Improving searching skills and evidence retrieval

ABSTRACT - **Objective:** To evaluate the effect of a three-hour training session in formulating questions and searching databases.

- **Design:** A randomised controlled trial and before and after study, with blinded outcome assessment.
- **Setting:** Oxford University Medical School, first clinical year.
- **Subjects:** Altogether 108 medical students were randomly assigned to an experimental group (54) or a control group (54), and all were given the task of searching for evidence around an ulcer related problem or a cardiac problem. Students in the experimental group were randomly allocated to research one of the two problems before training and the remaining problem afterwards. Control students received no training and were randomly allocated to search for evidence around either of these problems.

- **Main outcome measures:** Searching performance; the quality of evidence retrieved; student satisfaction.
- **Results:** Training improved the students' search performance and the quality of evidence retrieved. Students' satisfaction with the training was high.
- **Conclusions:** A three-hour interactive training session improved the students' ability to search databases and retrieve evidence and was well received by the students.

Changes in undergraduate medical education advocated by the General Medical Council's Education Committee include a reduction in didactic teaching, the promotion of self-directed learning and the development of information technology skills. At the same time there have been increasing demands for junior doctors' training to integrate education with service delivery. As yet, little support has been provided to enable trainees to learn, or for career grade doctors to function as educators within the context of continued service provision. The traditional didactic teaching programmes of continued medical education are impractical as they take both learners and teachers away from providing essential clinical service, and, furthermore, they have been shown to be ineffective. By contrast, strategies employed in the practice of evidence based medicine are effective in improving the performance of doctors.

Evidence based medicine is an approach to learning that can be integrated into busy clinical practice. It satisfies many of the requirements of *Tomorrow's doctors* by promoting self-directed learning, learning through enquiry and the use of information technology. The practice of evidence based medicine involves five linked steps as follows:

- Question formulation
- Searching for evidence
- Critical appraisal
- Implementing evidence
- Evaluation of the performance and impact of this process.

The evidence based medicine approach can be used to convert the increasing volume of research evidence from information overload to a valuable resource. Advances in computer aided information management have changed radically the ease with which the medical literature can be searched and relevant articles retrieved. Learning how to access these resources rapidly and effectively has been shown to benefit patients, and should form an integral part of undergraduate and postgraduate medical training.

We have devised a short programme for training medical students to formulate questions and undertake a computer search of Medline via WinSpirs, Silver Platter's Windows Medline searching software. In a randomised controlled trial, we tested the effectiveness of the teaching package by studying changes in searching behaviour and the quality of evidence retrieved by clinical medical students who had been given a three-hour interactive training session compared with those who had not received this training.

**Methods**

- **Design**

Students were randomly allocated to receive training or no training. Within each group, students were randomly allocated to work on different scenarios (see below). Before the study, most students in both groups had performed searches on a weekly basis without having received formal training. During the study, students in all groups were free to practise searching at will, and so the intervention and control groups differed only in their exposure to training. Evaluation of searching performance was conducted before and after training for the intervention group, and once for the control group - contemporaneously with the intervention group's second evaluation. Three separate analyses were performed as follows:
A before and after analysis of the change in performance as a result of training in the intervention group.

- Comparison of the performance of students randomised to the intervention group with those randomised to the control group (a randomised controlled trial of the training).

- Comparison of the performance of the intervention group before training with the control group to test whether the effect of training spread from the intervention group to student colleagues without formal training.

**Subjects and randomisation**

All 108 first year clinical students at Oxford University Medical School were invited to participate. Randomisation was performed using a standard random numbers chart and blocked to ensure that equal numbers of students were allocated to receive the training intervention (experimental group) or no special training during the study period (control group). One subject in each group had received training in question formulation and searching before the study. Some of the other students had experience of searching, but none had been trained formally in question formulation or searching.

**Procedure**

Two clinical situations encompassing problems encountered in the first clinical year were devised as follows:

- **Ulcer scenario:** 'You have a 70 year old patient with a history of gastric ulcer for whom you are considering long term acid suppression treatment. You wonder if any randomised controlled trials have been conducted on this.'

- **Cardiac scenario:** 'A 65 year old patient is about to undergo cardiac surgery and you are concerned about postoperative bleeding. You wonder if there are any randomised controlled trials on the control of postoperative bleeding relevant to your elderly patient.'

Students in the experimental group were randomly assigned to one or other scenario and required to perform a Medline database search to find relevant evidence as the first step in their training intervention. They then attended a three-hour workshop on question formulation and searching skills given by two librarians (AL and RS). Within three months of completing the training programme, students in the study group were asked to perform searches on the scenario they had not researched. The control group – who had also been allocated randomly to either the ulcer or cardiac scenario – were asked to undertake a single search within the same three month period. The effectiveness of the training intervention could be estimated by comparing the outcomes before and after training in the intervention group or by comparing the outcomes after training in the intervention group with those of the control group. Furthermore, any diffusion of training from students in the intervention group to their colleagues in the control group could be estimated by comparing the values for the intervention group before training with those of the control group.

**Training session**

The three-hour training session aimed to teach question formulation and computer bibliographic database searching, including basic and advanced operations. No handouts were provided. The training programme was delivered to 54 students in small groups (four to seven students) by two experienced librarians acting as trainers and facilitators.

**Primary objectives:** the primary objectives of the session were as follows:

- The ability to convert a clinical scenario into a structured question\(^\text{11}\)
- Deriving search terms before and during searching
- Basic computer skills
- Use of three main database search features: (a) free-text or natural language searching; (b) controlled vocabulary index searching (Thesaurus on Medline); and (c) limiting the search by language, by species and by publication type to include trials and reviews.

A secondary objective was to improve the quality of evidence retrieved by searchers.

**Session activities**

The activities involved in the training were structured using the Kolb learning cycle\(^\text{12}\). Students first gained concrete experience of hands-on searching before any training. This was followed by an interactive training session that encouraged them to reflect on their experience and learn question formulation and searching skills. In the final step, students completed their learning through active experimentation. The students all had online access to the Silver Platter's Windows version of Medline (WinSpirs) on personal computer.

**Detailed session activities**

- Learners were taught to convert a clinical scenario into a four part structured clinical question\(^\text{13}\) using a worksheet that required them to identify a patient or problem, an intervention or exposure, a comparison and clinical outcome(s).

- They worked to produce as many search terms as possible for each part of the question using natural language synonyms and related terms. Boolean operators ('and', 'or', 'not') were explained and used to link terms into a testable free-text search strategy, to be searched and reviewed several times, with Thesaurus searching and Limits added in as the search progressed.

- The bulk of the session was spent in hands-on search-
ing, with skilled facilitators to guide learners. This was broken up by short bursts of theory or demonstration, with the main emphasis on learners developing their own strategies and styles to acquire transferable cognitive search skills rather than system specific computer tricks.

- Participants were asked to complete questionnaires rating the usefulness of the session and inviting comments.

**Evaluation**

Students were asked to return computer printouts from their searches. The WinSpirs software package records the search terms entered and the references yielded by the chosen search strategy. All searches and findings were evaluated by the investigators (WR, RS and AL), who were blind to which group the student belonged or whether the search was conducted before or after training.

- **Search score** – the search method was evaluated by investigators using a structured score sheet (Fig 1) and each search was assigned a search score.
- **Yield score** – the titles and abstracts of the articles retrieved by each student were evaluated by doctors. A score was assigned to each search according to a predetermined hierarchy from 4 (best) to 0 (worst) for the single highest quality article retrieved (Table 1).
- **Student satisfaction** – student satisfaction was assessed by a simple questionnaire asking students to rate the session from 1 (not very useful) to 6 (extremely useful).

**Statistical methods**

The search and yield scores were summarised using medians and interquartile ranges. Data on those who completed the assessment before training, but failed to complete the training or assessment afterwards, were included in the analysis where possible. The effect of training on the total search and yield scores was estimated as a difference in the median scores before and after training. Similarly, the effect of any diffusion of training was estimated by comparing scores before training on the experimental group with scores of the control group. Analyses were performed separately for each clinical problem. Statistical significance was assessed by the Mann-Whitney U test (with adjustments for ties), and confidence intervals for the difference in medians were calculated.
using the CIA software\textsuperscript{14}. Comparisons of changes in the individual components of the search score were made within each student; the Wilcoxon matched pairs rank sum test (with adjustment for ties) was used to assess statistical significance.

**Results**

The flow of students through the trial is shown in Fig 2. Eighteen percent of students failed to attend any of their allocated sessions, while 84% of those who attended the assessment session before training went on to complete both the training and the assessment afterwards.

*Search scores*

The distribution of students’ total search scores and evidence yield scores is shown in Table 2. The students’ median search scores before training and those of students in the control group were similar for both clinical situations.

**Table 2. Distribution of total search and yield scores (median (interquartile range)).**

| Group            | Search score (0–18) | Yield score (0–4) |
|------------------|---------------------|------------------|
| **Ulcer problem** |                     |                  |
| Intervention group: |                     |                  |
| Before training* (n=20) | 4 (3–6)             | 1 (0–2.5)        |
| After training (n=22)  | 9.5 (9–11)          | 4 (3–4)          |
| Control group (n=22)   | 4 (3–7)             | 0 (0–2)          |
| **Cardiac problem**   |                     |                  |
| Intervention group: |                     |                  |
| Before training* (n=25) | 5 (3–6)             | 4 (4–4)          |
| After training (n=16)  | 8 (7–9.5)           | 4 (4–4)          |
| Control group (n=22)   | 4 (4–6)             | 4 (3–4)          |

* Baseline scores of students who were allocated to training but did not attend all sessions are included in these figures.
Table 3. Analysis of the effect of training and diffusion on total search and yield scores (median difference (95% confidence interval)).

| Group                      | Search score | Yield score |
|----------------------------|--------------|-------------|
| Ulcer problem              |              |             |
| Training effect            | 5 (4–6)      | 2 (1–3)     |
| (before and after estimate*)| \(p<0.0001\) | \(p=0.0005\) |
| Training effect            | 5 (4–6)      | 3 (1–4)     |
| (comparison with control)  | \(p<0.0001\) | \(p<0.0001\) |
| Diffusion effect           | 0 (-1–2)     | 0 (-1–0)    |
|                           | \(p=0.7\)    | \(p=0.5\)   |
| Cardiac problem            |              |             |
| Training effect            | 3 (2–5)      | 0 (0–0)     |
| (before and after estimate*)| \(p=0.0003\) | \(p=0.6\)   |
| Training effect            | 3 (2–4)      | 0 (0–0)     |
| (comparison with control)  | \(p<0.0001\) | \(p=0.5\)   |
| Diffusion effect           | 0 (-1–1)     | 0 (0–0)     |
|                           | \(p=0.9\)    | \(p=0.5\)   |

* Baseline scores of students who were allocated to training but did not attend all sessions are included. Statistical significance calculated using the Mann-Whitney U test with adjustment for ties. Sample sizes are shown in Fig 2.

Training resulted in a statistically significant 5 point improvement in the search score for the ulcer scenario and a significant 3 point improvement for the cardiac scenario (Table 3). Comparison of the intervention group scores after training and the control group scores showed the same effect. There was no evidence of a diffusion effect of training information passing from experimental to control students for either scenario.

The effect of training on the performance of the individual components of searching was analysed for each situation separately, and for both combined (Table 4). Statistically significant improvements were seen for components C (terms that increase sensitivity) and D (terms that increase specificity) for both scenarios. In addition, there was some evidence of an increase in category B components (sensitive free text searching) when the ulcer scenario was the second scenario searched. Averaging the two allocation orders suggests a significant 0.5 point increase in search components B, a 3 point increase in C and a 1 point increase in D.

**Yield scores**

The yield scores for students performing searches before and after training and for students in the control group are presented in Table 2, and the effects of training on students’ yield scores are presented in Table 3. Yield scores for the ulcer scenario before or without training were low (median score of 1 out of a possible 4), while those for the cardiac scenario were high (median score of 4 out of a possible 4). After training, a significant increase of 2 points was seen in students’ yield scores for the ulcer problem (Table 2), but not the cardiac one (median increase=0). The same effect was observed when the intervention group scores after training were compared with the control group scores.

**Student satisfaction**

Overall, a 67% response rate was achieved, with 96% of responders rating the session as extremely or very useful. Repeated written comments emphasised the value of hands-on practice, enforced thinking about search strategy, clinical relevance and the ready availability of skilled help. Many students expressed the wish that the course had been delivered during their preclinical training.

Table 4. Within student changes for individual components of the search score (median (interquartile range)).

| Individual search score components* | A (0–2) | B (0–4) | C (0–6) | D (0–6) |
|------------------------------------|---------|---------|---------|---------|
| Ulcer then cardiac problem (n=16) |         |         |         |         |
| Scores on ulcer problem (before training†) | 1 (1–1.5) | 0.5 (0–1) | 1 (1–2.5) | 2 (0–2) |
| Scores on cardiac problem (after training) | 0.5 (0–1) | 1 (0–1.5) | 5 (4–5)  | 2 (2–2) |
| Median change (95% CI)             | \(p=0.005\) | \(p=0.3\) | \(p=0.0007\) | \(p=0.08\) |
| Cardiac then ulcer problem (n=22) |         |         |         |         |
| Scores on cardiac problem (before training†) | 1 (1–1)  | 0 (0–1)  | 2 (1–3)  | 1 (0–2) |
| Scores on ulcer problem (after training) | 1 (1–1)  | 2 (1–2)  | 5 (4–6)  | 2 (2–3) |
| Median change (95% CI)             | \(p=0.6\) | \(p=0.003\) | \(p=0.0001\) | \(p=0.0002\) |
| Both groups combined (n=38)        |         |         |         |         |
| Median change (95% CI)             | \(p=0.11\) | \(p=0.001\) | \(p=0.0001\) | \(p=0.0001\) |

*See Fig 1 for details of search score components A-D.
†Baseline scores of students who did not attend all sessions are excluded from these figures.
Statistical significance derived using a Wilcoxon Matched Pairs Rank Sum test, with adjustment for ties.
Discussion

Evidence on the effectiveness of teaching evidence based medicine is growing. Previous studies have investigated the effects of teaching critical appraisal\(^{15-19}\). Other studies of the teaching of searching skills to North American medical graduate trainees have used access to Medline via Gratefulmed as database and searching software. Their initial findings suggested that brief training in searching was ineffective\(^{20}\) but further studies suggested that, with experience, trainees became as competent as trained librarians\(^{8,11}\). These studies tended to score favourably the acquisition of searching techniques specified beforehand such as selection of an appropriate database, the use of strategies that broaden and restrict searches and the use of Boolean operators and limiters, as evidence of effective training. In addition, some studies evaluated student satisfaction.

We have developed a scoring system that evaluates all the key strategies tested in earlier studies. We set out to develop a training programme that would instruct learners in the first two steps of evidence based medicine, question formulation and searching for evidence, using effective educational strategies\(^5\). We chose to train students in the use of Medline as it is the database most frequently accessed by our students. WinSpirs was selected as a searching engine, as we have found that it provides an ideal framework for teaching the principles and practice of structured question formulation\(^3\) and database searching.

The programme was easily delivered to 54 students in 10 weeks, and we have gone on to train all 108 students over a six month period. Responses to the evaluation forms show that the teaching sessions were popular with the students. Both students and instructors felt that it was important to concentrate on question formulation and search strategy design before using the computer. Practical experience and the readiness of experts to guide and troubleshoot were highly valued. The success of this training, delivered as an interactive group activity, is in line with the findings of Davis et al\(^6\).

The trial has established that the teaching session is an effective way of improving students' searching skills with statistically significant gains in search scores after training. Significant improvements were seen also in the quality of evidence retrieved after training, as evidenced by gains in search yield scores, but only for the ulcer problem. The high quality of evidence retrieved for the cardiac scenario by a single term free text search obscured the effect of training. The ulcer problem was more exacting. It demanded a good searching technique, as the identification of good quality evidence depended on using more sophisticated searching methods. In a previous study, Wildemuth and Moore found little evidence of a relation between search behaviour and effectiveness, and they asserted that no generalisations could be made\(^8\). Our study throws light on this problem, as the two different clinical situations we used produced very different relations between search strategy and the quality of evidence retrieved. Thus, only an exacting scenario will reveal the relations between search behaviour and effectiveness.

In this study we used one learning package for all participants. Wood et al found that the choice of computer-assisted learning packages correlates with cognitive and learning styles\(^22\). Future learning programmes, developed for preclinical students and postgraduate trainees, might include the design of alternative learning packages.

We were interested to note that there was no difference in the search or yield scores achieved by the experimental group before training and the control group tested some months after the initial instruction sessions. This suggests that there was little diffusion of the training among students who were not directly exposed to it, emphasising the need to offer these training interventions to all students as part of their curriculum.

We wish to emphasise the benefits that come from clinicians and librarians collaborating in this training. From our own experience of working with preclinical students, postgraduate trainees and established clinicians, we believe that this intervention can be delivered equally effectively to doctors at all stages of training.

Acknowledgements

The authors thank Sian Jones for help in conducting the trial and Judy Palmer for encouragement.

References

1. Doctors' Recommendations on undergraduate medical education. London: General Medical Council, 1993.
2. Delamothe T. Junior's new deal on hours. Br Med J 1991;302:1482.
3. Bulstrode CJ, Muir Gray AJ, Anderson M, Hawke CJ. New deal for junior doctors' hours: how to achieve it. Br Med J 1992;305:1203–5.
4. Department of Health. Hospital doctors training for the future: the report of the Working Group on Specialist Medical Training. London: Department of Health, 1993.
5. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA 1995;274:700–5.
6. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. Br Med J 1995;310:1122–6.
7. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Br Med J 1996;312:71–2.
8. Sackett DL, Rosenberg WM. The need for evidence-based medicine. J R Soc Med 1995;88:620–4.
9. Haynes RB, McKibbon KA, Walker CJ, Ryan N, et al. Online access to MEDLINE in clinical settings. A study of use and usefulness. Ann Intern Med 1990;112:78–84.
10. Klein MS, Ross FV, Adams DL, Gilbert CM. Effect of online literature searching on length of stay and patient care costs. Acad Med 1994;69:489–95.
11. Haynes RB, Johnston ME, McKibbon KA, Walker CJ, Willan AR. A program to enhance clinical use of Medline. A randomized controlled trial. Online J Curr Clin Trials 1993;Doc no 56[4005].
12. Kolb DA. Experiential learning as the source of learning and development. Englewood Cliffs, NJ: Prentice Hall, 1984.
13. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995;123:A12–3.
Organophosphate sheep dip

CLINICAL MANAGEMENT OF LONG-TERM LOW-DOSE EXPOSURE

A joint report of the Royal College of Physicians and the Royal College of Psychiatrists

Organophosphate (OP) pesticides are used widely for treating crops and farm animals, eg in dipping sheep, and are associated with a variety of symptoms in those who use them either intermittently or continuously. Acute exposure to OPs is known to cause neurological syndromes. However the mechanism whereby low-dose OPs might cause reported symptoms such as cognitive impairment, psychiatric morbidity and minor sensory changes is presently unknown. Such uncertainty has led to inadequate management and to the perception by many sufferers that they are not taken seriously. This report, commissioned by the Department of Health, includes evidence given by patients, patient groups and doctors and analyses the clinical symptoms they describe and which they have attributed to working with OP sheep dip. An account of the toxicology of OP pesticides and toxic syndromes is also included.

Extensive recommendations are made for the management and diagnosis of symptoms, improved patient facilities and for further research. A bibliography covers most of the literature on what is known about OP pesticide-related symptoms and similar symptoms not necessarily related to exposure to organophosphates. It is hoped that this report will bring some clarity and objectivity to this contentious area of medicine with consequent help to those who suffer.

Summary  Abbreviations and glossary

PART 1 SUMMARY OF EVIDENCE:  Background and terms of reference
  - Toxicology of organophosphate pesticides and toxic syndromes  - Evidence received concerning symptoms
  - Evidence received concerning diagnosis and management

PART 2 RECOMMENDATIONS:  Diagnosis and guidelines  Management  Facilities  Research
  - References  Appendix: Sources of information relating to organophosphate sheep dip exposure

Price:  UK £12.50  Overseas £14.00  ISBN 1 86016 080 0  A4 report, soft cover  54 pages

AVAILABLE FROM THE ROYAL COLLEGE OF PHYSICIANS