

To what extent do clinical notes by general practitioners reflect actual medical performance? A study using simulated patients

JAN-JOOST RETHANS

ERIC MARTIN

JOB METSEMAKERS

Keywords: record content; patterns of work; consultation process; performance review.

SUMMARY

Background. Review of clinical notes is used extensively as an indirect method of assessing doctors' performance. However, to be acceptable it must be valid.

Aim. This study set out to examine the extent to which clinical notes in medical records of general practice consultations reflected doctors' actual performance during consultations.

Method. Thirty nine general practitioners in the Netherlands were consulted by four simulated patients who were indistinguishable from real patients and who reported on the consultations. The complaints presented by the simulated patients were tension headache, acute diarrhoea and pain in the shoulder, and one presented for a check up for non-insulin dependent diabetes. Later, the doctors forwarded their medical records of these patients to the researchers. Content of consultations was measured against accepted standards for general practice and then compared with content of clinical notes. An index, or content score, was calculated as the measure of agreement between actions which had actually been recorded and actions which could have been recorded in the clinical notes. A high content score reflected a consultation which had been recorded well in the medical record. The correlation between number of actions across the four complaints recorded in the clinical notes and number of actions taken during the consultations was also calculated.

Results. The mean content score (interquartile range) for the four types of complaint was 0.32 (0.27–0.37), indicating that of all actions undertaken, only 32% had been recorded. However, mean content scores for the categories 'medication and therapy' and 'laboratory examination' were much higher than for the categories 'history' and 'guidance and advice' (0.68 and 0.64, respectively versus 0.29 and 0.22, respectively). The correlation between number of actions across the four complaints recorded in the clinical notes and number of actions taken during the consultations was 0.54 ($P < 0.05$).

Conclusion. The use of clinical notes to audit doctors' performance in Dutch general practice is invalid. However, the use of clinical notes to rank doctors according to those who perform many or a few actions in a consultation may be justified.

Introduction

GENERAL practitioners spend about one hour a day recording information about patients seen during surgery consultations;¹ medical records are thus an important source of information about general practitioners' performance. National working bodies have agreed standards for record keeping^{2,3} that include the patient's identity (name, address, telephone number, marital status, occupation), the patient's medical problem list, relevant medical history and medication taken but recorded somewhere in the notes, as well as details about the last consultation(s) with the general practitioner. As well as providing information for good medical care, medical records are a useful resource for research, education and management.⁴ This raises the question whether medical records can be used to assess medical performance.

Assessment instruments of doctors' performance include direct methods where research workers observe a doctor managing patients, and indirect methods where direct observation of doctor-patient contact is not possible.⁵ Direct methods include video- or audio-tapes of consultations and simulated patients.

Review of clinical notes made by doctors is an indirect method extensively used in quality assurance programmes and by licensing boards.^{6,7} The MRCGP examination is exploring ways of judging clinical performance including the use of a workbook recording the detailed events in a consultation.⁸ For review of clinical notes to be acceptable, it should possess high validity and reliability. Research reveals that general practitioners record only part of the available information on patients.⁹⁻¹² While sex and age are recorded for nearly all patients,⁹ the level of recording is lower for history summary (36%), marital status (30%), occupation (30%), and repeat medications (23%).¹¹ Only just over half of patients with a chronic disease ever have it recorded in their general practitioner's record.¹² These studies, based on notes in records, are unable to show to what extent clinical notes in medical records reflect content of consultations or how they reflect the doctors' true performance.⁵ For example, absence of a record of 'repeat medication' may reflect the doctor's omission or indicate that the patient does not take medicine. A recording of 'bronchitis' does not indicate why the doctor wrote bronchitis or what information was given to the patient.

It is important to evaluate the relationship between clinical notes and performance in order to decide whether review of clinical notes is a valid research method for investigating doctors' performance. In the Netherlands no formal pressures exist on general practitioners to record information, apart from the fact that patients have a legal right to see their notes. Meyboom audiotaped and observed general practitioners' consultations and compared their content with patients' records against detailed standards of care.¹³ He found that 69% of doctors' actions were recorded. However, the doctors knew that they were being observed and that their records were to be examined, leading to possible bias in record keeping performance.

A study was therefore undertaken using simulated patients to

J-J Rethans, PhD, assistant professor, Centre for Quality Assurance in Research in General Practice, University of Limburg, Maastricht, Netherlands. E Martin, medical student and J Metsemakers, MD, assistant professor, Department of General Practice, University of Limburg, Maastricht, Netherlands.

Submitted: 14 May 1993; accepted: 20 August 1993.

© British Journal of General Practice, 1994, 44, 153-156.

determine the extent to which clinical notes of consultations reflected the actual performance of doctors.

Method

During 1989, 39 general practitioners were each visited by four different simulated patients, yielding 156 consultations. Selection of doctors, training of simulated patients in role play and reporting of doctor's performance has been described elsewhere.^{14,15} The complaints presented by the simulated patients were tension headache, acute diarrhoea and pain in the shoulder, and one presented for a check up for non-insulin dependent diabetes. Of the 39 general practitioners 35 were contacted two years later (one had moved and three promised only to be involved in the pilot phase) to request the medical records of those patients. The records were retyped to facilitate reading and scoring, and the doctors' names replaced with random numbers to ensure blind analysis. Where items could not be read a photocopy was sent to the general practitioner for clarification. The frequency of recording of certain patient characteristics was examined.

Immediately after the consultations, the simulated patients reported on the consultations using extensive checklists, based upon accepted standards of care. These standards described essential actions (considered to be necessary for good quality care) and intermediate actions (not essential for good care but not harmful either) and an example is given in Appendix 1.¹⁶ The four standards each included between 25 and 36 actions. Identical standards were used to score the content of the medical records.

Comparison of the doctors' performance, assessed by the simulated patients with the clinical notes for each item of the standards, could be made in four ways: performed and recorded; performed but not recorded; not performed and not recorded; and not performed but recorded. The last two, necessary for scoring, were not used in further calculations in this study.

EM scored all the medical records and a 10% sample of the records was scored independently by J-JR. Their coefficient of agreement was 0.96 (kappa 0.93). Subsequently, an index was calculated as a measure of agreement between actions which had actually been recorded and actions which could have been recorded in the clinical notes, that is, 'performed and recorded' divided by 'performed and recorded' plus 'performed but not recorded'. This index was called the content score and a high content score reflected a consultation which had been recorded well in the medical record.

Frequency calculations were made of content scores for each complaint and across the four complaints for all actions together and for individual categories: history, physical examination, laboratory examination, guidance and advice, medication and therapy and follow up. The same was done for the essential and intermediate actions.

Correlations (Pearson product moment correlations) were calculated between number of actions across the four complaints recorded in the clinical notes and number of actions taken during the consultations, in order to establish the relationship between clinical notes in medical records and actual performance.

Results

From the 35 practitioners contacted 101 records of the possible 140 consultations were received (72%). Eighteen doctors sent records of all four patients, five of three patients, seven records of two patients and in five cases no records were found. Five doctors used a computer to record patient information. Age was recorded in 99% of records, sex in 94%, a medical summary in 51%, occupation in 47%, and marital status in 16%.

The content scores of records for the different categories of

each complaint, as well as across the four complaints, are shown in Table 1. Of all actions undertaken in the four types of consultation only 32% were recorded, varying from 25% for the shoulder pain case to 36% for the diarrhoea case. Not only did this figure vary between complaints, but also between the different categories. For example, 68% of actions taken concerning medication and therapy and 64% of actions taken concerning laboratory examination had been recorded in the notes compared with only 22% of actions concerning follow up.

The content scores of the medical records for the essential and intermediate action categories of each complaint as well as across the four complaints are also shown in Table 1. Intermediate actions rather than essential actions were recorded most accurately. For example, of actions undertaken in the four types of consultation, only 30% of essential actions had been recorded compared with 46% of intermediate actions.

The correlations between what had been recorded in the medical records and what had been done in actual practice across the four complaints by the 18 doctors who sent in all four records were calculated. The correlation for the history category was 0.17, physical examination 0.45, laboratory examination 0.75 ($P < 0.01$), guidance and advice 0.50 ($P < 0.05$), medication and therapy 0.43 and follow up -0.04 . The correlation for the categories overall was 0.54 ($P < 0.05$). There was therefore a moderate correlation between performing and recording for all actions together, but particularly for the laboratory examination category.

Discussion

This is the first comparison of the content of general practitioners' medical records of a consultation with the content of the same consultations in an unbiased situation. Since only four cases per doctor were used and all patients were effectively new patients, one should be careful when generalizing beyond the specific sample of cases in this study.¹⁷ The finding that 28% of all records were missing two years after the visits and that only 18 of the 35 general practitioners were able to forward all four records shows that doctors are not good at saving records. Some doctors said they delayed making a complete record until the patient consulted for a second time. However, this does not explain why the notes were lost, since Dutch law requires that doctors keep records for a least five years.

The finding that the mean content score for all the categories of actions ranged from 0.25 to 0.36 shows that little can be concluded from records about what doctors actually do during consultations. It seems that only for the categories medication and therapy, laboratory examination and, to a much lesser extent, physical examination can one be more certain that absence of recording means this action did not take place during the consultation. It may be that the category of laboratory examination reflects the recording abilities of medical secretaries rather than the general practitioners themselves, since in the Netherlands most laboratory results are filled out by medical secretaries. It is remarkable that the history category, considered to be the most important part of the consultation, showed a mean content score lower than the scores for all actions together. It may be that when one performs many different actions in a consultation there is less time to record all the information. This may explain why the mean content scores for the essential actions, originally meant as the most important part of the consultation, were lower than the intermediate actions scores.

While there were high mean content scores in the medication and therapy category for the patients with tension headache, acute diarrhoea and shoulder pain, this was not so for the patient with diabetes, showing that content specificity (that is, what sort of problem) is important.¹⁸ Other results using simulated patients

Table 1. Agreement between actions which had actually been recorded and actions which could have been recorded by general practitioners in the clinical notes of simulated patients, by category and by priority.

| | Mean content score (interquartile range) for the care of simulated patients | | | | |
|-------------------------------------|---|---------------------------------------|-------------------------------------|---|--|
| | With tension headache (<i>n</i> = 28) | With acute diarrhoea (<i>n</i> = 27) | With shoulder pain (<i>n</i> = 24) | Having diabetes check up (<i>n</i> = 22) | Across all four cases (<i>n</i> = 18) |
| <i>Category</i> | | | | | |
| History | 0.27 (0 to 0.42) | 0.32 (0.14 to 0.50) | 0.23 (0.04 to 0.33) | 0.08 (0 to 0) | 0.29 (0.18 to 0.40) |
| Physical examination | 0.42 (0 to 1.0) | 0.50 (0 to 1.0) | 0.09 (0 to 0.20) | 0.92 (0.92 to 1.0) | 0.31 (0.20 to 0.42) |
| Laboratory examination ^a | 0.33 (0 to 1.0) | 1.0 (1.0 to 1.0) | – | 0.75 (0.33 to 1.0) | 0.64 (0.14 to 1.0) |
| Guidance and advice | 0.18 (0 to 0.33) | 0.23 (0 to 0.33) | 0.30 (0.21 to 0.33) | 0.04 (0 to 0) | 0.22 (0.15 to 0.29) |
| Medication and therapy | 0.72 (0.25 to 1.0) | 0.88 (0 to 1.0) | 0.90 (1.0 to 1.0) | 0.07 (0 to 0) | 0.68 (0.53 to 0.83) |
| Follow up | 0.05 (0 to 0) | 0 | 0.22 (0 to 1.0) | 0.15 (0 to 0) | 0.22 (0 to 0.47) |
| Total | 0.30 (0.15 to 0.46) | 0.36 (0.24 to 0.50) | 0.25 (0.19 to 0.28) | 0.26 (0.13 to 0.30) | 0.32 (0.27 to 0.37) |
| <i>Priority</i> | | | | | |
| Essential action | 0.28 (0.11 to 0.38) | 0.36 (0.22 to 0.50) | 0.22 (0.14 to 0.29) | 0.25 (0.09 to 0.41) | 0.30 (0.23 to 0.37) |
| Intermediate action | 0.38 (0.15 to 0.67) | 0.51 (0.33 to 0.75) | 0.67 (0.50 to 1.0) | 0.90 (0.75 to 1.0) | 0.46 (0.37 to 0.55) |

n = number of general practitioners. ^aLaboratory examination not used in shoulder pain category.

in general practice suggest that doctors perform differently when patients present with chronic diseases compared with acute problems.¹⁵ This may therefore affect their recording.

Although records may lack many details of a consultation, the relationship between what doctors actually did and what they recorded was moderate, as indicated by the overall correlation of 0.54. However, the high correlation for laboratory examination seems to have been mainly responsible for the overall significance level. Again, it is remarkable that the history category showed such a low correlation. Doctors seemed not to record history details in a systematic matter. Thus, at the content level clinical notes inadequately reflected the actual consultation and thus were insufficiently valid. However, if doctors had to be ranked according to those performing many or a few actions in the consultation, their clinical notes could serve as a basis for this ranking.

The finding of Meyboom (who compared doctors' actions against similar standards) that 69% of the doctors' actions were recorded is at variance with the findings presented here.¹³ There is no reason to believe that Meyboom's scoring technique and the techniques used here differed. However, his study differed from the present study in the number of consultations sampled and in the fact that doctors were aware that they were being audio-taped and observed, leading to possible bias in behaviour. Bearing in mind other surprising results from studies using simulated patients, the findings from the present study may show how important it is to measure general practitioners' real performance with methods that do not change the behaviour of doctors.^{19,20}

This study shows that review of clinical notes to assess whether doctors have performed certain actions in their consultations is an invalid assessment method.

Appendix 1. Accepted standard of care for the simulated patient presenting with pain in the shoulder.

History

1. Essential. How long has the complaint existed?
2. Essential. Did he pain develop gradually or acutely?
3. Essential. Where is the pain?
4. Essential. Can a history of the cause be found?
5. Essential. Is there any movement limitation: when, which movements?
6. Essential. How do complaints relate to life and work?
7. Essential. Are there any accompanying symptoms such as paraesthesiae?
8. Intermediate. Is patient self-medicating?
9. Intermediate. Are there any general illness symptoms?

Physical examination

10. Essential. Patient to undress above waist; inspect.
11. Essential. Ask for pain locations to be indicated.
12. Essential. Palpate pain points.
13. Essential. Palpate joint during movement.
14. Essential. Check passive movement: extension, flexion, rotation.
15. Essential. Check active movement: painful arc, flexed biceps, lower arm extensors.

Guidance and advice

16. Essential. Report findings and/or diagnosis.
17. Essential. Discuss prognosis.
18. Essential. Give movement and exercise advice.
19. Intermediate. Give ergonomic advice.

Medication and therapy

20. Intermediate. Apply sling, immobilization bandage, splint.
21. Intermediate. Prescribe non-steroidal anti-inflammatory drugs; local corticosteroid injections.
22. Intermediate. Refer to physiotherapist.

Follow up

23. Essential. Indicate whether return visit necessary, in relation to prognosis.
24. Essential. Return visit after 2-3 weeks.
25. Essential. Return visit if movement sufficiently limited to make working impossible.

References

1. van Zutphen WM. *De taken van de huisarts* [General practitioners' job description, dissertation]. Maastricht, Netherlands: Rijksuniversiteit Limburg, 1984 (English summary).
2. Royal College of General Practitioners. *Evidence to the Royal Commission on the National Health Service. Policy statement 1*. London: RCGP, 1985.
3. Meyboom WA, Metsemakers JFM, Hofstra ML, Beusmans GHMI. Standaard medische verslaglegging [a standard for record keeping]. *Huisarts Wet* 1990; **33**: 114-117 (English summary).
4. Metsemakers JFM, Höppener P, Knottnerus JA, et al. Computerized health information in the Netherlands: a registration network of family practices. *Br J Gen Pract* 1992; **42**: 102-106.
5. Rethans JJ, van Leeuwen Y, Drop R, et al. Competence and performance: two different concepts in the assessment of quality of medical care. *Fam Pract* 1990; **7**: 168-174.
6. Shaw CD. Criterion based audit. *BMJ* 1990; **300**: 649-651.
7. Norman GR, Davis DA, Painvin A, et al. Comprehensive assessment of clinical competence of family/general physicians using multiple measures. In: Hart R, Rarden R, Des Marchais J (eds). *Current developments in assessing clinical competence*. Montreal, Canada: Heal Publications, 1992.
8. Irvine D. General practice in the 1990s: a personal view on future developments. *Br J Gen Pract* 1993; **43**: 121-125.
9. Dawes KS. Survey of general practice records. *BMJ* 1972; **286**: 219-223.

10. Lönberg NC, Bentsen BG. Record keeping in Norwegian general practice. *Scand J Prim Health Care* 1984; **2**: 151-157.
11. Mansfield BG. How bad are medical records? A review of the notes received by a practice. *J R Coll Gen Pract* 1986; **36**: 405-406.
12. Mant D, Tulloch A. Completeness of chronic disease registration in general practice. *BMJ* 1987; **294**: 223-224.
13. Meyboom WA. *Verlaglegging van huisartsgeneeskundig handelen* [Recording clinical notes, dissertation]. Meppel, Netherlands: Krips Repro, 1990 (English summary).
14. Rethans J-J, Sturmans F, Drop R, *et al.* Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. *BMJ* 1991; **303**: 1377-1380.
15. Rethans J-J, Sturmans F, Drop R, van der Vleuten C. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. *Br J Gen Pract* 1991; **41**: 97-99.
16. Grol R, Mesker P, Schellevis F (eds). *Peer review in general practice*. Nijmegen, Netherlands: University Department of General Practice, 1988.
17. van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teaching and Learning in Medicine* 1990; **2**: 925-934.
18. Elstein AS, Shulman LS, Sprafka SA. *Medical problem solving: an analysis of clinical reasoning*. Cambridge, MA: Harvard University Press, 1978.
19. Norman GR, Neufeld VR, Walsh A, *et al.* Measuring physicians' performances by using simulated patients. *J Med Educ* 1985; **60**: 925-934.
20. Rethans JJE, van Boven CPA. Simulated patients in general practice: a different look at the consultation. *BMJ* 1987; **294**: 809-812.

Acknowledgements

We thank the participating doctors, Trudie Seegers and Robin Hull of the General Practice Writers Association for their help.

Address for correspondence

Dr J-J Rethans, Department of General Practice, University of Limburg, Postbus 616, 6200 MD Maastricht, Netherlands.

RCGP
ROYAL COLLEGE OF GENERAL PRACTITIONERS

14 Princes Gate, Hyde Park, London SW7 1PU

RCGP SALES Tel: 071 823 9698 Fax: 071 225 3047
**OCCASIONAL PAPER 63 PORTFOLIO-BASED
LEARNING IN GENERAL PRACTICE** £9.00

This document comes from a College Working Group which was set up by the Education Division of the College to explore how higher professional education might be provided, and recognized, in general practice. The paper not only describes the essential features of portfolio-based learning, such as modularisation and the use of a mentor, but offers concrete examples of their application both in vocational training and continuing medical education.

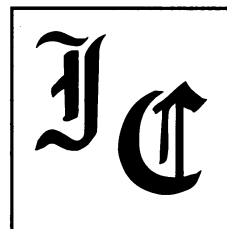
JOURNAL BINDERS £6.50
Blue Cortex Binder holds 12 Journals, easy to use.

RCGP COURSES Tel: 071 823 9703 Fax: 071 228 3047
PERFORMANCE APPRAISAL

Course Tutor: Hilary Haman
17/18 May 1994

Two day course open to general practitioners and practice and health centre managers. It aims to provide an appreciation of performance appraisal, the opportunity to practise interviewing skills, including role play and the use of video, and to develop a plan to allow participants to introduce a performance appraisal scheme into their practice. This course is highly intensive and therefore limited to twelve participants only.

The fee is £325.00



INSTITUTE OF COUNSELLING

Skills-based Courses

COUNSELLING

- *Certificate Courses*
- *1 Year Diploma Courses*
- *2 Year Diploma Courses*
- *Training Resources*

(Our bookshop stocks one of the largest ranges of books on therapy in Scotland)

All courses on a home study basis with tutor and video support.

COUNSELLING IN

A New Series of 6 Training Videos

- 1 **Counselling in Perspective (31 min)**
Defining Counselling Contrasting 4 Client-centred models.
- 2 **Counselling in Action (48 min)**
Example of bad & good practice with supervision.
- 3 **Counselling in Depth (50 min)**
3 counselling sessions demonstrating empathy, positive regard & genuineness. They tackle issues such as blindness, alcoholism, serious illness and depression.
- 4 **Counselling in Focus (50 min)**
One to one counselling and group work using active listening and advanced empathy and looking also at cultural counselling, unemployment.
- 5 **Counselling in Analysis (48 min)**
One to one counselling and group discussion on 2 important counselling models which assist clients to re-evaluate and play action.
- 6 **Counselling in Role Play (Psychodrama) (76 min)**
Part 1. A formal introduction to Psychodrama.
Part 2. 40 mins from an actual Psychodrama.

Series produced for the Institute of Counselling
☆ Series cost £150 per set ☆ Also available to hire ☆
Accredited by The CACC; recognised by ACC

For a Prospectus Contact:

Dept GP, Institute of Counselling,
6 Dixon St, Glasgow, G1 4AX. Tel: 041-204 2230