Evaluation of audit of medical inpatient records in a district general hospital

John Gabbay, Amanda J Layton

Abstract

Objective — To evaluate an audit of medical inpatient records.


Setting — Central Middlesex Hospital.

Materials — Random sample of 188 notes per year drawn systematically from notes from four selected one month periods and audited by two audit nurses and most hospital physicians.

Main measures — General quality of routine clerking, assessment, clinical management, and discharge, according to a standardised, criterion based questionnaire developed in the hospital.

Results — 1988 was the year preceding the start of audit in the hospital, 1989 the year of active audit with implicit and loosely defined criteria, and 1990 the year after introduction of explicit criteria for note keeping. There was a significant trend over the three years in 21/56 items of the questionnaire, including recording of alcohol intake (χ² = 8.4, df = 1, p = 0.001), ethnic origin (χ² = 57, df = 1, p = 0.001), allergies and drug reactions (χ² = 10, df = 1, p = 0.001) at admission and of chest x-ray findings (χ² = 8, df = 1, p = 0.01), final diagnosis (χ² = 5.6, df = 1, p = 0.025), and signed entries (χ² = 11.3, df = 1, p = 0.001). Documentation of discharge and notification of discharge to general practitioners was not significantly improved.

Conclusions — Extended audit of note keeping failed to sustain an initial improvement in practice; this may be due to coincidental decline in feedback to doctors about their performance.

Introduction

Over the past two years medical audit has been promoted widely and is now an expected part of medical practice. Much of the boost came from the 1989 government white paper Working for Patients and from a series of royal college reports, including that from the Royal College of Physicians in 1989. The models adopted by physicians for conducting audit were drawn from the experience of a few pioneers, which centred mainly on audit of medical notes. There has as yet been no evaluation of the impact of a longer term of such audit on improving standards.

Since 1989 we have been working with physicians at Central Middlesex Hospital, a 580 bed acute general hospital in north London, to carry out a research and development programme in audit. The work had several closely related aims: firstly, to help introduce audit; secondly, to develop robust methods of auditing the process of hospital care; and, thirdly, to evaluate the impact of the audit.

We report an evaluation of the audit of medical inpatient records. Data were collected with the Central Middlesex Hospital physicians’ audit questionnaire, produced from a set of guidelines developed in the spring of 1990 after a year of regular feedback of information from a general audit of medical records on inpatient admissions. This standardised criterion based questionnaire assesses the general quality of the routine clerking, assessment, clinical management, and discharge of the patient. It was developed to reflect aspects of the inpatient record which were thought to be important by the physicians and audit steering group, which had been found to be practical as agreed criteria of good care, and which were also amenable to reliable audit using closed questions. This questionnaire was used to compare documented standards of inpatient medical admissions over three years. The first year, 1988, preceded the start of audit in the hospital. The second, 1989, was when audit was being actively carried out, but with only implicit and loosely defined criteria of good practice. The third, 1990, followed the introduction and distribution of the explicit criteria for note keeping, but audit meetings were less often concerned with standards of note keeping than in 1989 and more often with care of specific conditions.

Methods

We sampled notes of patients admitted during four selected one month periods (March, May, June, September, and November – December) in each of the three years. A sample size of 188/year was achieved by sampling about one in eight patient records drawn systematically from the patient administration system using a random starting point and stratified by medical consultant firms. Oversampling of up to 40% was needed to achieve the sample size because of difficulty in locating notes. The sample size, based on the results of a previous study, was designed to detect changes in major variables of at least 10% with a power of 80% at a 2% significance level.
level, which was chosen in preference to the conventional 5% level because many variables were being tested.

The audit was carried out by two audit nurses and by most physicians in the hospital. Every month the consultant physicians and senior registrars were sent a maximum of two photocopies of admission forms to audit. Other junior doctors were sent one every alternate month. Photocopies in which all reference to year of admission had been deleted were used. The notes were audited with the Central Middlesex Hospital physicians’ audit questionnaire. The physicians but not the audit nurses were also sent a short supplementary set of open questions which allowed them to express any concerns about the clinical management of the patient which might merit further discussion.

In comparing the results from unmatched pairs of data from any two years the χ² test was used or, when the data consisted of widely distributed continuous variables such as the elapsed time or the percentage of unsigned notes, the Mann-Whitney U test was used. To test for changes over the full three years the χ² for trend was used. Reliability of the questionnaire was tested by randomly selecting 20 case notes from those previously reviewed by physicians for independent re-audit by the audit nurses and comparing the results. When possible k analysis for comparing reliability between the auditors was used. This test is useful for comparing agreement between two observers classifying nominal categories, taking into account the agreement that might be expected by chance alone. We accepted questions as reliable when the k score achieved a significance level of p < 0.01. The 11 questions not amenable to k analyses were accepted as reliable if the auditors disagreed in three or less of the 20 sets of records.

Table 1 Items recorded in case notes of medical inpatients by doctor on admission. Figures are numbers (percentages of questions answered).

<table>
<thead>
<tr>
<th>Year</th>
<th>Presenting complaint</th>
<th>History of presenting complaint:</th>
<th>Social history:</th>
<th>Recorded by doctor:</th>
<th>Recorded in notes:</th>
<th>Treatment plan</th>
<th>Investigation plan</th>
<th>Review of systems:</th>
<th>Gastrointestinal system</th>
<th>Genitourinary system</th>
<th>Neurological system</th>
<th>Locomotor system</th>
<th>Physical examination (&gt; 5 remarks):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>182(97)</td>
<td>167(89)</td>
<td>63(37)</td>
<td>159(85)</td>
<td>93(52)</td>
<td>156(83)</td>
<td>166(88)</td>
<td>101(56)**</td>
<td>56(32)**</td>
<td>41(25)**</td>
<td>29(16)**</td>
<td>12(9)</td>
<td>107(58)***</td>
</tr>
<tr>
<td></td>
<td>183(97)</td>
<td>176(95)</td>
<td>75(43)</td>
<td>165(88)</td>
<td>109(60)</td>
<td>163(87)</td>
<td>168(89)</td>
<td>130(70)</td>
<td>80(50)</td>
<td>81(45)</td>
<td>80(50)</td>
<td>13(9)</td>
<td>131(70)</td>
</tr>
<tr>
<td></td>
<td>184(98)</td>
<td>171(92)</td>
<td>74(42)</td>
<td>161(87)</td>
<td>123(69)†</td>
<td>160(86)</td>
<td>179(93)</td>
<td>132(71)†</td>
<td>97(52)†</td>
<td>81(44)†</td>
<td>97(52)†</td>
<td>13(9)</td>
<td>121(65)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Reliability (Range of K when p &lt; 0.01)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>§</td>
</tr>
<tr>
<td>1989</td>
<td>†</td>
</tr>
<tr>
<td>1990</td>
<td>††</td>
</tr>
</tbody>
</table>

Notes: *p < 0.005, **p < 0.01, ***p < 0.001, χ² 1988 versus 1989.
†p < 0.025, ††p < 0.01, †††p < 0.001, χ² test for trend.
§ Auditors disagreed in ≥ 3 items out of 20.
Results

A total of 564 sets of notes were assessed: 141 and 139 records were audited by each of the two audit nurses; 79 by senior registrars; 36 by registrars; and 29 by more junior grades.Audit nurses took a median of 10 minutes to audit each set of notes and the physicians a median of 15 minutes.

The age range of the patients in the sample was 16-106 (median 68); 288 were male and 276 female. There were no significant differences between the age or sex distribution of the patients in the three years.

Tables 1 to 3 indicate the questions which proved reliable according to the criteria stated previously; the \( \kappa \) values ranged from 0.4 to 0.9. Questions that required the auditors to count data (for example, the number of entries in the notes or of remarks noted on the physical examination) proved much less reliable than direct yes/no questions. A question on patients' occupation was unreliable in the elderly patients because of lack of prior agreement as to whether the patients' former occupation should be noted.

For most questions there was a clear improvement between 1988 and 1989, but only rarely was the rate of improvement maintained between 1989 and 1990 (tables 1-3). Between 1988 and 1989 there were significant improvements in 27 of the 56 items, and between 1989 and 1990 there were only four significant improvements but six significant deteriorations (p < 0.025). The three year trend analysis, as indicated in the final column of the tables, was significant in only 21 out of 56 items in the questionnaire (p < 0.025).

Table 1 deals with the admission procedure. A significant trend of improvement was found in the recording of several variables such as alcohol intake (\( \chi^2 = 8.4, df = 1, p = 0.01 \)) and patient's ethnic origin (\( \chi^2 = 57, df = 1, p = 0.001 \)) although ethnic origin was still not recorded in over half of the admissions. Allergies and drug reactions were better noted (\( \chi^2 = 10, df = 1, p = 0.01 \)), as was much of the review of systems, although at best this was checked in only 70% of patients admitted. Results of urine analysis were noted in only a third of admissions throughout the period of audit.

Table 2 highlights note keeping during admission. Over the three years the notes were increasingly likely to contain written results of a chest radiograph (\( \chi^2 = 8, df = 1, p = 0.01 \)) but not of an electrocardiogram. The noting of the final diagnosis at the end of the inpatient notes improved (\( \chi^2 = 5.6, df = 1, p = 0.025 \)), but the record of information given to patients did not. The doctors' entries in the notes were more likely to be signed (\( \chi^2 = 11.3, df = 1, p = 0.001 \)). Table 3 shows that there was no significant trend in improvement in the documentation of discharge, nor of its notification to general practitioners over the three year period.

Auditing physicians expressed concern about clinical management in about 10% of admissions in each of the three years: in 12 out of 140 (8.6%) admissions for consultants, 12 out of 140 (8.6%) admissions for registrars, and in three out of 36 (8.3%) for registrars. Doubts included the necessity of the admission, the need for ITU care, poor or absent recordings of important information, uninvestigated clinical problems, questionable treatment regimes and concern about the information given to general practitioners.

Discussion

CHANGES IN STANDARDS OF NOTE KEEPING

Improvements were achieved in the agreed criteria of good note keeping during the first year of audit (1989), when note keeping was the central topic of audit meetings. Over the second year (1990), when explicit criteria and

<table>
<thead>
<tr>
<th>Year</th>
<th>1988</th>
<th>1989</th>
<th>1990</th>
<th>Reliability (K value to ( p &lt; 0.01 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chest radiograph requested</td>
<td>143(76)</td>
<td>137(74)</td>
<td>132(70)</td>
<td>0.6</td>
</tr>
<tr>
<td>Result noted</td>
<td>85(59)</td>
<td>82(59)</td>
<td>100(76)**</td>
<td>0.5-0.9</td>
</tr>
<tr>
<td>Electrocardiography requested</td>
<td>115(61)</td>
<td>112(61)</td>
<td>111(59)</td>
<td>0.9</td>
</tr>
<tr>
<td>Result noted</td>
<td>88(76)</td>
<td>84(75)</td>
<td>88(79)</td>
<td>0.7-0.8</td>
</tr>
<tr>
<td>No illegible words</td>
<td>144(77)</td>
<td>151(80)</td>
<td>154(82)</td>
<td>**</td>
</tr>
<tr>
<td>( &lt; 10% ) Entries unsigned</td>
<td>81(43)**</td>
<td>110(58)</td>
<td>115(61)***</td>
<td>**</td>
</tr>
<tr>
<td>( \geq 2 ) Days with no notes written</td>
<td>81(43)</td>
<td>85(45)</td>
<td>73(39)</td>
<td>**</td>
</tr>
<tr>
<td>Patient informed of diagnosis, prognosis, or treatment</td>
<td>32(17)</td>
<td>34(19)</td>
<td>35(19)</td>
<td>0.7</td>
</tr>
<tr>
<td>Final diagnosis written at end of notes</td>
<td>31(16)</td>
<td>45(24)</td>
<td>50(27)†</td>
<td>**</td>
</tr>
</tbody>
</table>

* p < 0.025, ** p < 0.01, *** p < 0.001, \( \chi^2 \) 1988 versus 1989.
† p < 0.025, †† p < 0.01, ††† p < 0.001, \( \chi^2 \) for trend.
‡ Auditors disagreed in \(< 3 \) items out of 20.

Table 3 Discharge documentation in patient's notes. Figures are numbers (percentages of questions answered)

<table>
<thead>
<tr>
<th>Year</th>
<th>1988</th>
<th>1989</th>
<th>1990</th>
<th>Reliability (K value to ( p &lt; 0.01 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP discharge proforma present</td>
<td>128(68)</td>
<td>128(68)</td>
<td>141(75)</td>
<td>0.7-0.8</td>
</tr>
<tr>
<td>No sent within 3 days of discharge</td>
<td>104(81)</td>
<td>102(79)</td>
<td>111(79)</td>
<td>✧</td>
</tr>
<tr>
<td>Discharge summary present</td>
<td>130(69)**</td>
<td>152(82)</td>
<td>139(74)</td>
<td>0.9</td>
</tr>
<tr>
<td>No sent within 1 month of discharge</td>
<td>82(63)</td>
<td>80(52)</td>
<td>83(60)</td>
<td>✧</td>
</tr>
</tbody>
</table>

** p < 0.01, \( \chi^2 \) 1988 versus 1989.
✧ Auditors disagreed in \(< 3 \) items out of 20.
guidelines had been distributed, some of these changes were sustained. Overall there was little further improvement and often a partial relapse, and yet, even after the significant improvements in 1989, there was still plenty of scope for further progress.

Other changes could also have influenced the standards of note keeping during the three years of the study. The NHS and Community Services Act was introduced, and Central Middlesex Hospital became an NHS trust with clinical directorates managerially committed to improvements in the quality of care during the three years of the study. Such managerial changes would have been expected to augment the effects of audit and so do not explain the lack of progress in the third year.

A Hawthorne effect, arising from the introduction of audit not only locally but nationally, could have been partly responsible for the initial changes. This effect might explain the early success and could have been unlikely to be sustained into the second year. Another possible explanation is the change in junior staff. This was partly controlled for by using sample months away from the times when junior doctors change jobs, and the consultant staff had not noticed any subjective differences in the abilities of the junior staff in the three years of the study.

Two possible sources of bias also need to be considered. Firstly, unavailable notes may have been those of patients who were less well managed; but the proportion of cases is unlikely to have varied from year to year. Secondly, the inability to eliminate all clues to the year of the notes may have biased the auditor; but this might be expected to have led to more recent notes being rated more highly. Neither of these possibilities explains the pattern of the results.

During 1989, when audit meetings were introduced, one meeting a month was concerned with the standards of admission records. At this time the junior staff attended audit meetings regularly and were initially provoked into active participation by emphasis on the poor standard of note keeping. Both junior and consultant staff, however, became disaffected and bored. By 1990, when the guidelines for note keeping had been agreed and issued, audit meetings focused on specific clinical topics with only occasional discussions about note keeping. There was little feedback to doctors about their performance in note keeping. Interviews with the junior staff disclosed that consultant staff rarely reinforced the standards set out in the guidelines during routine work.

The change in emphasis of audit meetings may have contributed to the failure during 1990–1 to sustain the rate of improvement which occurred in 1989–90. The method of feedback is an important factor in influencing clinical practice; the use of guidelines alone is unlikely to bring about change.

**Audit Questionnaire**
The audit questionnaire proved a successful tool for reviewing the standard of note keeping. The questionnaire's well piloted closed questions were useful and reliable. The more detailed questions in which procedures were itemised, for example, a series of questions on the review of systems asking whether a particular item had been recorded, were more reliable than attempts to summarise, for example, the physical examination by asking how many physical findings were noted for each system. To improve reliability further, questions that involve accurate counting (for example, the numbers of entries in the notes) will need to be replaced. The questionnaire could be further refined by taking out non-discriminating questions about practices that are almost universal (for example, writing up the presenting complaint).

That only the most tightly defined questions stood up to the test of interobserver reliability suggests that audit questionnaires based on the Checklist of the American College of Physicians, which in general has very loose questions, are unlikely to produce reliable results. Unreliable questionnaires are of value for opening up the process of audit or indicating possible areas of concern but cannot reliably be used to compare practice between firms or between provider units. This has important implications for approaches to monitoring quality in hospital contracts.

**Implications for Development of Audit**
It would be wrong to conclude from our evaluation alone that audit can sustain the standards of care. There is a distinction between the merit of audit as a general activity and the specific audit methods reported in this study. Moreover, we have discussed the audit of record keeping and not the audit of the quality of care. However, medical notes are a source of information for audit of the process of care and for this they need to be accurate and complete.

We showed that audit of medical note keeping can result in significant, although limited, results. Many hospitals are practising audit along similar lines, which may also be of similarly limited value. This suggests an urgent need to evaluate audit under many different circumstances and to explore the factors that lead to its success or failure, rather than let the growth of ineffective audits continue unchecked.

In our view further improvements may be seen with attention to certain interpersonal and managerial aspects of the conduct of the audit. A study of the effects of audit in general practice outlined four key features leading to sustained improvements in practice: relevant clinical knowledge, interpersonal skills in managing small groups, technical expertise in devising guidelines, and full involvement in standard setting by those whose practice is being audited. Only the first of these key features is easy to find in hospital settings. The hierarchical and transient nature of many of the relationships in hospital practice may militate against true involvement of junior staff in the audit process. Many of the junior staff...
thought that audit was being done to them and not by them. The hospital is now looking at different approaches to audit which encourage junior staff to take the initiative in reviewing the standards of practice.

The central practice that was being audited was clerking and record keeping. This is surely taught at medical school? Perhaps the audit feedback, if it is to achieve maximal effect, should also be directed at the medical schools.

We thank the physicians at Central Middlesex Hospital for their cooperation and help, Jane Wadsworth for statistical advice, and Jean Bodimeade for retrieving the notes.