



IMI Work Package 5: Report 2:b:ii Benefit - Risk Wave 2 Case Study Report: Rosiglitazone

Benefit-Risk Analysis Rosiglitazone

Lawrence Phillips, Billy Amzal, Alex Asimwe, Edmond Chan, Chen Chen, Diana Hughes, Juhaeri Juhaeri, Alain Micaleff, Shahrul Mt-Isa, Becky Noel, Susan Shepherd, Nan Wang
On behalf of PROTECT Work Package 5 participants

Version 1 Date: 25 Feb 2013 Date of any subsequent amendments below	Person making amendments	Brief description of amendments
25 th May 2013	Shahrul Mt-Isa	Order of authorship was amended to first author, then alphabetically

https://erombayer.de/eRoomReq/Files/PH-GDC-PI-SID/IMI-PROTECT/0_f9082/PROTECT_WP5_report_template.docx

Disclaimer: The processes described and conclusions drawn from the work presented herein relate solely to the testing of methodologies and representations for the evaluation of benefit and risk of medicines. This report neither replaces nor is intended to replace or comment on any regulatory decisions made by national regulatory agencies, nor the European Medicines Agency

Acknowledgements: The research leading to these results was conducted as part of the PROTECT consortium (Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium, www.imi-protect.eu) which is a public-private partnership coordinated by the European Medicines Agency. The PROTECT project has received support from the Innovative Medicines Initiative Joint Undertaking (www.imi.europa.eu) under Grant Agreement n° 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution

TABLE OF CONTENTS

1.	EXECUTIVE SUMMARY	4
2.	GLOSSARY	6
3.	PARTICIPANTS	8
4.	ROSIGLITAZONE BENEFIT-RISK APPRAISAL	10
5.	PART I: THE DETERMINISTIC BENEFIT-RISK MODEL.....	12
5.1	Alternatives	12
5.2	Effects Criteria	12
5.3	Weighting.....	12
5.4	Scoring.....	20
6.	RESULTS.....	21
6.1	Cumulative weights.....	21
6.2	Overall benefit-risk balance.....	22
6.3	Comparative Analyses	23
6.4	Sensitivity Analyses.....	24
6.5	Non-linear value functions	26
7.	DISCUSSION	31
8.	PART II: THE PROBABILISTIC BENEFIT-RISK MODEL	33
8.1	Data.....	33
8.2	Meta Analysis.....	33
8.3	Cumulative Weights.....	38
8.4	Simulations.....	38
9.	RESULTS.....	41
9.1	MCDA Results, revised deterministic model.....	41
9.2	SIMULATION RESULTS	45
10.	DISCUSSION AND CONCLUSIONS.....	48
11.	STRENGTHS AND LIMITATIONS OF THE APPROACHES USED.....	50
11.1	META-ANALYSIS.....	50
11.2	MCDA MODEL.....	50
11.3	PROBABILISTIC SIMULATION	51
12.	RECOMMENDATIONS AND FURTHER WORK.....	52
12.1	PATIENT LEVEL DATA.....	52
12.2	DATA FORMAT AND QUALITY.....	52
12.3	ELICITATION OF VALUE FUNCTIONS AND WEIGHTS	52
12.4	SOFTWARE PACKAGES	53

13. REFERENCES.....	55
APPENDIX A—DECISION CONFERENCING.....	57
Overall Favourable-Unfavourable Effects Balance	59
Favourable Effects.....	59
Unfavourable Effects	60
APPENDIX C: COMPREHENSIVE META-ANALYSIS INPUT AND RESULTS.....	63
META-ANALYSIS FOR MEAN TREATMENT EFFECT.....	63
Glycaemic efficacy % for rosiglitazone plus adjunct	63
Result for combined effect	63
Forest plot for combined effect.....	64
META-ANALYSIS FOR ODDS RATIO	64
Data input window for cv death %.....	64
Result for CV death % odds ratio	65
Forest plot for cv death % odds ratio.....	65

1. EXECUTIVE SUMMARY

During two decision conferences at Imperial College on 28 June and 24 July 2012, members of the Rosiglitazone Case Study Team of PROTECT's Work Package 5 developed a decision-theory-based model¹ for evaluating the benefit-risk balance of rosiglitazone, a drug used in the treatment of type II diabetes. Participants in the decision conferences took the role of regulators in a hypothetical European country, today, considering all publicly-available data from clinical trials and as available post-authorisation. This report summarises the process of creating and exploring the model, and drawing conclusions from it.

Alternatives considered were:

1. Rosiglitazone in a fixed dose combination with metformin and/or glimepiride, referred to as rosi + adjunct in this document.
2. Metformin and/or glimepiride alone, referred to as adjunct only in this document.

The group considered two favourable effects and nine unfavourable effects, the latter representing four effects based mostly on the clinical trials, and four MACE criteria from post-marketing observational data (p. 14, Effects Tree, Figure 1). Each criterion was carefully defined to enable meaningful evaluations of the alternatives (p. 15, Effects Table, Table 1). Measurement scales were identified for all the effect criteria.

The group assessed relative weights for all the criteria (p. 19, Figure 7). The method of swing weighting, which requires comparative judgments about the ranges of effects and clinical judgements about how much they matter relative to each other, made it possible to assign meaningful relative weights to all scales (pp. 16-18, Figures 2-6). These weights reflect both the range from the least to most preferred effects on each scale, a matter of fact, and how much those effect differences matter, a consideration of clinical relevance that takes the context for decision making into account (such as unmet medical need). The swing-weighting method results in numbers that are, basically, scale constants, which equate the units of preference value across all the criteria.

The overall weighted preference value for fixed dose combination of rosiglitazone + metformin or glimepiride was 35 compared to metformin or glimepiride alone at 43, indicating that the benefit-risk balance is better for adjunct only (p. 22, Figure 10). However, a breakdown of those overall preference values into the weighted contribution of each effect showed that rosiglitazone + adjunct is safer on three MACE criteria, Non-CV death, CV death and stroke, with a very small superiority for Microvascular events (p. 24, Figure 12). But the superiority on those four effects is overbalanced by the poorer safety of rosiglitazone + adjunct on other unfavourable effects. The finding that adding rosiglitazone to adjunct therapy reduces the overall benefit-risk balance was maintained under many sensitivity analyses examining imprecision in the data and differences in judgements.

Participants found that the decision conference, with its combination of on-the-spot computer modelling, engagement of participants with different perspectives, and impartial facilitation, was a useful and instructive process that deepened understanding about the topic.

To explore the value of incorporating uncertainty about all effects, all the statistical data summaries used in the MCDA model were replaced with their underlying probability distributions. Monte Carlo analyses implemented probabilistic simulation of each option's performance, enabling the computation of a single distribution of the benefit-risk difference of rosiglitazone plus adjunct compared to adjunct only. This distribution gave a probability of 0.998 that the benefit-risk difference of adjunct only is better than rosiglitazone plus adjunct. This high level of certainty surprised the team, as it was not apparent in the deterministic MCDA model, and it remained relatively unaffected by sensitivity analyses on the criteria weights.

¹The model represents value preferences, as in multi-criteria decision analysis (MCDA), and their uncertainties (as in decision tree analysis), so can be considered a mixed model, though it will be referred to in this report as an MCDA model.

2. GLOSSARY

CHF	Congestive heart failure
CI	Confidence Interval
CMA	Comprehensive Meta-Analysis (software)
CV	Cardio-vascular
EMA	European Medicines Agency
GSK	GkaxoSmithKline
MACE	Major adverse cardiovascular events
MCDA	Multi-criteria decision analysis
MI	Myocardial infarction (heart attack)
NCBI	National Center for Biotechnology Information
NEJM	New England Journal of Medicine
OR	Odds Ratio
PML	Progressive multifocal leukoencephalopathy, a usually-fatal virus disease of the white matter of the brain
PrOACT-URL	An eight-stage process for structuring a decision as an aid to decision makers
QTc	Corrected QT interval, an electrocardiographic measure of both depolarization and repolarization within the heart
RCT	Random Controlled Trial
SD	Standard Deviation
SE	Standard Error
WP5	Work Package 5 of the PROTECT project

3. PARTICIPANTS

(Some face-to-face at Imperial, others by teleconference full or part-time.)

28 June

Billy Amzal, LA-SER Analytica
Edmond Chan, Imperial College
Diana Hughes, Pfizer
Alain Micallef, Merck Serono
Shahrul Mt-Isa, Imperial College
Nan Wang, Imperial College

24 July

Alex Asimwe, Lilly
Edmond Chan, Imperial College
Chen Chen, London School of Economics
Kimberley Hockley, Imperial College (observer)
Diana Hughes, Pfizer
Juhaeri Juharei, Sanofi
Alain Micallef, Merck Serono
Becky Noel, Lilly
Shahrul Mt-Isa, Imperial College
Susan Shepherd, Amgen
Nan Wang, Imperial College

Dr Larry Phillips facilitated both decision conferences.

4. ROSIGLITAZONE BENEFIT-RISK APPRAISAL

This report documents the process and results of two decision conferences (a facilitated, group-modelling process described in Appendix A), one on 28 June, the other on 24 July 2012, and a subsequent study examining the effects of uncertainty in the data. The purpose of the first decision conference was to create the structure of a benefit-risk model for the anti-diabetic drug rosiglitazone, as used in fixed dose combinations with an adjunct, either metformin or glimepiride, as compared to the use of the adjunct only. The second decision conference revised and completed the model, and explored its results to attain a better understanding of the benefit-risk balance of the drug. Work following the decision conferences extended the model to include a probabilistic simulation in which all the statistical summary data used in the benefit-risk model were replaced with their corresponding probability distributions. Monte Carlo analyses enabled the benefit-risk balances of the drug plus adjunct to be compared with the effects of the adjunct only.

Rosiglitazone received marketing authorisation in 1999 in the United States and in 2000 in the European Union. New data subsequently emerged about possible heart problems associated with rosiglitazone, confirmed by a meta-analysis in 2007 (Nissen & Wolski, 2007), which resulted in a European suspension late in 2010. This suspension included its use as a fixed dose combination with metformin or glimepiride which had been approved in 2003 for metformin and 2006 for glimepiride. The drug remains available in the United States, but only under a restricted-access program.

Participants in the decision conferences took the role of regulators. The focus in the June decision conference was on structuring the problem, agreeing what favourable and unfavourable effects should be taken into account, developing definitions for the effects, establishing measurement scales for the effects, assessing swing-weights for the effect scales and deciding what alternatives were to be compared. The July decision conference began by looking at some apparent anomalies in the weights and revising them, followed by agreeing the data to be input. Considering the data led to several revisions of the model. The overall result showed that the benefit-risk balance of rosiglitazone + adjunct therapy is less than adjunct therapy only. Sensitivity analyses revealed that model results are very robust to imprecision and disagreements about weights. Even non-linear value functions on the most discriminating effects did not tip this balance.

Work over succeeding weeks replaced statistical summaries of all the effects data with probability distributions calculated from meta-analyses of all published trials that were sufficiently detailed to recover the data, 22 clinical trials altogether. This enabled probabilistic simulation to establish the overall distributions of favourable and unfavourable effects, distributions for the favourable-unfavourable effect difference between the drug plus adjunct and adjunct only, and, finally, a single probability distribution for the benefit-risk difference of rosiglitazone plus adjunct compared to adjunct only. This latter distribution provides a quantitative answer to the question, "Does the benefit-risk balance favour the drug?", in the form of a probability that the benefit-risk balance is greater for the drug than for the comparator.

5. PART I: THE DETERMINISTIC BENEFIT-RISK MODEL

In the first decision conference, after each participant explained what they could bring to the task of creating the structure of the rosiglitazone model, the group began their work with a discussion of the decision context (the first step in the informal agenda provided by the PrOACT-URL framework²). The group agreed that we are the regulators for a hypothetical country, many of whose inhabitants suffer from Type II diabetes. The time is now and any publically-available data can be accessed.

5.1 Alternatives

In the first decision conference considerable discussion attended the task of identifying the alternatives to be compared, as rosiglitazone had been approved in the UK and USA to be used as a monotherapy, or in combination with metformin or glimepiride, or in triple therapy with metformin and sulphonylurea. Studies have also compared rosiglitazone to pioglitazone. Eventually, the group agreed three alternatives:

1. Rosiglitazone in a fixed dose combination therapy with metformin, and/or glimepiride, referred to as rosi + adjunct in this document
2. Pioglitazone in a fixed dose combination therapy (with metformin and/or glimepiride)
3. Metformin and/or glimepiride alone referred to as adjunct in this document.

However, at the second decision conference, as data were being considered for the favourable effects, two problems arose: (1) it was not possible to decide about two alternative interpretations of the microvascular event data for pioglitazone because the clinical study publication expressed the results apparently unaware that the text was open to two interpretations, and (2) the selection of patients for the pioglitazone study made it impossible to reconcile the very different response of the control patients compared to the rosiglitazone studies. As a result, the group decided to delete option 2, the pioglitazone alternative.

5.2 Effects Criteria

Two favourable effects and nine unfavourable effects characterised the final model. The Effects Tree, Figure 1, shows favourable and unfavourable effects at the nodes, and criteria against which the drugs were evaluated at the extreme right. The two clusters of unfavourable effects are intended to separate post-marketing data under the MACE (Major Cardiac Event) heading, from the Other heading, which reports data drawn mostly from clinical trials conducted prior to approval. Although the available documentation reports many effects, the group chose to model only those effects that might affect the benefit-risk balance; thus, many unfavourable effects were not included in the model.

Definitions of the criteria are given in Table 1. This Effects Table shows the short name given in Figure 1, the description of the effect, fixed upper and lower input-data values that define a plausible range for the data, the cumulative weights that resulted from weighting the criteria swings between the upper and lower values of the associated metric scales, an indication of how the measured data will be converted into preference values, the units of measurement, and, finally, columns of statistical summaries of the data for the two alternatives.

5.3 Weighting

Since all input measures are converted to preference values on scales that range from 0 to 100, the addition of individual preference values to create an overall preference value showing the benefit-risk balance requires the units of preference on all scales to be comparable. As an analogy, both Fahrenheit and Celsius scales contain 0 to 100 portions, but the swing in temperature from 0 to 100 on the Fahrenheit scale is, of course, a smaller swing in

² See Appendix 5.1 of the Work Package 4 report from the EMA's Benefit-Risk Project, at http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/03/WC500123819.pdf.

temperature than 0 to 100 on a Celsius scale; the comparability of the scales is that it takes 5 Celsius units to equal 9 Fahrenheit units. It is this 'trade-off' or 'statement of equivalence' that is established by the assessment of swing weights, which are, in essence, scale factors that equate the units of preference value from one scale to the next.

To assess these scale factors two steps in thinking must be separated. First, it is necessary to think about the difference between the measured effects represented by preference values of 0 compared to 100. That is a straightforward assessment of a difference in effect, from the least preferred effect to the most preferred effect on that criterion. The next step is to think about how much that difference in effect matters; this is essentially a judgement of the clinical relevance of the difference in effect size. "How big is the difference and how much do you care about it?" This is the question that was posed in comparing the 0-to-100 swing in effect on one scale with the 0-to-100 swing on another scale. This process is referred to as 'swing-weighting,' and is now the method preferred by most decision analysts for assessing criterion weights.

Swing weights apply to the range of the measurement scale; they should not be interpreted as the importance of the criterion, which has no meaning in MCDA unless a range is specified.

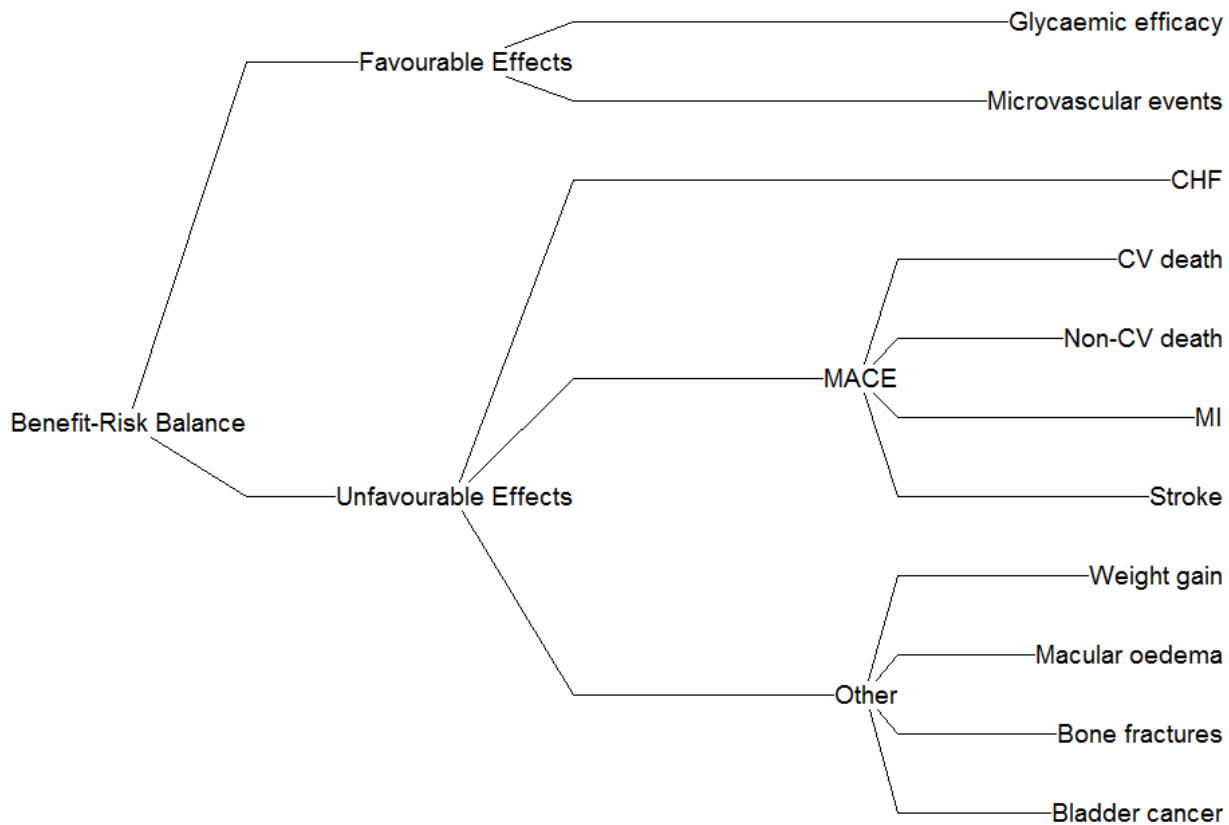


Figure 1: The Effects Tree, showing the two Favourable Effects and the nine Unfavourable Effects, with eight of the latter clustered under MACE and Other effects.



Table 1: Effects Table for Rosiglitazone.

	Name	Description	Fixed Upper	Fixed Lower	Weight	Value function*	Units	Rosi + adjunct	Adjunct only	
Fav'ble Effects	Glycaemic efficacy	(A surrogate marker of the quality of glucose regulation.) Mean change from baseline in the proportion of Hb in which A1c is greater than 48 mmol/ml.	5.00	-5.00	0.9	Inverse linear	%	-1.18	0.06	
	Micro-vascular events	Incidence of new cases of microvascular events compared to baseline (Retinopathy requiring photocoagulation, vitreous haemorrhage, & fatal or non-fatal renal failure.)	20.00	0.00	8.8	Inverse linear	%	2.70	3.50	
Unfavourable Effects	CHF	Proportion of patients experiencing congestive heart failure during the study period.	4.00	0.00	13.4	Inverse linear**	%	3.69	1.89	
	CV death	The proportion of patients who died from any cardiovascular event including stroke.	4.00	0.00	16.7	Inverse linear	%	2.70	3.19	
	Non-CV death	The proportion of patients who died from any non-cardiovascular event including stroke.	4.00	0.00	16.7	Inverse linear**	%	2.97	3.86	
	MI	Proportion of patients who experience a non-fatal heart attack.	5.00	0.00	5.9	Inverse linear	%	3.33	3.01	
	MACE	Stroke	Proportion of patients who experience a non-fatal ischemia stroke.	5.00	0.00	5.0	Inverse linear	%	1.94	2.83
	Weight gain	Mean change from baseline in weight gain at 1 yr.	10.00	-5.00	5.9	Inverse linear	Kg	3.80	0	
	Macular oedema	Proportion of patients who experience macular oedema. [Are data available?]	1.00	0.00	5.0	Inverse linear	%	1.27	0.23	
	Other	Bone fractures	Proportion of patients experiencing bone fractures.	3	0	13.4	Inverse linear	%	8.33	5.3
	Bladder	Proportion of patients contracting bladder cancer.	1.00	0.00	8.4	Inverse	%	0.27	0.22	



* Determines how measured values are to be converted into 0-100 preference values. Can be direct (larger measures are more preferred) or indirect (smaller numbers are more preferred) and either of these can be linear or non-linear.

** The effect of non-linear value functions on the benefit-risk balance were considered. See text at p. 21, Figures 17-19.

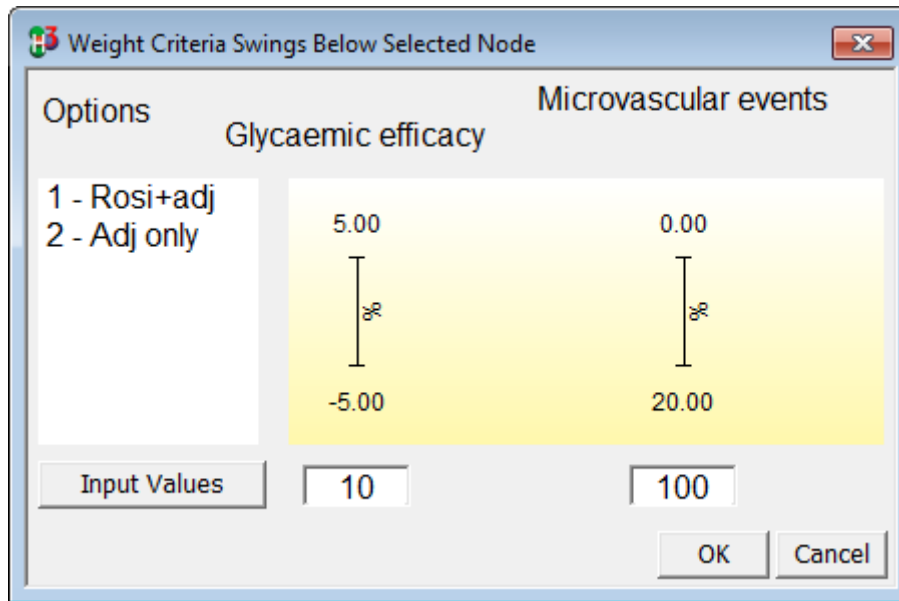


Figure 2: The swing-weights assigned to the two Favourable Effects scales.

During the June decision conference participants began the weighting process by assessing weights for the favourable effects; Figure 2. The group agreed that the swing from 20% to 0% on the Microvascular events scale was better than the -5% to +5% increase in Glycaemic efficacy improvements, so the Microvascular events scale was assigned a weight of 100. Compared to that, the group judged the swing on the Glycaemic efficacy scale to be low, for it is a surrogate endpoint, so participants agreed a weight of 10.

Swing weights for the four MACE criteria are given in Figure 3. Once again, the largest swing was given a weight of 100, though that 100 does not necessarily represent the same magnitude of preference difference as the 100 in Figure 2, an issue addressed at a later stage, below.

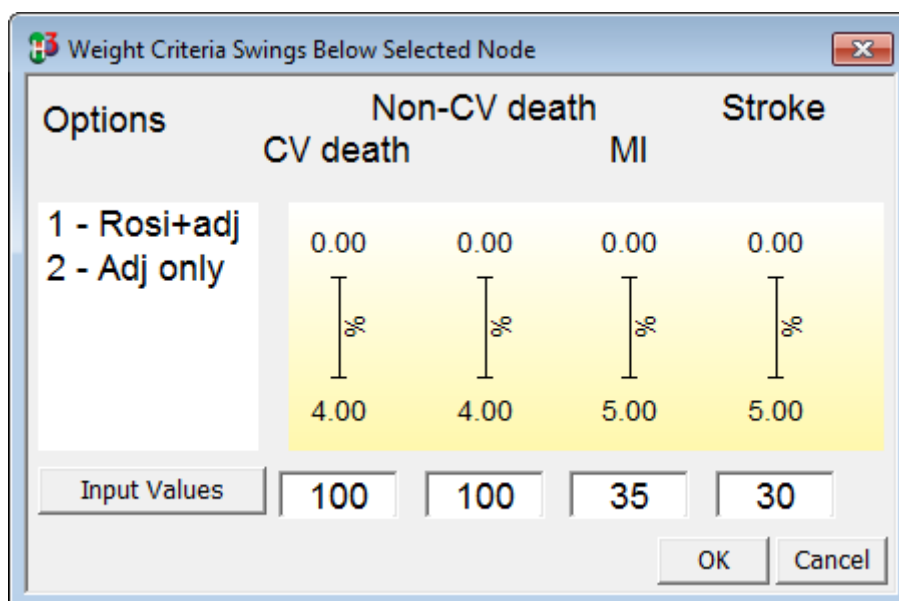


Figure 3: Swing weights assigned to the four MACE criteria.

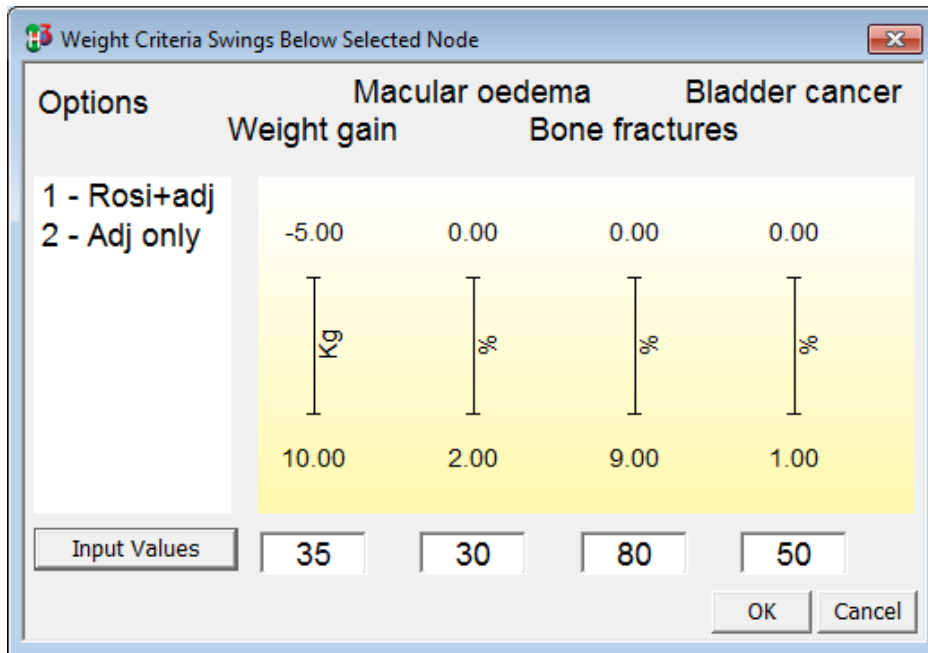


Figure 4: Swing weights for the four Other criteria.

Swing weights for the four Other criteria were revised at the July meeting. The group decided that the 0-3% difference in Bone fractures was more clinically relevant than the other three differences, so that effect scale was given a weight of 100; the weight gain weight was reduced and the bladder cancer scale's weight was reduced. Eventually, the weights in Figure 4 were agreed.

So why is the weight on Bone fractures in Figure 4 not 100? The reason is that the next assessment compared the swing on CHF with the two 100-weighted criteria, one each under MACE and Other, Figure 5. Note that the swing on Non-CV death was judged to be the most clinically relevant, followed closely by a tie at a weight of 80 for CHF and Bone fractures. The weight of 80 on Bone fractures reduced all four swing weights on the Other effects by 80%, which accounts for the 80 on this effect in Figure 4. The other three weights in Figure 4 were arrived at by the group only after the original 100 had been reduced to 80, and were the result of consistency checks. For example, after the group had reassessed the other three weights, Larry asked if reducing both Macular oedema from 2 to 0% and reducing Bladder cancer from 1 to 0% was equivalent in added clinical value to reducing Bone fractures from 9 to zero percent. After some thought, the two clinicians in the group said that felt about right.

The final, and most difficult comparison, is shown in Figure 6: Microvascular events versus Stroke, the most heavily weighted Favourable and Unfavourable Effects, respectively. After considerable debate, the group agreed that the Microvascular events' swing, from 20% down to zero patients, was half the clinical relevance of the reduction in Stroke, from 5% to 0% of patients, so weights of 50 and 100, respectively, were assigned.

It is this process of comparing swings from least to most preferred positions on the criteria associated with a node, assigning one criterion swing a weight of 100, then comparing the 100-weighted criteria across the nodes, which ensures the comparability of the units of preference values across all the criteria.

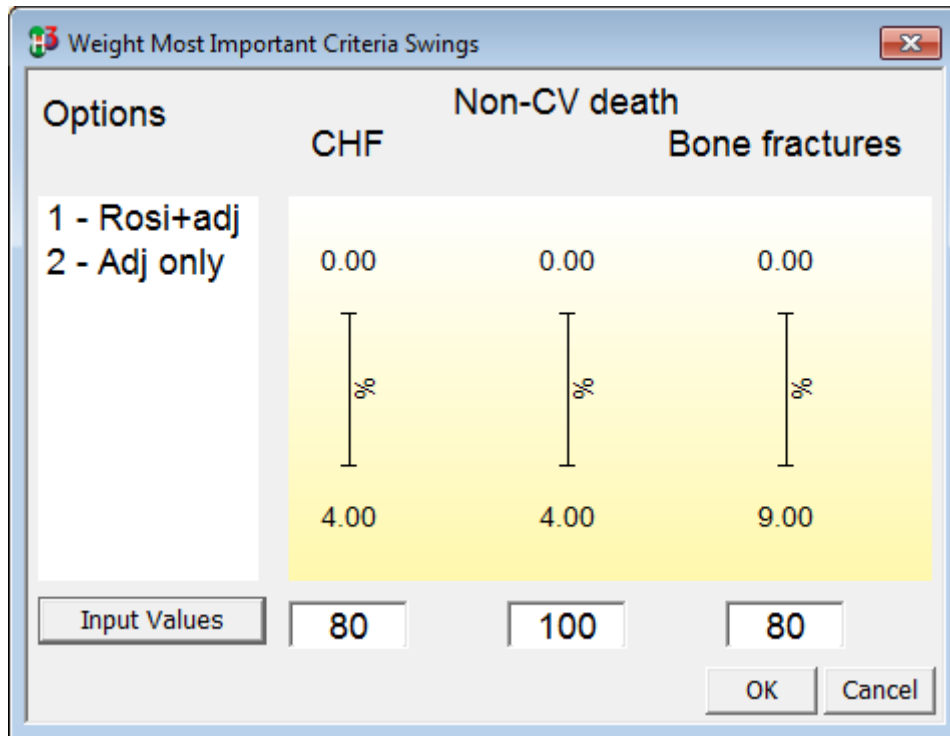


Figure 5: Swing weights for the two 100-weighted scales under MACE and Other, compared to each other and to the CHF scale.

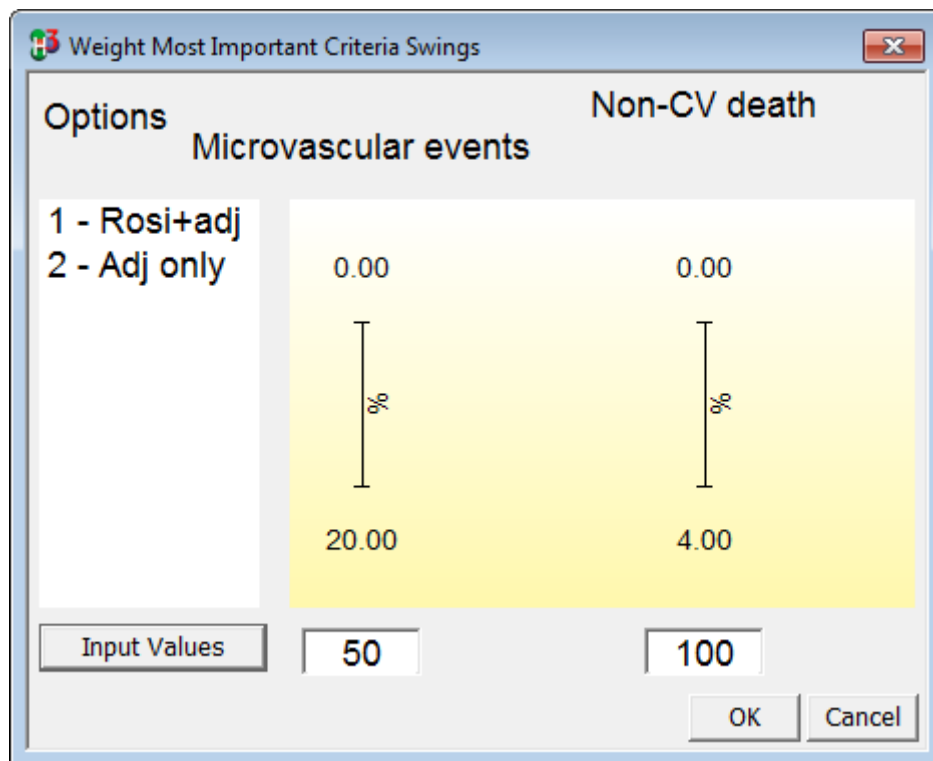


Figure 6: Swing weights comparing Microvascular events to Non-CV death.

It is easy to become lost in attempting to understand the weighting process by reading about it, so Figure 7 shows all the originally-assessed weights (with the exception noted in the legend), each divided by 100, on the value tree. Hiview multiplies these weights of 1.0 or less along each path through the tree, sums the products for all 11 criteria and divides each product by the sum. This gives the cumulative weights shown in Figure 9, re-normalised to 100, with the criteria sorted in order of the cumulative weights.

What prompted the revisions in the July decision conference of weights for the Unfavourable Effects? It was the result of checking the realism of the cumulative weights. For example, in the June decision conference MI was positioned third from the bottom when the effects were ordered by their cumulative weights. This seemed too low; the clinical relevance of that scale's range was judged to be higher than the range for Weight gain. And the weight for Stroke was judged to be too high compared to MI. Several iterations of the weights were required before the cumulative weights produced an ordering of scale ranges that seemed realistic, particularly to the clinicians in light of their experience.

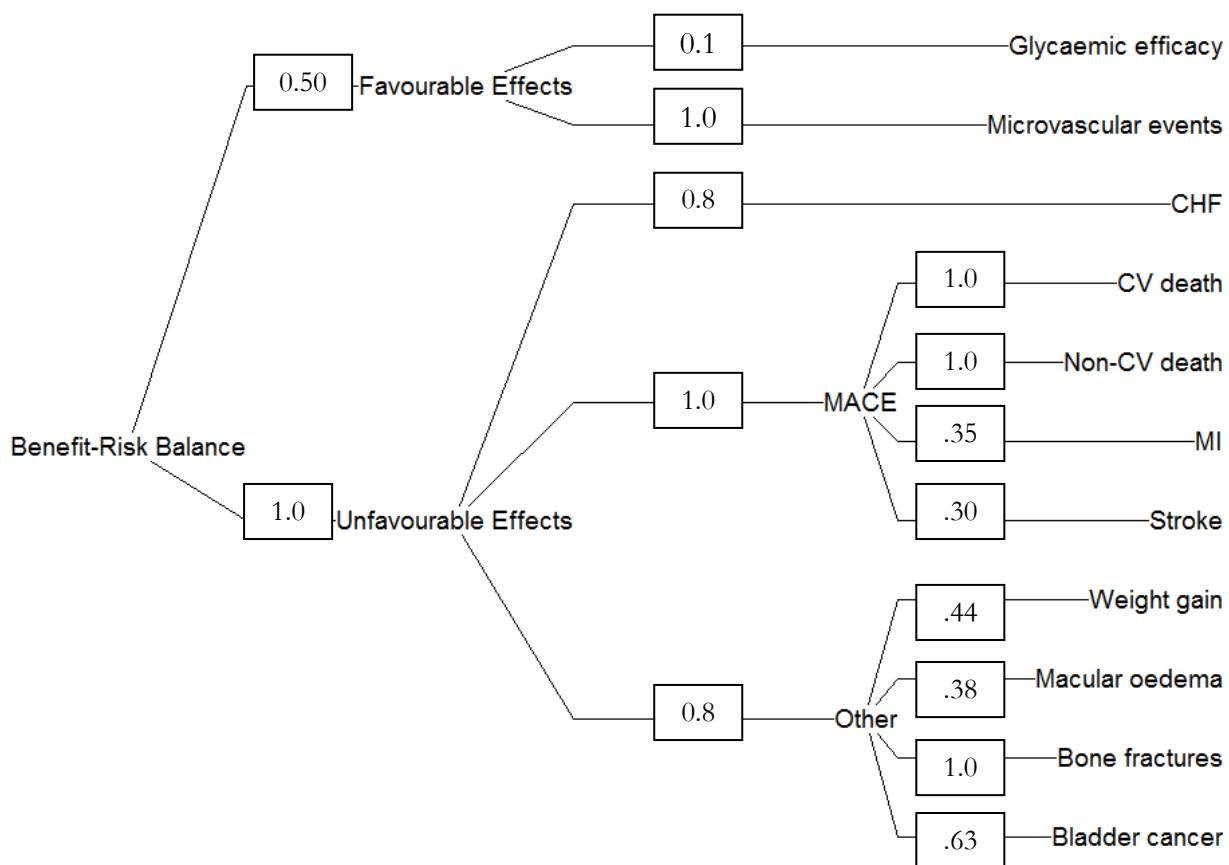


Figure 7: The originally-assessed swing-weights (except for the Other criteria, where 0.8 times the weight in the box gives the weights as assessed in Figure 4), divided by 100, assigned at all the nodes. The products of the weights on any path equal the cumulative weights shown in Figures X and X.

5.4 Scoring

MCDa converts all input data into preference values. In general, favourable effects are converted so that larger input metrics are represented by higher preference values, while for unfavourable effects an inverse relationship is used so that larger input metrics result in lower preference values. The relationships for all effects are usually linear, but some effects, like QTc prolongation or numbers of PML cases are typically non-linear.

The mapping of input metrics to preference values is a matter of clinical judgement, another way in which MCDa helps to make explicit what is usually implicit. As a result, deciding whether or not the conversion should be linear or non-linear is a difficult call to make. So, to start, the group decided to leave all value functions as linear. Thus, once the high and low values of metric scales were defined and data entered into Hiview, the computer carried out the conversion. Figure 8 gives an example for CHF, which illustrates a linear, inverse relationship. Later, the group explored non-linear functions for CHF and Non-CV deaths, reported in the Non-linear value functions section, below.

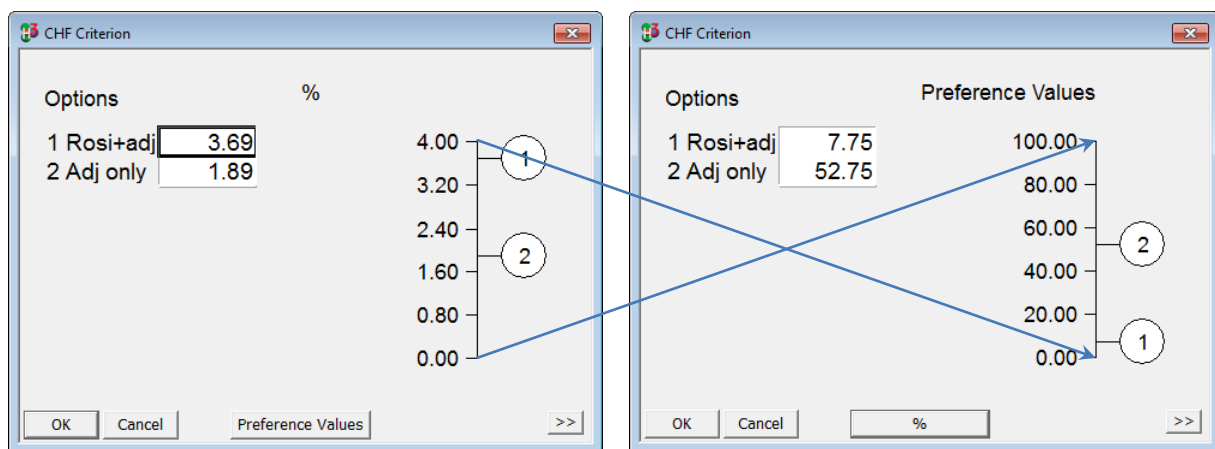


Figure 8: The inverse linear relationship for CHF that converts the input data on a percentage scale ranging from 0 to 4%, left, onto a preference value scale ranging from 0 to 100.

6. RESULTS

6.1 Cumulative weights

As explained above, the group often carried out consistency checks by looking at the cumulative weights for all the criteria to see if they seemed realistic. Recall that cumulative weights are the final normalised weights associated with the effects themselves, obtained by multiplying the weights in Figure 7 along each branch of the Effects Tree, adding them and dividing each by that sum. The finally-agreed weights are shown in Figure 9.

		Cum Wt	Sum	
MACE	CV death	16.8	16.8	
MACE	Non-CV death	16.8	33.6	
Other	Bone fractures	13.4	47.1	
Unfavourable Effects	CHF	13.4	60.5	
Other	Bladder cancer	8.4	68.9	
Favourable Effects	Microvascular events	8.4	77.3	
MACE	MI	5.9	83.2	
Other	Weight gain	5.9	89.1	
MACE	Stroke	5.0	94.1	
Other	Macular oedema	5.0	99.2	
Favourable Effects	Glycaemic efficacy	0.8	100.0	
		100.0		

Figure 9: Cumulative weights of all the criteria, with the criteria ordered by the size of their cumulative weights, which represent the swings in preference from the least to the most preferred positions on the scales.

It is important to keep in mind that a cumulative weight represents the total added preference value in moving from the least to most preferred positions on a scale; it is not defined as the difference between the options. These weights represent the relative importance of the 0-100 preference value ranges on the scales, not the relative importance of favourable and unfavourable effects, and particularly not the relative importance of those effects for the drug and comparator. By summing cumulative weights, it is possible to see the weights at each node. For example, the sum of all the favourable effects weights is 9.2 with 90.8 for the unfavourable effects. In other words, the added preference value of the 0-100 differences in preference values on the unfavourable effects is ten times the range of that on the favourable effects.

Note that the scale ranges for two of the MACE criteria were considered by the group to be most clinically relevant: 4% to 0% for each of CV death and non-CV death. (And, that the lowest cumulative weight appears on Glycaemic efficacy, the primary endpoint!)

In short, consistency checks using the cumulative weights helped participants to construct realistic and consistent weights for the individual scale ranges. It is these numbers that enabled intuition based on clinical experience to be made explicit, debated and shared.

6.2 Overall benefit-risk balance

With scoring and weighting completed, it was possible to calculate sums of weighted scores and display preliminary results at any node. The key calculation behind MCDA adds all weighted value score by applying the equation,

$$\text{Aggregate Value} = \sum_{i=1}^n w_i \times v_i(c_i) \quad (1)$$

where c_i is the input value of criterion i , $v_i(c_i)$ is the preference score for criterion i and w_i is the cumulative weight for criterion i . The sum is taken across all $n=11$ criteria.

Figure 10 shows the relative scores at the Benefit-Risk Balance node of Figure 1 as stacked bar graphs. Note that longer green bars represent *more benefit*, while longer red bars represent *more safety*. Adding rosiglitazone to the adjunct therapy provides hardly any beneficial advantage and substantially increases the risks.



Figure 10: The overall weighted preference values for rosiglitazone plus adjunct compared to adjunct only. Favourable effects, the upper green bars, show slightly more effectiveness from adding the drug to the adjunct, while the shorter red bar of rosiglitazone + adjunct shows that the slight added effectiveness is accompanied by substantial extra risk. The Cumulative Weight column shows the normalised weight on the FE and UFE nodes: favourable effects weighted about one-tenth as much as unfavourable effects.



Figure 11: The overall weighted preference scores of the two options, with the stacked bar graphs showing the contribution of each effect to the overall score. The right column shows the cumulative weights, normalised to 100, of each of the criteria. Bone fractures (green), for example, is 13.4.

The stacked bar graphs can also be shown for their separate contributions from the criteria, as seen in Figure 11. This instructive display shows the problem with rosiglitazone: the small CHF (yellow) and Bone fracture (light green) slices, indicating greater risk, for rosiglitazone plus adjunct.

6.3 Comparative Analyses

A display of the differences between the two options can be seen broken down into the contribution of each effect in Figure 12. This shows the ways in which rosiglitazone plus adjunct compares to the adjunct only, taking account of both the data and the clinical relevance of the data. The Diff column shows the difference in the preference scores, while the Wtd Diff column multiplies that difference by the cumulative weight on the criterion. It is this weighted difference display that reveals the advantages and disadvantages of the comparisons.

Interestingly, the right-extending (green) bars show that rosiglitazone plus adjunct is better than adjunct only for three MACE effects and for lower microvascular events, though the latter is a negligible advantage. The left-extending (red) bars show the many effects that favour adjunct only. The sum of the lengths of the green bars minus the sum of the lengths of the red bars equals the overall 8-point difference (ignoring some rounding difference) in the overall weighted preference values of the two options.

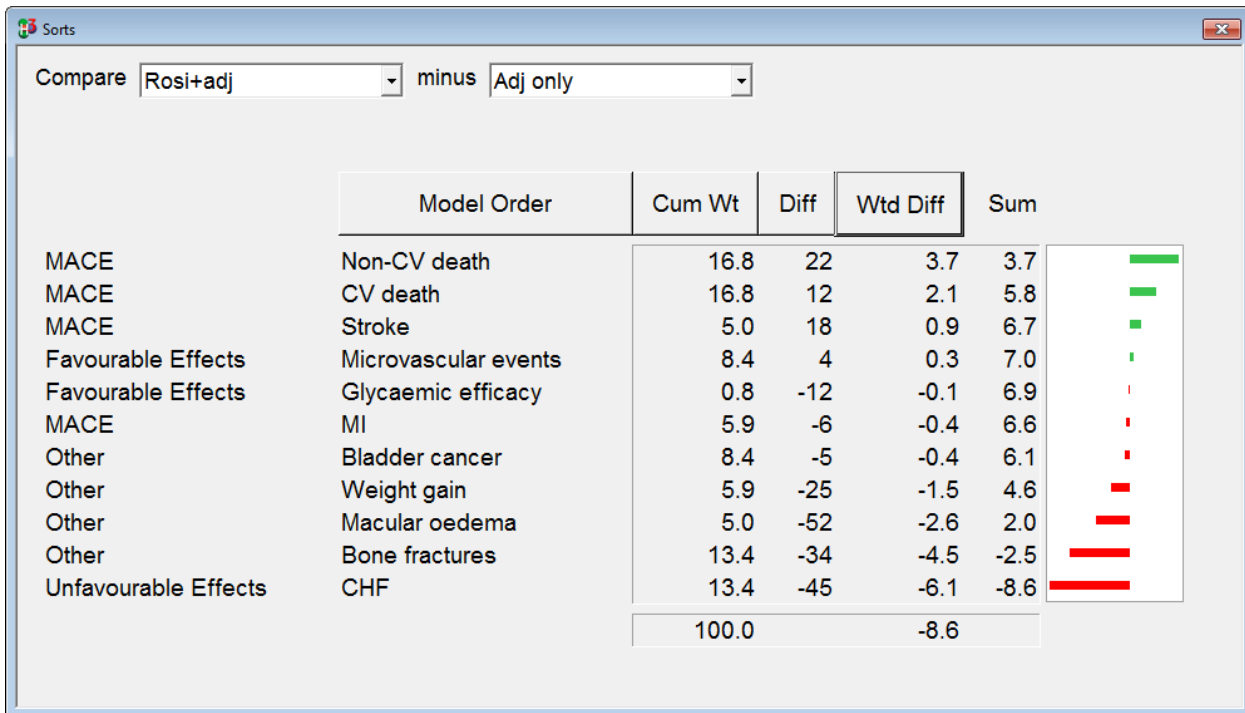


Figure 12: Rosiglitazone plus adjunct compared to the adjunct only. Rosiglitazone as adjunct is preferred to the adjunct only on three MACE effects, with a tiny advantage on Microvascular events, for a total weighted score of 7.0 (shown in the Sum column). But that seven-point advantage is outweighed by the total weighted difference scores on the Other unfavourable effects and particularly by CHF.

Note that the first two advantages of rosiglitazone plus adjunct, Non-CV death and CV death, sum to 5.8, which is overbalanced by the single CHF advantage of adjunct only. And adjunct only even has a slight advantage on the primary endpoint, Glycaemic efficacy.

6.4 Sensitivity Analyses

These analyses explore the sensitivity of the overall results to changes in weights on the criteria, which were the source of much of the debate about the balance of benefits and risks. The first analysis examined the weight on the CHF effect. The normalised weight on the base-case model described above was 13.4, as shown in the right column of Figure 11. The computer varied that weight over its entire feasible range, 0 to 100, with the result shown in Figure 13.

The vertical red line intersects the horizontal axis at 13.4, and its intersection with the red and green lines give the overall scores for the two options, 35 for rosiglitazone plus adjunct and 43 for adjunct only. Either decreasing the weight or increasing it makes no difference: adjunct only remains the more preferred option.

Would increasing the weight on Non-CV death tip the benefit-risk balance in favour of rosiglitazone plus adjunct? Figure 14 shows that a substantial increase, to more than double the current value, would be required to tip the balance. Figure 12 shows that three of the four MACE events favour rosiglitazone plus adjunct. It's an interesting thought that post-marketing data suggest that what would normally be considered risks are actually benefits for this drug.

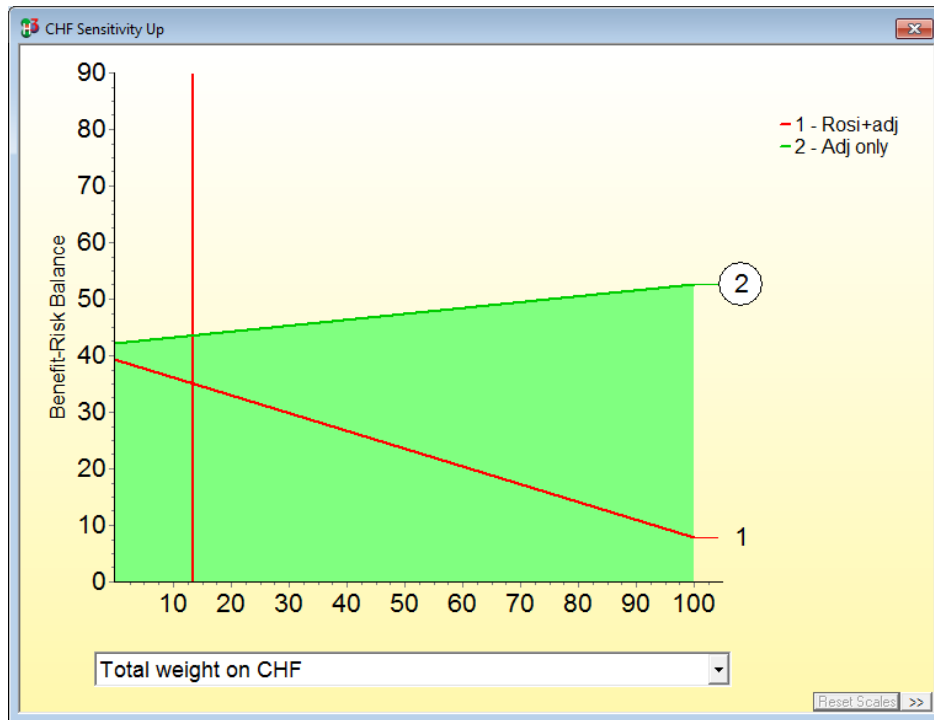


Figure 13: Neither decreasing nor increasing the weight on CHF from its current value of 13.4, the vertical (red) line, will affect the preference for adjunct only over rosiglitazone plus adjunct.



Figure 14: Only a substantial increase in the cumulative weight on Non-CV death, so this criterion alone accounts for 40% of the overall results, would rosiglitazone plus adjunct become more preferred than adjunct only. (The green shading does no more than identify crossover points.)

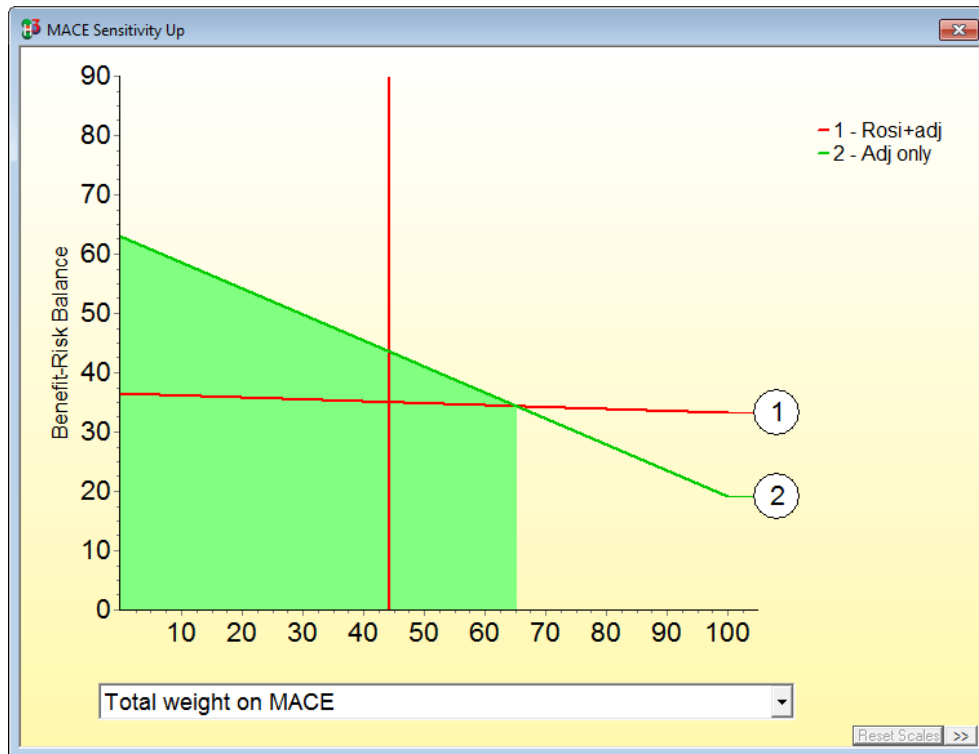


Figure 15: Increasing the weight on MACE from its current value of 44.3 by about 20 points will just tip the FE-UFE balance in favour of rosiglitazone plus adjunct.

A further sensitivity analyses at the MACE node shows the effect of making all four MACE events simultaneously more important. Figure 15 indicates that a 20-point increase in the weight on that node, which currently accounts for about 45% of the overall results, could indeed tip the benefit-risk balance favouring rosiglitazone plus adjunct.

Additional sensitivity analyses showed that the current weights would have to be very seriously changed before the benefit-risk balance would tip in favour of the drug. Figure 16 summarises the results of separate sensitivity analyses on each of the criteria. As indicated in the legend, this model is very robust to changes in weights—any reasonable set of weights will always show the overall superiority of adjunct only.

6.5 Non-linear value functions

Near the end of the July decision conference, participants explored the effects of non-linear value functions, starting with CHF. The clinicians were mainly responsible for this graph. They started by saying that a value of 1% would not be considered too serious, so moved that point to a preference value of 90. They judged the next drop in preference value, in going from 1% to 2%, to be twice as serious, so moved down to 70. A further increase in the drop-rate attended an increase from 2% to 2.5%, but then the rate started to decrease as the data approached 4%. Although still unsure about this, the clinicians judged Figure 17 to be a plausible non-linear function.

The effect of this function on the translation of input data to preference values can be seen in Figure 18; it has clearly increased the gap between rosiglitazone plus adjunct and adjunct only. The consequence for the overall result is that the overall weighted preference value for rosiglitazone plus adjunct remained at 35, but adjunct only increased to 46, an 11-point gap whereas before it was 8 points.

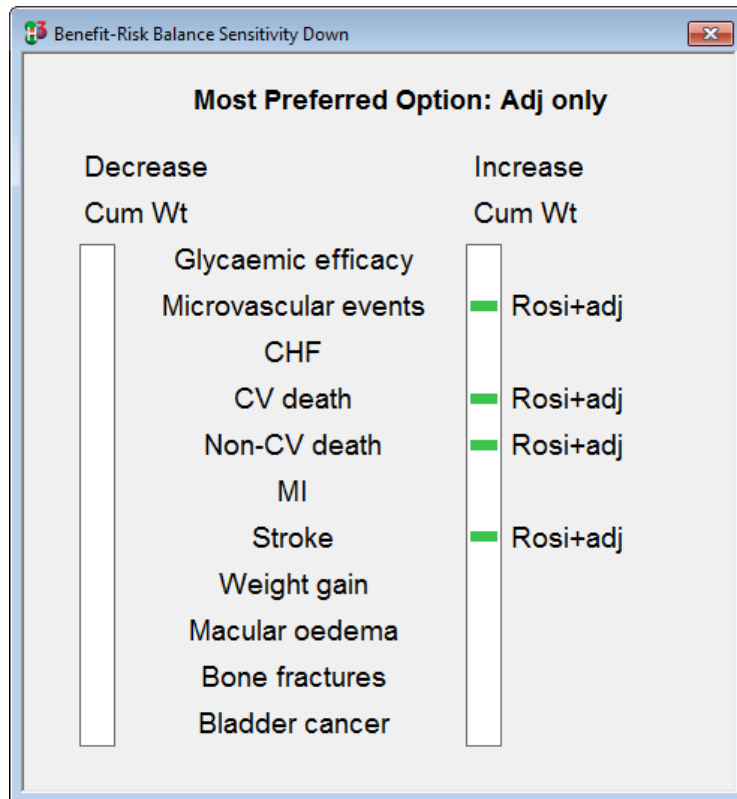


Figure 16: Separate sensitivity analyses on each of the criteria shows how the most preferred option, adjunct only, would change as the cumulative weight on a criterion is decreased or increased. Green bars show cumulative weight changes greater than 15 points are needed to shift the overall preferences. The absence of yellow bars, which signal a shift to the drug resulting from a weight change of 5 to 15 points, or of red bars, which signal a shift from a weight change of less than 5 points, indicates a very robust model; precision of weights isn't needed. No bars at all mean that changing that weight over its entire range results in no change of the more preferred option (as was demonstrated for CHF in Figure 13).

The comparator was better because it benefitted from the concave (looking upward) portion of the value function, which increased its preference value compared to the linear value function. This result prompted the group to explore the effect of a convex value function, which was constructed for Non-CV deaths, shown in Figure 19. Because the input data for both alternatives were between 3% and 4%, this function decreased the difference between the preference values, resulting in a further increase in the overall difference in preference values of 32 and 46.

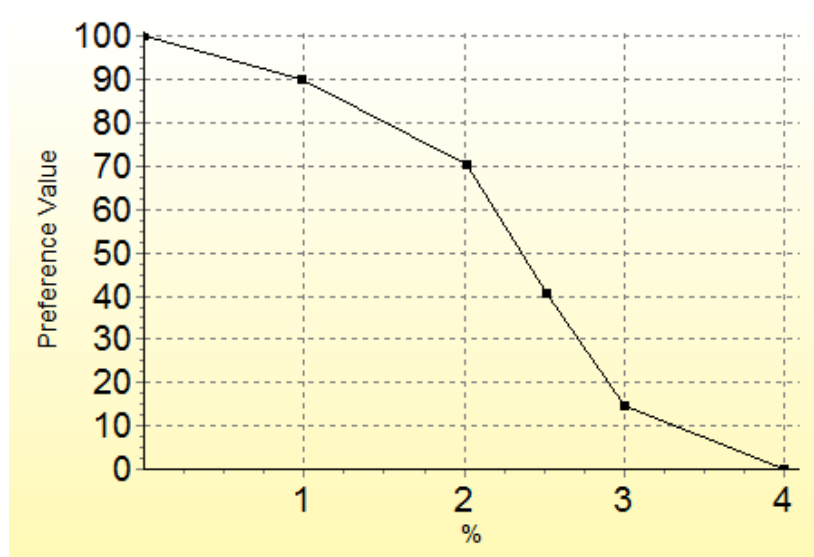


Figure 17: A non-linear, inverted S-shaped value function for CHF.

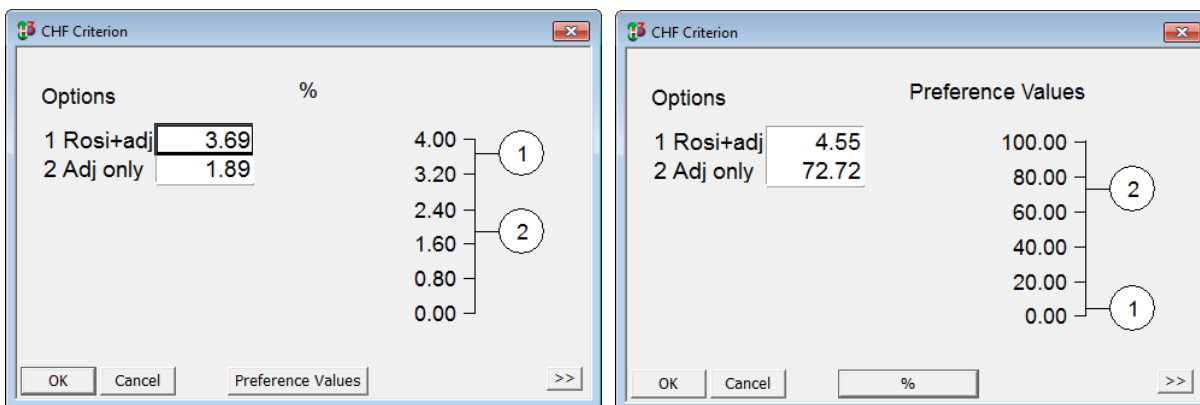


Figure 18: Input data for CHF, left, and the corresponding preference values, right, when the mapping is based on the inverse, non-linear value function shown in Figure 17. The inverted S-shape of the function has widened the difference between the two alternatives.

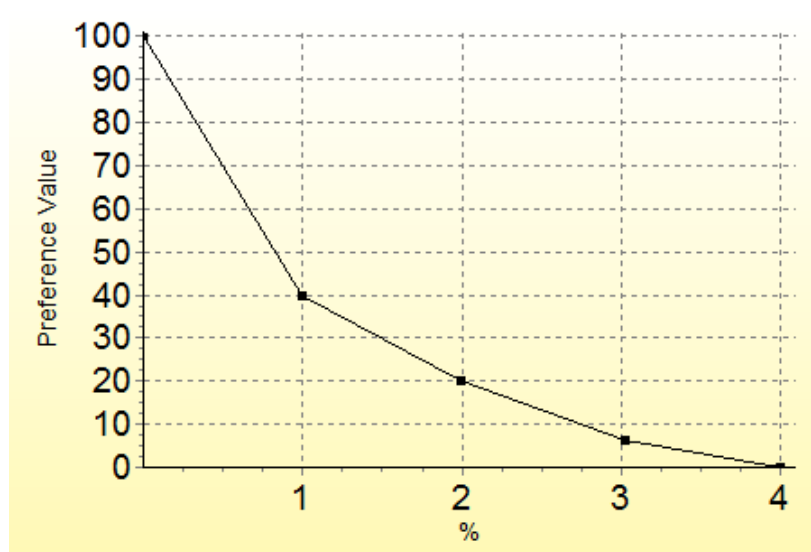


Figure 19: A non-linear, convex inverse value function for Non-CV deaths.

7. DISCUSSION

Perhaps the most notable feature of the two days the rosiglitazone team spent working together was not the model and its results, which confirmed the decision by European regulators to withdraw the use of rosiglitazone, but rather the process by which we arrived at a workable model. By bringing together participants with a wide variety of perspectives, scientific, epidemiological, statistical and clinical, and providing impartial (for the most part) facilitation, everyone could contribute to the discussion and engage in vigorous discussions that were focussed on the specific topics derived from the PrOACT-URL framework. For example, debate about data for Micro-vascular events led to the realisation that the definition of this favourable effect developed in the June decision conference was not clear. Once a new, more satisfactory definition was agreed, it eventually became clear that available data for pioglitazone plus adjunct could not be compared to the available data for the other two alternatives, with the result that the pioglitazone alternative was dropped.

Even the comparatively simple task of entering data brought about some changes. For example, the original 0 to 2% ranges for the measurement scales for CV-death and Non-CV death proved to be too narrow, for the data ranged up to 3.69%. Thus, the ranges were changed to 0-4%, which required an adjustment of their weights, as these larger swings were judged by the group to be the most important. Giving each of those effect ranges weights of 100 requires lowering the weights of the other two MACE criteria. Further shaping of preferences came about by testing different assumptions and judgements using the model, and then seeing the effects on the overall results. This on-the-spot feedback provided useful learning and helped us to form realistic and consistent preferences.

This quantification and shaping of judgements through a process of deliberative discourse and model-testing helps to make explicit what is normally implicit. Two people can agree that something is 'important', but when swing-weights are assessed and one person gives a weight of 40 and another person gives a 90, it is worth exploring the difference. An exchange of views and further discussion by other participants often uncovers reasons that some had not thought about, and a consensus emerges, not as a compromise that satisfies no one, but as an agreed 'good enough' value that reflects the collective experience and judgements of the group. In this way, assessing the benefit-risk balance becomes a collaborative process.

It is worth commenting here on the occasional frustration experienced by the group at the difficulty of interpreting published data. Authors of published papers often report only sample sizes, means, confidence intervals and significance levels. While this may be sufficient for making statistical inferences, it may not be adequate for the purposes of an MCDA, particularly for sensitivity analyses. A confidence interval for a mean is, of course, an interval describing the sampling distribution, not the underlying distribution of patients' data. Without more information about that distribution, it is not possible to determine the range of patients' data or any percentiles of that underlying distribution. Thus, even with confidence intervals, exploring the sensitivity of the overall benefit-risk balance to lower data values for favourable effects and higher values for unfavourable effects, will lead to limited conclusions.

At this stage in the project, we were unable to find any individual patient data. This limited the next stage of the project, which was to replace all input scores with probability distributions so that simulations of the difference in the benefit-risk balance of rosiglitazone plus adjunct compared to adjunct only could be determined. Thus, we had to make realistic assumptions about those distributions to explore how the balance might be changed.

Was the time spent on modelling rosiglitazone worth the trouble? One participant observed that working in a group brings more than the sum of individual efforts. Another reported having learned a lot from the experience. The following are un-edited comments from participants, obtained for the first draft of this report, who answered the question, "What are the advantages and disadvantages of the decision conferencing approach (on-the-spot modelling in a workshop with impartial facilitation) to improving decision making about new drugs?"

“It requires each participant to explain publicly his/her opinion to other stakeholders, it requires that participant to construct the argument with some rationale instead of a rough global intuition.”

“In addition the public sharing of discussion leads to a cross fertilisation in individual experiences or knowledge, which in return contributes to the evidence-based or rationale of the expressed opinion.”

8. PART II: THE PROBABILISTIC BENEFIT-RISK MODEL³

Currently most regulators of medicinal products rely heavily on statistical averages in judging the benefit-risk balance for approving drugs. We used such data for the model in Part I. However, because the statistical average fails to represent the full probability distribution of the effect, a skewed distribution can lead to an erroneous impression of a drug's effect. Therefore, decisions made regarding to the approval of drugs based purely on statistical summaries might not be satisfactory.

This section reports our attempt to discover whether or not the additional information provided by probability distributions for all rosiglitazone effects could alter the drug's modelled benefit-risk balance. We investigate whether or not incorporating probability distributions in the MCDA framework provides a more intuitive evaluation of the benefit-risk balance for rosiglitazone as an adjunct therapy.

In order to test this approach, we carried out a meta-analysis of the clinical trial data to combine data from various studies with respect to the uncertainties for the MCDA model's 11 criteria. We modified the MCDA model reported in Part I to incorporate probability distributions for each of the favourable and unfavourable effects. This was accomplished by exporting the deterministic MCDA model to EXCEL, replacing all the input data with the appropriate probability distributions, and then using @RISK, an EXCEL add-in program, to carry out a Monte Carlo analysis over all the benefit-risk effects. This resulted in a distribution of the benefit-risk difference between the two treatment alternatives. Sensitivity analyses tested the robustness of the probabilistic model. We conclude this section with a discussion of our findings, including the limitations of this model and key areas for improvements and future work.

8.1 Data

We dealt with two types of data: continuous and binary data endpoints. Due to their differing characteristics, the two types of data needed to be handled differently. Continuous data endpoints refer to data that can be primarily measured as a number, often as the average level of specific effects. For example: the mean Glycaemic efficacy percentage increased in trial 1 is 8.9%. Binary data endpoints usually refer to the data which cannot be primarily measured on a continuous scale. They were typically obtained with a yes or no question at a patient level, for example: has the patient experienced bone fracture? The total number of patients in that trial who suffered from bone fracture is counted and the data are usually presented as 56 patients out of the sample size of 189 having experienced bone fractures. It is important to distinguish these two types of data since they were dealt with differently in the meta-analysis and simulation process.

8.2 Meta Analysis

Guo et al (2010) point out that "data extraction from clinical trials is critical for the internal validity assessment of the MCDA technique". This is because clinical trial data is characterised by high levels of uncertainty because the same drug is very likely to have different effects for different patients. The accuracy of each trial depends on the sample size and trial design, which typically differ from one trial to the next. Therefore, simply averaging the summary data from different trials requires a sophisticated and systematic method of combining different clinical data.

The data used for assessing new drugs are usually sourced from multiple studies. In the Book *Clinical Trials*, Wang and Bakhai (2006) describe meta-analysis as "a systematic method of combining the results of multiple similar studies to allow more accurate conclusions to be drawn from a larger pooled number of studies." By including the population of each study, the treatment effect can be estimated with greater accuracy. Although the process can be time consuming, it is still less expensive than conducting a new, larger trial. Meta-analysis typically produces the results of an overall mean treatment difference between two treatment groups and a confidence interval. Wang and Bakhai (2006) give five basic steps involved in a meta-analysis:

³ This section is based on Chen Chen's project report, *Modelling the benefit-risk balance of rosiglitazone based on statistical summaries and probability distributions; accommodated by MCDA and probabilistic simulations*, that was part of her MSc thesis awarded in 2013 by the London School of Economics and Political Science.

8.2.1 Step 1: Formulation of the Study Question

This step requires the aim of the analysis to be clear and specific. In this case, the aim of our meta-analysis is to explore whether patients using Rosiglitazone plus adjunct therapy perform better than patients with adjunct therapy only in the 11 effects identified.

8.2.2 Step 2: Literature Research

It is important to gather as much information as possible. Some common sources for clinical trial data are electronic medical databases, published books on similar subjects, National or European registers and specific pharmaceutical companies.

The Clinical trial data used in this project was gathered by Dr Edmond Chan. It initially contains 32 clinical trials for the period of 2000-2009. The trial data were sourced from the European Medical Agency (EMA), the National Center for Biotechnology Information (NCBI), the New England Journal of Medicine (NEJM) and the GlaxoSmithKline (GSK) Clinical Study Register. A summary of the initial trial data is shown in Table 2.

Table 2: List of studies from the literature search

Study	Date	Reference	Source	Type
EPAR	2006	European Medicine Agency Scientific Discussion Report	EMA website	EPAR
Fonseca	2000	The Journal of American Medical Association 2000 Apr 5;283(13):1695-702.	NCBI website	RCT
Kipnes	2001	The American Journal of Medicine 2001 July; 111(1): 10-7	NCBI website	RCT
Einhorn	2000	Clinical Therapeutics. 2000 Dec; 22(12):1395-409.	NCBI website	RCT
Gomez-Oerez	2002	Diabetes/ Metabolism research and Review 2002 Mar-Apr;18(2):127-34.	NCBI website	RTC
Vongthavarat	2002	Current Medical Research & Opinion 2002;18(8):456-61	NCBI website	RCT
Richter	2009	Cochrane Database of Systematic Reviews. 2007 Jul 18;(3):CD006063.	NCBI website	Meta Analysis
Home	2007	New England Journal of Medicine 2007 July 5; 357:28-38	NEJM website	RCT
Nissen *	2007	New England Journal of Medicine 2007 June 14; 356:2457-2471	NEJM website GSK CLINICAL REGISTER	RCT
Bakris	2003	Journal of Human Hypertension. 2003 Jan;17(1):7-12.	NCBI Website	RCT
Bakris	2006	Journal of Human Hypertens. . 2006 Oct;24(10):2047-55.	NCBI Website	RCT
Home et al RECORD	2009	Lancet. 2009 Jun 20;373(9681):2125-35. Epub 2009 Jun 6.	NCBI Website	RCT
<p>* Nissen 2007 consists 21 clinical trials, and they are treated as separate trials throughout this project. Study in light green shade were selected and used after step 3 Nissen 2007 was reduced to 18 trials after step 3. EPAR = European Public Assessment report RCT = Randomised Controlled Trial</p>				

8.2.3 Step 3: Study Selection

Not all studies were retained for the meta- analysis; some selection was required to identify quality studies to be combined. Some common criteria identified by Wang and Clemens (2006) are:

Trial type : randomised controlled trials (RCTs) vs non-RCTs

Treatment Strategies: Rosiglitazone plus Adjunct vs Adjunct only

Primary Outcomes: criteria reported by the study, such as bladder cancer.

Data Availability: whether there are missing values on important attributes.

After step 3, the 5 studies identified to have similar features were: Fonseca (2000), Home et al (2007), Nissen (2007), Bakris (2006), and Home et al (2009). In particular Nissen (2007) reduced from 21 trials to 18 trials, plus the other remaining 4 studies gives a total number of 22 clinical trials. All clinical trials used are in the randomised controlled trial (RCT) type and the sample sizes vary from 43 to 2227. All trials provide one or more data sets on the various effects of patients using rosiglitazone plus adjunct therapy and using adjunct therapy only.

8.2.4 Step 4: Data Extraction and Quality Assessment

After we identified the final set of studies, we extracted information on the 11 effects in a standard form for the purpose of further statistical analysis. The data are most commonly presented in the form of mean effect size, standard deviation (S.D), standard error (S.E) and 95% confidence interval (CI). Since different studies reported different combinations of these statistics, we converted them into a standard form for ease and clarity of subsequent statistical analyses. The formulae used to convert between S.D, S.E and CI are:

$$S.E = S.D/N$$

$$S.E = (UL-LL)/1.96/21$$

where N is the sample size, UL is the confidence interval upper limit, and LL is the confidence interval lower limit. The statistical summaries for continuous and binary data endpoints are presented in a standard format in tables 3 and 4. These are the input values for the final step of meta-analysis.

Table 3: Extracted data summary for binary data

Study	Year	Type	Rosi+Adjunct			Adjunct Only		
			Sample Size	Effect Mean	S.E	Sample Size	Effect Mean	S.E
Fonseca	2000	RCT	110	8.900	0.143	113	8.600	0.151
Nissen	2007	49653/093 (1)	105	-0.700	0.127	95	0.100	0.121
		49653/094 (2)	113	-0.780	0.115	110	0.450	0.111
		49653/211 (5)	108	7.700	0.125	109	7.300	0.115
		49653/125 (11)	175	-1.130	0.116	173	0.090	0.075
		49653/127 (12)	56	-0.140	0.147	58	0.000	0.125
		49653/147 (15)	89	-1.200	0.143	88	0.450	0.115
		49653/162 (16)	168	-0.910	0.076	172	-0.140	0.070
		49653/234 (17)	59	-1.174	0.160	61	-0.079	0.162
		49653/132 (21)	221	-1.900	0.093	112	-0.500	0.114
Bakris	2006	RCT	194	0.720	0.007	180	0.920	0.006
Home et al	2009	RCT (1)	1117	-0.280	0.030	1105	0.050	0.040
		RTC (2)	1103	-0.440	0.030	1122	-0.180	0.040
Weight Gain [kg]								
Bakris	2006	RCT	194	1.940	2.362	180	1.500	1.801
Home et al	2009	RCT (1)	1117	3.800	0.240	1105	0.000	0.200
		RTC (2)	1103	4.100	0.200	1122	-1.500	0.200

Table 4: Extracted data summary for continuous data

Study	Year	Type	Rosi+Adjunct		Adjunct Only	
			Sample Size	Event	Sample Size	Event
Home et al	2009	RTC	2220	59	2227	78
CHF [%]						
Home	2007	RTC	2220	38	2227	17
Nissen	2007	49653/094 (2)	113	0	110	0
		49653/125 (11)	175	0	173	1
		49653/127 (12)	56	0	58	0
		49653/135 (13)	116	2	111	3
Home et al	2009	RCT	2220	82	2227	42

CV Death [%]						
Home	2007	RTC	2220	29	2227	35
Nissen	2007	100684 (3)	43	0	47	1
		49653/143 (4)	121	0	124	0
		49653/211 (5)	108	3	109	2
		49653/284 (6)	382	0	384	0
		712753/008 (7)	284	0	135	0
		AVM100264 (8)	294	2	302	1
		49653/125 (11)	175	0	173	0
		49653/135 (13)	116	1	111	1
		49653/147 (15)	89	0	88	0
		49653/162 (16)	168	0	172	0
		49653/234 (17)	59	0	61	0
		49653/137 (18)	204	0	185	1
		SB-712753/002 (19)	288	1	280	0
		SB-712753/003 (20)	254	0	272	0
		49653/132 (21)	221	1	112	0
Bakris	2006	RCT	194	1	180	0
Home et al	2009	RCT	2220	60	2227	71
Non CV Death [%]						
Home	2007	RTC	2220	45	2227	45
Nissen	2007	49653/135 (13)	116	0	111	0
		49653/147 (15)	89	0	88	0
		49653/162 (16)	168	1	172	0
		49653/234 (17)	59	0	61	0
Home et al	2009	RCT	2220	66	2227	86
MI [%]						
Home	2007	RTC	2220	43	2227	37
Nissen	2007	49653/093 (1)	105	0	95	1
		100684 (3)	43	0	47	0
		49653/143 (4)	121	1	124	0
		49653/211 (5)	108	5	109	2
		49653/284 (6)	382	1	384	0
		712753/008 (7)	284	1	135	0
		AVM100264 (8)	294	0	302	1
		49653/125 (11)	175	0	173	1
		49653/127 (12)	56	1	58	0
		49653/135 (13)	116	2	111	3
		49653/147 (15)	89	1	88	0
		49653/162 (16)	168	1	172	0
		49653/234 (17)	59	0	61	0
		49653/137 (18)	204	1	185	2
		SB-712753/002 (19)	288	1	280	0
		SB-712753/003 (20)	254	1	272	0
		49653/132 (21)	221	1	112	0
Home et al	2009	RCT	2220	74	2227	67
Stroke [%]						
Nissen	2007	49653/094 (2)	113	0	110	1
		49653/135 (13)	116	1	111	1
		49653/147 (15)	89	0	88	1
		49653/162 (16)	168	0	172	1
Home et al	2009	RCT	2220	43	2227	63
Macular edema [%]						
Nissen	2007	49653/094 (2)	113	1	110	0
Bone Fractures [%]						
Nissen	2007	AVM100264 (8)	294	2	302	1
Home et al	2009	RCT	2220	185	2227	118
Bladder Cancer [%]						
Home et al	2009	RCT	2220	6	2227	5

8.2.5 Step 5: Statistical Analysis

The final step of meta-analysis was to adopt statistical tools to combine data extracted from various clinical trials to enable the uncertainty in the studies to be taken into account. Either patient-level data or aggregated trial data were used as the input. The two basic data inputs for a meta-analysis are:

1. An estimation of treatment effect, such as Odds Ratio, Risk Ratio or Mean Difference.
2. The Standard Error or Variance of the treatment effect for each trial.

There are two types of statistical models for a meta-analysis: the fixed effects model and the random effects model. Both provide a combined measurement of treatment effect but with different assumptions. The Fixed Effects Model assumes that all the studies included in the meta-analysis are drawn from the same population hence implying that all studies have the same features and underlying treatment effect. The Random Effects Model assumes that the studies involved in the meta-analysis have different characteristics and allow the true treatment effect to vary from study to study. This is a much wider assumption as it assumes additional variation. Since the 22 clinical trials selected for this project do have different features such as sample sizes, it is more appropriate to adopt the Random Effects Model Approach in which both within study variation and between studies variation is incorporated.

Since the statistical procedure for meta-analysis is rather complex and time consuming, statistical software is typically acquired to perform the task. For this project, Chen Chen used the software called “Comprehensive Meta-Analysis” (CMA) to perform the statistical analysis procedure. This software is designed specifically for the purpose of estimating the combined treatment effect from various clinical studies. The standardised summary data shown in table 3 and table 4 were used as the input into the CMA software package; by selecting the random model, CMA automatically calculate the estimation of combined effect as the output. The detailed steps of using CMA to generate the combined effects can be found in Appendix C. The basic concept behind meta-analysis is to assign weight to each study in a way that reflects the precision of the study.

By importing the extracted data from Table 3 and Table 4 into CMA, we obtained the results shown in Table 5. We replaced the input values for the model in Part I with the Table 5 means, creating a revised MCDA model that is based on more data.

The results presented in Table 6 are the estimation of combined treatment effects in terms of the Odds Ratio between the two alternatives for the data in the form of binary endpoints. These results are used for the binary endpoints during the simulation procedures.

$$OR = \frac{\text{Odds of event in treatment group}}{\text{Odds of event in controlled group}}$$

A logarithmic transformation of the OR provided a symmetric confidence interval.

Table 5: Estimates of the treatment effect for each criterion.

Effects	Rosi+Adjunct				Adjunct Only			
	Mean	LL	UL	S.E*	Mean	LL	UL	S.E*
Glycaemic Efficacy	0.00664	-0.00106	0.01433	0.003926	0.01308	0.00571	0.02045	0.00376
Microvascular Events	0.027	0.021	0.034	0.00332	0.035	0.028	0.044	0.004082
CHF	0.019	0.010	0.036	0.00663	0.012	0.007	0.023	0.004082
CV death	0.009	0.005	0.014	0.00230	0.009	0.005	0.015	0.002551
Non-CV death	0.021	0.014	0.032	0.00459	0.020	0.011	0.038	0.006888
MI	0.012	0.007	0.018	0.00281	0.011	0.007	0.018	0.002806
Stroke	0.017	0.011	0.026	0.00383	0.017	0.008	0.034	0.006633
Weight gain	3.968	3.668	4.269	0.15332	-0.490	-1.875	0.893	0.706122
Macular oedema	0.009	0.001	0.060	0.01505	0.005	0.000	0.068	0.017347
Bone fractures	0.027	0.002	0.256	0.06480	0.016	0.001	0.202	0.051276
Bladder cancer	0.003	0.001	0.006	0.00128	0.002	0.001	0.005	0.00102

LL = Lower Limit (95% Confidence Interval)
 UP = Upper Limit (95% Confidence Interval)
 * S.E = Standard Error = (UL -LL)/1.96/2

Table 6: Estimates of odds ratios for effects with binary data

Effects	O.R	LL	UL	ln(O.R)	ln(LL)	ln(UL)	ln(S.E)
Microvascular Events	0.752	0.534	1.060	-0.285	-0.627	0.058	0.175
CHF	1.968	1.445	2.680	0.677	0.368	0.986	0.158
CV death	0.869	0.661	1.142	-0.140	-0.414	0.133	0.139
Non-CV death	0.853	0.660	1.103	-0.159	-0.416	0.098	0.131
MI	1.156	0.900	1.485	0.145	-0.105	0.395	0.128
Stroke	0.662	0.453	0.969	-0.412	-0.792	-0.031	0.194
Macular oedema	2.947	0.119	73.115	1.081	-2.129	4.292	1.638
Bone fractures	1.629	1.284	2.065	0.488	0.250	0.725	0.121
Bladder cancer	1.204	0.367	3.952	0.186	-1.002	1.374	0.606

*ln(S.E) = Natural log of Standard Error = [ln(UL)-ln(LL)]/1.96/2

8.3 Cumulative Weights

The cumulative weights in Figure 9 were calculated from the swing weights given to each of the criterion scales. The two reference points on those scales, shown in Table 1, remained unchanged for the probabilistic model. Since all the scales were linear, if any simulation generated values outside the range, extrapolation by the computer appropriately generated the corresponding preference values.

8.4 Simulations

Once the estimation of treatment effects were made available by meta-analysis and the MCDA model validated, we incorporated probability distributions for all effects in the MCDA model. Because no patient-level data were available, the probabilistic distribution can only be recovered backwards based on the available statistical summary data mainly produced from the meta-analysis. The first step was to export the deterministic MCDA model into MS Excel, provide functions for translating the input data into preference values, then use @RISK to define the distributions and perform simulations on the preference values. Chen Chen created linear equations from the upper and lower limits of the scales to define functions for Excel that converted input data into preference values; these are shown in Table 7.

A major task of this project was to determine the type of distribution for each effect and to transform the data in the appropriate form for simulations. Since there are no patient-level data available and no indications of the distribution forms, we worked backwards and assumed the distribution shape based on the statistical summary data.

Table 7: Linear equations for converting input data into preference values.

	Weight	Cumulative Weight w_i	Fixed Lower	Fixed Upper	Value Function $v_i(c_i)^*$	
					a^{**}	b^{**}
Glycaemic efficacy	5	0.84%	-0.05	0.05	1000	50
Microvascular events	50	8.40%	0.00	0.20	-500	100
CHF	80	13.45%	0.00	0.04	-2500	100
CV death	100	16.81%	0.00	0.04	-2500	100
Non-CV death	100	16.81%	0.00	0.04	-2500	100
MI	35	5.88%	0.00	0.05	-2000	100
Stroke	30	5.04%	0.00	0.05	-2000	100
Weight gain	35	5.88%	-5.00	10.00	-6.67	66.67
Macular oedema	30	5.04%	0.00	0.02	-5000	100
Bone fractures	80	13.45%	0.00	0.09	-1111.11	100
Bladder cancer	50	8.40%	0.00	0.01	-10000	100
	595	100.00%				

* The value function can be written as: $v_i(c_i) = a + b c_i$
 **Coefficient a and b were determined from the linear relationship between fixed lower and fixed upper, with value score ranging between 0 and 100.

Initially, we considered the normal distribution as a reasonable fit for the data, working backwards from the mean and standard error of the treatment effects as shown in Table 5. Given the large sample sizes, that assumption seemed reasonable for the one effect measured on a continuous scale, weight gain. But for all the other criteria, the measures were percentages, whose values are restricted to the range from 0 to 100, inclusive. For those, we chose a beta distribution, one form of which extends over the range for proportions: 0 to 1.0. The beta distribution is characterised by just two parameters, which for our purposes are simply the numbers of patients who experienced the event and the number who didn't. Thus, knowing the sample size we could easily translate the observed percentages into a distribution that represented our uncertainty about the true number of patients who would experience the event⁴.

⁴ See Section 3.13.1, pp. 80-82 if Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, UK: John Wiley & Sons.

9. RESULTS

9.1 MCDA Results, revised deterministic model

The revised deterministic MCDA model, using the statistical summaries from the meta-analysis, gave the results shown in Figure 20. The overall weighted preference value for rosiglitazone plus adjunct is 64 compared to adjunct only at 72, indicating that the overall benefit-risk balance is better for adjunct therapy only. Figure 20 shows a breakdown of those overall preference values into the weighted contribution of favourable and unfavourable effects. Once again, rosiglitazone plus adjunct provides a little more benefit than adjunct only, with favourable effect scores 7.7 and 7.5, respectively. However the superiority on the benefit is overbalanced by the poor safety of rosiglitazone plus adjunct on unfavourable effects, with unfavourable effects scoring 56.3 and 64.4, respectively. Recall that longer green bars represent more benefit and longer red bar represent more safety. Therefore adding rosiglitazone to adjunct therapy appears to provide a tiny amount of benefit but with a significant increase in risks, just as with the original deterministic model in Part I.

The stacked bar graph in Figure 21 shows the contribution of each criterion to the benefit-risk balance.

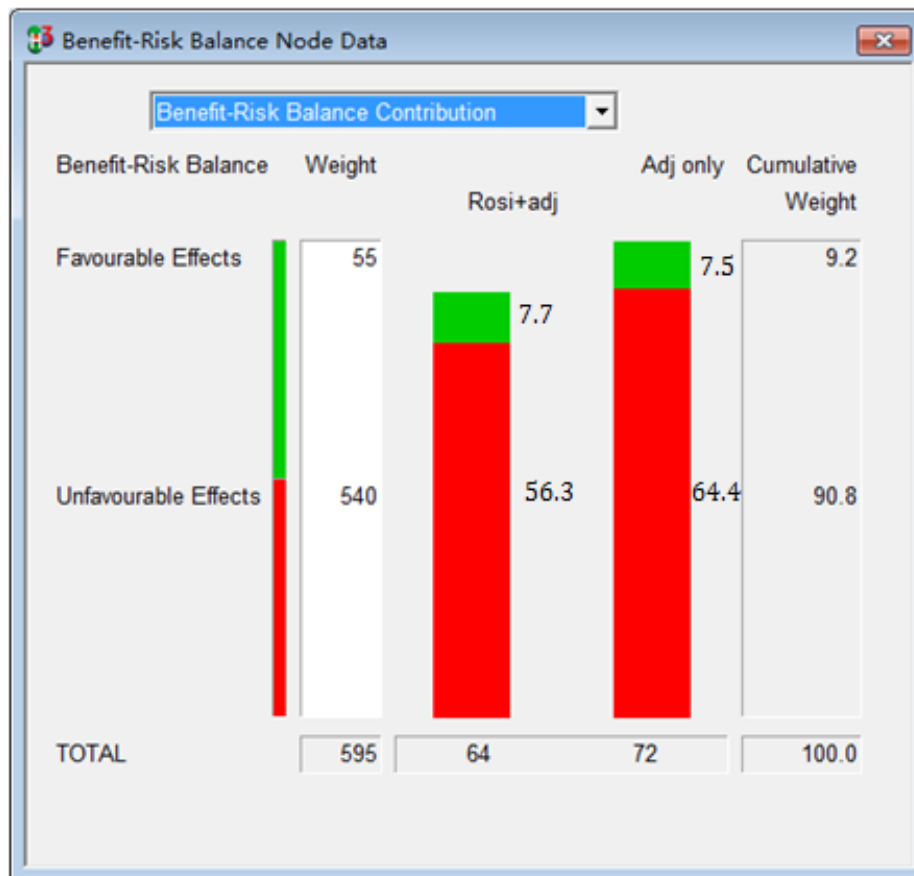


Figure 20: The overall weighted preference values for rosiglitazone plus adjunct compared to adjunct only, for the probabilistic model. Note the difference in overall weighted values here compared to Figure 10, the result for the deterministic model. The figures here are larger mainly because the probabilistic model included more data. However, the original eight-point difference is also obtained here.

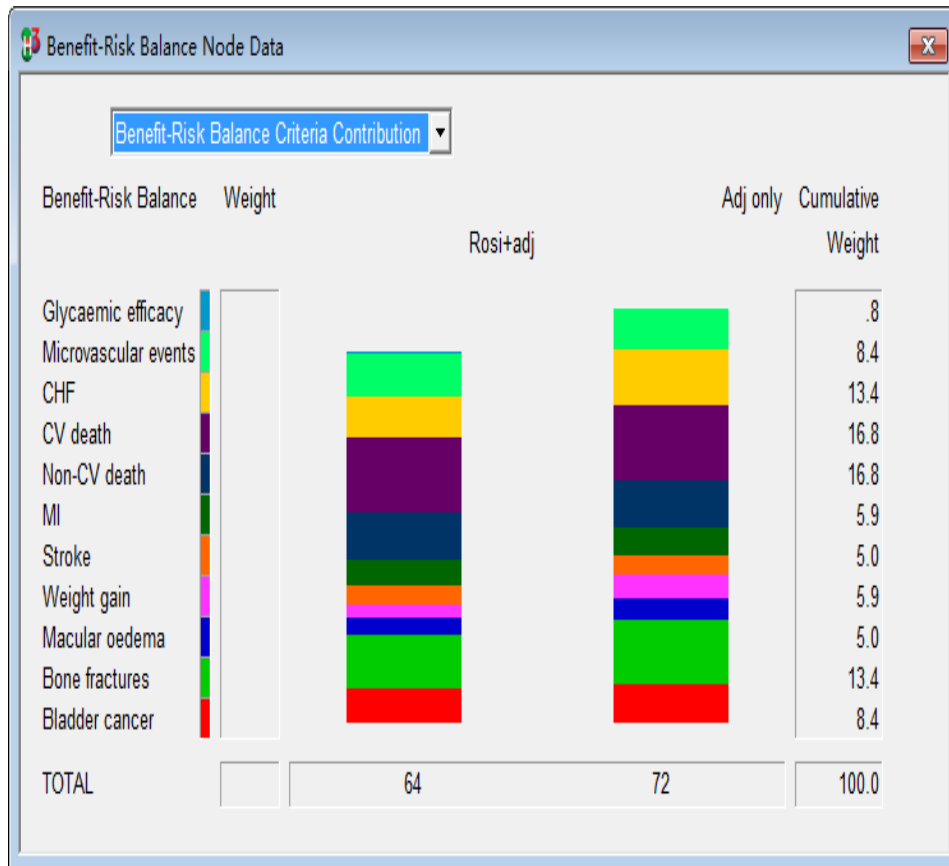


Figure 21: The weighted contribution of each effect to the overall weighted preference values.

9.1.1 Difference display

A display of the differences between the heights of the same-coloured slices in Figure 21 can be seen in Figure 22. The coloured bars in the right panel show the weighted differences for each effect, ordered from the largest difference favouring rosiglitazone plus adjunct down to the largest difference favouring the adjunct only. It is now clear that the only beneficial effect for the drug plus adjunct is Microvascular events. In this analysis, based on more data than in the Part I model, the primary endpoint, Glycaemic efficacy, very slightly favours adjunct only.

The two treatment alternatives score the same for CV death and Stroke and all the remaining effects favour adjunct only with red bars extending to the left. The sum of the weight difference adds up to -7.9, equivalent to the 8-point difference in the overall weighted preference value of the two alternatives. The small Microvascular-events advantage is easily overbalanced by the poorer safety of rosiglitazone on other unfavourable effects.

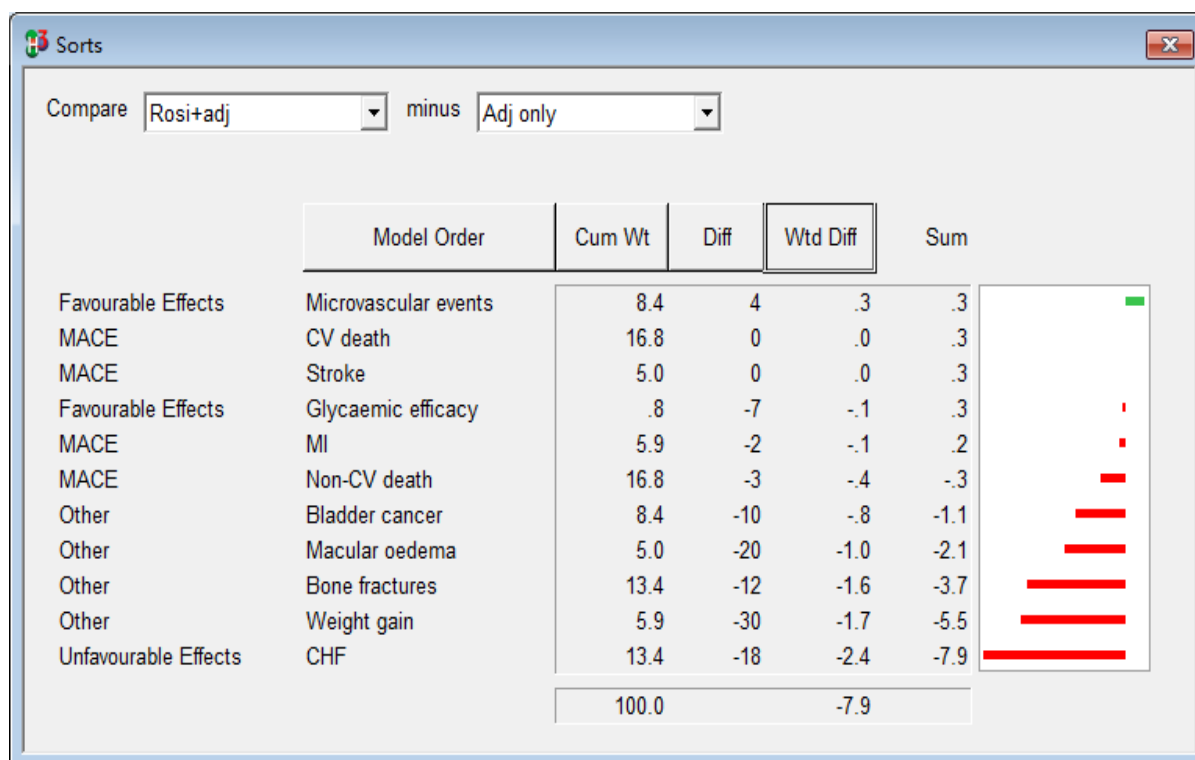


Figure 22: The difference in weighted preference scores between the two alternatives.

Comparing this result with the model in Part I (Figure 12) shows that the additional data have retained the 0.3 weighted difference on Microvascular events for rosiglitazone plus adjunct, but has halved the unfavourable effect differences for Bone fractures and CHF. The net effect of a reduced benefit favouring the drug along with the reduced side effects is to retain the 8-point difference favouring adjunct only.

9.1.2 Sensitivity analyses

The sensitivity down analysis shows how the most preferred option, adjunct only, would change as the cumulative weight on a specific criterion is decreased or increased. The green bars in Figure 23 indicate that the cumulative weight will need to change more than 15 points in order to alter the overall preference. No bars means that changing the weight over the entire 0-100 range will result in no change of the preferred option.

Separate sensitivity analyses for these two effects are shown in Figure 24. The left graph shows that the current weight on Microvascular events, 8.4%, would have to be increased to 69% to make rosiglitazone plus adjunct the more preferred option. The right graph indicates that the current weight on Stroke, 5%, would have to be increased to 100% for the same shift to occur. Clearly, the model is quite insensitive to any sensible changes in weights. The extra data have added robustness to the model.

At this stage, the interested reader may wish to compare these results with the forest plots in Appendix C to explore whether just displaying data (as illustrated by Glycaemic efficacy and CV death in Appendix C) separately for the studies considered for those effects, would allow an assessor to carry out an intuitive synthesis of the data.

The reader is also invited, before reading on, to consider all the analyses conducted up to this point, and assess the probability that the benefit-risk balance of rosiglitazone plus adjunct is better than that of the adjunct only (provided that you have forgotten the figure given in the Executive Summary).

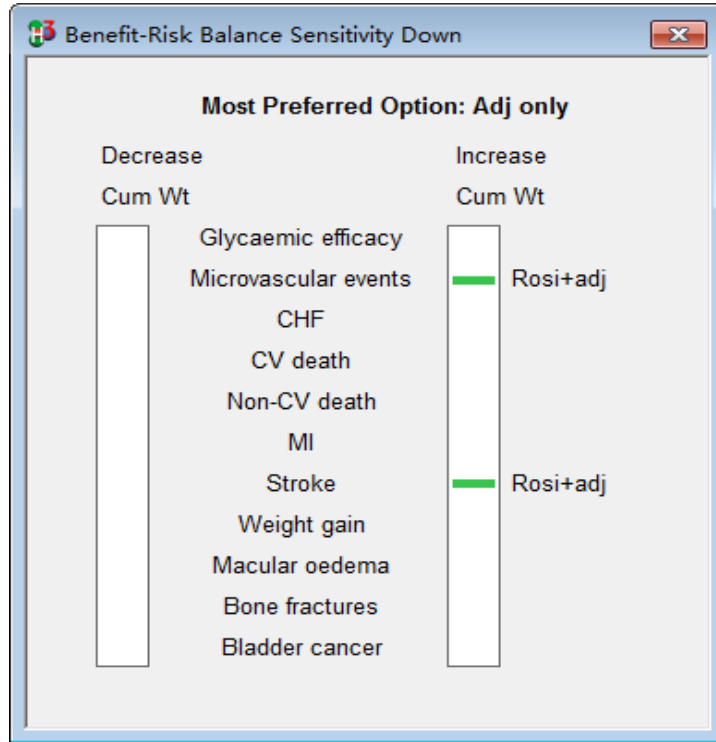


Figure 23: Sensitivity analysis on each of the effects. Only by increasing the cumulative weight on either Microvascular events or on Stroke by more than 15 points would the model favour rosiglitazone plus adjunct.

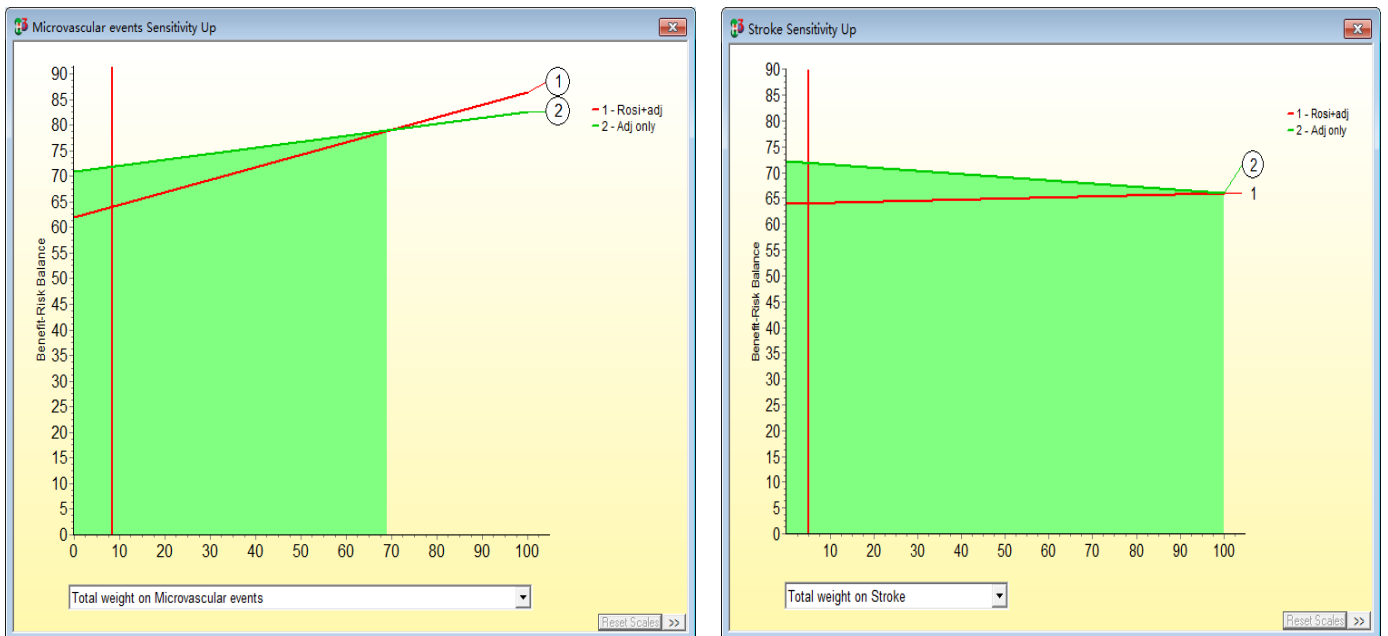


Figure 24: Sensitivity analyses on Microvascular effects (left) and on Stroke (right).

9.2 SIMULATION RESULTS

9.2.1 Probability Distribution Results

The complete probability distribution model was simulated for 100,000 iterations in @RISK. The result is shown in Figure 25. Rosiglitazone plus adjunct has a mean overall value of 53, a minimum of 12, a maximum of 66 and a 95% confidence interval of [44.9, 59.7]. Adjunct only has a mean overall value of 61, a minimum of 50, a maximum of 67 and a 95% confidence interval of [57.5, 64.3]. The probability distribution for rosiglitazone plus adjunct has a higher standard deviation of 3.79 compared to the standard deviation of adjunct only at 1.72.

As shown in Figure 25, the red probability distribution, for rosiglitazone plus adjunct, has a wider and shorter bell shape compared to the shape for adjunct only. The probability distribution for rosiglitazone plus adjunct is also skewed to the left with a significant low minimum overall value of 12. In contrast, the probability distribution of adjunct only has a narrower and taller bell shape and is more symmetrical. There is also some overlap in the middle between the two probability distributions. At this stage of the analysis, confidently claiming superiority for rosiglitazone plus adjunct may not be fully justified.

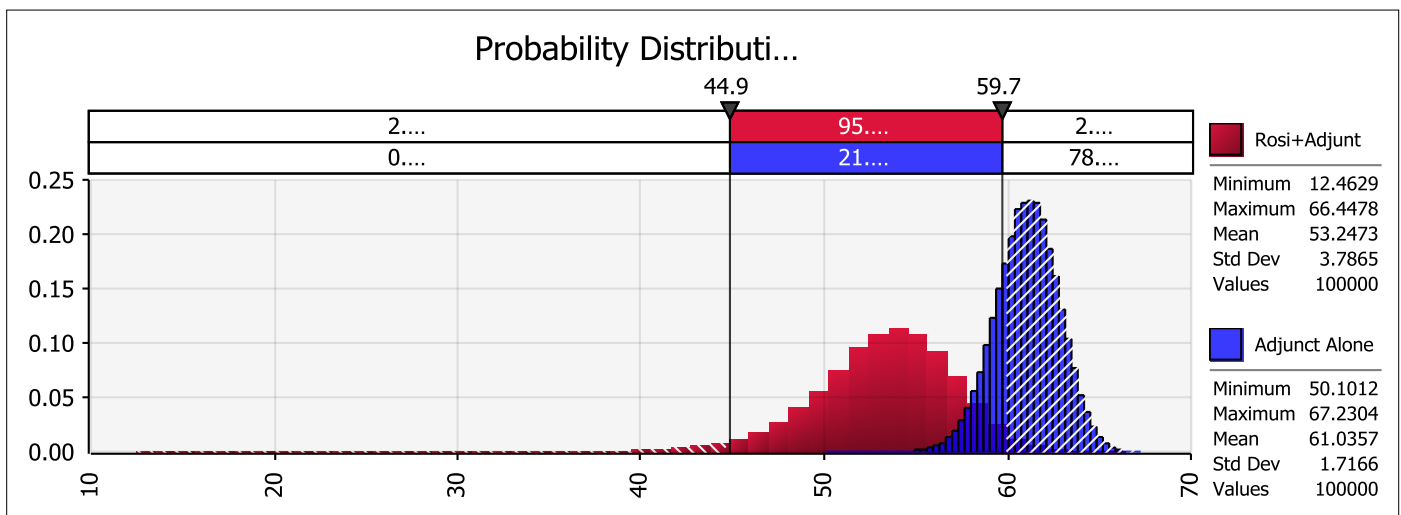


Figure 25: Cumulative probability for the two treatment alternatives

Since the two probability distributions overlap in the middle, it is possible for rosiglitazone plus adjunct to score a higher overall value than adjunct only. This raises the important question of how likely this is.

To answer this question, 100,000 iterations were also simulated for the difference between the overall score of rosiglitazone plus adjunct compared to adjunct only. The summary statistics presented in Figure 26 shows that the mean overall difference between the two treatment alternatives is around -7.8 with a 95% confidence interval extending from -14.85 to -2.33. The shape of the probability distribution for the difference is also skewed to the left with a minimum of -44.7 and a maximum of 3.3.

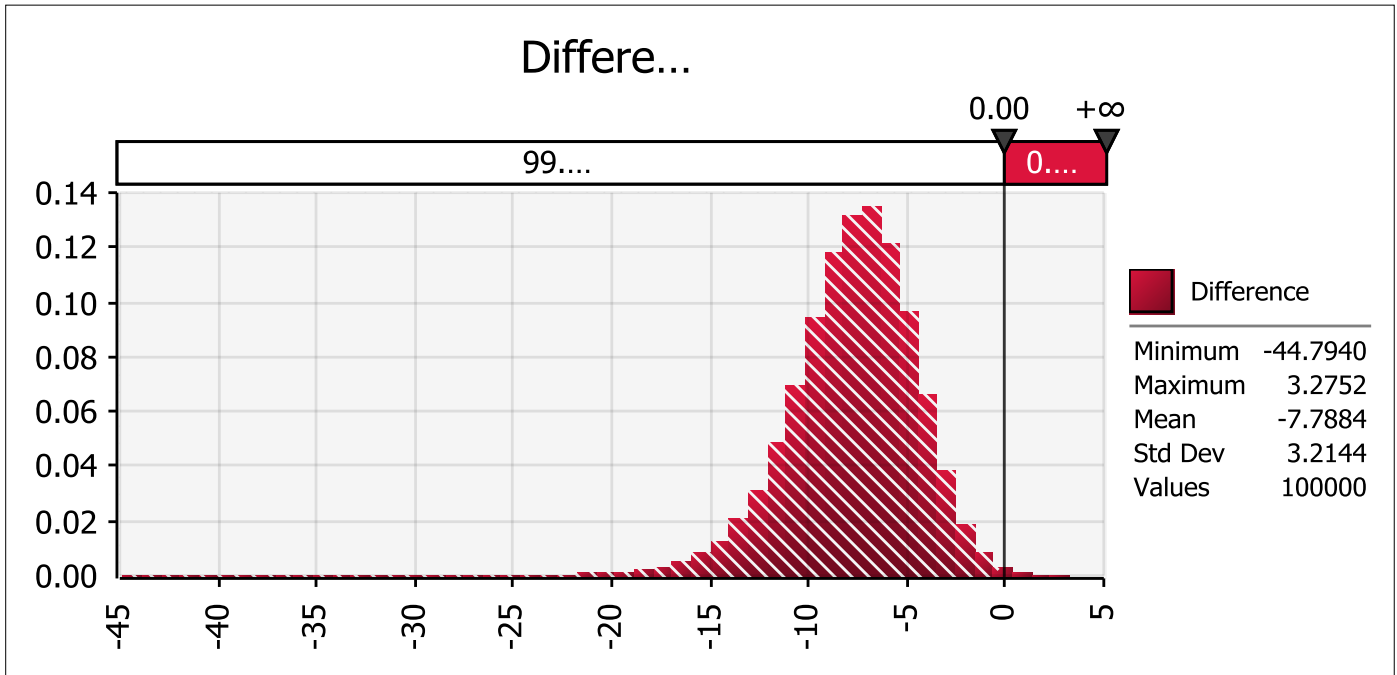


Figure 26: Probability distribution for the value difference between rosiglitazone plus adjunct and adjunct only

The important result from this simulation is that our uncertainty about the possible superiority of rosiglitazone plus adjunct over adjunct alone can be represented by a single probability: the area of the probability density function shown in Figure 26 that is to the right of zero: it is 0.2%. In other words, based on the simulation runs and given the validity of assumptions regarding the additive value model, there is only a 0.2% chance that the benefit-risk balance of rosiglitazone plus adjunct is better than adjunct only. For all 100,000 runs of the simulation, rosiglitazone plus adjunct showed a better overall benefit-risk than adjunct only just 200 times. The reason for this extreme result is to be found in the very low probability that a simulation run will choose a value in the right tail of the rosiglitazone plus adjunct distribution and in the same run a value in the left tail of adjunct only because both are very low probability possibilities. Multiplying two small probability events yields a very low probability event.

9.2.2 Sensitivity Analyses

Sensitivity analysis tested the effect of changes in the cumulative weights for all criteria on the value difference between the two treatment alternatives. The result is shown in the 'tornado diagram' of Figure 27. The vertical green bar represents the initial mean difference of -7.8 and the vertical red bar represents the position where the mean difference is zero. The length of the horizontal blue bars represents the resulting changes in mean difference when the cumulative weight for each criterion was varied from 0 to 100 by the computer. The node closer to the green bar shows the mean difference when the criterion was assigned a cumulative weight of 0% and the node extending away from the green bar shows the mean difference when the criteria was assigned a cumulative of 100%.

The mean difference between the two alternative treatments is most sensitive to the change in weight of criteria CHF, bone fractures, weight gain and stroke. However, stroke appears now to be the only criterion for which a cumulative weight increase can cause the difference between the two alternative treatments to turn positive, with rosiglitazone plus adjunct becoming the preferred option.

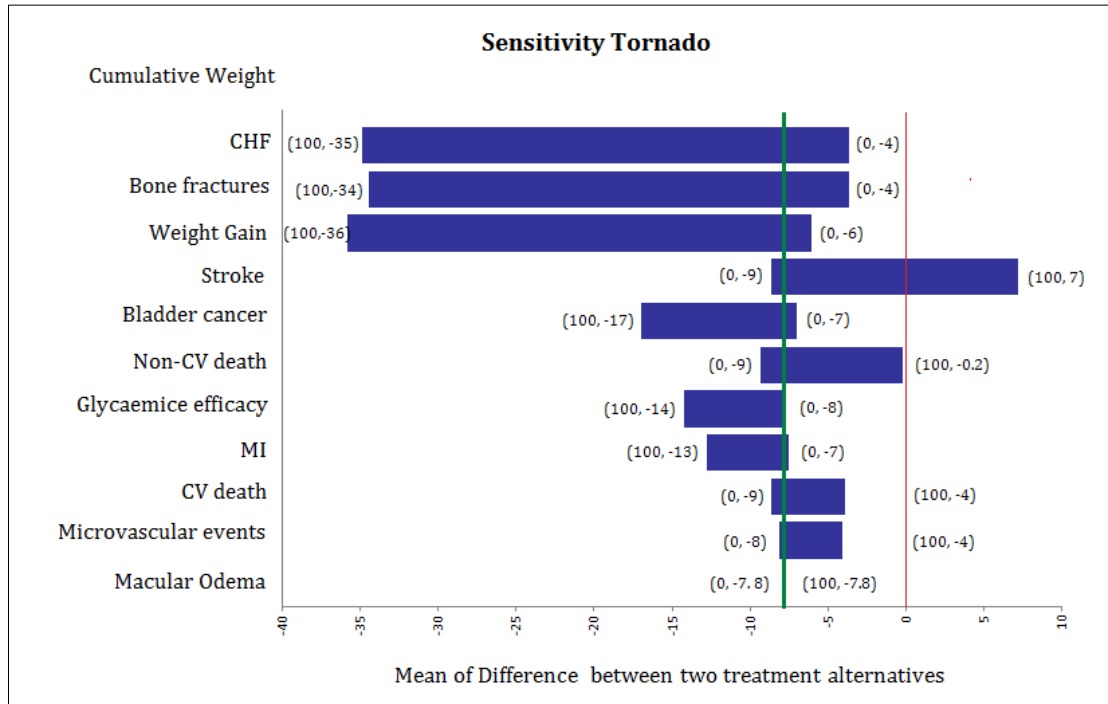


Figure 27: Sensitivity tornado diagram

We conducted two additional iterations of 100,000 runs each to explore the overall results with stroke assigned to a cumulative weight of 55% or 100%. The left probability distribution for the weighted value difference shown in Figure 28 shows that when stroke was assigned a cumulative weight of 55%, there is a 53% chance that rosiglitazone plus adjunct is better than adjunct only. In an extreme situation where Stroke was assigned a cumulative weight of 100%, there is an 84 % chance that the benefit-risk balance of rosiglitazone plus adjunct will obtain a higher overall value.

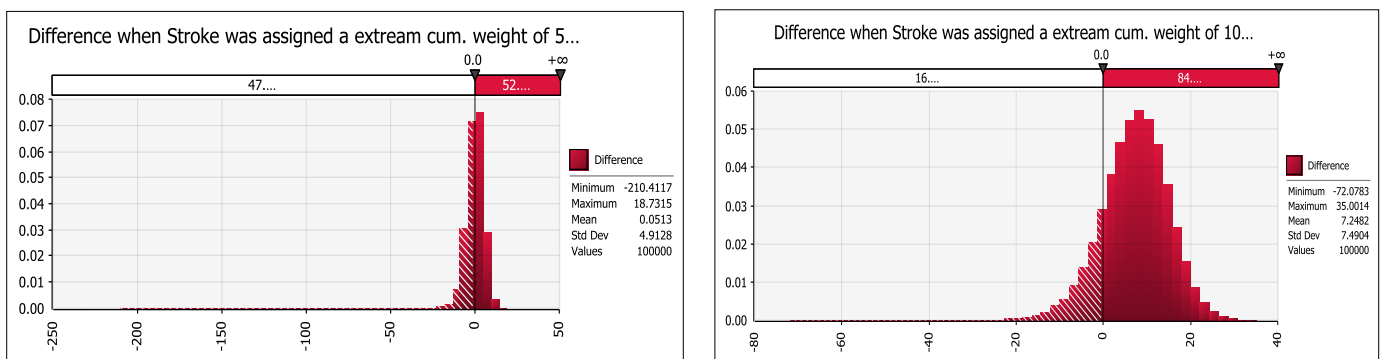


Figure 28: Probability distributions for the weighted value difference when when the cumulative weight of stroke is at 55% or 100%.

The simulation model appears to be insensitive to the change in cumulative weights. The only way for rosiglitazone plus adjunct to achieve better than a 50% of scoring a higher overall value than adjunct only is for the cumulative weight of stroke to increase above 55%, which is a substantial amount of increase from its original cumulative weight of 5%.

10. DISCUSSION AND CONCLUSIONS

The primary objective of the Part II study was to test a quantitative approach that incorporates the probability distributions of the various effects on the benefit-risk balance as compared to a deterministic model whose inputs are statistical summaries of all effects. The probabilistic model was based on 22 clinical trial studies, so how the data were extracted from these clinical trials is critical for the internal validity assessment of the MCDA and simulation models. Therefore it is important to carry out the meta-analysis in a systematic way to combine clinical data from multiple studies in order to achieve more accurate summary statistics of the treatment effects.

The deterministic MCDA model based on these 22 studies provided an overall weighted preference value of 64 for rosiglitazone plus adjunct compared to 72 for adjunct only, indicating that the benefit-risk balance is better for adjunct only. The contribution breakdown for the overall value showed that adding rosiglitazone to adjunct therapy appears to provide a negligible amount of benefit but with a significant increase in risks. The conclusion that adding rosiglitazone to adjunct therapy reduces the overall benefit-risk balance was maintained under the sensitivity analyses. The MCDA model was insensitive to the changes in cumulative weight for all criteria, and it would require a substantial amount of increase in the weight of Microvascular events or Stroke to alter the preferred option.

Based on the MCDA framework, probabilistic simulation was performed to address the uncertainty in the data by replacing the point estimates with full probability distributions for all effects. The simulation aimed at estimating the probability distribution of the value difference between the two treatment alternatives. The resulting probability distribution for the value difference has a mean of -7.8 which suggests that adding rosiglitazone to the adjunct therapy lowers the overall value of adjunct therapy only by 7.8 points. The probability distribution also indicates that based on the 100,000 iterations simulated, there is only 0.2% probability that rosiglitazone plus adjunct is higher in overall value than adjunct only. Sensitivity analysis showed that the probabilistic distribution of the value difference is also very insensitive to the changes in the cumulative weight of all effects. The estimated probability of 0.2% appears to be rather definitive and hence under the current uncertainty, confidently claiming superiority for adding rosiglitazone to the existing adjunct therapy is not fully justified. This high confidence in adjunct alone is the result of two factors: more data and skew in the raw data.

It is worth noting that the means of the overall weighted benefit-risk preference value for the two alternative treatments from the simulations (53, 61) are different from those of the deterministic MCDA model (64,72). A large assumption was made to assume that all studies used have the same feature and characteristics, so the studies were given equal weights in generating beta distributions. Therefore the estimates for means are different because different rates and confidence intervals were obtained. The means are different purely due to the different assumptions and approaches across the clinical studies that were adopted to obtain the most accurate estimations for both models. Interestingly, the value differences for the two treatment alternatives are very similar, both around -8 for the MCDA model and the simulations.

The important insight of the findings is that by working with point estimates of the data instead of with the probability distributions could restrict the medical regulators from seeing the full picture. Despite the fact of being an adequate benefit-risk balancing technique, using a deterministic MCDA alone is not sufficient to tell the whole story. That may not matter for simpler cases, but for complex ones, like rosiglitazone, a fuller consideration of all the clinical data could lead to a different conclusion.

Since rosiglitazone is still available in the United States for use with adjunct therapy, regulators there must have judged there to be some positive benefit-risk difference. However, with the new information provided by the probabilistic distribution of value differences, using the meta-analysis data, the medical regulators could now see that there is only a 0.2% chance that the combination might actually do better. Therefore, this project shows that it can be risky for the medical regulators to rely heavily on point estimates only when it comes to addressing the benefit-risk balance of drugs.

It is important to keep in mind that this project is based on a single drug and one comparator only, the structure of the MCDA model, and the elicited weights of the medical experts on the PROTECT WP5 team. Although it could take many modifications and validations for this model to be officially applicable by the medical regulator, the ability to obtain a solid percentage value that indicates the probability of one drug performing better than the comparator is a new contribution to deliberative discourse about the benefit-risk balance.

Pharmaceutical companies could also explore the MCDA-plus-simulation approach during the process of drug development. This way, the company can stop the development and production process as soon as evidence becomes available indicating that the new product comes with a low probability of performing better than existing products. This could potentially save any pharmaceutical company hundreds of millions of pounds. For both pharmaceutical companies and regulators, deterministic modelling coupled with simulation can provide an answer to the question they both ultimately have to answer: What is the probability that the benefit-risk balance of this new drug is better than the comparator? That question is simply not answered with significance levels and confidence intervals associated with the data.

11. STRENGTHS AND LIMITATIONS OF THE APPROACHES USED

11.1 META-ANALYSIS

Meta-analysis provides a more sophisticated and systematic way of combining medical data while allowing the uncertainty to be taken into consideration. It increases the precision of measuring the true treatment effect and can accommodate, with suitable weighting, studies of varying quality. In this way, no data are lost, and possibly less data need to be obtained.

For this project, due to the existing amount of data availability, Microvascular events, Macular oedema and Bladder cancer data are based on only one available trial after the data selection and quality assessment procedure, which limit the validity of the results. We originally planned that patient-level data for all the effects would be collected and used as inputs to the probabilistic simulations. However, we found that there is no patient level data that is available to the general public. All available studies on the performances of rosiglitazone are in the form of statistical summaries. This lack of raw data forced us to reconstruct the probability distributions based on the existing statistical summaries. As it stands, the probabilistic model of rosiglitazone demonstrates satisfactorily what could be achieved if raw data were available.

11.2 MCDA MODEL

MCDA is regarded as a logical, coherent and flexible model for decisions with multiple objectives. It decomposes a complex decision problem into smaller and easier parts. The additive value function is then used to link the separate parts together and produce an aggregate numerical value for each option.

The MCDA model is comprehensive in the way that it is able to accommodate all types of data, uncertainties and risk attitudes. It takes into account individuals' preferences and risk attitude by providing a way of transforming input values into utilities. The MCDA model also enables both the decision maker and analyst to gain a deeper understanding of the decision context through the model-building process. The results produced by the MCDA model are visual and easy to read so that the rationale can be explained, understood and used as evidence for future justifications.

One criticism of the approach is that it is limited to cases in which assessors are knowledgeable about the criterion weights, so can readily make these explicit. Usually this is not the case, so critics have argued that the assessment of swing weights can give spurious results. Indeed, the SMAA software provides a function for developing probability distributions for weights.

This criticism seems to be motivated by worry about the validity of MCDA results, which presumes that the purpose of an MCDA is to provide 'the right answer'. However, the approach taken for this case study assumes that the purpose of MCDA is to provide a tool for a group of assessors to explore different data sets, assumptions and value judgements so that the group can gain a shared understanding of the features that affect the benefit-risk balance and develop a sense of common purpose about the decision or recommendation they wish to make. All members of the group are encouraged to share their perspectives and experiences, to test assumptions and differences of opinion with the model, and allow preferences to be constructed during the modelling process. Initially labile preferences gradually become more stable, enabling the group eventually to agree about decisions or recommendations without necessarily gaining consensus about all aspects of the model. In short, MCDA provides structure to thinking while participants shape and re-shape their preferences.

It was difficult for our group to elicit value functions for medical effects and to assess swing weights. This is a common experience for assessors when they first are exposed to MCDA because these are new ways of thinking about the clinical value of evidence. And it was particularly difficult for our group because most of us had no clinical experience with type 2 diabetes. It is abundantly clear that MCDA requires a degree of expert input, that the group should be composed of people representing all the relevant expertise about the drug and disease state it treats. Our

results represent the view of the few medical experts among us, who tried to look at the problem from a regulatory point of view. It does not necessarily represent the preference of any regulator or the general public.

Perhaps the most serious limitation of MCDA is its unfamiliarity to the medical profession. Indeed, most of the concepts, effects tree, effects table, scoring and weighting, are new, and the graphical presentations of results, stacked bar graph, difference display or sensitivity analysis plot, are rare in the medical research literature. As the narrator of Abraham Verghese's novel, *Cutting for Stone*, discovers as a young man in his studies of medicine:

"I found that the bricks and mortar of medicine (unlike, say, engineering) were words. You needed only words strung together to describe a structure, to explain how it worked, and to explain what went wrong."

A culture of words sees MCDA as alien, so although this quantitative model can deepen understanding and facilitate communication, perhaps the added benefit is seen as not sufficient to justify the cost of learning a new language.

11.3 PROBABILISTIC SIMULATION

Probabilistic simulation is able to accommodate various types of distribution and can be implemented using any existing software for probabilistic simulation, such as @Risk, in Excel. The probability distribution outputs were presented in a readily and easy interpretable manner.

This approach provides full probability distributions of the benefit-risk balance of each option as well as benefit-risk differences between options. Statistical summaries and confidence intervals can be misleading when the actual shape of the population distribution is skewed⁵. Although the simulation procedure can be complex and time consuming, its ability to provide a single percentage value that indicates the probability of one drug performing better than the comparator could prove to be useful not only to the drug regulator but also to pharmaceutical companies.

As no patient-level data were available for our project, the probability distributions were assumed to be normal or beta distributed, and we made the best use possible of the statistical summaries to infer the population distributions. We assumed that all the clinical studies are subject to the same features and characteristics so that the beta distributed rate can be defined by adding up all the events and non-events. However, this assumption might have led to biases that distort the actual distribution shape in reality.

⁵ A point elaborated in Savage, S. L. (2009). *The Flaw of Averages*. Hoboken, NJ: John Wiley & Sons.

12. RECOMMENDATIONS AND FURTHER WORK

Many of the limitations of this study originate from data either being unavailable or being presented in a format that is different and difficult to incorporate into a decision model. This section aims to make recommendations as to what type of data and what additional work may be useful for further developing the model. It also foresees a future in which quantitative modelling of benefit-risk transforms development activities in pharmaceutical industries approaches to drug approvals by regulators, and scientific research about medicinal products.

12.1 PATIENT LEVEL DATA

As we were unable to resource any patient level data for the purpose of this project, we ended up estimating the probability distributions from the evidence based on summary statistics. Making the data from scientific investigations available at least to other responsible researchers is considered as best practice within the scientific community.

Recommendation 1: That the European Commission investigate this issue of data availability and take steps to ensure that patient-level data about clinical studies of medicinal products are properly archived and made accessible.

12.2 DATA FORMAT AND QUALITY

During the meta-analysis process, the clinical trial data were presented in different formats and it can be time consuming to pick out the useful data from the full clinical trial report. Sometimes, treatment effects are presented in different types of units and form such as measures of central tendency and variability, standard errors, confidence intervals, significance levels and meta-analyses results in different reports.

Recommendation 2: That the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) or some other statutory body develop and recommend a standardised format for all clinical trial reports so that data can be extracted quickly and in a unified form.

It would also be useful if a cohesive set of treatment effects could be established for a given disease state. For example, now that we have examined drugs for type 2 diabetes, we have provided a set of favourable and unfavourable effects that might be of use in assessing a new treatment for this condition. Each time a type 2 diabetic treatment is modelled using MCDA, the previous model structure could be updated. Eventually, a generic model would emerge, and could then be applied quickly and confidently for assessing new drugs.

It would also be possible, then, for regulators to take the generic model of all effects, their measurement scales and swing weights for the scales, but with no data, and give this as a template to pharmaceutical companies. The company would then know more about the relative importance of the criteria to be considered by the regulator, which should assist the company in designing a new drug that would present an attractive profile.

Recommendation 3: That the ICH or other body, such as CIOMS or university research centres, take on the task of developing templates of criteria, measurement scales and swing weights that can be used in modelling drugs indicated for specific disease states.

12.3 ELICITATION OF VALUE FUNCTIONS AND WEIGHTS

The value functions and weights elicited for the MCDA model were based on the individual preferences of only a few medical experts in the PROTECT WP5 team. In future, it would be constructive to involve more medical experts in the weighting process to generate a more representative set of weights and thus improve the model.

Recommendation 4: That MCDA modelling of the benefit-risk of medicinal products be conducted in a facilitated workshop, like a decision conference, composed of assessors whose experience and expertise cover all the

perspectives needed to populate the model. Face-to-face interaction is essential so that participants can share their views and experiences, and work towards constructing the preferences that are necessary for a valid model.

Furthermore, since the medical experts assigned the weights from a regulatory point of view, it may be worth investigating the preferences from a patient point of view by engaging patients to elicit the weights.

Recommendation 5: That the IMI sponsor research on obtaining meaningful swing weights from patients and prospective patients for the criteria in MCDA models that have been created in the current PROTECT project. These weights could then serve in the templates being created in Recommendation 4, and they should be of interest to Health Technology Assessment organisations.

12.4 SOFTWARE PACKAGES

Effective modelling of the benefit-risk balance of drugs requires the help of specialised software; Excel by itself is too limiting. We used Hiview3 in this project, whose deterministic results were exported to Excel with @Risk as an add-in to carry out the probabilistic simulation. It would be more convenient if both MCDA and simulation could be part of just one software package, as has been attempted in the open-source software product, SMAA.

Recommendation 6: Explore the possibility of developing a stand-alone software product that is specifically designed for modelling the benefit-risk of drugs and supports probabilistic simulation. It should incorporate features and displays in other software products that were found to be useful throughout WP5.

13. REFERENCES

- Bakhai A, Chhabra A, Wang D. (2006). Meta- Analysis. In: D Wang, F Clemens & A Bakhai, (Ed.s). *Clinical Trials: A practical guide to design, analysis and reporting*. London: Remedica. 439-461
- Deeks JJ, Higgins JPT, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1 [updated September 2008]. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.
- EMA CHMP. (2008). *Reflection paper on benefit-risk assessment methods in the context of the evaluation of marketing authorisation applications of medical products for human use in London*. European Medicines Agency.
- Felli, J., Noel, R., & Cavazzoni, P. (2009). A Multi-attribute Model for Evaluating the Benefit-Risk Profiles of Treatment Alternatives. *Medical Decision Making*, vol. 29 , 104-115
- GlaxoSmithKline (GSK) (2010) *Information for healthcare professionals-Avandia*: [online] available from <<http://en.wikipedia.org/wiki/Rosiglitazone>> [15 August 2012]
- Guo, J., Pandey, S., Doyce, J., Bian, B., & Raish, Y. L. (2010). *A Review of Quantitative Risk-Benefit Methodologies for Assessing Drug Safety and Efficacy – Report of the ISPOR Risk-Benefit Management Working Group*. International Society for Pharmacoeconomics and Outcomes Research (ISPOR).
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: John Wiley, republished in 1993 by Cambridge University Press.
- Lynd, L., & O'Brien. B (2004). Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. *Journal of Clinical Epidemiology*, 795-803.
- Nissen, S. E., & Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, 356(24), 2457-2471.
- Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (2012) *About Protect*: [online] available from < <http://www.imi-protect.eu/about.shtml>>[21 August 2012]
- Phillips, L. D. (2007). Decision Conferencing. In W. Edwards, R. F. Miles & D. von Winterfeldt (Eds.), *Advances in Decision Analysis: From Foundations to Applications*. Cambridge: Cambridge University Press.
- Raiffa, H. (1968). *Decision Analysis*. Reading, MA: Addison-Wesley.
- Phillips, L. D., Fasolo, B. & Zafiroopoulos, N. (2010). *Work Package 2 Report: Applicability of current tools and processes for regulatory benefit-risk assessment*. European Medicines Agency (EMA).
- Phillips, L. D. (1984). A theory of requisite decision models. *Acta Psychologica*, 56, 29-48.
- Wikipedia (2012) *Rosiglitazone*: [online] available from <<http://en.wikipedia.org/wiki/Rosiglitazone>>[15 August 2012]

APPENDIX A—DECISION CONFERENCING

The approach taken to constructing a benefit-risk model is based on *decision conferencing* (Phillips, 2007). This is a socio-technical process that combines working in groups helped by an impartial facilitator, on-the-spot computer-based modelling of data and participants' judgments, and continuous visual display of the model and its results. The 'socio' aspect of the process relies on mobilizing the right people at the right time to give the right inputs to the model. The 'technical' part refers to the model itself. This is based on decision analysis, first introduced in 1968 by Howard Raiffa (1968), and extended by Ralph Keeney and Howard Raiffa (Keeney & Raiffa, 1976) to cover decisions with multiple objectives, which is now an accepted methodology for dealing with decisions that are characterized by uncertainty and multiple objectives⁶.

The generic purposes of decision conferencing are to achieve a shared understanding of the issues (though not necessarily consensus), a sense of common purpose (while preserving individual differences of opinion) and a commitment to the way forward (though allowing individual differences in the paths). The idea is to encourage individual creativity, and to use differences of perspective to find ways forward that will gain support from those implementing the actions. A key assumption of decision conferencing is the notion of 'requisite modelling' (Phillips, 1984): that a model should be just sufficient in form and content to resolve the issues at hand. For benefit-risk analysis of drugs, the model need not be more complex than is sufficient to determine if the benefits outweigh the risks and to determine what additional information might be necessary. The model is a 'tool for thinking' enabling participants to see the logical consequences of differing viewpoints, and the effects of uncertainty on the benefit-risk balance.

A decision conference typically moves through four stages. The first stage is a broad exploration of the issues and context. In the second stage, a model is constructed of the favourable and unfavourable effects, incorporating available data and participants' judgements about clinical relevance of the effects. In the third stage, the model combines the effects and shows the benefit-risk balance. Extensive sensitivity analyses examine the effects on the balance of imprecision in the data, uncertainties, and differences in participants' risk tolerance. Discrepancies between model results and members' judgements are examined, causing new intuitions to emerge, new insights to be generated and new perspectives to be revealed. Revisions are made and further discrepancies explored; after several iterations the new results and changed intuitions are more in harmony. Then the group moves on to the fourth stage summarising key issues and conclusions, formulating next steps and, if desired, agreeing recommendations. The facilitator prepares a report of the event's products after the meeting and circulates it to all participants.

⁶ For additional information about benefit-risk methodologies for regulators see the WP2 report on the EMA public website. Click on the Special Topics tab, then on Benefit-Risk Methodology in the left column, and choose the pdf file "Benefit-risk methodology project work package 2 report".

APPENDIX B—THE MCDA MODEL

Each of the following matrices corresponds to a node in the value tree of Figure 1. The weights shown in the left column are the sums of the original weights at lower nodes. Next, the preference values based on the metric input data are shown, with total weighted preference values given at the bottom in the TOTAL row, e.g., $35=(82 \times 0.097)+(30 \times 0.903)$. The right column shows the cumulative weights, obtained after normalising the weights in the left column, e.g., $9.7=100 \times (55 \div 569)$, to ensure all criterion weights sum to one. Asterisks identify criteria at the extreme right of the Effects Tree.

Overall Favourable-Unfavourable Effects Balance

Benefit-Risk Balance Node Data				
Benefit-Risk Balance Data Breakdown				
Benefit-Risk Balance	Weight	Rosi+adj	Adj only	Cumulative Weight
Favourable Effects	55	82	80	9.7
Unfavourable Effects	514	30	40	90.3
TOTAL	569	35	44	100.0

Favourable Effects

Favourable Effects Node Data				
Favourable Effects Data Breakdown				
Favourable Effects	Weight	Rosi+adj	Adj only	Cumulative Weight
Glycaemic efficacy*	5	38	51	0.9
Microvascular events*	50	87	83	8.8
TOTAL	55	82	80	9.7

Unfavourable Effects

Unfavourable Effects Node Data

Unfavourable Effects Data Breakdown

Unfavourable Effects	Weight	Rosi+adj	Adj only	Cumulative Weight
CHF*	80	8	53	13.4
MACE	265	33	19	44.3
Other	195	35	62	32.6
TOTAL	540	30	40	90.3

MACE Node Data

MACE Data Breakdown

MACE	Weight	Rosi+adj	Adj only	Cumulative Weight
CV death*	100	33	20	16.7
Non-CV death*	100	26	4	16.7
MI*	35	33	40	5.9
Stroke*	30	61	43	5.0
TOTAL	265	33	19	44.3

Other Node Data

Other Data Breakdown

Other	Weight	Rosi+adj	Adj only	Cumulative Weight
Weight gain*	35	41	67	5.9
Macular oedema*	30	37	89	5.0
Bone fractures*	80	7	41	13.4
Bladder cancer*	50	73	78	8.4
TOTAL	195	35	62	32.6

APPENDIX C: COMPREHENSIVE META-ANALYSIS INPUT AND RESULTS

META-ANALYSIS FOR MEAN TREATMENT EFFECT

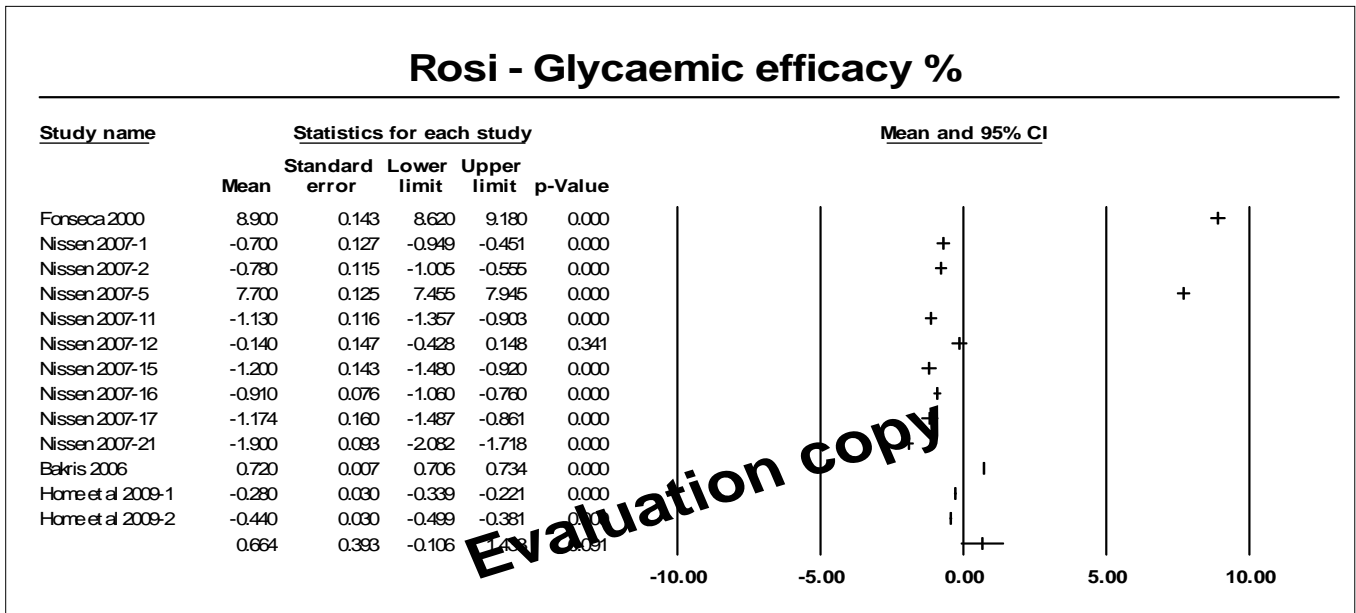
Glycaemic efficacy % for rosiglitazone plus adjunct

Comprehensive meta analysis - [Data]															
File Edit Format View Insert Identify Tools Computational options Analyses Help															
Run analyses → [Icons]															
	Study name	Mean	Standard deviation	Sample size	Mean	Std Err	G	H	I	J	K	L	M	N	O
1	Fonseca 2000	8.900	1.500	110	8.900	0.143									
2	Nissen 2007-1	-0.700	1.300	105	-0.700	0.127									
3	Nissen 2007-2	-0.780	1.219	113	-0.780	0.115									
4	Nissen 2007-5	7.700	1.300	108	7.700	0.125									
5	Nissen 2007-11	-1.130	1.530	175	-1.130	0.116									
6	Nissen 2007-12	-0.140	1.100	56	-0.140	0.147									
7	Nissen 2007-15	-1.200	1.350	89	-1.200	0.143									
8	Nissen 2007-16	-0.910	0.990	168	-0.910	0.076									
9	Nissen 2007-17	-1.174	1.227	59	-1.174	0.160									
10	Nissen 2007-21	-1.900	1.380	221	-1.900	0.093									
11	Bakris 2006	0.720	0.100	194	0.720	0.007									
12	Home et al 2009-1	-0.280	1.003	1117	-0.280	0.030									
13	Home et al 2009-2	-0.440	0.996	1103	-0.440	0.030									
14															

Result for combined effect

Comprehensive meta analysis - [Analysis]															
File Edit Format View Computational options Analyses Help															
← Data entry ↻ Next table [Icons] + Effect measure: Mean [Icons]															
Model	Study name	Statistics for each study							Mean and 95% CI					Weight (Random)	
		Mean	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	-10.00	-5.00	0.00	5.00	10.00		
	Fonseca	8.900	0.143	0.020	8.620	9.180	62.229	0.000						+	7.66
	Nissen	-0.700	0.127	0.016	-0.949	-0.451	-5.518	0.000			+				7.68
	Nissen	-0.780	0.115	0.013	-1.005	-0.555	-6.802	0.000			+				7.69
	Nissen	7.700	0.125	0.016	7.455	7.945	61.554	0.000					+		7.68
	Nissen	-1.130	0.116	0.013	-1.357	-0.903	-9.770	0.000			+				7.69
	Nissen	-0.140	0.147	0.022	-0.428	0.148	-0.952	0.341			+				7.66
	Nissen	-1.200	0.143	0.020	-1.480	-0.920	-8.386	0.000			+				7.66
	Nissen	-0.910	0.076	0.006	-1.060	-0.760	-11.914	0.000			+				7.72
	Nissen	-1.174	0.160	0.026	-1.487	-0.861	-7.352	0.000			+				7.64
	Nissen	-1.900	0.093	0.009	-2.082	-1.718	-20.468	0.000			+				7.71
	Bakris 2006	0.720	0.007	0.000	0.706	0.734	100.284	0.000							7.74
	Home et al	-0.280	0.030	0.001	-0.339	-0.221	-9.333	0.000							7.74
	Home et al	-0.440	0.030	0.001	-0.499	-0.381	-14.667	0.000							7.74
	Random	0.664	0.393	0.154	-0.106	1.433	1.690	0.091			+				

Forest plot for combined effect



META-ANALYSIS FOR ODDS RATIO

Data input window for cv death %

Comprehensive meta analysis - [C:\Users\mac\Desktop\Summer Project\Meta-Analysis\Model 2\CMA files\CV Death [%].cma]

File Edit Format View Insert Identify Tools Computational options Analyses Help

Run analyses → [Icons]

	Study name	Rosi Treated Group Events	Rosi Treated Group Total N	Adjunt Control Group Events	Adjunt Control Group Total N	Odds ratio	Log odds ratio	Std Err	I	J	K	L
1	Home 2007	29	2220	35	2227	0.829	-0.188	0.253				
2	Nissen 2007 - 3	0	43	1	47	0.356	-1.032	1.647				
3	Nissen 2007 - 4	0	121	0	124							
4	Nissen 2007 - 5	3	108	2	109	1.529	0.424	0.923				
5	Nissen 2007 - 6	0	382	0	384							
6	Nissen 2007 - 7	0	284	0	135							
7	Nissen 2007 - 8	2	294	1	302	2.062	0.724	1.227				
8	Nissen 2007 - 11	0	175	0	173							
9	Nissen 2007 - 13	1	116	1	111	0.957	-0.044	1.420				
10	Nissen 2007 - 15	0	89	0	88							
11	Nissen 2007 - 16	0	168	0	172							
12	Nissen 2007 - 17	0	59	0	61							
13	Nissen 2007 - 18	0	204	1	185	0.301	-1.202	1.636				
14	Nissen 2007 - 19	1	288	0	280	2.927	1.074	1.635				
15	Nissen 2007 - 20	0	254	0	272							
16	Nissen 2007 - 21	1	221	0	112	1.531	0.426	1.637				
17	Bakris 2006	1	194	0	180	2.798	1.029	1.636				
18	Home et al 2009	60	2220	71	2227	0.844	-0.170	0.178				
19												

Result for CV death % odds ratio

Comprehensive meta analysis - [Analysis]												
File Edit Format View Computational options Analyses Help												
← Data entry Next table High resolution plot Select by ... Effect measure: Odds ratio												
Model	Study name	Statistics for each study					Odds ratio and 95% CI					Weight (Random)
		Odds ratio	Lower limit	Upper limit	Z-Value	p-Value	0.01	0.10	1.00	10.00	100.00	Relative weight
	Home 2007	0.829	0.505	1.361	-0.742	0.458						30.42
	Nissen	0.356	0.014	8.982	-0.627	0.531						0.72
	Nissen	1.529	0.250	9.334	0.460	0.646						2.28
	Nissen	2.062	0.186	22.859	0.589	0.556						1.29
	Nissen	0.957	0.059	15.482	-0.031	0.975						0.96
	Nissen	0.301	0.012	7.428	-0.734	0.463						0.73
	Nissen	2.927	0.119	72.153	0.657	0.511						0.73
	Nissen	1.531	0.062	37.876	0.260	0.795						0.73
	Bakris 2006	2.798	0.113	69.137	0.629	0.529						0.73
	Home et al	0.844	0.595	1.196	-0.956	0.339						61.42
Random		0.869	0.661	1.142	-1.009	0.313						

Forest plot for cv death % odds ratio

