Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling

R Garside,[1*] M Pitt,[1] M Somerville,[1] K Stein,[1] A Price[2] and N Gilbert[1]

[1] Peninsula Technology Assessment Group, Peninsula Medical School, Universities of Exeter and Plymouth, Exeter, UK
[2] Southampton Health Technology Assessments Centre, University of Southampton, UK

* Corresponding author

Surveillance of Barrett's oesophagus

Executive summary

Health Technology Assessment 2006; Vol. 10: No. 8

Health Technology Assessment
NHS R&D HTA Programme

Executive summary

Background

The NHS Health Technology Assessment (HTA) programme commissioned this project, having established the need for research in this area as a priority. It was, however, unsuccessful in commissioning primary research. The reason for the previous lack of success was thought to be the lack of clarity about the current state of knowledge and areas of uncertainty that might be of most importance to the NHS.

Barrett's oesophagus is a histological diagnosis and occurs when the normal squamous epithelial cells lining the oesophagus are replaced with columnar cells. This metaplasia gives a red appearance to the oesophagus on endoscopic examination. The risk of developing adenocarcinoma of the oesophagus (ACO) is increased with Barrett's oesophagus, although the size of this increased risk is unknown.

Gastro-oesophageal reflux disease (GORD) is associated with Barrett's oesophagus. However, some people with Barrett's oesophagus may be symptom free, and it is probable that only a small minority of those with GORD will have Barrett's oesophagus.

Given a known, unquantified increased risk of ACO with Barrett's oesophagus, endoscopic surveillance of this condition is common and three-quarters of UK gastroenterologists believe it to be worthwhile. However, evidence for this practice is lacking.

Aims of the project

The aims of the project were as follows:

1. to assess what is known about the effectiveness, safety, affordability, cost-effectiveness and organisational impact of endoscopic surveillance in preventing morbidity and mortality from adenocarcinoma in patients with Barrett's oesophagus
2. to identify important areas of uncertainty in current knowledge for these programmes
3. to identify important areas of research for the HTA Prioritisation Strategy Group to consider addressing by commissioning further research.

Methods

Three strands of enquiry were used to address these aims:

1. A systematic review of the effectiveness of endoscopic surveillance of Barrett's oesophagus was carried out following the methodological guidelines set out by the Centre of Reviews and Dissemination Report No. 4. Electronic databases were searched for published surveillance studies, economic evaluations and current research. Inclusion criteria were broad, reflecting the known lack of randomised trials or other well-designed or controlled studies in this field.
2. We invited experts in Barrett's oesophagus from the UK to contribute to a workshop on surveillance of Barrett's oesophagus, which was held in London in May 2004. At this stage, the systematic review was not complete and the cost–utility model was still in development. We divided the topic of Barrett's oesophagus into four broad sections and asked four individual experts to summarise the current state of knowledge in each section. Small group discussion, using a modified nominal group technique, then identified key areas of uncertainty within each section and ranked them for importance. The subsequent plenary discussion identified some additional questions, but no attempt was made to rank the questions for overall importance.
3. A Markov model was developed in Microsoft Excel by the Peninsula Technology Assessment Group (PenTAG) to assess the cost-effectiveness of a surveillance programme for patients with Barrett's oesophagus compared with no surveillance and to quantify important areas of uncertainty. The model estimates incremental cost–utility and expected value of perfect information for an endoscopic surveillance programme compared with no surveillance. A cohort of 1000 55-year-old men with a diagnosis of Barrett's oesophagus was modelled for 20 years. The base case used costs in 2004 and took the perspective of the UK NHS. Estimates of expected value of information were included.

Results

Systematic review of clinical and cost-effectiveness of surveillance programmes

Clinical effectiveness

No randomised controlled trials (RCTs) or well-designed non-randomised controlled studies were identified, although two comparative studies and numerous case series were found. Only the comparative studies and seven case series with >300 patients were included in the review.

Reaching clear conclusions from these studies was impossible owing to lack of RCT evidence. In addition, there was incomplete reporting of data, particularly clinical details of the subjects under surveillance and their follow-up, details of the diagnostic methods and protocols used, details of treatment for gastro-oesophageal reflux disease (GORD), policies for offering treatment for adenocarcinoma of the oesophagus or high-grade dysplasia (HGD) and mortality from adenocarcinoma of the oesophagus and from other causes. In addition, changes in surveillance practice over time were mentioned but not explained in several studies.

Limiting the included case series to those with >300 patients did not result in better quality studies; choosing other criteria for limiting inclusion such as length of follow-up or the same definition of Barrett's oesophagus might have made synthesis of the results easier, but probably would not have altered the conclusions in the absence of agreed quality criteria by which to assess case series.

Cost-effectiveness

Three cost–utility analyses of surveillance of Barrett's oesophagus were identified, of which one was a further development of a previous study by the same group. Both sets of authors used Markov modelling and confined their analysis to 50- or 55-year-old white men with GORD symptoms. The models were run either for 30 years or to age 75 years.

The first two studies used a Markov model to examine various surveillance and treatment strategies for Barrett's oesophagus. The earlier study found that surveillance of Barrett's oesophagus every 5 years compared with no surveillance was cost-effective, but that the model was very sensitive to the incidence of adenocarcinoma and quality of life (utility value) in the post-oesophagectomy state. The later study from the same authors reached similar conclusions, but the incremental cost-effectiveness ratio for 5-yearly surveillance was no longer within the range usually considered cost-effective.

The third study also used a Markov model to examine various surveillance strategies. The authors concluded that the only cost-effective strategy was once in a lifetime screening of 50-year old white men with GORD, followed by surveillance of those with dysplasia only. Surveillance of non-dysplastic Barrett's oesophagus was not found to be cost-effective.

Both of these models are American, so there are almost certainly differences in practice from the UK and possible underlying differences in the epidemiology and natural history of the disease. In the UK, there is a major difficulty in knowing what proportion of patients with GORD have an endoscopy and at what stage of the disease, whereas in the USA, those who present to health services are more likely to be investigated at an earlier stage. The costs of the procedures involved are also likely to be very different.

Expert workshop

The group which discussed the epidemiology and natural history of Barrett's oesophagus identified six possible questions concerning areas of uncertainty, of which the following was rated as the clear key priority:

- What contributions do risk factors (demographic, environmental, genetic, molecular) for progression of Barrett's oesophagus make, individually and together, to the development of HGD and adenocarcinoma of the oesophagus?

The group that discussed diagnostic tests for Barrett's oesophagus identified seven possible areas of uncertainty. The key priority recognised that the ultimate aim of surveillance of Barrett's oesophagus is to reduce the risk of ACO:

- Is there a technique that we can use in the general population to identify patients with high risk of adenocarcinoma?

The group discussing treatment of Barrett's oesophagus identified seven possible areas of uncertainty and rated two of them as top priorities:

- How effective are any treatments for Barrett's oesophagus in altering cancer incidence?
- How can we best identify those at risk in order to target treatment?

►

The group discussing the potential impact of surveillance programmes identified nine possible areas of uncertainty and rated two of them as top priorities:

- Should we survey at all?
- Are there clinical subgroups at higher risk of adenocarcinoma?

The final plenary discussion at the workshop brought out questions that were of specific concern for the patient representatives and also identified additional questions that the small group discussions had not raised. No attempt was made to allocate an overall priority to these areas of uncertainty.

Cost–utility model

PenTAG's Markov model suggests that the base case scenario of endoscopic surveillance of Barrett's oesophagus at 3-yearly intervals, with low-grade dysplasia (LGD) surveyed yearly and HGD 3-monthly, does more harm than good when compared with no surveillance. Surveillance produces fewer quality-adjusted life-years (QALYs) for higher cost than no surveillance, therefore it is dominated by no surveillance. The cost per cancer identified approaches £45,000 in the surveillance arm and there is no apparent survival advantage owing to high recurrence rates and increased mortality due to more surgical interventions (i.e. oesophagectomies) in this arm.

The input parameters to which the model is most sensitive, in some cases reversing the results so that surveillance becomes cost-effective, are as follows:

1. the rate of recurrence of adenocarcinoma after oesophagectomy in the surveillance compared with the no surveillance arm
2. the rate at which adenocarcinoma becomes symptomatic once it has developed
3. the utility value (quality of life) attached to the health states for Barrett's oesophagus.

According to one-way sensitivity analyses, which vary just one model input while all the others are fixed, for 3-yearly surveillance to become cost-effective at usual levels of willingness to pay (£30,000 per QALY), the following parameters would need to achieve the following values:

1. if the rate of recurrence of adenocarcinoma after oesophagectomy reduces to 4.5% in the surveillance arm (from the base case of 9.3%) or
2. if the rate of recurrence of adenocarcinoma after oesophagectomy reduces to 7% in the non-surveillance arm (from the base case of 26%) or
3. if progression from undetected to symptomatic adenocarcinoma increases to at least 23% per year (from the base case of 14.3%) or
4. if utility values for Barrett's oesophagus health states fall to ≤ 0.63 (from the base case of 0.81).

These need to be viewed with caution given the uncertainty around many of the model variables. Less drastic alterations in the inputs made in combination could also change the model results. Nonetheless, these scenarios may well be realistic, given the current uncertainty in the literature about the true values for many parameters. The only inherently unrealistic scenario, in current practice, is a utility (quality of life) value for the post-oesophagectomy of nearly unity, which would imply that most people recover from this major procedure to virtually perfect health – an assumption not supported by the literature.

There must be considerable uncertainty about the impact of Barrett's oesophagus on quality of life, given that many people may be asymptomatic. Our model assumed that patients with Barrett's oesophagus referred for endoscopy would have symptoms, and that there would be equal numbers of those with mild, moderate and severe symptoms of GORD as rated by PenTAG's Value of Health Panel (this is a general population panel trained in standard gamble methods). A utility value of 0.81 was given for Barrett's oesophagus. Population norms for the relevant age range are 0.8 using a UK sample and derived from the EQ5D.

Non-surveillance continues to cost less and result in better quality of life whatever the surveillance intervals for Barrett's oesophagus and dysplastic states and whatever the costs (including none) attached to endoscopy and biopsy as the surveillance test.

The probabilistic analyses assess the overall uncertainty in the model. According to this, it is very unlikely that surveillance will be cost-effective even at relatively high levels of willingness to pay. The simulation showed that, in the majority of model runs, non-surveillance continued to cost less and result in better quality of life than surveillance.

At the population level (i.e. people with Barrett's oesophagus in England and Wales), a value of £6.5 million is placed on acquiring perfect information about surveillance for Barrett's oesophagus using expected value of perfect information (EVPI) analyses. This is if the technology (surveillance) is assumed to be relevant over 10 years. As with the one-way sensitivity analyses, the partial EVPI

highlighted recurrence of ACO after surgery and time taken for ACO to become symptomatic as particularly important parameters in the model.

Gaps in the evidence

Most of the published data on Barrett's oesophagus and surveillance come from uncontrolled case series. Reporting of data was generally poor in the studies included in this review.

Few data are available in the literature on the natural history of Barrett's oesophagus, particularly around the progression of Barrett's oesophagus through dysplastic states to ACO and then progression to symptomatic adenocarcinoma. Prevalence of Barrett's oesophagus in the general population and the clinical characteristics of the population presenting for endoscopy are also not well described. Follow-up in most studies is relatively short.

No data were identified on the performance of endoscopy as a test for identifying progression of Barrett's oesophagus to dysplasia or adenocarcinoma.

The current evidence base suggests that there is no intervention yet proven to reduce cancer risk in patients with Barrett's oesophagus, regardless of control of symptoms or regression of Barrett's oesophagus changes to normal.

The major gap in the evidence is the lack of RCT data on the effectiveness of surveillance programmes in reducing morbidity and mortality from adenocarcinoma. The lack of standard diagnostic criteria, diagnostic methods and surveillance intervals all hamper comparison between studies of surveillance programmes.

Possible specific harms of surveillance, either due to physical or psychological/emotional adverse effects, of Barrett's oesophagus are not generally reported in the studies identified here.

Conclusion

The systematic review concludes that there is insufficient evidence available to assess the clinical effectiveness of surveillance programmes of Barrett's oesophagus. There are numerous gaps in the evidence, of which the lack of RCT data is the major one. The expert workshop reflected these gaps in the range of topics raised as important in answering the question of the effectiveness of surveillance. Previous models of cost-effectiveness have most recently shown that surveillance programmes either do more harm than good compared with no surveillance or are unlikely to be cost-effective at usual levels of willingness to pay.

The PenTAG cost–utility model has shown that, across a range of values for the various parameters that have been chosen to reflect uncertainty in the inputs, it is likely that surveillance programmes do more harm than good. They cost more and confer lower quality of life than no surveillance.

Probabilistic analysis shows that, in most cases, surveillance does more harm and costs more than no surveillance. It is unlikely, but still possible, that surveillance may prove to be cost-effective. The cost-effectiveness acceptability curve, however, shows that surveillance is unlikely to be cost-effective at either the 'usual' level of willingness to pay (£20,000–30,000 per QALY) or at much higher levels. The expected value of perfect information at the population level is £6.5 million.

Recommendations for further research

Further research is required before the question of the effectiveness and cost-effectiveness of surveillance of Barrett's oesophagus in reducing morbidity and mortality from ACO can be answered with confidence. In addition, such evidence may form a vital part of any education programme for clinicians to support the decision to continue or cease surveillance. Future research should target both the overall effectiveness of surveillance and the individual elements that contribute to a surveillance programme, particularly the performance of the test and the effectiveness of treatment for both Barrett's oesophagus and ACO. In addition, of particular importance is the clarification of the natural history of Barrett's oesophagus. More detailed research proposals will be discussed separately with the HTA programme to inform their commissioning process.

Publication

Garside R, Pitt M, Somerville M, Stein K, Price A, Gilbert N. Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling. *Health Technol Assess* 2006;**10**(8).

NHS R&D HTA Programme

The research findings from the NHS R&D Health Technology Assessment (HTA) Programme directly influence key decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC) who rely on HTA outputs to help raise standards of care. HTA findings also help to improve the quality of the service in the NHS indirectly in that they form a key component of the 'National Knowledge Service' that is being developed to improve the evidence of clinical practice throughout the NHS.

The HTA Programme was set up in 1993. Its role is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The HTA Programme commissions research only on topics where it has identified key gaps in the evidence needed by the NHS. Suggestions for topics are actively sought from people working in the NHS, the public, service-users groups and professional bodies such as Royal Colleges and NHS Trusts.

Research suggestions are carefully considered by panels of independent experts (including service users) whose advice results in a ranked list of recommended research priorities. The HTA Programme then commissions the research team best suited to undertake the work, in the manner most appropriate to find the relevant answers. Some projects may take only months, others need several years to answer the research questions adequately. They may involve synthesising existing evidence or conducting a trial to produce new evidence where none currently exists.

Additionally, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme is able to commission bespoke reports, principally for NICE, but also for other policy customers, such as a National Clinical Director. TARs bring together evidence on key aspects of the use of specific technologies and usually have to be completed within a short time period.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees 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.

The research reported in this monograph was commissioned and funded by the HTA Programme on behalf of NICE as project number 03/49/01. The protocol was agreed in March 2004. The assessment report began editorial review in September 2004 and was accepted for publication in August 2005. 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 referees 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.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme, NICE or the Department of Health.

Editor-in-Chief: Professor Tom Walley
Series Editors: Dr Peter Davidson, Dr Chris Hyde, Dr Ruairidh Milne, Dr Rob Riemsma and Dr Ken Stein
Managing Editors: Sally Bailey and Sarah Llewellyn Lloyd

ISSN 1366-5278

© Queen's Printer and Controller of HMSO 2006

This monograph may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Applications for commercial reproduction should be addressed to NCCHTA, Mailpoint 728, Boldrewood, University of Southampton, Southampton, SO16 7PX, UK.

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.