available at www.sciencedirect.com

SCIENCE DIRECT®

journal homepage: www.elsevier.com/locate/jval

The CHD challenge: Comparing four cost-effectiveness models

David Turner, MSc[a,], James Raftery, PhD, MA[a], Keith Cooper, PhD, MSc[a], Eleanor Fairbank, MSc[a], Stephen Palmer, MSc[b], Sue Ward, BA[c], Roberta Ara, MSc[c]*

[a] *Wessex Institute University of Southampton, Southampton, UK*
[b] *Centre for Health Economics, University of York, York, UK*
[c] *School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK*

Keywords:
Coronary disease
Cost-utility analysis
Economic modeling
Model validation

ABSTRACT

Objectives: To compare four UK models evaluating the cost-effectiveness of interventions in coronary heart disease (CHD), exploring the relative importance of structure and inputs in accounting for differences, and the scope for consensus on structure and data.

Methods: We compared published cost-effectiveness results (incremental cost, quality-adjusted life year, and cost-effectiveness ratio) of three models conforming to the National Institute for Health and Clinical Excellence guidelines dealing with three interventions (statins, percutaneous coronary intervention, and clopidogrel) with a model developed in Southampton. Comparisons were made using three separate stages: 1) comparison of published results; 2) comparison of the results using the same data inputs wherever possible; and 3) an in-depth exploration of reasons for differences and the potential for consensus.

Results: Although published results differed by up to 73% (for statins), standardization of inputs (stage 2) narrowed these gaps. Greater understanding of the reasons for differences was achieved, but a consensus on preferred values for all data inputs was not reached.

Conclusions: We found that published guidance on methods was important to reduce variation in important model inputs. Although the comparison of models did not lead to consensus for all model inputs, it provided a better understanding of the reasons for these differences, and enhanced the transparency and credibility of all models. Similar comparisons would be aided by fuller publication of models, perhaps through detailed web appendices.

Introduction

Reimbursement agencies are continuously assessing new health-care technologies and need to ensure robustness, transparency, and consistency in their decisions. Health economic models can help inform these decisions and aid the efficient use of limited National Health Service (NHS) resources. However, if these models are to be used they must also be credible to those making policy decisions. This credibility can be hindered because these models are often complex and are the result of the different expertise of statisti-

Funding: The authors have no other financial relationships to disclose.

* *Corresponding author*. David Turner, Wessex Institute University of Southampton, Alpha House, Southampton Science Park, Chilworth, Southampton SO16 7NS, UK.
E-mail address: dturner@soton.ac.uk (David Turner).

1098-3015/$36.00 – see front matter
doi:10.1016/j.jval.2010.10.009

cians, clinicians, and health economists. There may also be a range of models that evaluate the same technology but use different methods and produce different results.

One way to increase the perceived credibility of a model is to test and demonstrate its validity. Four methods of validation have been suggested and these have been summarized by Philips et al. [1]. First, internal consistency implies that the practical model should behave as the theoretical model predicts, and that it is "debugged." Second, external consistency implies that the model should demonstrate face validity; that the outputs of a model are consistent with our knowledge of a disease or intervention. Third, between-model consistency implies that different independent models addressing the same question should give similar results. Fourth, predictive validity involves testing the results of a model against observable data or a prospective study to ascertain that the results are similar.

The fourth test of validity could be thought of as a "gold standard test" and would be the most credible evidence for the veracity of a model. If the results of a model matched real-world observations then we might have greater confidence. However, this is not generally possible because these models are often used to combine evidence from multiple sources, to extrapolate the results of a short-term clinical trial to the lifespan of patients or to generalize results to "real-world" settings. In these situations, data are unlikely to be available to formally assess the predictive validity of models and checking for between-model consistency may be the most feasible way of validating a model. If structure, inputs, and results are similar between models it implies there is general agreement on how to model a particular intervention or disease area; hence, it may mean that there is a clear preference for methods and data inputs. If there are disagreements between models in data inputs or structure then checking between-model consistency can highlight the important differences in terms of altering results. Effort can then be concentrated on those differences which may lead to either consensus over the use of existing data or more effort to generate better data. It is increasingly common to identify sets of alternative models that address a similar decision problem or disease area. For example, in a recent assessment of the cost-effectiveness of antivirals for the treatment of influenza, a total of 22 separate studies were identified (including 7 from a UK perspective) [2].The existence of alternative models provides an important opportunity to explore similarities and differences between models.

Differences between models can be due to differences in parameter values, methods, and structures. These are similar to the sources of uncertainty in models [3]. Parameter differences could include state transition probabilities and the quality-adjusted life year (QALY) losses associated with those states. Data on the costs of being in particular states and the costs of transition between those states could also be included in this category as could the data and assumptions used to characterize uncertainty in a probabilistic model. Many potential sources of data exist, particularly for a disease which is both common and well researched. Different data sets may be particularly suited for answering specific questions.

Differences between models may also be attributed to methodological differences including: the methods used to derive utility values, the perspective of the analysis (either NHS or a societal perspective), and the discount rate used. Some of these will be reduced by closer adherence to guidelines [1] and the National Institute for Health and Clinical Excellence (NICE) reference case [4]. Differences may also arise due to structural issues, including the modeling approach used (e.g., Markov, decision tree, or discrete event simulation). They may also be because of differences in the questions addressed or the health states included in models. Models may cover a single technology or may model a disease or population.

Some differences between models may be legitimate because two models set up to answer different questions may have different structures. Also, some differences may occur because analysts have correctly followed different sets of guidelines applicable to separate jurisdictions, or the analyses have been conducted at different times. Other differences may arise because there is no obvious "best" approach; these may require a need for clarification and future research to obtain more reliable or appropriate sources of data. Identifying these differences would be a useful outcome of any checks of between-model consistency.

To examine the feasibility and usefulness of a check of between-model consistency, we compared the Southampton CHD treatment model with three previously published models. The research question of the Southampton treatment model was "What are the relative cost-effectiveness ratios of a wide range of commonly used treatments for coronary heart disease for a UK population?" This involved using data on the clinical effectiveness and cost of a number of coronary care interventions and meant that comparisons could be made with many other CHD models, providing they addressed any of the interventions covered by the Southampton model. The Southampton treatment model was developed as part of a study that modeled CHD [5,6].

Comparator models were selected from the literature, restricting comparisons to models specific to the UK and conforming to the NICE reference case. Furthermore, we restricted comparisons to models covering one of the interventions evaluated by the Southampton treatment model. This meant that each model focused on NHS practice and followed similar guidelines for economic modeling. This increased the comparability of the models as there were a number of characteristics that would be shared; two examples are the use of a cost per QALY approach and a health service perspective. It also meant that there were differences between models published at different times.

One of the comparator models was developed at the School of Health and Related Research at the University of Sheffield to look at the cost-effectiveness of statins (School of Health and Related Research [ScHARR]-statins model) [7]. Specifically, this model was constructed to answer the question: "at what level of CHD/cardiovascular disease (CVD) risk are statins cost effective in the United Kingdom?" The other two comparator models were developed at the University of York. The York percutaneous coronary intervention (York-PCI) model [8] was designed to "explore the cost effectiveness of thrombolysis compared to primary angioplasty in acute myocardial infarction (MI) patients." The last model (York-clopidogrel) [9] was designed to "explore the cost-effectiveness of clopidogrel plus