Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project

J Lord,[1]* S Willis,[2] J Eatock,[3] P Tappenden,[4]
M Trapero-Bertran,[1] A Miners,[2] C Crossan,[1]
M Westby,[5] A Anagnostou,[3] S Taylor,[3] I Mavranezouli,[6]
D Wonderling,[5] P Alderson[7] and F Ruiz[7]

[1]Health Economics Research Group, Brunel University, Uxbridge, UK
[2]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK
[3]Information Systems, Computing and Maths, Brunel University, Uxbridge, UK
[4]School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK
[5]National Clinical Guideline Centre, Royal College of Physicians, London, UK
[6]National Collaborating Centre for Mental Health, University College London, London, UK
[7]National Institute for Health and Care Excellence, London, UK

*Corresponding author

Scientific summary

Diagnostic and treatment pathways in NICE clinical guidelines

Scientific summary

Background

National Institute for Health and Care Excellence (NICE) clinical guidelines (CGs) make recommendations for the care of people with defined diseases or conditions in the NHS in England and Wales. The recommendations are drawn up by a Guideline Development Group (GDG) after consideration of evidence of the benefits, harms and costs of various diagnostic and treatment options. As NICE CGs are usually very broad, covering large complex pathways of care for heterogeneous groups of patients, the approach to assessing clinical and economic evidence is inevitably selective. The GDG identifies key 'clinical questions' within the scope, and evidence for these questions is identified, critiqued and summarised in a series of systematic reviews. Evidence of cost-effectiveness is drawn from relevant literature or from original evaluations conducted by guideline economists. Timelines and resources are not usually sufficient to build new models to evaluate cost-effectiveness for every clinical question, and so specific topics for economic analysis are prioritised. There are three potential risks with this approach: (1) estimates of cost-effectiveness might not be available for some topics with potentially important resource consequences; (2) methods and assumptions used for economic analysis at different points in the pathway might be inconsistent; and (3) systemic effects and interactions between decisions might be neglected.

Another approach to economic evaluation in CGs would be to build a model of the full care pathway and to use this as a platform to evaluate changes that are being contemplated. This has some putative advantages. Though a big initial investment of time and resources, once developed the full guideline model should enable analysis of a range of cost-effectiveness questions and provide the flexibility to address unanticipated questions, which often arise during or after guideline development. The full guideline model would also provide a common framework of methods, baseline data and assumptions, to improve the consistency of cost-effectiveness estimates. Furthermore, embedding the decision options within the context of the whole pathway provides the potential to account for systemic interactions between options. However, there are some technical and practical obstacles to the adoption of a full guideline modelling approach in the NICE CGs programme. To reflect guideline pathways, the models would most likely need to be large and complex, and would therefore be challenging to build and to validate. It is therefore unclear whether or not full pathway modelling in NICE CGs is realistic, and what value it would add to the more conventional piecewise approach to economic analysis.

Objectives

The Modelling Algorithm Pathways in Guidelines (MAPGuide) project was set up to test the feasibility and usefulness of full pathway modelling in NICE CGs. We chose to test the concept first outside of routine guideline development by building models for two existing guidelines. The guidelines chosen as case studies were both due to be reviewed by NICE to determine whether or not they should be updated. This offered an opportunity to identify new evidence and suggested changes to the guidelines raised by experts and stakeholders, which provided some real cost-effectiveness questions to be addressed by our models. Stakeholders for the two guidelines were surveyed to elicit their views on the relative importance of the topics as a comparison for the economic priorities identified by the models.

The aims of the project were therefore to:

- investigate the feasibility of modelling whole service pathways from NICE CGs to estimate the cost-effectiveness of possible changes to the pathways

- use this approach to estimate the potential value of updating selected topics within the guidelines, as estimated by the likelihood of a change in recommendation and the potential impact of any such change on expected net benefits (NBs)
- compare the economic priorities for updates obtained from formal modelling with those elicited during the routine NICE guideline review process.

Methods

Two published NICE guidelines were selected as case studies: prostate cancer and atrial fibrillation (AF). These guidelines were chosen from the 17 guidelines due to be reviewed by NICE within our timelines (between January and September 2011). The criteria for selecting the case studies were (a) that the published guideline contained a relatively well-formulated pathway suitable for modelling; (b) that they addressed important topics that were likely to be updated; (c) that there was thought to be uncertainty or controversy over which topics should be updated; and (d) the guidelines would address different patient groups or disease areas, and hence would present different challenges for the modellers.

The project consisted of three streams of work led by different teams of researchers: the first team (CC and MW) identified the potential update topics and conducted and analysed the surveys of stakeholders; the second team (SW, PT and AM) developed the prostate cancer model and used it to evaluate possible update topics for this guideline; and the third team (MTB, JE, AA and JL) developed the AF model and evaluated possible update topics for this disease area. The modelling teams did not influence the choice of topics and did not see their list of topics until the design of the base-case model had been finalised. They were also not shown the results of the stakeholder surveys until after the base-case models had been implemented. The two modelling teams met regularly to discuss progress and difficulties encountered, and to agree the general principles to be followed.

The potential update topics were elicited from the review decision documents published on the NICE website. One researcher (CC) read the documents and collated a list of suggested topics with advice from a second researcher who has extensive experience in systematic reviewing and guideline development (MW). A shortlist of topics for each guideline was agreed by members of the research team who were not involved in developing the models. Information about the selected topics and relevant new evidence cited in the review documents was collated. Surveys were then conducted with registered stakeholders for the two guidelines to elicit their opinions about the importance of including the selected topics in a future update of the guidelines. Participants were presented with a short summary of the shortlisted topics and then asked to rate them in terms of importance (using a five-point Likert scale) and to rank them in order of priority for inclusion in a future update. Results were summarised using simple descriptive statistics and graphs.

The modelling teams began with background reading to familiarise themselves with the guideline and current issues in the field. The boundaries for the modelling exercises were defined by the scopes for the original guidelines. The model designs consisted of two main elements: a 'service pathway model', which specifies the expected care that patients with defined characteristics would receive according to the current guideline recommendations; and a 'disease process model', which specifies how patients' health status or risk of events is expected to change conditional on their characteristics and treatment. The pathway models were developed following detailed examination of the guideline documents, and then checked with clinical experts. Published NICE Technology Appraisal (TA) guidance within the scope of the guidelines was added to the care pathways. The disease process models were developed following review of other published models, descriptions of epidemiology and discussions with clinical experts. The implemented models were designed to estimate the mean discounted health-care costs and outcomes (quality-adjusted life-years; QALYs) over a lifetime horizon for a representative heterogeneous cohort of patients treated according to the defined pathway, and following the NICE reference case for economic evaluations. The models were implemented as individual-level discrete event simulations (DES) programmed using a specialist software package (SIMUL8 Professional version 15.0; SIMUL8 Corporation, Boston, MA, USA).

The starting characteristics for patients entering the models were taken from samples of individual data from a general practice database for AF and a disease register for prostate cancer. It was beyond the scope of this project to conduct systematic reviews to inform all model parameters. Information was therefore identified from the original guideline, supplemented with new evidence from rapid reviews of the literature and discussions with clinical experts. The impact of uncertainty over model parameters was considered through probabilistic sensitivity analysis. The models were each evaluated over two nested levels of iteration: an internal level to reflect variability between individual patients (first-order uncertainty); and an external level to represent uncertainty over population parameters (second-order uncertainty). The modelling teams checked for errors and inconsistencies during model development. The models were also verified to test for correct programming and validated to ensure consistency with expected results (e.g. that survival times and levels of service use were realistic).

The modelling teams then attempted to estimate the cost-effectiveness of questions related to their list of possible update topics. This involved running the models for the base-case pathways, and then for a series of strategies reflecting possible changes to the pathways. Additional data required for the alternative strategies were obtained from the original guideline or new evidence identified by stakeholders or experts, or by rapid literature searches. It is important to note that the economic analyses presented in this report were not based on full evidence reviews and have not been informed by an expert guideline group. Our focus was to assess the modelling methods and their usefulness in the context of identifying priorities for inclusion in a guideline update. The results should not be used to inform clinical practice.

The optimal strategy within each topic was identified by incremental cost-effectiveness analysis (CEA), or equivalently by calculation of incremental net benefits (INBs). Results are reported using a primary cost-effectiveness threshold of £20,000 per QALY gained. The relative 'economic priority' for which topics should be included in an update was determined on the basis of two pieces of information: (i) the probability that the currently recommended base-case option is not the optimal strategy within that topic and (ii) the magnitude of the potential gain in NB (difference between the optimum and base-case strategies). These statistics were used by the modelling teams to rank the topics in order of importance for inclusion in an update. This ranking was then compared with that obtained from the survey of guideline stakeholders.

Results

Both modelling teams succeeded in building a DES representing the guideline pathway and disease process. Development took nearly 24 months, and required more than one whole-time equivalent year of analyst time for each full guideline model. The scope of both models was very broad, and included the large majority of the guideline pathways and recommendations, though with some exceptions.

Nine potential update topics were identified from the NICE review documents for the prostate cancer guideline, and eight topics for the AF guideline. Survey responses were received from 27 out of 239 registered stakeholders for prostate cancer. Two topics were rated as 'very important' for inclusion in an update of the guideline by the majority of respondents: active surveillance (AS) in previously unscreened 'low-risk' men, and effective techniques for performing radical prostatectomy. For AF, 32 out of 182 stakeholders responded to the survey, and two topics were rated as 'very important' by the majority of these respondents: new oral anticoagulants (OACs), and stratification tools to assess bleeding risk before antithrombotic medication.

Cost-effectiveness analysis was conducted with the prostate cancer model for six of the nine potential update topics. Two of the remaining topics were not modelled because of an absence of necessary information, and there was no need for CEA of the final topic owing to likely dominance of one of the options. The model indicated that for three topics in particular, there was a high potential for increased NB: brachytherapy with external beam radiotherapy for localised or locally advanced disease; pelvic

radiotherapy with adjuvant hormone therapy for localised disease; and continuous versus intermittent hormone therapy for metastatic prostate cancer.

For AF, cost-effectiveness estimates were obtained for five out of eight topics. Two topics were not evaluated as they related to areas excluded from the scope of the model, and one topic was omitted due to time constraints. The results for the five modelled topics suggested that stroke and bleeding risk thresholds for oral anticoagulation and the comparison of rate and rhythm control strategies should be treated as priorities for inclusion in an update; as there was a fair probability that the current guideline recommendations were not optimal and a high potential for health gain and cost savings. The estimated economic priorities for update topics differed from those elicited from stakeholders.

Conclusions

This project has demonstrated that models can be developed to cover the large part of complicated CG pathways, and that these models can be used to evaluate a range of cost-effectiveness questions across the pathways. This approach has the potential to produce consistent estimates of cost-effectiveness that account for systemic effects of placing decision options within a broad pathway of care. Implementation of the models was facilitated by the use of DES, which can provide a relatively compact representation of the flow of a heterogeneous cohort of patients through a very complex care pathway.

However, there are some barriers to the routine adoption of this approach in the NICE CGs programme. The most obvious is that it is unlikely to be possible within current timelines and health economic resources, because of the demands of developing such large models and the need for specialist DES expertise. It is also unlikely that this approach will work for all CGs, due to data limitations. Nevertheless, we believe that the approach does show sufficient promise to warrant further investigation.

Learning from this project should enable faster development of full guideline models in the future. The two case studies provide a set of methods and terminology that could be adapted for other applications. We also identify some considerations for future development, including the need for clarity about the boundaries of the model; the need for clarity over whether the model pathway is meant to reflect recommended or current practice; and the importance of conceptual modelling of both the service pathway and of the disease process.

We recommend five priorities for further research:

1. The case study models have been made available to the National Collaborating Centre (NCC) teams developing updates of the guidelines. Research should be conducted to seek the opinions of the members of the NCC technical teams, the GDGs, stakeholders and the NICE guidelines team to determine whether or not they made use of the models, and, if so, whether or not the results were useful.
2. Further development of the case study models to assess the existence and magnitude of possible interactions between changes to different parts of the care pathways.
3. Extension of the case study models to estimate budget and health impacts across a whole patient population (rather than for single incident cohort).
4. To apply the full guideline modelling approach to a new NICE guideline.
5. Further development of methods:

   i. methods for eliciting expert opinion and reaching consensus about the structure of disease process and service pathway models to inform guideline development and economic modelling
   ii. methods for robust model calibration to infer missing or unobservable parameters in complex decision models

   iii. development of standardised software templates or methods of presentation to help guideline economists to develop flexible and accessible guideline models in consultation with other methodologists, GDGs and stakeholders

   iv. methods to test the internal and external validity of full guideline models.

Funding

Health Technology Assessment

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme or, commissioned/managed through the Methodology research programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: www.hta.ac.uk/

This report

This issue of the Health Technology Assessment journal series contains a project commissioned/managed by the Methodology research programme (MRP). The Medical Research Council (MRC) is working with NIHR to deliver the single joint health strategy and the MRP was launched in 2008 as part of the delivery model. MRC is lead funding partner for MRP and part of this programme is the joint MRC–NIHR funding panel 'The Methodology Research Programme Panel'.

To strengthen the evidence base for health research, the MRP oversees and implements the evolving strategy for high quality methodological research. In addition to the MRC and NIHR funding partners, the MRP takes into account the needs of other stakeholders including the devolved administrations, industry R&D, and regulatory/advisory agencies and other public bodies. The MRP funds investigator-led and needs-led research proposals from across the UK. In addition to the standard MRC and RCUK terms and conditions, projects commissioned/managed by the MRP are expected to provide a detailed report on the research findings and may publish the findings in the HTA journal, if supported by NIHR funds.

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health.