NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

Briefing paper for methods review working party on surrogate outcomes

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

The choice of outcome(s) is a key factor in any technology appraisal. In assessing the clinical and cost-effectiveness of technologies, the principal health outcome(s) should be clinically relevant, i.e. measures of health benefits and adverse effects that are important for patients and/or their carers. A clinically important (or 'final') outcome would typically include survival and/or health-related quality of life (HRQoL) that can be directly translated into quality-adjusted life years (QALYs). However, the evidence available at the time of appraisal for some (new) technologies may be solely (or largely) based on effect on surrogate outcomes (or intermediate outcomes), rather than final outcomes. In the absence of the final outcome, a surrogate outcome is defined as an outcome that is intended to both *substitute* for and *predict* the final outcome (Elston and Taylor, 2009; PBAC, 2008).

Surrogate outcomes are used as they may occur faster (than final outcomes) or may be easier to assess, thereby shortening the duration of clinical trials. In the context of In the context of health technology assessment (HTA), a surrogate outcome can include a 'biomarker' (e.g. LDL cholesterol or glycated haemoglobin [HbA_{1C}] as substitute for and predictor of future cardiovascular mortality or future major diabetic complications respectively) and also an intermediate measure of health outcome (e.g. progression-free survival as a substitute for and predictor of overall survival in cancer).

Thus, a key question for a technology appraisal, where the clinical effectiveness evidence base is principally based on a surrogate outcome, is how accurately that evidence can be used to predict the final outcomes? Or, in other words, what is the level of uncertainty associated with using a proposed surrogate outcome(s) to assess the clinical effectiveness and cost-effectiveness of a technology?

2.2 Introduction to surrogate outcomes

The use of surrogate outcomes in health policy has been controversial. Their use, at least in some applications, has led to erroneous or even harmful conclusions (Fleming and DeMets, 1996; Gotzsche *et al*, 1996).

There are a number of specific issues surrounding the use of surrogate outcomes in HTA, the first being the appropriate definition to use within this context, i.e. what meets the definition of surrogate outcome in the context of a technology appraisal? According to the US National Institutes of Health Biomarkers Definitions Working Group, a surrogate outcome is a biomarker intended to substitute for a clinical endpoint, which is ”a characteristic or variable that reflects how a patient feels or functions, or how long a patient survives” (Biomarkers Definitions Working Group, 2001). For example, the biomarkers of HbA_{1c} and LDL-cholesterol have been accepted in licensing as surrogate outcomes for risk of diabetes complications and cardiovascular disease respectively. However, a broader surrogate outcome definition is needed in the context of HTA and reimbursement that includes not only biomarkers but also what might be regarded as intermediate measures of health outcome. A common example seen in NICE appraisals is the use of the intermediate outcome of progression (or disease-free) survival to predict overall mortality (the final outcome) in cancer (Sargent *et al*., 2005; Bowater *et al*, 2008). Bone mineral density is often used as the surrogate in licensing decisions for osteoporotic treatments. However, in the context of a cost-effectiveness analysis, hip fracture risk (an intermediate outcome) may also be regarded as surrogate outcome in that it is used to substitute (and predict) for the principal health benefits related to the treatment, namely survival and HRQoL (Stevenson *et al*, 1995). Clarification at the scoping stage of an appraisal as to which outcomes are surrogate is important to inform future technical and methodological discussions for that appraisal.

A second issue is the assessment of the validity of the surrogate outcome, i.e. in a technology appraisal what evidence should be used to assess whether a proposed outcome can reasonably accepted as a surrogate outcome (or not)? A large literature has been written about the validation of surrogate outcomes,

particularly in terms of statistical approaches. In brief, three broad validity criteria have been proposed (Bucher *et al*, 1999; Lassere, 2008; Elston and Taylor, 2009):

(1) biological reasoning – is there evidence of biological plausibility of relationship between surrogate and final outcome (from pathophysiological studies and/or understanding of the disease process)?

(2) epidemiological evidence – is there evidence demonstrating a consistent association between surrogate outcome and final outcome (from epidemiological/observational studies)?

(3) trial-based evidence – is there evidence demonstrating treatment effects on the surrogate correspond to effects on final outcome (from clinical trials)? Trial-based evidence is usually not available for the specific technology in question so instead this evidence is sourced from another technology within the same class or a different technology class.

Several statistical methods have been proposed to assess these criteria, particularly for trial-based evidence (for review see Weir and Walley, 2006).

In order to appropriately assess the validity of proposed surrogate outcome in the context of a technology appraisal, a recent HTA review of surrogate outcomes has proposed that a systematic review of the evidence for each of these three criteria is needed (Elston and Taylor, 2009).

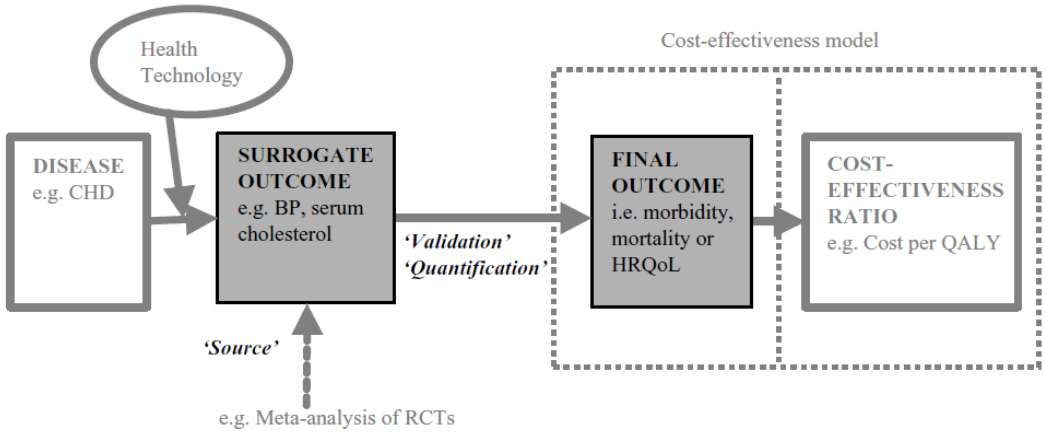
In a technology appraisal it might be expected that for an outcome to be deemed a 'valid' surrogate, it should fulfil each of the above three criteria. However, as there is currently no consensus in the HTA community on the minimum level of evidence for validation of surrogate outcomes, it could more conservatively be argued that these criteria instead need to be considered on a case-by-case basis.

The final issue relates to the prediction and *quantification* of the surrogate-final outcome relationship and how this is captured in the cost-effectiveness analysis, i.e. in a technology appraisal, how is the treatment effect on

surrogate outcome used to predict the final outcome and, thus, assess the incremental cost per QALY? As outlined above, various statistical approaches have been used to validate surrogate outcomes. In doing so, these methods effectively quantify the relationship between the treatment effect on surrogate and final outcome. For example, regression-based methods can use trial level data (meta-regression) or individual patient data from a single trial or combination of both (e.g. Johnson *et al*, 2009; Molenberghs *et al*., 2002).

Economic modelling typically involves extrapolating the clinical effectiveness evidence in order to estimate QALYs, e.g. extrapolation of trial-based observed mortality or attributing utility values to cardiovascular events observed in the trial. In doing so, such models are used to set out the potential relationship(s) between surrogate/intermediate and final endpoints (this is part of what makes them models). As such the role of surrogates is relevant to any NICE appraisal. However, in appraisals where the clinical effectiveness evidence is based only (or principally) on a surrogate outcome there is an additional element of uncertainty specifically associated with the prediction of the (unobserved) final outcome (typically survival or HRQoL) (see Figure 1). There may or may not be evidence to support such relationships. The impact of this uncertainty on cost-effectiveness needs to be fully explored, such as through the extensive use of sensitivity analyses (Elston and Taylor, 2009).

Figure 1. Schematic representation of the use of a surrogate in an HTA cost-effectiveness model (from Elston & Taylor, 2009)



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

There is limited discussion on the use of surrogate outcomes in the current Methods Guide.

In the 'Suppliers of evidence, commentary and analysis' section, the methods guide says:

> *4.4.3 The written submissions […] include evidence that relates to some or all of the following. […] The identification of appropriate outcome measures and the appropriate use of surrogate outcome measures.*

In the 'Modelling methods' section, it states:

> *5.7.2 Situations when modelling is likely to be required include those where […] intermediate outcomes measures are used rather than effect on HRQoL and survival*

Furthermore, the definition of 'intermediate outcome' is given in the Glossary:

'Intermediate outcome: Outcomes that are related to the outcome of interest but may be more easily assessed within the context of a clinical study; for example, blood pressure reduction is related to the risk of a stroke.'

The methods guide also adds a 'process' consideration:

> *2.2.6 As far as possible, principal measures of health outcome are identified in the scope. For the valid analysis of clinical effectiveness, the principal outcome(s) will be clinically relevant; that is, they measure health benefits and adverse effects that are important to patients and/or their carers.*

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.

- Which definition of surrogate outcome is most suitable in the technology appraisal context?
    - Should NICE's definition of surrogate outcomes be limited to biomarkers or should they include a wider category of intermediate health outcomes (e.g. fracture rate, progression free survival)?
    - Should the scoping exercise be used to clarify if the clinical effectiveness evidence in support of a technology appraisal is likely to be based principally on a surrogate outcome?

What are the potential consequences of a revision of the classical definition of surrogate outcomes in the HTA context?

- Should the methods guide require a review of the evidence to support the use of a surrogate outcome in place of a final outcome during the appraisal?
    - Does this review of evidence have to be systematic?
    - Should there be a minimum level of evidence for an outcome to be accepted as a surrogate and thereby inform the estimation of a technology's clinical effectiveness and cost-effectiveness? Should specific statistical approaches to surrogate validation be recommended/prescribed?

What could be the impact of always requesting a synthesis of evidence for the use of the surrogate outcomes in the technology appraisal process? What could be the impact of specifying a minimum level of evidence needed?

- *Should there be an explicit quantification of the uncertainty related to the use of surroga*te outcomes in the cost-effectiveness analysis?

- How should this uncertainty be estimated and presented?

What could be the impact of always requesting an explicit quantification of the uncertainty around the relationship between the surrogate and the final outcomes?

4 References

Bucher, HC, Guyatt, GH, Cook, DJ, Holbrook, A, McAlister, FA. (1999) Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA*, 282, 771-8.

Biomarkers Definitions Working Group (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther,* 69, 89-95.

Bowater RJ, Bridge LJ, Lilford RJ. (2008) The relationship between progression-free and post-progression survival in treating four types of metastatic cancer. *Cancer Letters*, 272,48-53.

Elston J, Taylor RS (2009) Use of surrogate outcomes in cost-effectiveness models: a review of United Kingdom health technology assessment reports*. Int J Technol Assess Health Care*, 25, 6-13.

Fleming TR, DeMets DL. (1996) Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med*, 125:605-613.

Gotzsche PC, Liberati A, Torri V, Rossetti L. (1996) Beware of surrogate outcome measures. *Int J Technol Assess Health Care*, 12:238-246.

Johnson K, Freemantle N, Anthony D, Lassere M (2009) LDL-cholesterol differences that predicts benefit in statin trials as determined by the surrogate threshold*. J Clin Epidemiol*, 63, 328-336.

Lassere MN. (2008) The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based,

quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med* Res, 17, 303-40.

Molenberghs G, Buyse M., Geys H, Renard D, Burzykowski T, Alonso A. (2002) Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Controlled Clinical Trials,* 23, 607-625

Pharmaceutical Advisory Benefits Committee (PBAC). *A framework for evaluating proposed surrogate measures and their use in submissions to PBAC*. December 2008 Available at: http://www.pbs.gov.au/industry/useful-resources/PBAC_feedback_files/STFOWG%20paper%20FINAL.pdf

Sargent DJ, Wieand HS, Haller DG *et al*. (2005) Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Onc.* 23:8664-8670

Stevenson M, Jones ML, De Nigris E, Brewer N, Davis S, Oakley J. (2005) A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis. *Health Technol Assess* 9:1-160.

Weir, CJ, Walley, RJ. (2006) Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med,* 25, 183-203.

5 Authors

This briefing paper has been prepared by Oriana Ciani and Rod Taylor of the Peninsula Technology Assessment Group (PenTAG), Peninsula College of Medicine & Dentistry, University of Exeter. Thanks to Paul Tappenden for commenting on the document.

11th November 2011