A Critique and Impact Analysis of Decision Modeling Assumptions

Jonathan Karnon, PhD, Alan Brennan, MSc, Ron Akehurst, Hon MFPHM

Background. Numerous guidelines have been published defining good practice for the conduct of economic evaluations in general and model-based evaluations in particular. The extent to which guidelines are accepted is unknown, and the impact of deviations from good practice is not generally recorded. The authors identified 4 specific issues in applied studies that may affect the accuracy and comparability of different evaluations. Methods. A descriptive analysis of 4 modeling issues (inclusion of incident cases over a model time horizon, appropriate time horizon, parsimonious model structure, and the handling of age-specific subgroups) is presented. A case study model is analyzed to illustrate the quantitative impact of 3 of the issues. Results. In the case study model, alternative specifications of the modeling framework are shown to alter the estimated cost-effectiveness by large percentages. The combined effect of including incident cases and reduced follow-up yielded the highest divergence from the reference case results, by between 20% and 40%, depending on the age group. Reference case results of an age-weighted population were almost 14% different from the middle single age cohort. Discussion. The identified issues are all generalizable to a wide range of treatment areas and are, or should be, addressed by evaluative guidelines. The authors call for the continued development, dissemination, and application of guidelines for the conduct of economic evaluation in general and model-based economic evaluations in particular. Key words: guidelines; models; decision support; cost-effectiveness; antiplatelet drugs. (Med Decis Making 2007;27:491–499)
considered infeasible for this study). The aim is to highlight the need for the common application of guidelines that address these specific modeling issues.

IDENTIFYING THE ISSUES

The following sections describe the basis for our assessment of the 4 identified issues: inclusion of incident cases over the time horizon of an evaluation, appropriate time horizon, a parsimonious model structure, and the handling of age-specific subgroups.

Incident Cases

Gaspoz and others\(^3\) used the CHDP model to estimate the incremental cost-effectiveness of aspirin, clopidogrel, or combined aspirin plus clopidogrel for the secondary prevention of coronary heart disease. It is stated that the evaluation “modelled U.S. patients, 35 to 84 years of age, in whom coronary disease developed during or before 2003 to 2027 and who survived their first month with it.”

The evaluation configured the CHDP model to predict the costs and quality-adjusted life years (QALYs) associated with coronary heart disease (CHD) in the US population over a 25-year period. This means that incident patients are added to the model every year over the 25-year period but that the model follows all patients only to the end of the 25-year period (i.e., patients diagnosed in year 20 are followed up for 5 years, irrespective of their age at the end of the follow-up period). This approach provides an incomplete analysis of the full impact of the interventions being evaluated, for which the downstream costs and benefits of avoiding secondary cardiovascular events in the short term are excluded.

If incident cases are to be included in a model-based evaluation, then the time horizon of the evaluation should be extended such that the costs and effects are measured over the same timeframe (i.e., to the same upper age limit) for every individual who enters the model. If this rule is implemented, in most treatment evaluations, the modeling of incident cases will affect only the absolute cost and effect estimates as the relative cost-effectiveness of alternative interventions will remain the same. Exceptions will occur in cases in which the characteristics of the patient population vary over time, such as changes in the age distribution or disease severity. If identifiable characteristics are shown to significantly alter cost-effectiveness results, then subgroup analyses may be relevant.

Time Horizon

A commonly identified area of variation between modeling studies is the defined time horizon of models, which describes the maximum age to which individuals are followed. In the applied CHDP model,\(^3\) the choice of a 25-year time horizon for the evaluation of a population aged between 35 and 84 years may also be questioned. The youngest individuals entering the model are followed up to the age of only 60 years, and even individuals aged 50 years (an age at which CHD incidence is considerable) may gain significant QALYs after the age of 75 years.

There are numerous examples of the specification of nonlifetime horizons in models evaluating interventions in a wide range of disease areas.\(^4,5\) The impact on the accuracy of the cost-effectiveness results will vary, although it will increase in line with decreasing mortality rates, time horizons, and discount rates.

 Parsimony

Weinstein and others\(^6(12)\) advise that “the structure of the model should be as simple as possible, while capturing underlying essentials of the disease process and interventions,” in their report on principles of good practice for decision-analytic modeling.

The CHDP model aimed to estimate the health effects and costs of coronary heart disease in the US population.\(^2\) This is a complex state transition model comprising 3 linked models describing the incidence of coronary disease in the unaffected population, the first 30 days of an initial coronary event, and the disease history of patients who survive the first 30 days following a coronary event. The final section of the CHDP model (the disease history model) includes 4 health states in months 2 to 12 following diagnosis (angina, acute myocardial infarction [AMI], cardiac arrest, or AMI plus cardiac arrest). Eight states describe the subsequent natural history, representing patients’ coronary disease history as alternative combinations of the experience of 3 events: cardiac arrests, AMIs, and revascularization in the form of a coronary artery bypass graft (CABG). Patients not experiencing any of these events have, by default, angina. From these 8 states, patients are assigned annual transition probabilities of experiencing an AMI, cardiac arrest, CABG, or noncoronary death.

The evaluation of antiplatelet therapies using the CHDP model illustrates the link between unnecessary model complexity and potential modeling errors. First, as a result of identified errors, the authors were
required to revise the originally published results. In particular, 1 error occurred in the estimation of the benefits in reducing the risk of stroke, an area that was not covered by the original CHDP model and was incorporated into the model in the form of a derived reduction in the overall rate of death from noncoronary causes.

Second, a lifetime duration of clopidogrel therapy is assumed in the original Gaspoz analysis, although in response to de Lemos and McGuire, Gaspoz and others specify a therapy option of clopidogrel plus aspirin for all patients in the 1st year post-CHD diagnosis. The results of the short-term treatment option are counterintuitive, showing that clopidogrel is less cost-effective when treatment is confined to the 1st year postdiagnosis— the time period over which patients are at increased risk of further vascular events. Gaspoz and others do not discuss the foundations of this counterintuitive result, and one cannot discount the possibility that a more parsimonious model, describing only the relevant events and providing more flexibility in the representation of the transition probabilities, would have produced alternative results.

Subgroup Analysis

Dewilde and Anderson present a very interesting study that shows the importance of incorporating the age distribution of the eligible population when modeling the cost-effectiveness of screening, rather than simply modeling the pathways of patients from the age at which they first become eligible for screening. Using cervical cancer screening as a case study, they show that the standard approach of modeling a single screening cohort (from age 15 years) and an age-weighted approach in which costs and effects are estimated for 11 separate age-based cohorts result in mean incremental cost-effectiveness ratios of Aus$47,361 and Aus$61,031, respectively (discounted at 5%).

The single cohort approach estimates the cost-effectiveness of screening once the screening program reaches equilibrium (i.e., all eligible women were first offered screening at age 15 years). In the case of cervical cancer screening, equilibrium would not be reached for 45 years based on 5-year intervals between age cohorts and a final screening age of 65 years. An alternative approach would be to treat the different age cohorts as subgroups of the eligible population, with the possibility of defining age-based screening policies. This latter approach has been adopted in various model-based screening evaluations.

The findings of Dewilde and Anderson are likely to be applicable to treatment evaluations, particularly in areas such as CHD, where event rates have been shown to vary significantly with respect to age. Where age-based subgroups are feasible from a policy implementation perspective, separate analyses by age should be undertaken. If it is not possible to implement age-specific policy recommendations, the age distribution of the eligible population should be represented, and the approach presented by Dewilde and Anderson provides a suitable method for undertaking such analyses.

MODELING THE IMPACT OF ALTERNATIVE APPROACHES

A comparison of the use of a parsimonious versus a nonparsimonious model structure is difficult to undertake and requires the application and independent analysis of multiple models as the hypothesized advantages of parsimony relate to a reduced probability of error. However, the 3 other issues raised above may be investigated further by the application of alternative methods of analysis to a single case study model, which is used to demonstrate the impact of including incident cases over the course of the specified time horizon, and the use of a short time horizon relative to expected survival. The case study model is also analyzed to estimate the impact of modeling alternative age-based patient subgroups:

- a single age cohort based on the mean age of the patients informing the input data,
- multiple age cohorts with the results presented separately for each cohort, and
- multiple age cohorts with the results aggregated based on the age distribution of the eligible population.

METHODS

The case study model is 1 of our own previously published models incorporating a wide range of age-specific data. The chosen model evaluates the cost-effectiveness of 1 year of clopidogrel in addition to aspirin compared to aspirin alone in patients with newly diagnosed acute coronary syndromes (ACSs). A decision tree covering the 1-year treatment period describes patient pathways with respect to alternative forms of revascularization over the 1-year treatment period (Figure 1). Major bleeding event rates are linked to the incidence of CABG and percutaneous...
coronary intervention (PCI) rates. The summed proportions of patients in each of the end states of the decision tree (event free, nonfatal myocardial infarction [MI], nonfatal stroke or vascular death) inform the distribution of patients across the start states in a lifetime Markov model. The Markov model describes the annual probability of patients experiencing a MI, stroke, or death from vascular or nonvascular causes (Figure 2). Following the experience of either a MI or stroke, patients may either remain in that state or progress to the death state. Alternative states are used to describe mortality rates for patients in the 1st year of a MI or stroke event and for patients who have survived their 1st year following a MI or stroke.

Data sources used to populate the model are described in more detail elsewhere. In summary, revascularization and bleeding event rates were based on aggregate (non–age-specific) analysis of prospective and retrospective studies of ACS patients in the United Kingdom. Patient-level data from the Prospective Registry of Acute Ischaemic Syndromes in the UK and the Nottingham Heart Attack Register (D. Gray, personal communication, 2003) were informed age-specific event rates for ACS patients from diagnosis, as well as following the experience of an MI. The estimated event rates accounted for higher event rates in the period immediately following an event by undertaking logit regression analysis to estimate initial event rates, as well as exponential survival analysis to estimate constant longer term event rates (time-varying distributions were tested and were not found to be statistically significantly different). General population mortality rates, minus the age-specific proportion of deaths due to vascular causes, were also applied in the model.

No data were identified describing subsequent event rates in ACS patients experiencing stroke. Data were available describing event rates in all stroke patients (i.e., not limited to stroke patients previously diagnosed with ACS), but these rates were unrealistically low in comparison to the observed event rates in patients remaining event free following ACS. A conservative assumption was made that ACS patients experiencing stroke during the 1-year treatment period had the same subsequent event rates as ACS patients remaining event free.

Treatment effectiveness was represented as the relative risk (RR) of 4 clinical outcome events: vascular death, nonfatal MI, nonfatal stroke, and major bleeding. The RRs were obtained from the Clopidogrel in
Unstable Angina to Prevent Recurrent Events (CURE) trial, which randomized 12,562 patients within 24 hours of the onset of non–ST-segment elevation ACS to standard therapy, including aspirin (75–325 mg), or to clopidogrel (300 mg loading dose and then 75 mg daily) on top of standard therapy, including aspirin.15 Cost and utility data were obtained from published sources and are described in Table 1, as are the RRs.

MODEL ANALYSES

The reference case analysis of the model is defined to exclude incident cases, following a patient cohort of 60-year-olds newly diagnosed with ACS to a maximum age of 100 years. Costs and QALYs are both discounted at an annual rate of 3.5% (currently recommended in the United Kingdom). Subsequent analyses cover all combinations of including and excluding incident cases; modeling patient cohorts of 50-, 60-, and 70-year-old patients; and maintaining the maximum age 100-years’ timeframe or reducing the timeframe to a 20-year follow-up period. The results of these analyses are also combined to estimate age-weighted aggregate estimates of costs and QALYs.

RESULTS

The results of the analyses based on alternative sets of framework assumptions are presented in Table 2 for a single age patient cohort. The reference case analysis (number 1: 60-year-old patients, excluding incident cases, followed to a maximum age of 100 years) shows average costs per patient of almost £16,000 (difference between treatment groups of £393) and QALY gains of around 9 QALYs (difference between treatment groups of 0.036), leading to an incremental cost-effectiveness ratio (ICER) of £11,054.

Compared to the reference case, analysis number 2 (in which a shorter time horizon of 20 years is specified) shows a very small increase in the ICER (0.1%), which is due to the high mortality rates associated with ACS—after 20 years, more than 85% of a 60-year-old cohort receiving acetylsalicylic acid + clopidogrel are expected to have died. In analysis number 3, the lone effect of modeling incident cases over a 40-year time horizon is not large (7.3%) as the effects on patients presenting beyond the first 20 years of the time horizon are reduced due to discounting, whereas patients presenting up to 20 years are mostly followed up to death.

Analysis number 4 shows that the combined impact of reducing the time horizon and including incident cases across the time horizon has a more substantial impact on the estimated ICER than the sum of the parts, increasing it by more than £3000 (over 27%).

Comparisons of the same set of analyses within the 50- and 70-year-old patient cohorts show a similar pattern. The difference from the reference case results is highest for the combined case (incident cases and short time horizon) for the cohort of 50-year-old patients as the shorter time horizon assumption has a greater proportional impact in the case study model. The absolute effects in the 60- and 70-year-old cohorts are similar for the shorter time horizon assumption,

### Table 1 Relative Risks, Costs, and Utility Values Used in the Model

<table>
<thead>
<tr>
<th>Relative risk</th>
<th>Mean</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonfatal MI (clopidogrel + standard therapy v.</td>
<td>0.71</td>
<td>15</td>
</tr>
<tr>
<td>standard therapy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonfatal stroke (clopidogrel + standard therapy v.</td>
<td>0.73</td>
<td>15</td>
</tr>
<tr>
<td>standard therapy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular death (clopidogrel + standard therapy v.</td>
<td>0.93</td>
<td>15</td>
</tr>
<tr>
<td>standard therapy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revascularization (clopidogrel + standard therapy v.</td>
<td>0.976</td>
<td>15</td>
</tr>
<tr>
<td>standard therapy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major bleed (clopidogrel + standard therapy v.</td>
<td>1.38</td>
<td>15</td>
</tr>
<tr>
<td>standard therapy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost (£s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASA 300 mg/d (pa)</td>
<td>3.47</td>
<td>21</td>
</tr>
<tr>
<td>Clopidogrel 75 mg/d (pa)</td>
<td>460.32</td>
<td>21</td>
</tr>
<tr>
<td>PCI revascularization</td>
<td>2455</td>
<td>22</td>
</tr>
<tr>
<td>(HRG E15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CABG revascularization</td>
<td>6275</td>
<td>22</td>
</tr>
<tr>
<td>(HRG E05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major bleeding event</td>
<td>2064</td>
<td>23</td>
</tr>
<tr>
<td>ACS event free (year 1)</td>
<td>1421</td>
<td>16</td>
</tr>
<tr>
<td>ACS event free (post–year 1)</td>
<td>1421</td>
<td>16</td>
</tr>
<tr>
<td>MI (year 1)</td>
<td>3966</td>
<td>16</td>
</tr>
<tr>
<td>MI (post–year 1)</td>
<td>1587</td>
<td>16</td>
</tr>
<tr>
<td>Stroke (year 1)a</td>
<td>7466</td>
<td>24</td>
</tr>
<tr>
<td>Utility value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS event free (year 1)</td>
<td>0.80</td>
<td>25–27</td>
</tr>
<tr>
<td>ACS event free (post–year 1)</td>
<td>0.93</td>
<td>28–30</td>
</tr>
<tr>
<td>MI (year 1)</td>
<td>0.80</td>
<td>25–27</td>
</tr>
<tr>
<td>MI (post–year 1)</td>
<td>0.93</td>
<td>28–30</td>
</tr>
<tr>
<td>Stroke (year 1)a</td>
<td>0.69</td>
<td>31</td>
</tr>
</tbody>
</table>

Note: a. ACS, acute coronary syndrome; ASA, acetylsalicylic acid; CABG, coronary artery bypass graft; MI, myocardial infarction; PCI, percutaneous coronary intervention. Based on 35% of new stroke patients being dependent.24 No utility values are described for the post–year 1 stroke state because the model assumes patients experiencing a stroke return to the ACS event-free state due to a lack of relevant prognostic data for these stroke patients.

Unstable Angina to Prevent Recurrent Events (CURE) trial, which randomized 12,562 patients within 24 hours of the onset of non–ST-segment elevation ACS to standard therapy, including aspirin (75–325 mg), or to clopidogrel (300 mg loading dose and then 75 mg daily) on top of standard therapy, including aspirin.15 Cost and utility data were obtained from published sources and are described in Table 1, as are the RRs.
and the percentage change differs only due to the magnitude of the reference case estimate.

Table 3 presents a comparison of the results within the separate age groups with the results of an age-weighted aggregate patient cohort that reflects the age distribution of the patient cohort within the general population. Analyses numbers 1, 5, 9, and 13 show that the reference case ICER (no incident cases and a time horizon that follows patients to a maximum age of 100 years) for the youngest of the 3 single age patient cohorts was almost double the value of the ICER for the oldest age cohort, and there was a 14% difference between the age-weighted ICER and the ICER for the middle age cohort (the most likely single age cohort to be presented in a single age cohort analysis). A 60% difference in the ICER is noted when comparing the reference case results for the 50-year-old cohort and the age-weighted cohort.

Analysis number 8, involving a 20-year time horizon, incident cases, and a 50-year-old cohort, increases the ICER by more than 120% relative to the reference case age-weighted analysis. Comparing the age-weighted cohort with the 60-year-old cohort, analysis number 12 shows that if incident cases were combined with a shorter time horizon, the estimated ICER would be 45% higher than the age-weighted reference case.

DISCUSSION

This article explores the impact of implementing alternative modeling assumptions relating to the inclusion of incident cases over the course of the specified time horizon, the use of a nonlifet ime horizon, and the handling of age-specific patient subgroups. The analysis of a published model that incorporated significant age-specific parameter values shows that the effect of assumptions alternative to defined reference case scenarios can have a significant impact on the resulting cost-effectiveness ratios.

The case study evaluation involves relatively high levels of disease-specific mortality. This would be expected to reduce the estimated impact of including incident cases and short time horizons as fewer long-term costs and effects are omitted. Other models in disease areas with lower mortality effects would be expected to identify even more significant differences in the incremental cost per QALY gained estimated by alternative sets of assumptions.

Philips and others,1 in their review of guidelines for good practice in decision-analytic modeling in health technology assessment, state that existing guidelines (including their own amalgamated guideline) require further development but that guidelines are likely to be “too general to pick up on the specific nuances of each model.” These guidelines identify the appropriate time horizon as one that extends “far enough into the future to reflect important differences between options”1(p21) and recognize that a lifetime horizon is the most relevant option for most models. Weinstein and others6 also state that “lifetime horizons are appropriate for many models and are almost always required for models in which options have different time-varying survival rates.” We have identified various model-based evaluations that have not adopted these recommendations,3–5 a factor that reduces the comparability of published evaluations and may contribute to flawed policy decisions.

The issue of modeling incident cases over the duration of a model is not explicitly addressed by current guidelines. It is linked to the time horizon of the model as all important effects of an intervention for each patient included in a model should be described (i.e., all patients entering the model should be followed for their lifetime). Few cost-effectiveness models have included incident cases to the end of the specified time horizon, as done by Gaspoz and others,3 although this study serves to clarify the erroneous nature of this
approach and to demonstrate the potential effects on the cost-effectiveness results.

The definition of the appropriate patient cohort, specifically with regard to the age distribution of an eligible patient population, is important. Many model-based treatment evaluations do not include any age-specific data other than general population (non–disease-specific) mortality rates, which may reduce the impact of modeling separate age cohorts. However, in cases in which more widespread age-specific data are available, the impact on aggregate cost-effectiveness may be significant.

The 4th issue raised in this article concerns the parsimony of decision-analytic models for the cost-effectiveness analysis of health care interventions, as recommended by Weinstein and others. The model that highlighted this issue was an application of the CHDP model for the cost-effectiveness analysis of antiplatelet therapies. This initial publication of this model was amended due to the postpublication identification of modeling errors, and the amended version retained some seemingly counterintuitive results that were not addressed by the authors. Although it is not possible to prove a causal link between a nonparsimonious model and model errors, we believe this example does confirm the value of model parsimony.

The potential advantage of using a broad model such as the CHDP model to evaluate a range of alternative interventions in the same broad disease area is that it may provide a more suitable basis for the comparison of their relative cost-effectiveness. There is a balance to be struck regarding the benefits derived from increased comparability of different evaluations and the extent to which reduced parsimony complicates the analysis and increases the probability of error.

The increased awareness of the economic aspects of health care, as well as the need for a guided rationale in allocating scarce health care resources, led to a rapid increase in the number of published economic evaluations over the past decade. Although there is a growing acceptance of the role for economic evaluation in informing clinical and policy decisions over the provision of alternative technologies, there remains some suspicion around the methodological validity of economic evaluations.

Critics believe that the results of model-based evaluations can be manipulated more easily and with...
more subtly than other forms of evaluation\textsuperscript{18} and that the “black box” feel about the analysis of such models does not enhance confidence in the derived results.\textsuperscript{19} Such criticisms have less foundation when a transparent description of model assumptions and data sources is provided. Halpern and others\textsuperscript{20} accept that there are problems associated with modeling studies but argue that many of the cited problems could be countered through “rigorous peer review and methodological development.”

The main limitation of this article is that the effects of 3 of the identified issues were only examined in a single case study model, and so the study is not able to generalize the expected magnitude of the observed effects across a range of case studies in different disease areas. It was also not feasible to undertake a quantitative analysis of the effects of model parsimony, which, in our view, would require a multicenter study, in which separately developed models are subject to blinded review by independent expert reviewers.

This article has demonstrated the effects of specific issues to the extent that they are important issues that have the potential to significantly affect cost-effectiveness results and that they should now be included in modeling guidelines. The continued development, dissemination, and application of guidelines for model-based economic evaluations are necessary to further enable evidence-based decision making.

REFERENCES


