

**REVIEW OF SOFTWARE FOR DECISION MODELLING.**

DECISION SUPPORT UNIT

Jon Tosh and Allan Wailoo

Health Economics and Decision Science, School of Health and Related Research,  
University of Sheffield

29 July 2008

## 1. INTRODUCTION

This report provides a review of the use of software packages for decision modelling as part of the NICE Technology Appraisal (TA) process. Current NICE guidance requires models to be submitted using “standard software (for example Excel (without Visual Basic Programming) or DATA<sup>1</sup>)”(NICE, 2004)<sup>1</sup>. This requirement is reiterated in the Single Technology Appraisal (STA) Specification for manufacturer/sponsor submission of evidence update<sup>2</sup>:

*“The model should be constructed using standard software, such as Excel or DATA. If non-standard software is required for the construction of the model, please discuss this with the Institute at the earliest opportunity in advance of submission.” page 3*

The aim of this review is to provide advice for updating that guidance. The objectives are to assess the frequency with which particular packages are used (both in NICE submissions and other assessments), identify the capacity for assessment groups and submitting organisations to build and/ or review in different packages, and to consider the key strengths and weaknesses of the various software packages available.

We conducted a survey of key stakeholders submitting to the technology appraisal process (Technology Assessment Groups (TAGs)/Evidence Review Groups (ERGs), manufacturers and consultancies). Survey methods are described in the following section followed by results from the survey. A brief review of the key features of each of the software packages identified in the survey is then provided. Issues for consideration are presented in the final section.

## 2. SURVEY

### a. Methods

An online survey was developed and an invitation to participate sent by email to a selection of organisations that submit models as part of a NICE TA. This included all seven TAGs, all consultancy firms that had disclosed that they, not a manufacturer, had produced a model, and all manufacturers from the post 9<sup>th</sup> wave list of TAs that

---

<sup>1</sup> It is worth noting that “DATA” refers to the software now known as TreeAge Pro.

had submitted a model. The survey was logic-based and so certain questions were only asked depending on the response to a previous question. This was developed to make the survey quicker to complete, to ask only questions that were relevant to specific respondents and to generally streamline the survey instrument for respondents, as well as assist in filtering out respondents and analysing results by responder subgroup. The survey instrument is attached as Appendix 1.

### **b. Results**

44 organisations (7 TAGs, 29 Manufacturers and 8 Consultancy Firms) were invited to complete the survey. Of the 44 invited organisations there were 33 completed responses of which 29 had produced models for a NICE TA (the remaining 4 respondents had produced models through an external consultancy). There were 7 responses from TAGs: 6 different TAGs responded with one TAG responding twice. These responses were amalgamated. There were 8 responses from consultancy firms, and 14 different health industry responses. Therefore a total of 28 responses are reviewed in the analysis.

The respondents were asked which specific software packages they had used to submit a model for a TA. Table 1 shows which specific software packages had been used by each respondent. All respondents had used MS Excel, and 16 of 28 (57%) had used TreeAge Pro.

**Table 1 - Software used for NICE Technology Appraisals**

Software	Respondents that used this software		Number of TAGs	Number of Manufacturers	Number of Consultancies
	n	%			
MS Excel	28	100%	6	14	8
TreeAge Pro	16	57%	6	7	3
WinBUGS	6	21%	1	2	3
R	5	18%	1	2	2
Arena	3	11%	0	2	1
SAS	3	11%	0	1	2
Crystal Ball	2	7%	1	0	1
Simu8	2	7%	1	0	1
STATA	1	4%	1	0	0
RevMAN	1	4%	1	0	0
Borland Delphi	1	4%	1	0	0
S-PLUS	1	4%	1	0	0
@risk	1	4%	0	0	1
STELLA	0	0%	0	0	0
Witness	0	0%	0	0	0

The survey informed respondents that it was concerned only with computer software packages used for cost-effectiveness analysis, defined as the ‘modelling platform’, and not other software packages such as specialist statistical software that may be used to derive parameters that are subsequently used within the cost effectiveness model. The inclusion of RevMAN, a systematic review package, and comments that software had been used in conjunction with modelling software suggests that respondents included software packages that were used for specific parts of the analysis, and not solely for the development of the base cost-effectiveness model. Statistical software such as STATA, R and WinBUGS can be programmed to produce cost effectiveness models.

Respondents were asked if they had submitted a model to NICE in a package other than the recommended packages, MS Excel and TreeAge Pro. 8 of the 28 (29%) respondents had submitted using a package other than MS Excel or TreeAge Pro (3 Assessment Groups, 4 Industry, 1 Consultancy). These 8 respondents had submitted a total of 9 models in packages other than MS Excel or TreeAge Pro.

**Table 2 - Specific software used for NICE TAs other than MS Excel and TreeAge Pro**

<b>Package</b>	<b>Frequency</b>
R	3 (33%)
Simul8	1 (11%)
Borland Delphi	1 (11%)
VBA	1 (11%)
CORE web-based model	1 (11%)
RevMAN	1 (11%)
STATA	1 (11%)
<b>Total</b>	<b>9 (100%)</b>

The differences between Table 1 and Table 2 suggests that there may not have been a consistent distinction between the software used as the platform for the model, and that used for specific components that are populated and visualised in the model platform. VBA was a response in the “other” part of the answer, although we assumed that almost all MS Excel models require VBA programming, especially for Probabilistic Sensitivity Analysis (PSA)<sup>2</sup>.

<sup>2</sup> There is therefore an inconsistency between the guidance to manufacturers (quoted on page 2) and the recommendation for PSA in the methods guide.

Respondents were then asked if they had ever felt it would be more appropriate to build a model in a format other than MS Excel/TreeAge Pro. 19 (68%) reported that they had felt it would be more appropriate, with the remaining 9 (32%) reporting that they had not. Table 3 reports the specific reasons for the 19 respondents who felt it would be more appropriate.

**Table 3 - Reasons for feeling submitting in a package other than MS Excel or TreeAge Pro would be more appropriate**

<b>Reason</b>	<b>Frequency</b>
Familiarity of the software	0 (0%)
Suitability to the type of model	7 (37%)
Computational requirements of the model	9 (47%)
Other	3 (16%) (2 = All of the above, 1 = “combine data analysis with modelling capacity”)
<b>Total</b>	<b>19 (100%)</b>

Respondents were given the chance to expand on their feelings around this issue. 5 respondents referred to individual sampling models. In this situation an essential component of the model is that individuals are tracked, one at a time. Each individual can only take a single specific path through the model but sufficiently large numbers of patients are sampled to give precision to estimated costs and benefits. When coupled with PSA there can be a substantial computational burden with these types of models, since there can be a need to sample a large number of individual patients a large number of times. It is probably for this reason that a review identified that only one of six individual sampling models submitted to NICE TAs performed PSA<sup>3</sup>. Individual sampling models can provide extra functionality and are sometimes more suitable to particular decision problems. The respondents reported that these models are best developed in a dedicated simulation software packages.

“I would not use TreeAge for anything other than the simplest decision tree or markov model.”

*Consultancy Organisation*

“Advice from statistical colleagues suggests that the use of Excel for running complex simulation models is inefficient.”

*Consultancy Organisation*

“DES [Discrete Event Simulation] packages are not standardised so there are big issues with how easy it is for others to view and experiment with a developed model.”

*Technology Assessment Group*

“With the requirement for probabilistic sensitivity analysis it appears that Excel is no longer fit for purpose for building complex patient level simulation models.”

*Healthcare Industry*

Respondents were asked what their key reason was for selecting a specific software package when submitting a model to NICE.

**Table 4 - Key Reason for Software Choice**

<b>Reason</b>	<b>Frequency</b>	<b>Details</b>
Familiarity of the software	6 (21%)	
Suitability to the type of model	12 (41%)	
Computational requirements of the model	5 (17%)	
Other	6 (21%)	2 = “transparency of model” 1 = “acceptability to NICE” 1 = “all of the above except familiarity” 1 = “consistency for users/reviewers” 1 = “there is no choice!”
<b>Total</b>	<b>29 (100%)</b>	

Many respondents specifically referred to primarily considering building the model in MS Excel, due to its alleged transparency, availability and for reviewing and evaluation purposes:

“...it is not really possible to separate the above reasons. The decision is made on the basis of a combination of the three.”

*Technology Assessment Group*

“Ability of others to utilize the model”

*Healthcare Industry*

“...our primary consideration is the evaluating agency and their requirements.”

*Consultancy Organisation*

“We would always try to produce a model for NICE in excel, in line with requirements. Choice of software is a mixture of ensuring a) software would be acceptable for submission and b) suitability of software and familiarity.”

*Consultancy Organisation*

Respondents were then asked if they built models for other purposes, and what software they used. 27 of the 28 respondents built models for reasons other than for NICE TAs, and 18 had used software other than MS Excel and TreeAge Pro for these non NICE projects.

**Table 5 - Software Packages other than MS Excel and TreeAge Pro used for non-NICE models**

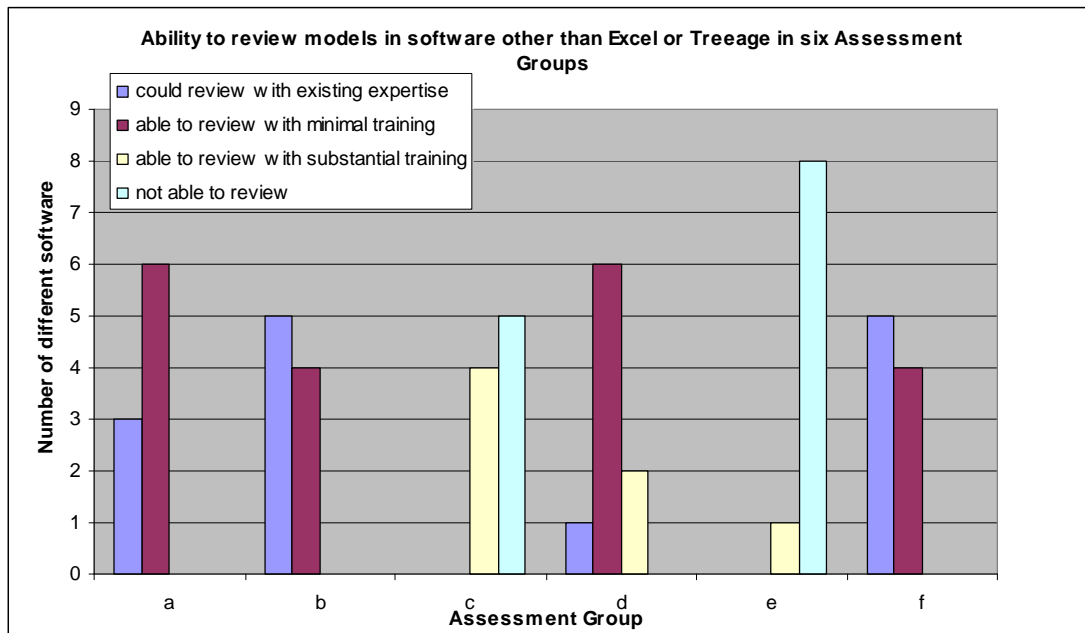
Software	Number of Organisations (n=27)		TAGs	Manufacturers	Consultancies
	n	%			
WinBUGS	9	33	4	2	3
R	6	22	2	1	3
Arena	5	19	0	4	1
STATA	4	15	1	0	3
Simul8	3	11	2	0	1
SAS	3	11	0	2	1
VBA	2	7.4	0	2	0
Crystal Ball	2	7.4	1	1	0
Mathematica	2	7.4	0	1	1
S-PLUS	1	3.7	1	0	0
Borland Delphi	1	3.7	1	0	0
@risk	1	3.7	0	0	1

As well as the consideration of the suitability of packages to develop models, it is also important that the skills are available for TAGs/ERGs to fully review models as part of an STA or MTA. The 6 TAGs that responded were asked about their ability to review certain software packages

**Table 6 – Ability of Assessment Groups (AGs) to review models in different software packages (for MTA, STA)**

Software	Number of AGs			
	That have expertise	That would require minimal training	That would require substantial training	That could not review
MS Excel	6	0	0	0
TreeAge	6	0	0	0
Simul8	3	1	1	1
Arena	0	4	0	2
R	2	2	1	1
WinBUGS	2	2	0	2
Crystal Ball	1	3	2	0
Witness	2	2	0	2
S-PLUS	1	3	0	2
SAS	2	1	2	1
STELLA	1	2	1	2

**Figure 1**



The distribution of these responses by assessment group (labelled a to f) is displayed in figure 1. One AG responded that they would not be able to review models in any software other than MS Excel, TreeAge or Crystal Ball, the latter with substantial training. One other AG reported that they would not be able to review models in any software other than MS Excel or TreeAge. Simul8, R, Crystal Ball and SAS would require substantial training for this group and Arena, WinBUGS, Witness, S-Plus and STELLA could not be reviewed even with substantial training.



Three groups responded that they could review models in any of the suggested software packages either with existing expertise or with minimal training time. The final group reported that models in every software type could be reviewed either with existing expertise or minimal training with two exceptions: SAS and STELLA would require substantial training.

It is important to note that there are apparent inconsistencies in the responses given from assessment groups in tables 5 and 6. For example, 4 groups claim to have built models in WinBUGS but only 2 report having the capacity to review models with current expertise. It may be the case that respondents are thinking of their own abilities in the second question, or those of a more limited group. Alternatively, it could be that members of staff no longer in the unit developed previous models.

Furthermore, it is questionable why some groups believe they could master certain pieces of software which they are currently unfamiliar within a short amount of time, whilst others consider a substantial amount of time would be required, or that they simply would not be able to review models in certain software. It seems unlikely that the capabilities of individuals to learn how different software packages operate differs substantially.

Finally respondents were given the opportunity to make any other comments in free text boxes. Extracts from a selection of the comments are given below, and are grouped by their general theme. We have also supplied a full copy of these free text quotes, suitably censored to ensure anonymity of responders, to NICE for consideration alongside this report.

Several suggested specific reasons for why NICE should allow models to be submitted in other software packages:

“NICE should stimulate the use of R. Official recognition and governmental acceptance will help its further development. Further, allowing people to step outside the restriction of pre-programmed packages will force them think again about what they are doing.”

*Consultancy Organisation*

“It would be helpful if NICE could expand the list of approved software to include R and other commonly used modelling software. It appears that assessment groups are not restricted to only using Excel and Treeage, therefore it is unfair for manufacturers to have this additional restriction.”

*Healthcare Industry*

However, some respondents commented on why NICE should not allow submissions to be made in other software packages:

“Transparency of modelling is already an issue within the process, I would recommend not exacerbating this by widening the range of options.”

*Healthcare Industry*

“...makes good sense to have limited choice for consistency/replication and transparency issues.”

*Healthcare Industry*

“...this is not the time to be imposing additional recruitment/training constraints by opening the flood-gates to a great multiplicity of additional software platforms. Benefits for authoring modellers are likely to be marginal, and the consequences for Review Groups substantial.”

*Technology Assessment Group*

Several respondents suggested alternative ways that NICE could update their requirements for submissions of models:

“An alternative approach to reviewing models constructed with 'non-standard' software is to recast the major structures / assumptions in the form of an Excel 'validation' model, which can be used to verify that submitted model results (both absolute values and sensitivities) can be readily replicated on another platform. This works well when manufacturers provide such a 'validation' spreadsheet alongside their preferred model.”

*Technology Assessment Group*

“I think it would be very difficult for NICE to allow some Discrete Event packages and not others for Tech. Ass. submissions. Vendors and developers who were excluded in this way would probably feel aggrieved. One potential solution would be request/require that all simulation vendors should make run-time versions of the software available free. This would allow third parties to view models produced although not author or significantly change the structure etc.”

*Technology Assessment Group*

Several respondents made comments relating to specific modelling challenges:

“...important to separate interfaces from analytical approaches. Confusion with several issues: (i) need for patient level analysis (e.g. SIMUL8) vs cohort analysis; (ii) fit for purpose - to what extent can Excel/Treeage deal with indirect/MTC comparisons; (iii) statistical inputs - to what extent is 'any' statistical analysis conducted within Excel/Treeage - I don't remember either of their ability to deal with standard approaches to evidence synthesis or survival analysis! I'm unclear why there is such a desire to have models built and analysed in 1 package.”

*Technology Assessment Group*

“In some cases, the problem is not the software itself, but the capability of the software to be used in non-standard ways. For example, a complex model in any sort of programming language (including Visual Basic macros for Excel, R, and Borland Delphi) can be almost impenetrable without extra time. Simple models can be built in these languages which can be reviewed with no difficulty whatsoever.”

*Technology Assessment Group*

A number of TAGs made comments relating to the review of models:

“...however, the person with the expertise may not be available to review the model or tutor colleagues.”

*Technology Assessment Group*

“I have responded that currently we would not be able to review models in these other packages. It may be possible to train people in other software packages if additional funding were provided for a substantial training programme and to obtain copies of other packages.”

*Technology Assessment Group*

Finally, a number of respondents made general comments about software packages:

“Currently we support the use of transparent, verifiable models that are fit for purpose - not unduly complex, and where possible meet the needs of reimbursement submissions and market access in a variety of countries, not just UK. In the majority of cases Excel-based models are adequate.”

*Healthcare Industry*

“We are thinking about using some of the discrete event simulation software but have not so far.”

*Healthcare Industry*

“I believe the use of script based languages such as STATA, R, SAS have an important role in the development of robust and easily validated cost-effectiveness models and their dissemination.”

*Consultancy Organisation*

### **3. SOFTWARE REVIEW**

The following table gives a review of a number of commonly used software packages for cost-effectiveness modelling in healthcare. It provides a basic overview of the capabilities and transparency of each package, along with approximate costs (may vary by edition, licensing requirements and commercial requirements). A detailed objective comparison would require familiarity with each package and testing of a consistent model across all platforms. This was beyond the scope of this report.

<b>Software Package</b>	<b>Software Type</b>	<b>Specific Analysis</b>	<b>Cost</b>	<b>Transparency</b>
MS Excel	Spreadsheet	When integrated with VBA programming language, can perform cohort, simulation and Markov models	2007 Full Version £200 (available free on most PC's)	“Cell chasing” can be challenging and more complex models require VBA coding
TreeAge Pro	Decision Tree/Decision Modelling	A pre-programmed package that is designed for decision tree, cohort and Markov models. Has the ability to perform simulation models and PSA	Standard \$1145, Academic Annual Unit Price \$205, 10-user annual license \$1,950. Commercial Annual Unit Price \$345	Being a pre-programmed package there is no coding to review. Some find large models and/or equations can be unwieldy
Simul8	Simulation package	Uses bespoke programming language (Visual Logic). Can perform Markov, Simulation models, PSA and complex systems such as infectious disease modelling	£895 standard, £2995 professional. Simul8 for education site license £1250	Visual Logic language is simple and easy to review using code step-through. The Graphical User Interface (GUI) allows simple debugging
Arena	SIMAN based simulation package	Arena is a simulation package, includes VBA and can easily be incorporated into MS Excel.	Basic Single User \$795, Academic version + textbook £60	Unable to comment
R	Programming language and statistical package	R is a free and open-source statistical package and programming language. It is extremely powerful, with the language allowing flexibility and complex models to be coded very quickly and simply	Free	R is script-based and so models are generally transparent. It has a very easy to follow syntax. As free software it is open to review from all stakeholders
WinBUGS	Bayesian statistics	WinBUGS allows statistical	Free	WinBUGS is script-based and so

	package	analysis in a Bayesian framework, and is used for indirect/mixed-treatment comparison and other Bayesian statistical analyses		models are generally transparent. As free software it is open to review from all stakeholders
Crystal Ball	MS Excel simulation add-on	An add-on for MS Excel, it allows distributions to be added to parameters and Monte-Carlo simulations to be run	Academic quote unavailable. Standard £505, Professional £962	The assumptions used to provide parameterisations are transparent. The software package is required to review the model and so as a separate add-on is must be disclosed
SAS	Script-based simulation and analytics package	SAS is based on the S programming language and is a dedicated analytics and simulation software package. It is particularly used in business and forecasting specialties	Unknown	Being script-based, models are generally transparent and easy to review
STATA	Script-based statistical package	STATA is not a dedicated package for simulations or cost-effectiveness modelling, although it can be used for these applications. It is widely used for survival, epidemiological and meta-analyses as well as data management and so is an important component of many models	£580 Stata 10 SE.	Being script-based, models are generally transparent and easy to review
Borland Delphi	Script-based software development package, incorporating the	Borland Delphi produces executable software applications that allow models to be run based on the C++ programming	£649 - £1789	There are potential issues about interrogation of the underlying model structure and assumptions

	C++ and C# languages	language		
--	-------------------------	----------	--	--



#### 4. DISCUSSION

The survey and software review have highlighted divisions between stakeholders regarding the appropriateness of limiting submissions to MS Excel and TreeAge Pro.

Whilst there was consensus that other packages may be more appropriate to particular types of decision problems, there was no agreement as to the frequency with which these situations arise. There were some concerns that the benefits of allowing a wider array of software would not justify the implications, including training costs, for assessment groups. However, there were also voices that strongly favoured the use of other software.

Increasingly, advanced statistical techniques are required to estimate model parameters. There can be advantages to also developing the decision model in the same software. For example, packages such as WinBUGS can be used to appropriately reflect the full uncertainty and correlations in the underlying data within the decision model<sup>4</sup>, particularly when indirect and mixed treatment comparisons are required. This may not be possible when separating the model building software from the statistical analysis providing inputs to the model.

In the case of individual sampling models, several respondents indicated that bespoke packages are more user friendly and can be more transparent in their methods. Running PSA in MS Excel or TreeAge Pro can take an extremely long time, with bespoke packages often requiring a fraction of the computing time. This may be of particular importance where individual sampling models are considered appropriate. However, it should be noted that there is debate about the extent to which such model types are required and the computational burden associated with performing PSA<sup>3,5,6</sup>. Furthermore, there are developing methods to address this computational burden through the use of more efficient simulation<sup>7</sup> and emulators<sup>8</sup>.

MS Excel is often seen as a particularly transparent package for model building. Whilst it is true that most individuals have some level of familiarity with Excel, this is not necessarily the same as transparency per se. In fact, 'cell-chasing' and validation can be extremely time consuming and complex. It could be argued that well annotated script-based models may be considered more transparent and easier to validate than

equivalent Excel based models once familiarity and experience with the software or programming language is gained. Good practice guidelines in the format of program code, annotation of the code and its publication within reports would enhance transparency further.

Two respondents proposed the use of MS Excel “validation” models when non standard software is used. This would allow the fundamental structure and parameter values to be interrogated and basecase and univariate sensitivity analyses to be replicated or conducted. Of course, where the motivation for the use of non standard software is to enable PSA, this may leave a substantial gap between the elements of the model that are reviewed by the Assessment Group and the analysis of parameter uncertainty that may be pivotal to the appraisal. A “validation model” in MS Excel was supplied in the case of abatacept for rheumatoid arthritis whilst the main model was built in R due to its complexity<sup>9</sup>. In this case the ERG considered the Excel model<sup>10</sup>.

---

<sup>1</sup> NICE (2004) Contributing to a Technology Appraisal: A Guide for Manufacturers and Sponsors (reference N0518). Available at

[http://www.nice.org.uk/niceMedia/pdf/TAprocessmanual\\_manufacturers\\_sponsors.pdf](http://www.nice.org.uk/niceMedia/pdf/TAprocessmanual_manufacturers_sponsors.pdf)

<sup>2</sup> NICE (2008) Single Technology Appraisal (STA) specification for manufacturer/sponsor submission of evidence. Available at

<http://www.nice.org.uk/media/45D/42/SpecificationForManufacturerOrSponsorSubmissionOfEvidenceUpdate.pdf>

<sup>3</sup> Griffin S, Claxton K, Hawkins N, Sculpher M. (2006) Probabilistic Analysis and Computationally Expensive Models: Necessary and Required?, *Value in Health*, Vol. 9:244-252

<sup>4</sup> Cooper NJ, Sutton AJ, Abrams KR, Turner D, Wailoo A. (2004) Comprehensive decision analytical modelling in economic evaluation: A Bayesian approach, *Health Economics*, Vol.13: 203-226.

<sup>5</sup> Caro, J., Getsios, D., and Möller, J. (2007) Regarding Probabilistic Analysis and Computationally Expensive Models: Necessary and Required? *Value in Health*, Vol.10:317-318

<sup>6</sup> Griffin, S., Claxton, K., Hawkins, N., and Sculpher, M. (2007) Decision Models Need to be “Fit for Purpose” for Decision-Making: Response to Caro et al., *Value in Health*, Vol.10:319

<sup>7</sup> O'Hagan, A., Stevenson, M. and Madan, J. (2007). Monte Carlo probabilistic sensitivity analysis for patient level simulation models: Efficient estimation of mean and variance using ANOVA, *Health Economics*, Vol.16: 1009-1023.

<sup>8</sup> Stevenson, M., Oakley, J., Chilcott, J. (2004) Gaussian process modelling in conjunction with individual patient simulation modelling; a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis”, *Medical Decision Making*, Vol.24:89-100.

<sup>9</sup> Bristol-Myers Squibb Pharmaceuticals Ltd (2007) Abatacept (Orencia®) for the Treatment of Rheumatoid Arthritis. Single Technology Appraisal Submission to the National Institute for Health and Clinical Excellence. Available at

<http://www.gserve.nice.org.uk/nicemedia/pdf/ACDAbataceptOrenciaSTABMSSubmission.pdf>

<sup>10</sup> Liverpool Reviews and Implementation Group (2007) Abatacept for the treatment of rheumatoid arthritis. ERG Report. Available at

<http://www.nice.org.uk/nicemedia/pdf/ACDAbataceptFinalERGReport.pdf>