OHE *Briefing*

Research
Office of Health Economics

Challenges and Opportunities for Improving Benefit-Risk Assessment of Pharmaceuticals from an Economic Perspective

James T. Cross, MS, Louis P. Garrison, PhD
University of Washington, Seattle, USA

Contents

EXECUTIVE SUMMARY

In October 2007, the Office of Health Economics (OHE) hosted a workshop in London on Benefit-Risk Assessments for Drugs. Researchers (see Box 1) presented ideas and case studies to show how certain tools and methods used in economic analysis and the decision sciences could improve the methodology that regulatory authorities such as the European Medicines Agency (EMEA) and the US Food and Drug Administration (FDA) use to evaluate drug benefit and risk. (EMEA 2007)

Four methods were presented and discussed:

1. Health Outcomes Modelling Using Quality-Adjusted Life Years (see Box 3)
2. Incremental Net Health Benefit (see Box 4)
3. Stated Benefit-Risk Preferences (see Box 5)
4. Multi-criteria Decision Analysis (see Box 6)

The first two—representing the health impact dimension of cost-effectiveness analysis-- are quite similar, comparing risks and benefits in a common metric. Health outcomes modelling can be seen as an extension or broadening of the incremental net benefit approach, which itself involves the modelling of outcomes, to include population level impacts such as the health benefits forgone in delaying an approval decision. Stated preference measurement of benefit-risk involves eliciting patient preferences among hypothetical health states. Multi-criteria decision analysis is a structured group decision-making process based on decision science.

There was agreement that each of these tools offers the potential to improve regulatory benefit-risk assessment by addressing three major issues with the current regulatory approach:

1. *Transparency:* To what extent are benefit-risk assessments and decisions clearly defined and justified for patients and clinicians, and others outside the regulatory process?
2. *Preferences:* Do or should regulators' preferences in benefit-risk trade-offs reflect those of patients or clinicians? What is the appropriate perspective that should be taken in regulatory assessments? How and should "societal preferences" be estimated?
3. *Consistency:* Should the same benefit-risk assessment method be used for all therapeutic areas given differences in the quality of evidence available?

No single tool has yet emerged as the clear choice to address these three issues. Yet, strong arguments were made at the OHE meeting for regulators to consider the preferences of patients and the community in some manner and each of the methods provides some scope for this paradigm. The workshop presentations and discussion highlighted the strengths and limitations of each. Further developmental work and research is needed to refine and test each of the methods and to identify the circumstances in which each would be used.

CURRENT DRUG BENEFIT-RISK ASSESSMENT: MOTIVATION FOR INCLUDING A HEALTH ECONOMIST'S PERSPECTIVE

When a physician prescribes a drug product and a patient decides to use it, both assume that the product is generally safe and effective for its intended use-- that, on average, the product's benefits (e.g. increases in life expectancy or quality-of-life) outweigh the product's risks (e.g. arrhythmia, liver failure). The EMEA's regulatory review process, first established in 1993 and since revised to its current form, specifies that the risk-benefit balance of products must be assessed for marketing authorisation. (Council of the European Union 1993, 2004) Product-specific assessments are in the form of European Public Assessment Reports (EPARs). (EMEA 2007) An EPAR reflects the scientific conclusion of the Committee for Medicinal Products for Human Use (CHMP), summarising the grounds for the CHMP opinion in favour of granting a marketing authorisation for a specific product. The EPAR is updated throughout the authorisation period as changes to the original terms and conditions of the authorisation. Regulations do not specify the methods for conducting risk-benefit assessment, nor do they provide guidelines as to what constitutes a positive or negative "risk-benefit balance." This regulatory scenario also exists in the US, where government regulations and guidance from the FDA do not specify how data should be weighed or quantified to form an overall risk-benefit balance.

The ability of regulators adequately to assess drug benefit and risk has been repeatedly called into question. The Council for International Organizations of Medical Sciences (CIOMS) issued a report in 1998 which declared that no defined and proven method exists to evaluate benefit and risk. (CIOMS 1998) In the decade since this report, a series of increasingly publicised product withdrawals and controversial postmarketing labelling decisions has occurred, most notably the market withdrawal of rofecoxib. This ignited controversy over how new postmarketing trial data affect a product's perceived overall benefit-risk profile and methods for synthesizing various pieces of information on benefit and risk. Like the earlier CIOMS report, the recent Institute of Medicine (IOM) report on drug safety has challenged US regulators to develop novel methods of benefit-risk assessment. (IOM 2006)

Current regulatory decisions are often not transparent because they attempt to integrate the quality and quantity of a heterogeneous body of evidence presented to demonstrate efficacy and safety. (FDA 1998; International Conference on Harmonisation 1998) These decisions have not explicitly weighed the relative value of each piece of evidence (e.g. risk estimates for different health outcomes) that inform the decision maker. The regulatory process is based upon a multidisciplinary review of the evidence submitted by drug developers and is heavily reliant on phase 3 randomised clinical trials. Statistical tests of inference, to which so much attention is paid during regulatory review, can indicate whether observed therapeutic or adverse effects seen in licensing studies are attributable to chance alone. However, these tests do not indicate whether the sum of observed positive effects outweighs a countervailing sum of harms (assuming the harms have been identified). So how do regulators then weigh all of the evidence of benefit and risk to determine a product's overall value? Despite decades of pharmaceutical regulation, there is no generally accepted, systematic method for conducting benefit-risk assessment (see Box 2).

In principle, statistical confidence in estimates of a drug's risks and benefits should also be part of a transparent benefit-risk calculation. The relative uncertainty associated with these estimates should be part of the calculation, as should the willingness of patients or clinicians to assume the risk of harm for potential therapeutic benefit. For instance, when regulators decided to add strong warnings about suicidal ideation in children and adolescents to the prescribing information of selective serotonin reuptake inhibitors, (EMEA 2005, FDA 2007) some clinicians began to ask whether a potential loss in benefit would result from decreased drug use among populations for whom benefit was still believed to outweigh risk (Valuck et al 2007). Perhaps the most depressed paediatric patients derive more benefit than harm from these drugs? It is unclear whether or to what extent regulators estimated the forgone benefit as a result of their "risk management" efforts. It is not clear how regulators weighed various perspectives on the value of these drugs (i.e. the clinician's, parents' or children's perspective). A role for basing such decisions in part upon formal evaluation of patient or physician preferences for the benefit-risk trade-off could be explored, for instance, by surveying identified subpopulations to see which are most willing to accept potential risks for potential benefit.

These regulatory scenarios would appear to provide an ideal opportunity for using some of the traditional tools and methods of health economic evaluation such as mathematical modelling and utility measurement (see Box 2). These approaches offer a systematic way of informing decision-makers about the value of a technology. They also permit exploration of a product's value from various

perspectives. Yet, in general, regulators are charged with considering only health outcomes and not costs or resource utilization. They must consider benefit-risk in the terms of clinical endpoints and do not attempt to monetize health states such as adverse events. To this extent, new methods or variations on existing methods of health economic evaluation would be needed if health economists are to help inform decision-makers in medical product regulation. This was the rationale behind the OHE meeting.

Given the limitations to the current method of assessing drug benefit-risk, the EMEA's CHMP set up a working group in May 2006. The working group described the feasibility of four different approaches for regulatory assessments. (EMEA 2007) None of these was explicitly an economic approach: the October 2007 OHE workshop was organized to discuss these approaches as well as what an economic perspective might add. The OHE workshop participants debated whether an economic perspective could help to improve the current regulatory paradigm. For example, can tools commonly used by health economists and other decision science disciplines improve the *transparency* of the benefit-risk assessment process by explicitly linking how evidence of risk and benefit was weighted in the final assessment and regulatory decision? Whereas the current approach is often accused of applying inconsistent standards for regulatory approval, might new methods bring about greater *consistency* when defining benefit and risk across a range of products? If a rare adverse event (e.g. severe rhabdomyolysis for a new cholesterol lowering drug) was reported in a clinical trial for two persons out of 1,000 taking an unapproved drug but for none among the 1,000 on placebo, how would the risk of that event be weighted in the ensuing regulatory decision? How does uncertainty around this risk differ from uncertainty around risks observed for previously approved products of the same indication? How much uncertainty in the estimate of risk versus benefit can stop a product from licensure? What are the important factors in that benefit-risk decision and how are the attributable effects of each factor weighted? Lastly, to what degree does regulator acceptance of a given degree of risk for a given degree of benefit reflect what patients (or consumers in general) are willing to accept? Health economic and decision analytic approaches can address many of these regulatory concerns regarding benefit-risk trade-offs.

Dr. Eric Abadie, Chair of the CHMP, set the stage for the discussion and debate of these issues with his opening presentation. He remarked on the CHMP's efforts to improve benefit-risk assessment methodology, *"Today the benefit-risk balance of new chemical entities is based on evaluation of extensive evidence, based on clinical efficacy and safety, but*

also, at the end of the day, on subjective judgment. This subjective judgment could in fact more or less preclude some transparency and some consistency. That is why we decided to undertake this work--in order to be more transparent and more consistent."

OVERVIEW OF NEW APPROACHES TO BENEFIT-RISK ASSESSMENT

Four approaches to aid regulatory benefit-risk assessment were discussed at the October 2007 OHE meeting.

Health Outcomes Modelling with QALYs
Modelling with quality-adjusted life years (QALYs; see Box 3) involves the construction of a disease-state model that links various health states associated with the disease under consideration (e.g. metastatic cancer, remission, death) as well as the various treatment-related adverse effects. Life expectancy adjusted for health-related quality of life (i.e. QALYs) is used to integrate all outcomes into a single metric. By virtue of constructing the model, the links among the different health states, the treatment effects considered relevant to benefit-risk, the link between surrogate and clinical outcomes, and the probabilities for developing various outcomes are all defined explicitly. Garrison and colleagues have argued for constructing these models at a population level, considering multiple subgroups and potential losses in health net-benefit due to delays in reaching a decision. (Garrison et al. 2007) Such modelling can also inform decision-makers about the value of collecting additional information to reduce uncertainty in the risk-benefit decision, by identifying which effects most influence the overall estimate. (Briggs 2006). Health outcomes modelling can be seen as an extension or broadening of the incremental net benefit approach, discussed next, to include population-level impacts, such as the health benefits forgone in delaying an approval decision.

Incremental Net Benefit
The incremental net benefit approach (see Box 4) aims to quantify the difference in net benefit for an intervention relative to a comparator. Data sources and model construction must be justified, just as with health outcomes modelling with QALYs. Incremental differences in outcomes can be quantified in terms of QALYs or in terms of clinical events. (Lynd 2007, 2004) Computer simulations of outcomes can be used to calculate probabilities of an intervention exceeding a specified threshold of incremental benefit. (Lynd 2004) This can be informative when the threshold is linked to acceptability of benefit-risk, as it can then quantify the probability of positive or

negative benefit-risk decisions. Stratum-specific estimates for health outcomes can also be incorporated in the analysis to provide subpopulation level estimates of net benefit.

Stated Preference

The stated preference approach (see Box 5) aims to address the potential discrepancy between regulators and patients or clinicians in what is considered to be an acceptable benefit-risk trade-off. Since the public entrusts a regulatory authority to make benefit-risk decisions on its behalf, the regulatory authority is assumed to act in a manner that consistent with what its constituents seek. However, some would argue that regulatory agencies may be more conservative in their licensing decisions than their constituents might like given the negative attention drawn to regulators by "incorrect" decisions. This is most evident when patient advocacy groups lobby regulators to reinstate a product (e.g. alosetron, natalizumab). The stated preference approach surveys subjects in order to elicit—hypothetically—the maximum acceptable risk and minimum acceptable benefit of a product. This approach can be particularly useful to evaluate how risk acceptance changes as a function of the severity of harm, or as a function of therapeutic effect size or baseline patient covariates such as age. (Johnson 2007)

MCDA

Multi-criteria decision analysis (see Box 6), a tool from the decision sciences, aims to define more consistently and transparently the criteria for technology evaluation. (Phillips and Costa 2007) It provides a systematic group process for identifying attributes of product risk and benefit that are deemed relevant for making a regulatory decision. The attributes are weighted by the decision-makers to provide a transparent valuation of various outcomes or product features (e.g. adherence). Because the same regulatory-product scenario can be presented to different decision-makers, MCDA has also been proposed as a way to explain why different agencies make different decisions, given the same set of data. (Walker and Cone 2004)

Discussion of these four approaches at the October 2007 OHE workshop led to the emergence of three major themes: (1) transparency of benefit-risk assessments, (2) role of preferences in assessments, and (3) the extent of consistency and impact of uncertainty in risk or benefit on decisions.

1. TRANSPARENCY

Both US and European regulatory authorities have indicated that one important way to improve the benefit-risk decision-making process is to increase its transparency. (Galson 2007, EMEA 2007)

Dr. Abadie provided a working definition of transparency, "Being transparent means to explain to the outside world how we assess benefit-risk. We regulators have been accused by some stakeholders who disagreed with our past opinions, of being not so transparent. These stakeholders did not understand why we had taken a negative or positive decision."

The proposed methodological approaches address the issue of transparency in benefit-risk assessments from different angles, by explicitly stating: (1) the criteria of risk and benefit that are relevant to the assessment, and (2) how relevant data are weighted in the decision.

With regard to establishing the relevant criteria, it is "extremely important describe all potential criteria for benefit or risk that will be used to make a licensing determination" said Dr. Abadie. Dr. Larry Phillips of the London School of Economics reiterated Dr. Abadie's point, "Whatever you consider to be a risk or a benefit, one must have a fundamental understanding of risk and benefit as they relate to the scenario at hand if you are going to balance benefit against risk."

Under conventional regulatory decision-making, an array of data will be analysed by reviewers to inform the decision. But, as Dr. Larry Lynd of the University of British Columbia noted, the conventional approach does not render this very clear to stakeholders. In the case of rofecoxib's removal from the market, "the conclusions of the expert panel were in fact that the potential risks outweighed the potential benefits but this conclusion was reached relatively subjectively; they did not explicitly quantitatively trade them off simultaneously, but certainly evaluated all the data and made a decision." Thus, it was not transparent which data were used and how they were used.

Dr. Abadie noted that multi-criteria decision analysis is an approach under discussion in Europe and that it "will define the list of benefits and the list of risks, which is one of the objectives that we should have." Dr. Phillips said that "Multi-criteria decision analysis is a methodology for appraising options on the individual, often conflicting, criteria and combining them into one overall appraisal." Component criteria are defined through consensus panels of relevant experts.

The incremental net health benefit (INHB) and health outcomes modelling approaches address the transparency of criteria informing the decision. These approaches explicitly quantify the outcomes data that are included in determining benefit and risk estimate for each intervention. That is, the summary estimates of risk and benefit from the model outputs are each composed of various model inputs of probabilities and weights for harms and benefit that should be clearly presented. The MCDA approach uses a consensus of experts to estimate "preference values" associated with various clinical outcomes. For INHB benefit-risk models, the selection of inputs usually comes from clinical trials, retrospective studies, and patient- or community-reported "preference" studies, estimating rankings for different health outcomes. It can be debated whether MCDA or INHB modelling is more subjective, as all modelling requires that assumptions be made. "A virtue of modelling", noted Dr. Louis Garrison of the University of Washington, "is that the reviewers can examine the entire model, so they can see how the surrogate outcomes are linked to the final outcomes estimates of benefits and risks." Often it is left for the modeller to decide which parameter values to use in the end, yet model inputs should remain transparent, have clinical face validity, and be defensible.

A distinction between the modelling and MCDA approaches can also be seen in how evidence informing the decision is integrated in the final decision. The MCDA approach is driven by expert panel weighting and scoring of the component criteria. In the case of health outcomes modelling, probabilities and weights (e.g. the utility of health states) are obtained from available data or specific studies. Dr. Lynd felt that "one of the key issues is that modelling does make the data explicit and it does make the preference weights explicit. If at any point anybody disagrees with any of the data that one uses, you can change the analysis: you can incorporate different data and see if that changes the outcome and it does make it explicit."

The criteria for risk and benefit in the stated-preference approach differ from the MCDA or modelling approaches. Criteria are derived from what patients perceive to be clinically relevant. However, the options presented to them in the trade-off scenario are ultimately determined by the assessor. Overall risk and benefit are not evaluated separately and then aggregated, as is done for MCDA and modelling. Rather, they are presented as a composite, a sort of "holistic approach" in which the overall clinical trade-off is described. An advantage of disaggregating before presenting an intervention's net impact on health outcomes is that, as Dr. Garrison said, "It is important to show the intermediate outcomes before showing the final calculated result because the calculation is complex

and may not be very intuitive for people." On the other hand, Dr. Reed Johnson of Research Triangle Institute indicated that the point of eliciting preferences in a "holistic" manner is to, "offer people alternatives or scenarios that replicate something that conceivably one could encounter in a real clinical setting and these alternatives consist of attributes or features of particular treatment in this case, and there can be positive and negative consequences of treatment," which may or may not necessarily be perceived by patients by their individual components but in aggregate.

The proposed methods have advantages over some of the existing approaches used to describe benefit-risk. This is most evident when trying to describe the totality of data or complexity of the clinical scenario. "The number needed to treat and the number needed to harm--to me as a clinician--is something that I like. I like it because it is easy to calculate, it speaks for itself, and it is increasingly used in the scientific literature," said Dr. Abadie. But, "one of the shortcomings is that it does not take into account the whole dossier but it is mostly adapted for one clinical trial and essentially for binary endpoints."

With regard to the stated-preferences approach, Dr. Brett Hauber of Research Triangle Institute described maximum acceptable risk and minimum acceptable benefit as "analogous to the number needed to treat and number needed to harm, except that they are preference-based rather than event-based." If this is the case, then it raises the issue that stated preference approaches may be similarly challenged in their ability to consider all data that are potentially relevant to the trade-off between benefit and risk. However, the stated preference approach offers the possibility of eliciting preferences that mimic as much as possible real medical decision scenarios, which may not be achieved with methods of assessment that examine the trade-off at the trial-level or endpoint level.

In terms of the transparency of the weighting process, each method presents its own opportunity and challenge. The weakness of the current approach was summarised by Dr. Garrison, "The experts and regulators have subjective and unobservable weights on the pieces of information that they are considering, so their underlying framework is inherently implicit."

The MCDA approach offers a potential improvement over the current approach in this respect. As Dr. Phillips pointed out, "it incorporates judgments about the impact of data, converting the measured performance of a product, in terms of benefit-risk, into what it is actually valued as. It allows for differential importance of decision criteria, and it is based on a sound theory that only assumes that decision makers wish their decisions to fit together."

However, one potential drawback of MCDA, shared by Dr. Abadie and others is that "the consensus panel assigns weights to each of the criteria that comprise benefit and risk. So, the weight of this tree [of benefit-risk information] is totally subjective." Yet as Dr. Phillips indicated, each method introduces subjectivity into the benefit-risk assessment process, and the advantage of the MCDA approach is that "it makes the subjective explicit and defendable." For instance, the modelling approaches and stated preference approach each involve assumptions about what should be included in the model and what should be asked of the patient.

This begs the larger question: if each proposed method introduces its own form of subjectivity, whether it be the inputs chosen for the model or the preference options they chose to elicit, then how do we capture the effect of this subjectivity? Can it be minimised? Dr. Phillips stated, "MCDA or not, regulatory decision-makers currently make implicit judgments about which risks and benefits are more important than others."

Modelling with QALYs uses utility weights derived from approaches such as standard gamble, time trade-off, or visual-analogue scale, which have a number of limitations. (McGregor and Caro 2006) As Dr. Johnson noted, "the standard-gamble preference-elicitation format does not reveal anything about risk aversion. Yet risk aversion is a fact of daily life. It is inherent in the way physicians think about prescribing medication and it is inherent in the way patients think about the medication prescribed for them. So it is essential to incorporate that reality in how we evaluate products," (see Box 5) Dr. Garrison agreed that risk aversion is an important consideration but argued that it can be incorporated into the health outcomes modelling approach by requiring a margin of positive expected net benefit or by explicit inclusion in the valuation of health states.

There was concern that the approaches may simply introduce more complexity to the decision-making process, and reduce the transparency that each approach seeks to provide. Dr. Andrew Briggs of the University of Glasgow, asked that "we should try not to confuse transparency with complexity. I think you can have a fairly transparent but quite complex model or the other way round… There are many ways in which we can learn about how better to present our methods and maybe also ask ourselves the difficult question about 'Can we make it simpler without becoming simplistic?'"

Dr. Phillips suggested that simplifying the regulatory benefit-risk decision should not itself be a primary goal. "Why do we expect decisions about the outcome of that comparably complex process to be so simple that you can reduce it to just a few criteria and numbers to then make a decision? I think the regulators have a complex job and they need models to help them make it simple and transparent; as simple as possible but not too simple."

2. PREFERENCES

The second major theme of the workshop was that of preferences: Whose preferences matter for benefit-risk assessment? How can we determine whether different perspectives result in different determinations of benefit-risk? In what way are these issues manifested in current and proposed approaches?

Much of this discussion centred on perspective. Dr. Phillips asked, "Whose preferences do we use? Do we use society's, the regulator's or patients' preferences?" The reason for much debate on this topic was captured succinctly by Dr. Hauber, who related differences in preferences to differences in regulatory decisions, "It is possible that regulators' preferences actually reflect societal benefit-risk preferences… However, in some cases the general population may be more risk-averse than regulators. In that situation some drugs would be approved that would not be acceptable to the general population… Likewise, it is possible that…risk tolerance is greater among the general population or among particular stakeholders and there then could be cases where people believe they are being unjustly denied access."

Regulators, as Dr. Abadie noted, are grappling with how to better reflect patient preferences. This would entail, "taking into consideration the presence of alternative therapies and involving industry in the reflection." This will present a challenge where placebo-controlled trials are conducted since these do not yield direct-comparison data that reflect the actual choices among therapeutic substitutes facing clinicians and patients.

Multi-criteria decision analysis is one approach for tackling differences in preferences. Dr Phillips stated that, unlike other proposed methods, "The purpose of MCDA, is not to get the right answer, because when there are multiple criteria there cannot be a right answer… It is to provide a structure for thinking, so that a group of people with different perspectives on the issues can use it to construct their preferences." The MCDA approach can compare and contrast how different perspectives and preferences result in different criteria, weights for those criteria, and resultant scores for benefits and risks. Variation in outcomes is a reflection of differences in perspectives and preferences.

Given the potential for different preferences, how might benefits and risks be valued? According to Dr. Phillips, "We can accommodate the differing perspectives by subjecting them to sensitivity analysis: we can also accommodate uncertainty in that way but if you want to incorporate uncertainty more formally we can also do that replacing some of these value judgments with certainty equivalents in a variety of ways that are all consistent with decision theory." Briefly, a certainty equivalent is a single-point value judged by a decision maker as equivalent in preference to the future uncertain values that might actually occur.

Dr. Garrison noted that we need to account both for differences in both health state preferences and in risk aversion among patients. One point to consider is that bad outcomes may have very low probabilities attached to them. As Dr. Garrison emphasized, "people are generally poor at weighing low probability events, and they are generally poor at predicting ex post utility for a hypothetical state. And their ex ante and their ex post ratings differ. Those of us who advocate these preference-based approaches--either QALYs or risk-risk trade-offs-- have some measurement challenges ahead."

For products such as alosetron (for irritable bowel syndrome), which was launched, withdrawn, and relaunched in certain markets, we can evaluate the effect of preferences," Dr. Lynd argued. He pointed out that "Certainly alosetron was withdrawn originally due to the concern over adverse outcomes," and suggested that "it was likely voluntarily withdrawn based on the risk preferences from more of a societal or regulatory perspective. Then it was reintroduced following the FDA review and a patient lobby where the patients basically said, 'We are willing to accept the risk, we would like to have the drug back, please.' So with the reintroduction, there was a look at alosetron from a different perspective, now more from the perspective of the patients' risk preferences as to how much risk were they willing to accept in order to potentially realize a benefit. Thus, we had somewhat of a measure of patients' revealed risk preferences." Although explicit regulatory statements declaring risk acceptance levels and the perspective taken are rare, it might seem logical that for severe or life-threatening conditions, or those for which good substitutes are not available, higher rates of risk acceptance may be tolerated for product approval, versus conditions where the opposite circumstances exist.

Through the elicitation of stated-preferences, Dr. Lynd felt that one could "use the same model with different preference weights and see how that changes the result. Wouldn't it be great if we could somehow go back and measure the regulators' risk preference when they were reviewing alosetron?"

In discussing his study examining a hypothetical treatment for Alzheimer's disease, Dr. Johnson found that "Physicians tend to be more risk-tolerant in treating elderly patients. The primary treatment goal for such patients may be to alleviate symptoms. This priority is consistent with assigning a lower relative weight to treatment-related mortality risks for patients with relatively short life expectancy than for younger patients." This argues the case for exploring the potential for heterogeneity in preferences in different populations and their effect on regulatory determinations.

There are drawbacks to quantitative methods that elicit preferences or utility. As Dr. Hauber indicated, "When we ask people to tell us about their preferences, there is a *cognitive burden* involved. We have to reconcile the need to provide enough information for subjects to make 'informed' trade-offs with the potential of providing more information than subjects can absorb and use effectively to express their actual preferences. Additionally, most people have difficulty evaluating small probabilities. Whether we are eliciting trade-offs for health-state utilities or for conjoint analysis, there is *measurement error* which… is a direct function of the complexity of the trade-off task. In essence, we want to get as much information as possible from subjects before reaching a point where we generate bad data because researchers have worn people out or made the evaluation task too complex."

Dr. Hauber continued: "The role of preference data will vary depending on the magnitude and character of the societal impact of the decision. Adding to Dr. Garrison's comment, quantitative benefit-risk methods offer opportunities to identify those patient groups for whom the net benefit is greatest." This would help the regulatory process by better defining sub-populations for the product prescribing information, rather than relying on overall population means of benefit and risk. Dr. Briggs noted that one could "make sure that the products get into those patients for whom the benefits are going to be the greatest as opposed to…pulling the drug off the market so those who might derive some benefit actually do not have any access to it."

3. CONSISTENCY AND UNCERTAINTY

Why does consistency matter in regulatory decision-making? Dr. Abadie noted that inconsistency is symptomatic of subjectivity in decision-making, which led the EMEA's CHMP to form a working group on benefit-risk assessment. Could methods or standards for benefit-risk assessment be applied consistently

irrespective of therapeutic class? The CHMP working group's review of EPARs for 33 new chemical entities revealed an inconsistency in how benefit-risk assessment was conducted. The purpose of the alternative approaches now being discussed is to improve consistency. As Dr. Johnson indicated, "making regulatory judgments involving benefit-risk trade-offs more methodical, more consistent, more transparent and more explainable to the public will help manage inevitable post-marketing surprises. When surprises occur, regulators then can say, 'These are the criteria we applied to the available data in making that decision,' instead of saying, 'I wish we hadn't made that decision'."

Consistency is intertwined with transparency and preferences. As Dr. Johnson noted, "The emergence of promising new treatments such as biologics, which also may have potential ill effects, makes difficult decisions unavoidable. Currently, such decisions are made inconsistently across therapeutic areas, regulatory institutions, and clinical settings, so there is a lack of transparency about the relationship between regulatory and therapeutic means and ends." So, what drives this inconsistency and how can it be mitigated?

A separate question is whether a level of inconsistency in benefit-risk decisions is acceptable to society. For example, it may be acceptable to have a lower risk-benefit threshold for cancer products or treatments for other life-threatening diseases, and higher thresholds for non-life-threatening conditions. Stated preference surveys, value weighting of outcomes in multi-criteria decision-analysis, and utility-based modelling could permit us to quantify the margin of benefit-risk trade-off acceptable to the public and the clinician community. Where more severely ill patients or those with fewer available remedies derive benefit from treatment, greater risk may be accepted for that benefit, as was the case with alosetron.

Uncertainty, and not only therapeutic-specific idiosyncrasies, may drive the issue of inconsistency in drug evaluation. Different evaluators may have different implicit thresholds at which they approve products. This inconsistency could arise from differences in how each regulator weighs the evidence (e.g. preferences for risk versus benefit), or how comfortable the regulator is with uncertainty surrounding certain outcomes, such as long-term cardiovascular morbidity. This could, in turn, lead to inconsistency in requests for additional phase 3 or post-marketing trials. There is a need for transparency in how evidence of benefit or risk (or conversely, uncertainty of benefit or risk) is evaluated in order to make the decision process more consistent. Thus, there are issues of consistency of process, by types of products, and across regulators or advisors with

different preferences, technical training and experience. While manufacturers and regulators will typically agree upon endpoints and treatment effect sizes needed to establish evidence of therapeutic effect, there is less agreement on how to handle uncertainty in estimates of benefit and risk.

To address this, stated preference approaches allow one to plot maximum acceptable risk and minimum acceptable benefit trade-off curves. One can then analyse the effects of underlying uncertainty if you change the nature of the trade-off, e.g. the duration of benefit or the probability of an adverse event. Furthermore, by eliciting the preferences of the patient, the regulator and the clinician, regulatory decision-makers can quantify both the magnitude and the uncertainty between these perspectives and how they might affect benefit-risk determinations. It tackles the issue of "how much are people willing to accept," a question that is usually fraught with uncertainty. MCDA approaches the issue by allowing different groups to construct their preferences around the clinical decision scenario.

Modelling approaches can address uncertainty by allowing one to explore the effect of consistency through sensitivity analyses of individual estimates (1st-order uncertainty) and of the overall model (2nd-order uncertainty). Dr. Lynd provided case studies using alosetron, enoxaparin and rofecoxib as examples. Where this approach is adopted, attention must be paid to the method of utility elicitation, since that can affect the resultant values. Both the stated preference and decision-analytic modelling approaches could address the effects of such uncertainty by incorporating preferences from different perspectives, either directly or through utility estimates derived from different subject groups (physicians, regulators, patients).

Additional information on drug risk and benefit can be garnered either through real-world use, or through formal phase 3 or post-marketing studies. The burden of collecting this information then falls either on society or the manufacturer, which can beget another kind of inconsistency. The health outcomes modelling approach allows one to specify the amount by which variance in risk or benefit estimates must be reduced to ensure that risk does not exceed benefit. Dr. Garrison indicated, "These regulatory decisions are often complicated and include an important question of whether it is the responsibility of the company to fund any of the follow-on studies, or is it society? Suppose a company develops a product that is barely over the risk-benefit bar in part because there is a high variance in the risk estimate. We might argue that they need to fund follow-on studies to reduce that variance." He noted that value of information analysis, a tool being used increasingly in

health economics, could provide a more consistent and transparent approach for determining when additional information should be collected as well as who should pay for it under a consistent regulatory regime.

All agreed that addressing uncertainty in risk-benefit analysis is important but also very challenging given the numerous sources of uncertainty and variability. First, there is the ex ante individual uncertainty of whether or not one will experience a sick health state. Second, if diagnosed with illness, there is often uncertainty associated with the disease progression timing and severity level. Third, if there are treatments available, there is uncertainty associated with the heterogeneity of treatment response. Fourth, there is uncertainty (and heterogeneity) related to individual and societal preferences, both ex ante and ex post in relation to experiencing a health detriment or ill condition.)

4. SUMMARY

Several methods were proposed to improve on current approaches to benefit-risk assessment. As Dr. Garrison noted, "It is probably not reasonable to assume that you would use the same methodology for every product." Can these methods or combination of methods replace human judgment? Clearly not, as Dr. Johnson reiterated that "Unfortunately, there is no single, quantitative, analytical solution to the benefit-risk trade-off problem. While such methods can help inform and improve the consistency of regulatory decision-making, these difficult decisions will always require judgments based on multiple considerations." We learned from the OHE workshop that the different methods had distinct strengths and weaknesses that lend themselves to being used in a complementary manner, rather than alone or as a substitute for the current process. Regulatory authorities in both Europe and North America, are currently investigating the extent to which these methods can be applied. The May 2008 Annual meeting of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) in Toronto, Canada devoted several sessions to this very topic. Dr. Lynd summarised the vision held by many, "I think we are going to be working collaboratively to develop a collective group of tools that might be applied in different situations."

ABOUT THE AUTHORS

James Cross is a PhD candidate in the Pharmaceutical Outcomes Research and Policy Program at the University of Washington (Seattle, USA). His research concerns policy analysis and modelling for regulatory benefit-risk decision-making. He was a fellow at the Georgetown University Center for Drug Development Science and subsequently a project manager at the Food and Drug Administration in Washington DC.

Louis Garrison is Professor of Pharmacy and Associate Director of the Pharmaceutical Outcomes Research and Policy Program at the University of Washington. He currently serves as a director for the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). He was formerly the Vice President and Head of Global Health Economics and Strategic Pricing at Roche Pharmaceuticals.

Box 1
OHE Workshop Contributors:

Eric Abadie, MD, EMBA	European Medicines Agency
Andrew Briggs, MSc, DPhil	University of Glasgow
Louis Garrison, PhD	University of Washington
A. Brett Hauber, PhD	Research Triangle Institute
F. Reed Johnson, PhD	Research Triangle Institute
Larry Lynd, PhD	University of British Columbia
Lawrence Phillips, PhD	London School of Economics

Box 2
Benefit-Risk Assessment

Definition of benefit-risk assessment:
Per the EMEA Working Group on Benefit-Risk Assessment, this is a regulatory process of evaluating the balance of "observed benefits and harms, as well as the uncertainties and risks" associated with a particular product. The EMEA adds that "There are no standard quantitative methods to be recommended for evaluating the balance of benefits and risks. Generally, the evaluation of the balance relies on balancing as objectively as possible benefits and harms, each consisting of several different events of different importance and estimated with variable precision. The estimation of the balance is often not precise and large approximations are commonly used."

Definition of risk:
"Risk" in the regulatory context describes an effect that is harmful to the patient's or public's health and which can relate to the safety, efficacy or quality of a product. Risks are frequently thought of as product-related adverse events, which can be serious (e.g. causing hospitalization) or unexpected (not previously observed). Benefit consists of all the disease preventive, mitigating or therapeutic effects.

Current methods of benefit-risk assessment
The EMEA and FDA currently evaluate the benefit-risk balance through a multi-disciplinary scientific review process for assessing evidence of drug safety and efficacy, as well as other sources of safety and efficacy evidence that stem from manufacturing, nonclinical (i.e. animal) and pharmacokinetic data. From these evaluations, regulators then decide whether to authorize an investigational product for commercial use. No guidance or directive currently exists on how regulators might integrate a product's benefits versus risk for decision-making purposes.

Benefit-risk assessment versus cost-benefit analysis and cost-effectiveness analysis:
Cost-benefit analysis and cost-utility analysis are two forms of health technology evaluation used by health economists to value a technology, using monetized outcomes and utility as measures. Regulatory health agencies are charged with interpreting the evidence of a product's effects (health risks and health benefits) to decide whether to permit product commercialisation. The methods of evaluation used in health economics and other decision sciences offer a structured means to evaluate a trade-off that a decision maker faces. Yet, in contrast to economic evaluations, regulatory evaluations are restricted to scientific data which are not monetized and which do not consider non-scientific issues such as resource utilization, which are common to economic evaluations. Despite the inherent differences between the goals and current methods used in health economic evaluation versus regulatory review, both involve making and optimizing a trade-off. Thus, the models and tools used in economics and other decision sciences could be helpful in making these regulatory decisions more systematic, structured and transparent.

Box 3

Health Outcomes Modelling Using Quality-Adjusted Life-Years

The models used in health outcomes research often use the unitary metric of the quality-adjusted life-year (QALY) to compare nonmonetary outcomes across different types of interventions. The QALY represents an adjustment to length of life for the quality of life experienced. This measure can be easily adapted to benefit-risk assessment by separating the outcomes into expected health improvements with positive QALYs (benefits) and adverse health impacts with negative QALYs (risks) to yield an incremental net health benefit comparing two interventions. – L. Garrison

Incremental net health benefit (INHB) of new Drug 2 versus conventional Drug 1 can be expressed as:

$$INHB = (E2–E1) – (R2–R1)$$

where effectiveness (E) is measured in QALYs and risk (R) can also be measured in QALYs.

Strengths
- Can compare benefit versus risk quantitatively in terms of aggregated QALYs for a population, thereby assigning a weight to each outcome using utility.
- Explicitly considers uncertainty and heterogeneity of preferences across individuals.
- Considers the costs and benefits of gathering additional information (that is, the value of information) where regulatory decisions are delayed or post-licensing studies are mandated.
- Provides a model structure with quantitative parameters that an advisory committee could explore in its deliberations using sensitivity analyses.

Current Limitations
- Utility does not usually explicitly capture patient risk aversion.
- QALYs less well understood as a decision-making "yardstick" by regulators, physicians and patients.

Box 4

Incremental Net Benefit (INB) for Quantitative Benefit-Risk Assessment

Is this a tool that can help us evaluate whether the benefits outweigh the risks from a regulatory perspective or a clinical perspective? I think one of the key issues is it does make the data explicit and it does make the preference weights explicit and if at any point anybody disagrees with any of the data that one uses in one of these analyses you can change the analysis, you can incorporate different data and see if that changes the outcomes and it does make it explicit. – L. Lynd

$$INB = (E_A - E_S) - (H_A - H_S)$$

where E = effectiveness, H = harm, A = alternative product, S = standard/comparator

Expected INB $= \sum$expected treatment benefit - $\sum$expected treatment harm

Strengths
- Utilises all available epidemiological evidence.
- Can incorporate multiple outcomes that might affect decision.
- Can perform sensitivity analyses on effect of uncertainty in model inputs, including risk and benefit.
- Can incorporate alternative patient-preference weights, including conventional health-state utility weights and generalised preference weights.
- Can produce stratified estimates of benefit-risk (see figure below).

Current Limitations
- Results are model-derived, and therefore subject to assumptions and to model validity.
- Net benefit may not be easily translatable to end-users (clinicians, patients).
- Weights for outcomes subject to method of ascertainment.
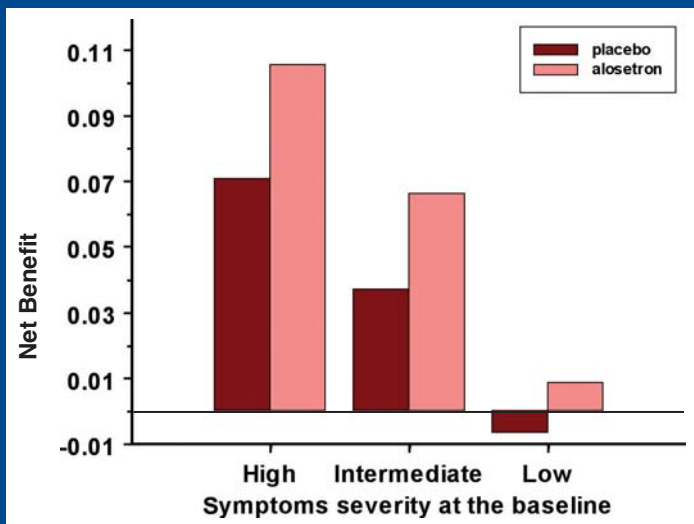
Incremental net benefit of alosetron for irritable bowel syndrome, stratified by baseline disease severity. Treatment benefit is positively associated with baseline severity. Thus INB can also guide regulatory decisions at a subpopulation level of benefit-risk where subpopulation specific determinants of heterogeneity are known.

Box 5
Stated-Preference Methods for Benefit-Risk Assessment
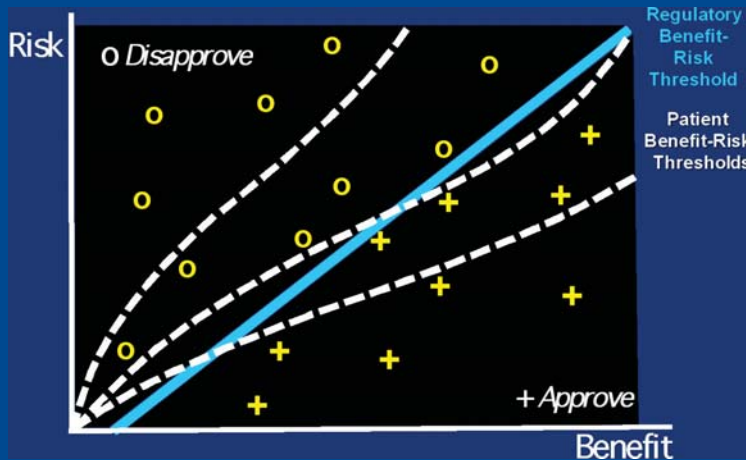
Risk-benefit functions indicate that population preferences (white lines) for benefit-risk trade-offs change nonlinearly, based on the magnitude and severity of risks and benefits conferred by the product. Preferences elicited from multiple populations or subgroups can be compared to a regulator's valuation of the trade-off (blue line), to ascertain when and for whom the two functions might lead to discordant valuations of overall benefit-risk.

Quantifying Patients' Benefit-Risk Threshold
• Preferences among alternatives depend on the relative importance of attributes.
• Hypothetical alternatives consist of combinations of attributes.
• Subjects state preferences in series of choices involving hypothetical alternatives.
• Pattern of choices identifies willingness to accept trade-offs.
• Statistical model identifies implicit preference weights.

Strengths
• Satisfies the regulatory need to understand patient preferences before the health outcomes of a product are accurately identified.
• Does not require data on observed outcomes. Preferences, not clinical evidence, are assessed.
• Realistic incorporation of nonlinear preferences:
   o A given change in benefit is valued differently depending on where it occurs on the severity scale.
   o Conventional health utility is assumed to be a linear function.
   o Risk aversion requires non-linear preferences.
• Provides estimates of maximum acceptable risk, minimum acceptable benefit, maximum acceptable number needed to treat, minimum acceptable number needed to harm, net benefit, net safety margin.

Current Limitations
• *Hypothetical bias*: patients may choose differently among real treatment alternatives.
• *Measurement error*: data quality may be poor if trade-off tasks are too difficult.
• *Innumeracy*: people have a poor understanding of small probabilities.

Box 6
Multi-Criteria Decision Analysis

This is a simplification of a part of something that the regulators need to do. It does not give you the answer as to what to do but it can clarify thinking to the extent that you will now find it easier as a group of regulators to agree amongst yourselves as to what to do. It is a constructive model, it is helping people to be clear about their preferences and to construct preferences that they feel will be reasonably robust, can be reported in the public, can be made available. – L. Phillips

Steps to Conducting MCDA
- Identify and organise a list of relevant benefit and risk criteria for determining the benefit-risk profile.
- Score the options on each criterion, using numerical values between two reference points and either a fixed, but not necessarily linear, scale or a relative preference scale.
- Assign weights to each criterion to reflect their relative importance in the decision.
- Multiply the options score by the weight for each criterion; sum for both benefits and risks.
- Examine the result, compare the total scores of benefits and risks.
- Use sensitivity analyses to explore the effects on the overall results of imprecision and of differences of opinion about the scores and weights.

Strengths
- Can be used across disease states, treatments and populations.
- Incorporates new data from various sources. Able to incorporate uncertainty.
- Accommodates multiple risks and benefits and various comparators. Characterises objective risks (e.g. mortality) as well as subjective benefits (e.g. health-related quality of life).
- Considers preferences of decision-makers (subjective perceptions of weights) by applying protocols.

Current Limitations
- Benefit-risk profiles are snapshots in time: new evidence may demand revised assessments.
- Further efforts are required to develop appropriate scales for the different benefit and risk attributes.
- Resource-consuming to develop a MCDA model.
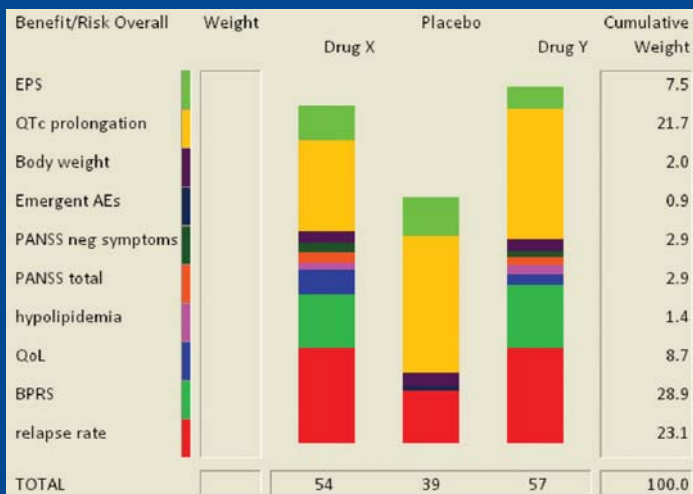
Overall preferences for benefits and risks for a new antipsychotic drug X compared to existing drug Y and placebo. The left column lists criteria: the first four are risks, the next six are benefits. Segments of the stacked bar graphs show the contributions of each criterion to the total weighted scores (bottom row). Weighted preferences are displayed: longer segments for risks signify lower risks; longer segments for benefits signify higher benefits. Judged relative weights of the criteria (right column) indicate the relative importance of the ranges of the criterion scales on which the three options were scored. (Walker et al. 2006)

REFERENCES

A Briggs, M Sculpher, K Claxton. Decision Modeling for Health Economic Evaluation. Oxford University Press, New York. 2006.

Council for International Organizations of Medical Sciences IV Working Group. Benefit-risk balance for marketed drugs: evaluating safety signals. World Health Organisation. 1998, Geneva.

European Council Regulation (EEC) No 2309/93. Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Agency for the Evaluation of Medicinal Products. 22 July 1993.

European Council Regulation (EC) No 726/2004 . Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency. 31 March 2004.

European Medicines Agency. European Medicines Agency finalises review of antidepressants in children and adolescents. EMEA/CHMP/128918/2005corr. 25 April 2005, London UK.

European Medicines Agency. Report of the CHMP Working Group on Benefit-Risk Assessment Models and Methods. EMEA/CHMP/15404/2007. 19 January 2007, London UK.

European Medicines Agency. EPARs for authorised medicinal products for human use. London. Accessed March 24, 2007 http://www.emea.europa.eu/htms/human/epar/epa rintro.htm

European Parliament and Council of the European Union. Parliament Directive No. 2001/83/EC. The Community code relating to medicinal products for human use. 6 November 2001.

FDA. US Food and Drug Administration. Guidance for Industry: Providing clinical evidence of effectiveness. May 1998, Rockville, USA. www.fda.gov/CDER/GUIDANCE/1397fnl.pdf

FDA. FDA proposes new warnings about suicidal thinking, behavior in young adults who take antidepressant medications. May 2007 Rockville, USA. Accessed: February 24, 2007. http://www.fda.gov/bbs/topics/NEWS/2007/NEW0 1624.html

Galson S. Reducing the risks and increasing the benefits of marketed drugs—FDA response to the IOM report. Presentation to the American College of Preventative Medicine, 24 February 2007. www.fda.gov/cder/present/galson/2007/ACPM022 407FINAL.pdf

Garrison L. A structured health outcomes approach to the pre-marketing quantitative risk-benefit modeling of new pharmaceuticals. Presented at the Benefit-Risk Assessments for Drugs Workshop. Office of Health Economics. 24 October 2007. London, UK.

Garrison LP, Towse A, Bresnahan, B. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff, 2007. 26(3):684-95.

Gold MR, Siegel JE, et al. Cost-effectiveness in health and medicine. Oxford University Press 1996. New York.

Institute of Medicine. Drug safety report. September 2006, Washington DC.

International Conference on Harmonisation. E9-Statistical principles for clinical trials. September 1998. http://www.fda.gov/cder/guidance/ICH_E9-fnl.PDF

Johnson FR, Fillit H, et al. Alzheimer's disease progression healthy-year equivalents: Stated risk-benefit trade-off preferences. International Conference on Prevention of Dementia, Washington, DC. June 2007. http://www.rtihs.org/request/index.cfm?fuseaction= display&PID=9517

Johnson FR, Hauber AB. Quantitative risk-benefit analysis for regulatory decision making: Alternatives and challenges. Presented at the Benefit-Risk Assessments for Drugs Workshop. Office of Health Economics. 24 October 2007. London, UK.

Johnson FR, Ozdemir S. et al. Women's willingness to accept perceived risks for vasomotor symptom relief. J Womens Health, 2007. 16(7):1028-40.

Lynd LD. An Example of incremental net benefit analysis using Lotronex, in Drug Information Association 43rd Annual Meeting. 2007. Atlanta, USA.

Lynd LD. Using incremental net benefit for quantitative benefit-risk analysis – Case studies of Vioxx and Lotronex. Presented at the Benefit-Risk Assessments for Drugs Workshop. Office of Health Economics. 24 October 2007. London, UK.

Lynd LD, O'Brien BJ. Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. J Clin Epidemiol, 2004. 57(8):795-803.

McGregor M, Caro JJ. QALYs: are they helpful to decision makers? Pharmacoeconomics, 2006. 24(10):947-52.

National Institute for Clinical Excellence. Guide to the methods of technology appraisal. April 2004. London.

Phillips L. Multi-criteria decision analysis. Presented at the Benefit-Risk Assessments for Drugs Workshop. Office of Health Economics. 24 October 2007. London, UK.

Phillips LD, Costa CAB. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. Ann Operations Res, 2007. 154(1):51-68.

Valuck RJ, Libby AM, et al. Spillover effects on treatment of adult depression in primary care after FDA advisory on risk of pediatric suicidality with selective serotonin reuptake inhibitors. Am J Psychiatry 2007. 164(8):1198-205.

Walker S, Cone M. Benefit-Risk Assessment: Summary Report on the Workshop on the Development of a Model for Benefit-Risk Assessment of Medicines Based on Multi-Criteria Decision Analysis. Centre for Medicines Research International, 2004. Surrey, UK.

Walker, S., Phillips, LD, et al. Benefit-Risk Assessment Model for Medicines: Developing a Structured Approach to Decision Making. Epsom: Centre for Medicines Research International, 2006. Surrey, UK.

As with all OHE publications, this briefing was peer reviewed by its Editorial Board and by other experts in the field and is intended to be a contribution to research and to public policy making. It does not represent the views of the OHE or of its funding body the ABPI.

About the Office of Health Economics

The Office of Health Economics was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and by revenue from sales of its publications, consulting services and commissioned research.

The research and editorial independence of the OHE is ensured by its Policy Board. This OHE publication has been peer reviewed by members of its Editorial Board. Further information about the OHE, including details of these boards, can be found at **www.ohe.org**.

Price: £7.50

© Office of Health Economics