Why is evaluation of the cost effectiveness of audit so difficult? The example of thrombolysis for suspected acute myocardial infarction

Michael B Robinson, Elizabeth Thompson, Nick A Black

Abstract

Background—Cost effectiveness analysis is an established technique for evaluation of delivery of health care, but its use to evaluate clinical audit is rarely reported. Thrombolysis for suspected acute myocardial infarction is a commonly used therapy of established effectiveness and an appropriate subject for audit in many healthcare settings.

Objective—To measure the cost effectiveness of audit of thrombolysis in some district general hospitals.

Main outcome measure—Cost of audit per extra patient treated with thrombolysis (incremental cost effectiveness ratio).

Design—Prospective agreement with physicians to undertake repeated audits of a specific aspect of the management of patients with acute myocardial infarction. Baseline measurement of the proportion of these patients given thrombolysis in each hospital were made, as were three subsequent retrospective audits, giving time series of measurements. Costs were estimated from records of staff time and other resources used in each hospital; effectiveness was estimated by fitting the results to a model which assumed a uniform rate of increase over time in the proportion of eligible patients given thrombolysis which might be accelerated by regular audit. Upper and lower limits for main outcome measure were derived from sensitivity analysis of costs and logistic regression of time series data.

Setting—Five district general hospitals in North West Thames Regional Health Authority including one control hospital were used, starting in April 1991 when widespread medical audit was first introduced.

Results—Between the first and last audits, the proportion of patients with suspected acute myocardial infarctions receiving thrombolysis rose in three of the hospitals undertaking audit by 20% to 37% and fell by 6% in the fourth (although this hospital started with a rate in excess of 90%). The corresponding change in the control hospital was an increase of 15%. The differences between each of the auditing hospitals and control hospital were not significant, except in one case, where 51 extra treatments per year were attributable to audit (95% confidence intervals (95% CIs) 0.5 to 61 cases per year).

Estimated overall costs in each hospital ranged from £3700 to £5200 for data collection, a series of four audit meetings, and subsequent actions. The central estimate of cost effectiveness in the three responsive hospitals ranged from £101 to £392 per extra case given thrombolysis, with very wide 95% CIs. In the fourth hospital audit had zero effectiveness as defined in this study.

Conclusions—Methodological difficulties were encountered which need to be considered in future economic evaluations of clinical audit and related activities. These were: (a) adequate control for other factors influencing clinical behaviour; (b) uncertainties about the sustainability of changes in behaviour associated with audit; and (c) the relative infrequency in a single hospital of specific clinical events leading to small numbers for analysis. These difficulties constitute major challenges for the economic evaluation of clinical audit. They are most likely to be overcome in a large study which compares clinical audit with other interventions aiming for the same quality improvement, such as patient specific reminders or educational programmes.

Keywords: audit; acute myocardial infarction; thrombolysis; cost effectiveness

Introduction

From 1991 onwards, the United Kingdom health departments have directed substantial resources to the development of clinical audit, and its introduction throughout the National Health Service (NHS). Although audit has become an accepted part of good clinical practice,1 there are still doubts about whether it is achieving its goal of ensuring high quality care.2–4 Similar uncertainties have been expressed in the United States,5 where audit has been a routine requirement since 1974.6

Rigorous evaluation of audit is therefore required, if the best use is to be made of overall healthcare resources.7 The diversity of benefits expected from audit means that a range of evaluative methods is desirable,8 including cost effectiveness analysis which was originally developed to inform investment decisions in the public sector.9 It has been widely used to evaluate specific aspects of delivery of health care.10 Although there are difficulties, particularly with the generalisation of results,11 standard techniques have been established.12 Eco-
Table 1  Types of costs identified

<table>
<thead>
<tr>
<th>Type of cost</th>
<th>Audit itself</th>
<th>Resulting patient care</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour</td>
<td>Time required by various staff to set up audit programme (meetings to agree criteria, and to obtain access to registers and case notes)</td>
<td>Nursing and medical time associated with additional use of thrombolysis (setting up infusions, treating side effects)</td>
</tr>
<tr>
<td></td>
<td>Preparation for each audit meeting Attendance at audit meetings</td>
<td>Nursing and medical time required to look after increased numbers of survivors of acute myocardial infarction associated with additional thrombolysis</td>
</tr>
<tr>
<td></td>
<td>Time required by medical staff to comply with audit intervention (form filling, extra history, and examination)</td>
<td></td>
</tr>
<tr>
<td>Supplies</td>
<td>Paper and overheads for analysing and presenting observations and implementing agreed changes (circulation of guidelines)</td>
<td>Extra doses of thrombolytic drugs General supplies associated with hospital stays of extra survivors</td>
</tr>
<tr>
<td>Overheads</td>
<td>Consumption of extra heat and lighting Cost of occupying physical space (lecture theatres for audit meetings, desks for case note abstraction, etc)</td>
<td>Extra heat and lighting Increase in overheads due to additional bed days</td>
</tr>
<tr>
<td>Equipment</td>
<td>Faster wear and tear on photocopiers, medical records, and computers</td>
<td>Wear and tear on ward and nursing equipment</td>
</tr>
</tbody>
</table>

nomic evaluation of the use of clinical guidelines has been proposed, but there are no reports of its specific application to clinical audit in the United Kingdom.

A study was therefore designed to carry out a cost-effectiveness analysis of one part of the clinical audit programme which began to be developed in district general hospitals in the United Kingdom after the 1991 NHS reforms. If satisfactory results could be obtained from this initial study, there would be considerable scope for more widespread use to influence subsequent developments.

Two major difficulties had to be considered in designing the study. Firstly, clinical audit within the NHS was set up under professional not managerial control. Imposition of a specific model or method of audit on participants therefore ran the risk of creating an artificial and atypical example for evaluation, severely limiting external validity. It was thought that this ruled out an experimental design—such as a randomised trial. However, at the time of the study most of the participating physicians had little practical experience of audit so some guidance from the researchers on what might be expected on theoretical grounds to make a successful audit seemed appropriate.

The second anticipated difficulty was to ensure that the audit activity would be suitable for economic evaluation. Many of the quality improvements sought through clinical audit have no clear link to better health outputs or increased efficiency in economic terms. Unless the standards sought by the audit were known to be associated with good outcomes, the relevance of a cost-effectiveness analysis would be questionable.

The following solution was adopted. (1) An agreement with participating physicians in each hospital for repeated audits of the same topic every three to four months for one year, with the option of changing the topic if further audit of the original one seemed unnecessary. (2) Choice of topic was restricted to those with established evidence of clinical effectiveness. For the chosen topic of thrombolysis, a meta-analysis including 58,600 patients has shown an 18% reduction in 35 day mortality. Moreover, a related economic model has shown that the range of assumptions under which the treatment has a positive benefit-cost margin is very wide. Therefore the assumption that an increase in the rate of use of thrombolysis for suspected acute myocardial infarction leads to better overall outcome seemed highly reasonable.

Methods

SELECTION OF HOSPITALS, AUDIT TOPIC, AND PATIENTS

These have been described in detail elsewhere. Briefly, four district general hospitals in the North West Thames Region were identified about one year after the launch of the Department of Health’s audit initiative. Hospitals were selected if the consultant physicians were neither opposed to nor very advanced in their implementation of audit. After a small retrospective case note survey and a review of the scientific literature on the appropriate clinical management of patients presenting with acute chest pain, a shortlist of possible topics for audit was discussed at a meeting attended by a leading consultant from each hospital. The use of intravenous thrombolysis in patients with suspected acute myocardial infarction was chosen because its effectiveness had recently been confirmed by several large trials, yet the case note survey had suggested low rates of use. The sampling frame used in each hospital was the accident and emergency department register, supplemented where necessary by ward registers. The aim was to identify retrospectively a consecutive series of patients presenting with chest pain initially diagnosed as acute myocardial infarction. After the patients were discharged, data were abstracted from all case notes where acute myocardial infarction was mentioned at least as a possible initial diagnosis. Those patients who seemed to have been eligible for thrombolysis during their admission were identified with a predefined algorithm and agreed list of contraindications, common to all hospitals.

A fifth, control hospital was chosen in a similar manner to the four study hospitals, but was not approached until the audits elsewhere had been completed, to avoid any change in clinical behaviour arising in the control hospital simply from the knowledge that their practice was being monitored. Patients in the control hospital were identified in the same way and for the same periods as had been used in the study hospitals.

The reliability of case note abstraction and classification of eligible cases into three categories—definite, likely, and possible—has been reported elsewhere.

AUDIT CYCLES

Baseline data were collected from the case notes of patients admitted during a two month period during April and May 1991. Available research assistant time limited this to a maximum of 50 consecutive cases in each of the four study hospitals. The information obtained was presented by a member of the department register, supplemented where necessary by ward registers. The aim was to identify retrospectively a consecutive series of patients presenting with chest pain initially diagnosed as acute myocardial infarction. After the patients were discharged, data were abstracted from all case notes where acute myocardial infarction was mentioned at least as a possible initial diagnosis. Those patients who seemed to have been eligible for thrombolysis during their admission were identified with a predefined algorithm and agreed list of contraindications, common to all hospitals.

A fifth, control hospital was chosen in a similar manner to the four study hospitals, but was not approached until the audits elsewhere had been completed, to avoid any change in clinical behaviour arising in the control hospital simply from the knowledge that their practice was being monitored. Patients in the control hospital were identified in the same way and for the same periods as had been used in the study hospitals.

The reliability of case note abstraction and classification of eligible cases into three categories—definite, likely, and possible—has been reported elsewhere.
Why is evaluation of the cost effectiveness of audit so difficult

Table 2  Summary of audit costs in each hospital (4 meetings in each hospital)

<table>
<thead>
<tr>
<th>Hospital</th>
<th>H1</th>
<th>H2</th>
<th>H3</th>
<th>H4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs per meeting:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparation time</td>
<td>£374</td>
<td>£367</td>
<td>£498</td>
<td>£425</td>
</tr>
<tr>
<td>Attendance time</td>
<td>£285</td>
<td>£282</td>
<td>£410</td>
<td>£283</td>
</tr>
<tr>
<td>Materials</td>
<td>£5</td>
<td>£5</td>
<td>£5</td>
<td>£5</td>
</tr>
<tr>
<td>Compliance time</td>
<td>£328</td>
<td>£188</td>
<td>£288</td>
<td>£297</td>
</tr>
<tr>
<td>Total</td>
<td>£992</td>
<td>£842</td>
<td>£1201</td>
<td>£1010</td>
</tr>
<tr>
<td>Total costs of meetings</td>
<td>£3966</td>
<td>£3370</td>
<td>£4803</td>
<td>£3040</td>
</tr>
<tr>
<td>Initial set up costs</td>
<td>£323</td>
<td>£323</td>
<td>£354</td>
<td>£323</td>
</tr>
<tr>
<td>Overall cost of audit</td>
<td>£4290</td>
<td>£3693</td>
<td>£3517</td>
<td>£4363</td>
</tr>
<tr>
<td>Sensitivity analysis:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper limit (compliance time per case × 3)</td>
<td>£5745</td>
<td>£4156</td>
<td>£6109</td>
<td>£5332</td>
</tr>
<tr>
<td>Lower limit (compliance time per case × 0.5)</td>
<td>£3926</td>
<td>£3577</td>
<td>£4918</td>
<td>£4121</td>
</tr>
</tbody>
</table>

The value of labour was calculated from basic annual salaries, taking the mid-point of the relevant scales, and adding 15% to cover superannuation. Consultants’ salaries were further inflated to take account of the average prevalence of merit awards within the specialty.20 Audit officers, who were involved in setting up meetings in some of the study hospitals, were assumed to be paid at the same rate as a research assistant. Market values were used for the cost of materials, which in any case was a small fraction of the total.

sensitivity analysis

The most uncertain part of the costing was the extra time needed by junior doctors to process each patient with suspected acute myocardial infarction as a result of the audit—for example, in writing more detailed case notes knowing that they were to be reviewed, or in re-registering hospital guidelines before giving care. In some cases, no extra time may have been needed, because junior doctors were writing adequate notes and checking guidelines even before the audit began. The total costs were therefore recalculated, with a lower estimate (of 50% of the central estimate) and an upper estimate (of three times the central estimate, table 2). The lead consultant in each hospital was asked to check the completed cost analysis for face validity.

The aim was to identify, measure, and value the resources consumed as a result of the audits which would otherwise have been available for other activities. Methods were designed to give an estimate of the costs involved and a sensitivity analysis was undertaken to explore the uncertainty of these estimates.

Table 1 shows the types of costs which were initially considered, divided into those arising from the audit itself, and those due to resultant changes in patient care. The purpose of the evaluation was to assess cost effectiveness of audit, in comparison with other mechanisms to change patient care. Hence the costs of changes in direct patient care arising from audit were not considered further, as these would have arisen whatever method of quality assurance was being used. A model for a wider economic evaluation of audit in comparison with other NHS activities, incorporating treatment as well as audit costs, is described elsewhere.19

Certain costs were incurred because the audits were being set as part of a research study—for example, preliminary meetings with the consultant staff in each hospital and the maintenance of diaries of time spent on the project. These would not have been incurred if the same audit method had been used again elsewhere in routine practice and so they were also excluded from the analysis.

Some of the costs shown in table 1 were left out, either because they were likely to be trivial (wear and tear of equipment) or because the audit was taking up unused capacity rather than displacing other work (physical space). Overheads were omitted because they constituted a small and fairly constant fraction of hospital expenditure.

Labour costs were judged likely to be much larger than any other type, and so most effort was focused on these. A log of the time spent by the research team on audit as opposed to activities related to the evaluation was kept. Observations of time spent by medical staff—for example, attendance at audit meetings—were similarly noted. Some relevant activity—such as discussion to set standards and compliance with agreed interventions—could not be directly measured, so estimates were made after questioning the staff.

The value of labour was calculated from basic annual salaries, taking the mid-point of the relevant scales, and adding 15% to cover superannuation. Consultants’ salaries were further inflated to take account of the average prevalence of merit awards within the specialty.20 Audit officers, who were involved in setting up meetings in some of the study hospitals, were assumed to be paid at the same rate as a research assistant. Market values were used for the cost of materials, which in any case was a small fraction of the total.

MEASUREMENT OF COSTS

The system was organised as part of the hospital’s audit programme and was chaired by one of the consultant physicians. Each meeting focused on the results for that particular hospital, although an anonymised comparison of the four study hospitals was also included. The doctors were asked to set explicit standards for the proportion of definite, likely, and possible patients with acute myocardial infarctions who should receive thrombolysis, taking account of their baseline results and the recent and extensive medical literature on the subject.

The baseline audit was followed by three audit cycles each of about four months in duration. Each cycle consisted of a period of observation, data collection, a feedback meeting, and subsequent intervention if any were judged necessary by the doctors concerned.

Each meeting focused on the results for that hospital, which all relevant medical staff were invited. Each meeting was organised as part of the hospital’s audit programme and was chaired by one of the consultant physicians. Each meeting focused on the results for that particular hospital, although an anonymised comparison of the four study hospitals was also included. The doctors were asked to set explicit standards for the proportion of definite, likely, and possible patients with acute myocardial infarctions who should receive thrombolysis, taking account of their baseline results and the recent and extensive medical literature on the subject.

The baseline audit was followed by three audit cycles each of about four months in duration. Each cycle consisted of a period of observation, data collection, a feedback meeting, and subsequent intervention if any were judged necessary by the doctors concerned.

The value of labour was calculated from basic annual salaries, taking the mid-point of the relevant scales, and adding 15% to cover superannuation. Consultants’ salaries were further inflated to take account of the average prevalence of merit awards within the specialty.20 Audit officers, who were involved in setting up meetings in some of the study hospitals, were assumed to be paid at the same rate as a research assistant. Market values were used for the cost of materials, which in any case was a small fraction of the total.

SENsITIVITY ANALYSIS

The most uncertain part of the costing was the extra time needed by junior doctors to process each patient with suspected acute myocardial infarction as a result of the audit—for example, in writing more detailed case notes knowing that they were to be reviewed, or in re-registering hospital guidelines before giving care. In some cases, no extra time may have been needed, because junior doctors were writing adequate notes and checking guidelines even before the audit began. The total costs were therefore recalculated, with a lower estimate (of 50% of the central estimate) and an upper estimate (of three times the central estimate, table 2). The lead consultant in each hospital was asked to check the completed cost analysis for face validity.

The system was organised as part of the hospital’s audit programme and was chaired by one of the consultant physicians. Each meeting focused on the results for that particular hospital, although an anonymised comparison of the four study hospitals was also included. The doctors were asked to set explicit standards for the proportion of definite, likely, and possible patients with acute myocardial infarctions who should receive thrombolysis, taking account of their baseline results and the recent and extensive medical literature on the subject.

The baseline audit was followed by three audit cycles each of about four months in duration. Each cycle consisted of a period of observation, data collection, a feedback meeting, and subsequent intervention if any were judged necessary by the doctors concerned.

The value of labour was calculated from basic annual salaries, taking the mid-point of the relevant scales, and adding 15% to cover superannuation. Consultants’ salaries were further inflated to take account of the average prevalence of merit awards within the specialty.20 Audit officers, who were involved in setting up meetings in some of the study hospitals, were assumed to be paid at the same rate as a research assistant. Market values were used for the cost of materials, which in any case was a small fraction of the total.

Models used to calculate effectiveness of audit.
### Table 3 Period of observation for each audit, number eligible, number (%) treated, and annual number eligible by hospital

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Audit (n)</th>
<th>Start and end dates</th>
<th>Eligible (n)</th>
<th>Treated (n (%))</th>
<th>Annual eligible (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>1</td>
<td>1.4–17.5.91</td>
<td>17</td>
<td>16 (94)*</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9–3.11.91</td>
<td>28</td>
<td>24 (86)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.1–1.3.92</td>
<td>31</td>
<td>29 (94)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8.4–8.6.92</td>
<td>35</td>
<td>30 (86)</td>
<td></td>
</tr>
<tr>
<td>H2</td>
<td>1</td>
<td>1.4–11.5.91</td>
<td>15</td>
<td>9 (60)*</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9–10.10.91</td>
<td>24</td>
<td>21 (87)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.1–20.2.92</td>
<td>23</td>
<td>19 (83)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10.4–26.6.92</td>
<td>28</td>
<td>26 (93)</td>
<td></td>
</tr>
<tr>
<td>H3</td>
<td>1</td>
<td>1.4–19.4.91</td>
<td>19</td>
<td>11 (58)*</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9–7.10.91</td>
<td>26</td>
<td>23 (88)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9.4–1.6.92</td>
<td>26</td>
<td>20 (77)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10.4–26.6.92</td>
<td>28</td>
<td>26 (93)</td>
<td></td>
</tr>
<tr>
<td>H4</td>
<td>1</td>
<td>1.4–22.5.91</td>
<td>14</td>
<td>8 (57)*</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9–17.10.91</td>
<td>22</td>
<td>16 (73)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.1–20.2.92</td>
<td>19</td>
<td>17 (89)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9.4–1.6.92</td>
<td>26</td>
<td>20 (77)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>1.4–4.5.91</td>
<td>19</td>
<td>10 (53)*</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9–20.10.91</td>
<td>28</td>
<td>17 (61)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.1–20.2.92</td>
<td>24</td>
<td>17 (71)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10.4–29.5.92</td>
<td>25</td>
<td>17 (68)</td>
<td></td>
</tr>
</tbody>
</table>

*% Treated in first audit in each hospital.

### MEASUREMENT OF EFFECTIVENESS

Effectiveness was defined as the number of extra patients in each hospital expected to receive thrombolysis during a nominal one year period after the series of audits was completed, beyond that predicted from the control hospital data (figure). Calculation required fitting the data to an empirical model. Their limitations and those of alternative models will be considered in the discussion.

The main assumptions of the calculations were:

1. The proportion of patients treated during a given audit could be represented by a point estimate at the mid-point of the observation period of that particular audit.

2. The period during which audit was actively influencing the rate of thrombolysis was exactly one year in all hospitals, starting at the mid-point of the baseline observation period (about 1 May 1991). The actual interval between the first and last audits was slightly longer than one year.

3. During the auditing period, the use of thrombolysis, expressed as the log odds of the proportion treated, would increase at a constant rate. This rate was estimated by logistic regression of the point estimates obtained from each of the four audits carried out during the year.

4. Without audit, the use of thrombolysis would have increased at a rate given by logistic regression of the control hospital’s data.

5. After the active period of one year, the proportion of patients given thrombolysis would remain constant for the next year. The effect of audit would last only for this second year. Clinical performance in subsequent years was assumed to be too distant from the series of audits to be influenced by it, given the rate of turnover of junior staff and the frequency of publication of new research evidence.

### DETAILS OF CALCULATION

Standard spreadsheet software was used. Separate calculations were performed for each hospital and copies of the spreadsheets are available on request. The mid-point of each observation period was expressed as a Julian date—that is, the number of days elapsed since 1 January 1900. This date and the proportion of patients treated for each audit were then transferred to a statistical software package for logistic regression.

The following analyses were performed.

1. Logistic regression of the proportions treated against time. This regression coefficient ($R_2$) represented the average change in log odds of treatment per day between the first and last audits.

2. Multiple logistic regression, combining data used in (1) with equivalent data from the control hospital, using a dummy variable. This produced a common intercept (not used in the model) and a regression coefficient ($R_2$) representing the difference in the rate of change of use of thrombolysis between the two hospitals, and 95% confidence intervals (95% CIs) for this difference ($R_{2u}$ and $R_{2l}$).

The results were transferred back to the spreadsheet. The baseline observation for each study hospital (p in the figure) and the control hospital gradient ($R_1$ control) were used to calculate the proportion expected after one year if no audit had taken place (p in the figure), with the formula:

$$\ln(p_{o}/1-p_{o}) = \ln(p/1-p) + 365(R_1 \text{control}).$$

The gradient derived from the study hospital’s own audits was likewise used to calculate the proportion expected with audit after one year (p in the figure). The difference, $p_{o} - p$, was multiplied by the annual number of eligible cases (N) to give the extra number of cases treated with thrombolysis in one year; N was calculated as a weighted average of the numbers observed during each audit in the hospital concerned—that is:

$$N = \sum_{i=1}^{4} \frac{\text{duration in days of audit } i}{\text{duration of all 4 audits}} \times \text{No of cases observed in audit } i.$$

Upper and lower limits for the extra number of cases treated were derived from the sum of the control gradient and the 95% range for the difference between control and audit gradients in place of the audit gradient in the spreadsheet—that is, $R_{2u}(\text{hosp } x)$ was substituted by $(R_{2u}(\text{control}) + R_{2u}(\text{hosp } x))$ to give the upper limit, and $(R_{2l}(\text{control}) + R_{2l}(\text{hosp } x))$ for the lower limit. When these calculations suggested a negative effect for audit this was ignored and the lower bound for the additional effect of audit was set at zero.

### COST EFFECTIVENESS

The central estimate of cost effectiveness for each hospital was calculated by dividing the total cost of audit by the estimated number of extra cases given thrombolysis over one year. The upper limit was calculated by substituting the upper estimate of costs from the sensitivity analysis and the lower 95% CI for effectiveness, and vice versa for the lower bound. Where the lower 95% CI suggested a negative effect of audit on the treatment proportion this was replaced by zero, representing no change as a result of audit and no cost effectiveness.
Table 4 Regression coefficients; predicted one year treatment rates; equivalent number of extra cases treated, by hospital

<table>
<thead>
<tr>
<th>Hospital</th>
<th>H1</th>
<th>H2</th>
<th>H3</th>
<th>H4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression coefficients:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Change in log odds treatment/day=x10^4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual hospitals data (R_i)</td>
<td>1.4</td>
<td>4.7</td>
<td>7.3</td>
<td>3.0</td>
</tr>
<tr>
<td>(Standard error of R_i)</td>
<td>(2.5)</td>
<td>(2.1)</td>
<td>(2.2)</td>
<td>(2.0)</td>
</tr>
<tr>
<td>Difference between individual hospital and control hospital:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower 95% limit (R_{lo})</td>
<td>-9.0</td>
<td>-2.4</td>
<td>0.03</td>
<td>-3.9</td>
</tr>
<tr>
<td>Central estimate</td>
<td>-3.3</td>
<td>2.8</td>
<td>5.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Upper 95% limit (R_{uo})</td>
<td>2.4</td>
<td>7.9</td>
<td>10.7</td>
<td>6.0</td>
</tr>
<tr>
<td>Treatment rates after one year (%):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without audit (p_i)</td>
<td>97</td>
<td>75</td>
<td>74</td>
<td>73</td>
</tr>
<tr>
<td>With audit (p_{i+1})</td>
<td>97*</td>
<td>75*</td>
<td>74*</td>
<td>73*</td>
</tr>
<tr>
<td>Central estimate</td>
<td>97*</td>
<td>89</td>
<td>95</td>
<td>80</td>
</tr>
<tr>
<td>Upper 95% limit</td>
<td>99</td>
<td>98</td>
<td>99</td>
<td>96</td>
</tr>
<tr>
<td>Extra cases due to audit:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower 95% limit</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>Central estimate</td>
<td>0</td>
<td>21.5</td>
<td>51.3</td>
<td>11.1</td>
</tr>
<tr>
<td>Upper 95% limit</td>
<td>3.0</td>
<td>35.1</td>
<td>61.0</td>
<td>37.3</td>
</tr>
</tbody>
</table>

*Audit ineffective.

Symbols (R_{i}, R_{lo}, R_{uo}, P, P_i) explained in text. Regression coefficients derived from data in table 3. Treatment rates after one year calculated from P (shown in table 2), R_{i}, R_{lo}, R_{uo}, as described in text. Extra cases due to audit = (P−P_i) × annual number eligible (shown in table 2).

Results

Evidence of the effectiveness of audit is presented first, then indication of the costs which were incurred to achieve this, and the resultant cost effectiveness ratio.

Effectiveness

The baseline rate of use of thrombolysis in suspected acute myocardial infarction was 53% in the control hospital and varied from 57% to 94% in the study hospitals. One year later, the rate in the control hospital had increased by 15%, whereas the rates in the study hospitals (each of which had completed three audit cycles) had increased by 20% to 37% in three hospitals and decreased by 69% in the fourth. Table 3 shows details of data from all four audits in each hospital, and table 4 shows the regression output produced. Differences in the slope of the regression between the study and control hospitals produced an estimate of the actual number of extra cases in one year given thrombolysis as a result of the audits. Central estimates of this varied from no extra cases in hospital 1 to 51.3 in hospital 3. Table 4 shows the 95% CIs for the number of extra cases, and details of the interim steps in the calculation.

Costs

The overall cost of the programme of four audit meetings and associated preparation varied from £3693 in hospital 2 to £5157 in hospital 3 (at 1990 prices). Most of the individual cost items were similar across the four hospitals. Preparation time for obtaining case notes and abstracting data varied by a factor of 1.5, reflecting differences in the organisation and ease of location of record folders in the medical records department. The cost of staff attendance at meetings varied by a factor of 1.8, reflecting differences in the number of doctors involved. These were due to differences in the sizes of the medical establishments in general medicine rather than in the proportion attending audit meetings, which was about 70% in all hospitals. Table 2 shows details of the costs.

Altering assumptions about the extra time needed to process each admission (the most uncertain cost element) produced no change in the ranking of hospitals by the cost of the audit. Hospital 3 was always the most expensive and hospital 2 the least expensive. Table 2 shows details.

Cost effectiveness

In hospital 1 no extra patients had been given thrombolysis as a result of the audit, so there was no cost effectiveness. In the other three hospitals the central estimate varied roughly fourfold, from audit costs of £101 per extra case given thrombolysis in hospital 3 to £392 in hospital 4. Table 5 shows details.

The lower bounds in these three hospitals showed less variation, from £81 to £110, but the upper bounds were infinite in two of the three hospitals and over £12 000 per extra case in hospital 3. These infinite upper bounds arose because there was doubt whether any extra cases were treated as a result of audit in these two hospitals.

Discussion

The objective of the study was only partially met, in that cost effectiveness of the audit programme was measured in only one of the four hospitals studied, and then only with very limited precision. Given these disappointing results, why should attempts to measure cost effectiveness be pursued any further?

The answer is that implicit judgements of cost effectiveness of clinical audit programmes are being made every time a budget is agreed for audit or a specific programme is being set up. Intangible costs and benefits—such as the effect on morale and patient satisfaction—are being made every time a budget is agreed for audit or a specific programme is being set up. Intangible costs and benefits—such as the effect on morale and patient satisfaction—are being made every time a budget is agreed. Given that the broad objective of the study is sensible, the next major concern is whether the results of the evaluation have any relevance beyond the specific time and place where they were obtained. The study was undertaken at the beginning of the development of routine audit in the NHS, and the strategy of providing...
feedback on the same specific topic every three months is not one which has been widely adopted subsequently. In retrospect, it constitutes an unusually intense programme for one aspect of the management of a single condition. However, regular and frequent feedback was expected on theoretical grounds to be most likely to lead to changes in behaviour. The intensity of the programme was reflected in the costs measured, and the approach was practical for wider implementation if it were shown to be a cost effective one.

The remainder of this discussion will focus on why the results were uncertain, and how this might be avoided in future evaluations.

Four main areas of uncertainty can be distinguished, each of which will be considered in turn. These are:

1. Adequacy of control for other factors besides audit which could influence clinical behaviour in each hospital.
2. Statistical uncertainties, given the relatively few patients included in each audit, and the few audits conducted in each hospital.
3. Assumptions about the time course of changes in clinical performance with and without audit, especially their sustainability once direct observations had stopped.
4. Validity of using a process measure (% of eligible patients receiving thrombolysis) as a proxy for health outcome.

In each case, the critical question is whether the problem is specific to this particular study, or a more general one which will affect other attempts to use economic evaluation for clinical audit.

ADEQUACY OF CONTROL FOR CONFOUNDING FACTORS

The study design included only one control hospital because it was expected that this would confirm that with no change in clinical behaviour occurred. The unexpected result of an increasing trend in treatment rate in the control as well as in the auditing hospitals shows that other factors besides audit can change clinical performance. In retrospect, a more even distribution of auditing and non-auditing hospitals, perhaps with matched pairs would have been preferable, as the measurement of the changes to be expected without audit would then have been more widely based. This might have shown either a greater or lesser cost effectiveness than that quoted, depending on whether the additional controls showed less or more of a trend than the one actually used. This is a problem specific to this study which, given sufficient resources, can easily be overcome by future evaluations. However, comparing audit in hospitals against controls undertaking no quality assurance activity whatsoever is impractical, given that this is now a policy requirement. One solution is to compare audit programmes against other quality assurance activities with the same objectives—such as automatic patient specific reminders or educational programmes.

Another possible approach would be a case-control framework, retrospectively comparing hospitals which did and did not change practice, to look for differences in exposure to audit activities. However, such differences might be subtle, and so difficult to measure. Also, there would be numerous other confounding exposures which would need to be identified and taken into account.

In a matched pairs study some of these difficulties would be reduced, in that the audit exposure and related confounders could be followed prospectively. The ideal controls would be hospitals which were identical in every way to the study hospitals except in their audit arrangements. This is unattainable in reality except in exceptional and hence unrepresentative instances.

In summary, although it would be easy to increase the number of controls in subsequent studies, their optimum nature remains uncertain. This needs careful consideration in the design of future studies, and in the critical appraisal of their results.

STATISTICAL UNCERTAINTY

Having chosen a specific model as described above, why were the 95% CIs for the estimates of cost effectiveness so wide? Uncertainty about both costs and effectiveness was involved, but the element due to costs was relatively small—the sensitivity analysis ranged from 15% to 42% of the central cost estimates, depending on the hospital concerned.

Uncertainty about the effectiveness of audit was much greater. In one hospital the initial rate of use of thrombolysis was sufficiently high (over 90%) that there was little potential for further improvement. Further audits were bound to show limited effectiveness as defined in this study so it could be argued that comparison with alternative hospitals was unfair. Among these other hospitals, two showed a significant upward trend in rates of treatment over the course of the study, but when the rate of change in the control hospital was taken into account, this was significant in only one. This lack of power arose because (a) the numbers of eligible cases found in each audit were smaller than anticipated; and (b) the control hospital showed a greater than expected increase in the use of thrombolysis over the year of observation; (c) only four time intervals were studied in each hospital.

One way to increase statistical power would have been to compare the aggregate effect in all four study hospitals against the control. This was not considered appropriate because each study hospital was encouraged to tailor the details of the audit intervention to their own wishes and needs. This fitted with the concept of audit as a professionally led, not managerially driven, activity. In this sense audit is different from most direct patient care, which is externally managed (by a doctor) and hence more uniform in its process if not in its outcome.

A better way for future economic evaluations to be able to increase statistical power will be to study more patients for longer periods in each hospital. There is, however, a natural limit imposed by the frequency of specific clinical
Why is evaluation of the cost effectiveness of audit so difficult

events in any one hospital. Thrombolysis for acute myocardial infarction is a relatively common event compared with others which have been popular choices for audit in general medicine, such as gastrointestinal haemorrhage or diabetes. Aggregating data from several hospitals should increase numbers but there is the risk already explained that the intervention may differ between hospitals or if constrained to be uniform by the research design, then it may be unrepresentative of audit as currently practised. Also, relevance to medical staff in a given hospital and their sense of responsibility for what the data show will be reduced.

The best approach for future evaluations may be to concentrate on audits of processes which are common to patients with a wide range of diagnoses—such as the use of laboratory tests or non-specific treatments. This may be feasible in a wide range of medical specialties. The disadvantage of this approach is uncertainty about the health output associated with such measures.

TIME COURSE OF CHANGES IN CLINICAL BEHAVIOUR

Calculation of the effectiveness of audit required the adoption of a model of how clinical performance changed over time, the variables of which could be estimated from the collected data. The main features of the chosen model were (a) a linear trend in log odds of treatment during the data collection period, the gradient of which was increased by audit; (b) no further change in rates of treatment during the subsequent year; and (c) equalisation of rates of treatment with and without audit more than one year after the last audit.

An alternative approach to (a) would have been to assume that audit was associated with a one off increase in rates of treatment, followed by a linear trend identical to that found in the control hospital. Given that there were only four points in the time series for each hospital, it was not possible to discriminate between these alternatives. The single linear trend was preferred on the theoretical grounds that each audit, not just the first one, could be expected to change the behaviour of a proportion of the medical staff who were involved. Also, due to the rotation of junior staff, there were some newcomers at each audit meeting.

Feature (b) was critical to the size of effectiveness ascribed to audit, because this was defined in terms of extra patients treated during this 12 month period. No observations were made during this time, so an assumption was unavoidable. The alternatives to the chosen option of no change were a continuation of the linear trend from the previous 12 months, (i) in the auditing hospitals only and (ii) in both the auditing and the control hospital. Option (i) would have produced a higher figure for cost effectiveness, the exact size of which depended on the gradient from the auditing hospital. Option (ii) would have produced an estimate exactly 50% greater than that quoted in the results. These conclusions are reached by simple geometric considerations of how the shaded area in the figure would vary.

In a similar way, the inclusion of the difference in the year of observations as well as the subsequent year can be shown to lead to an increase in cost effectiveness of exactly 50%.

Which option is most valid is unknown. Resource constraints prevented measurement of the rate of treatment continuing into the second year, and existing publications provide no empirical evidence about the sustainability of changes in behaviour due to audit. The chosen option was preferred because it was simple to calculate and to show graphically (figure). A different option instead of feature (b) would not have affected the relative cost effectiveness across the four hospitals, only their absolute magnitude.

Feature (c) of the model—that is, equivalence with and without audit more than 12 months later—resembled feature (b) in that there was no empirical evidence to base it on. If audit were considered to have a more prolonged effect, this would have considerably increased cost effectiveness, the exact extent depending on the discount rate used.

The lack of empirical evidence about the sustainability of audit effectiveness is not a problem specific to this study. Longitudinal studies are required, which take into account other influences on practice as well as audit.

PROCESS MEASURES AS PROXIES FOR HEALTH OUTCOMES

Clinical audit may be effective in changing clinical practice, but such changes in practice may not lead to more effective and efficient health care. In this study there is unusually strong evidence that the process being initiated is an effective one. This justifies the assumption that reductions in the rate of thrombolysis, as was actually found in hospital 1 or shown to be a statistical possibility in the others, could be disregarded.

The possibility that baseline practice included some inappropriate use of thrombolysis which was subsequently corrected as a result of audit cannot be totally excluded, but seems unlikely given the relatively clear consensus about contraindications which had developed in 1991. However, the assumption that “more of a specific process of care is better” is not necessarily true of all that clinical audit seeks to bring about. Future evaluations should be restricted to audits which are clearly linked to evidence of effectiveness. These should become more frequent with current NHS policy initiatives—such as evidence-based medicine—and clinical effectiveness.

Given the difficulties encountered in this study, what is the future scope for economic evaluation of clinical audit? On the negative side, economic evaluation will not be suitable for routine priority setting within local audit programmes, because of the scale and type of data collection that would be required and the problem of selecting controls. However, when the audit examines an area of established clinical effectiveness, the outlook is better. This study shows the type of results which could be obtained and some of the methodological problems which need to be considered. These
constitute major challenges to the future use of this technique, but without such results, the appropriate level of audit activity will remain uncertain.

We thank all the medical and audit staff in the five hospitals who participated in the study, Charles Normand for economic advice, and the Department of Health for funding. This study was carried out in collaboration with the Research Unit of the Royal College of Physicians (London).

26 Withdrown.