Use of surrogate outcomes in cost-effectiveness models: A review of United Kingdom health technology assessment reports[1]

Julian Elston

Peninsula Medical School, Universities of Exeter and Plymouth and
Devon Primary Care Trust

Rod S. Taylor

Peninsula Medical School, Universities of Exeter and Plymouth

Objectives: The aim of this study was to explore the use of surrogate outcomes—a substitute outcome that predicts final patient-related outcomes—in cost-effectiveness models (CEM) within health technology assessment (HTA) reports and provide guidance for their future use.
Methods: Our sampling frame was all UK HTA Program monograph series reports published in 2005 and 2006. Reports were included if they addressed a treatment effectiveness/efficacy question and included a CEM based on a surrogate outcome. The two authors independently applied inclusion and exclusion criteria, and the following data was extracted from included reports: source of surrogate outcome, level of evidence for validation of the surrogate outcomes, methods used in report to quantify link between surrogate outcome and final outcome, and consideration of the uncertainty associated with using surrogate outcomes in the results or conclusions of report.
Results: Of 100 HTA reports, 35 complied with the inclusion criteria. Of these, four (11 percent) reports included a CEM based on a surrogate outcome. All four reports sourced treatment-related changes in surrogate outcome through a systematic review of the literature. One provided Level 1 surrogate evidence (randomized controlled trial data showing a strong association between the change in surrogate outcome and change final outcome); two reported Level 2 evidence (observational study data); and one provided Level 3 evidence (disease natural/ history data). The transparency of quantification and exploration of uncertainty of the surrogate and final outcome relationship varied considerably across all four reports.
Conclusions: Recommendations are made for the use of surrogate outcomes in future HTA reports.

Keywords: Surrogate outcomes, Intermediate outcomes, Biomarkers

One of the most important factors in determining the duration, size, and cost of a clinical trial of a new or existing treatment is the choice of outcome. Ideally, decisions on the use of treatment should be based on well-conducted randomized controlled trials (RCTs) that assess clinically important "final" patient-relevant outcomes. That is, outcomes of which the patient is aware and wants to avoid, for example, death

or morbid end-points (such as myocardial infarction, stroke, or impaired quality of life) (3;13).

However, conducting trials with final patient-relevant outcomes can require a very large sample size and/or periods of long follow-up if differences in outcome are to achieve statistical significance, particularly in case of chronic diseases. To overcome these practical limitations other end points can be used to substitute, or act as a "surrogate" for the final outcome. This may lead to shorter studies and, therefore, faster time to licensing and dissemination of new treatments. In conditions where a patient's risk of serious morbidity or mortality is high and/or their illness is rare, use of surrogate outcomes provide an attractive option when it comes to approval of new treatments for market access. Some common surrogate outcomes that have been used to gain regulatory approval include: CD4 count, (tumor) progression free survival, prostatic specific antigen, blood pressure, cholesterol level, intraocular pressure and bone density. Two key tenets of a surrogate outcome are that it represents an end-point that is intended to *substitute* and to be *predictive* for a final patient-relevant clinical outcome (12).

However, the use of surrogate outcomes in trials is controversial. Their use, at least in some applications, has led to erroneous or even harmful conclusions (4;5). Despite their potential appeal, and success in some areas, there are potential risks in using surrogate outcomes. Fleming and DeMets catalogued several examples from the cardiovascular, AIDS, orthopedic and infectious disease literature where surrogates have failed to be an effective substitute for the final outcome, that is, an improvement in the surrogate outcome was not linked with an equivalent improvement in final outcome (4). More recently, Ridker and Torres reviewed the various characteristics of 324 consecutive cardiovascular trials published in three major general medical journals (*JAMA*, *The Lancet*, and *N Engl J Med*) between January 2000 and July 2005 (10). The authors found that trials reporting a surrogate outcome as a primary outcome were more likely to report a positive treatment effect (77 of 115 trials [67 percent]) than those trials that reported a final patient-related primary outcome (113 of 209 trials [54 percent], $p = .02$).

These reviews suggest that the use of surrogate outcomes in health technology assessment (HTA) may lead to two types of error: (i) a conclusion that a new treatment has greater health benefit than risk when the opposite is true ("false positive"); (ii) an overestimate of the true level of benefit of new treatment ("bias"). Furthermore, at least theoretically, the use of a surrogate outcome could lead to a false negative or underestimate of treatment effect. Where possible, policy makers and HTA analysts would seek to avoid such errors. Nevertheless, little guidance is currently available on what might be deemed as the "appropriate use" of the surrogates in cost-effectiveness models (CEMs) in HTA (12). The aim of this study was to explore the use of surrogate outcomes in CEMs within UK HTA Program reports and by doing so

provide a basis for guidance for their future use, validation and reporting.

METHODS

We sampled UK HTA Program monograph series reports published in 2005 and 2006. This period was chosen to reflect recent HTA practice and limited to two years because of time and resources available for this project.

Reports were included if they addressed a treatment effectiveness/efficacy question, included CEM and the CEM was primarily based on a surrogate outcome. Reports addressing a diagnostic, screening, etiology, prognostic or methodological question were excluded.

We developed a conceptual framework for surrogate outcomes in an HTA cost-effectiveness model (see Figure 1). This was used as the basis for developing a structured proforma which contained a series of questions to be addressed to each report. This approach also ensured the consistent application of selection criteria. The proforma was piloted on five HTA reports. Piloting identified that it was not always possible to judge whether the CEM in an HTA report was based on a surrogate outcome. We initially used the U.S. NIH Biomarker Group definition of surrogate end point—"a biomarker that is intended to substitute for a clinical (final) outcome, and that a surrogate end point is expected to predict clinical benefit" (1). However, this definition was difficult to operationalize in practice, as the outcomes used in HTA reports were often not what could be described as a "biomarker" but rather a patient-related end point. A pragmatic approach was, therefore, taken which permitted such reports to be included if they otherwise fulfilled the definition of a surrogate outcome (i.e., substitution and prediction of a final outcome). The inclusion and exclusion criteria were applied independently to all reports by the two authors (J.E. and R.S.T.) and discrepancies resolved by discussion. The following categories of information were extracted by one of the authors using a standardized proforma (and checked by the second):

- Characteristics of report (i.e., type of technology, disease area, and whether the report was on behalf of the National Institute for Health and Clinical Excellence [NICE]).
- Overview of the CEM (i.e., type of model and base case incremental cost-effectiveness ratio(s)).
- Characteristics of surrogate outcome used in CEM and identification of derived final outcome.
- Source of surrogate outcome evidence used in CEM (e.g., systematic review of clinical trials).
- Evidence base for validating surrogate outcome, rated according to our three-level hierarchical system developed using the U.S. NIH Biomarkers Definitions Working Group framework (1) and ICH-9 guidelines (see Figure 2) (6).
- Methods used in report to quantify the link between surrogate outcome and final outcome (e.g., regression based approach).
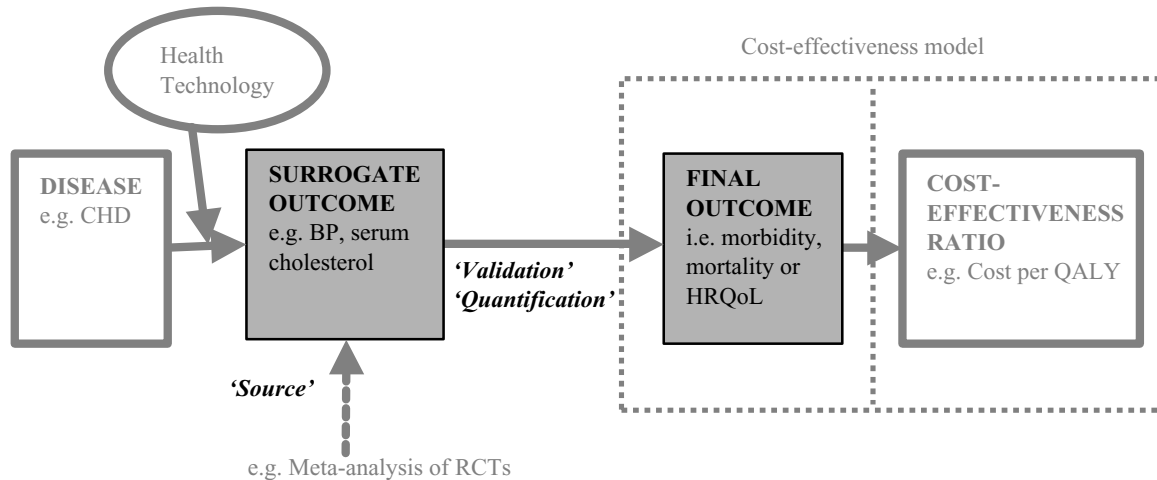
Figure 1. Schematic representation of the use of a surrogate outcome in a HTA cost-effectiveness model. Key: "Source" refers to where the surrogate outcome data has been derived (usually by a systematic review/meta-analysis of clinical effectiveness literature in an HTA); "Validation" refers to the evidence provided supporting the relationship between the surrogate and final outcome; "Quantification" refers to how this relationship has been quantified. The two dotted boxes illustrate that the quantification of the surrogate to final outcome may take place either within or outside the cost-effectiveness model per se.

Hierarchical level	Evidence requirement	Source of evidence
Level 1	Treatment effects on the surrogate correspond to effects on the patient-related outcome	Clinical trial(s) showing that change in surrogate outcome with treatment is associated with a commensurate change in final patient-related outcome
Level 2	Consistent association between surrogate outcome and final patient-related outcome	Epidemiological (observational) studies demonstrating an association between the surrogate outcome and final related outcome
Level 3	Biological plausibility of relationship between surrogate and final patient-related outcome	Pathophysiologic studies and understanding of disease process

Figure 2. Level of evidence used to validate surrogate outcomes. This hierarchy was developed primarily from the ICH-9 guidelines (6) and US NIH Biomarkers Definitions Working Group framework (1). To fulfill the evidence requirement for Level 1 or 2 necessitates the fulfillment of the requirements of the lower levels of evidence.

- Consideration of the uncertainty associated with using surrogate outcomes in the results or conclusions or elsewhere in the report.

This was supplemented with a narrative analysis which focused on how surrogate validation, quantification and uncertainty relating to their use was reported in the text. These were tabulated to aid comparison, with quotations used to illustrate specific points.

In addition, we evaluated the adequacy of surrogate outcomes using the criteria and scoring schema proposed by: (i) Journal of the American Medical Association (*JAMA*) User's Guide to the Medical Literature series XIX: use of Surrogate End-points (2); and (ii) Outcomes Measures in Rheumatology Clinical Trials (OMERACT) Biomarker and Surrogate Endpoint Evidence Schema (7). These two key publications

assess the strength of evidence that links a surrogate outcome and a final outcome. The *JAMA* Guide evaluates evidence for surrogates according to three levels, depending on whether it is observational or trial-based (within or outside drug class), while the OMERACT scheme awards points for type of surrogate, study design, and statistical strength of evidence base, with penalties for contradictory evidence (see Tables 1a and 1b).

RESULTS

A total of 100 UK HTA reports were published between 2005 and 2006. The process of report selection is summarized in Figure 3. Thirty-three reports were initially excluded as they either addressed a methodological or
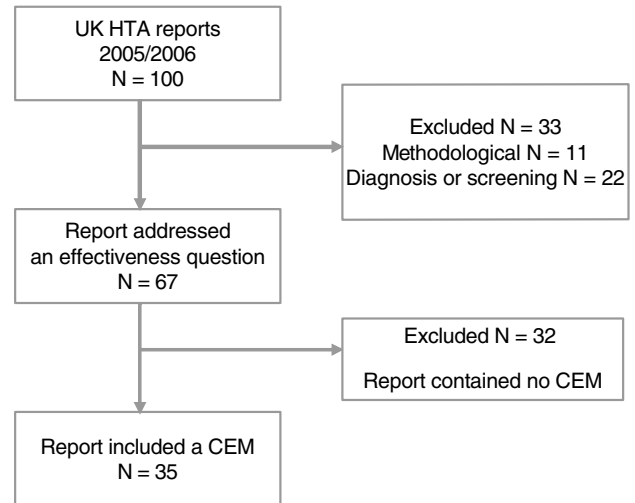
Table 1a. JAMA Guide for Surrogate Outcomes[a]

Guide	Requirement questions
1	Is there a strong, independent, consistent association between the surrogate outcome and the final outcome?
2	Is there evidence from randomized controlled trials in other drug classes that improvement in the surrogate outcome has consistently led to improvement in the final outcome?
3	Is there evidence from randomized controlled trials within the same drug class that improvement in the surrogate outcome has consistently led to improvement in the final outcome?

[a]Modified from Table 1, Bucher et al. (1999) (2).

Figure 3 Selection of reports.

diagnosis/screening question. Of the remaining sixty-seven HTA reports a further thirty-two were excluded as they did not contain a CEM. Details of the remaining thirty-five HTA CEM reports are summarized in Supplementary Table 1 (which can be found at www.journals.cambridge.org/thc). Following a detailed review of these reports, four (11 percent) were identified as using an outcome in the CEM based on prediction of a different end point reported in the clinical effectiveness section (Woodroffe et al, 2005 (14); Loveman et al, 2006 (8); Shepherd et al, 2006 (11); Yao et al, 2006 (15)). These four reports were, therefore, judged to be examples of the use of surrogate outcomes.

Characteristics of Report and Surrogate Outcome

Woodroffe et al. (2005) (14). This report examined the clinical and cost-effectiveness of several new immunosuppressive therapies (tacrolimus, basiliximab, daclizimab,

mycophenolate mofeitil, and sirolimus) compared with existing therapy (ciclopsorin and azathrioprine) in adults undergoing kidney replacement. A systematic review identified a total of 33 randomized controlled trials across the various drugs comparisons. Most trials were short-term (≤12 months) and therefore were of insufficient sample size and duration to detect differences between drugs in terms of relevant patient-related final outcomes, that is, survival of the kidney graft and patient mortality. However, virtually all trials reported biopsy confirmed acute rejection (BPAR). BPAR was used as surrogate outcome to predict graft survival.

Yao et al. (2006) (15). This sister report to the study by Woodroffe et al. examined the clinical and cost-effectiveness of several new immunosuppressive therapies in children. The same group of drugs were compared, and the

Table 1b. OMERACT Surrogate Schema Scoring

Domain	Woodroffe et al. (2005) (14)	Yao et al. (2006) (15)	Shepherd et al. (2006) (11)	Loveman et al. (2006) (8)
A. Target[a]	4 (renal graft survival)	4 (renal graft survival)	4 (chronic hepatitis, liver cancer)	4 (need for full-time care)
B. Study design	2 (at least one pre-specified population based study)	2 (at one least pre-specified population based study)	0 (review of disease natural history)	2 (at least one prespecified population based study)
C. Statistical strength	3 (at least good association between marker change and target change in all individual studies)	3 (at least good association between marker change and target change in all individual studies)	0 (no relevant data)	3 (at least good association between marker change and target change in all individual studies)
Penalties	0	0	0	0
Overall score	9	9	4	9

[a]Where a Target score of '4' represents "at least one patient-centered target of irreversible organ morbidity or major irreversible clinical burden of disease."

systematic review identified 14 RCTs and non-RCTs. As in the report by Woodroffe et al., BPAR was used as surrogate outcome to predict graft survival.

Shepherd et al. (2006) (11). This study assessed the clinical effectiveness and cost-effectiveness of two antiviral agents (adefovir dipivoxil [ADV] and pegylated interferon alfa-2a [PEG]) for the treatment of adults with chronic hepatitis B infection. The report's systematic review identified seven RCTs that assessed the effectiveness of ADV and three trials evaluating the effectiveness of PEG. These trials reported treatment effects as short-term biochemical response (e.g., levels of alanine aminotransferase for liver function), virological response (e.g., presence of HBV DNA as evidence of viral replication), and seroconversion (e.g., HBeAg loss/anti-HBe; HBsAg loss/anti-HBs). The authors used seroconversion rates as a surrogate outcome in a transition natural history model to predict liver cirrhosis, liver cancer, liver transplant, and death.

Loveman et al. (2006) (8). The study assessed the clinical and cost-effectiveness of new drugs (donepezil, rivastigmine, galantamine, and memantine) for Alzheimer's disease. A total of twelve RCTs were included. The four drugs were shown to be effective when assessed by cognitive function outcome measures, that is, Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog) score. ADAS-cog score was used as surrogate by the authors to predict the outcome of needing full-time care.

Validation of Surrogate Outcomes

Woodroffe et al. provided evidence from a systematic review and meta-analysis of observational studies to demonstrate the relationship between BPAR (surrogate outcome) and graft survival, that is, Level 2 evidence.

A key assumption in the cost-effectiveness modelling framework of this review is the linkage between BPAR, graft and patient survival, quality of life and costs. The selection of acute rejection is supported by a systematic review of potential prognostic predictors for graft survival. (pg. 68) (14)

Yao et al. updated this systematic review to include evidence in children. To limit bias and confounding, the authors restricted the systematic review to observational studies with multivariate analyses with 5-year or longer follow-up. The authors identified one of two studies in children that confirmed the relationship between the surrogate (BPAR) and final outcome (graft survival) – Level 2 evidence.

In summary, this updated review of surrogate outcome predictors in children appears to support the findings that acute rejection is a strong predictor of future graft loss. (pg. 7) (15)

In addition, Yao et al. examined if the surrogate and final outcome relationship held up in a trial setting:

To investigate the level of extrapolation between observational data and RCTs for this review, we compared the change in surrogate levels to the change in graft survival seen in the paediatric RCT by Filler and colleagues (pg. 7) (15)

and found that:

In this trial, an improvement in 2-year graft survival with tacrolimus (p = .04) was associated with improvements in both GFR and the incidence in acute rejection at 6 months to 1 year in the tacrolimus group. (pg. 7) (15) that is, Level 1 evidence

The report of Shepherd et al. (2006) recognized the limitation of the outcomes assessed in the trials and the need to predict a more final outcome of chronic hepatitis B (CHB).

Clinical trial data relating to the effectiveness of interventions included in this appraisal are limited to measurements of short-term serological, virological and histological changes. In order to estimate the impact of these intermediate effects on final outcomes for patients, a natural history model for CHB was required. (pg. 81) (11)

The authors developed a Markov state transition disease model following a literature search on the natural history and epidemiology. This epidemiological data was judged to represent Level 3 evidence.

The principal effect of antiviral treatment is to change patients' serological, biochemical, histological or virological status to place them in health states where they are less likely to develop progressive liver disease. (pg. 82) (11)

Loveman et al. based their decision to use cognitive function as a predictor for full time care on a previously developed cost-effectiveness model for Alzheimer's disease. The authors state that the relationship between cognitive function and full-time care is based on individual patient data analysis undertaken by the developers of the economic model. On checking this source reference, the study concerned was identified to be a cohort comparison of cognitive function outcome and full-time care in Alzheimer's disease, that is, Level 2 evidence.

Surrogate Quantification

A range of approaches was used to quantify the relationship between the surrogate and final outcome across the four reports. The CEM in both the Woodroffe and Yao reports used a hazard ratio (derived from a systematic review of observational studies examining the patient-level relationship between the surrogate [BPAR] and final outcome [graft survival]) to numerically represent this relationship.

The authors reported that the pooled hazard ratio (HR) for allograft survival based on an acute rejection episode was 1.95 (95 percent confidence interval (CI), 1.42 to 2.67). (pg. 6) (15)

The adult BSA model was adapted for paediatrics. . .[and]. . .use made of a paediatric-specific HR of 1.41, 95 percent CI, 1.15 to 1.74 (pg. 43) (15).

Shepherd et al. assessed the relationship between seroconversion rate (surrogate outcome) and final outcomes (e.g., chronic hepatitis, liver cancer) within a natural history CEM. The link between the surrogate and final outcomes was quantified as transition probabilities within this model.

Loveman quantified the impact of cognitive function (surrogate outcome) on full-time care (final outcome) using a predictive risk equation developed for the economic model. This equation was developed using a Cox proportional hazards model and contains coefficients for cognitive function, age at disease onset and the presence of psychotic symptoms and extrapyramidal syndromes and treatment duration.

Handling Uncertainty

Woodroffe et al. flagged the link between surrogate outcome (BPAR) and final outcome (graft survival) in their model as a potential limitation in the discussion.

In contrast, certain limitations were placed on the review. . .to estimate long-term effectiveness (and cost-effectiveness), extrapolation from trial 1-year BPAR to graft survival was undertaken (pg. 68) (14)

and in the executive summary of their report:

The absence of both long-term outcome and quality of life from trial data makes assessment of the clinical and cost-effectiveness on the newer immunosuppressants contingent on modelling based on extrapolations from short-term trial outcomes. (pg. xi) (14)

Yao et al. took a quantitative approach to handling the uncertainty associated with the use of a surrogate outcome in their CEM. Using sensitivity analysis, they explored how the incremental cost-effectiveness ratio (ICER) would alter when varying the hazard ratio for the relationship between the surrogate and final outcome. Furthermore, in the report's discussion the authors raise the dependence on surrogate outcome as a specific limitation of the CEM.

Surrogate outcomes – The short duration of follow-up of RCTs necessitated the prediction of long-term graft loss [final outcome] and all cause mortality from 1-year BPAR [surrogate outcome] (pg. 55) (15).

Shepherd et al. quantified the impact of uncertainty associated with the use of surrogates through sensitivity analyses varying the assumptions of the structure of the CEM such

as setting the transition probability between seroconversion and liver cancer to zero.

Also through sensitivity analysis, Loveman et al. assessed the impact of a one-point shift (in both directions) for the surrogate outcome (ADAS-cog). Furthermore, in the discussion section, the authors highlighted the limitation of the use of surrogate outcomes.

It is difficult to know what the changes [in cognitive function] demonstrated on each measure really mean. (pg. 14) (8).

OMERACT Scoring Schema and Adapted JAMA Criteria

The scoring on the OMERACT surrogate schema domains for the four reports is summarized in the Table 1b. The maximum potential OMERACT score is 15. The low score for the Loveman et al. report (four) reflects that although the authors "embedded" the relationship between seroconversion (surrogate outcome) and chronic hepatitis/liver cancer (final outcome) in their disease history CEM, they did not present specific biological or epidemiological evidence to support this link. The reports of Woodroffe, Loveman, and Yao each scored nine. All reports failed to meet the threshold score of ≥10 that schema's authors deemed to represent the minimum level of evidence that an end point should reach to support its use as a surrogate outcome.

The studies of Woodroffe, Loveman, and Yao were judged to meet the JAMA criteria for surrogate validation to level Guide 1. However, only the report by Yao et al. attained level Guide 2—the minimum requirement for surrogate validation (equivalent to Level 1 in our framework).

DISCUSSION

Of a total sample of 100 HTA UK reports published between 2005 and 2006, 35 addressed an effectiveness/efficacy question and contained a CEM. Of these, four (11 percent) reports were found to have based their cost-effectiveness analysis on a surrogate outcome—two reports in patients undergoing kidney transplant using biopsy-confirmed acute rejection (BPAR) outcome (final outcome—graft survival) (11;15); one report on Alzheimer's disease using cognitive function score (final outcome—need for full time care) (8); and one report on chronic hepatitis B using seroconversion (final outcome—chronic hepatitis/liver cancer) (11). All four reports sourced treatment-related changes in surrogate outcome through a systematic review of the literature, in some cases also undertaking meta-analysis. However, there was some variability in the consistency and transparency by which these reports provided evidence of the validation for the surrogate/final outcome relationship. Most usefully some reports used sensitivity analyses to explore the impact of the potential uncertainty of the surrogate to final outcome relationship on cost-effectiveness. Only one of the reports

undertook a systematic review to specifically seek the evidence base for the surrogate/final outcome link (15). Furthermore, this was the only report to provide Level 1 surrogate/final outcome validation evidence, that is, RCT data showing a strong association between the change in surrogate outcome (BPAR) and change final outcome (graft survival) at an individual patient level. It was also the only outcome in reports considered to be a valid surrogate when assessed using the JAMA criteria. Two of the other three reports reported Level 2 evidence, that is, observational study data showing the relationship between the surrogate and final outcome (8;14). By contrast, none of the reports achieved a sufficient score on the OMERACT schema to be judged to have acceptable evidence of a surrogate outcome by its authors. Having only been recently developed, the OMERACT schema requires further testing against a range of surrogate outcomes to fully assess its suitability as a practical tool.

It is interesting to note that the four reports based on the use of surrogate outcomes were all undertaken on behalf of NICE whose reference case seeks a cost per quality-adjusted life-year (QALY) analysis (9). This might reflect a pressure on HTA analysts to extrapolate from surrogate outcomes to QALYs to formally quantify the cost-effectiveness of a health technology when undertaking work directly for policy makers.

Strengths and Limitations of Study

We believe this to be the first empirical study of the use of surrogate outcomes in CEMs in HTAs. Previous surrogate outcome surveys have focused on their use in clinical trials and, unlike this study, often used a purposive sampling strategy to identify examples that have led to surrogate failure (4;5). This report highlights several issues relating to the use of surrogate outcomes in CEM in HTAs (and elsewhere), including definitional uncertainty, an inconsistent approach to surrogate identification and validation and a lack of recognition of the uncertainty surrounding their use.

However, because of limited time and resources, the sample of HTA reports surveyed was relatively small and restricted to the UK. The small sample size and the limited number of HTA reports with a CEM based on a surrogate outcome, potentially limits the generalizability of the findings of this study. The report focused on inclusion of HTA reports where there was clear evidence of the dependence of the CEM on a surrogate outcome. We may have, therefore, excluded reports that used surrogate outcomes but were unclear about this in their CEM description or where the CEM depended on a mix of final and surrogate outcomes, or reports or outcomes that were not clearly surrogates (in terms of the operational definition of this review). Documentary analysis was used to assess the content of included reports. It is, therefore, important to acknowledge that the absence of an issue in the text does not necessarily suggest the absence of consideration of that issue by the report's authors, or, for

example, that there is no evidence for surrogate validation (especially if a systematic review to identify studies linking surrogate and final outcomes was not undertaken). Thus, OMERACT scores may not reflect the "true score." Finally, for the purposes of this report we have focused on identifying HTA reports with a CEM that have used definitive examples of surrogate outcomes. However, we recognize that rather than a dichotomy there is effectively a continuum between what might be regarded as "true" surrogate outcomes and "true" final outcomes. Nevertheless, we would contend that the recommendations remain applicable.

Recommendations for the Use of Surrogate Outcomes in HTA Reports

Recommendations were formulated from the findings of a review of literature on the use of surrogate outcomes (12), the experience of this survey and feedback and discussion on a draft of the recommendations with the UK HTA groups who undertake technology assessment reports commissioned by NIHR HTA program and the NICE technology assessment team.

These recommendations are intended to act as a list of considerations that policy makers and HTA analysts should take into account when faced with the use of surrogate outcomes in cost-effectiveness models in HTA reports. It is acknowledged that the practicality and resource implication of implementing these recommendations has not been formally tested within this project.

RECOMMENDATION I

Ideally, the assessment of clinical effectiveness and cost-effectiveness of a health technology should be based on final patient-related outcomes (i.e., mortality, important clinical events, and health-related quality-of-life). To minimize the risk of bias, this evidence should be identified from a systematic review (and meta-analysis) of well-conducted randomized clinical trials.

RECOMMENDATION II

Where this is not possible and there is a requirement to use a surrogate outcome, the following should be undertaken: (i) A review of the evidence for the validation of the surrogate/final outcome relationship. To minimize the risk of bias, such a review should be systematic. (ii) The evidence on surrogate validation should be presented according to an explicit hierarchy such as the following:

Level 1: evidence demonstrating treatment effects on the surrogate correspond to effects on the patient-related outcome (from clinical trials);

Level 2: evidence demonstrating a consistent association between surrogate outcome and final patient-related outcome (from epidemiological/observational studies);

Level 3: evidence of biological plausibility of relationship between surrogate and final patient-related outcome (from pathophysiologic studies and/or understanding of the disease process).

(iii) Consideration for undertaking a CEM analysis based on a surrogate outcome when there is Level 1 or 2 validation evidence.

RECOMMENDATION III

When a CEM analysis based on a surrogate outcome is undertaken: (i) Provide a transparent explanation as to how the relationship of the surrogate and final outcome is quantified within the CEM. (ii) Explicitly explore and discuss the uncertainty associated with use of the surrogate outcome in the CEM, especially through sensitivity analysis. In accord with recent HTA methodological developments, such uncertainty may be quantified using probabilistic sensitivity analysis. (iii) Make specific research recommendations regarding the need for future research on the surrogate/final outcome relationship. In accord with recent HTA methodological developments, the impact of the surrogate outcome on decision uncertainty may be quantified by a value of information analysis. (iv) Include the term "surrogate outcome" in the report executive summary/abstract to assist bibliographic identification.

CONTACT INFORMATION

Julian Elston, PhD (julian.elston@pms.ac.uk), Honorary Research Fellow, Institute of Health Service Research, Peninsula Medical School, Universities of Exeter and Plymouth, Noy Scott House, 3rd Floor, Barrack Road, Exeter EX2 5DW, UK; (julian.elston@nhs.net), Academic Specialty Registrar in Public Health, Department of Public Health, Devon Primary Care Trust (PCT), County Hall, Topsham Road, Exeter EX2 4QL, UK

Rod S. Taylor, PhD (rod.taylor@pms.ac.uk), Associate Professor in Health Services Research, Department of Primary Care, Peninsula Medical School, Universities of Exeter and Plymouth, Noy Scott House, 3rd Floor, Barrack Road, Exeter EX2 5DW, UK

REFERENCES

1. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: Preferred and conceptual framework. *Clin Pharmacol Ther.* 2001;22:485-502.
2. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. 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.* 1999;282:771-778.
3. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess.* 1998;2:1-7.
4. Fleming TR, DeMets DL. Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med.* 1996;125:605-613.
5. Gotzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *Int J Technol Assess Health Care.* 1996;12:238-246.
6. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised tripartite guideline. *Statistical principles for clinical trials* E9. February 1998. http://www.ich.org/LOB/media/MEDIA485.pdf. Accessed October 17, 2008.
7. Lassere MN, Johnson KR, Boers M, *et al.* Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol.* 2007;34:607-615.
8. Loveman E, Green C, Kirby J, *et al.* The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease. *Health Technol Assess.* 2006;10:1-160.
9. NICE. *Guide to the methods of technology appraisal.* London: National Institute for Health and Clinical Excellence; April 2008. http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf. Accessed October 17, 2008.
10. Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000-2005. *JAMA.* 2006;295:2270-2274.
11. Shepherd J, Jones J, Takeda A, Davidson P, Price A. Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: A systematic review and economic evaluation. *Health Technol Assess.* 2006;10(28):1-183.
12. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: A review of UK health technology assessment reports. *Health Technol Assess.* In press.
13. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health Qual Life Outcomes* 2006;4:79.
14. Woodroffe R, Yao GL, Meads C, *et al.* Clinical and cost-effectiveness of newer immunosuppressive regimens in renal transplantation: A systematic review and modelling study. *Health Technol Assess.* 2005;9:1-179.
15. Yao G, Albon E, Adi Y, *et al.* A systematic review and economic model of the clinical and cost-effectiveness of immunosuppressive therapy for renal transplantation in children. *Health Technol Assess.* 2006;10: iii-iv, ix-xi, 1-157.