Benefit-Risk Assessment Model for Medicines: Developing a Structured Approach to Decision Making

Report of the Workshop organised by the

CMR International Institute for Regulatory Science

at the **Georgetown Inn, Washington D.C., USA**

13-14 June 2005

Stuart Walker
Larry Philips
Margaret Cone

January 2006

CMR International

Institute for Regulatory Science

CMR INTERNATIONAL INSTITUTE FOR REGULATORY SCIENCE

The CMR International Institute for Regulatory Science has been established as a not-for-profit division of the Centre for Medicines Research International Ltd in order to continue its work in the regulatory and policy arena, and to maintain the well-established links that the Centre has with regulatory authorities around the world. The Institute operates autonomously, with its own dedicated management, and funding that is provided by income from a membership scheme. The Institute for Regulatory Science has a distinct agenda dealing with regulatory affairs and their scientific basis, which is supported by an independent Advisory Board of regulatory experts.

CMR International, Institute for Regulatory Science,

Novellus Court, 61 South Street Epsom, Surrey KT18 7PX, UK

Tel: +44 (0) 1372 846100, Fax: +44 (0) 1372 846101

E-mail: institute@cmr.org

Website: www.cmr.org/institute

CMR International Institute for Regulatory Science

Workshop on

BENEFIT-RISK ASSESSMENT MODEL FOR MEDICINES –
DEVELOPING A STRUCTURED APPROACH TO DECISION MAKING

13-14June 2005, The Georgetown Inn, Washington, D.C.

CONTENTS

ANNEXES

CMR International Institute for Regulatory Science

Workshop on

BENEFIT-RISK ASSESSMENT MODEL FOR MEDICINES –
DEVELOPING A STRUCTURED APPROACH TO DECISION MAKING

13-14June 2005, The Georgetown Inn, Washington, D.C.

SUMMARY REPORT

OVERVIEW

As public accountability grows for decisions taken in both public and private sectors, so the need develops for organisations to provide an audit trail for important decisions. Decisions leading to drug approvals demand careful attention with respect to balancing the benefits and risks, which in turn requires a structured process that leaves an audit trail.

The Workshop brought together twenty-seven senior participants drawn from regulatory agencies in the USA and Europe, academia and the pharmaceutical industry (see *Annex 1*). The purpose was to explore a structured model for benefit-risk assessment and consider its potential for providing support to decision-makers in both regulatory agencies and pharmaceutical companies.

This was the second Workshop on this topic, hosted by the CMR International Institute for Regulatory Science, the first having been held in March 2004, in the UK. The success of this meeting led recommendations for a further Workshop to be convened in the USA

The Workshop was organised to encourage frank and open discussions about the way in which important, far-reaching decisions are taken within both companies and regulatory agencies and the potential shortcomings of such processes. The main component of the meeting, however, was an interactive demonstration of the application of the technical model 'multi-criteria decision analysis' (MCDA). The model was discussed using a hypothetical scenario based on safety and efficacy data relating to an atypical antipsychotic agent that was clinically tested against placebo and a comparator compound.

The MCDA methodology provides a way of looking at complex problems and breaking them down into more manageable pieces that can be studied using a mixture of data and judgement. The components are then 're-assembled', using computer software, to present a coherent overall picture for the decision makers. The purpose of this tool is to serve as an aid to thinking and decision-making, but not to take the decisions.

At the conclusion of the meeting it was, as at the 2004 meeting, unanimously agreed that the methodology had great potential and it was recommended that it should be explored further as an adjunct to the decision-making process for both companies and regulatory agencies. An important part of this would be to undertake retrospective testing using actual case studies.

It was also recommended that a third Workshop should be convened at a later stage to discuss the results of such studies and explore further the practical application of the methodology and its place in preparing regulatory submissions as well as in the regulatory review process.

BACKGROUND

The CMR International institute for Regulatory Science became actively involved in the application of MCDA methodology to the benefit-risk assessment of medicines through the research project undertaken by Dr Filip Mussen, MSD Europe, as part of his doctoral thesis, when studying for a PhD as an Institute Scholar[1].

[1] PhD thesis entitled 'The evaluation of methods for benefit-risk assessment of medicines and the development of a new model using multi-criteria decision analysis'. Studies undertaken at the University of Wales, Cardiff, under the supervision of Professor Sam Selek, Welsh School of Pharmacy, and Professor Stuart Walker, CMR International

Benefit-Risk Assessment Model for Medicines –Developing a Structured Approach to Decision-Making, 13-14 June 2005, Washington D.C.

The previous Workshop held in March 2004 was convened at the London School of Economics when the MCDA model was demonstrated using a case study based on a novel treatment for active rheumatoid arthritis. As background to the discussions a summary of criteria and current methods for benefit-risk analysis was presented by Filip Mussen and this is given in *Annex 2* to this report.

Participants at the first Workshop were enthusiastic about the MCDA methodology but it was acknowledged that participation had been predominantly European and one of the recommendations was to hold a similar meeting in the US in order to capture the North American perspective, and with the active participation of the FDA.

FORMAT AND STYLE OF THE WORKSHOP

The second Workshop started from the premise, agreed at the London meeting, that decisions on benefit-risk that decide the fate of a new medicine must be made in a consistent and transparent manner that weighs up the evidence of efficacy and safety objectively, whilst allowing judgement to be exercised.

In the introductory session, participants were asked to outline their experiences and concerns that made them interested in participating in the Workshop. An important driver was the fact that an increasing number of individuals within both companies and agencies are becoming involved in decision-making processes, at all management levels, and there were concerns about the lack of structure and guidance on such processes.

In industry, assessment of the benefit and risks is a continuous process throughout R&D and extends into risk management strategies for products once they are authorised. Management decisions on whether to terminate the development of a new product must take account of commercial risks as well as scientific views on the safety of the product.

The benefit-risk assessments made by regulators may have more clearly defined criteria for efficacy and safety, resulting from regulations and guidelines. There were, however, concerns about the need to ensure that judgement and flexibility can be exercised and that pivotal decisions are always subject to peer review.

There is also a need to take account of the concept of 'benefit' from the patients' point of view, especially in terms of quality of life. In addition, sufferers from life-threatening or chronic diseases may have a different attitude from companies and regulators when it comes to 'acceptable risk'.

Case Study Scenario

The case study for the scenario used to develop the model for benefit-risk assessment related to a hypothetical 'Drug X', an atypical antipsychotic that may offer important clinical benefits to the approximately 1%-2% of the population that suffer from the devastating effects of acute and chronic schizophrenia and schizoaffective disorder.

The following gives the Introduction and Executive Summary from the scenario. The figures and tables referenced in the summary are given in *Annex 3* to this report.

Case Study: Introduction

Schizophrenia is a devastating, debilitating illness for patients, their families and caregivers, and exposes those afflicted individuals to higher degrees of morbidity and mortality than is seen in the general population. While the current generation of atypical agents represents an improvement over older neuroleptics, some combination of extrapyramidal symptoms, tardive dyskinesia, QTc prolongation, weight gain, hypercholesterolemia, hypertriglyceridemia, hyperprolactinemia, and diabetes attends the use of any single or combination antipsychotic regimen. Clearly, the treatment armamentarium is by no means complete or satisfactory, and individuals suffering from this disease require more, not fewer, options. A definite need exists for newer agents whose pharmacological and side effect profiles differ from, and offer improvements over, those of the currently marketed antipsychotics

DRUG X – EXECUTIVE SUMMARY

The case study summarizes safety and efficacy data available for an atypical antipsychotic Drug X, with particular emphasis on ECG findings and other important cardiovascular risk factors.

Efficacy and Tolerability in the Short-Term and Long-Term Management of Psychosis.

In short-term (4 to 6 week), double-blind, fixed-dose, placebo-controlled trials Drug X was superior to placebo in treating the positive, negative, and depressive symptoms associated with an acute exacerbation of schizophrenia or schizoaffective disorder (Fig1). In two one-year, double-blind, placebo-controlled maintenance trials, Drug X significantly reduced the risk of recurrence of acute exacerbation in hospitalized patients with chronic or subchronic schizophrenia (Figs 2 & 3; 4 & 5). Drug X was well tolerated in both the short-term and long-term placebo-controlled trials, with a low overall incidence of adverse events (Figs 6 & 7). In a 196 - week, double-blind parallel group comparison with Drug Y, similar improvements in psychosis were seen with both agents (Figs 8 and 9), along with relatively greater improvements in quality of life and drug acceptability (Figs 10 and 11). Drug X demonstrated a low liability for movement disorder adverse events as evidenced by a low use of antimuscarinic drugs, compared to Drug Y (Fig 12). Drug X treatment was not associated with any laboratory test abnormalities indicative of clinically relevant toxicity (data not shown).

Pharmacokinetics of Drug X have been Studied in Individual Trials as well as in a Population Pharmacokinetic Database (data not shown).

Drug X displays linear pharmacokinetics over the recommended dose range (80 to 160 mg daily) and has a mean half-life of 6.6 hrs. Its relative oral bioavailability is increased by up to 100% in the presence of food. In multiple dose studies, the C_{max} typically occurs at approximately 6 hrs after dosing in the fed state, with steady state attained within 1 to 3 days. Drug X is extensively metabolized by both aldehyde oxidase and P-450 mixed function oxidases (predominantly CYP3A4). Co-administration of CYP3A4 inhibitors or inducers with Drug X results in limited (~35%) increases/decreases in Drug X exposure. No other clinically significant drug-drug interactions have been observed.

The Effect of Drug X on the QTc.

In the short-term, double-blind, placebo-controlled trials submitted with the NDA, doses of Drug X from 80 to 160 mg daily were associated with a mean increase in QTc relative to baseline of 5.9 to 9.7 msec (Bazett correction or 4.4 to 9.3 msec (Baseline correction – data not shown).

Study QT was designed to measure the effects, at peak drug exposure after dosing, of Drug X, Drug M, Drug N, Drug O, Drug P and Drug Y on the QTc (Fig 13). Electrocardiograms were recorded under fasting conditions and at the time of estimated maximum exposure to each study drug, in the absence and presence of a metabolic inhibitor. QT interval measurements were made using standardized 12-lead ECG methodology. Although selected as a comparator in part because it was expected to have no effect on QTc, a relationship between concentration and QTc effect was detected for Drug Y. The QTc effect of Drug X 160 mg was found to be approximately 10 msec greater than the effects of four of the comparative antipsychotics (Drug Y, Drug O, Drug M and Drug N) and 10 msec less than the QTc effect of Drug P (Fig 14).

Drug X Demonstrated no Further QTc Prolongation in the Presence of Metabolic Inhibition.

The metabolism of Drug X is mediated by aldehyde oxidase and by CYP3A4. There are no clinically recognized inhibitors or inducers of aldehyde oxidase. Inhibition of CYP3A4 resulted in an increase in the concentrations of Drug X, but no further increase in QTc effect (Fig 14). Concentration – QTc data collected in Phase 2/3 clinical trials, and review of the clinical experience of those individuals with highest concentrations of Drug X reveal no evidence of clinical events associated with QTc effect in clinical trials (Fig 15).

The Drug X Database Showed No Signal of Increased Cardiovascular Risk.

The QTc interval is highly variable and is affected by a broad set of both internal and external influences. The precise relationship between cardiac repolarization and the risk of serious adverse cardiac events remains unsettled. What is certain is that QTc prolongation is of concern because of its potential to induce syncope, *Torsade de Pointes* and sudden death. Since the overwhelming majority of reported cases of Torsade de Pointes are seen in individuals with measured QTc values of 500 msec or greater, the Drug X database and Study QT were examined for QTc measures >500 msec. In the Phase 2/3 development program overall, 0.06% (2/3095) of patients had a QTc (Bazett) interval ≥500 msec. In Study QT, no Drug X-treated patient had a QTc ≥500 msec, despite coadministration of the metabolic inhibitor ketoconazole to patients receiving the highest recommended dose of Drug X.

DRUG X – EXECUTIVE SUMMARY *cont.*

There is no evidence in the clinical database of excess total mortality, sudden deaths or syncope for Drug X patients compared with patients given placebo or other commonly prescribed antipsychotics. In fact, the mortality rate in the Drug X group has declined slightly over the past four years, and is less than that measured in the placebo group in each reporting category (data not shown). This has occurred while the number of patients receiving Drug X has more than doubled, and the cumulative patient-years of exposure to Drug X have increased nearly three-fold. No episodes of Torsade de Pointes have been reported among the 4571 patients treated with Drug X for a cumulative total of 1733 patient-years exposure. No significant cardiac events were associated with Drug X in the ten overdose cases.

Drug X Is Neutral or has Beneficial Effects on Body Weight and Serum Lipid Levels

Drug X can be differentiated from other atypical antipsychotics with respect to its low propensity to cause weight gain and its beneficial effect on serum lipids. Patients with schizophrenia are likely to have a higher Body Mass Index than individuals in the general population, a trend that is aggravated by the tendency of antipsychotic medications to cause weight gain. In long-term trials, Drug X-treated patients had a lower incidence of clinically significant weight gain than Drug M-treated patients (Fig16).

A favourable effect of Drug X on serum cholesterol levels was first noted in studies in which patients were switched to Drug X from other antipsychotics. Study QT (short-term treatment) demonstrated that Drug X produced decreases from baseline in fasting total cholesterol, LDL cholesterol, and triglycerides, while having no impact on HDL cholesterol (Fig 17). The effect of Drug X on total cholesterol has been demonstrated over 52 weeks of therapy (Fig 18).

Demonstration of the model

As on the previous occasion, the MCDA model was explored using a live, interactive demonstration in which delegates participated in the role of the assessors and the model was developed and tested 'on screen' using customised software.

The approach presented at the Workshop consists of two components, namely a technical model, multi-criteria decision analysis (MCDA), which is based on the theory of decisions with multiple objectives[2], and a social process, decision conferencing, which engages the right people to provide the right data and judgments at the right time[3]. The MCDA modeling approach, originally developed at Harvard University, is now widely taught at university level in management science courses and is the technique used by many other industries as a support to senior decision makers. Combining it with the social component, decision conferencing, which was developed at the London School of Economics and Political Science, enables groups of key players to generate a shared understanding of the benefit-risk issues, to create a sense of common purpose and to gain commitment to the way forward. The process provides a tool to aid thinking, so providing decision makers with a solid basis for effective and smarter decisions that can withstand public scrutiny. Thus, the combination of MCDA and decision conferencing provides more than a model for benefit-risk assessment; it is a process that is intended to support the decision maker by providing an in-depth analysis of the problem, thereby enabling the decision maker to take a more informed decision.

[2] Keeney, R. and Raiffa, H., *Decisions with Multiple Objectives*, New York: John Wiley, 1976.
[3] Phillips, L. D. (1989). People-centred group decision support. In G. Doukidis & F. Land & G. Miller (Eds.), *Knowledge-based Management Support Systems*. Chichester: Ellis Horwood.

THE MCDA MODEL FOR BENEFIT-RISK ASSESSMENT OF DRUG X

The process of developing the benefit-risk model in the Workshop followed the steps for a multi-criteria decision analysis outlined in Multi-Criteria Analysis: A Manual[4], reproduced in Table 1.

Table 1. A summary of the MCDA process. Reproduced from the MCA Manual, p. 50.

Applying MCDA: Detailed steps
1. Establish the decision context. 1.1 Establish aims of the MCDA; identify decision makers and other key players. 1.2 Design the socio-technical system for conducting the MCDA. 1.3 Consider the context of the appraisal.
2. Identify the options to be appraised.
3. Identify objectives and criteria. 3.1 Identify criteria for assessing the consequences of each option. 3.2 Organise the criteria by clustering them under high-level and lower-level objectives in a hierarchy.
4. 'Scoring'. Assess the expected performance of each option against the criteria. Then assess the value associated with the consequences of each option for each criterion. 4.1 Describe the consequences of the options. 4.2 Score the options on the criteria. 4.3 Check the consistency of the scores on each criterion.
5. 'Weighting'. Assign weights for each of the criteria to reflect their relative importance to the decision.
6. Combine the weights and scores for each option to derive an overall value. 6.1 Calculate overall weighted scores at each level in the hierarchy. 6.2 Calculate overall weighted scores.
7. Examine the results.
8. Sensitivity analysis. 8.1 Conduct a sensitivity analysis: do other preferences or weights affect the overall ordering of the options? 8.2 Look at the advantage and disadvantages of selected options, and compare pairs of options. 8.3 Create possible new options that might be better than those originally considered. 8.4 Repeat the above steps until a 'requisite' model is obtained.

[4] Dodgson, J., Spackman, M., Pearman, A., & Phillips, L. (2000). *Multi-Criteria Analysis: A Manual*. London: Department of the Environment, Transport and the Regions. Download by doing a Google search on the manual's title, or from www.odpm.gov.uk/stellent/groups/odpm_about/documents/page/odpm_about_608524.hcsp. Chapter 6 presents a brief overview of MCDA, and Chapter 7 gives case studies.

1. Establish the decision context

The case study on Drug X, outlined above, was provided to Workshop participants. It was noted that Drug X may offer important clinical benefits to the approximately 1% to 2% of the population that suffer from the devastating effects of acute and chronic schizophrenia and schizoaffective disorder. The document summarized safety and efficacy data available for Drug X, Drug Y and a placebo, with particular emphasis on ECGT findings and other important cardiovascular risk factors.

The case study included information about efficacy and tolerability in the short-term and long-term management of psychosis, pharmacokinetics studied in individual trials as well as in a population pharmacokinetic database, effect on QTc, bodyweight and serum lipid levels, and cardiovascular risk.

2. Identify options to be appraised

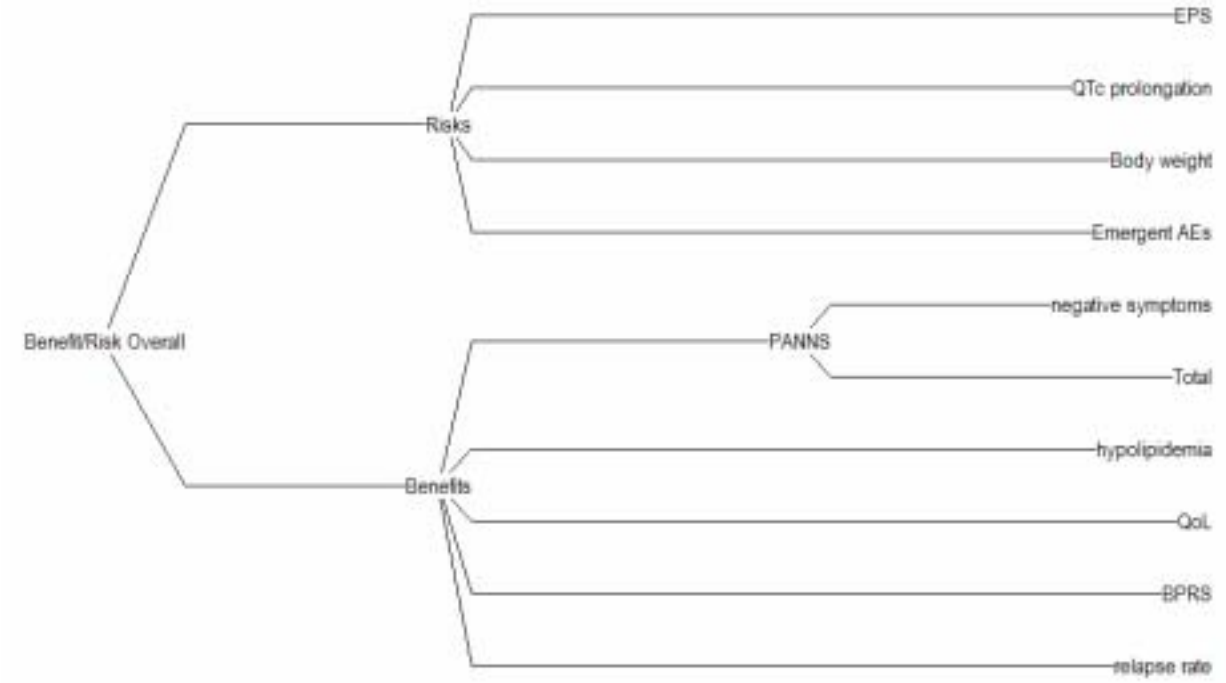
Three options were identified, Drug X, placebo and Drug Y:

1. Drug X
2. Placebo
3. Drug Y

3. Identify objectives ad criteria

The group agreed objectives common to all medicines, namely to maximize benefits while minimizing risks. For this study those objectives were decomposed into concrete benefit criteria associated with the efficacy data, and risk criteria capturing the various safety factors. The structuring of objectives and criteria into a hierarchy is shown in Figure 1.

Figure 1: The value hierarchy showing benefit and risk criteria.



EPS: Extrapyramidal Symptoms. Percent of subjects exhibiting EPS.

QTc prolongation: Prolongation of the QT interval on the surface electrocardiogram (ECG).

Body weight: Incidence of greater than 7% increase. Inverse linear function over a fixed range from 10% (best) to 60% (worst). Data from Figure 16 *(Annex 3 page 17)*, with placebo as base rate.

Emergent AEs: Directly judged preference scale, based on data from Figures 6 & 7 *(Annex 3 pages 6 & 7)*. Placebo is best at 100, Drugs X and Y worst at 0.

PANNS negative symptoms: Positive and Negative Syndrome Scale, a 30-item rating instrument for evaluating symptoms on a scale of 1 (absent) to 7 (extreme). The Negative Subscale Score equals the sum of the seven PANSS Negative Subscale items. Here, the change from the starting or baseline position was entered into the model, so larger numbers are more preferred.

PANNS total: The sum of scores for both negative and positive symptoms. The change from baseline was entered into the model. Larger numbers are more preferred.

Hypolipidemia: Direct preference judgments, based on data shown in Figures 17 and 18 *(Annex 3 pages 18&19)*.

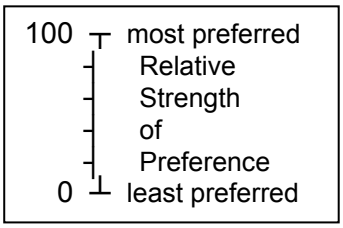
QoL: Quality of Life total score at week 40 in Protocol 300, Figure 10 *(Annex 3 page 9)*.

BPRS: Brief Psychiatric Rating Scale (BPRS), an 18-item inventory of psychotic symptoms for evaluating the effects of drug treatment in psychosis.

Relapse rate: Probability of relapse at 52 weeks. Data from Figure 2 *(Annex 3 page 2)*.

4. Score the options

The purpose of scoring in MCDA is to achieve for each criterion a scale extending from 0 to 100, with those two points on the scale clearly defined. They can be defined independently of the options under consideration, or with reference to the least and most preferred options, as was done for this study. Scoring is the process of assigning numbers to the options on the criterion under consideration. The scale can be thought of as representing relative value or 'strength of preference'. Since it is an interval scale, like Celsius or Fahrenheit, whose zero points and units of measurement are arbitrarily chosen, it is important to appreciate that only differences in value can be compared on such scales. For example, if options A, B and C have been scored at 100, 80 and 0, respectively, then the difference in preference value between B and C is four times as big as the difference in value between A and B. It is wrong to suggest that option B, scoring 80, is 80% as good as option A, scoring 100, or that option D, scoring 40, is half as good as option B. It is ratios of differences in the scores that can be compared, not ratios of the scores themselves.



In this study, two approaches were used, direct and indirect scoring. Direct scoring required the group to identify the most and least preferred options on a given criterion, assign these scores of 100 and 0, then to score the remaining option so that differences in the scores reflected differences in preference. This approach was taken for the AEs and hypolipidemia criteria. All the other criteria used indirect scoring. Measurable data for these criteria provided inputs to the model, and the software converted these inputs to preference value on 0 – 100 scales. The least and most preferred measures on those scales, assigned preference values of 0 and 100, respectively, had been agreed by the group as defining a range within which the measurable data for the options would fall. In the case of benefits, the translation was usually direct: higher numbers (e.g., higher PANNS scores) converted to higher preference values. An exception was the relapse rate; lower relapse rates are more preferred. For all the risk criteria (e.g., EPS), the conversion was inverse: lower input scores mapped to higher preference values. Thus, direct scoring enabled preference judgments to be input to the model in the absence of hard data, whereas indirect scoring used available data. In all cases but one, the input data were converted linearly to preference scores. The one exception was QTc prolongation, for which a regulator assessed a non-linear inverse value function, shown in Figure 2.

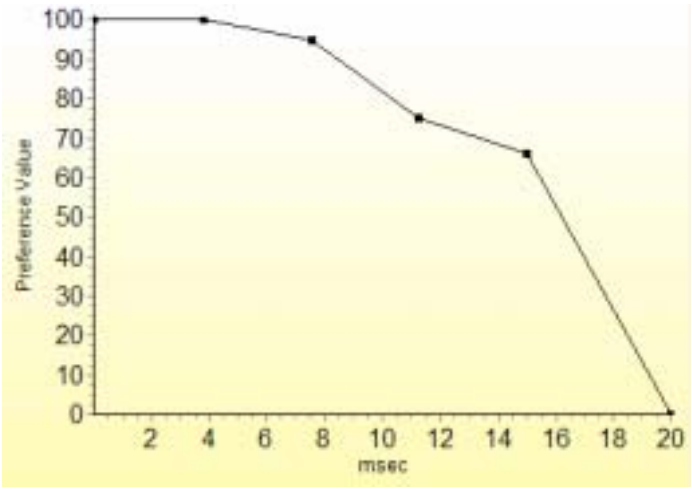


Figure 2: Value function relating preference to QTc prolongation.

This rather surprised some of the participants from the pharmaceutical industries, as they had assumed that regulators would exhibit a sharp decline after about 5 to 10 msec, given the potential for fatalities above 20 msec, whereas the assessed function is still at about 67 at 15 msec. Of course, this being the first time the regulators had encountered the concept of a value function relating hard data to preference values, this function should not be at all considered as anything other than a first rough judgment that would, necessarily, benefit from being refined in consultation with other colleagues.

One problem that arose during the scoring was the absence of some data for some options. The current version of Hiview automatically inserts preference values of zero for missing data, but for this application that might benefit risk criteria and penalize benefits. Thus, the group assigned values that neither penalized nor benefited the option, using their judgments to determine an appropriate value. There is, of course, no problem for MCDA to handle missing data; this is merely a correctable problem with the current version of Hiview.

5. Weight the criteria

Weighting in MCDA is the process of ensuring the equality of a unit of preference value on all the 0 to 100 scales. The scoring process results in a relative scale for each criterion, but the value difference between 0 and 100 may be different for each scale, as a Celsius degree represents a different unit of temperature from a Fahrenheit degree. The process of equating the units of value was accomplished by asking participants to compare the swings in preference from 0 to 100 on all the scales. This is a process of identifying the options associated with 0 and with 100 on a particular scale, then asking the group how big the difference is between those options and how much they care about that difference, as compared to 0-100 differences on other criteria. The process is called 'swing-weighting'[5].

It is this step that is perhaps the most misunderstood in MCDA, for weights are often thought to reflect the absolute importance of the criteria. Not so. The following example was given in the workshop. If you were to purchase a new car, would you consider price to be important to your decision? Most people answered, 'yes'. Since you can't consider all possible cars on the market, imagine that you construct a short list that includes just five cars. Suppose the difference in cost between the least and most expensive is $200. Now is price an important consideration in your decision? Most participants said 'no'. But if the difference in price were $2,000, then many said that would be more important. Unless, we pointed out, you are very wealthy, in which case a difference of $2,000 might not have much impact. The point of this example is two-fold: first, a criterion's weight depends on the range of difference in the input data, and secondly, on how much you care about that difference. Inevitably, that has to be a judgment. Thus, balancing risks and benefits requires judgment; it is inescapable, whether MCDA is used or not. While objective data may be available to establish the size of the difference between least and most preferred options on a criterion, the assessment of how much that difference matters requires, inevitably, an act of judgment. MCDA makes those judgments explicit.

Weights in MCDA represent trade-offs. Once the weights are established, they show how much an increase on one criterion is equal to an increase on another. An increase of 9° Fahrenheit is equal to an increase of 5° Celsius. Swing-weighting was applied in the workshop by comparing the 0 to 100 preference value ranges on all the scales. This required looking at the ranges of the input measures on all the fixed scales as compared to the preference-value ranges on the directly-assessed scales. First, the largest swing was identified. It was given a weight of 100. Then all other swings on the other criteria were compared to 100 and assigned appropriate weights. Thus, if the swing on another criterion was judged to provide half the swing in preference value, then that criterion was given a weight of 50. Weights are ratio scale numbers, since they compare differences, so there need not be a zero-weighted criterion.

Comparing the criteria two at a time, the group first agreed that the range from 0 to 20 points on the fixed BPRS scale was the largest swing that they cared about. The swing on each of the other

[5] A brief introduction to swing weighting is given in Section 6.2.10, pp. 62-3, of the MCA Manual.

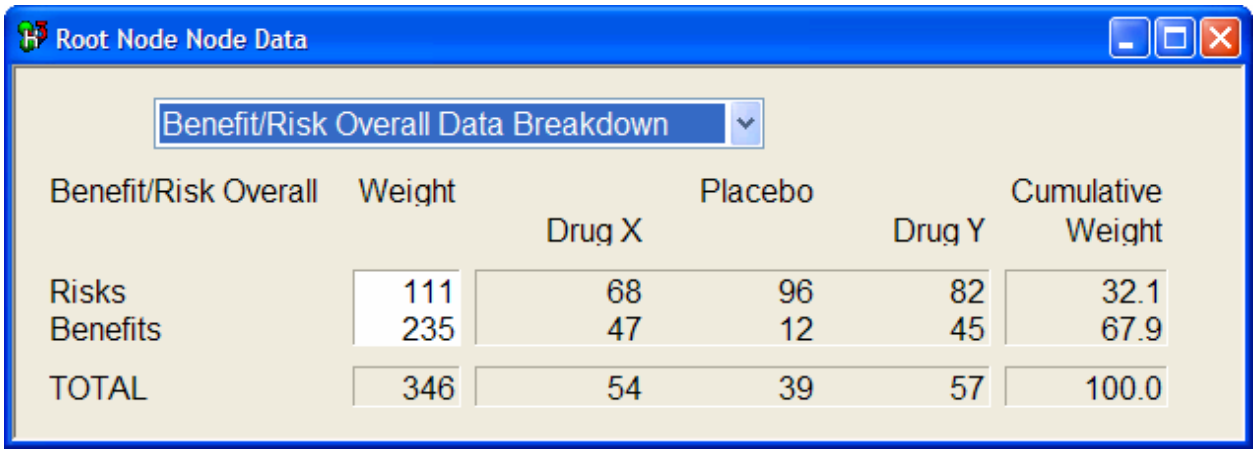
criteria was then compared to this one. For example, the swing on relapse rate, a range from 0 to 100%, was judged to be 80% as great as the 0-to-20-point BPRS swing, so was given a weight of 80. The swings on each of the PANNS criteria were judged to be 10, so those swings contribute a total value of 20. Adding the relapse rate swing of 80 gives a total of 100. So, a consistency check asked if the sum of swings on both PANNS criteria and the relapse rate was equivalent to the BPRS swing. That was judged to be so. Thus, weighting was accompanied by consistency checks, which occasionally led to revisions of the weights, thus helping to ensure the realism and consistency of the weights. Weights for lower-level criteria were then summed by Hiview to give the weights at the level of Risks and Benefits.

Although the explanation of weighting seems complex, in practice it was taken one step at a time, never requiring the group to compare more than two scales. The result is a set of very concrete comparisons, which while not necessarily easy to make, are considerably easier to do than making a holistic judgment of the relative importance of benefits and risks. Summing the criterion weights gives a total weight of 235 on Benefits and 111 on Risks, about a 2:1 ratio between the two clusters of criteria. These weights represent the total added value from the 0 to the 100 positions on the criterion scales. Rationale for the concrete comparisons of two criteria at a time could have been captured and written down, providing an audit trail for the final weights.

6. Combine the weights and scores for each option to derive an overall value.

The Hiview3 computer software does this. It normalizes the weights at each node (making them add to 1.0 while preserving their ratios) and then multiplies the preference values for each option by the weights on the respective criteria. This process is repeated up through the value tree, giving a final set of weighted preference values.

The final preference values and weights for all the criteria and higher-level nodes are shown in Figure 3.



Benefit/Risk Overall	Weight	Drug X	Placebo	Drug Y	Cumulative Weight
Risks	111	68	96	82	32.1
Benefits	235	47	12	45	67.9
TOTAL	346	54	39	57	100.0

Risks Node Data

Risks Data Breakdown

Risks	Weight	Drug X	Placebo	Drug Y	Cumulative Weight
EPS*	26	75	82	49	7.5
QTc prolongation*	75	66	100	96	21.7
Body weight*	7	92	100	92	2.0
Emergent AEs*	3	0	100	0	0.9
TOTAL	111	68	96	82	32.1

Benefits Node Data

Benefits Data Breakdown

Benefits	Weight	Drug X	Placebo	Drug Y	Cumulative Weight
PANNS	20	53	0	38	5.8
hypolipidemia*	5	80	0	100	1.4
QoL*	30	45	0	20	8.7
BPRS*	100	30	0	35	28.9
relapse rate*	80	65	36	65	23.1
TOTAL	235	47	12	45	67.9

PANNS Node Data

PANNS Data Breakdown

PANNS	Weight	Drug X	Placebo	Drug Y	Cumulative Weight
negative symptoms*	10	50	0	30	2.9
Total*	10	55	0	45	2.9
TOTAL	20	53	0	38	5.8

Figure 3: Preference values and weights for the Drug X benefit/risk model.

7. Examine results

The overall result is shown in the top matrix of Figure 3. Drug X scores 54 compared to Drug Y at 57, suggesting that, overall, Drug X is not an improvement over Drug Y. The extra 2 points of benefits is more than overbalanced by the 14-point loss in risk. Both drugs are better than the Placebo, and it is worth noting that the placebo's score, 39, is largely the result of its low risk. These results can be seen graphically in Figure 4.

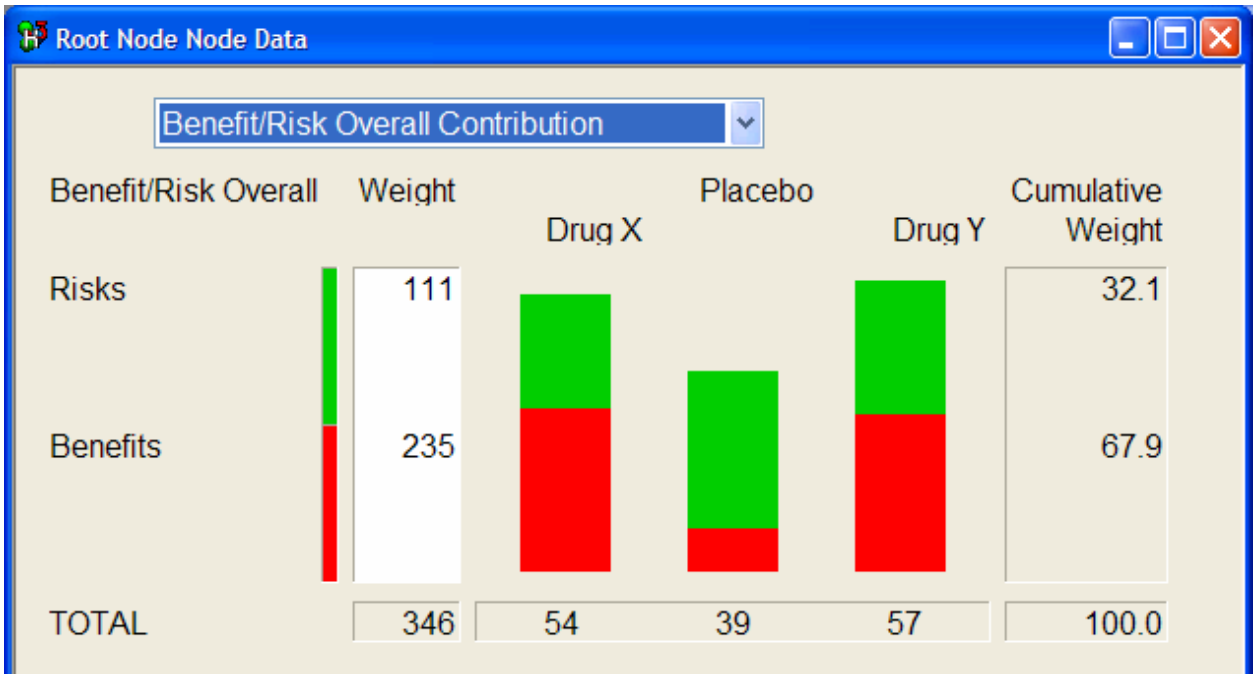


Figure 4: Overall results for the Drug X benefit/risk model. Note that more red indicates more benefits, but because less risk is preferred to more, and preference is shown here, more green means less risk. Obviously, the placebo is least risky, with Drug X and Drug Y about equally beneficial. Overall, Drug Y is best because it is less risky.

It is clear that Drug Y is overall most preferred. While it is about the same overall benefit as Drug X, it is less risky. The relative benefits outweigh the relative risks for all three options, but this is largely due to the relatively higher weight placed by the group on benefits than risks.

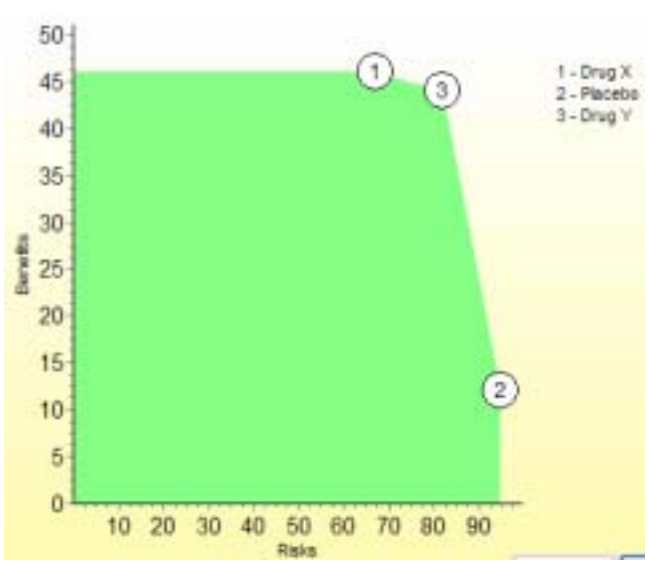


Figure 5: Overall benefits versus risks for the three options.

Another way to see these results is to examine the overall weighted preference values versus the overall weighted risk preference, Figure 5. The placebo is lowest in risk (a high preference value) but lowest in overall benefits. Option 1, Drug X is highest in benefits, but less good than Drug Y in risks. The added benefit of Drug X is bought at a disproportionate increase in risk, leaving Drug Y overall most preferred.

This display and the top matrix of Figure 3 illustrate the trade-off between risks and benefits in a way that is difficult to do with words. Recall that the weighting process equated a unit of benefit to a unit of risk, so it is possible to see clearly in Figure 5 that the 14-point increase in risk of Drug X over Drug Y exceeds the 2-point gain in benefits.

8. Sensitivity analyses

Since the final result is sensitive to the relative weight on risks and benefits, a sensitivity analysis showed the extent to which the final result was determined by these weights. This is shown in Figure 6.

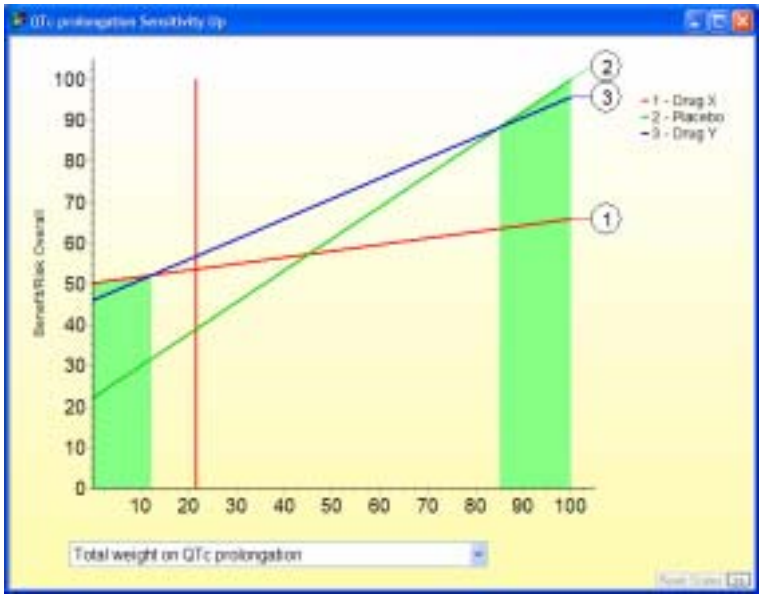


Figure 7: Sensitivity analysis on QTC prolongation.

The vertical red line indicates the relative weight on Risk, 32%, leaving 68% on benefits. At that weight, Option 3, Drug Y, is most preferred. That 32% has to be more than doubled, leaving less than 35% on benefits, for the most preferred option to become the Placebo. (The green shading only indicates the transition in overall preference.) Thus, a wide range of disagreement on these relative weights can be tolerated while those disagreeing can agree about the best option. This is an important consideration in allowing a committee to arrive at agreement about the way forward short of consensus about details.

Another sensitivity analysis showed the overall results as the weight on QTc prolongation was increased. Figure 7 shows that the current weight on this node is about 21%. That weight would have to be very greatly increased before the Placebo would be preferred. Clearly, a decision about the benefit/risk trade-off does not depend on agreement about the weight on this criterion.

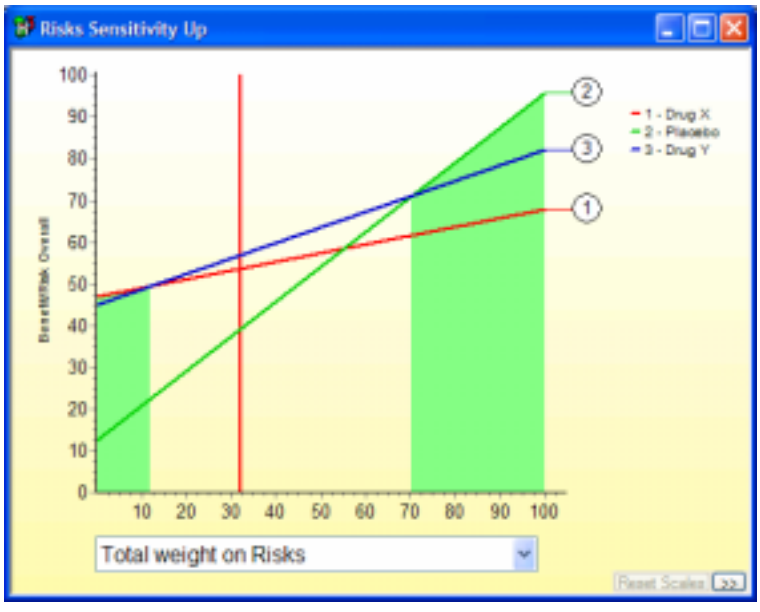


Figure 6: Sensitivity analysis on Risks.

A sensitivity analysis on the BPRS criterion revealed an insensitivity of overall results to the weight on this criterion. Figure 8 shows that over the entire range of weight on BPRS, Drug Y remains the most preferred option. Data on this variable does not discriminate the options, so further or better data about BPRS would not assist decision making.

Hiview can carry out sensitivity analyses on the weights for all bottom-level criteria to see which ones might change the overall results. This is shown in Figure 9. This analysis proved that the model is robust to many different viewpoints about the relative importance of criteria and nodes. This is not, of course, always the case. Situations will arise in which one or more weights will make a difference. In this case, the MCDA model will not resolve those differences because they are essentially judgments. However, the sensitivity analyses are still useful in identifying areas in which more data might be required in order to resolve those differences in judgment. More importantly, the sensitivity analyses show that differences of opinion on many weights do not need to be resolved, so that more debate and discussion is unnecessary.

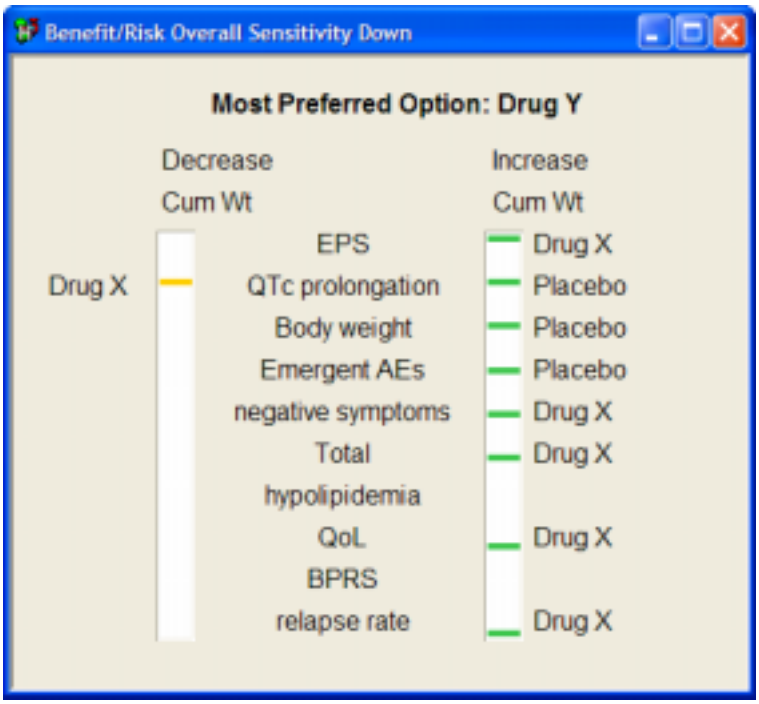


Figure 9: Sensitivity analyses on all bottom-level criteria show the magnitude of change in a criterion weight required for the most preferred option overall to change to the option on the right (for an increase in weight) or the left (for a decrease in weight). The green bars on the right indicate that large changes would be needed, while the yellow bar on the left identifies a modest change. The absence of red bars, flagging small changes that would change the overall result, demonstrates that the model is robust to changes in the weights.

Sorts

These analyses provide comparisons of one option to another, and are useful in diagnosing why a particular option is better or worse than another option. However, before exploring these analyses, it is helpful to see the extent to which the 10 criteria are true discriminators among the options. This is shown in Figure 10.

The cumulative weights shown are based on the swing weights judged by participants. Hiview normalizes the weights so they sum to 100 across all the criteria, while preserving the ratios of the weights as they were assessed. Note from the CumWt column, where the criteria have been sorted from the largest to the smallest cumulative weight, and from their cumulative sum shown in the Sum column, that the first three criteria, BPRS, relapse weight and QTc prolongation, account for nearly 74% of the total weight on all the criteria. Ninety percent of the total weight is captured in half the criteria, so those criteria will be the ones that have a major impact on the overall results. Additional data on the other criteria are unlikely to contribute very much at all to final decisions. So, this kind of analysis could help regulators identify where more data might make a difference to their decisions.

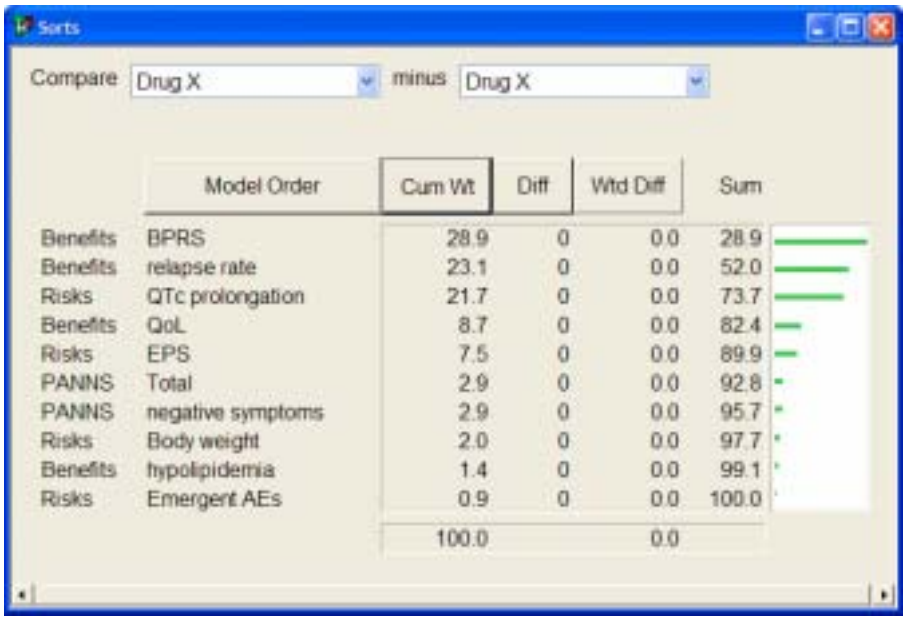


Figure 10: Cumulative weight associated with each bottom-level criterion.

The most useful comparisons explored during the decision conference were a comparison of Drug to the Placebo, Figure 11, and of Drug X to Drug Y, Figure 12. In both figures the difference in scores between the two options for a given criterion are shown in the Diff column, with that difference multiplied by the cumulative weight on the associated criterion shown in the Wtd Diff column. The sum of the weighted differences equals the difference between the options shown in the Total row of the top matrix in Figure 3.

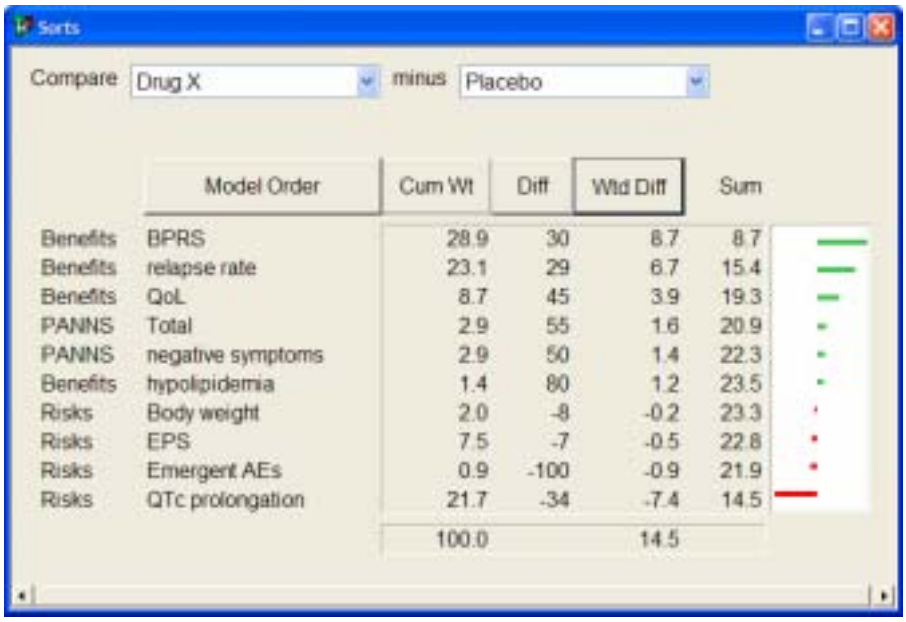


Figure 11: Comparison of Drug X with the placebo.

The main take from Figure 11 is that the key advantages of Drug X over the placebo are the BPRS score, relapse rate and quality of life, three big differences on two highly weighted criteria and one criterion with modest weight. Of course, the placebo's only serious advantage is in the QTc prolongation. All the other weighted differences make virtually no contribution to the discrimination between the two options. The overall difference of 14.5 is the same as the rounded-off 15 point difference shown in the top matrix of Figure 3. The figure shows the part-differences associated with each criterion. Thus, the cumulative advantages of 23.5 points of Drug X are reduced but not overtaken by the smaller advantages of the placebo.

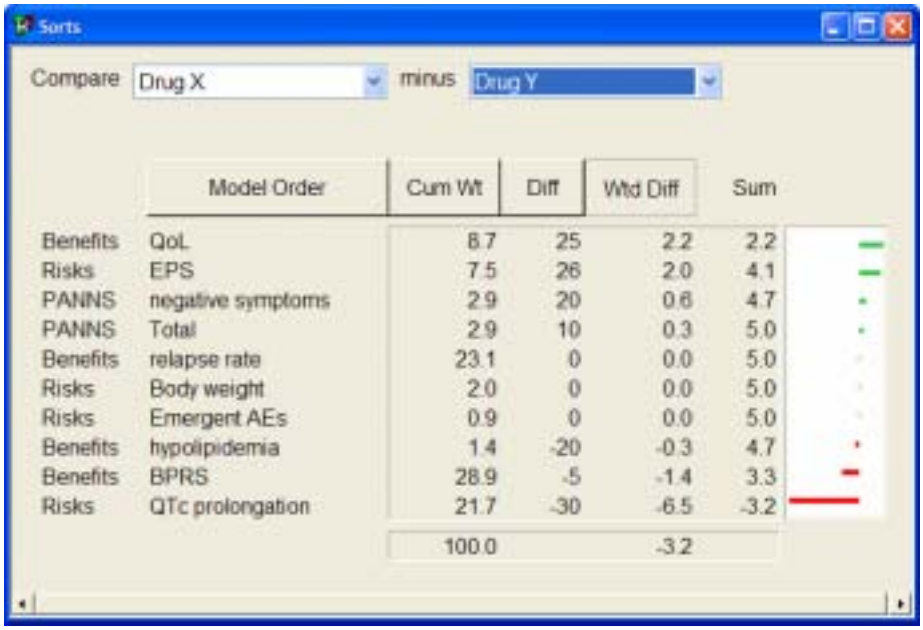


Figure 12: Comparison of Drug X with Drug Y.

The comparison of Drug X with Drug Y, shown in Figure 12, shows that the three-point difference in overall score favoring Drug Y is due to the substantially better QTc prolongation of Drug Y, helped slightly by its better BPRS score, which together overcome the two main advantages of Drug X, quality of life score and EPS. Note that large advantages on criteria with small weights, such as the better hypolipidemia of Drug Y, do not contribute to discriminating the criteria. The best discriminators are large difference scores on heavily weighted criteria. It is the ability of MCDA to properly combine these two factors that provides a substantial advantage of the approach over unaided human judgment.

Summary

This modeling exercise, while somewhat hypothetical for the group, served to demonstrate several features of multi-criteria decision analysis as it might be applied to examining the benefit/risk balance for a medicine.

1. It shows the importance of establishing the appropriate benefit and risk criteria.
2. It distinguishes between two types of judgment needed to establish the benefit and risk: assessments of the values of each option on the criteria:
   - the process of scoring and establishing value functions for converting data to preference values, and
   - the relative weights of the criteria, i.e. the swing-weighting process.
3. It shows how a higher-level view can emerge from aggregating the benefits and the risks, a view that was not obvious at the level of the individual criteria.
4. It demonstrates the value of the social process, which engaged people with differing perspectives on the issues, using the MCDA model to serve as an agenda for a structured conversation and a tool for thinking.

Finally, the process had demonstrated how the decision conferencing process can create aligned commitment in a group of key players who can now agree the way forward even if they don't agree about all the details.

WORKSHOP CONCLUSIONS AND RECOMMENDATIONS

In the final session of the Workshop, participants were asked to reflect on the demonstration of the model and provide their observations on its application in drug development and review.

Observations

- Company members could envisage the approach being used to shape the way in which arguments on the benefit-risk balance of a new product were presented in a regulatory submission;
- FDA expressed the view that the methodology could have a valuable place in in-house decision-making by both companies and regulatory agencies but that, at present, it would not be a suitable way to present data for assessment or at a hearing.
- It was acknowledged that there might be significant organisational and 'cultural' challenges to introducing the methodology and it would require individuals with enthusiasm and conviction to overcome these.
- Consistency among agencies is a problem and the MCDA model could have a place in research into inconsistencies within agencies and differences of opinion between agencies.
- The methodology might provide greater insight into why products fail in Phase III.

Recommendations

- There was a clear consensus that the participants wished to see the CMR International Institute carrying this project forward with further validation of the MCDA model.
- It was strongly recommended that CMR should undertake a specific study of the methodology, in collaboration with regulatory agencies and pharmaceutical companies, using data from 'real-world' examples from past and current cases. This might include:
  - A retrospective study, including products that had failed at the pre-and post-submission stage, using the model to re-evaluate the benefit-risk decisions that had been taken;
  - A pilot study, starting with early portfolio projects, in a therapeutic area such as diabetes;
  - A comparison of outcomes, using the MCDA model, when applied in the US and in the EU.
- It was agreed that the 'learning' process would benefit from a further interactive Workshop, organised on a similar basis to the current meeting, with a limited number of participants and a further 'hands-on' demonstration of the methodology.

The Way Forward

Project: To develop a structured approach to decision-making for benefit-risk assessment of medicines

Following the Workshop in June 2006 several immediate actions have been agreed and initiated to take this project forward:

Publications

The publication of two papers is proposed under the title '*A structured approach to the Benefit-Risk Assessment of Medicines':*

1. Development of a new model using multi-criteria decision analysis (MCDA)
2. The practical application of a new model

It has been proposed that these should be published in the journal of Clinical Pharmacoepidemiology and Drug Safety, ideally in the first half of 2006.

Further doctoral study

In April 2006, a PhD student will commence a 3-year research study. The objectives will be to:

- Identify and determine how decision-making is implemented in established drug regulatory authorities;
- Identify, by means of a questionnaire, how 'best practices' are incorporated into decision-making by industry and regulatory authorities;
- Refine and develop the MCDA model for benefit-risk assessment in order to optimise its practical application in real-life situations.

A pharmacy graduate with a first class honours degree has been identified for this study.

Incorporating the views of patients

CMR International will continue the successful formula of interactive workshops to demonstrate the model to regulatory and industry participants, as recommended by the Workshop. It is proposed that these should focus on:

- Comparing outcomes from the same scenario evaluated by different stakeholders;
- Seeking ways to incorporate the views of patients, using the methodology, and evaluating the impact on outcomes.

Institute Workshop, December 2006

The CMR International Institute will be holding a 'regular' Workshop, in the UK in December 2006 that will look at *Criteria for quality decision-making during drug development and review.*

The Workshop will provide an opportunity to discuss the outcomes and recommendations from the two special focus Workshops (March 2004 and June 2005) on building a model for benefit risk-assessment as well as other aspects of building quality into decision-making. The focus will not only be on the decision-making process by the regulatory agencies at the end of review, but also on the way companies make decisions at key milestones in the drug development process.

Further details will be posted on the website www.cmr.org/institute.

ANNEX 1

WORKSHOP PARTICIPANTS

Dr Andre Broekmans	Vice President, Regulatory Affairs & Pharma Policy	Organon International Inc, USA
Dr George Butler	Vice President, Customer Partnerships	AstraZeneca Pharmaceuticals, USA
Dr Lyn Caltabiano	Vice President, Clinical Pharmacology Operations	GlaxoSmithKline, USA
Dr Patrizia Cavazzoni	Director, Global Drug Safety – Neuroscience	Eli Lilly & Company, USA
Dr Janine Collins	Director, Drug Safety	Celgene Europe Ltd, UK
Ms Margaret Cone	Director of Regulatory Science	CMR International Institute
Prof Bruno Flamion	Chair, Scientific Advice Working Party and CHMP Member	EMEA, UK
Dr Trevor Gibbs	Senior Vice President, Senior Physician, Medical Governance and Pharmacovigilance	GlaxoSmithKline, UK
Dr Edmund Harrigan	Senior VP, Worldwide Regulatory Affairs & Quality Assurance	Pfizer Inc, USA
Dr John Howell III	President	Portfolio Decisions Inc, USA
Dr David Jefferys	Senior Regulatory Strategic Adviser	Eisai Europe, UK
Dr John Jenkins	Director, Office of New Drugs, Center for Drug Evaluation & Research	Food and Drug Administration, USA
Dr Sandra Kweder	Deputy Director, Office of New Drugs, Center for Drug Evaluation & Research	Food and Drug Administration, USA
Prof Jan Liliemark	Scientific Director	Medical Products Agency, Sweden
Dr Murray Lumpkin	Acting Deputy Commissioner	Food and Drug Administration, USA
Dr Freeman Marvin	Principal	Innovative Decisions Inc, USA
Dr Neil McAuslane	Chief Scientific Officer	CMR International Institute
Patricia McGovern	Director, Special Projects	Novartis Pharmaceuticals Corporation, USA
Charles Monahan	Senior Regulatory Affairs Associate	Millennium Pharmaceuticals Inc, USA
Dr Filip Mussen	Associate Director, Regulatory Affairs	Merck Sharp & Dohme (Europe) Inc, Belgium
Dr Thierry Nebout	Senior Consultant	Institut de Recherches Internationales Servier, France
Professor Larry Phillips	Professor of Decision Analysis	London School of Economics, UK
Dr Atsuko Shibata	Associate Medical Director	Amgen Inc, USA
Dr Robert Temple	Associate Director for Medical Policy, Center for Drug Evaluation	Food and Drug Administration, USA
Dr Mary Ellen Turner	Vice President	Wyeth Research, USA
Prof Stuart Walker	President and Founder	CMR International
Dr Wayne Wallis	Head, Global Safety	Amgen Inc, USA

Benefit-Risk Assessment Model for Medicines –Developing a Structured Approach to Decision-Making, 13-14 June 2005, Washington D.C.

ANNEX 2

CRITERIA AND CURRENT METHODS FOR BENEFIT-RISK ANALYSIS[6]

Definitions and terminology in benefit-risk assessment

- *Benefit*: the proven therapeutic good of a product; should also include the patient's subjective assessment of its effects (WHO Collaborating Centre)
- *Risk*: the probability of harm being caused; the probability (chance, odds) of an occurrence (WHO Collaborating Centre)
- *Benefit and risk* are evaluative terms which contain value judgements (clinical studies cannot determine whether an effect is a benefit/risk, and how beneficial/harmful the effect is)
- *Benefit-risk balance*: more accurate than benefit-risk ratio (benefits and risks are not of the same nature)

Five concepts in benefit-risk assessment

- A separate benefit-risk balance for each indication
- All available data should be considered in benefit-risk assessment
- The nature of the disease should be taken into account for benefit-risk balance
- Absolute versus relative benefit-risk balance (compare with alternative therapies?)
- The benefit-risk balance is dynamic and evolves over time

Criteria to consider in benefit-risk assessment

The criteria selected are based on EU, FDA and ICH guidance. (The numbers in brackets refer to the 20 responses - 14 companies, 6 agencies - to questions in a CMR survey in 2002 that asked which factors should be included in a model for benefit risk assessment)

Benefit

For each pivotal trial:

- Efficacy (primary endpoint) versus comparator and its clinical relevance *(20/20)*
- Statistical significance of the efficacy results *(18/20)*
- Clinical relevance of the primary endpoints *(19/20)*
- Representativity of the studied population for the population targeted in the label *(18/20)*
- Evidence for the efficacy in relevant subgroups *(14/20)*
- Design, conduct and statistical adequacy of the trial *(18/20)*
- Confirmation of treatment effect by results of non-primary endpoints *(16/20)*

General benefit criteria:

- Confirmation of efficacy by results of relevant non-pivotal trials and extensions *(16/20)*
- Anticipated patient compliance *(11/20)*
- Clustering (consistency) of results of the pivotal trials

Risk

- Overall incidence of adverse effects (from clinical trials) *(16/20)*

[6] Extracts from a paper presented by Filip Mussen at the first CMR International Institute Workshop on developing a model for benefit-risk assessment, March 2004

- Overall incidence of serious adverse effects (from clinical trials) *(20/20)*
- Discontinuation rate due to adverse effects (from clinical trials) *(15/20)*
- Incidence, seriousness and duration of specific adverse effects (from clinical trials and post-marketing surveillance) *(20/20)*
- Interactions with other drugs and with food *(18/20)*
- Safety in subgroups (e.g., age, race, sex) *(20/20)*
- Potential for off-label use leading to safety hazards *(12/20)*
- Potential for non-demonstrated additional risk due to limitations of clinical trials and/or short market exposure
- Potential for non-demonstrated additional risk due to safety issues observed in preclinical safety studies but not in humans
- Potential for non-demonstrated additional risk due to safety issues observed with other medicines of the same pharmacological class

(The latter three criteria were previously clustered in the survey as 'generalizability of the safety profile to the general population' (18/20))

Why would models for benefit-risk assessment be useful?

- Enhance consistency in expressing the benefit-risk balance of a product
- Enhance objectivity in recommendations/ decisions on the benefit-risk of a product (by Registration Committees and in Marketing Authorization Applications)
- Increase transparency of regulatory decisions (approval and post-approval)
- Force the assessor to focus on benefits and risks
- Ideally, could be used as a tool to compare products
- Can be used as a tool for regulators and industry, but cannot substitute for the final decision-making

Objectives for a new model

A model which

- Is able to take into account the data in the MAA or otherwise available to regulatory agencies (i.e., safety and efficacy data from multiple clinical trials, post-approval AE data); No cost-benefit data
- Requires no additional analyses of source data (safety and efficacy), or meta-analyses
- Closely matches the current regulatory agency practices for benefit-risk assessment
- Can be used during initial registration and post-approval
- Can be validated
- Is applicable to all kind of drugs, including vaccines and OTC drugs

Which models are currently available?

Currently there are no well-established, validated models (qualitative or quantitative) although a few models are described in the literature:

- 'Principle of threes' (Edwards et al, 1996)
- TURBO model (Amery, 1998)
- Evidence-based benefit and risk concept (Beckmann, 1999)

These three models were mainly developed for pharmacovigilance purposes - post-marketing re-assessment (the 'Principles of Threes' model and the TURBO model are described in the CIOMS IV report)

Benefit-Risk Assessment Model for Medicines –Developing a Structured Approach to Decision-Making, 13-14 June 2005, Washington D.C.

Other models have been developed to assess the benefit-risk based on one clinical trial, e.g.:

- 'Benefit-Less-Risk Analysis' (Chuang-Stein)
- Mathematical model based on Numbers Needed to Treat (NNT) & Numbers Needed to Harm (NNH) (Schulzer & Mancini)
- 'Principle of threes' grading system (Edwards et al.)

Weaknesses of the current models

- Many criteria in the models are not well defined with regard to the type, quality and relative importance of the data to be taken into account
- Models do not take into account many of the benefit and risk criteria previously identified
- Models are not very sophisticated and allow only a very crude benefit-risk assessment
- Models have not been validated nor broadly used in practice

Benefit-Risk Assessment Model for Medicines –Developing a Structured Approach to Decision-Making, 13-14 June 2005, Washington D.C.

ANNEX 3

Case Study

Figures and data from the scenario used to develop the model for benefit-risk assessment of Drug X

Figure 1

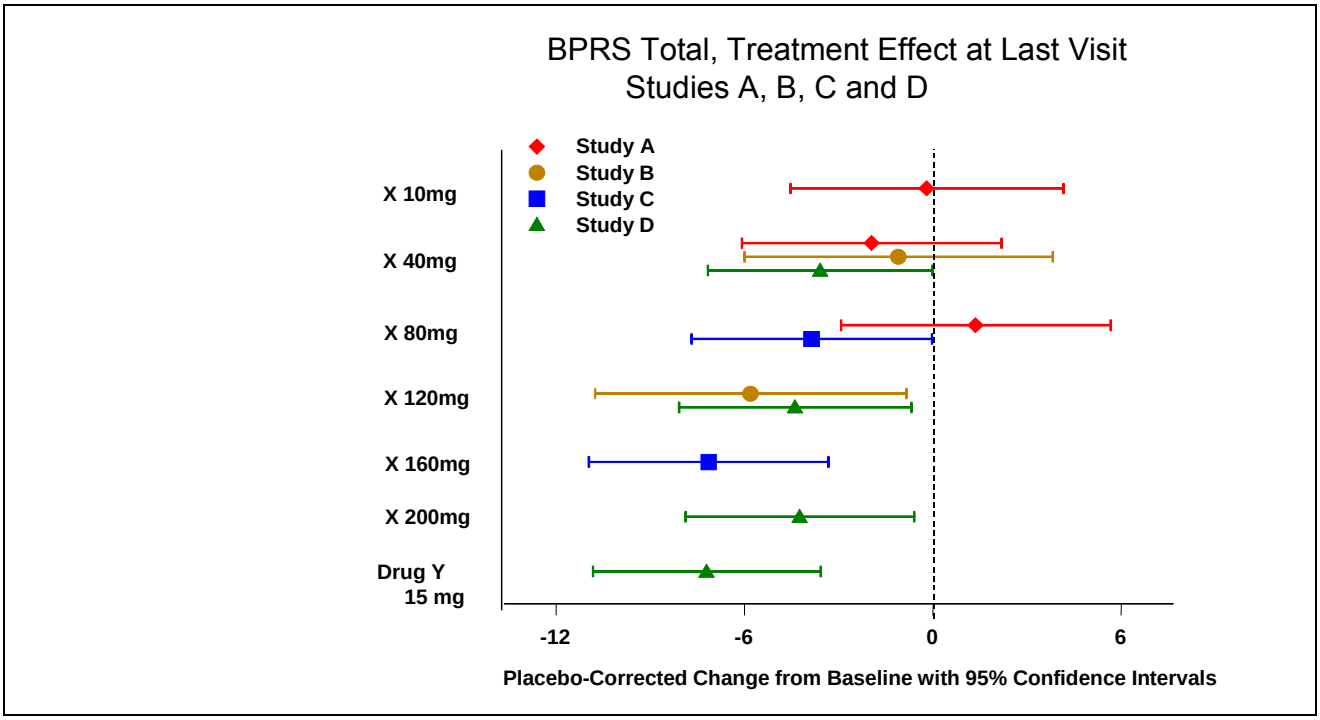


Figure 1. Summary of Estimated Treatment Effects on BPRS Total at Last Visit in Short-Term, Placebo-Controlled Trials

Figure 2

Analysis of Time to Relapse – All subjects
Product X Protocol 100

PROTOCOL 100:	MULTICENTRE DOUBLE-BLIND STUDY OF DRUG X VERSUS PLACEBO IN RELAPSE PREVENTION FOR HOSPITALISED PATIENTS WITH CHRONIC OR SUBCHRONIC SCHIZOPHRENIA
Study Objectives:	Compare the efficacy of Drug X 40 mg BID with placebo
	Investigate the dose-response relationship of three fixed dose regimens of, Drug X (20, 40 and 80 mg BID) in the prevention of psychotic relapse
	Evaluate the treatment effects on a subgroup of subjects with predominantly negative symptoms
	Measure concentrations of Drug X in the serum and provide data for the evaluation of the population pharmacokinetics of Drug X
	Assess the safety and toleration of the three Drug X doses
Study Design	A randomized, double-blind study of up to 52 weeks treatment commencing with placebo run-in.

Analysis of Time to Relapse – All subjects
Product X Protocol 100

		Cum Incidence (%) of Relapse		Probability of Relapse		95% Confidence Limits			
Treatment Group	**N**	**<= 28 Wks**	**<=52 wks**	**<=28 wks**	**<= 52 wks**	**Rel Risk**	**Lower**	**Upper**	**p-value**
Product X									
20 mg BID	75	23 (31)	27 (36)	0.339	0.405	0.481	0.296	0.781	0.003
40 mg BID	72	21 (29)	22 (31)	0.326	0.346	0.414	0.247	0.693	0.001
80 mg BID	71	22 (31)	24 (34)	0.324	0.358	0.411	0.249	0.68	0.001
Placebo	75	35 (47)	34 (54)	0.428	0.638				
Overall									<0.001

Figure 3
Kaplan Meier Curves for Time to Relapse by Treatment Group
All Subjects
Product X Protocol 100

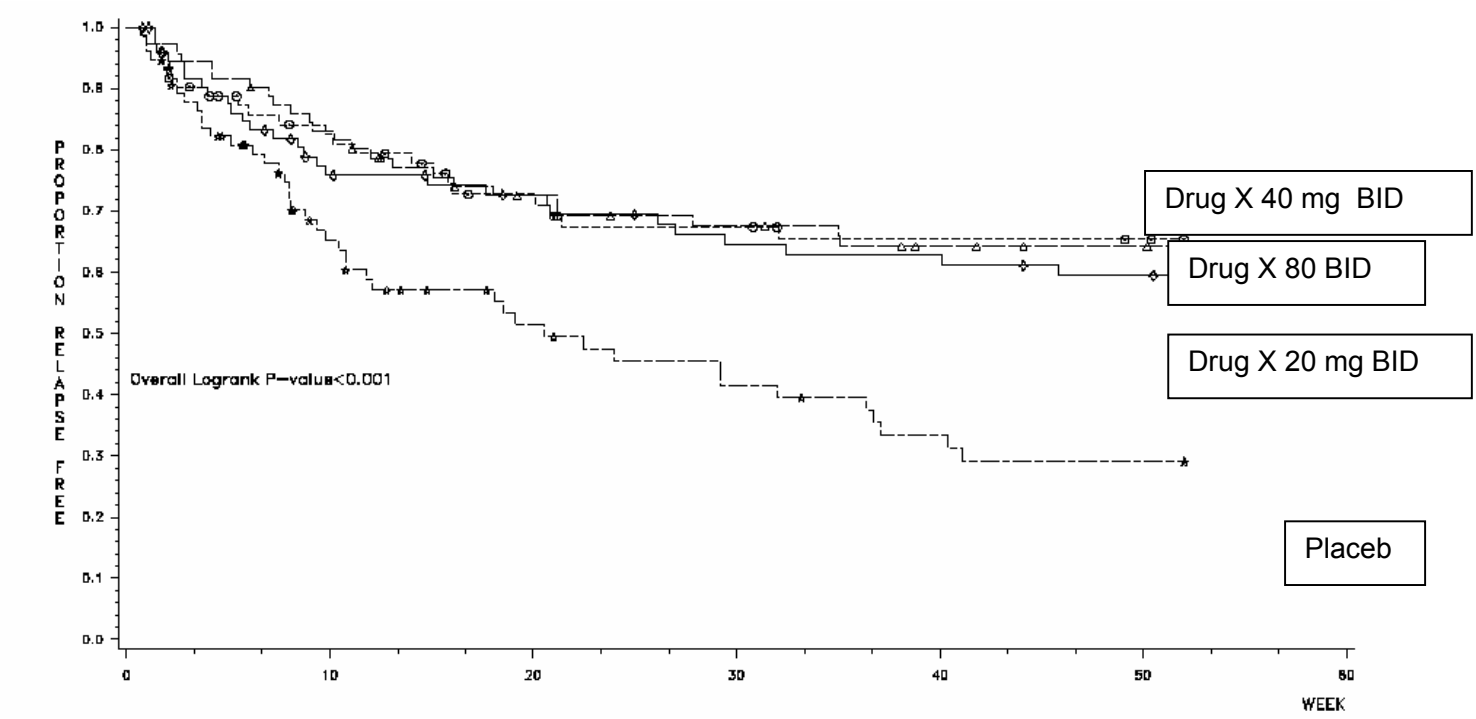


Figure 4

PROTOCOL 200: ONE YEAR MULTICENTRE DOUBLE-BLIND STUDY OF TWO FLEXIBLE DOSES OF DRUG X ONCE DAILY VERSUS PLACEBO IN RELAPSE PREVENTION FOR PATIENTS WITH CHRONIC OR SUBCHRONIC SCHIZOPHRENIA

Phase of Development: Phase III

Study Objectives: To compare the efficacy of a flexible dose regimen of Drug X once daily (80-100 mg QD) to placebo in the prevention of psychotic relapse in patients with chronic of subchronic schizophrenia.

To examine the dose-response relationship using an additional flexible dose regimen of Drug X (40-60 mg QD).

To compare the concentration-response relationship, population pharmacokinetics, safety and toleration of the two dosage regimens of Drug X to placebo when administered for up to one year.

To determine whether Drug X, when administered for up to one year at a flexible dose of 40-60 mg QD or 80-100 mg QD, is more efficacious than placebo in treating negative symptoms.

Study Design: A multicenter, double-blind, placebo-controlled, randomized study in which subjects received Drug X (40-60 mg QD or 80-100 mg QD) or placebo for up to 52 weeks. The primary efficacy outcome was time to impending psychotic relapse.

Analysis of Time to Relapse – all subjects
Product X 200

		Cum Incidence (%) of Relapse		Probability of Relapse		95% Confidence Limits			
Treatment Group	**N**	**<= 24 Wks**	**<=52 wks**	**<=24 wks**	**<= 52 wks**	**Rel Risk**	**Lower**	**Upper**	**p-value**
Product X									
40-60 mg QD	63	14 (22)	19 (30)	0.246	0.384	0.445	0.253	0.783	0.005
80-100 mg QD	62	14 (22)	23 (37)	0.234	0.412	0.553	0.325	0.941	0.029
Placebo	63	24 (38)	34 (54)	0.428	0.638				
Overall									0.008

Figure 5
Kaplan Meier Curves for Time to Relapse by Treatment Group
All Subjects
Product X Protocol 200

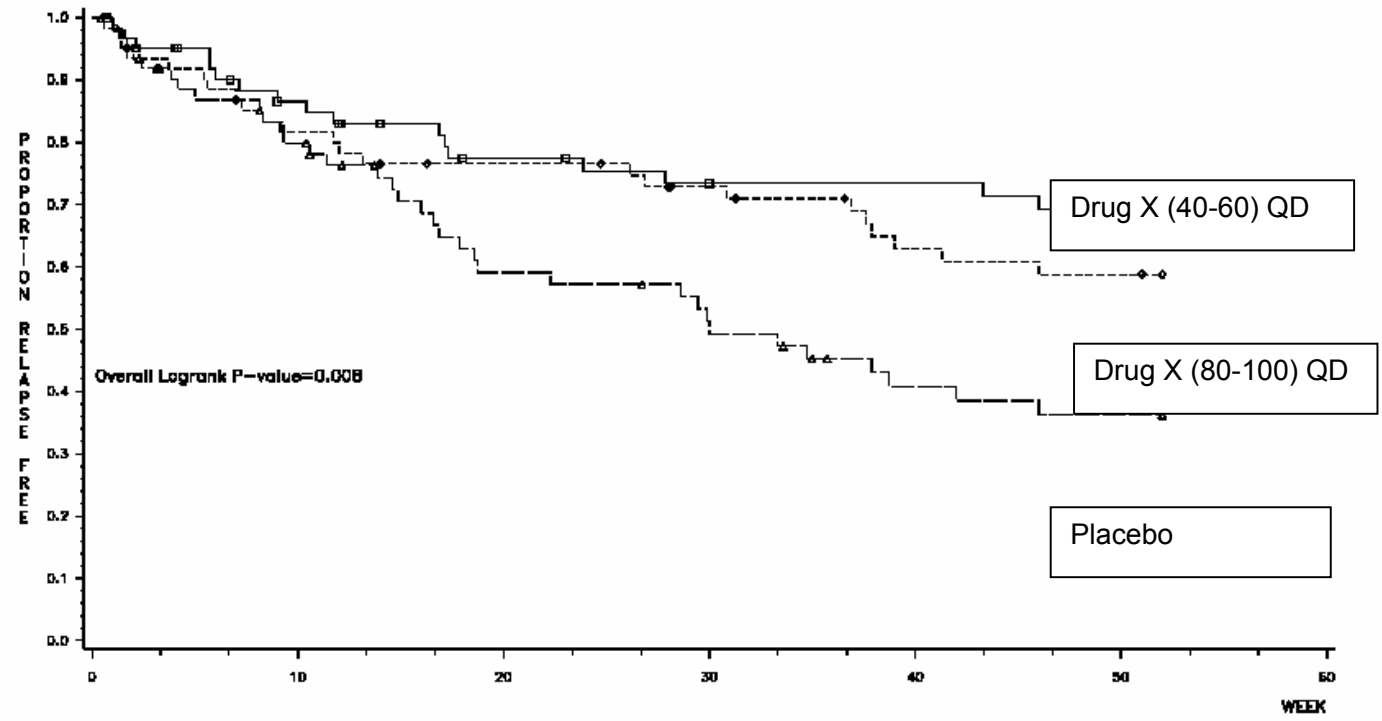


Figure 6.
Protocol 100
Incidence of Treatment – Emergent Adverse Events
(All Causalities)

	Product X 20mg BID (n=76) % of Subjects	Product X 40mg BID (n=72) % of Subjects	Product X 80mg BID (n=71) % of Subjects	Placebo (n=75) % of Subjects
Headache	3.9	6.9	9.9	5.3
Back Pain	5.3	4.2	1.4	0.0
Infection	3.9	5.6	2.8	0.0
Flu Syndrome	0.0	0.0	7.0	2.7
Accidental Injury	2.6	5.6	4.2	0.0
Hypertension	5.3	4.2	8.5	6.7
Vomiting	3.9	5.6	5.6	4.0
Diarrhea	6.6	6.9	7.0	4.0
Tooth Disorder	3.9	2.8	5.6	4.0
Weight Loss	6.6	9.7	7.0	8.0
Dyskinesia	3.9	4.2	2.8	6.7
Tremor	1.3	4.2	5.6	2.7
Anxiety	10.5	9.7	14.1	16.0
Agitation	13.2	12.5	14.1	17.3
Depression	7.9	5.6	11.3	5.3
Insomnia	30.3	30.6	46.5	32.0
Hallucinations	15.8	2.8	8.5	5.3
Manic Reaction	14.5	13.9	14.1	17.3
Personality Disorder	5.3	2.8	1.4	6.7
Libido Increased	6.6	1.4	1.4	1.3
Hostility	5.3	6.9	8.5	5.3
Extrapyramidal Syndrome	2.6	2.8	7.0	6.7
Delusions	10.5	4.2	7.0	10.7
Akathisia	9.2	8.3	11.3	5.3
Pharyngitis	3.9	5.6	2.8	1.3
Bronchitis	3.9	4.2	5.6	2.7
Respiratory Tract Infection	2.6	5.6	2.8	5.3
Rash	5.3	5.6	7.0	1.3

* Subjects with multiple occurrences of the same adverse event are counted only once (at the maximum severity) for that adverse event. Subjects with multiple occurrences of adverse events in the same body system are counted only once (at the maximum severity) for the body systems totals. Adverse events with unknown severities are classified as severe. Only adverse events occurring while on study treatment with within six days after the last day of study treatment were included in this table.

Figure 7.
Protocol 200
Incidence of Treatment-Emergent Adverse Events
(All Causalities)

	Product X 40- 60 mg QD % of Subjects	Product X 80- 100 mg QD % of Subjects	Placebo % of Subjects
Flu Syndrome	6.3	9.5	7.8
Hyperkinesia	3.2	1.6	6.3
Anxiety	7.9	14.3	23.4
Agitation	7.9	9.5	7.8
Psychosis	3.2	4.8	1.6
Insomnia	30.2	28.6	34.4
Somnolence	3.2	6.3	3.1
Hallucinations	1.6	7.9	12.5
Manic Reaction	22.2	14.3	25.0
Personality Disorder	0.0	1.6	6.3
Thinking Abnormal	3.2	1.6	6.3
Hostility	3.2	1.6	12.5
Extrapyramidal Syndrome	7.9	6.3	4.7
Delusions	4.8	14.3	18.8
Akathisia	3.2	6.3	4.7

* Subjects with multiple occurrences of the same adverse event are counted only once (at the maximum severity) for that adverse event. Subjects with multiple occurrences of adverse events in the same body system are counted only once (at the maximum severity) for the body systems totals. Adverse events with unknown severities are classified as severe. Only adverse events occurring while on study treatment with within six days after the last day of study treatment were included in this table.

Figure 8.

PROTOCOL 300: ONE HUNDRED AND NINETY SIX WEEK, DOUBLE-BLIND STUDY EVALUATING THE SAFETY AND EFFICACY OF THE TWO DOSE REGIMENS OF ORAL DRUG X (80-120 MG, QD, AND 40-80 MD, BID) AND DRUG Y (5-20 MG DAILY) IN THE MAINTENANCE TREATMENT OF OUTPATIENTS WITH SCHIZOPHRENIA OR SCHIZOAFFECTIVE DISORDER

Phase of Development: Phase III

Study Objectives: This one hundred and ninety six week, double-blind, randomized, flexible dose, parallel-group study evaluated the safety, tolerability, and efficacy of two flexible-dose regimens of Drug X (80 to 120 mg QD or 40 to 80 mg BID) and one flexible-dose regimen of Drug Y (5 to 20 mg daily) in subjects with chronic or subchronic schizophrenia or schizoaffective disorder who were believed to be treatment-responsive and in whom outpatient neuroleptic maintenance therapy was indicated.

Study Design: Following a 3 to 14-day stabilization period, subjects receive either Drug X (80 to 120 mg QD or 40 to 80 mg BID) or Drug Y (5 to 20 mg daily) for up to 196 weeks. The protocol medication replaced the subject's pre-existing neuroleptic medication with no wash out period.

Subjects were evaluated at screening, baseline (i.e., within 48 hours before the first dose of study drug), and at regular intervals throughout the study.

PANSS Total Score** – Change from Baseline by Week – All Subjects, Observed Cases

		PANSS Total Score**					
		Baseline*	Week 6	Week 40	Week 92	Week 196	Last
Drug X	QD Mean	72.45	-6.88	-12.17	-7.32	-12.81	-5.01
	Std. Dev.	17.74	13.17	17.47	22.02	20.18	21.68
	N	199	168	78	60	37	199
Drug X	BID Mean	73.68	-7.87	-16.19	-12.55	-17.37	-11.10
	Std. Dev.	18.26	14.36	17.76	22.05	21.67	23.82
	N	197	166	81	67	35	197
Drug Y	Mean	72.61	-6.98	-10.51	-9.05	-12.55	-8.57
	Std. Dev.	18.14	13.42	15.86	19.47	18.23	21.13
	N	140	118	61	41	22	140

*Baseline = last visit prior to double-blind treatment: Weeks are determined by visit designators; Last = last visit, planned or unplanned (excluding partial data for ad hoc evaluation of relapse).
**PANSS Total Score equals the sum of the 30 PANSS items.

Figure 9.
Protocol 300
PANSS Negative Subscale Score** - Change from Baseline by Week – All Subjects, Observed Cases

PANSS Negative Subscale Score**

		Baseline*	Week 6	Week 40	Week 92	Week 196	Last
Drug X	QD Mean	21.13	-2.46	-4.40	-3.43	-5.00	-2.87
	Std. Dev.	6.16	4.03	5.25	6.56	6.69	6.48
	N	199	168	78	60	37	199
Drug X	BID Mean	21.37	-3.04	-5.14	-4.49	-7.43	-4.86
	Std. Dev.	6.25	4.74	5.41	7.36	6.82	7.64
	N	197	166	81	67	35	197
Drug Y	Mean	20.75	-2.01	-3.05	-2.93	-3.59	-3.17
	Std. Dev.	6.43	4.97	5.26	6.22	5.81	6.15
	N	140	118	61	41	22	140

*B/L= last baseline visit prior to double-blind treatment; Weeks are determined by visit designators.
**Last – last visit, planned or unplanned excluding partial data for ad hoc evaluation of relapse.
***PANSS Negative Subscale Score equals the sum of the 7 PANSS Negative Subscale items.

Figure 10. Protocol 300
Quality of Life Total Score – Change from Baseline by Week – All Subjects, Observed Cases

Quality of Life Total Score**

		Baseline*	Week 40	Week 92	Week 196	Last
Drug X	QD Mean	57.5	9.8	8.32	15.19	8.92
	Std. Dev.	21.9	21.1	20.98	18.37	20.67
	N	158	77	54	36	158
Drug X	BID Mean	54.9	8.9	10.59	15.59	9.59
	Std. Dev.	20.1	20.6	22.15	21.51	23.45
	N	163	80	65	33	163
Drug Y	Mean	57.8	4.1	1.97	-0.93	1.74
	Std. Dev.	21.6	13.5	16.50	24.33	20.37
	N	120	61	38	21	120

*Baseline = last visit prior to double-blind treatment; Weeks are determined by visit designators.
**Quality of Life Total Score equals the sum of the 21 QLS items.

REFER TO INSTRUCTION MANUAL Page 28

INVESTIGATOR _____ DATE OF VISIT (month/day/year) ___/___/___

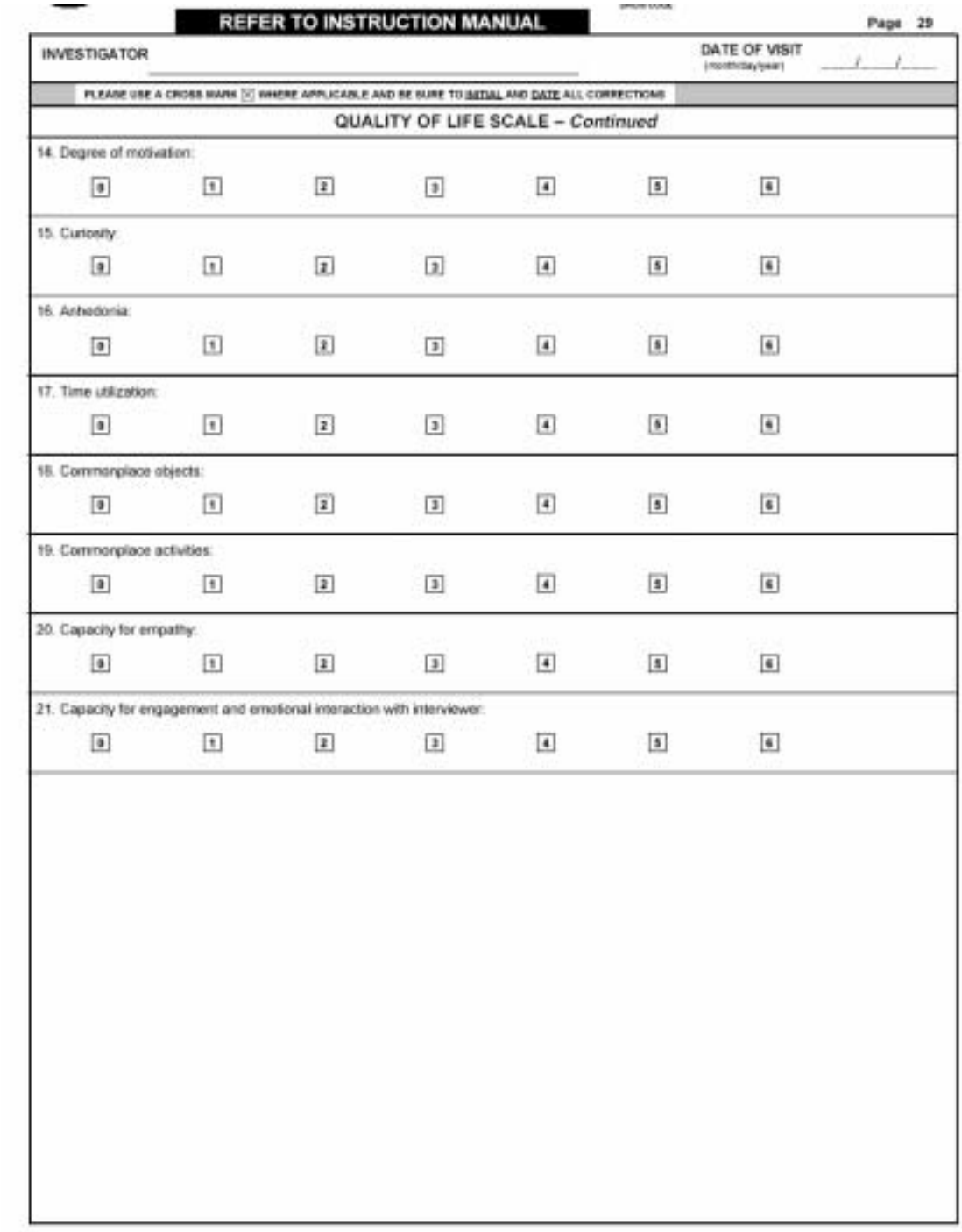
PLEASE USE A CROSS MARK ☒ WHERE APPLICABLE AND BE SURE TO INITIAL AND DATE ALL CORRECTIONS

☐ NOT DONE

QUALITY OF LIFE SCALE

1. Intimate relationships with household members:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	[9]
2. Intimate relationships:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
3. Active Acquaintances:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
4. Level of social activity:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
5. Involved social network:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
6. Social Initiatives:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
7. Social withdrawal:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
8. Sociosexual relations:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
9. Extent of occupational role functioning:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
10. Level of accomplishment:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
11. Degree of underemployment:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	
12. Satisfaction with occupational role functioning:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	[9]
13. Sense of purpose:	[0]	[1]	[2]	[3]	[4]	[5]	[6]	

Continued on next page



REFER TO INSTRUCTION MANUAL

Page 29

INVESTIGATOR _____ DATE OF VISIT (month/day/year) ___/___/___

PLEASE USE A CROSS MARK ☒ WHERE APPLICABLE AND BE SURE TO INITIAL AND DATE ALL CORRECTIONS

QUALITY OF LIFE SCALE – *Continued*

14. Degree of motivation:

[0] [1] [2] [3] [4] [5] [6]

15. Curiosity:

[0] [1] [2] [3] [4] [5] [6]

16. Anhedonia:

[0] [1] [2] [3] [4] [5] [6]

17. Time utilization:

[0] [1] [2] [3] [4] [5] [6]

18. Commonplace objects:

[0] [1] [2] [3] [4] [5] [6]

19. Commonplace activities:

[0] [1] [2] [3] [4] [5] [6]

20. Capacity for empathy:

[0] [1] [2] [3] [4] [5] [6]

21. Capacity for engagement and emotional interaction with interviewer:

[0] [1] [2] [3] [4] [5] [6]

Figure 11.
Protocol 300
Drug Attitude Inventory Total Score** - Change from Baseline by Week – All Subject, Observed Cases

		Drug Attitude Inventory Total Score**				
		Baseline*	Week 40	Week 92	Week 196	Last
Drug X	QD Mean	7.1	1.0	0.89	1.44	0.90
	Std. Dev.	2.1	2.3	2.13	1.50	2.05
	N	168	77	56	36	168
Drug X	BID Mean	7.2	0.9	0.63	0.57	0.68
	Std. Dev.	2.2	2.0	2.09	1.99	2.03
	N	167	78	64	35	167
Drug Y	Mean	7.2	0.3	0.30	0.24	0.04
	Std. Dev.	2.1	2.0	1.92	2.07	2.06
	N	120	60	40	21	120

*Baseline = last visit prior to double-blind treatment; Weeks are determined by visit designators.
**Drug Attitude Inventory Score equals the sum of the 10 DAI items.

DRUG ATTITUDE INVENTORY

The purpose of this questionnaire is to gain some understanding of how patients view the use of psychiatric medications and the nature of their experiences on these drugs. Your responses are used for research purposes only, are strictly confidential, and will in no way affect your treatment.

Read each statement below and decide whether it is true as applied to you or false as applied to you. If a statement is TRUE or MOSTLY TRUE, circle the T following the statement. If a statement is FALSE or NOT USUALLY TRUE, circle the F following the statement. If you want to change an answer, mark an X over the incorrect answer and circle the correct answer.

Please answer every question. If a statement is worded not quite the way you would express it yourself, decide whether it is mostly true or mostly false. Remember to give YOUR OWN OPINION --- there is no right or wrong answer. Do not spend too much time on any one item.

The medications referred to in the statements are psychiatric medications only.

DAI10

1	For me, the good things about medication outweigh the bad. ---------->	T F
2	I feel weird, like a 'zombie', on medication. ------>	T F
3	I take medications of my own free choice. ------>	T F
4	Medications make me feel more relaxed. ------>	T F
5	Medication makes me feel tired and sluggish. ------>	T F
6	I take medication only when I am sick. ------>	T F
7	I feel more normal on medication. ------>	T F
8	It is unnatural for my mind and body to be controlled by medications. ---------->	T F
9	My thoughts are clearer on medication. ------>	T F
10	By staying on medications I can prevent getting sick.	T F

If you have any further comments about medication or this questionnaire, please write them below or overleaf.

Figure 12.
Product X – Protocol 300
Antimuscarinic Drugs Used in Parkinsonism – Usage by Week – All Subjects, Observed Cases

			Baseline*	Week 6	Week 40	Week 92	Week 196	Last
Product QD	X	n	103	88	35	21	9	129
		(%)	46.6	41.7	39.8	42	31	58.4
		N	221	211	88	50	29	221
Product BID	X	n	108	82	24	13	8	126
		(%)	47.6	39.2	26.1	24.5	29.6	55.5
		N	227	209	92	53	27	227
Drug Y		n	73	69	39	14	4	109
		(%)	48.3	48.6	55.8	40	23.5	72.1
		N	151	142	70	35	17	151

n is the number of subjects taking antimuscarinics during the week.
N is the number of subjects participating in the study during the week.

Figure 13

Study QT Design

Study Design

Study QT was an open-label, parallel-group study in patients with schizophrenia to assess the effect of oral doses of Drug X and five other antipsychotic agents, on the QT interval. The times of ECG measurements were estimated to correspond with the mean Tmax $\pm$ 30 to 60 min for each study drug. ECGs and blood samples for pharmacokinetic analysis were obtained at baseline, during dose escalation, and at steady state (in the absence and presence of a metabolic inhibitor). All ECGs were manually read in a blinded fashion by a central reader. A correction formula derived from the population ECG data at baseline has been used to calculate QTc. Other safety assessments (adverse events, clinical laboratory tests, vital signs) were made at intervals throughout the study; efficacy was not assessed.

Benefit-Risk Assessment Model for Medicines –Developing a Structured Approach to Decision-Making, 13-14 June 2005, Washington D.C.

Figure 14

Study QT: QTc Mean Change from Baseline in Absence and Presence of Metabolic Inhibitor

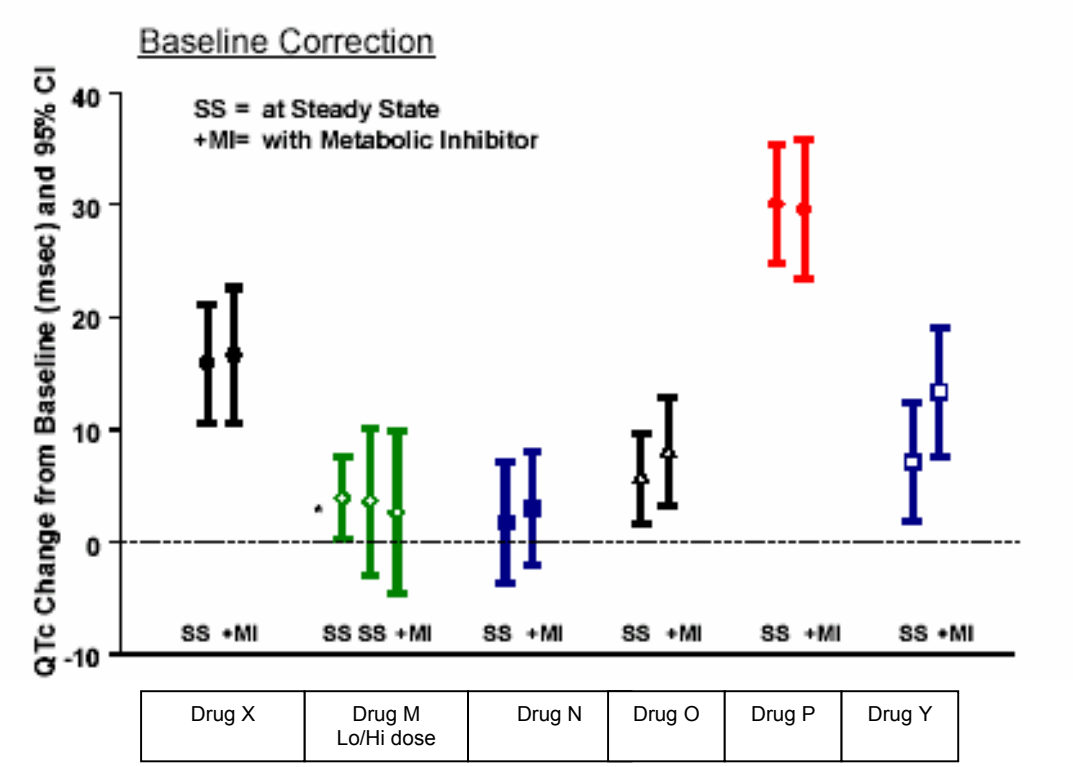
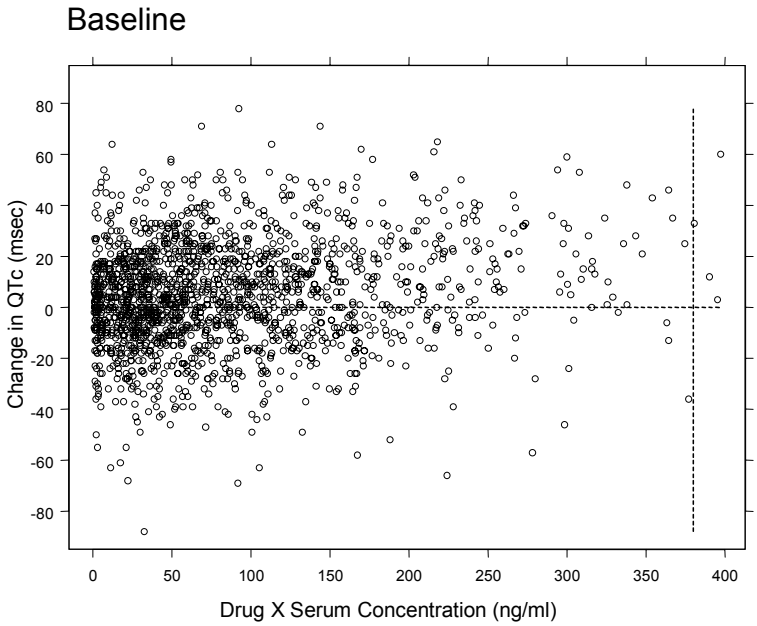


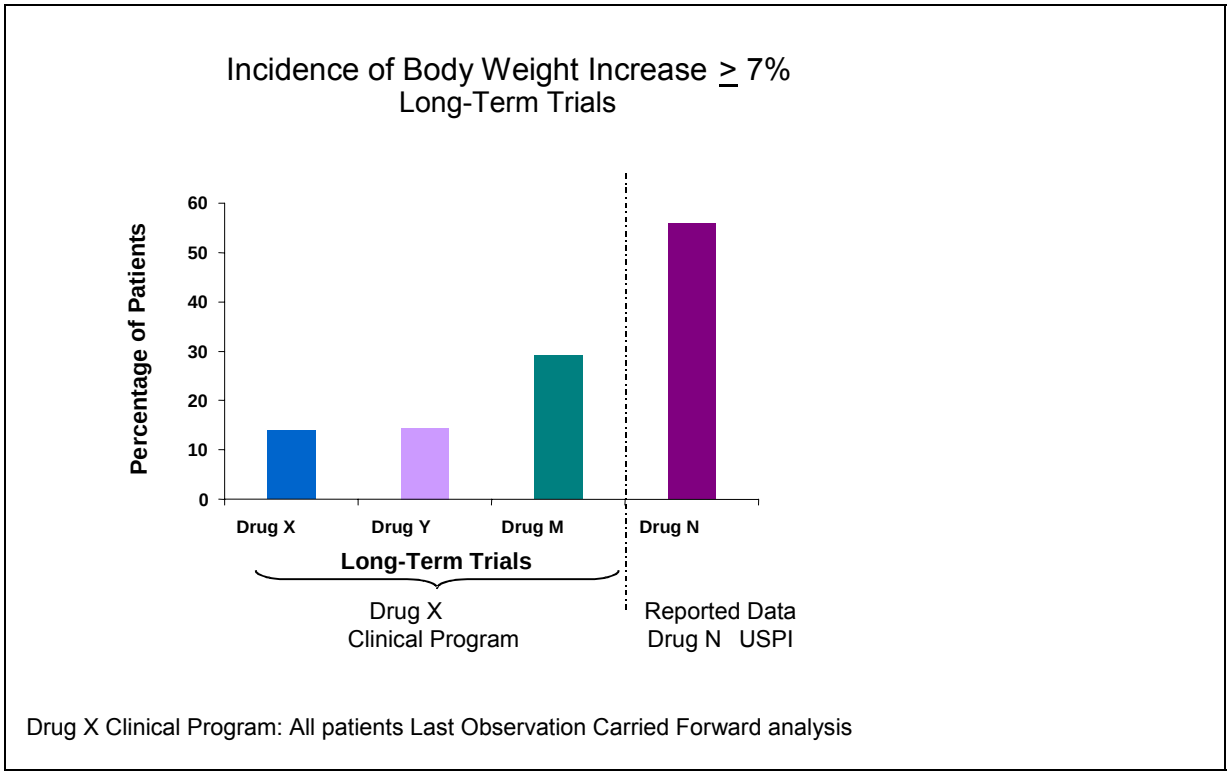
Figure 1. Mean QTc Change from Baseline in Absence and Presence of Metabolic Inhibtor; Study QT

Figure 15.
QTc Change vs. Drug X Serum Concentration
Phase 2/3 Population PK Database



QTc Change from Baseline vs Drug X Serum Concentration for Samples Collected within 1 Hour of ECG Measurement

Figure 16
Long Term Weight Gain
All Patients



Incidence of Weight Gain ≥7% of Baseline Weight; Long-Term Trials

Figure 17

Table 1. Change in Fasting Lipids; Study QT

Lipids (mg/dl)	*Drug X*	*Drug M*	*Drug N*	*Drug O*	*Drug P*	*Drug Y*
Total Cholesterol						
N	***34***	***28***	***27***	***29***	***31***	***29***
Median Baseline	***197.5***	***204.0***	***201.0***	***196.0***	***186.0***	***193.0***
Median Change	–14.5***	–3.0	4.0	5.0	21.0***	–22.0***
Median % Change	–7.5**	–1.6	2.1	2.4	13.7***	–11.5***
LDL Cholesterol						
N	33	25	26	28	29	29
Median Baseline	122.0	125.0	128.0	117.0	121.0	121.0
Median Change	***–11.0***	***9.0***	***1.5***	***–0.5***	***20.0******	***–14.0******
Median % Change	–8.5	6.5	1.1	–0.3	18.6***	–10.5***
HDL Cholesterol						
N	34	27	27	29	30	29
Median Baseline	43.5	41.0	44.0	45.0	41.0	43.0
Median Change	***0.0***	***-2.0***	***-2.0***	***-3.0***	***1.5***	***–3.0****
Median % Change	0	–4.9	–4.6	–8.6	3.0	–6.0*
Triglycerides						
N	***34***	***28***	***27***	***29***	***31***	***29***
Median Baseline	***141.0***	***158.0***	***148.0***	***124.0***	***120.0***	***118.0***
Median Change	–37.0***	–17.0	43.0***	25.0***	9.0	–18.0**
Median % Change	–28.0***	–6.7	31.0***	18.3***	7.9	–18.0**
Total Cholesterol/HDL ratio						
N	***34***	***27***	***27***	***29***	***30***	***29***
Median Baseline	***4.31***	***5.43***	***5.14***	***4.42***	***4.61***	***4.26***
Median Change	***–0.33****	***0.31***	***0.28***	***0.48****	***0.41****	***–0.22****
Median % Change	–7.5**	5.9*	5.4*	10.8**	12.4***	–7.0*

**** p<0.05, **p<0.01, ***p<0.001 Wilcoxon signed rank test on change from baseline values vs. 0 and percent change from baseline values versus 0.***

Figure 18.

Median Change in Total Cholesterol; Long-Term Active-Comparator Trials

	Total Cholesterol mg/dl								
	Drug X			*Drug Y*			*Drug M*		
	*N**	*Baselin*	*Change*	*N**	*Baselin*	*Change*	*N**	*Baselin*	*Change*
Week 28	*334*	*198.5*	*-10.0*	*68*	*203.0*	*2.5*	*81*	*202.0*	*0.0*
Week 40	*168*	*204.0*	*-12.0*	*59*	*201.0*	*1.0*	*3*	*NA*	*NA*
Week 52	*104*	*187.0*	*-10.0*	*0*	*NA*	*NA*	*51*	*208.0*	*1.0*
LOCF	*1009*	*196.0*	*-10.0*	*141*	*195.0*	*-6.0*	*134*	*206.0*	*-2.0*

Serum cholesterol measured from samples collected at random times.
**** Number of patients with baseline and on treatment cholesterol measurement in completed or ongoing open-label, active-comparator trials with duration ≥6 months up to 5 February.***
Weeks 28, 40, 52: All patients with a cholesterol measurement from Days 183-210, 267-294, and 351-378, inclusive; LOCF: Last Observation Carried Forward.
NA: not applicable due to small sample size.