 30 patients randomised to placebo, 5, 10, or 20mg/kg infliximab	• Placebo identical in appearance to infliximab solution • Patients blinded • Study investigators/personnel blinded • Pathologist assessing biopsy specimens blinded	Unclear if missing data or how missing data was handled; states that second colonoscopy could not be performed in 2 patients; states that "only biopsy specimens from patients who underwent two endoscopic procedures and biopsy sampling were used for the final analysis (n=9)" Unclear which analysis this refers to.	% of withdrawal/loss to follow-up unclear.

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
Targan 1997[54] (Infliximab)	4 week multi-centre, randomised, double-blind, placebo-controlled trial; 108 patients randomised to placebo, 5, 10, or 20mg/kg infliximab Patients without a response at week 4 were enrolled in a parallel, open-label study and were followed for 12 additional weeks	(Refers to first 4 weeks) • Placebo identical in appearance to infliximab solution • Patients blinded • Study investigators/personnel blinded	Unclear how missing data was handled. States that the original study protocol did not specify the use of intention-to treat analysis, but that patients were analysed according to assignment (except 2 patients who did not receive treatment and were excluded from the analysis). For assessing the response and remission rates in all evaluation periods after the initial blinded infusion, patients who received an open-label infusion or those with a change in concomitantly administered medication were considered non-responders. It is unclear if patients who did not contribute data during blinded period where also counted as non-responders. Not clear how missing continuous data handled.	Based on data in figure 1 appears that at most 7/108 (6%) patients not evaluated for results (at week 2) 100% of patients completed 4 weeks of double-blind therapy.
MAINTENANCE				

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
Hanauer 2002[4] & Rutgeerts 2004[3] ACCENT I (Infliximab)	54 week, multi-centre, randomised, double-blind placebo-controlled trial; all patients initially received 5 mg/kg (week 0), then randomised to placebo ('episodic treatment') or 5 mg/kg (5 mg/kg week 2 and 6, then every 8 weeks) or 10mg/kg infliximab (5 mg/kg week 2 and 6, then 10 mg/kg every 8 weeks) All patients included in analysis - responders and non-responders (in Rutgeerts et al., 2004) Responders only analysis in Hanauer et al., 2002[4] NB at week 14 or later patients who had responded at any time to infliximab therapy but then worsened were eligible to crossover to 'active episodic' treatment *as needed* with infliximab 5, 10, 15 mg/kg for patients originally assigned to episodic, 5 mg/kg or 10 mg/kg respectively.	• Placebo identical in appearance to infliximab solution • Patients blinded (until crossover if applicable) • Study investigators blinded (until patient crossover if applicable)	Rutgeerts 2004 -responder and non-responder analysis: Data from the patients who participated in the crossover to treatment with a higher dose, upon loss of response, were analysed under the original treatment group assignment. Patients who withdrew from the study, or did not have a value at an originally scheduled visit because of crossover, and those with missing CDAI or IBDQ scores had their last value carried forward for these analyses. Hanauer 2002-reponder analysis: Data obtained after episodic re-treatment were not included in the efficacy analysis. Patients who crossed over to episodic infliximab retreatment, who received a protocol –prohibited drug, who had surgery for Crohn's disease, or who discontinued follow-up due to lack of efficacy or loss of response were judged to have failed treatment, irrespective of the CDAI score. Patients who discontinued the study for reasons other than lack of efficacy or loss of response and those with missing CDAI scores were censored in the analysis of time to loss of response up to week 54. These patients were treated as not in clinical response or clinical remission for other analyses.	124/573 (22%) patients had withdrawn by week 54; 201/573 (35%) had crossed over to active episodic treatment by week 54 (92/188 (49%) of patients crossed over from placebo to episodic treatment) No loss to follow-up Results include any patients with a response or in remission at different time points, not just patients maintaining a response (also includes non-responders)

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
Rutgeerts 1999[55] (Infliximab)	36 week, multi-centre, randomised, double-blind placebo-controlled trial; patients randomised to placebo or 10mg/kg infliximab; eligible patients had previously shown a response in the RCT by Targan 1997[54] (see induction trials) or, if initial non-response, a response in an 8 week open label extension of Targan 1997[54]; unclear if this included any patients who had shown a response to placebo, or if all had received infliximab Responders only randomised	• Patients blinded No explicit statement regarding blinding of other parties.	Treatment was considered a failure in patients who underwent surgery or were treated with medication regimens excluded from the study regardless of CDAI. Last measure carried forward for continuous measures (CDAI, IBDQ) in patients who discontinued follow-up, had a CD related surgical procedure or non-permitted medication change.	Results include any patients with a response at different time points, not just patients maintaining a response; unclear why data does not start with 100% of patients with a response as only responders included 24/73 (33%) patients withdrawn by end of study. No details regarding potential crossovers.

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
Colombel 2007[62] CHARM (Adalimumab)	56 week, multi-centre, randomised, double-blind placebo-controlled trial; all patients received adalimumab 80 mg subcutaneously at week 0, followed by 40 mg dose at week 2; randomisation at week 4, stratified by responder status to placebo or 40 mg adalimumab weekly or 40 mg adalimumab every other week (eow); responders only included in efficacy analysis States that secondary efficacy analyses include non-responders also, but present results for responders only in this publication; only fistula results include non-responders NB those patients who experienced a disease flare or sustained non-response at or after week 12 were switched to open label treatment (40mg eow, which could be escalated to 40mg weekly)	• No details regarding placebo • Patients blinded (until open label if applicable) • Study investigators and coordinators blinded (until open label if applicable)	Patients who switched to open label therapy or withdrew from the study were counted as remission failures. Patients without CDAI assessments at weeks 26 or 56 were classified as remission failures. (remission =primary endpoint) Unclear if patients also counted as treatment failures for secondary outcomes (response). No details on how continuous data was handled.	Results include any patients in remission at different time points, not just patients maintaining remission 505/778 (65%) of patients completed study (NB paper states 59%-unclear); 50% of these remained on double-blind therapy, 50% completed the study on open label treatment Of 499 patients (responders) included in efficacy analysis, 29% withdrew and 38% were still on double-blind therapy. Assume 33% therefore on open-label treatment though not clearly stated. No details on loss-to follow-up.

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
Sandborn 2007[61] CLASSIC II (Adalimumab)	56 week, multi-centre, randomised, double-blind placebo-controlled trial; all patients from CLASSIC I[58] trial eligible if they demonstrated remission at weeks 0 and 4 (unclear if this includes patients from placebo group in remission); randomisation to placebo, 40 mg adalimumab weekly or 40 mg adalimumab every other week (unclear if placebo weekly or eow) Randomised patients in remission only NB Randomised patients experiencing a flare or with continued non-response could switch to open label adalimumab 40mg eow; patients on OL adalimumab eow could switch to adalimumab 40mg weekly	• No details regarding placebo (unless assume same as in CLASSIC I[58]) • Patients blinded (until open label if applicable) • Study investigators and coordinators blinded (until open label if applicable)	Efficacy analysis included all randomised patients; patients who switched to open label or with missing data were classified in a 'no maintenance of remission' category Secondary analyses used 'last observation carried forward'	10/55 (18%) withdrew (of these 1 lost to follow-up) 32/55 (58%) of patients completed 56 weeks of double-blind therapy (6/18, 33% of patients in placebo group completed 56 weeks of double-blind therapy); remainder completed study on open label therapy Results include any patients in remission at different time points, not just patients maintaining remission
FISTULISING				
Present 1999[57] (Infliximab)	18 week multi-centre, randomised, double-blind placebo-controlled trial; patients randomised to placebo, 5 mg/kg infliximab or 10 mg/kg infliximab	• Placebo identical in appearance to infliximab solution No details on blinding (other than to state that this was a double-blind trial).	Treatment considered to have failed in patients who had changes in medication that were not permitted, who underwent surgery related to CD or who did not return for follow-up visits. For continuous variables measurements from the last evaluation were carried forward.	88/94 (94%) completed trial; no loss to follow-up Appear to be no crossovers

Study	Trial design	Blinding	Handling of missing data (binary and continuous); ITT	% of withdrawals and/or crossovers and loss to follow-up Other comments
Sands 2004[60] ACCENT II (Infliximab)	54 week multi-centre, randomised, double-blind placebo-controlled trial; all patients received 5 mg/kg infliximab at week 0, 2 and 6; responders at week 14 randomised to placebo or 5 mg/kg infliximab every 8 weeks Responders only included in primary analysis Non-responders also randomised for secondary analysis From week 22, patients could crossover from placebo to 5 mg/kg or from 5 mg/kg to 10 mg/kg infliximab	• Placebo identical in appearance to infliximab solution • Patients blinded • Study investigators blinded	All patients included in analysis. Data for patients who crossed over from placebo to infliximab were censored before crossover occurred. Not stated for patients who crossed over from lower to higher infliximab dose. For continuous variables (CDAI, IBDQ) measurements from the last evaluation were carried forward.	95/282 (34%) crossed over (total randomised population; 2223 from placebo group to treatment) 78/195 (40%) crossed over (responder only) by week 54 (28% from placebo group to treatment) No details on withdrawals or loss to follow-up post-randomisation
PAEDIATRIC				
Baldassano 2003[43] (Infliximab)	12 week multi-centre, randomised (no placebo control); 21 patients randomised to 1 mg/kg, 5 mg/kg or 10 mg/kg infliximab	• No placebo, all received infliximab • Patients blinded to dose • Study investigators blinded to dose	No details. Numbers included in different analyses vary at different time-points.	19/21 (90%) of patients completed trial. No further details.
Hyams 2007[42] REACH (Infliximab)	54 week multi-centre, randomised, open-label (no placebo control); 112 patients received induction therapy (5mg/kg infliximab) for 10 weeks; only patients with response (n=103) randomised at week 10 to 5mg/kg infliximab every 8 weeks or 5mg/kg infliximab every 12 weeks; patients losing clinical response eligible to cross over one time to receive treatment more frequently or at higher dose (10mg/kg every 8 weeks)	No blinding: open-label study	All analyses based on ITT principle. Patients who lost response and crossed over were considered non-responders (treatment failures) for the remainder of the study. Last non-missing score used for continuous data where patients discontinued study or had insufficient data.	59/103 (57%) patients in treatment arms as randomised at study end. 35 patients (34%) crossed over in total and 9 (9%) withdrew. No loss to follow-up.

Appendix 13. Rates of response and remission in placebo arms of induction trials for anti-TNF interventions.

Figure 76. Placebo rates for response 100 (upper panel) and response 70 (lower panel) in induction trials.

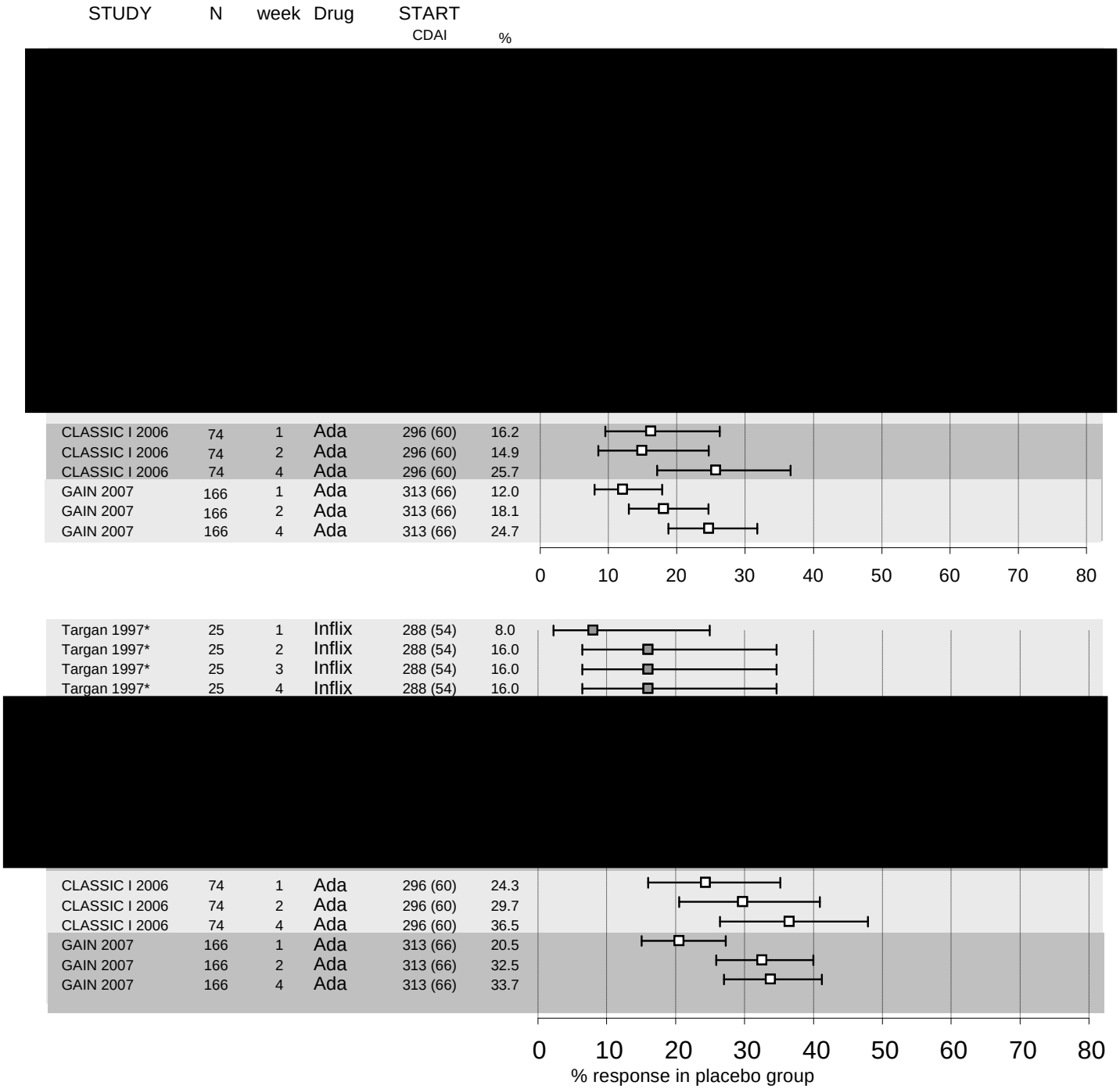
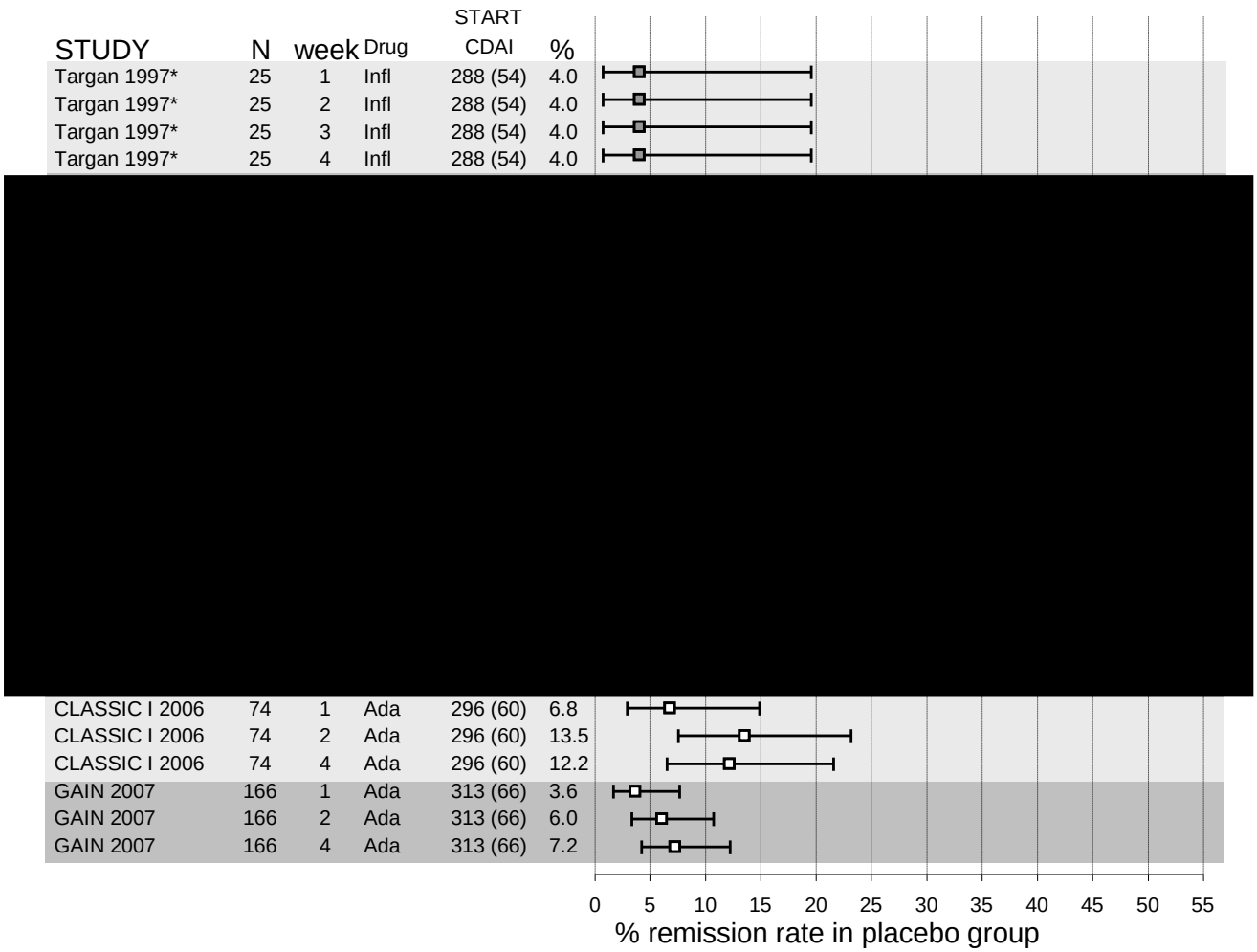


Figure 77. Placebo rates for remission in induction trials



Appendix 14. Search strategy for economic evaluation

Note: certolizumab pegol and natalizumab were originally part of this appraisal; they were subsequently excluded after searching had been completed

Source – MEDLINE (Ovid) 1950 to May Week 4 2007

1 (adalimumab or humira).mp. (540)
2 (certolizumab or cimzia).mp. (19)
3 (infliximab or remicade).mp. (3096)
4 (natalizumab or tysabri).mp. (208)
5 or/1-4 (3473)
6 Crohn Disease/ (21624)
7 crohn$.mp. (25626)
8 or/6-7 (25626)
9 5 and 8 (1046)
10 economics/ (24885)
11 exp "costs and cost analysis"/ (129414)
12 cost of illness/ (9149)
13 exp health care costs/ (28541)
14 economic value of life/ (4847)
15 exp economics medical/ (11355)
16 exp economics hospital/ (14731)
17 economics pharmaceutical/ (1764)
18 exp "fees and charges"/ (22970)
19 (econom$ or cost or costs or costly or costing or price or pricing or pharmacoeconomic$).tw. (244897)
20 (expenditure$ not energy).tw. (10410)
21 (value adj1 money).tw. (10)
22 budget$.tw. (10892)
23 or/10-22 (358461)
24 9 and 23 (51)
25 limit 24 to yr="2000 - 2007" (48)

Source – MEDLINE (Ovid) 1950 to June Week 3 2007*

1 ca2.mp. (105908)
2 d2e7.mp. (23)
3 cdp870.mp. (26)
4 pha-738144.mp. (0)
5 pha 738144.mp. (0)
6 (anti adj2 4 integrin).mp. (45)
7 anti alpha4 integrin.mp. (49)
8 anti alpha 4 integrin.mp. (32)
9 or/1-8 (106047)
10 crohn disease/ (21699)
11 crohn$.mp. (25732)
12 or/10-11 (25732)
13 9 and 12 (66)
14 economics/ (24922)
15 exp "costs and cost analysis"/ (130028)
16 cost of illness/ (9244)
17 exp health care costs/ (28753)

18 economic value of life/ (4854)
19 exp economics medical/ (11385)
20 exp economics hospital/ (14773)
21 economics pharmaceutical/ (1786)
22 exp "fees and charges"/ (23036)
23 (econom$ or cost or costs or costly or costing or price or pricing or pharmacoeconomic$).tw. (246746)
24 (expenditure$ not energy).tw. (10484)
25 (value adj1 money).tw. (10)
26 budget$.tw. (10945)
27 or/14-26 (360657)
28 13 and 27 (1)

* Additional search to account for alternative terminology used for the drugs.

Source - EMBASE (Ovid) 1980 to 2007 Week 22

1 (adalimumab or humira).mp. (2036)
2 (certolizumab or cimzia).mp. (230)
3 (infliximab or remicade).mp. (7811)
4 (natalizumab or tysabri).mp. (843)
5 or/1-4 (8685)
6 Crohn Disease/ (20817)
7 crohn$.mp. (23756)
8 or/6-7 (23756)
9 5 and 8 (2554)
10 cost benefit analysis/ (26197)
11 cost effectiveness analysis/ (48867)
12 cost minimization analysis/ (1140)
13 cost utility analysis/ (1927)
14 economic evaluation/ (3621)
15 (cost or costs or costed or costly or costing).tw. (146502)
16 (economic$ or pharmacoeconomic$ or price$ or pricing).tw. (70477)
17 (technology adj assessment$).tw. (1366)
18 or/10-17 (223990)
19 9 and 18 (151)
20 limit 19 to yr="2000 - 2007" (149)

Source – EMBASE (Ovid) 1980 to 2007 Week 25*

1 ca2.mp. (115879)
2 d2e7.mp. (65)
3 cdp870.mp. (16)
4 pha-738144.mp. (1)
5 pha 738144.mp. (1)
6 (anti adj2 4 integrin).mp. (9)
7 anti alpha4 integrin.mp. (37)
8 anti alpha 4 integrin.mp. (2)
9 or/1-8 (116001)
10 crohn disease/ (20928)
11 crohn$.mp. (23876)
12 or/10-11 (23876)
13 9 and 12 (72)
14 cost benefit analysis/ (26342)

15 cost effectiveness analysis/ (49154)
16 cost minimization analysis/ (1156)
17 cost utility analysis/ (1947)
18 economic evaluation/ (3637)
19 (cost or costs or costed or costly or costing).tw. (147257)
20 (economic$ or pharmacoeconomic$ or price$ or pricing).tw. (70847)
21 (technology adj assessment$).tw. (1367)
22 or/14-21 (225192)
23 13 and 22 (3)

* Additional search to account for alternative terminology used for the drugs.

Quality of life:

Source – MEDLINE (Ovid) 1950 to May Week 4 2007

1 (adalimumab or humira).mp.
2 (certolizumab or cimzia).mp.
3 (infliximab or remicade).mp.
4 (natalizumab or tysabri).mp.
5 or/1-4 (3473)
6 Crohn Disease/ (21624)
7 crohn$.mp. (25626)
8 or/6-7 (25626)
9 5 and 8 (1046)
10 quality of life/ (59486)
11 life style/ (25902)
12 health status/ (33125)
13 health status indicators/ (10984)
14 value of life/ (4847)
15 quality adjusted life.mp. (3912)
16 or/10-15 (124425)
17 8 and 16 (427)
18 limit 17 to yr="2000 - 2007" (246)
19 from 18 keep 1-246 (246)

Source – EMBASE (Ovid) 1980 to 2007 Week 22

1 (adalimumab or humira).mp. (2036)
2 (certolizumab or cimzia).mp. (230)
3 (infliximab or remicade).mp. (7811)
4 (natalizumab or tysabri).mp. (843)
5 or/1-4 (8685)
6 Crohn Disease/ (20817)
7 crohn$.mp. (23756)
8 or/6-7 (23756)
9 5 and 8 (2554)
10 quality of life/ (75452)
11 quality adjusted life year/ (3013)
12 health status/ (31455)
13 health status indicator$.mp. (129)
14 or/10-13 (104174)
15 8 and 14 (624)

16 limit 15 to yr="2000 - 2007" (481)

Source - HEED June 2007

Search terms: (adalimumab OR humira OR certolizumab OR cimzia OR infliximab OR remicade OR natalizumab OR tysabri OR ca2 OR d2e7 OR cdp870 OR pha-738144 OR pha 738144 OR anti 4 integrin OR anti alpha4 integrin OR anti alpha 4 integrin) AND (crohn OR crohns)

Cohort studies of Infliximab and Crohns Disease

Source – MEDLINE (Ovid) 1950 to May Week 4 2007

1 (infliximab or remicade).mp. (3096)
2 crohn$.mp. (25626)
3 crohn disease/ (21624)
4 or/2-3 (25626)
5 1 and 4 (992)
6 cohort studies/ (73136)
7 Risk/ (74606)
8 cohort$.mp. (132834)
9 risk$.mp. (848754)
10 or/6-9 (922221)
11 5 and 10 (186)

Clinical guidelines

Source - MEDLINE(Ovid) 1950 to May Week 5 2007

1 Crohn Disease/ (21659)
2 crohn$.mp. (25672)
3 or/1-2 (25672)
4 exp "guideline [publication type]"/ (15854)
5 exp "consensus development conference [publication type]"/ (5531)
6 guideline$.mp. (137797)
7 recommend$.mp. (215591)
8 consensus.mp. (63448)
9 or/4-8 (381390)
10 3 and 9 (631)
11 or/4-5 (20617)
12 3 and 11 (51)

Appendix 15. Details of studies included in cost-effectiveness review

Table 76. Study characteristics of studies in cost-effectiveness review

Study	Type of evaluation and synthesis	Interventions	Study population	Country	Duration of study
Jaisson-Hot (2004[8]; non-fistulising)	CUA	A) Surgery and medical treatment (without infliximab) B) Infliximab (infusions + episodic reinfusions for relapse) C) Infliximab (maintenance)	Adult patients with non-responsive, non-fistulising CD, (CDAI between 220 and 440) 38 years old at baseline.	France	Lifetime
Clark (2003[6]; non-fistulising)	CUA	A) Placebo. B) Single dose: Infliximab C) Episodic: Initial infliximab + (for responders) up to three treatments at subsequently relapses (flare).	Adult patients (70kg) with non-responsive, non-fistulising CD, 37 years at baseline.	UK	Unclear, probably 1 year.
Marshall (2002[7]; non-fistulising)	CUA	A) Usual care B) Single dose: Infliximab infusion at Week 0, relapses treated with usual care C) Episodic: Infliximab infusion at Week 0, relapses treated with single infusion of infliximab D) Maintenance: Infliximab infusion at Week 0, with responding patients (CDAI drop of 70) receiving maintenance infusions of infliximab 5mg/kg every 8 weeks starting at Week 12. Non-responding or subsequently relapsing patients receive usual care.	Adult patients (70kg) with CD resistant to conventional medical therapy.	Canada	1 year
Clark (2003[6]; fistulising)	CUA	A) Placebo. B) Initial treatment only.	Adult patients with fistulising Crohn's disease.	UK	1 year
Arseneau (2001[76]; fistulising)	CUA	A) 6-mercaptopurine and metronidazole as first and second line treatment B) Infliximab infusions (infliximab infusions + 6MP/met for treatment failures/relapse) C) Infliximab (infusions + episodic infliximab reinfusions for treatment failures/relapse) D) Second-line infliximab (6-MP/met + episodic infliximab reinfusions for treatment failures/relapse)	Adult patients (70kg) with symptomatic perianal fistulas.	USA	1 year

Table 77. Type of model used in studies in cost-effectiveness review

Study	Type of model	Perspective	Model assumptions	
			Outcomes	Costs and resource use
Jaisson-Hot (2004[8]; non-fistulising)	Markov model, cycle length of 2 months	Third-party payer perspective	Lifetime model but no stated mortality assumptions.	Infliximab dose at 5mg/kg per infusion. Maintenance treatment every 8 weeks.
Clark (2003[6]; non-fistulising)	Modified industry submission. Markov model, cycle length of 2 months	Unclear.	Benefits related to the numbers in remitted health state (CDAI <150). Report also gives outcomes under industry assumption (benefit = reduction of 70 CDAI points).	Unclear.
Marshall (2002[7]; non-fistulising)	Markov model: initial cycle length of 12 weeks, with subsequent cycles at 8 weeks.	Third-party (Canadian provision ministry of health) perspective.	US data (Olmstead County) used to estimate transition probabilities in usual care. No transitions between Remission and Drug Responsive states (due to data limitations). Re-treatment strategy assumed to have equivalent effectiveness to initial dosage. All infliximab dosages (5mg/kg, 10mg/kg, 20mg/kg) treated as equally effective.	20% of patients in drug refractory state would be admitted to hospital, with the remaining 80% receiving outpatient care. Only 5mg/kg infliximab dosages used. Acute infusion reactions are mild, and have no effect on treatment efficacy or cost. Methotrexate and cyclosporine not used by the model cohort. No medication given in the period following surgery as post-operative prophylaxis.
Clark (2003[6]; fistulising)	Unclear.	Unclear.	Time spent with fistulas closure.	Infliximab dose (unclear) offset by possible savings in surgery.
Arseneau (2001[76]; fistulising)	Markov model, cycle length of 1 month	Third-party payer perspective	Episodic remission figures assumed to equal remission from initial infusion. Benefits from initial infusion assumed to occur within the first month following infusion. The chance of fistula recurrence increases by 3% per month after 4 months. Pancreatitis state includes 1 week with acute pancreatitis, and 3 weeks of fistula/improved fistula.	Initial infliximab infusions at 5mg/kg, (Weeks 0, 2, 6) according to FDA-approved protocol.

Table 78. Cost and resource use data sources for studies in cost-effectiveness review

Study	Cost items	Cost data sources	Resource use	Resource data source	Currency and currency year	Discount rate
Jaisson-Hot (2004[8]; non-fistulising)	Hospitalisations. Outpatient care (physicians' visits, nursing care, laboratory), medications, and patient transportation.	Some unit costs based on diagnosis related group (DRG) estimates and negotiated prices.	Based on expert opinion. Not details given.	Not given.	Not given.	5%
Clark (2003[6]; non-fistulising)	Drug and administration costs. Other items unclear.	Not given.	Not given.	Not given.	Not given.	Not discounted (probably 1 year)
Marshall (2002[7]; non-fistulising)	Infliximab infusion; CD-related outpatient prescriptions; outpatient physician visits; medical hospital admissions for CD; surgical hospital admissions for CD	Unit costs based on: 2001 Drug Benefits Formulary, McMaster University Medical Centre outpatient pharmacy;	Appears in appendices to CCOHTA report.	Three member expert panel of Gastroenterologists based on text description. Surgical costs from patient-level database.	Canadian dollars, 2001	Not discounted (1 year)
Clark (2003[6]; fistulising)	Drug costs. Surgery. Other items unclear.	Not given.	Not given.	Not given.	Not given.	Not discounted (1 year)
Arseneau (2001[76]; fistulising)	Diagnostic, physician, medication. Surgical costs in abscess state only.	Administrative database of hospital and physician billing data. Cost data calculated according to hospital cost-charge ratios.	Usage split by state and treatment.	Not given.	US dollars, 1999	3%

Table 79. Efficacy data and health outcomes/utility for studies in cost-effectiveness review

Study	Efficacy data	Efficacy data sources	Health outcomes/utility	Health outcome data sources	Discount rate
Jaisson-Hot (2004[8]; non-fistulising)	Derived from published data and expert opinion.	Unclear. Some figures from Targan 1997[54]	QALY. Quality of life figures unclear.	Gregor et al (1997)	5%
Clark (2003[6]; non-fistulising)	Response from treatment continued for 80 days (median) in both initial and subsequent treatment. 100% success of initial responders in re-treatment. Large amounts data removed due to confidentiality. Scenario 1: uses company's effectiveness estimates. 19.4% more patients achieve remission (CDAI < 150) over infliximab arms. Scenario 2: uses estimates on remission at different dosages to infer the proportion of those achieving mild disease at a 5mg dosage. 28.7% more patients achieve remission (CDAI < 150) under 5mg infliximab.	Olmstead County data (usual care). Infliximab data from clinical trials but not ACCENT I[3,4].	Interpolation used to SG utilities. Mild Disease: 0.86 Drug-refractory disease: 0.74	Gregor et al (1997) plus Olmstead County data.	Not discounted (probably 1 year)
Marshall (2002[7]; non-fistulising)	8 week transitions: Usual care Drug refractory from remission: 0.2150 Remission from drug refractory: 0.0524 Remission from drug dependent: 0.0540 8 week transitions: Infliximab Probability of remaining in clinical response (remission or drug responsive) over 8 weeks = 0.796 (single dose) 0.937 (maintenance). Remission at CDAI<150.	Olmstead County data (usual care) Targan 1997[54] trial (infliximab, initial values 46) Rutgeerts trial (infliximab, after 12 weeks)	Mild (0.82) used for remission states and mild disease. Moderate (0.73) used for drug responsive/dependent states. Severe (0.54) used for drug refractory and surgery states.	Gregor et al (1997) provides SG values for three states (Mild, Moderate, and Severe)	Not discounted (1 year)
Clark (2003[6]; fistulising)	Based on Present 1999[57] study into time spent with close fistulas in first 12 months after treatment.	Present 1999[57] study	Based on CDAI and PDAI scores using an unpublished algorithm provided in industry submission.	Unpublished data.	Not discounted (1 year)
Arseneau (2001[76]; fistulising)	Fistula recurrence based on clinical data in first four months (18% per month), then 3% in subsequent months. Monthly transitions: Fistula improves (complete closure or symptomatic improvement) after infliximab: 0.70	Various studies (named).	QALY. Quality of life figures from patients: Infliximab: Fistula: 0.73	Standard gamble utilities from 32 CD patients (17 fistulising, 15 non-fistulising).	3%

	Recurrent fistula after infliximab (≤4 months): 0.18 Recurrent fistula after infliximab (>4 months): 0.03 Abscess after infliximab 0.06 Abscess recurs after incision and drainage: 0.03 Fistula improves after 6MP/met: 0.48 Recurrent fistula after 6MP/met is stopped: 0.14 Recurrent fistula whilst taking 6MP/met: 0.01 Pancreatitis: 0.03 Paresthesias: 0.10		Improved fistula: 0.85 Perianal abscess: 0.62 6MP/met: Fistula: 0.69 Improved fistula: 0.81 Pancreatitis + fistula: 0.61 Pancreatitis alone: 0.70 Paresthesias + fistula: 0.66 Paresthesias: 0.75	Descriptions of valued states not given.	

Table 80. Cost-effectiveness ratios for studies in cost-effectiveness review

Study	Cost of anti-TNF-α therapy	Total costs	Total incremental costs	Total outcome	Total incremental outcomes	Cost-effectiveness ratios
Jaisson-Hot (2004[8]; non-fistulising)	Not given.	A) Surgery + medical management €71,296.44 B) Infliximab (episodic) €119,801.60 C) Infliximab (maintenance) €687,086.96	B vs A: €48,505.16 C vs A: Infliximab (maintenance) €615,790.52	Not given.	Not given.	B vs A: Infliximab (episodic) €63,700.82/QALY C vs A: Infliximab (maintenance) versus usual care €784,057.49/QALY
Clark (2003[6]; non-fistulising)	£1,457 per dose.	Not given.	vs Placebo <u>Single Treatment</u> £1,457 per patient. <u>Episodic Treatment (vs placebo)</u> £3,861	Not given.	QALY vs Placebo <u>Single Treatment</u> Scenario 1: 0.006 Scenario 2: 0.009 <u>Episodic Treatment</u> Scenario 1: 0.043 Scenario 2: 0.067	vs Placebo <u>Single Treatment</u> Scenario 1: £244,756 per QALY Scenario 2: £165,445 per QALY <u>Episodic Treatment</u> Scenario 1: £72,261 per QALY Scenario 2: £62,016 per QALY
Marshall (2002[7]; non-fistulising)	Single dose cost C$5064.11.	A) C$9,940 B) C$12,702 C) C$13,739 D) C$21,597	B vs A: C$2,762 C vs B: C$1,037 D vs C: C$7,858	A) 0.6281 B) 0.6433 C) 0.6455 D) 0.6568	B vs A: 0.0152 C vs B: 0.0022 D vs C: 0.00132762/.	B vs A: C$181,201/QALY C vs B: C$480,111/QALY D vs C: C$696,078/QALY
Clark (2003[6]; fistulising)	Unclear.	Not given.	Not given.	Not given.	Not given.	Initial treatment versus placebo is £102,000-123,000 per QALY depending on cost offsets.
Arseneau (2001[76]; fistulising)	Single dose cost $2,030 for 5mg.kg dose, 70kg person.	A) $2,894 B) $10,003 C) $10,112 D) $6,664	All vs comparator: B vs A. $7109 C vs A. $7218 D vs A. $3770	A) 0.76 B) 0.78 C) 0.78 D) 0.77	Not given.	All vs comparator: B vs A. $355,450 C vs A. $360,900 D vs A. $377,000

Table 81. Sensitivity analyses for studies in cost-effectiveness review

Study	Sensitivity analysis methods	Sensitivity analysis results
Jaisson-Hot (2004[8]; non-fistulising)	"Influential" variables considered, but choice of variables not justified. Tornado diagram used to identify utility weights for "post-surgical remission" and "remission not following surgery" as important. Only one-way sensitivity analyses reported.	Surgery and non-infliximab treatment becomes dominant where postsurgical remission receives utility value 0.92. No dominance found when varying the value for non-surgical remission utility.
Clark (2003[6]; non-fistulising)	One-way sensitivity analyses for utility (to 0.20 from 0.12), duration of response (120 days from 80 days), averted surgery (50% averted surgeries).	None of the one-way sensitivity analyses reduced the ICERS below£40,000 per QALY.
Marshall (2002[7]; non-fistulising)	Probabilistic sensitivity analysis conducted in addition to one way sensitivity analysis: use of medical/surgical treatment in drug refractory state (varying 0% to 100% from 20% baseline). Surgical admissions varied (0% to 100%, 13% baseline). Infliximab cost (0% to 100% of baseline cost).	Rate of surgical admission for drug-refractory CD found to have little effect on ICER. Proportion of patients with drug-refractory disease treated medically fell to C$39,000/QALY at 60% for B vs A. At 75% of baseline cost, ICERs are: (B vs A) C$98,186, (C vs B) C$329,204, (D vs C) C$522,511 /QALY. Usual care dominated by Strategy B (one single dose) where prices reduced to 25% of baseline cost. Usual Care favoured for maximum WTP per QALY (□) < C$180,000. One single dose of infliximab (B) favoured for C$180,000 < □ □C$430,000
Clark (2003[6]; fistulising)	Success rate for re-treatment and re-closure of fistulas varied, alongside the level of costs offset due to averted surgery.	Even at the most favourable assumptions, the ICER remains above £80,000 per QALY.
Arseneau (2001[76]; fistulising)	One-way sensitivity analyses for all cost, probability, and utility estimates in the model, as well Cost estimates varied by25%, probability and utility estimates over 95% CI. One-way sensitivity analyses as assumptions varying fistula recurrence >4 months after infliximab usage. (0% or 18% recurrence) One-way sensitivity analysis on the effectiveness of infliximab as first- and second-line therapy (0% to 100%). Tornado diagram used to identify influential variables (not given) Threshold analysis on the cost of a single dose of infliximab. Utility estimates from healthy volunteers.	All ICERs remain above $100,000 per QALY, except where comparator treatment dominates (equal or more effective, lower cost). ICER above $100,000 per QALY even with 100% chance of improvement following either 1st line, 2nd line, or reinfused infliximab. Assuming 18% recurrence rate of fistulas after infliximab following month 4 increases ICERS to: $736,400, $409,500, $412,700 per QALY (Int I, II, III versus comparator). Assuming 0% recurrence rates of fistulas after infliximab following month 4 decreases ICERS to: $339,450, $218,133, and $361,200 per QALY (Int I, II, III versus comparator). Intervention II) Infliximab (infusions + episodic infliximab reinfusions for treatment failures/relapse) falls beneath $100,000 per QALY where infliximab dose is reduced in price by 75% (to $508/dose).

Table 82. Author conclusions for studies in cost-effectiveness review

Study	Author Conclusions	Industry author affiliation
Jaisson-Hot (2004[8]; non-fistulising)	Infliximab treatment (episodic) could be cost-effective but infliximab treatment (maintenance) may not justify increased cost.	None declared.
Clark (2003[6]; non-fistulising)	Re-estimation of the cost-effectiveness using company estimates for the proportion of patients gave a cost/QALY for episodic treatment of £72,000 when using efficacy data from all infliximab arms, and £62,000 when using 5mg/kg dosing. These findings were relatively insensitive to major changes in key assumptions. The key issue appears to be the duration of benefit from treatment.	None declared. Study funded by UKHTA
Marshall (2002[7]; non-fistulising)	For cost-effectiveness thresholds less than C$180,000, usual care was more likely to maximise net benefit than infliximab treatment strategies.	None declared. Study funded by CCOHTA.
Clark (2003[6]; fistulising)	The cost-per-QALY estimates from the industry model were high, at £82,000 even in the most favourable re-treatment assumptions on closure rates.	None declared. Study funded by UKHTA
Arseneau (2001[76]; fistulising)	The ICER for infliximab is above $350,000 per QALY, driven by both the high cost of infliximab and the similar effectiveness of infliximab and 6MP/metronidazole treatment strategies.	None declared.

Table 83. Quality assessment for studies in cost-effectiveness review

	Jaisson-Hot (2004)[8]	Marshall (2002) [7]	Arseneau (2001) [76]
(1) The research question is stated	Yes	Yes	Yes
(2) The economic importance of the research question is stated	Yes	Yes	Yes
(3) The viewpoint(s) of the analysis are clearly stated and justified	Yes	Yes	Yes
(4) The rationale for choosing the alternative programmes or interventions compared is stated	Yes	Yes	Unclear
(5) The alternatives being compared are clearly described	Yes	Yes	Yes
(6) The form of economic evaluation used is stated	Yes	Yes	Yes
(7) The choice of form of economic evaluation is justified in relation to the questions addressed	Yes	Yes	Yes
(8) The source(s) of effectiveness estimates used are stated	Unclear	Yes	Yes
(9) Details of the design and results of effectiveness study are given (if based on a single study)	No	NA	NA
(10) Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)	NA	Yes	Yes
(11) The primary outcome measure(s) for the economic evaluation are clearly stated	Yes	Yes	Unclear
(12) Methods to value health states and other benefits are stated	Yes	Yes	Yes
(13) Details of the subjects from whom valuations were obtained are given	Yes	Yes	Yes
(14) Productivity changes (if included) are reported separately	NA	NA	NA
(15) The relevance of productivity changes to the study question is discussed	Yes	Yes	Yes
(16) Quantities of resources are reported separately from their unit costs	No	Yes	Yes
(17) Methods for the estimation of quantities and unit costs are described	No	Yes	Yes
(18) Currency and price data are recorded	No	Yes	Yes
(19) Details of currency of price adjustments for inflation or currency conversion are given	No	Yes	NA
(20) Details of any model used are given	Unclear	Yes	Yes
(21) The choice of model used and the key parameters on which it is based are justified	No	Yes	Yes
(22) Time horizon of costs and benefits is stated	Yes	Yes	Yes
(23) The discount rate(s) is stated	Yes	1 Year	Yes
(24) The choice of rate(s) is justified	No	1 Year	No
(25) An explanation is given if costs or benefits are not discounted	NA	1 Year	NA
(26) Details of statistical tests and confidence intervals are given for stochastic data	No	Yes	Partial
(27) The approach to sensitivity analysis is given	Yes	Yes	Yes
(28) The choice of variables for sensitivity analysis is justified	Yes	Yes	Yes
(29) The ranges over which the variables are varied are stated	No	Yes	Yes
(30) Relevant alternatives are compared	Yes	Yes	Yes
(31) Incremental analysis is reported	No	Yes	No
(32) Major outcomes are presented in a disaggregated as well as aggregated form	No	Yes	No

(33) The answer to the study question is given	Yes	Yes	Yes
(34) Conclusions follow from the data reported	Yes	Yes	Yes
(35) Conclusions are accompanied by the appropriate caveats	Yes	Yes	Yes
	18/35	34/35	26/35

Table 84. Included and excluded studies cost-effectiveness review

PAPER	INCLUDED/EXCLUDED	REASON
Arseneau et al (2001)	included	
Clark et al (2004)	included	
Jaisson-Hot et al (2004)	included	
Marshall et al (2002)	included	
Marshall (2002b)	excluded	see Marshall (2002)
Dubinsky et al (2005)	excluded	comparators not relevant
Williams et al (2000)	excluded	comparators not relevant
Condino et al (2005)	excluded	comparators not relevant
Harrison and Rubensteini (2003)	excluded	abstract only
Wong (1999)	excluded	abstract only
Andersson et al (2003)	excluded	not ee
Arnott et al (2001)	excluded	not ee
Balfour Sartor (2004)	excluded	not ee
Barkun (2002)	excluded	not ee
Bassi et al (2004)	excluded	not ee
Bernklev et al (2005)	excluded	not ee
Bernklev et al (2006)	excluded	not ee
Bodger (2002)	excluded	not ee
Bodger (2005)	excluded	not ee
Broering et al (2001a)	excluded	not ee
Broering et al (2001b)	excluded	not ee
Buller (2001)	excluded	not ee
Cadahia et al (2004)	excluded	not ee
Caprilli et al (2006)	excluded	not ee
Casellas (2000)	excluded	not ee
Casellas et al (2003)	excluded	not ee
Casellas et al (2005a)	excluded	not ee
Casellas et al (2005b)	excluded	not ee
Cohen (2002a)	excluded	not ee
Cohen (2002b)	excluded	not ee
Cohen (2003)	excluded	not ee
Cohen (2006)	excluded	not ee
Cohen et al (2002)	excluded	not ee
Colombel et al (2007)	excluded	not ee
D'Haens (2002)	excluded	not ee
Etienney et al (2004)	excluded	not ee
Feagan (2001)	excluded	not ee
Feagan et al (2003)	excluded	not ee
Feagan et al (2005)	excluded	not ee
Fleurence and Spackman (2006)	excluded	not ee
Garnett and Yunker (2001)	excluded	not ee
Ghosh (2003)	excluded	not ee
Goldfarb et al (2004)	excluded	not ee
Gregor et al (1997)	excluded	not ee
Hanauer (2005)	excluded	not ee

Hanauer (2007)	excluded	not ee
Hilsden (2002)	excluded	not ee
Hyams (2003)	excluded	not ee
Inadomi and Terdiman (2006)	excluded	not ee
Jewel et al (2005)	excluded	not ee
Kam (2000)	excluded	not ee
Kay (2003)	excluded	not ee
Kennedy et al (2000)	excluded	not ee
Kennedy et al (2004)	excluded	not ee
Koelewijn et al (2006)	excluded	not ee
Leshno (2001)	excluded	not ee
Lichtenstein (2004)	excluded	not ee
Lichtenstein (2005)	excluded	not ee
Lichtenstein et al (2004)	excluded	not ee
Lichtenstein et al (2006)	excluded	not ee
Luces and Bodger (2006)	excluded	not ee
Marshall (2002a)	excluded	not ee
Mealy and Bayes (2005)	excluded	not ee
Mitton (2002)	excluded	not ee
Nahar et al (2003)	excluded	not ee
Nash and Florin (2005)	excluded	not ee
Odes et al (2006)	excluded	not ee
Ollendorf and Lidsky (2006)	excluded	not ee
Rubenstein et al (2002)	excluded	not ee
Rutgeerts et al (2004)	excluded	not ee
Sartor (2004)	excluded	not ee
Siegel et al (2006)	excluded	not ee
Silverstein et al (1999)	excluded	not ee
Strong (2001)	excluded	not ee
Thaler et al (2005)	excluded	not ee
van Balkom et al (2002)	excluded	not ee
Wicks (2002)	excluded	not ee
Williams and Meyers (2002)	excluded	not ee

REFERENCES

1 Farrell RJ, Alsahli M, Jeen Y-T, Falchuk KR, Peppercorn MA, Michetti P. Intravenous hydrocortisone premedication reduces antibodies to infliximab in Crohn's disease: A randomized controlled trial. *Gastroenterology* 2003; **124**(4):917-924.

2 Lemann M, Mary J, Duclos B, Veyrac M, Dupas J, Delchier JC, *et al*. Infliximab Plus Azathioprine for Steroid-Dependent Crohn's Disease Patients: A Randomized Placebo-Controlled Trial. *Gastroenterology* 2006; **130**(4):1054-1061.

3 Rutgeerts P, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JF, *et al*. Comparison of Scheduled and Episodic Treatment Strategies of Infliximab in Crohn's Disease. *Gastroenterology* 2004; **126**(2):402-413.

4 Hanauer SB, Feagan BG, Lichtenstein GR, Mayer LF, Schreiber S, Colombel JF, *et al*. Maintenance infliximab for Crohn's disease: The ACCENT I randomised trial. *Lancet* 2002; **359**(9317):1541-1549.

5 Abbott Laboratories Ltd. Adalimumab (Humira®). *Submission to the National Institute for Health and Clinical Excellence* 2007;1-98.

6 Clark W, Raftery J, Song F, Barton P, Cummins C, Fry-Smith A, *et al*. Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn's disease. *Health Technology Assessment* 2003; **7**(3):1-67.

7 Marshall,JK. Infliximab for the treatment of Crohn's disease: a systematic review and cost utility analysis. (Abstract) *Canadian Coordinating Office for Health Technology Assessment* Technology report no 24

8 Jaisson-Hot I, Flourie B, Descos L, Colin C. Management for severe Crohn's disease: a lifetime cost-utility analysis. *International Journal of Technology Assessment in Health Care* 2004; **20**(3):274-279.

9 Jewell D. Crohn's Disease. *Medicine* 1998; **26**(9):87-92.

10 Carter MJ, Lobo AJ, Travis SP. On behalf of the IBD section of the British Society for Gatroenterology. Guidelines for the management of inflammatory bowel disease in adults. *Gut* 2004; **53** (Suppl V):v1-v16.

11 Hanauer SB, Sandborn W. Management of Crohn's disease in adults. *Am J Gastroenterology* 2001; **96**(3):635-643.

12 Jenkins HR. Inflammatory bowel disease. *Arch Dis Child* 2001; **85**:435-437.

13 Henriksen M, Jahnsen J, Lygren I, Aadland E, Schulz T, Vatn MH, *et al*. Clinical course in Crohn's disease: results of a five-year population-based follow-up study (the IBSEN study). *Scandinavian Journal of Gastroenterology* 2007; **42**:602-610.

14 Vermeire S, Van Assche G, Rutgeerts P. Review article: altering the natural history of Crohn's disease-evidence for and against current therapies. *Aliment Pharmacol Ther* 2006; **25**:3-12.

15 Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007; **369**(9573):1641-1657.

16 Hanauer SB, Meyers S. Managment of Crohn's disease in adults. *Am J Gastroenterology* 1997; **92**:559-565.

17 Irvine EJ. Usual therapy improves perianal Crohn's disease as measured by a new Disease Activity Index. *J Clin Gastroenterol* 1995; **20**(1):27-32.

18 Loftus EV. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* 2004; **126**:1504-1517.

19 Knigge KL. Inflammatory bowel disease. *Clin Cornerstone* 2002; **4**(4):49-60.

20 Gurudu S, Fiocchi C, Katz JA. Inflammatory bowel disease. *Best Practice & Research Clinical Gastroenterology* 2002; **16**(1):77-90.
21 Felley C, Mottet C, Juillerat P, Froehlich F, Burnand B, Vader JP, *et al*. Fistulizing Crohn's disease. *Digestion* 2005; **71**(1):26-28.
22 Munkholm P, Langholtz E, Davidsen M, Binder V. Disease activity courses in a regional cohort of Crohn's disease patients. *Scandinavian Journal of Gastroenterology* 1995; **30**:699-706.
23 Silverstein MD, Loftus EV, Sandborn WJ, Tremaine WJ, Feagan BG, Nietert PJ, *et al*. Clinical course and costs of care for crohn's disease: markov model analysis of a population-based cohort. *Gastroenterology* 1999; **117**:49-57.
24 Russel MGVM. Changes in the incidence of inflammatory bowel disease: what does it mean? *European Journal of Internal Medicine* 2000; **11**:191-196.
25 Hanauer SB. Inflammatory bowel disease: epidemiology, pathogenesis, and therapeutic opportunities. *Inflamm Bowel Dis* 2006; **12 (Suppl 1)**:S3-S9.
26 Ehlin AGC, Montgomery SM, Ekbom A, Pounder RE, Wakefield AJ. Prevalence of gastrointestinal diseases in two British national birth cohorts. *Gut* 2003; **52**:1117-1121.
27 Rubin,GP, Hungin,APS, Kelly,PJ, Ling,J. Inflammatory bowel disease: epidemiology and management in an English general practice population.
28 National Association for Colitis and Crohn's Disease. Inflammatory bowel disease basics. Access date: 25 Apr. 7 A.D., URL: http://nacc.org.uk/content/ibd.asp
29 Shivananda J, Lennard-Jones J, Logan R, Fear N, Price A, Carpenter L, *et al*. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European collaborative study on inflammatory bowel disease (EC-IBD). *Gut: The Journal of the British Society of Gastroenterology* 1996; **39**:690-697.
30 Stone MA, Mayberry JF, Baker R. Prevalence and management of inflammatory bowel disease: a cross-sectional study from central England. *European Journal of Gastroenterology & Hepatology* 2003; **15**(12):1275-1280.
31 Yapp TR, Stenson R, Thomas GAO, Lawrie BW, Williams GT, Hawthorne AB. Crohn's disease incidence in Cardiff from 1930: an update for 1991-1995. *European Journal of Gastroenterology & Hepatology* 2000; **12**(8):907-911.
32 National Association for Colitis and Crohn's Disease. Information Sheets, FAQs and Drug Treatment Information. Access date: 7 Nov. 7 A.D., URL: http://www.nacc.org.uk/content/services/infoSheets.asp
33 Blondel-Kucharski F, Chircop C, Marquis P, Cortot A, Baron F, Gendre J, *et al*. Health-related quality of life in crohn's disease: a prospective longitudinal study in 231 patients. *American Journal of Gastroenterology* 2001; **96**(10):2915-2920.
34 Bassi A, Dodd S, Williamson P, Bodger K. Cost of illness of inflammatory bowel disease in the UK: a single centre retrospective study. *Gut* 2004; **53**:1471-1478.
35 Sostegni R, Daperno M, Scaglione N, Lavagna A, Rocca R, Pera A. Review article: Crohn's disease: monitoring disease activity. *Aliment Pharmacol Ther* 2003; **17** (Suppl 2):11-17.
36 Best WR, et al. Development of a Crohn's Disease Activity Index. National Co-operative Crohn's disease study. *Gastroenterology* 1976; **70**:439-444.
37 Yoshida EM. The Crohn's Disease Activity Index, its derivatives and the Inflammatory Bowel Disease Questionnaire: a review of instruments to assess Crohn's disease. *Can J Gastroenterol* 1999; **13**(1):65-73.

38 Sandborn WJ, Feagan BG, Hanauer SB, Lochs H, Lofberg R, Modigliani R, *et al*. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. *Gastroenterology* 2002; **122**:512-530.

39 Best WR. Predicting the Crohn's Disease Activity Index from the Harvey-Bradshaw Index. *Inflamm Bowel Dis* 2006; **12**(4):304-310.

40 Irvine EJ, Feagan BG, Rochon J, Archambault A, Fedorak RN, Groll A, *et al*. Quality of Life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. *Gastroenterology* 1994; **106**:287-296.

41 Mary JY, Modigliani R. Development and validation of an endoscopic index for the severity for Crohn's disease: a prospective multicentre study. *Gut* 1989; **30**:983-989.

42 Hyams J, Crandall W, Kugathasan S, Griffiths A, Olson A, Johanns J, *et al*. Induction and Maintenance Infliximab Therapy for the Treatment of Moderate-to-Severe Crohn's Disease in Children. *Gastroenterology* 2007; **132**(3):863-873.

43 Baldassano R, Braegger CP, Escher JC, DeWoody K, Hendricks DF, Keenan GF, *et al*. Infliximab (REMICADE) therapy in the treatment of pediatric Crohn's disease. *American Journal of Gastroenterology* 2003; **98**(4):833-838.

44 Joint Formulary Committee. British National Formulary. *London: British Medical Association and Royal Pharmaceutical Society of Great Britain* 2008.

45 Loftus EV, Schoenfeld P, Sandborn WJ. The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. *Aliment Pharmacol Ther* 2002; **16**:51-60.

46 Royal College of Physicians. Unacceptable variation in care for people with inflammatory bowel disease. Access date: 22 May 7 A.D., URL: http://www.rcplondon.ac.uk/news/news.asp?PR_id=349

47 Schreiber S, Rutgeerts P, Fedorak RN, Khaliq-Kareemi M, Kamm MA, Boivin M, *et al*. A randomized, placebo-controlled trial of certolizumab pegol (CDP870) for treatment of Crohn's disease. *Gastroenterology* 2005; **129**:807-818.

48 Calabrese L. The yin and yang of tumor necrosis factor inhibitors. *Cleveland Clinic Journal of Medicine* 2006; **73**(3):251-256.

49 European Medicines Agency. Summary of Product Characteristics (Humira). Access date: 10 Oct. 7 A.D., URL: http://www.emea.eu.int/humandocs/PDFs/EPAR/Humira/H-481-PI-en.pdf

50 European Medicines Agency. Summary of Product Characteristics (Remicade). Access date: 10 Oct. 7 A.D., URL: http://www.emea.eu.int/humandocs/PDFs/EPAR/Remicade/H-240-PI-en.pdf

51 Centocor,I. Medication Guide Remicade. Access date: 10 Oct. 7 A.D., URL: http://www.remicade.com/remicade/assets/Med-Guide.pdf

52 Abbott Laboratories. Important Safety Information Humira. Access date: 10 Oct. 7 A.D., URL: http://www.humira.com

53 Han PD, Cohen RD. Managing immunogenic responses to infliximab. *Drugs* 2004; **64**(16):1767-1777.

54 Targan S, Hanauer SB, van Deventer SJH, Mayer L, Present DH, Braakman T, *et al*. A short term study of chimeric monoclonal antibody cA2 to tumour necrosis factor alpha for Crohn's disease. *New England Journal of Medicine* 1997; **337**(15):1029-1035.

55 Rutgeerts P, D'Haens G, Targan S, Vasiliauskas E, Hanauer SB, et al. Efficacy and Safety of retreatment with anti-tumor necrosis factor antibody (infliximab) to maintain remission in Crohn's Disease. *Gastroenterology* 1999; **117**:761-769.

56 Hommes DW, Baert F, Van Assche G, Caenepeel F, Vergauwe P, Tuynman H, *et al*. Management of recent onset Crohn's disease: a controlled, randomised trial

comparing step-up and top-down therapy (Abstract). *Gastroenterology* 2005; **129**(371).

57 Present DH, Rutgeerts P, Targan S, Hanauer SB, Mayer L, van Hogezand Rea. Infliximab for the treatment of fistulae in patients with Crohn's disease. *New England Journal of Medicine* 1999; **340**:1398-1405.

58 Hanauer SB, Sandborn WJ, Rutgeerts P, Fedorak RN, Lukas M, MacIntosh D, *et al*. Human anti-tumor necrosis factor monoclonal antibody (adalimumab) in Crohn's disease: The CLASSIC-I trial. *Gastroenterology* 2006; **130**(2):323-332.

59 Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Colombel JF, Panaccione R, *et al*. Adalimumab induction therapy for crohn disease previously treated with infliximab: a randomized trial. *Ann Intern Med* 2007; **146**(12):829-838.

60 Sands BE, Anderson FH, Bernstein CN, Chey WY, Feagan BG, Fedorak RN, *et al*. Infliximab Maintenance Therapy for Fistulizing Crohn's Disease. *New England Journal of Medicine* 2004; **350**(9):876-885.

61 Sandborn WJ, Hanauer SB, Rutgeerts PJ, Fedorak RN, Lukas M, Macintosh DG, *et al*. Adalimumab for Maintenance Treatment of Crohn's Disease: Results of the CLASSIC II Trial. *Gut* 2007.

62 Colombel J, Sandborn WJ, Rutgeerts P, Enns R, Hanauer SB, Panaccione R, *et al*. Adalimumab for Maintenance of Clinical Response and Remission in Patients With Crohn's Disease: The CHARM Trial. *Gastroenterology* 2007; **132**(1):52-65.

63 D'Haens G, van Deventer S, van Hogezand R, Chalmers D, Kothe C, Baert Fea. Endoscopic and histological healing with infliximab anti-tumour necrosis factor antibodies in Crohn's disease. *Gastroenterology* 1999; **116**:1029-1034.

64 Egger M, Davey Smith G, Altman DG. Systematic Reviews in Health Care; meta-analysis in context. *2'nd edition* 2001; BMJ Books**.**

65 Wu EQ, Yu A, Atanasov P, Tang J, Pollack P, Lomax K, *et al*. Lower disease activity and clinical remission are associated with reduced hospitalisation risk in Crohn's disease. *ISPOR 12th Annual international Meeting, 19-23 May, Arlington, USA* 2007.

66 Feagan B, Panaccione R, Sandborn W, D'Haens G, Schreiber S, Rutgeerts P, *et al*. An evaluation of adalimumab on the risk of hospitalisation in patients with Crohn's disease, data from CHARM. *Digestive Disease Week (DDW), 19-24 May 2007, Washington DC, USA* 2007.

67 Behm BW, Bickston SJ. Tumor necrosis factor-alpha antibody for maintenance of remission in Crohn's disease. *Cochrane Database Syst Rev* 2008;(1):CD006893.

68 Sands BE, Blank MA, Patel K, Van Deventer SJ. Long-term treatment of rectovaginal fistulas in Crohn's disease: Response to infliximab in the ACCENT II study. *Clinical Gastroenterology & Hepatology* 2004; **2**(10):912-920.

69 Sands BE, Blank MA, Diamond RH, Barrett JP, Van Deventer SJ. Maintenance infliximab does not result in increased abscess development in fistulizing Crohn's disease: Results from the ACCENT II study. *Alimentary Pharmacology & Therapeutics* 2006; **23**(8):1127-1136.

70 Su C, Lichtenstein G, Krok K, Brensinger CM, Lewis JD. A meta-analysis of the placebo rates of remission and response in clinical trials of active Crohn's disease. *Gastroenterology* 2004; **126**:1257-1269.

71 Chang M, Chow S-C. Adaptive design methods in clinical trials - a review. *Orphanet journal of rare diseaeses* 2008; **3**:11-36.

72 Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al*. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. 2006;10(42). *Health Technol Assess* 2006; **10**(42):1-248.

73 Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R. Indirect comparisons of competing interventions. *Health Technology Assessment* 2005; **9**(26):1-148.

74 Brodsky RA, Young NS, Antonioli E, Risitano AM, Schrezenmeier H, Schubert J, *et al*. Multicenter phase 3 study of the complement inhibitor eculizumab for the treatment of patients with paroxysmal nocturnal hemoglobinuria. *Blood* 2008; **111**(4):1840-1847.

75 Drummond M.F., Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* 1996; **313**:275-283.

76 Arseneau KO, Cohn SM, Cominelli F, Connors AFJ. Cost-utility of initial medical management for Crohn's disease perianal fistulae. *Gastroenterology* 2001; **120**:1640-1656.

77 Gregor J, McDonald JWD, Klar N, Wall R, Atkinson K, Lamba B. An evaluation of utility measure in Crohn's disease. *Inflamm Bowel Dis* 1997; **3**:265-276.

78 Organisation for economic co-operation and development. Statistics Portal. Access date: 15 Nov. 7 A.D., URL: http://www.oecd.org/statistics

79 Schering-Plough Ltd. Remicade® in the treatment of Crohn's disease in the United Kingdom. *Submission to the National Institute for Health and Clinical Excellence* 2007;1-67.

80 Wolters FL, Russel MG, Sijbrandij J, et al. Crohn's disease: increased mortality 10 years after diagnosis in a Europe-wide population based cohort. *Gut* 2007; **55**:510-518.

81 Jess T, Riis L, Vind I, Winther KV, Borg S, Binder V, *et al*. Changes in clinical characteristics, course and prognosis of inflammatory bowel disease during the last 5 decades: a population-based study from Copenhagen, Denmark. *Inflamm Bowel Dis* 2007; **13**(4):481-489.

82 Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, *et al*. Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technology Assessment* 2006; **10**(31):1-258.

83 Casellas F, Arenas JI, Baudet JS, Fabregas S, Garcia N, Gelabert J, *et al*. Impairment of health-related quality of life in patients with inflammatory bowel disease: a Spanish multicenter study. *Inflammatory Bowel Diseases* 2005; **11**(5):488-496.

84 Dolan P, Gudex C, Kind P, Williams A, Centre for Health Economics YUoY. A social tariff of EuroQol: results from a UK general population survey. *Discussion Paper no 138* 1995.

85 NHS Reference cost database 2005/6. *http://www dh gov uk/en/publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_062884 HRG FF51,FF5* 2008; Accessed June 2008.

86 Curtis L, Netten A. Unit costs of of health and social care. *Personal Social Services Research Unit, UK* 2006;1-230.

87 TreeAge Software I. TreeAge Pro 2006 Health Care 1.2. *Williamstown MA, USA* 2006.

88 Wooltorton E. Paroxetine (Paxil, Seroxat): increased risk of suicide in pediatric patients. *CMAJ* 2003; **169**(5):446.

89 Rosenbaum PL, Saigal S. Measuring Health-Related Quality of Life in Paediatric Populations: Conceptual Issues. *In: Quality of life and Pharmacoeconomics in Clinical Trials* 1996 Editor**:** B Spilker.

90 Harris M, Buttterworth G. Developmental Psychology: A student's handbook. *Psychology Press* 2002; Hove.

91 Matza LS, Swensen AR, Flood EM, Secnik K, Leidy NK. Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value Health* 2004; **7**(1):79-92.
92 Fox-Rusby J, Parker M. Culture and the measurement of health-related quality of life. *Revue Europeenne de Psychologie Appliquee* 1995; **45**(4):257-263.
93 Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. *Health Econ* 2003; **12**(8):697-702.
94 Petrou S, Henderson J. Preference-based approaches to measuring the benefits of perinatal care. *Birth* 2003; **30**(4):217-226.
95 Hansard source. NHS Drug Budget. Access date: 29 Oct. 7 A.D., URL: www.theyworkforyou.com
96 Moore TJ, Cohen MR, Furberg CD. Serious adverse drug events reported to the Food and Drug Administration, 1998-2005. *Arch Intern Med* 2007; **167**(16):1752-1759.
97 The Information Centre N. Hospital Prescribing, 2006: England. *Statistical Bulletin IC 2007* 2007; **12**:1-24.
98 Hanauer SB. Turning traditional treatment strategies on their heads: Current evidence for 'step-up' versus 'top-down'. *Reviews in Gastroenterological Disorders* 2007; **7**(SUPPL. 2):S17-S22.