MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1626

Published online 15 June 2014 in Wiley Online Library

Subgroup analyses in cost-effectiveness analyses to support health technology assessments

Christine Fletcher,[a]* Christy Chuang-Stein,[b] Marie-Ange Paget,[c] Carol Reid,[d] and Neil Hawkins[e]

'Success' in drug development is bringing to patients a new medicine that has an acceptable benefit–risk profile and that is also cost-effective. Cost-effectiveness means that the incremental clinical benefit is deemed worth paying for by a healthcare system, and it has an important role in enabling manufacturers to obtain new medicines to patients as soon as possible following regulatory approval. Subgroup analyses are increasingly being utilised by decision-makers in the determination of the cost-effectiveness of new medicines when making recommendations. This paper highlights the statistical considerations when using subgroup analyses to support cost-effectiveness for a health technology assessment. The key principles recommended for subgroup analyses supporting clinical effectiveness published by Paget *et al*. are evaluated with respect to subgroup analyses supporting cost-effectiveness. A health technology assessment case study is included to highlight the importance of subgroup analyses when incorporated into cost-effectiveness analyses. In summary, we recommend planning subgroup analyses for cost-effectiveness analyses early in the drug development process and adhering to good statistical principles when using subgroup analyses in this context. In particular, we consider it important to provide transparency in how subgroups are defined, be able to demonstrate the robustness of the subgroup results and be able to quantify the uncertainty in the subgroup analyses of cost-effectiveness. **Copyright © 2014 John Wiley & Sons, Ltd.**

Keywords: subgroups; cost-effectiveness; health technology assessment; uncertainty

1. INTRODUCTION

Health technology assessment (HTA) is a form of policy research that examines the short-term and long-term social consequences (e.g. societal, clinical, economic, ethical and legal) of the application of technology [1]. The goal is to provide payers with information on technology alternatives. Because of the uncertainties regarding the outcome of health care and the absence of true competitive markets for many healthcare technologies [2], many HTA bodies use cost-effectiveness analysis (CEA) as a key component of technology assessments of healthcare products and services. CEA is a tool used to increase allocative efficiency, where technologies are identified which the healthcare society feels it needs and are in great demand. It seeks to identify those healthcare technologies that, if funded, would maximise total population health. This is achieved by estimating differences in costs and effects between technologies and choosing those with the lowest opportunity costs, that is, those that, if funded, will provide more health benefit than those that would no longer be funded. In addition to allocative efficiency, the decision process needs to consider other issues, such as diversity, to ensure that funding decisions reflect societal preferences.

Subgroup analyses that assess cost-effectiveness both within and between groups of patients are being used by manufacturers and HTA authorities to understand key drivers of cost-effectiveness. The magnitude of clinical effectiveness in different potential subgroups of interest can be compared with different WTP thresholds. These analyses help to identify what components of costs and effectiveness significantly influence the overall cost-effectiveness decisions. Where the choice of specific subgroups is driven more by cost considerations, a manufacturer needs to provide HTA authorities with a clear definition and justification for the subgroups and articulate the uncertainty and extent of heterogeneity across the subgroups to enable appropriate interpretation and final reimbursement decision(s). A recent systematic review conducted by Ramaekers *et al*. [3] of national pharmacoeconomic guidance on how to deal with patient heterogeneity in economic evaluation noted that the majority of guidance documents provide recommendations on how to assess patient heterogeneity. Approximately a third of guidance documents refer to the use of subgroup analyses; however, a lack of consensus was concluded on which specific methods are most appropriate.

In Paget *et al*. [4], we presented the use of subgroup analyses for clinical effectiveness to support an HTA. We summarised the

[a] *Amgen Ltd, Cambridge, UK*

[b] *Pfizer Inc., Kalamazoo, MI, USA*

[c] *Eli Lilly and Company, Contrexeville, France*

[d] *Roche, Burgess Hill, UK*

[e] *London School of Hygiene & Tropical Medicine, London, UK*

**Correspondence to: Christine Fletcher, Amgen Ltd, 240 Cambridge Science Park, Milton Road, Cambridge, UK, CB4 0WD.*
E-mail: fletcher@amgen.com

Copyright © 2014 John Wiley & Sons, Ltd.

requirements by HTA authorities on the use of subgroup analyses in HTA, we presented good statistical and scientific principles that can be applied in such analyses, we discussed recent HTA case studies where subgroup analyses had been used to support a reimbursement decision and we recommended how statisticians working in the pharmaceutical industry could influence HTA strategy and the use of subgroup analyses to support reimbursement dossiers. Indeed, some HTA authorities, for example, Germany [5], now have specific requirements to explore subgroups to assess clinical effectiveness as part of their HTA process. In this follow-on paper, we focus on the statistical considerations relating to the use of subgroup analyses in cost-effectiveness analyses to support an HTA (Section 2), and we review the applicability and relevance of the statistical and scientific guiding principles discussed in Paget *et al*. (2011) from a cost-effectiveness perspective (Section 3). The review of a published HTA case study discusses the challenges of incorporating subgroup analyses to support CEA and how the evidence was viewed by the decision-makers (Section 4). We conclude with further discussion of key aspects and provide recommendations (Section 5).

2. STATISTICAL CONSIDERATIONS IN COST-EFFECTIVENESS ANALYSES

2.1. A brief overview of the role of subgroups in cost-effectiveness analysis

Where there are identifiable subgroups of patients that differ in terms of costs or effects, population health is in principle maximised by choosing the treatment with the lowest opportunity cost within each subgroup. However, we also need to consider whether there is a 'cost' in identifying subgroups. For example, if the factors that define a subgroup are known at the time of treatment choice without incurring additional cost, we can simply estimate cost-effectiveness and choose the optimal treatment within each subgroup. An example of the latter is subgroups defined by previous therapies, as illustrated by the case study in Section 4. On the other hand, if the factors are 'costly' to evaluate, for example, when a pharmacogenetic screening test is required to identify subgroups, we would need to compare 'test and treat' and 'treat without test' strategies [6]. In this case, additional time or costs related to the tests would be included in the cost-effectiveness assessment. In addition to cost, there may also be other important considerations, for example, ethical issues concerning genetic testing.

Irrespective of whether or not a treatment is cost-effective on average over the entire population, Claxton [7] argued that cost-effectiveness should always account for subgroups. In practice, there is often uncertainty as to whether the differences between subgroups represent true differences or random variation. This can lead to decision-makers and other stakeholders disagreeing on whether subgroups should be taken into account in CEA. These disagreements can occur in both directions: Sometimes, decision-makers will not accept evidence that a technology is cost-effective within a subgroup where it is not cost-effective across the whole population; in other cases, some stakeholders will dispute the restriction of the use of technology to a specific subgroup where its use is deemed cost-effective over the whole target population. Also, the generalisability to clinical practice of cost-effectiveness results estimated for particular subgroups in clinical trials depends on the ability to identify those subgroups appropriately using standard procedures within the healthcare system. The case study in Section 4 provides an example where this was not deemed to be possible. Furthermore, applicability of the results to the healthcare system of interest will depend on where the data for the cost-effectiveness assessment was obtained.

It is important to consider some of the criteria suggested to reduce the risk of erroneously identifying subgroups within HTA. These criteria include the prespecification of a (hopefully small) number of subgroups, prespecification of the expected direction of effect and the statistical significance of the interaction effect when assessing effectiveness. However, the number and nature of the prespecified subgroups, and the number of subjects included in a trial, are often beyond the control of the decision-makers who are interested in the cost-effectiveness of treatments. This may lead to decision-makers being reluctant to be rigidly bound to apply these criteria that focus on prespecification.

Cost-effectiveness analyses can also become unwieldy as the number of potential subgrouping factors becomes large, especially when additional sensitivity analyses are required. The increase in the number of potential subgrouping factors will also lead to an increase in uncertainty around the estimates of cost-effectiveness for individual subgroups, as more 'main effects' and 'interaction terms' are included in the model. Subgroups based on clinical or biological plausibility will have more credibility and should be favoured over subgroups identified by other means.When there are a number of potential subgrouping factors identified a priori, we may need to consider these factors in combination in a CEA. The techniques of multivariate statistical analysis, such as PCE and factor analysis or multidimensional scaling may make the consideration of multiple predictive factors more tractable by grouping factors or patients together. Subgroups based on clinical or biological plausibility will have more credibility and should be favoured over subgroups identified by other means.

We should also note that subgroups do not need to be discrete. The factor defining a subgroup might be continuous, in which case we will be interested in identifying the threshold at which a treatment becomes cost-effective (Figure 1).

In Figure 1, the subgrouping factor is assumed to be a linear predictor of treatment response for one of the treatments. The top graph in Figure 1 shows the population distribution of treatment response as predicted by the subgrouping factor. The second graph shows the proportion of patients above a given response threshold. As the response threshold decreases, the proportion of patients above the threshold increases. The third graph shows the incremental costs and effects of the treatment as a function of the response threshold. As the response threshold decreases, the total incremental costs rise directly in proportion to the size of the population who receive the treatment (assuming that all patients cost the same to treat). However, as the threshold decreases, the incremental effects initially rise faster than the incremental costs as we are initially only treating patients who experience a better than average response. This results in the incremental cost-effectiveness ratio (ICER), the ratio of the change in costs to incremental benefits, increasing as the threshold decreases until the ICER reaches the population average when all patients are being treated.

2.1.1. Prognostic versus predictive factors. Subgroup effects may be prognostic or predictive (Figure 2). If the average treatment effect (e.g. the difference in mean response between treatment

Copyright © 2014 John Wiley & Sons, Ltd.

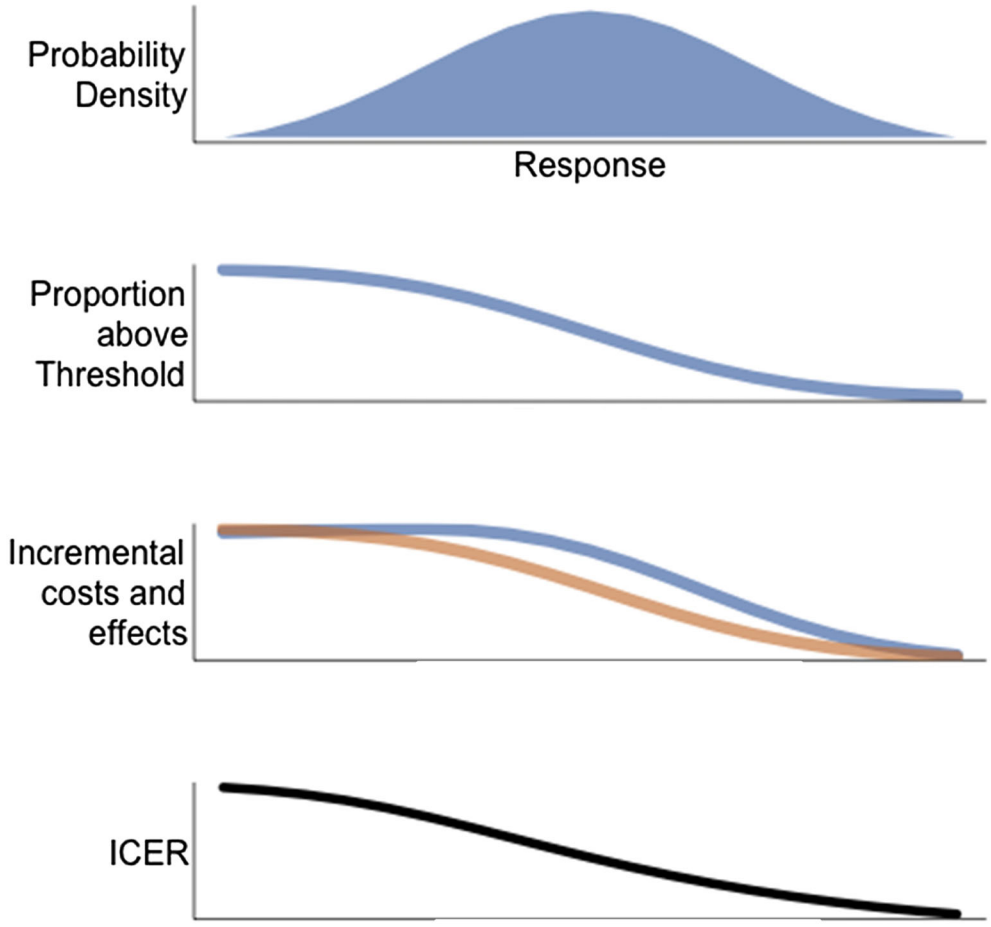

Figure 1. Subgrouping by a continuous factor.

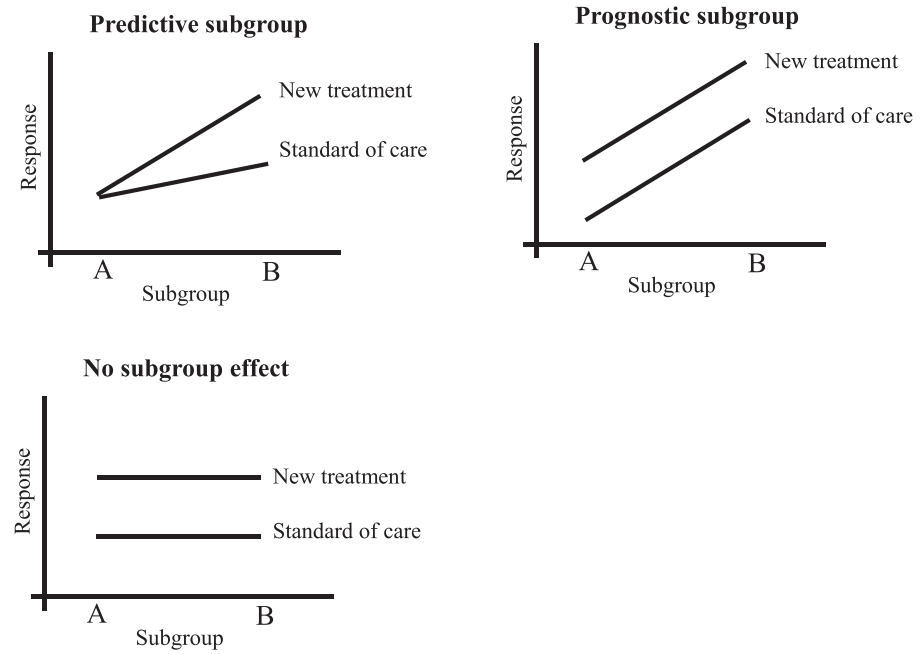

Figure 2. Subgroup effects that are predictive, prognostic or have no effect.

Copyright © 2014 John Wiley & Sons, Ltd.

and comparator groups) varies between subgroups, the subgroup effect is predictive. If there is a subgroup effect and the average treatment effect is constant across subgroups, but the response to treatment differs, the subgroup effect is prognostic. If the response to treatment is constant across subgroups, there is no subgroup effect. We can think of this in terms of a regression model: If there is an interaction between the subgrouping factor and the treatment effect, the subgroup effect is predictive; if there is a treatment effect and a subgroup effect, but no treatment by subgroup interaction, then the subgroup effect is prognostic. It is important for HTA agencies to understand whether a subgroup is predictive or prognostic before arriving at a final reimbursement decision.

A decision on whether candidate subgroup effects (at least in terms of clinical effectiveness) are real or because of random variation may vary according to whether the effect is seen as prognostic or predictive. Buyse *et al*. [8] suggest that candidate prognostic markers are relatively easy to identify, whereas predictive markers require extensive data for validation. Sleight [9] goes further to suggest that the identification of prognostic factors might be the best way to identify subgroups in a statistically reliable way and that we should be very cautious when identifying predictive effects.

Whether a subgroup effect is regarded as prognostic or predictive may depend on the scale used to measure treatment effects. For example, consider a trial with two subgroups: The probabilities of response to experimental treatment and comparator are 20% and 10% in one subgroup and 36% and 20% in the other. The subgroup effect would be prognostic on the odds ratio scale (odds ratio is 2.25 in both subgroups) but predictive on the relative response rate scale (relative risk is 2 and 1.8, respectively) and on the absolute difference scale (absolute difference is 10% and 16%, respectively).

It should be noted that in CEA, we need to consider all subgrouping factors jointly when estimating cost-effectiveness as we need to estimate the cost-effectiveness for a patient with given characteristics. For example, if both sex and age are prognostic in terms of cost-effectiveness, we need to estimate the cost-effectiveness of patients with each possible combination of age and sex. This is at variance with trial analyses as commonly published where each subgrouping factor is typically considered individually.

For the purposes of estimating cost-effectiveness, the key consideration is whether there are differences between subgroups in terms of costs and effects expressed as a cost-effectiveness ratio or incremental net monetary benefit. For example, consider a treatment for a chronic disease where eligible patients receive treatment for a short 'trial' period, and only those who respond continue to receive the treatment beyond the trial period [8]. Suppose in a clinical trial of such a treatment, the probabilities of response to experimental treatment and comparator were 20% and 10% in one subgroup and 60% and 50% in another. In the first subgroup, we would need to continue to treat two patients with the experimental treatment beyond the trial period to obtain one additional responder over the control treatment. In the second group, we would need to continue to treat six patients for each incremental responder. Although the subgroup effect is not predictive on the absolute risk difference scale, the cost-effectiveness of the active treatment varies between the two subgroups.

2.1.2. The impact of subgroup size on cost-effectiveness analysis. Outside of the explicit use of cost-effectiveness to improve technical efficiency, decision-makers may seek to restrict the use of a technology to those patients who are expected to experience the greatest clinical benefit in order to limit the total population receiving the technology and hence limit the total cost of funding the new technology (budget impact). In this case, the estimated number of patients within each subgroup who might be expected to receive the new technology becomes important.

The number of potential patients within each subgroup is important to manufacturers when setting prices for a new technology. If the different subgroups under consideration are arranged in order of decreasing incremental health benefit, by taking account of the relative size of the subgroups, a 'price elasticity' curve showing the responsiveness of the quantity demanded to a change in price can be defined. This could help manufacturers identify the revenue maximising price [10]. An example of the curve is shown in Figure 3. The top graph shows how demand for a technology varies with price and the lower graph shows how revenue (demand $\times$ price) varies with price. At a price of zero, the demand is maximal as the treatment will be used for all subgroups that receive any clinical benefit. As the price increases, demand remains constant and revenue increases until the technology becomes not cost-effective for the subgroup that receives the least benefit. Then, there is a step down in demand as the technology is no longer used for that subgroup. As the price further increases, the demands remain constant at the new level, and revenue increases until the technology becomes not cost-effective for the subgroup that receives the next least benefit and so on. The red dot marks the price at which revenue is maximised. The shape of the demand curve, and hence the revenue maximising price, is determined by the number of patients in each subgroup and the magnitude of the differences in effect between subgroups.

For a given WTP for a unit of health benefit, the number of subgroups in which a technology is cost-effective decreases as the price of a technology increases. In principle, to maximise revenue, the price should be set such that if the price were to be increased any further, the decrease in revenue from the subgroup in which the technology would be no longer cost-effective would be greater than the increase in revenue because of the increase in price across the subgroups in which the treatment remains cost-effective. In this context, the subgroups should represent different groups between which it is not possible to price discriminate. These subgroups might represent patient subgroups, difference in indications or even different markets if they have common pricing structure, as a result of reference pricing for example.

2.1.3. Contrasting subgroup analysis for health technology assessment and for marketing authorisation. The evaluation of subgroups within CEA may differ from that for regulatory purposes in a number of ways. Regulators will use subgroups predominantly to assess the consistency of treatment effects for a set of prespecified subgroups as this will aid their assessment of benefit–risk for defining a product label. However, in a CEA, multiple subgrouping factors defined within the product label will often be considered jointly rather than separately so the total budget impact of reimbursing the treatment can be assessed having taken into account the benefits, risks and costs. The number of patients within each subgroup will be important when considering budget impact or price elasticity. Subgrouping factors may also be considered as continuous variables rather than being categorised as shown in Figure 1.

Copyright © 2014 John Wiley & Sons, Ltd.

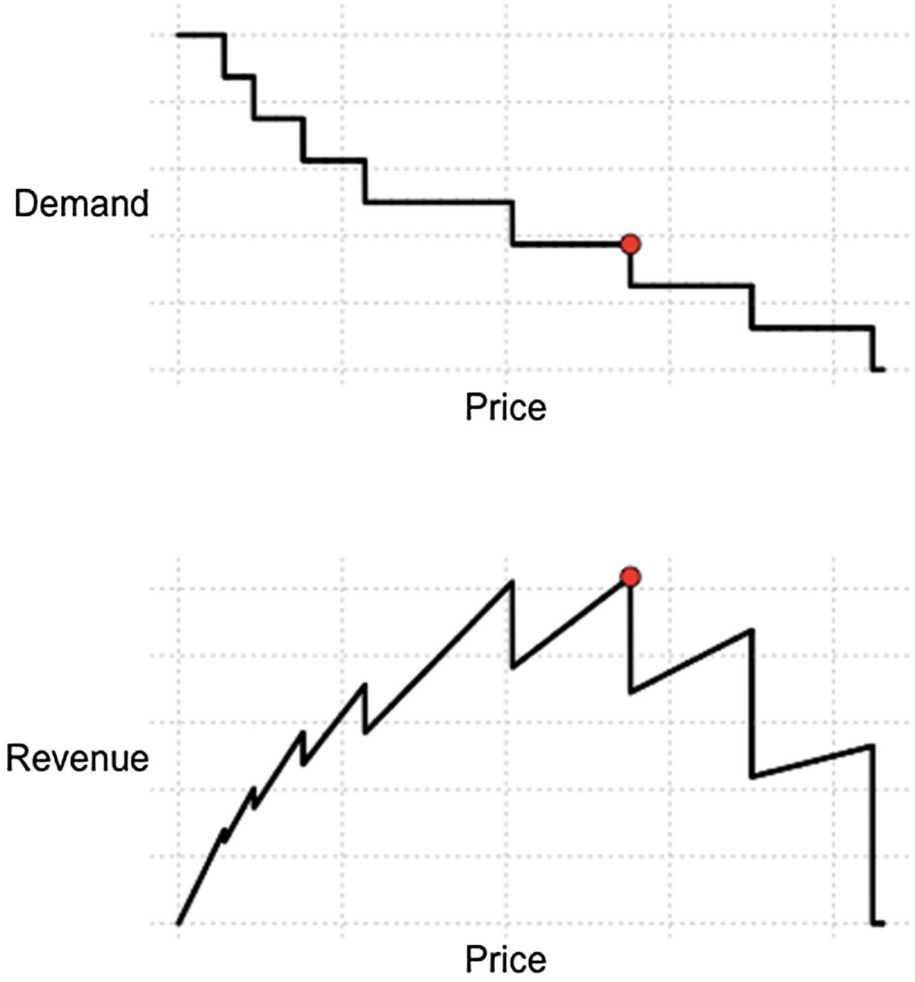

Figure 3. Subgroups and price elasticity.

On the other hand, subgroup analysis for CEA also shares some similarities with that for regulatory purposes. In both instances, we should be wary of 'optimism' bias when considering the results of subgroup analysis where there is increased expectation that the subgroup results will conclude cost-effectiveness. When considering the biological plausibility of a subgroup effect, it is important to consider subgroup results as though the results of the trial were unknown. It may be useful to ask whether the biological rationale would likely have been formulated before the trial was run. Sleight [9] offers a variation on this theme when he suggests that the direction of subgroup effects should be prespecified. We should also be wary of the effects of using the same dataset to both select the subgroups and to estimate cost-effectiveness model parameters [11]. This can lead to 'selection' bias and also overconfidence in our results if we do not adjust for the number of competing models considered.

Finally, we need to be mindful of the practicalities of implementing the results of the subgroup analysis in clinical practice. Can we evaluate a cost-effectiveness model each time a patient is treated to determine the most effective and cost-effective treatment based on multiple characteristics? Or do we need a simpler scoring algorithm based on one or two characteristics? Future improvements in information technology may allow more sophisticated models to be generated on the basis of multiple patient characteristics as observed in routine clinical practice.

2.2. Challenges in incorporating subgroup analyses in designing clinical trials that meet the needs of regulators and payers

Ideally, planning subgroup analyses in the phase III (confirmatory) designs should aim to meet the needs of both the regulators and the payers to optimise the chance for regulatory success and favourable market access decisions. For regulatory submissions, it has become a tradition to present the primary efficacy and the key safety endpoints for major subgroups based on demographics, disease stage and biomarkers considerations. The number of these subgroups could be large (e.g. around 20), and the subgroups are typically predefined in the protocol. If the analysis of a particular subgroup is meant to be confirmatory, that is, the evidence will form the basis of a benefit–risk assessment by the regulators to support a label claim for the subgroup, a manufacturer is expected to prespecify a multiple comparison procedure to handle the multiplicity in the statistical inference. In the latter case, sample size for the subgroup needs to be carefully planned to ensure adequate power. Otherwise, the subgroup analyses

Copyright © 2014 John Wiley & Sons, Ltd.

will be viewed only as an effort to investigate the homogeneity (consistency) of the treatment effect across subgroups.

By contrast, subgroup analyses in HTA submissions are conducted to identify where a treatment has the most clinical and cost benefit or there is incremental clinical benefit worth paying for (preferably compared with existing therapies). In an HTA, the absolute treatment effect is often more meaningful than the relative treatment effect. So, if a treatment has a similar relative effect across different subgroups, subgroups with a higher baseline risk will often benefit more from the treatment on the absolute scale than those with a lower risk at baseline. As such, it is quite possible that some subgroups may have a more favourable cost-effectiveness profile than the general population. It is possible for a smaller relative effect to be observed in a subgroup that results in a bigger absolute difference between treatments. For example, in oncology studies, whilst the hazard ratios (measure of the relative effect) could be considered similar across subgroups, patients with less severe disease at baseline could have a slightly less favourable hazard ratio but demonstrate a larger and clinically important difference in median survival compared to patients with more severe disease at baseline. From a society perspective, the bigger absolute treatment effect may be more important, but this could incur more costs and therefore a less favourable ICER. However, as noted in the National Institute for Health and Care Excellence (NICE) methods guide [12], it is important to use 'measures of both relative and absolute effectiveness with appropriate measures of uncertainty' when deriving estimates of treatment effect for CEAs.

It is also possible that subgroups could be chosen on the basis of cost considerations alone. For example, treating patients with Alzheimer's disease (AD) earlier may prevent family members from missing work and hence would minimise loss of productivity. This could lead to a reimbursement decision to treat early onset of disease. How to incorporate the impact on society or on productivity into a CEA is currently being discussed, for example, in the UK and amongst individuals advocating the development of a value-based pricing system.

Sculpher [13] suggested potential subgroups that will support an HTA are prespecified as much as possible based on clinical and economic plausibility whilst acknowledging that prespecification of all sources of heterogeneity is often not possible. However, recognising clinical heterogeneity and clarifying its implications helps decision-makers to identify patients and patient populations who benefit the most, who benefit the least and who are at the greatest risk of experiencing adverse outcomes from a particular intervention [14].

A major challenge associated with prespecifying subgroups in a confirmatory programme is the limited knowledge on the treatment effect in subgroups at the end of phase II trials to investigate treatment effect modifiers and conduct economic modelling at this stage. One option for a manufacturer is to prespecify a strategy that will be used to identify patients with higher treatment effect using methodologies such as classification and regression trees. The factors chosen for inclusion in the methodologies should be based on clinical and biological rationale.

2.3. Methods for adjusting estimates of costs and effects from subgroup analyses to minimise bias

CEA in subgroups may be compared across different subgroups using the ICER. Research has shown that if the true average ICER is the same across subgroups, choosing subgroups with a higher estimated mean ICER value is similar to selecting the more extreme observations from a set of independently and identically distributed random observations [15]. Not only could the conclusion on the best-performing subgroups be misleading, the estimates on the ICER for the chosen subgroups could also overestimate the mean ICER for these subgroups. Some researchers have suggested using a Bayesian analysis [16] to shrink subgroup results towards the average population results. The assumption behind this Bayesian analysis is that the same response distribution applies to different subgroups. Whilst the latter will produce estimated subgroup results closer to each other, subgroups with a more favourable cost-effectiveness estimate under a non-Bayesian analysis will still enjoy a more favourable estimate under the Bayesian approach.

The use of probabilistic sensitivity analysis enables uncertainty to be incorporated in CEA. Nevertheless, an HTA body still needs to decide if the level of uncertainty is appropriate to support a favourable reimbursement decision. In reality, a universal cut-off to make this binary decision does not exist. Rather, the level of uncertainty together with other factors such as unmet medical needs and ethical considerations will be used to guide final decisions. Whilst there does not exist a well-accepted approach to address bias arising from multiple analyses (short of discounting or shrinking the observed results), it is important that manufacturers and HTA bodies are aware of these issues when conducting subgroup analyses to support HTA decisions.

3. GUIDING PRINCIPLES FOR SUBGROUPS IN COST-EFFECTIVENESS

In Paget *et al.* (2011), we recommended seven guiding principles for conducting subgroup analyses of clinical effectiveness to support an HTA. We now explore the relevance and importance of each of these guiding principles in using subgroup analyses in CEA to support an HTA.

3.1. Principle (1): definition of subgroups

In confirming the clinical effectiveness of a new therapy in specific subgroups, having the subgroups prespecified with a scientific and biological rationale is important for drawing credible conclusions on the clinical effectiveness for those subgroups. For CEA, whilst prespecification of those subgroups is desirable, *a more pragmatic approach* is needed that accommodates the more complex nature of the decisions when both costs and effectiveness need to be considered together. The identification of important subgroups for CEA may be derived from preliminary health economic models using phase 2 data, which may point to key drivers that influence cost-effectiveness. However, until the health economic model incorporating the phase III results is finalised, it will be difficult to prespecify all the potential subgroups that may contribute to determining whether cost-effectiveness can be demonstrated. HTA agencies do have different opinions on the definitions of subgroups for CEAs. For example, the Haute Autorité de Santé) guidelines notes that 'an analysis of the health effects for subgroups can be made if it is based on clinical studies or other types of studies which include a subgroup analysis in their protocol' [17]. The NICE guidelines note that 'when possible, potentially relevant subgroups will be identified at the scoping stage with consideration being given to the rationale for expecting a subgroup effect. However, none precludes the identification

Copyright © 2014 John Wiley & Sons, Ltd.

of subgroups later in the process, in particular during the deliberations of the Appraisal Committee' [12].

Whilst *a degree of flexibility* is required in establishing which subgroups are most important from the CEA perspective, the process in which they were identified, including reference to any pre-specified strategy used should be transparent and fully described.

3.2. Principle (2): subgroup by treatment interactions

For CEAs, formal subgroup by treatment interaction tests *is less important* compared with clinical effectiveness analysis. As discussed in Section 2.1.1, assessing whether subgroup factors are predictive or prognostic is important within a CEA, but only a predictive subgroup factor results in a significant subgroup by treatment interaction. Instead, an informal assessment of the heterogeneity or consistency of cost-effectiveness results across subgroups using estimation and with graphical displays of the uncertainty is often conducted. The size and extent of data available for each subgroup should also be described and summarised. As noted in the NICE methods guide (Section 5.10), 'relevant subgroups may be identified in terms of differences in 1 or more contributors to absolute treatment effects'. Statistical modelling of parameters thought to influence cost-effectiveness can help ascertain how specific parameters are correlated and how uncertainty in the parameters affects decisions. In addition, the latter can help identify which parameters may be the most important for cost-effectiveness.

3.3. Principle (3): multiplicity issues

In CEA, the focus for the analyses is on estimation rather than hypothesis testing, so adjusting for multiplicity arising from subgroup analyses *is less important* but cannot be ignored. If finding a subgroup with an acceptable cost-effectiveness profile is only achieved in post-hoc analyses with little reference or linkage to planned subgroups known to be important for the clinical setting, then the integrity of the subgroup finding may be in question.

3.4. Principle (4): sensitivity analyses

Sensitivity analyses are *extremely important* for subgroup analyses supporting cost-effectiveness decisions. In contrast to clinical effectiveness analyses where sensitivity analyses are mainly conducted to support confirmatory findings in subgroup analyses, sensitivity analyses for cost-effectiveness are vital in order to assess the robustness of cost-effectiveness results in subgroups and quantify the uncertainty. To conclude that a new treatment is cost-effective for a subgroup, we need to demonstrate the *strength of evidence, the precision of* the cost-effectiveness estimates and *the consistency of the evidence* under different but clinically plausible scenarios considered in the sensitivity analyses. This requires a large number of parameters in the cost-effectiveness model to be varied. This can be carried out using univariate analyses when factors are varied one at a time or in multivariate analyses. The latter is often achieved using probabilistic sensitivity analysis.

The impact of distribution assigned to each parameter in a cost-effectiveness model can be explored by fitting alternative distributions and quantifying how the estimates of cost-effectiveness vary. A variety of graphical displays can be used to summarise results from the sensitivity analyses, including tornado plots (univariate analyses), cost-effectiveness acceptability curves (CEAC) (multivariate analyses) and more recently cost-effectiveness acceptability frontier (see also principle (7)).

3.5. Principle (5): replication

Replication is important for concluding cost-effectiveness in a subgroup analysis. Being able to replicate cost-effectiveness results for a subgroup *across multiple data sources* increases the validity of the conclusions and reduces concerns that the subgroup is a chance finding. Replication of subgroup results for clinical effectiveness is often achieved using RCTs, where either the subgroup effect was observed in multiple trials involving the new treatment or in multiple trials involving other treatments for the same disease. Replication of subgroup results for cost-effectiveness may require additional data sources other than RCTs. For example, observational data summarising country-specific clinical practice data will be useful where cost-effectiveness is driven by cost parameters that are dependent on local clinical practice. This is especially true when there is only a single pivotal RCT that formed the basis of clinical effectiveness assessment.

Ultimately, some reassurance is needed that any assumptions regarding subgroups are applicable and can be verified as suitable to support a CEA. For example, it may be the case that similar effectiveness estimates can be demonstrated across subgroups, but there are different cost-savings across subgroups that leads to different cost-effectiveness results. In such cases, being able to define the mechanism that led to different costs, for example, linked to disease or treatment pathway that may be influenced by the model structure, increases the justification of the subgroup. The estimated treatment effect for a treatment (versus a particular comparator) may vary according to the point at which the treatment is used in care or treatment pathway. It is important to quantify and account for this potential source of variation in estimated treatment effect when synthesising data from multiple trials and when attempting to estimate cost-effectiveness for a treatment (used at a particular point in the care pathway). Subgroup analyses that report the effects of treatment according to prior treatment history, disease duration or disease stage are potentially valuable in obtaining unbiased estimates for treatment effects relevant to specific decision problems.

3.6. Principle (6): sources of evidence

In subgroup analyses for cost-effectiveness analyses, reimbursement agencies often prefer evidence from *pragmatic trials* [18] that provide evidence of cost-effectiveness in the real-world settings. For initial reimbursement assessments at the time of a product launch, the evidence supporting CEA in subgroups will generally rely on RCT evidence for estimates of clinical effectiveness and rely on observational research for estimates of costs. The ability to use network meta-analysis methodology to derive estimates of relative effectiveness for subgroups can be challenging. In addition to the limited subgroup analysis results published for existing therapies, it is important to keep in mind potential biases of network meta-analysis in estimating relative effectiveness because of different study designs, different patient populations and endpoints as well as how subgroups were defined in different studies. As the healthcare environment moves to adopt a data-sharing paradigm [19,20], we expect subgroup results will become more readily available in the future.

Copyright © 2014 John Wiley & Sons, Ltd.

When CEA is being reassessed postlaunch, *all data sources* should be evaluated and considered. Postmarketing studies can provide subgroup estimates of clinical effectiveness from the real-world settings and provide estimates of important cost parameters reflecting how the treatment has been incorporated into local medical practice.

3.7. Principle (7): presenting and reporting subgroup results

The principles for presenting and reporting CEA in subgroups include being transparent about all analyses conducted, whether they were planned or ad hoc and whether the results are favourable or unfavourable. The full body of evidence including an appropriate range of sensitivity analyses should be presented and interpreted with *considerations to potential biases and limitations*.

Graphical displays are useful in presenting CEA for subgroups. Examples and best practices are discussed in Briggs *et al*. [21]. For example, when comparing two technologies, one can represent different subgroups using different symbols on a cost-effectiveness plane, where difference in cost is plotted (*y*-axis) against difference in effectiveness (*x*-axis) [22]. Another commonly adopted graph is the CEAC that shows the probability that one technology is cost-effective compared with another one for a range of threshold values that a decision-maker is willing to pay. Different CEACs could be drawn for different subgroups to summarise the evidence in support of one technology relative to another within the subgroups. The cost-effectiveness acceptability frontier shows the probability that a technology with the highest expected net benefit is cost effective relative to other technologies. Tornado plots show the impact (change) on CEA by assessing the impact of the parameters individually, with the parameters presented in descending order of impact. As noted by Barton *et al*. [23], the cost-effectiveness plane has limited use in representing the uncertainty surrounding multiple options as it cannot represent correlation between the options. CEACs can represent decision uncertainty but should not be used to determine the optimal decision. Instead, the CEAF shows the decision uncertainty surrounding the optimal choice. Tornado plots do not incorporate measures that reflect the accuracy of the parameter estimates. Often, a wide range of possible scenarios will be modelled to assess how CEA is impacted.

4. CASE STUDY

An example of an HTA where subgroup analyses were important in the recommendations is summarised in the succeeding texts.

4.1. Background

The National Institute for Health and Care Excellence Guidance TA111 [24], issued in November 2006, reviewed the use of three acetylcholinesterase (AChE) inhibitors, donepezil, galantamine and rivastigmine and an *N*-methyl-D-aspartate receptor antagonist, memantine, for the treatment of AD. At the time of the NICE review, the three AChE inhibitors had marketing authorisations for mild to moderately severe AD, whilst memantine had marketing authorisation for moderately severe to severe AD. Guidance

Table I. Comparison of cost-effectiveness assessments.

	TA111 (Nov 2006)		TA217 (March 2011)	
Issue	Manufacturers	NICE	Manufacturers	NICE
Prespecified subgroups	No	Retrospective analysis suggested incremental cognitive benefit in more severely impaired patients	No, clinical effectiveness in subgroups not performed	No, clinical effectiveness in subgroups not assessed.
Studies included for clinical effectiveness	RCTs + OL and observational studies	RCTs, including some with doses not used in clinical setting	RCTs + OL & observational studies	RCTs
Assumptions on:				
Cost	100% of costs of care met by NHS	Only 70% of costs of full-time care would be met by the NHS	All costs included	Excluded costs to individual of institutional care
Discontinuation		Discontinuation of treatment not accounted for	Discontinuation assessed in sensitivity analyses	Discontinuation rate = 4% per month
Mortality	AHEAD risk equation (galantamine)	Annual mortality rate = 11.2%	Survival = 4.6 life years (moderate cohort)	Survival = 3.6 life years (moderate cohort)
Time horizon	5 years (donepezil and rivastigmine), 10 years (galantamine) and 2 years (memantine)	5 years, also 2 years for memantine	Lifetime (donepezil), 5 years (others)	20-year time horizon (mild–moderate)

Copyright © 2014 John Wiley & Sons, Ltd.

TA111 recommended the three AChE inhibitors for people with AD of moderate severity only (mini mental state examination between 10 and 20). Memantine was not recommended as a treatment option except as part of well-designed clinical trials. For people with mild AD who were on an AChE inhibitor and for people with moderately severe to severe AD receiving memantine at the time Guidance TA111 was issued, the guidance stated that they should continue on therapy until their carers and/or specialist consider it appropriate to stop.

4.2. Appraisal

There were several differences in the CEA provided by the manufacturers and those performed by the assessment group, some not directly pertaining to subgroup analyses, such as data sources and assumptions made in the CEA modelling, but which had an impact on the decision. Others were directly related to subgroups. For example, all three AChE inhibitors were shown to be clinically effective in the overall study populations, and no subgroup analyses were prespecified or submitted by the manufacturers in the original submission. Retrospective analyses based on severity of cognitive impairment, performed by the MRC Biostatistics Unit, suggested some differential advantage (in terms of mean changes) for more severely cognitively impaired subgroups. These differential benefits were incorporated into the assessment group's CEA and led to the initial conclusion that the ICERs for the mild group did not support extending treatment of AChE inhibitors to mild cases. The assessment group considered that the evidence to determine the clinical effectiveness of memantine in either the whole population of moderately severe to severe AD or in a retrospective subgroup analysis of people with behavioural symptoms was insufficient. As such, NICE did not consider memantine a cost-effective therapy.

A reappraisal led to TA217 [25] being issued in March 2011, which replaced TA111 (and its amendments in 2007 and 2009) with new guidance: Donepezil, galantamine and rivastigmine are recommended as options for managing mild, as well as moderate AD. Memantine is recommended as an option for managing moderate AD in patients who are intolerant of or who have a contraindication to AChE inhibitors and for managing severe AD.

In the revised review, no subgroup analysis had been identified as part of the updated systematic review of clinical effectiveness analysis by either the manufacturers or by the assessment group. It was also noted that clinicians found it difficult to differentiate between mild and moderate disease using the mini mental state examination and that many AChE inhibitor trials included patients with both mild and moderate disease.

Despite differences in the models and assumptions used by the manufacturers and the assessment group to assess cost-effectiveness (refer to Table I for details), sensitivity analysis using different assumptions did not alter the conclusion that all four treatments are cost-effective. The final recommendations in TA217 were very close to the marketing authorisations granted to these four products. The case study illustrates the importance of prespecifying subgroups and being able to easily identify subgroups. Equally important is prespecifying as much as possible the data sources and modelling techniques and performing sensitivity analyses under various assumptions.

5. DISCUSSION AND CONCLUSIONS

Subgroup analyses are an important consideration in CEA. There are differences in how subgroup analyses are incorporated into decision-making processes between regulators and payers. Statisticians need to understand that these differences exist and that both perspectives are important. Between the end of 2010 and early 2012, pilots of early multistakeholder consultations were conducted within the Tapestry network initiative [26]. These consultations involved regulators, HTA and coverage bodies, patient representatives, clinicians and pharmaceutical companies in France, Germany, Italy, the Netherlands, Sweden, the UK and Spain. The primary objective of the pilots was to improve clarity and alignment across stakeholders regarding what constitutes a medicine's value and what evidence is required to demonstrate that value in a most effective manner. Statisticians are encouraged to keep abreast of these and similar initiatives and incorporate recommendations in their drug development activities.

Many of the guiding principles recommended by Paget *et al*. [4] for using subgroup analyses to support clinical effectiveness in an HTA are relevant for subgroup analyses used to support CEA also. In particular, it is important to demonstrate replication of cost-effectiveness results in subgroup analyses, apply a comprehensive set of sensitivity analyses to help quantify the uncertainty in CEA of subgroups and appropriately report and interpret subgroup CEA results. A more pragmatic approach is needed on whether a subgroup important for cost-effectiveness has to be predefined, although it certainly helps. Assessing consistency and the degree of variability of subgroup results is more important than subgroup by treatment interactions. The evidence required to support CEA in subgroups at launch will continue to rely primarily on RCT evidence, although pragmatic trials are gaining interest, and observational research will provide important supportive data.

The case study highlights different approaches when considering subgroups in CEA. Whilst key aspects such as data sources and modelling assumptions were pre-specified, these could change over time, including incorporating subgroups which were originally thought not to be important to demonstrate CEA. Documenting key assumptions and, where possible, agreeing these with payers may help to minimise these differences. Whilst some of the differences between manufacturers and NICE assessments did not pertain to subgroups specifically, they could impact on subgroup decisions indirectly in the assessments. Most striking is how different stakeholders may derive different conclusions from the same evidence base by employing different strategies to address questions important in the decision-making process. It is possible that subgroups could be chosen on the basis of cost considerations. For example, treating patients with AD earlier may prevent family members from missing work and hence would minimise loss of productivity. This could lead to a reimbursement decision to treat early onset of disease.

It is highly recommended that statisticians and health economists work together in developing preliminary cost-effectiveness models and discuss potential subgroups of interest important for demonstrating cost-effectiveness before finalising the design of confirmatory trials. The discussions should include both the relative and absolute effects of clinical effectiveness across subgroups and how baseline covariates may influence cost-effectiveness estimates and the associated levels of uncertainty. As CEA continues to evolve, we recommend

Copyright © 2014 John Wiley & Sons, Ltd.

that statisticians, in collaboration with their health economic colleagues, develop a high level statistical analysis plan for CEA analyses that enable key strategies and principles be identified and aligned on the CEA analysis requirements.

In the future, the increasing focus on personalised medicine will make incorporating subgroups in CEA even more important and will also bring more challenges. For example, with smaller sample sizes assessing CEA in subgroups may lack the desired level of precision. As in the regulatory setting, it would be helpful if payers could develop disease guidelines that define key principles and elements important for the decision-making framework for that disease.

Whilst not discussed in this paper, manufacturers are increasing seeking other approaches, such as risk share agreements to maximise market access success, especially for personalised medicines. Risk share agreements can be either financially driven, for example, manufacturers offer a discount to the list price, or performance driven, for example, manufacturers pay for part or all of the new treatment costs if patients fail to respond. Subgroup analyses of confirmatory RCTs are now being incorporated in defining risk share agreements and are considered an important tool for deriving pricing strategies. This is another reason for early collaboration between statisticians and health economic colleagues.

In conclusion, subgroup analyses for CEA need to be planned early in the drug development process, especially in the design and analysis of confirmatory clinical trials. Adhering to good statistical principles for subgroup analyses in CEA is recommended to provide a robust evidence package for reimbursement decisions.

Acknowledgements

The authors would like to acknowledge Christoph Gerlinger and John Davis for their contributions and review of the paper.

REFERENCES

[1] Banta D. The development of health technology assessment. *Health Policy* 2003; **63**(2):121–132.

[2] Arrow KJ. Uncertainty and the welfare economics of medical care. *Bulletin of the World Health Organization* 2004; **82**(2):141–149, ISSN 0042-9686.

[3] Ramaekers BLT, Joore MA, Grutters JPC. How should we deal with patient heterogeneity in economic evaluation: a systematic review of national pharmacoeconomic guidelines. *Value in Health* 2003; **16**:855–862.

[4] Paget C-S, Fletcher R. Subgroup analyses of clinical effectiveness to support health technology assessments. *Pharmaceutical Statistics* 2011; **10**:532–538.

[5] General Methods. Version 4.1 Institute for Quality and Efficiency in Health Care. Available at: https://www.iqwig.de/download/IQWiG_General_Methods_Version_%204-1.pdf (accessed 6.6.2014).

[6] Woods B, Veenstra D, Hawkins N. Prioritizing pharmacogenetic research: a value of information analysis of CYP2D6 testing to guide breast cancer treatment. *JVAL* 2011; **14**(8):989–1001.

[7] Claxton K, Briggs A, Buxton MJ, Culyer AJ, McCabe C, Walker S, Sculpher MJ. Value based pricing for NHS drugs: an opportunity not to be missed? *BMJ: British Medical* 2008; **336**(7638):251–254.

[8] Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points—the challenge of statistical validation. *Nature Reviews Clinical Oncology* 2010; **7**:309–317.

[9] Sleight P. Debate: subgroup analyses in clinical trials: fun to look at—but don't believe them. *Current Controlled Trials in Cardiovascular Medicine* 2000; **1**(1):25–27.

[10] Hawkins N, Scott DA. Reimbursement and value-based pricing: stratified cost-effectiveness analysis may not be the last word. *Health Economics* 2011; **20**(6):688–698.

[11] Steyerberg EW. Overfitting and optimism in prediction models. In *Statistics for Biology and Health*, Statistics for Biology and Health. Springer US: New York, NY; 2008; pp. 83–100.

[12] Guide to the methods of technology appraisal 2013. NICE Website. Available from: http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9 (Accessed 8.03.14).

[13] Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics* 2008; **26**:799–806.

[14] Gartlehner G, West SL, Mansfield AJ, Poole C, Tant E, Lux LJ, Lohr KN. Clinical heterogeneity in systematic reviews and health technology assessments: synthesis of guidance documents and the literature. *International Journal of Technology Assessment in Health Care* 2012; **28**(1):36–43.

[15] Li Z, Chuang-Stein C, Hoseyni C. The probability of observing negative subgroup results when the treatment effect is positive and homogeneous across all subgroups. *Drug Information Journal* 2007; **41**:47–56.

[16] Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technology Assessment* 2000; **4**(38):1–130.

[17] Choices in Methods for Economic Evaluation, HAS, October 2012. HAS website. Available at: http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf (accessed 8.03.14).

[18] Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003 Sep 24; **290**(12):1624–32.

[19] Publication and access to clinical-trial data: an inclusive development process. EMA website. Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/general/general_content_000556.jsp (accessed 1.10.2013).

[20] Availability of masked and de-identified non-summary safety and efficacy data; request for comments. FDA website. Available at: https://www.federalregister.gov/articles/2013/06/04/2013-13083/availability-of-masked-and-de-identified-non-summary-safety-and-efficacy-data-request-for-comments (accessed 1.10.2013).

[21] Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. Oxford: Oxford University Press; 2006.

[22] Black WC. The CE plane: a graphic representation of cost-effectiveness. *Medical Decision Making* August 1990; **10**: 212–214.

[23] Barton Gr, Briggs AH, Fenwick EAL. Optimal cost-effectiveness Decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value in Health* 1 September 2008; **11**(5):886–897. **DOI:** 10.1111/j.1524-4733.2008.00358.x

[24] Alzheimer's disease—donepezil, galantamine, rivastigmine (review) and memantine (replaced by TA217) (TA111), Sept 2007. National Institute for Health and Care Excellence.

[25] Alzheimer's disease—donepezil, galantamine, rivastigmine and memantine (TA217), March 2011. National Institute for Health and Care Excellence.

[26] Pilots of multi-country, multi-stakeholder consultations in drug development: from proof of concept to tangible benefits, June 2012. Available at: http://www.tapestrynetworks.com/initiatives/healthcare/upload/Pilots-of-multi-stakeholder-consultations-in-drug-development-6-June-2012.pdf (accessed 1.10.2013).

Copyright © 2014 John Wiley & Sons, Ltd.