NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Briefing paper for methods review working party on extrapolation and crossover

The briefing paper is intended to provide a brief summary of the issues that are proposed for discussion by the Methods Review Working Party to inform an update to the Institute's Guide to Methods of Technology Appraisal. It is not intended to reflect a comprehensive or systematic review of the literature. The views presented in this paper are those of the authors and do not reflect the views of the Institute.

1 Review of the 'Guide to Methods of Technology Appraisal'

The Institute is reviewing the 'Guide to the methods of technology appraisal', which underpins the technology appraisal programme.

The original Methods Guide was published in February 2001, and revised versions were published in 2004 and 2008. The Methods Guide provides an overview of the principles and methods used by the Institute in assessing health technologies. It is a guide for all organisations considering submitting evidence to the technology appraisal programme and describes appraisal methodology.

The revised draft of the Methods Guide will be available for a 3-month public consultation, expected to begin in June 2012. We encourage all interested parties to take part in this consultation.

2 Background

2.1 Relevance of topic to NICE technology appraisals

When conducting appraisals of the clinical and cost effectiveness of technologies, the Appraisal Committee considers all relevant costs and consequences of treatment, often these will occur over a lifetime horizon. However, it is very rare that sufficient data on the effectiveness of a technology will be available at the time of an appraisal. Frequently within the pivotal trials of a technology the length of data follow-up is relatively short. This is often the case for chronic conditions as it is very rare for all patients to be followed up throughout the full course of their treatment and subsequent experience of the condition. It is possible that data may be available from non-randomised controlled trial (non-RCT) sources and these data may serve a role in informing estimates of the longer-term effects of a treatment. However, such data are rarely available and therefore extrapolation of the observed trial data is undertaken to estimate benefits within the unobserved period. A number of methods can be used to extrapolate available trial data; however limited alternatives are usually presented to the Appraisal Committee. Clear justification for the choice of the extrapolation model used, especially in those instances whereby different curve fits produce very different cost effectiveness estimates, is rarely provided for the Appraisal Committee. Additionally, where very short-term data are available for extrapolation it is challenging to ensure that the most appropriate method for extrapolation has been used. In these circumstances external data sources and full, detailed justifications for the method of extrapolation are rarely presented to the Appraisal Committee.

Another circumstance whereby the true treatment effect is essentially unknown is when patients in clinical trials may be switched from the placebo arm to the active treatment arm. This is particularly common when the active treatment is deemed effective early in the trial, and there is a perception that it would be unethical to retain patients on the less effective treatment. In order to control for this potential ‘diluting’ effect of the treatment crossover, a variety of statistical methods have been used and presented to the Appraisal

Committee. These techniques are being sometimes used within technology appraisals, and are generally presented without rationale or clear justification for the choice of analytical method.

2.2 Introduction to extrapolation

Often, follow-up data within clinical trials are short-term and incomplete and do not follow-up the long-term experiences of all of the participants in the trial. Frequently, important outcomes such as disease progression and overall survival are collected during the trial and for a limited period after the end of the trial but this data collection then stops (that is, the data are right censored). Particularly for chronic conditions, this means that only limited data on the number of people who progress and survive with and without treatment are available to inform the mean estimates of the clinical effectiveness and cost effectiveness of health technologies.

In instances whereby follow-up is incomplete, assumptions are regularly required to fully estimate the long-term benefits of a technology. Restricting decision making to the observed data available, especially in the presence of high levels of censoring, is likely to provide inaccurate and potentially biased estimates of the long-term effect of treatments and may ultimately lead to inaccurate estimates of the cost effectiveness of a technology.

In some instances, non-RCT evidence (or ‘real-world’ observations) are available and these can be used to estimate what would have happened if the participants in the trials had continued to be observed. It is however rare that these real-world observations are available; particularly in the case of newly licensed technologies, such long-term data simply do not exist. In these circumstances extrapolation of the observed data must be performed.

A number of methods are available for performing extrapolation. Exponential, Weibull, Gompertz, log-logistic or lognormal parametric models can be used. In addition, a number of more complex and flexible models are available such as piecewise exponential models. Some of these models allow for assumptions of proportional hazards between treatment arms, whilst others do not. The different methods have varying functional forms and the choice of

which model should be used varies according to the available data and each model has different characteristics which may make it more or less suitable for use in particular circumstances.[1;2]

The importance of extrapolation is often paramount in a technology appraisal. It is possible that the choice of the survival model can have a substantial effect on the resulting estimates of benefit (for example overall survival) which can subsequently have a dramatic effect on the mean cost effectiveness estimates. Therefore, the choice of extrapolation method is critical and should be considered a key issue for decision makers.

There are a number of techniques that can be used to determine which model is the most appropriate for extrapolating the data of interest. Firstly, a visual inspection (or 'eyeballing') the various curve fits to the observed data can be conducted. This method can be informative; however it is considered subjective and is therefore potentially inaccurate. Additionally, it is common that a number of parametric curves appear to fit the data well, and therefore visual inspection alone should be used with caution and is not considered sufficient for decision making purposes.[2]

Further, a number of statistical tests can be used to compare alternative models and their relative fit to the observed trial data. Log cumulative hazard plots or plots of residuals can be used to ascertain the nature of the observed data which in turn can inform the suitability of particular functions that can or cannot be used given the data. Once the curves have been fitted to the data, the relative 'goodness of fit' of the curves can be tested using methods such as the Akaike's Information Criterion and Bayesian Information Criterion tests. In addition, several other methods have been used in previous technology appraisals to justify the choice of curve fit to the observed data. It should be noted that patient-level data are required to conduct many of these tests; these are often not available within NICE appraisals.

The major limitation of extrapolating and the subsequent justification of extrapolation method used is that the techniques all rely on the observed data. The curves that are fitted can only be tested for goodness of fit to the

observed data (rather than the unobserved period). Thus, whilst it is possible to assess how well alternative curves fit the observed data, this does not provide any information with respect to the plausibility of the extrapolated curve beyond the observed trial follow-up period. Frequently, it is the long-term effect of a treatment on survival that has the greatest impact on estimates of the cost-effectiveness of the technology; this is particularly the case in many technology appraisals of cancer treatments. In these circumstances, the justification for selecting a particular curve fit is challenging, but the use of expert opinion, external data sources (such as historical cohort datasets or other relevant trials), and an assessment of the biological plausibility of the projected curves is recommended.[2]

Introduction to treatment switching ('crossover')

In randomised controlled trials, it is possible that participants randomised to the control group can be allowed to switch treatment group and subsequently receive the active intervention. This most commonly happens in trials of cancer treatments, whereby the participants in the control arm are switched to the intervention arm after they have experienced disease progression. This means that estimates of progression-free survival are often considered accurate, but that the switching of participants may confound the overall survival treatment effect. Within NICE appraisals in which this issue arises, methods to control for treatment switching are sometimes used to modify the estimate of treatment effect to be used in the health economic model.

In general, the use of the intention to treat principle is used to evaluate treatment effects within randomised controlled trials. This principle dictates that the treatment groups are analysed according to the treatment that patients were randomised to, regardless of the treatment that the patient actually received. However, where treatment switching has occurred and the active intervention is considered effective, then undertaking this form of analysis can lead to a 'dilution' of the overall survival treatment effect. One approach that is considered is censoring participants that crossed over from the control arm to the active intervention arm. This method is seen regularly by the Appraisal Committee, but is associated with limitations if the treatment

switching is not random and/or if a large proportion of the trial participants switched treatment and there are too few data left to conduct meaningful analyses.

Novel statistical methods for controlling for treatment switching have been presented to the NICE Appraisal Committee. For example the Rank Preserving Structural Failure Time (RPSFT) and the Inverse Probability of Censoring Weight (IPCW) models have been recently presented to the Appraisal Committee. The RPSFT method uses the randomisation of the trial in its estimation procedure in order to estimate counterfactual survival times (survival times that would have occurred if treatment crossover had not occurred). This method does not however change the level of evidence against the null hypothesis and therefore will always produce very wide confidence intervals around the point estimates, even if the point estimate is much reduced when these methods are applied. The IPCW makes a 'no unobserved confounders' assumption in order to create a 'pseudo population' consisting of control group patients that did not crossover, by making use of measurements of prognostic covariates over time. These methods are considered very complex and there are few experts who are competent in their use.

A recent review by the Decision Support Unit[3] considered these methods, among a number of other statistical techniques that can be used to control for treatment switching. It is also possible that further techniques will be developed in the future. Whilst it is rare that such statistical techniques are used, when they are the justification for the choice of method is seldom given to the Appraisal Committee.

2.3 What the current Methods Guide advises with respect to extrapolation and crossover

The current Methods Guide is detailed with regards to the need for extrapolation:

> *3.2.3 ... However, it is important to recognise that, even for the analysis of relative treatment effects, RCT data are often limited to selected*

populations and may include comparator treatments and short time spans that do not reflect routine or best NHS practice. Therefore, good-quality non-randomised studies may be needed to supplement RCT data….

Section 5 (Time Horizon) says:

5.2.14 … For a lifetime time horizon, extrapolation modelling is often necessary. When the impact of treatment beyond the results of the clinical trials is uncertain, analyses that compare several alternative scenarios reflecting different assumptions about future treatment effects should be presented (see section 5.7 on modelling). Such assumptions should include the limiting assumption of no further benefit as well as more optimistic assumptions…

Section 5 (Modelling Methods) says:

5.7.3 Modelling is often required to extrapolate costs and health benefits over an extended time horizon. Assumptions used to extrapolate treatment effects should have clinical validity, be reported transparently and be clearly justified. Alternative scenarios should be considered to compare the implications of different assumptions around extrapolation for the results. For example, for the duration of treatment effects scenarios might include when the treatment benefit in the extrapolated phase is: (i) nil; (ii) the same as during the treatment phase and continues at the same level; or (iii) diminishes in the long term.

There is currently no discussion of methods to control for treatment switching ('crossover') in the 2008 Methods Guide.

3 Proposed issues for discussion

After consideration of the developments in this methodological area, the current Methods Guide and the requirements of the Institute's Technology

Appraisal Programme, it is proposed that the following key areas are discussed by the Methods Guide Review Working Party.

3.1 Extrapolation

Currently extrapolation is mentioned in reasonable detail in the methods guide. However, the consistency in submissions varies widely:

- Should further direction be given regarding the use of extrapolation?
- Should the choice of extrapolation methods and functions be explicitly specified (or at least preferences for particular methods and functions stated)?
- Should the methods guide be more explicit about distinctions between extrapolating the baseline curve and the relative treatment effect?

What are the potential consequences of explicitly specifying what methods and functions should be used?

- Should a number of alternative fits always be shown?

What could be the impact of specifying a minimum number of curve fits (for example, stating that a single curve fit is not acceptable)?

- Should the justification of choice of model be made more explicit
    - Should goodness of fit statistics be used?
    - Should external sources, such as clinical opinion, always be sought to support extrapolation?
    - How should the biological plausibility (face validity) of an extrapolation be demonstrated?
    - Can a ‘minimum’ for justifying the choice of extrapolation be defined? For example goodness of fit alone, clinical opinion alone, a mixture of this and other components?

- How should these concerns differ according to the availability of patient-level trial data?

What could be the impact of always requesting goodness of fit statistics and/or additional support for extrapolation methods? What could be the impact of defining a ‘minimum’ process of justifying extrapolation methods?

How should NICE Methods Guide draw on the information presented within the NICE DSU Technical Support Document on survival analysis when patient-level data are available?

3.2 Treatment switching (‘crossover’)

There is currently no mention of when it is appropriate to consider controlling for the effects of treatment switching and what methods should be considered. Given that treatment switching is being seen more and more commonly in clinical trials:

- Should any guidance be given with respect to when it may be appropriate to consider statistical methods to control for treatment crossover?

What could be the consequence of including such direction in the methods guide?

- If any guidance is to be given, should specific statistical methods be referred to? Note: it will be important to consider not restricting the methods guide in terms of potential methodological developments but also the likely impact of on the review groups when receiving submissions that employ the variety of available complex methods

- If statistical methods are to be used, should guidance on how the choice of method is justified be given?

What might be the impact of including explicit guidance on preferred methods that could be used? What might be the consequences of including guidance on how to justify the choice of any methods used?

4 References

1 Collett D. Modelling Survival Data in Medical Research 2009. Chapman and Hall: Florida

2 Latimer, N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Available from http://www.nicedsu.org.uk

3 Morden JP, Lambert PC, Latimer N, Abrams KR, Wailoo AJ. Assessing methods for dealing with treatment switching in randomised clinical trials. Available from http://www.nicedsu.org.uk/Crossover%20and%20survival%20-%20final%20DSU%20report.pdf

5 Authors

This briefing paper has been prepared by Rebecca Trowman, Nick Latimer, Andrew Stevens and Paul Tappenden.