Commentary

The time horizons of formal decision analyses

J. BENBASSAT¹ and R. BAUMAL²

From the ¹Myers-JDC-Brookdale Institute, The Smokler Center for Health Policy Research, Jerusalem, Israel, and ²Department of Laboratory Medicine and Pathobiology, Toronto Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

Summary

Clinical decision analyses use time horizons that vary from hours to the patient’s entire life. Analyses of decisions with a lifetime horizon commonly use Markov models, which simulate the patient’s lifespan by dividing it into equal periods (cycles). At each cycle, the model exposes a hypothetical cohort to the competing hazards of normal aging and of the disease in question (disease-specific hazards), and the results are presented as years of life expectancy. This paper highlights two limitations of lifetime Markov models that have been ignored in recent publications. First, since there are no readily available data on changes in disease-specific hazards over time, these hazards are often derived from short-term follow-up studies, and assumed to be constant over the patient’s entire life. Second, results may be better presented in terms of health states (i.e. proportions of patients expected to recover completely, recover with a disability or die) rather than life expectancy. Although well-known, these two limitations require re-emphasis. They may be avoided by restricting the time horizon of decision analyses and presenting results as health states as well as life expectancies. When a lifetime horizon is necessary, the performance of Markov models may be improved by the use of time-variant disease-specific hazards derived from long-term follow-up studies, or from theoretical models that simulate more closely the disease progression over time, rather than assuming constant disease-specific hazards.

Introduction

A decision tree models events and outcomes that may occur within a defined time horizon as the result of a clinical intervention. The time horizon of some clinical decisions (e.g. ‘should healthy women with lower urinary tract infections have a urine culture before treatment?’) may span days. Such decisions can be modelled by ‘simple’ trees consisting of chance nodes and terminal nodes that present the utilities of the outcomes in quantitative terms on an agreed utility scale. In other decisions (e.g. ‘should healthy persons with incidentally detected gallstones be treated?’), the time horizon spans the patient’s entire life. Such decisions commonly use Markov models¹ to simulate the patient’s lifespan by dividing it into equal periods (cycles). At each cycle, the model exposes a hypothetical cohort to the competing risks of normal aging and of the disease under consideration. The risk of dying from normal aging is derived from life-tables in terms of gender/race-stratified mortality at particular years of age. The risk of dying from the disease (disease-specific mortality) is derived from published studies of patients with the disease in question. As the cohort moves through the Markov model, it accumulates utilities (years of survival) that may be presented either as

Address correspondence to Professor J. Benbassat, Myers-JDC Brookdale Institute, PO Box 13087, Jerusalem 91037, Israel. email: benbasat@jdc.org.il

© The Author 2007. Published by Oxford University Press on behalf of the Association of Physicians. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
such (life expectancy), or corrected for quality of life and presented as quality-adjusted life years (QALYs).

Prolongation of life and improvement of its quality are the most important goals of medical care. Therefore, time horizons spanning the patient's entire life, and the presentation of clinical outcomes as life expectancies or QALYs, have gained acceptance in decision analyses in the clinical and public health domains. Markov lifetime simulations and QALYs are essential for analyses of the choice among 'one-time' clinical interventions (e.g. vaccination, screening for diseases, surgery) that are intended to affect long-term prognosis, and of their cost-effectiveness (i.e. the costs per QALY added to the patient's life expectancy). Furthermore, QALYs allow for the comparison of divergent health states, and as such, represent a standard measure of health outcomes in clinical and economic decision analyses.

Despite the undisputed importance of lifetime simulations in clinical decision analyses, the assumptions that underlie these simulations may result in systematic biases. Further, some physicians and patients may consider a presentation of the outcomes of clinical decisions in terms of health states (e.g. death, complete recovery, recovery with a disability) within a defined time horizon more clinically intuitive than a presentation of these outcomes in terms of QALYs. The objectives of this paper are to draw attention to some recently published clinical decision analyses that ignore either or both of two possible shortcomings of decision modelling, even though both have been discussed before: (a) the inappropriate use of time-constant risks of events in Markov models, instead of risks that vary with time, and (b) the presentation of health outcomes as QALYs only, instead of both in terms of QALYs and specific health states.

Possible limitations of Markov simulations

Limited validity of the assumption that disease-specific hazards are constant over time

The term ‘hazard’ refers to the probability of a subject developing an event (e.g. stroke) and its possible outcomes (e.g. death), during a discrete time interval, such as a Markov cycle. Average disease-specific hazards are derived from published studies of patients with the disease in question, by dividing the number of events by the duration of the follow-up. Markov simulations frequently assume that the average disease-specific annual hazards, which were derived in this way from a patient population of a mean age of (e.g.) 40 years, followed for (e.g.) 3 years, are constant over the patient’s entire lifetime, and therefore, may also be extrapolated beyond the first 3 years after diagnosis, as well as in patients belonging to other age groups.

However, this assumption (that disease-specific hazards are constant with respect to age and time since diagnosis) is inconsistent with the reported hazards in many diseases. Examples of time-variant hazards include: acute infections and traumatic injuries, in which mortality is higher after diagnosis and declines during recovery; hypertension, in which mortality increases with time since diagnosis, and cancer, in which mortality declines with time since diagnosis. Examples of age-variant hazards include: lower disease-specific mortality in younger patients with bladder or lung cancer than in older patients with these diseases; and the higher risk of recurrent stroke in older patients with chronic atrial fibrillation than in younger patients with this condition.

Data on time- and age-variant disease-specific hazards are often not readily available—hence the need for the simplifying assumption that they are constant. However, the time-variance of these disease-specific hazards indicates that duration of follow-up may confound the average hazard derived from a published series. In other words, because of this time-variance, it is possible to derive different average hazards from the same patient population simply by changing the duration of the follow-up. In diseases with declining hazards over time since diagnosis, the average hazard derived from a short-term follow-up would be higher than that derived from a longer follow-up of the same patients. Consequently, Markov models that extrapolated constant disease-specific mortality hazards beyond the duration of follow-up of the studies from which they were derived, would underestimate life expectancy. On the other hand, in diseases with increasing hazards over time since diagnosis, Markov models that extrapolated average annual disease-specific mortality hazards beyond the duration of study follow-up, would overestimate life expectancy. For example, the life expectancy of patients with lung and breast cancer predicted assuming a constant disease-specific mortality is different from that predicted by a Markov model that used the time-variant disease-specific mortality data derived from the Surveillance, Epidemiology and End Result (SEER) tumour registry.

The validity of the assumption of constant disease-specific hazards over time since diagnosis is particularly uncertain in the case of ‘acute on chronic’ types of disorders, such as ischaemic stroke.
in patients with chronic cardio- or cerebrovascular disease, where in follow-up studies, the risk of recurrence was highest during the first 30 days after a stroke,10–12 with about half11 to a quarter12 of the risk at 1 year being accrued during the first 30 days. The ratio between the observed rates of recurrent ischaemic strokes and the expected rates of first ischaemic stroke in the population has been reported to decline from 8.5 during the first year after the index stroke to about 2 during the second and third years, to 1.5 in the fourth year and 1 in the fifth year.11 Other authors have reported a similarly rapid decline in the observed/expected rate ratio from 15.4 during the first year after the index stroke to 8.5 during the second, 6.7 during the third, 4.5 in the fourth year and 2.0 in the fifth year.14 Similarly, in another study, the annual hazards of recurrent stroke initially declined from year one to year three and subsequently increased during the following six years.15 As noted by others, this biphasic time-variability of the probabilities of recurrence makes modelling the natural history of stroke difficult;16 and precludes any credible extrapolation of disease-specific hazards beyond the duration of follow-up of the patient population from which these hazards were derived.

Nevertheless, Markov simulations with a lifetime horizon have been used in published analyses of the choice among treatment options in patients with cerebrovascular disease.17–20 These analyses have assumed not only constant hazards of recurrent ischaemic stroke over time, but also constant bleeding hazards with or without anti-thrombotic therapy. This latter assumption is similarly incorrect, because the hazards of bleeding decline over time after beginning anti-thrombotic treatment,21 and the hazards of intra-cerebral haemorrhage (ICH) decline over time in untreated patients with an ischaemic stroke.22,23

For example, the analysis by Nandez et al.17 of the choice among the treatment options in patients with patent foramen ovale (PFO) and ischaemic stroke is based on two assumptions of questionable validity. First, the authors extrapolated the average annual risk of stroke recurrence (derived from two studies of patients followed for only 2 years24 and 3 years25) to the patient’s entire lifespan. Assuming constant recurrence hazards over time, the authors ignored the high risk of recurrence during the first month after an ischaemic stroke10, its decline during the first 3–5 years,13–15 and its increase after 3 years.15 Second, by assuming constant hazards of bleeding over time, the authors ignored the declining hazards of bleeding after beginning anti-thrombotic therapy.21 Third, the authors derived a lifetime 90–100% efficacy of surgical PFO closure in preventing recurrent ischaemic strokes from a follow-up study of only 2 years.26

Lifetime Markov simulations have also been used to model the choice between carotid endarterectomy and conservative treatment of patients at risk for ischaemic stroke.20 Here again, the authors assumed constant hazards of ischaemic stroke over the patient’s lifespan, and extrapolated a 50% efficacy of endarterectomy over the first 10 years after surgery from a study with only a 3.5 year follow-up.27 Similarly, the authors of two analyses of the trade-off between the risks and benefits of anti-thrombotic treatment in the case of patients with chronic atrial fibrillation and a past ICH18 or past bleeding events at other sites19 assumed constant disease-specific and treatment-related hazards over time. They extrapolated both of these hazards from studies of patients followed for only 1–2 years.7,28 The hazards of recurrent ICH in untreated patients were extrapolated from studies of patients followed for only 2 years29 and 3.4 years.30

Limited relevance of QALYs to clinical decision-making

It may be argued that the benefits obtained from lifetime Markov models justify the possible inaccuracies that are due to the simplifying assumption of a constant disease-specific hazard. This argument should certainly be considered when a presentation of the outcomes of a decision analysis in terms of life expectancies or QALYs is felt to be absolutely necessary. However, such a presentation is not always required. Most clinical decisions attempt to identify the management option with the most desirable outcome within a restricted time horizon. This horizon commonly spans over a few years, or even less, e.g. until the patient’s next medical re-evaluation, with a view of revising the previous decisions if necessary. In such cases, a presentation of the predictions of decision analyses as years of life expectancy or QALYs may be less relevant, and not as clinically intuitive as a presentation in terms of specific health outcomes (such as death, living and well, and living with a disability) within a restricted time horizon.

A restricted time horizon and a presentation of the results in terms of actual health states may have been more appropriate in several recently published decision analyses. Kim et al.31 analyzed the choice between surgery and medical treatment in the case of a 20-year-old patient with Marfan syndrome and aortic disease, and reported the predicted outcomes in terms of life expectancies (73.8 years for surgery and 71.4 years for medical treatment). We believe that a report of the proportion of patients, who are
expected to develop complications or die after surgery, bleed after anticoagulant treatment, sustain aortic dissection or thromboembolism, or stay well during a restricted time horizon of (e.g.) 5 years, would have been more informative. Similarly, Aujesky et al.\textsuperscript{32} analyzed the cost-effectiveness of various durations of anticoagulant treatment for the prevention of recurrent deep vein thrombosis (DVT) in the case of a 40-year-old man with a past DVT, and reported their findings as cost per QALYs added. Here again, we believe that an additional presentation of the results in terms of the costs (expenditures) and risks (number of bleeding events due to anticoagulation) needed to prevent one event of DVT or pulmonary embolism within a 5-year horizon would have provided information in more clinically relevant terms. Finally, Perlis et al.\textsuperscript{33} analysed the cost-effectiveness of pharmacogenetic testing in the management of patients with schizophrenia, and presented their results in terms of cost per QALY gained. Again, we believe that a time-restricted analysis in terms of cost, or number of pharmacogenetic tests needed to prevent one death or one psychotic bout, would have been more informative.

A presentation of the outcomes of the analyses as health states, rather than the presentation of their utilities in quantitative terms, would not only be more informative, but also more acceptable for those physicians who feel that making clinical decisions by multiplying probabilities by utilities is counterintuitive.\textsuperscript{34} However, reporting of the results as health states would also preclude sensitivity analyses. Furthermore, costs per added QALY are widely used to inform public health decisions about the adoption of new technologies and treatments. Therefore, we believe that the presentation of the outcomes of a decision analysis in terms of health states should supplement, rather than replace, their presentation in quantitative terms.

Conclusions
The limited validity of the assumption that disease-specific hazards are constant over time has been repeatedly discussed,\textsuperscript{1,3,7} and is well-known among health economists and other researchers as ‘mis-specification bias’.\textsuperscript{2} Similarly, the issue of presenting results as QALYs vs. health states vs. other utility-adjusted health measures has been discussed in the past. Some authors\textsuperscript{17,18,35,36} appear to be aware that decisions about the prevention of thromboembolism should be made using a restricted time horizon, and have presented their results both as QALYs and in terms of actual outcomes (proportion of patients expected to bleed while on anticoagulant therapy, sustain recurrent thromboembolism and die, recover completely or recover with a disability) after 1,\textsuperscript{18,36} and 5\textsuperscript{17,35} years. Furthermore, these two possible shortcomings are certainly not the only potential flaws of decision analysis modelling. However, we believe that the continuing use of lifetime Markov simulations and of the presentation of health outcomes in terms of QALYs justifies the following suggestions.

First, when a lifetime horizon is not necessary, or difficult to model as in the case of cerebrovascular diseases, both of the limitations of decision modelling that we have discussed in this paper may be avoided by (a) restricting the time horizons to the duration of the follow-up studies from which the disease-specific hazards were derived, and (b) presenting the utilities not only in quantitative terms, but also in terms of expected health states at the end of this restricted time horizon. Indeed, although many published decision analyses have used lifetime horizons and reported their results in terms of life expectancies or QALYs, other analyses using Markov models have restricted their time horizons and presented their results as health states, such as respiratory distress syndrome in newborns,\textsuperscript{37} disease-free survival in women with breast cancer,\textsuperscript{38} and mortality in patients with asymptomatic small gallstones.\textsuperscript{39} Examples of cost-effectiveness analyses within a restricted time horizon are those evaluating single- and dual-chamber pacemakers for the sick sinus syndrome,\textsuperscript{40} vaccination for influenza\textsuperscript{41} and the savings achieved by the control of tuberculosis in high-incidence countries.\textsuperscript{42}

Second, when a lifetime horizon and an expression of the results as QALYs are felt to be necessary, we suggest that the simplifying assumption of constant disease-specific hazards over time and the way that it may bias the analysis be made explicit and included in the list of the simplifying assumptions. Alternatively, whenever possible, the predictions of Markov models may be improved by the use of time-variant disease-specific hazards derived from long-term observational studies or theoretical models. We have already referred to the SEER tumour registry as a source of age- and time-variant cancer-mortality hazards. Other Markov simulations have derived time-variant mortality hazards from the literature to assess the cost-effectiveness of screening for lung cancer,\textsuperscript{43} and time-variant immunity levels after vaccination, in order to estimate the benefits, risks and costs of pertussis vaccination.\textsuperscript{44} Alternatively, the age- and time-variant disease-specific hazards may be derived from theoretical models\textsuperscript{45,46} that attempt to
simulate disease progression more closely, rather than assuming constant disease-specific hazards.

In the absence of a gold standard for assessing the validity of the results of decision analyses, the advantage of models that use restricted time horizons, or those using time-variant disease specific hazards, remains unproven. However, it stands to reason that the accuracy of the predictions of decision analyses is inversely related to the length of their time horizons, and that the predictions of models that simulate more closely the progression of a disease over time are more reliable than those assuming rough approximations. It may be argued that the possibility of errors due to the use of constant disease-specific hazards is appropriately addressed by sensitivity analyses of the robustness of the conclusions to changes in the estimated probabilities. But the take-home message of published decision analyses is frequently defined by their baseline probability values, rather than by the acknowledged uncertainties in the accuracy of these values. Therefore, we believe that efforts to improve the predictions of the disease models used in decision analysis are worthwhile, and the examples that we have presented in this paper indicate that they are feasible.

References


