

**USE OF TUMOUR NECROSIS FACTOR ALPHA (TNF A)  
INHIBITORS (ADALIMUMAB AND INFLIXIMAB ) FOR CROHN'S  
DISEASE**

**REPORT BY THE DECISION SUPPORT UNIT**

Allan Wailoo, Jon Tosh  
School of Health and Related Research, University of Sheffield

11 January 2009

## CONTENTS

<b>1. INTRODUCTION.....</b>	<b>3</b>
<b>2. COMPARISON OF ECONOMIC MODELS – STRUCTURAL ISSUES....</b>	<b>4</b>
2.1. THE DEFINITION OF RESPONSE.....	4
2.2. UTILITY VALUES .....	5
2.2.1. <i>The Leeds model</i> .....	5
2.2.2. <i>The Bodger model</i> .....	6
2.2.3. <i>The Schering Model</i> .....	8
2.2.4. <i>The Abbott Model</i> .....	8
2.2.5. <i>Summary on utility estimates</i> .....	8
2.3. TIME HORIZON.....	9
2.3.1. <i>Time horizon and the Leeds model</i> .....	9
2.3.2. <i>Time horizon and other models</i> .....	10
2.4. MODELLING STANDARD CARE.....	10
2.4.1. <i>The Silverstein cohort</i> .....	10
2.5. TREATMENT EFFECT.....	12
<b>3. SUMMARY .....</b>	<b>14</b>
3.1. STRUCTURAL MODELLING ISSUES .....	14
3.2. RECOMMENDATIONS .....	15
<b>4. REFERENCES.....</b>	<b>15</b>

### Figures

<i>Figure 1: Comparison of alternative utility models</i> .....	8
<i>Figure 2: ICER for maintenance adalimumab (severe patients) vs standard care as a function of time horizon in the Leeds model</i> .....	10
<i>Figure 3: Comparison of remission rates – Leeds and SP models</i> .....	11
<i>Figure 4: Comparison of combined remission rates – Leeds and SP models</i> .....	12

### Tables

<i>Table 1: Baseline costs, effects and ICERs from Leeds models</i> .....	3
---	---

# 1. INTRODUCTION

The aim of this report is to consider submitted cost effectiveness models for the appraisal of the anti-TNFs, adalimumab and infliximab, for the treatment of severe active Crohn's disease. Its focus is the independent assessment group models (the "Leeds model") for adult, non-fistulising Crohn's disease. The structure of the models used for this population are similar to that used to estimate the cost effectiveness of anti-TNF therapy in fistulising and paediatric Crohn's. Other submissions are used to provide comparisons. In particular, the two manufacturer submissions (the Schering Plough model and the Abbott model) are considered. A further independent analysis funded by the Welsh Office for Research and Development (the Bodger model) is also included in the review. The DSU did not have access to this model and as such the review is based only on the written account of the model. Furthermore, this report was provided to the Institute on an academic-in-confidence basis. As a consequence, all discussion of the Bodger model in this report should also be considered academic-in-confidence.

All comments received from stakeholders during consultation, and responses from the assessment group were taken into account in informing this review.

This review is based on the written account of the cost effectiveness analysis provided in the assessment report circulated to the appraisal committee in August 2008 (Assessment Report 2) and electronic versions of the TreeAge Pro model dated 07 August 08.

The deterministic results of the unaltered models for severe disease supplied to the DSU were as follows:

**Table 1: Baseline costs, effects and ICERs from Leeds models**

<b>Strategy</b>	<b>Cost</b>	<b>Incr Cost</b>	<b>Eff</b>	<b>Incr Eff</b>	<b>C/E</b>	<b>Incr C/E (ICER)</b>
<i>Infliximab severe</i>						
Episodic IXB	12,026		0.8948		13,440	
Standard Care	13,418	1,392	0.8121	-0.0827	16,523	(Dominated)
Maintenance IXB	19,138	7,112	0.8962	0.0014	21,354	5,024,522
<i>Adalimumab severe</i>						

Episodic ALB	7,037		0.8949		7,863	
Standard Care	13,418	6,381	0.8121	-0.0828	16,523	(Dominated)
Maintenance ALB	14,042	7,005	0.8963	0.0014	15,667	4,949,900

## 2. COMPARISON OF ECONOMIC MODELS – STRUCTURAL ISSUES

### 2.1. THE DEFINITION OF RESPONSE

The most widely used measure of disease activity in all the economic models, as well as the clinical trials, is the Crohn’s Disease Activity Index (CDAI).

In the Leeds model, the only means for a treatment to generate benefit is by moving patients to a remission state, defined as a CDAI score less than 150. Despite the fact that clinical trials typically report both partial response measures: CDAI70 and CDAI100 (a reduction of 70 and 100 points on the CDAI scale respectively) and mean CDAI as well as remission rates, the Leeds model does not include these potential treatment benefits. This is because the authors did not consider it feasible to “ascribe a robust utility value” (p.191) to improvements in CDAI short of remission. Utility gain from health improvements may be dependent on pre-treatment CDAI and, since no study reports the calculation of utility values adjusting for pre treatment health, the model includes only remission as the measure of treatment response. It is therefore clear that the Leeds model potentially underestimates the benefits of anti-TNF therapy.

[REDACTED]

Schering Plough also only include full remission as the measure of benefit in their model by distinguishing “active” and “remission” health states. No rationale is given for this approach.

The Abbott model includes the full remission state (CDAI < 150), along with a moderate (150 < CDAI < 300) state, a severe (300 < CDAI < 450) state, and a very severe (CDAI > 450) state. These are based on the Gregor et al. CDAI-interval defined disease states and are derived from the CHARM trial for the adalimumab arm. A probit regression derives the standard care arm from the CLASSIC I trial. The model estimates the average time spent in health states, as derived from the trial, and attaches health state utility values which when multiplied by the time spent in each state and summed, provides an estimate of the total QALY/benefit of the treatment. Abbott explain that the moderate health state allows the analysis of a moderate/severe patient population included in the clinical trials, as well as a severe patient population (the required population for the licensed indication of adalimumab).

## **2.2. UTILITY VALUES**

There are several issues relating to the use of utility values in the various models. First, different sources provide different estimates of the utility of a patient spending a period in health states which are similar or identical in terms of their clinical description. Second, the use of different sources permit modelling approaches which differ in terms of structure to be parameterised. In particular, it could be argued that in order to construct a model that reflects partial response to treatment requires an appropriate utility estimate. This issue is also discussed.

### *2.2.1. The Leeds model*

The health states “remission”, “relapse – moderate disease” and “relapse – severe disease”, defined in the Leeds model, are assigned utility values based on a study by Gregor et al. (1997). These values were obtained from a sample of patients with Crohn’s disease (n=180) who were asked to value three vignettes using time trade-off

(TTO), standard gamble (SG) and visual analogue scale (VAS) on two separate occasions. The Leeds model uses the first set of TTO values. The vignettes described “mild”, “moderate” and “severe” disease.

This approach to utility estimation has several potential weaknesses.

First, the description of disease in Gregor et al. has no direct relationship either with the intended states in the Leeds model (defined in terms of CDAI), or with the states from the original source for transition probabilities (the Silverstein study). There is an element of subjectivity in the translation of results between studies and model states.

Second, the valuations are derived from Crohn’s disease patients and not the general population as is required for reference case analyses (p.38 NICE Guide to the Methods of Technology Appraisal). There is a well documented tendency for patients to value health states higher than the general population and therefore departures from the reference case should be undertaken and interpreted with caution.

Third, and related to the previous point, the values obtained in this study are extremely high and of questionable validity. For example, a mean utility of 0.73 (95% CI 0.69 to 0.76) is used for a state described as “frequently experiences episodes of severe abdominal pain”, “always tired, has difficulty sleeping, feels depressed and frustrated”, and “unable to attend work or school or participate in social activities”.

Fourth, there are vast differences between the values obtained in the Gregor et al. study according to the valuation method. The time-trade-off (TTO) values used in the analysis yield the highest estimates for each state. Whilst the TTO method may be preferable in order to maintain consistency with the methods used in the EQ5D, the variation in results according to method do indicate a degree of uncertainty that is not currently reflected in the analysis.

### 2.2.2. *The Bodger model*

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

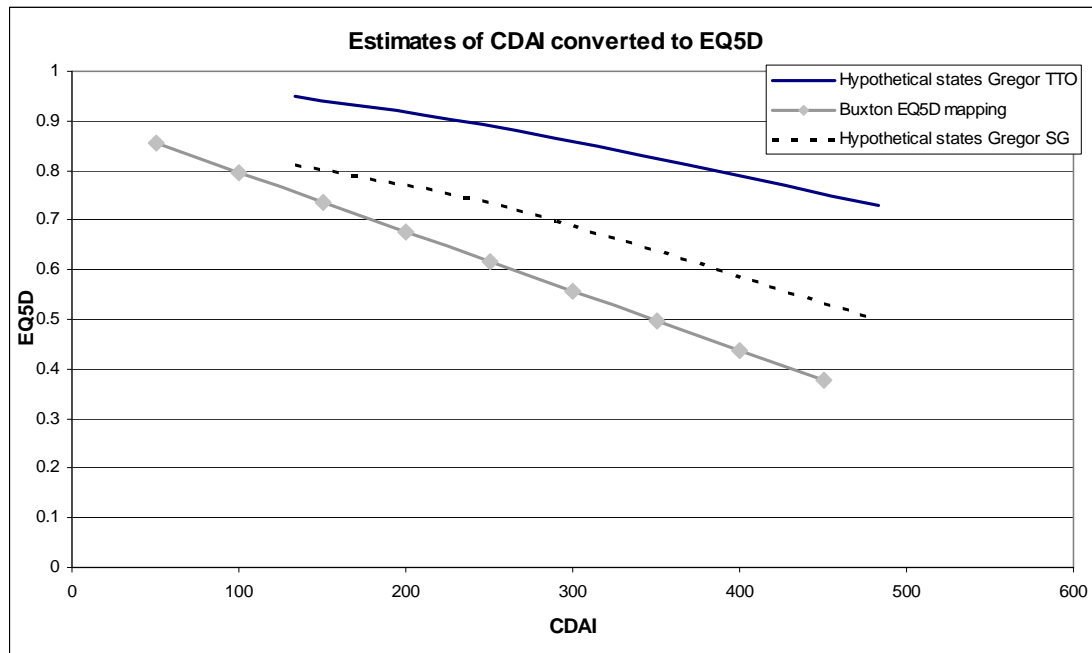
[REDACTED]

[REDACTED]

The analysis demonstrates that EQ-5D is moderately to strongly correlated with CDAI ( $r=-0.62$ ) and this is a statistically significant relationship. Only a simple linear regression is fitted for CDAI and only measures of goodness of fit ( $R^2$  and mean absolute percentage error) are reported. There is no further information provided to judge the performance of this model. The authors conclusion implies that the regression is not useful. However, this is a questionable conclusion in the context of the current appraisal. It appears that comparisons are made between models using CDAI as an explanatory variable versus those which use Inflammatory Bowel Disease Questionnaire (IBDQ) on the basis of the  $R^2$  statistic. It is questionable whether statistical models can be compared in this manner but in any event the “value” of the relationship must be judged in the context of the decision problem in hand. In particular, it is CDAI that is widely reported in the clinical trials and therefore the performance of regressions that consider EQ5D as a function of IBDQ is not relevant. In addition, there appears to be some inconsistency between the simple correlation coefficient and the  $R^2$  statistic for the relevant model. We would agree with the claim of the authors that the relationship between IBDQ and EQ5D has a better fit than the regression of CDAI against EQ5D. It cannot be claimed that the latter is invalid because of a lower  $R^2$ .

The impact of the Gregor et al. TTO values, versus Gregor et al. SG and Buxton et al. mapping approaches are shown in Figure 1. The CDAI scores are those estimated by the authors based on a database of patients that had participated in previous clinical trials. The plot illustrates that the Gregor TTO method results in the highest valuations for each health states and has the shallowest slope i.e. the smallest benefit in terms of health utility as a result of improvement in CDAI.

**Figure 1: Comparison of alternative utility models**



### 2.2.3. The Schering Model

The SP model is based mainly on a subset of data published in their entirety in Casellas et al. (2005). The data are from Spanish CD patients who completed the EQ5D and the UK tariff was applied to estimate utilities. The data used in the model differ from those published although the precise reasons for using a subset of patients (n=201) compared to those in the published study (n=628) is not stated. In general however, the approach taken to estimating utility values for remission and active disease meet the reference case requirements.

### 2.2.4. The Abbott Model

The Abbott submission states that their model uses health state utility values based on the Gregor et al. (1997) published study. The Gregor et al data are from Canadian CD patients (n=180) whose health states were valued using standard gamble methods, however it is not clear how Abbott have derived their utility values from the Gregor study. As well as standard gamble values, the Gregor et al study also reports utility scores using time-trade off methods, although Abbott report that they selected the standard gamble values due to it conforming to utility theory.

### 2.2.5. Summary on utility estimates

The source for estimating utility values is of crucial importance because this directly influences model results even when the same states are modelled and also because the



model structure may be determined by the availability of this data. [REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]  
[REDACTED]

## 2.3. TIME HORIZON

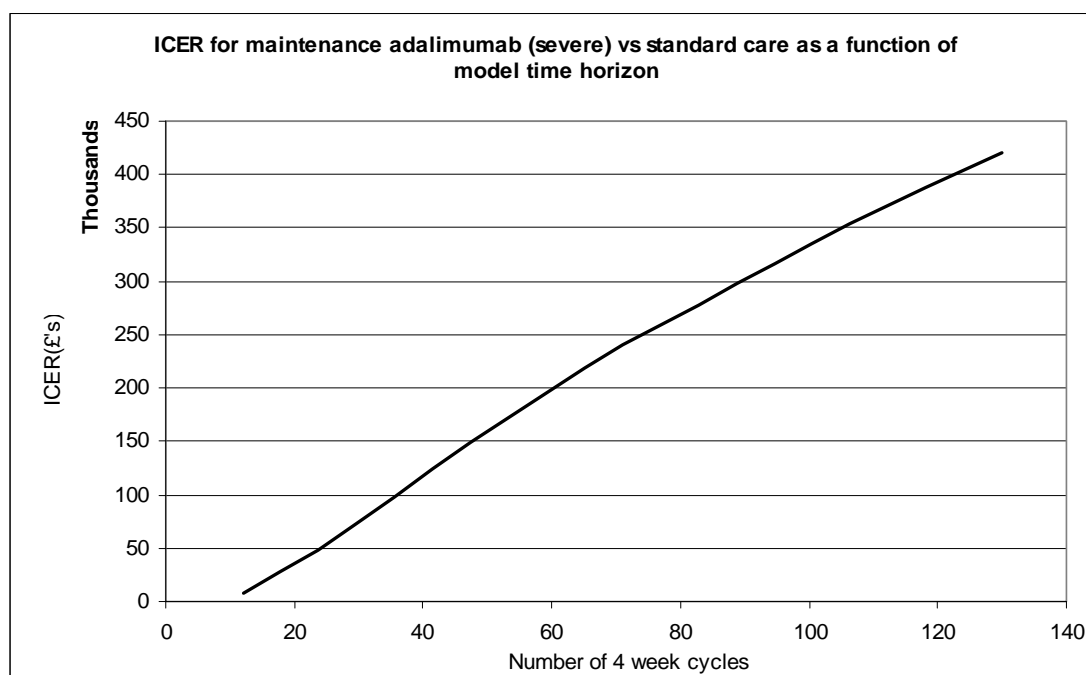
Analyses of cost effectiveness should adopt a time horizon that is sufficient to reflect any differences in costs or benefits between the treatments being compared. This is reflected in the NICE methods guide (p.27). However, the time horizons adopted by the various models differ substantially.

### 2.3.1. Time horizon and the Leeds model

The Leeds model adopts a one year time horizon on the basis that there is no impact of either disease or treatments on mortality. However, the lack of any impact on mortality is not a sufficient condition to ensure that a one year time horizon is sufficient to capture all differences in costs and benefits. Where the distribution of patients across different health states differs between the anti-TNF treated and standard care strategies, there will be continued differences in costs and/or benefits over the longer term. In this situation, a one year time horizon is unlikely to be sufficient.

In the Leeds model this distribution of patients does change and is not stable by one year (13 x 4 week cycles). Considering maintenance therapy alone, using the base case model for adalimumab, severe patients as an example, it can be demonstrated that the cost effectiveness results change substantially as a function of time horizon. Figure 2 demonstrates that using a one year time horizon, the ICER for maintenance therapy compared to standard care is £7410. As the model is run for longer periods, the ICER rises to £47808 at 24 cycles and to in excess of £420000 at 130 cycles (10 years).

**Figure 2: ICER for maintenance adalimumab (severe patients) vs standard care as a function of time horizon in the Leeds model**



### 2.3.2. Time horizon and other models

The SP model adopts a 1 year timeframe to estimate results within the clinical trials period, and extrapolated analyses that adopt a 5 year time horizon in the base case.

The Abbott model adopts a 1 year timeframe as its basecase analysis, although this was extrapolated to a lifetime timeframe assuming that the standard care and adalimumab patients had reached a steady state by week 56.

## 2.4. MODELLING STANDARD CARE

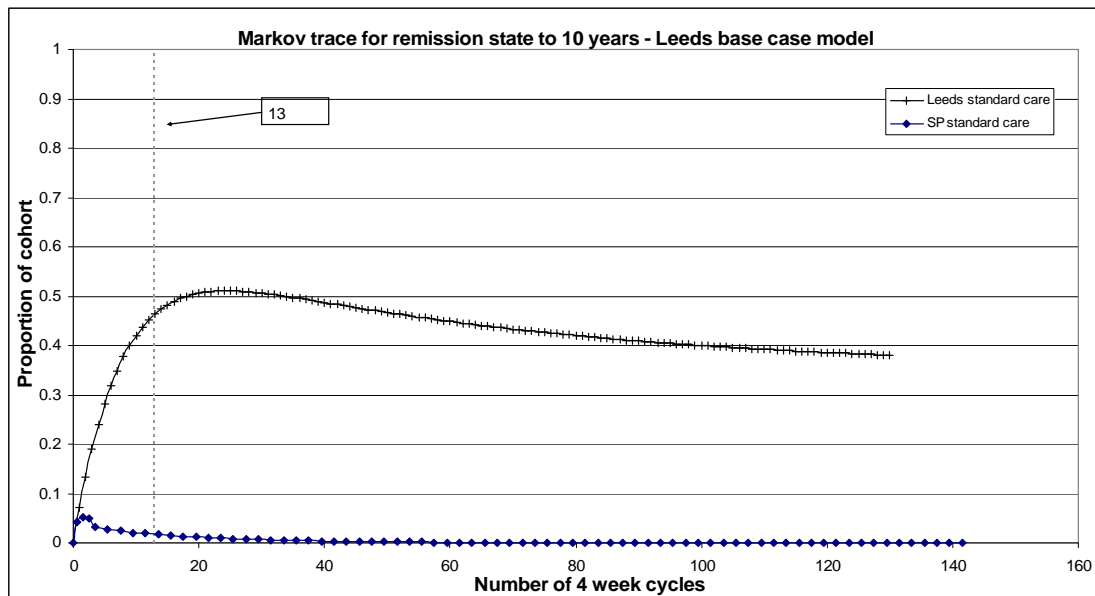
### 2.4.1. The Silverstein cohort

Both the Leeds and Bodger models are based on transition probabilities between health states that are derived from those published by Silverstein et al. (1999). This is a retrospective cohort study of all patients diagnosed with Crohn's disease between 1970 and 1993, resident in Olmsted County, Minnesota. A total of 174 patients were included ranging in age at diagnosis from 8.4 years to 83.8 years. This differs substantially from the population indicated for treatment with anti-tnfs: patients with severe, active disease who have not responded adequately to steroids and/or

immunosuppressants. This difference between the indicated population and that which forms the basis of the standard care costs and utilities, gives rise to a potential for the cost effectiveness of treatment to be underestimated. The use of the Silverstein dataset, without recognition of this potential bias and consideration of adjustments to deal with it, should be avoided.

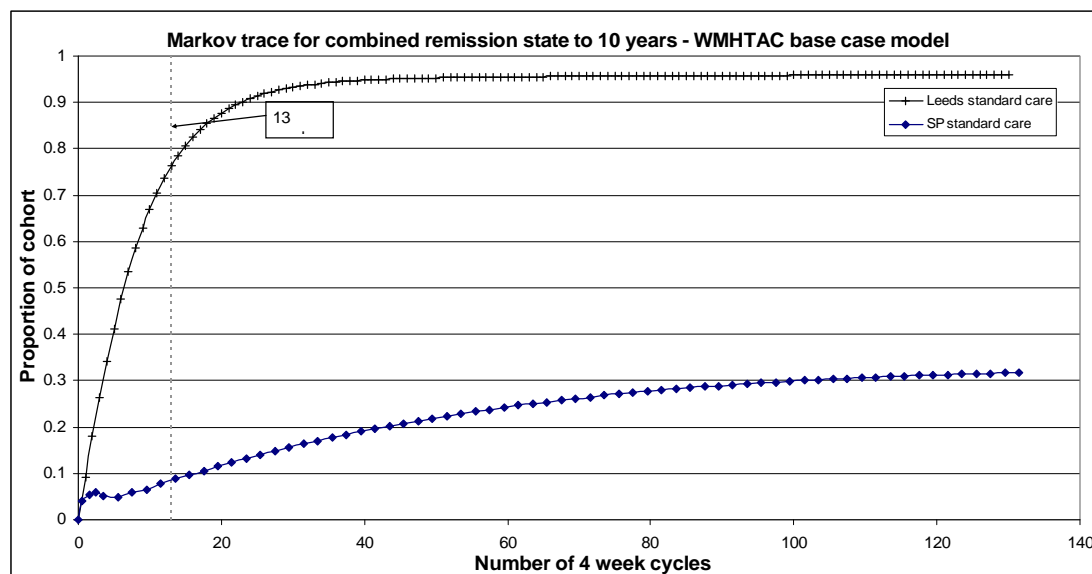
The impact of using the Silverstein data can be identified in the Markov trace from the Leeds model. Considering just standard care patients, the model starts with all patients in the relapse state in the model and it can be seen in Figure 3 that by 1 year, 46% are in remission. This percentage continues to rise up to 2 years where there are 51% of patients in remission. Contrast this with the Schering model in which transition probabilities are derived from the clinical trials where 2% are in remission at one year and this rate rapidly reduces.

**Figure 3: Comparison of remission rates – Leeds and SP models**



**Error! Not a valid bookmark self-reference.** shows a similar Markov trace but combines the remission and post surgical remission states. In the Leeds model, 76% of patients are in a combined remission state at year 1 and this percentage continues to rise up to 96%. The Schering model estimates a vastly lower number of patients in remission.

**Figure 4: Comparison of combined remission rates – Leeds and SP models**



The analysis serves to highlight the difference in the modelling approaches which are driven by using alternative sources to estimate disease progression in the absence of anti-TNF therapy. When the use of the Silverstein data is considered alongside the approach to reflecting treatment effect in the Leeds model, in subsequent sections, the implications for cost effectiveness can be further understood.

## 2.5. TREATMENT EFFECT

In the Leeds model, the anti-TNFs generate health benefits directly via two transition probabilities: from relapse (where all patients begin in the model) to remission, and from remission to relapse. The second probability is a simple function of the former. Whilst these probabilities are based on the Silverstein study for the standard care arm of the model, in the treatment arms (both episodic and maintenance) these probabilities are the absolute probabilities of response from the CHARM and ACCENT1 for adalimumab and infliximab respectively.

Furthermore, whilst the model is run for one year in the base case, the absolute probabilities are fixed using the 6 week response rates observed in the treatment arms of the trials. This is 0.56 for adalimumab (based on 6 week CHARM trial TNF arm events 96/172) and 0.56 for infliximab (based on 6 week ACCENT 1 TNF arm events 63/113).

These two approaches to modelling lead to some potential problems. The randomised nature of the clinical trial data is lost because standard care and anti-TNF arms are modelled independently from separate sources, rather than as a baseline and relative treatment effect. The approach also leads to a large treatment effect compared to the trial data. However, it must be recognised that the design and reporting of the key maintenance trials here is problematic and this has been extensively discussed in the assessment group report (p. 136). A particular feature of the trials and their reporting is the fact that all patients in these two trials received induction therapy and were then randomised. These studies do not, therefore, provide a relative treatment effect of no treatment compared to anti-TNF treatment.

[REDACTED]

Schering Plough use two RCTs to incorporate the treatment effectiveness of anti-TNF and standard care into their economic model. The Targan study (placebo arm for standard care, 5mg/kg arm for infliximab) was used for the initial induction period and initial response on treatment (0-2 weeks). The ACCENT I trial (placebo arm for standard care, 5mg/kg arm for infliximab) was used after 2 weeks for subsequent transitions. The absolute values of patients in the trials were used to populate the arms of the model, which were then converted to transition matrices. However it was not clear how the numbers had been derived from all of the published trial data.

Abbott also use two RCTs to incorporate the treatment effectiveness of anti-TNF and standard care into their economic model. The adalimumab arm of the model uses the CHARM trial and the CLASSIC I trial is used to estimate standard care outcomes.

CLASSIC I reports 4 weeks of standard care + placebo and these data were used in regression models to estimate the course of disease states over time. The probability of a patient being in a particular health state is estimated conditional on baseline CDAI, disease severity, previous exposure to anti-TNF and previous exposure to corticosteroids.

### **3. SUMMARY**

#### **3.1. STRUCTURAL MODELLING ISSUES**

There are some difficulties that arise from the approach adopted in the Leeds model that have been highlighted above. The assessment report and responses to consultation highlight that many judgements used to inform model development were based on particular views about available data and the numerous difficulties with many of the data sources. Nevertheless, there are several areas where alternative approaches may offer improvement.

First, the omission of treatment benefits less than full remission underestimates the benefits of anti-TNF therapy. A model structure that allows such benefits would be appropriate and would permit a fuller reflection of the benefits of therapy highlighted in the key clinical trials. [REDACTED]

Second, our interpretation of the regression model between EQ5D and CDAI, reported in Buxton et al. leads to different conclusions to those drawn by the assessment group. By using these results, a model based on CDAI could be developed and evaluated. This would allow the cost effectiveness model to reflect more completely the benefits of the alternative treatments and their use in clinical practice.

Third, any model based on the Silverstein cohort must recognise the substantial differences between the patient population included in this observational study and those indicated for anti-TNF therapy. Adjustments must be made to the transition probabilities accordingly. A failure to do so risks overestimating the extent to which patient benefit from standard treatments and biases against anti-TNF therapy. In

particular, the probabilities that patients in remission subsequently experience a relapse or require major surgery, may be substantially higher than observed in the Silverstein cohort. Both are potentially crucial drivers of cost effectiveness estimates.

Fourth, the use of an observational study to estimate transition probabilities for standard care and an absolute treatment effect from one arm of a trial for the treatment probabilities is a concern. The randomised nature of the trial data are ignored and the resulting differences between the estimates of effectiveness for anti-TNF treated patients versus standard care patients may be biased as a result.

### **3.2.RECOMMENDATIONS**

Whilst it is straightforward to make amendments to the Leeds model that reflect longer time horizons, fixed term maintenance therapy and a higher probability of relapse associated with standard care, many other important issues are more complex to implement. In addition, detailed reviews and comparisons with the other submitted modelling approaches may be required. The results of implementing partial adaptations of the Leeds model may be misleading and are therefore not presented here.

A full reconciliation between the models submitted to this appraisal may be required in order to develop estimates of cost effectiveness which are reliable and appropriately reflect areas of uncertainty. This reconciliation requires that structural issues, as well as individual parameter values, are fully considered and amended where appropriate.

## **4. REFERENCES**

Bodger, K., Kikuchi, T., and Hughes, D. (unpublished manuscript) Cost-effectiveness of biological therapy for Crohn's disease: Markov cohort analyses incorporating United Kingdom patient level cost data. ACADEMIC IN CONFIDENCE

Buxton, M.J., Lacey, L.A., Feagan, B.G., Niecko, T., Miller, D.W., and Townsend, R.J. (2007) Mapping from disease specific measures to utility: An analysis of the relationship between the inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's Disease and measures of utility, *Value in Health*, Vol.10(3):214-220.

Casellas, F., Arenas, J. I., Baudet, J. S., Fabregas, S., Garcia, N., Gelabert, J., et al. (2005) Impairment of Health-related Quality of Life in Patients with Inflammatory Bowel Disease: A Spanish Multicenter Study, *Inflammatory Bowel Disease*, Vol.11(5):488-496.

Gregor, J.C., McDonald, J.W.D., Klar, N., Wall, R., Atkinson, K., Lamba, B., and Feagan, B.G. (1997) An evaluation of utility measurement in Crohn's disease, *Inflammatory Bowel Disease*, Vol.3(4):265-76.

Silverstein, M.D., Loftus, E.V., Sandborn, W.J., Tremaine, W.J., Feagan, B.G., Nietert, P.J., Scott Harmsen, W.S., and Zinsmeister, A.R. (1999) Clinical course and costs of care for Crohn's disease: Markov model analysis of a population –based cohort, *Gastroenterology*, Vol.117:49-57.