Medical audit in general practice. I: Effects on doctors' clinical behaviour for common childhood conditions

North of England Study of Standards and Performance in General Practice

Abstract

***Objective*—To estimate the effects of medical audit, particularly setting clinical standards, on general practitioners' clinical behaviour.**

***Design*—Before and after study strengthened by a replicated Latin square.**

***Setting*—62 training general practices in the north of England.**

***Subjects*—92 general practitioner trainers, 84 (91%) of whom completed the study; random sample of 3500 children consulting one of these trainers for any of five conditions—acute cough, acute vomiting, bedwetting, itchy rash, and recurrent wheezy chest—stratified by doctor consulted, condition, and age.**

***Interventions*—Clinical standard set by each of 10 small groups of general practitioner trainers for one randomly selected childhood condition. Each group also experienced a different type of medical audit, randomly selected, for each of the four other study conditions (receiving a clinical standard set by another trainer group, tabulated data comparing clinical performance with that of all other groups, tabulated data from only their own group, and nothing ("control" condition)).**

***Main measures*—Content of initial consultation divided into: history, examination, investigation, diagnosis, and management (abstracted from medical records and "enhancement forms" completed by doctors).**

***Results*—There was increased prescribing of bronchodilators for acute cough, oral rehydration fluids for acute vomiting, antibiotics for itchy rash, and bronchodilators and oral steroids for recurrent wheezy chest and reduced prescribing of antibiotics for acute cough and recurrent wheezy chest and tricyclic antidepressants for bedwetting. Fewer children were "discharged." Each change was consistent with the standard and either limited to doctors who set a standard for that condition or significantly greater for them than all other doctors.**

***Conclusion*—Setting clinical standards improved prescribing and follow up.**

Participating departments and institutions: Division for General Practice of the Northern Regional Postgraduate Institute for Medicine and Dentistry, Department of Child Health, and Health Care Research Unit, University of Newcastle upon Tyne

Key people responsible for the paper:
Ian T Russell (Aberdeen); Julia M Addington-Hall (London); Peter J Avery, Claire H Bamford, Andrew G Barton (Newcastle upon Tyne); Cam Donaldson (Aberdeen); Martin P Eccles, Christopher J W Foy, Erica V Haimes (Newcastle upon Tyne); Jenny Hewison (Leeds); P P Mary Ho (Hong Kong); Robin D Humphrey, Allen Hutchinson (Newcastle upon Tyne); Donald H Irvine (Ashington); Elaine McColl, Louise Parker, I Nicholas Steen, Elizabeth M L Towner (Newcastle upon Tyne); John K G Webb (Lechlade).

Correspondence to:
Dr Ian Russell, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen AB9 2ZD.

BMJ 1992;304:1480-4

Introduction

The government has proposed that "every doctor should participate in regular systematic medical audit."[1] From 1984 to 1990 we undertook an extensive audit of the care of children in general practice. We also tested how effective the audit was in changing doctors' clinical behaviour and patients' health.

In the north of England vocational training for general practice began in 1969.[2] Many activities in small groups[3,4] convinced general practitioner trainers of the value of medical audit. Meanwhile, local paediatricians were considering ways of improving paediatric care in general practice after the Court report.[5] As a result of their keenness to explore the value of medical audit in the care of children trainers, paediatricians, and researchers combined to design and plan the north of England study of standards and performance in general practice.[6]

Expressed in 1992 terms, the study aimed at developing and evaluating methods of medical audit in general practice, notably the setting and dissemination of clinical standards. Reviews[7-9] of rigorous evaluations of standards[10-23] suggest that their effectiveness varies considerably.[9] However, few studies have evaluated medical audit in British general practice[24-27]; two that investigated standards drawn up by general practitioners were limited to prescribing and lacked a rigorous research design.[26,27] Thus our study was the first comprehensive evaluation of standard setting in British general practice.

Subjects and methods

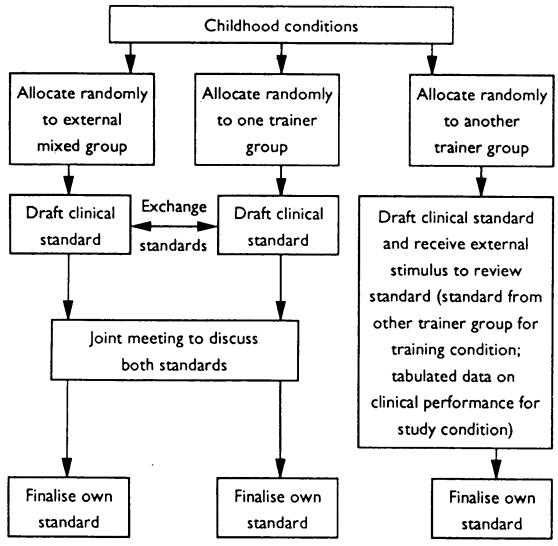
SETTING CLINICAL STANDARDS AND OTHER METHODS OF MEDICAL AUDIT

We invited all general practitioner trainers in the north of England who were neither singlehanded nor close to retirement to take part in the study.[28] Of 107 eligible trainers, 92 (86%) agreed to do so and 84 (79%) completed the study. Those who agreed formed 10 groups, each comprising about nine trainers practising within the same locality.

Each group set "clinical standards" for two of 10 childhood conditions, all chosen by the study organisers.[28] Groups were asked to define good primary care for children with the specified condition. They were encouraged to use algorithms (for example [29]) or any other format in which the recommended course of action depended critically on the information available. The first standards drawn up by these groups were for five "training" conditions: acute diarrhoea, acute earache, chronic handicap, fits and convulsions, and recurrent abdominal pain.[28] The second set of standards, on which our evaluation was based, were for five "study" conditions: acute cough, acute vomiting, bedwetting, itchy rash, and recurrent wheezy chest.

A standard for each condition was also drawn up by one of five "mixed groups" (each comprising two consultant paediatricians, two experienced general practitioners, and a researcher as resource)—that is, by doctors whose performance was *not* to be reviewed.[28] Before the standards were finalised each mixed group exchanged standards with, and then met, one of the two trainer groups working on the same condition; that trainer group was thereafter free to finalise its own standard (figure).

Each trainer group also experienced a different type of medical audit for three of the four other study conditions.[30] For one condition they received a standard set by another trainer group; for another condition tabulated data comparing their clinical per-



Process of setting standards for 10 childhood conditions

formance as a group with that of all other trainer groups; and for another tabulated data that merely described their performance as a group. The remaining condition served as a "control," for which they received neither a standard nor summary data. Although the study was designed to estimate how all four interventions affected performance, the main focus was on standard setting. Previous experience in the north of England had suggested that this was the activity most likely to change performance.[3]

EVALUATION OF STANDARD SETTING AND MEDICAL AUDIT

Our evaluation was based on the 62 practices of the 84 trainers who set standards. Initially, the trainers recruited 135 (75%) of their 181 partners to collect data[30]; this proportion remained constant through many partnership changes. In each practice, data were collected for one year, beginning in a random week between August 1984 and June 1985, and for another year, beginning in the same random week two years later (table I); thus the effects of medical audit were monitored for up to two years after the completion of standard setting. Data came from three sources: the records of children with the study conditions, surveys of these children's parents, and activity analysis within participating practices.

Children consulting for the two acute conditions (acute cough and vomiting) were identified prospectively by their general practitioners.[30] Those with the three chronic conditions (bedwetting, itchy rash, and recurrent wheezy chest) were identified retrospectively from a postal survey of parents of all 75 000 children registered with participating practices. The records of random samples of all these children (stratified[31] by doctor consulted, study condition, and age) were then flagged to enable doctors to recognise future consultations for that condition.

The content of each relevant consultation was recorded in two ways: doctors made notes in the child's medical record in the normal manner and also completed an "enhancement form."[30] This form asked them to record diagnosis or formulation; history, examination and investigation; management decisions; and reasons for these decisions. Samples of both types of record were abstracted by fieldworkers, who translated the information into numerical codes in a standardised manner.

TABLE I—*Basic design and timetable of north of England study*

Date	Standard setting groups	Research team
1982-4	(1) Train for standard setting (2) Prepare for data collection	Prepare and pilot research documents
1984-6	"Before" phase of data collection	
1985-6	Set standard for allocated study condition	Disseminate standards and baseline data on medical process
1986-8	"After" phase of data collection	
1988-90	Review definitive data on clinical performance (not part of formal evaluation)	Compare "before" and "after" phases to infer effect of setting standards and other types of medical audit

TABLE II—*Experimental design of north of England study: type of audit undertaken for each study condition by trainer groups A to K*

	Study condition				
Type of audit	Acute cough	Itchy rash	Acute vomiting	Recurrent wheezy chest	Bedwetting
Set clinical standard:					
Discussion with mixed group	G*	A	C	E	J
No discussion with mixed group	B	F	H	K	D
Receive clinical standard	J	G*	A	C	E
from another trainer group	F	H	K	D	B
Receive comparative data	E	J	G*	A	C
from all participating doctors	H	K	D	B	F
Receive descriptive data	C	E	J	G*	A
only from own trainer group	K	D	B	F	H
None	A	C	E	J	G*
	D	B	F	H	K

*From the top left hand corner, of the two trainer groups who set a clinical standard for acute cough, group G met the corresponding mixed group (figure) and group B worked entirely on its own; group G also received a standard for itchy rash, comparative data for acute vomiting, and descriptive data for wheezy chest but experienced no form of audit for bedwetting.

EXPERIMENTAL DESIGN AND SAMPLE SIZE

Table I shows that the basic design of the study was a "before and after" comparison.[30] This type of design has many weaknesses, including risk of bias from time trends in any of the data collected.[32] To ameliorate these weaknesses different groups of trainers had a different experience for each condition (table II). For each group there was a "control" condition for which trainers *received* neither the standard nor summary data (even though they had *provided* data for all five conditions); for each condition there were two "control" groups of trainers who received neither standard nor data. In this way we estimated what would have happened in the absence of medical audit.

The resulting design is known as a replicated Latin square.[33] Although complex (table II), it is easy to implement: each group is assigned a letter at random, stratified[31] by area to avoid adjacent groups working on the same condition. This randomisation guards against many sources of bias.

To calculate the sample size needed we judged that a general improvement of 10% in compliance with a standard would be clinically significant. From a pilot study[34] we estimated that we should therefore abstract the records of 10 children per condition per phase for each trainer, and 10 records from his or her partners taken together. However, interviews with trainers suggested that standard setting would be much more effective in stimulating change than the three other types of medical audit.[35] We therefore reduced our abstraction targets for each condition for which the trainer had *not* set a standard to five records per phase. This maintained the power of the study to detect changes arising from standard setting while reducing its power to detect other changes.

STATISTICAL ANALYSIS

The classic method of analysing the replicated Latin square displayed in table II is based on just 50 numbers that summarise differences in clinical performance between before and after phases, one for each of the 50 letters in table II.[33] Because the numbers of patients varied from cell to cell, however, we used the more robust technique of generalised linear modelling.[36] This had the additional advantage of allowing for differences between the children in each cell, notably in the doctor whom they consulted. We used the

TABLE III—*Effect of standard setting on adjusted percentages* of children recorded as being prescribed antibiotic at first consultation*

Study condition	Consultations after general practitioner had set standard for study condition	All other consultations (those before general practitioner had set standard for study condition and those with general practitioners who set standard for one of four other conditions)	Odds ratio (95% confidence interval) of antibiotic prescription after standard setting
Acute cough	31% (n=142)	41·8% (n=767)	0·62 (0·40 to 0·95)
Acute vomiting	24% (n=77)	21·5% (n=604)	1·18 (0·66 to 2·13)
Bedwetting	10% (n=36)	8·3% (n=305)	1·23 (0·38 to 3·93)
Itchy rash	17% (n=84)	5·6% (n=647)	3·38 (1·68 to 6·80)
Recurrent wheezy chest	24% (n=103)	34·8% (n=705)	0·61 (0·36 to 1·05)

*All adjusted for variation in recorded prescribing between doctors (the only significant factor other than study condition and standard setting).

GLIM statistical package[37] to analyse sequences of nested models before selecting that which best fitted the data.[38] For binary variables we adopted a binomial error structure with a logit link function; for most quantitative variables we adopted a normal error distribution with an identity link function.

This complex strategy was designed to answer a simple question—namely, whether standard setting had improved performance. The following tables summarise our findings as simply as possible without distortion. The final columns of tables III and IV record marginal odds ratios[33] with 95% confidence intervals derived from our GLIM analyses rather than from the percentages recorded in the tables.

Similar methods have shown that standard setting within this study had no effect on parents' reports of the prevalence of, and consultation rates for, the five study conditions.[39] This paper reports the effects on doctors' behaviour, and a subsequent paper reports those on patients' health.

Results

From nearly 12 000 children recorded by the 84 trainers and their partners as consulting with study conditions, the records of over 6000 were sampled randomly for statistical analysis, and the 3500 initial consultations with trainers by those children were analysed. (The number was smaller than our target because reported consultation rates, notably for acute vomiting and bedwetting, were lower than expected.)

We analysed the recorded content of these consultations under 16 headings. Six related to history: social history, family and genetic history, previous medical history, previous diagnoses, previous drug management, and previous other management; four to the diagnosis of the current episode: history, examination, investigations, and recorded diagnosis; and the remaining six to the management of that episode: advice and explanation, other doctor actions, drug management, referral decisions, follow up decisions, and reasons for management. All 16 types of recorded information showed variation that was significant at the 0·1% level; as expected there was substantial variation among conditions, among trainers, and between medical records and enhancement forms.

Only two types of information showed a significant change associated with the setting of clinical standards, those on drug management and follow up decisions. Only three specific changes could be attributed to standard setting: in antibiotic prescribing (table III), other therapeutic prescribing (table IV), and discharge decisions (table V). (Tables III to V summarise the best statistical models identified by our analysis and differ in format as the models differ.)

ANTIBIOTIC PRESCRIBING

After adjustment for variation between doctors table III shows that only two factors influenced the recorded prescribing of antibiotic drugs. Firstly, prescribing varied significantly ($p<0·001$) among conditions, from 8% for bedwetting to 42% for acute cough. Secondly, standard setting had three effects: trainers who set standards for acute cough thereafter reduced their prescribing by 11% ($p<0·05$); those who set standards for recurrent wheezy chest thereafter reduced such prescribing by 10%, which was nearly significant ($p>0·05$); and those who set standards for itchy rash thereafter increased prescribing by 11% ($p<0·01$).

The corresponding standards[28] enabled us to interpret these changes. All three standards for acute cough (two set by trainer groups and one by a mixed group; figure) and all three corresponding standards for recurrent wheezy chest cautioned against the indiscriminate prescribing of antibiotics. In contrast, all three standards for itchy rash advocated the use of antibiotics for infected eczema and impetigo. Finally, none of the four trainer group standards for acute vomiting or bedwetting gave unequivocal advice about use of antibiotics. Thus the changes in antibiotic prescribing were all consistent with the standards that had been set.[28] Furthermore, the general statistical test

TABLE IV—*Effect of standard setting on adjusted percentages* of children recorded as being prescribed a therapeutic drug (other than antibiotic, analgesic, or antipyretic) at first consultation*

Study condition	Before standard setting: All general practitioners	After standard setting		Odds ratio (95% confidence interval) of therapeutic prescription after standard setting
		General practitioners who set standard for that condition	Other general practitioners	
Acute cough	10·3% (n=471)	19% (n=142)	13·7% (n=296)	1·22 (0·70 to 2·12)
Acute vomiting	13·7% (n=397)	24% (n=77)	17·9% (n=207)	1·54 (0·84 to 2·82)
Bedwetting	29·0% (n=197)	28% (n=36)	36·0% (n=108)	0·27 (0·11 to 0·73)
Itchy rash	56·3% (n=421)	64% (n=84)	64·1% (n=226)	1·06 (0·62 to 1·80)
Recurrent wheezy chest	64·0% (n=428)	78% (n=103)	71·0% (n=277)	1·11 (0·64 to 1·90)

*All adjusted for variation in recorded prescribing between doctors (the only significant factor other than study condition, phase, and standard setting).

TABLE V—*Effect of standard setting on adjusted percentages* of children recorded as being discharged at first consultation*

Study condition	Routine medical records		Enhancement forms	
	Consultations after general practitioner had set standard for study condition	All other consultations (those before general practitioner had set standard for study condition and those with general practitioners who set standard for one of the four other conditions)	Consultations after general practitioner had set standard for study condition	All other consultations (those before general practitioner had set standard for study condition and those with general practitioners who set standard for one of the four other conditions)
Acute cough	2% (n=26)	3% (n=133)	18% (n=116)	26·5% (n=634)
Acute vomiting	4% (n=9)	6% (n=93)	31% (n=68)	42·0% (n=511)
Bedwetting	0% (n=24)	0% (n=100)	2% (n=12)	2·7% (n=205)
Itchy rash	1% (n=38)	1% (n=165)	10% (n=46)	14·9% (n=482)
Recurrent wheezy chest	1% (n=26)	1% (n=138)	9% (n=77)	13·1% (n=567)

*All adjusted for variation in recorded discharges between doctors (the only significant factor other than type of record, study condition, and standard setting).

of whether standard setting caused different changes in antibiotic prescribing for different conditions was significant at the 1% level. (Given the number of statistical tests implied by our analysis of 16 different types of recorded information, a significance level of 1% is more appropriate when testing for general effects than the traditional level of 5%. Nevertheless, the sample sizes for individual conditions were so small that a significance level of 5% and 95% confidence intervals are more appropriate for them.)

OTHER THERAPEUTIC PRESCRIBING

Of the 15 standards (10 set by trainer groups and five by mixed groups),[28] all six for respiratory conditions (acute cough and recurrent wheezy chest) advocated use of bronchodilators for children with wheeze or persistent cough. The three standards for acute vomiting advocated use of oral rehydration fluids. The three standards for bedwetting cautioned against prescribing of tricyclic antidepressants except as a last resort. Finally, the three standards for itchy rash advocated drugs such as benzylbenzoate for scabies but cautioned against indiscriminate prescribing of steroids for mild eczema.

After adjustment for variation among doctors table IV shows that only three factors influenced the recorded prescribing of therapeutic drugs other than antibiotics, analgesics, and antipyretics. Firstly, prescribing before standard setting varied between 10% for acute cough and 64% for recurrent wheezy chest ($p<0{\cdot}001$). Secondly, over the two years between the phases of data collection prescribing increased consistently for all five study conditions ($p<0{\cdot}001$). Thirdly, trainers consistently changed their prescribing for their allocated condition in the direction suggested by the standard they had set: for acute cough, acute vomiting, and recurrent wheezy chest this change reinforced an increasing trend and for bedwetting it counteracted that trend.

The confidence intervals within table IV suggest that only the change in prescribing for bedwetting was significant ($p<0{\cdot}01$). However, the general test of whether standard setting causes *different* changes in the recording of therapeutic prescribing for different conditions was also significant, although only at the 5% level. Fortunately, our preliminary analysis of the standards enabled us to use a test of fewer parameters whether standard setting causes *consistent* proportional changes in recording of therapeutic prescribing across different conditions, upwards for acute cough, acute vomiting, and recurrent wheezy chest and downwards for bedwetting; this was significant at the required 1% level.

The recording of prescriptions did not differ between medical records and enhancement forms. Furthermore, all the changes were entirely consistent between these two sources of data. Thus the enhancement forms introduced within this study and the existing records (often very abbreviated) each validated the other and led us to conclude that the changes presented in tables III and IV are real.

DISCHARGE DECISIONS

After adjustment for variation between doctors table V shows that recorded decisions to discharge children (that is, to advise them not to consult again unless their condition persisted or deteriorated) were influenced by three factors: condition, type of record, and whether the trainer had set a standard for that condition. Only 2% of entries in medical records mentioned discharge while 22% of enhancement forms did so ($p<0{\cdot}01$). The effect of standard setting was significant at the 1% level and consistent across all combinations of type of record and study condition. This effect may thus be summarised by a single odds ratio[33]: the odds that trainers who had set a standard would record that children with their condition had been discharged were 63% (95% confidence interval 45% to 87%) of the odds under all other circumstances. This finding is consistent with the emphasis on follow up throughout all 15 standards.

OTHER TYPES OF MEDICAL AUDIT

Although setting a standard had three distinct effects, there was no evidence that the other types of medical audit—receiving a standard, receiving comparative data, and receiving descriptive data—had any effect. Furthermore, the effects of standard setting were no greater for the five trainer groups who received specialist input (in the form of a meeting with a mixed group, table II) than for those who did not.

Discussion

EFFECTS OF MEDICAL AUDIT

In this study 84 general practitioner trainers in 10 groups set clinical standards for 10 childhood conditions, five of which were used to evaluate trainers' performance. After setting their standards, trainers changed their recorded practice for their condition in three ways and maintained these changes for up to two years. Firstly, the recorded prescription of antibiotics fell for acute cough and recurrent wheezy chest and rose for itchy rash. Secondly, the recorded prescription of bronchodilators rose for acute cough and recurrent wheezy chest, that of oral rehydration fluids rose for acute vomiting, and that of tricyclic antidepressants fell for bedwetting. Thirdly, trainers reduced the proportion of children with their study condition whom they recorded as having been discharged. All these changes were consistent with the corresponding standards.

There was no evidence that the five trainer groups whose standard setting culminated in a meeting with a mixed group, including paediatricians, subsequently performed better than the other five groups. There was also no evidence that receiving standards set by other trainer groups or receiving tabulated data on other groups' care of study conditions influenced trainers' practice. After the end of the study an attractive package of data in the form of personalised graphs was well received by trainers: three quarters of them used this feedback or made plans to do so.[35] However, well designed feedback has yet to be rigorously evaluated.[8]

How valid are these conclusions about the effectiveness of medical audit? Data were collected by two methods: participating doctors kept records in the normal manner and they also completed structured enhancement forms for a random sample of children. Reassurance about the validity of both methods comes from three sources. Firstly, the content of medical records is generally well correlated with that of the corresponding consultation.[40,41] Secondly, all three changes in recorded practice relate to objective decisions for which records should be accurate. Thirdly, records and enhancement forms yielded consistent estimates of all three positive changes. The most stringent test is whether the recorded change in practice leads to a change in patients' health; this is reported in the accompanying paper.

IMPLICATIONS FOR MEDICAL AUDIT

Working for Patients recognised that "access to a system of adequate medical records" is essential for successful medical audit.[1] Although medical records were useful in our study, they had to be augmented by structured enhancement forms; for example, decisions to discharge patients were mentioned in 22% of enhancement forms but only 2% of medical records. But the task of enhancement was very demanding, and

enthusiasm for it diminished over the long period of data collection.[35] In the long term the best solution may be to encourage doctors, through incentives or training, to adopt a structured and more comprehensive framework for their records, based on a problem oriented approach.[42,43]

Although trainers who set their own standard subsequently improved prescribing and follow up for that condition, they were not influenced by standards set by another trainer group for another condition. However, such "external" standards can be effective if they are disseminated and implemented so that doctors have opportunities to learn from, and even to assimilate, these standards.[9] Standards set by groups including representatives of the doctors who are to be audited can be even more effective.[9] Thus our findings are consistent with those of other evaluations: standards are effective in improving practice when they are set by, or otherwise made acceptable to, those doctors whose performance is to be reviewed.

1 Secretaries of State for Health, Wales, Northern Ireland, and Scotland. *Working for patients*. London: HMSO, 1989. (Cmnd 555.)
2 Walker JH, Smith A, Irvine DH. Vocational training for general practice: a preliminary report. *BMJ* 1971;i:41-3.
3 Irvine DH. Quality in general practice. In: McLachlan G, ed. *A question of quality?* London: Nuffield Provincial Hospitals Trust, 1975:63-96.
4 Ashton JR, Oliver G, Grant A, Taylor GK. An audit of deaths in general practice. *Update* 1976;**12**:1019-22.
5 Committee on Child Health Services. *Fit for the future*. London: HMSO, 1976. (Cmnd 6684.)
6 Irvine DH, Russell IT, Hutchinson A, Barton AG, Foy CJW, Haimes EV, *et al*. Performance review in general practice: educational development and evaluative research in the Northern region. In: Pendleton DA, Schofield TPC, Marinker ML, eds. *In pursuit of quality?* London: Royal College of General Practitioners, 1986: 146-67.
7 Haynes RB, Davis DA, McKibbon A, Tugwell P. A critical appraisal of the efficacy of continuing medical education. *JAMA* 1984;**251**:61-4.
8 Mugford M, Banfield P, O'Hanlon M. Effects of feedback of information on clinical practice: a review. *BMJ* 1991;**303**:398-402.
9 Russell IT, Grimshaw JM. *The effectiveness of referral guidelines: a review of the methods and findings of published evaluations*. Aberdeen: Health Services Research Unit, 1991. (Occasional paper 8.)
10 Morgan M, Studney DR, Barnett GO, Winickoff RN. Computerised concurrent review of prenatal care. *Quality Review Bulletin* 1978;**4**:33-6.
11 Sanazaro PJ, Worth RM. Concurrent quality assurance in hospital care. *New Engl J Med* 1978;**296**:1171-7.
12 Hopkins JA, Shoemaker WC, Greenfield S, Chang PC, McAuliffe T, Sproat RW. Treatment of surgical emergencies with and without an algorithm. *Arch Surg* 1980;**115**:745-50.
13 Linn BS. Continuing medical education: impact on emergency room burn care. *JAMA* 1980;**244**:565-70.
14 Lohr KN, Brook RH. Quality of care in episodes of respiratory illness among Medicaid patients in New Mexico. *Ann Intern Med* 1980;**92**:99-106.
15 Cohen DI, Littenberg B, Wetzel C, Neuhauser DB. Improving physician compliance with preventive medicine guidelines. *Med Care* 1982;**20**:1040-5.
16 Thompson RS, Kirz HL, Gold RA. Changes in physician behaviour and cost savings associated with organisational recommendations on the use of routine chest x-rays and multichannel blood tests. *Prev Med* 1983;**12**: 385-96.
17 Cohen SJ, Weinberger M, Hui SL, Tierney WM, McDonald CJ. The impact of reading on physicians nonadherence to recommended standards of medical care. *Soc Sci Med* 1985;**21**:909-14.
18 Norton PG, Dempsey LJ. Self-audit: its effect on quality of care. *J Fam Pract* 1985;**21**:289-91.
19 Palmer RH, Louis TA, Hsu L-N, Peterson HF, Rottrock JK, Strain R, *et al*. A randomised controlled trial of quality assurance in sixteen ambulatory care practices. *Med Care* 1985;**23**:751-70.
20 Putnam RW, Curry L. Impact of patient care appraisal on physician behaviour in the office setting. *Can Med Assoc J* 1985;**132**:1025-9.
21 Fowkes FGR, Davies ER, Evans KT, Green G, Hartley G, Hugh AE, *et al*. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986;i:367-70.
22 Tierney WM, Hui SL, McDonald CJ. Delayed feedback of physician performance versus immediate reminders to perform preventive care. *Med Care* 1986;**24**:659-66.
23 De Vos Meiring P, Wells IP. The effect of radiology guidelines for general practitioners. *Radiology* 1990;**42**:327-9.
24 Fleming DM, Lawrence MSTA. Impact of audit on preventive measures. *BMJ* 1983;**287**:1852-4.
25 Harris CM, Jarman B, Woodman E, White P, Fry JS. *Prescribing—a suitable case for treatment*. London: Royal College of General Practitioners, 1984. (Occasional paper 24.)
26 Grant GB, Gregory DA, van Zwanenberg TD. Development of a limited formulary for general practice. *Lancet* 1985;i:1030-2.
27 Anderson CM, Chambers S, Clamp M, Dunn IA, McGhee MF, Sumner KR, *et al*. Can audit improve patient care? Effects of studying use of digoxin in general practice. *BMJ* 1988;**297**:113-4.
28 North of England Study of Standards and Performance in General Practice. *Final report: I—setting clinical standards within small groups*. Newcastle upon Tyne: Health Care Research Unit, 1990. (Report 40.)
29 *Endocrine system: clinical algorithms*. London: *BMJ*, 1989.
30 North of England Study of Standards and Performance in General Practice. *Final report: II—methods for evaluating the setting and implementation of clinical standards*. Newcastle upon Tyne: Health Care Research Unit, 1990. (Report 41.)
31 Moser CA, Kalton G. *Survey methods in social investigation*. 2nd ed. Aldershot: Gower, 1971.
32 Russell IT. The evaluation of computerised tomography: a review of research methods. In: Culyer AJ, Horisberger B, eds. *Economic and medical evaluation of health care technologies*. Berlin: Springer-Verlag, 1983:299-316.
33 Armitage P, Berry G. *Statistical methods in medical research*. 2nd ed. Oxford: Blackwell, 1987.
34 North of England Study of Standards and Performance in General Practice. *Report on pilot study in South Cumbria and North Lancashire*. Newcastle upon Tyne: Health Care Research Unit, 1986. (Report 31.)
35 North of England Study of Standards and Performance in General Practice. *Final report: III—the effects of setting and implementing clinical standards*. Newcastle upon Tyne: Health Care Research Unit, 1990. (Report 42.)
36 McCullagh P, Nelder JA. *Generalised linear models*. London: Chapman and Hall, 1989.
37 GLIM Working Party. *The GLIM system release 3.77*. 2nd ed. Oxford: Numerical Algorithms Group, 1987.
38 Aitkin M, Anderson D, Francis B, Hinde J. *Statistical modelling in GLIM*. Oxford: Oxford University Press, 1989.
39 Ho PPM, Foy CJW, Avery PJ, Russell IT, Steen IN. Does quality assurance affect consultation rates in British family medicine? *Proceedings of the social statistics section of the American Statistical Association*. Washington, DC: American Statistical Association, 1990:270-5.
40 Lyons TF, Payne BC. The relationship of physicians medical recording to their medical care performance. *Med Care* 1974;**12**:463-9.
41 Hulka BS, Romm FJ, Parkerson GR, Russell IT, Clapp NE, Johnson FS. Peer review in ambulatory care: use of explicit criteria and implicit judgements. *Med Care* 1979;**17** (suppl).
42 Weed LL. *Medical records, medical education and patient care*. Cleveland, Ohio: Case Western University Press, 1969.
43 Petrie JC, McIntyre N. *The problem oriented medical record*. Edinburgh: Churchill Livingstone, 1979.

(Accepted 2 April 1992)

II: Effects on health of patients with common childhood conditions

North of England Study of Standards and Performance in General Practice

Abstract

***Objective*—To estimate the effects of medical audit, particularly setting clinical standards, on patients' health.**

***Design*—Before and after study strengthened by a replicated Latin square.**

***Setting*—62 training general practices in the north of England.**

***Patients*—Random sample of 9000 children with any of five conditions—acute cough, acute vomiting, bedwetting, itchy rash, and recurrent wheezy chest—stratified by doctor consulted, condition, and age.**

***Interventions*—Clinical standard set by each of 10 small groups comprising 84 general practitioner trainers for one randomly selected childhood condition. Each group also experienced a different type of medical audit, randomly selected, for each of the four other study conditions (receiving a clinical standard set by another trainer group, tabulated data comparing clinical performance with that of all other groups, tabulated data from only their own group, and nothing ("control" condition)).**

***Main outcome measures*—Condition specific, functional, psychological, and educational outcomes; together with parent satisfaction (recorded by home interviews and postal questionnaires).**

***Results*—Children consulting trainers for recurrent wheezy chest after those doctors had set a standard for that condition improved both in drug compliance (79% (n=33) before standard setting *v* 93% (30) after) and mean number of days of breathlessness (3·8 (SE 1·0) before *v* 1·7 (0·6) after) and**

Correspondence to:
Dr Ian Russell, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen AB9 2ZD.

BMJ 1992;**304**:1484-8