

Evidence-based guidelines: should guide us in asking for the evidence and information we need

Johannes G Schmidt

Critical appraisal of medical evidence and sound medical decision making are especially important for general practice.^{1,2} We need a good understanding of how important clinical findings and medical diagnoses are for the patients we face in practice. Is the (ICD-based) diagnosis we attribute to the patient of real importance to him or rather a reflection of the low sensitivity of our nosological instruments with which we try and often fail to categorise the patient's real problems?³ How well do we understand the importance of the pre-test likelihood of the disease tested before we apply diagnostic instruments? In how far do we consider the importance of the 'pre-treatment risk' on the degree of the effect of therapy in terms of the 'number-needed-to-treat' (number of patients that need to be treated to prevent one adverse event) before we administer treatment? Trained in conventional (hospital-based) medical thinking, it is not easy for GPs to grasp the crucial importance of such questions and concepts.

A paper or a lecture on the subject will not suffice to provide the insight we need in order to understand why conventional medical thinking may lead to inappropriate decisions and has become outdated. Our Foundation's training courses in critical appraisal of medical evidence, which are now sponsored and financially supported by the Swiss Doctors' Union FMH, last three days. These three days, containing intermittent individual work, are needed to at least make doctors understand a few fallacies of current medical reasoning. Often, doctors strongly reject statistics and reasoning based on trial evidence before they begin to understand that evidence-based medicine is more about proper 'thinking' than about 'statistics' and more about asking what evidence we need than what evidence we have.

Evidence-based medicine and guidelines for general practice

Clinical epidemiology is often wrongly perceived as a

Johannes G Schmidt, president foundation paracelsus today.
Family Practice and Institute for Clinical Epidemiology,
 Ilgenweidstrasse 3, CH-8840 Einsiedeln.
 E-mail: schmidt@paracelsus-heute.ch

Submitted: 27th March 1998; accepted in revised form: 13th January 1999.

'statistical' discipline instead of a discipline of logical thinking which cannot replace nor exclude clinical judgement.⁴ Judgement needs to be based on a sound logic and thoughtful critical appraisal of the validity of one's own perceptions and learned therapeutic beliefs, while knowledge on the latest publications is of secondary importance. There are good reasons for GPs to mistrust 'statistics' because current clinical research is so preoccupied with the specificity of isolated therapeutic factors that much of the current evidence lacks sensitivity or external validity for the questions that really matter in practice. Clinical research has credulously focused on 'hard' endpoints without analysing their limited predictivity for the suffering that patients experience themselves, and studies have often neglected important clinical distinctions which we find in our patients.⁵ Evidence-based medicine, especially in general practice, may degenerate to 'literature-based medicine' of little practical value if this important aspect is neglected. The great potential of evidence-based medicine is its methodological instruments to distinguish facts from fiction or hypotheses from validated evidence. Clinical epidemiology offers the instruments to define what really matters for the patient and how relevant it is, and how to make informed decisions about which of the many competing options to choose, and this may include complementary medicine.

A widely known example of guidelines that doctors were and still are advised to follow are those on cholesterol screening. The well-established evidence from randomised controlled trials that cholesterol-lowering treatment reduces coronary heart disease, however, does not mean that doctors should screen cholesterol levels and apply treatment if cholesterol is 'high'. This is not proper evidence-based medicine. The methodological concepts of evidence-based medicine require the positive predictive value and the number-needed-to-treat concepts being applied. It can easily be seen from the Framingham data, grouped into various risk groups (figure 1), that treating a 'very high' cholesterol in a low-risk patient is less effective than treating a 'normal' cholesterol in a high-risk patient.^{6,7} This is also illustrated in the Sheffield tables (table 1) which are based on the proper distinction between high risk and low risk and not on a fallacious cholesterol cut-off level.⁸ Low-risk individuals, healthy people, are no problem, their chol-

Table 1. Sheffield table for primary prevention of CHD.

| Men: cholesterol concentration (mmol/L). | | | | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|----|
| | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | No | No | Yes | No | No |
| Hypertension | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | No | No | Yes | No | No |
| Smoking | Yes | Yes | No | No | Yes | Yes | Yes | No | Yes | No | No | No | No |
| Diabetes | Yes | No | Yes | No | Yes | Yes | No | Yes | No | Yes | No | No | No |
| LVH | Yes | Yes | Yes | Yes | No | No | No | No | No | No | No | No | No |
| Age (yr) | | | | | | | | | | | | | |
| 70 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 6.0 | 6.5 | 7.7 | |
| 68 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.6 | 6.4 | 6.9 | 8.2 | |
| 66 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.7 | 5.9 | 6.8 | 7.3 | 8.7 | |
| 64 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 6.1 | 6.3 | 7.3 | 7.8 | 9.3 | |
| 62 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 5.6 | 6.5 | 6.7 | 7.8 | 8.3 | | |
| 58 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 6.1 | 6.5 | 7.4 | 7.7 | 8.9 | | | |
| 56 | 5.5 | 5.5 | 5.5 | 5.5 | 5.5 | 6.5 | 7.0 | 8.0 | 8.3 | | | | |
| 54 | 5.5 | 5.5 | 5.5 | 5.5 | 5.9 | 7.0 | 7.5 | 8.6 | 9.0 | | | | |
| 52 | 5.5 | 5.5 | 5.5 | 5.5 | 6.3 | 7.6 | 8.1 | 9.3 | | | | | |
| 50 | 5.5 | 5.5 | 5.5 | 5.7 | 6.9 | 8.2 | 8.8 | | | | | | |
| 48 | 5.5 | 5.5 | 5.5 | 6.2 | 7.5 | 8.9 | | | | | | | |
| 46 | 5.5 | 5.5 | 5.5 | 6.8 | 8.2 | | | | | | | | |
| 44 | 5.5 | 5.5 | 5.8 | 7.4 | 9.0 | | | | | | | | |
| 38 | 5.5 | 6.8 | 7.9 | | | | | | | | | | |
| 36 | 6.0 | 7.6 | 8.8 | | | | | | | | | | |
| 34 | 6.7 | 8.6 | | | | | | | | | | | |
| 32 | 7.6 | | | | | | | | | | | | |
| 30 | 8.7 | | | | | | | | | | | | |
| ≤29 | | | | | | | | | | | | | |
| Women: cholesterol concentration (mmol/L). | | | | | | | | | | | | | |
| | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | No | No | Yes | No | No |
| Hypertension | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | No | No | Yes | No | No |
| Smoking | Yes | No | Yes | Yes | No | Yes | No | Yes | No | Yes | No | No | No |
| Diabetes | Yes | Yes | No | Yes | No | Yes | Yes | No | Yes | No | No | No | No |
| LVH | Yes | Yes | Yes | No | Yes | No | No | No | No | No | No | No | No |
| Age (yr) | | | | | | | | | | | | | |
| 70 | 5.5 | 5.5 | 5.5 | 5.8 | 6.3 | 6.9 | 8.5 | 9.8 | | | | | |
| 68 | 5.5 | 5.5 | 5.5 | 5.8 | 6.4 | 7.0 | 8.6 | 9.9 | | | | | |
| 66 | 5.5 | 5.5 | 5.5 | 5.9 | 6.5 | 7.1 | 8.7 | 10.0 | | | | | |
| 64 | 5.5 | 5.5 | 5.5 | 6.1 | 6.6 | 7.2 | 8.9 | | | | | | |
| 62 | 5.5 | 5.5 | 5.5 | 6.2 | 6.8 | 7.4 | 9.1 | | | | | | |
| 58 | 5.5 | 5.5 | 5.5 | 6.7 | 7.3 | 8.0 | 9.8 | | | | | | |
| 56 | 5.5 | 5.5 | 5.5 | 7.0 | 7.7 | 8.4 | | | | | | | |
| 54 | 5.5 | 5.5 | 5.5 | 7.4 | 8.1 | 8.9 | | | | | | | |
| 52 | 5.5 | 5.5 | 5.9 | 7.9 | 8.7 | 9.4 | | | | | | | |
| 50 | 5.5 | 5.5 | 6.4 | 8.5 | 9.3 | | | | | | | | |
| 48 | 5.5 | 6.0 | 6.9 | 9.3 | | | | | | | | | |
| 46 | 5.5 | 6.7 | 7.7 | | | | | | | | | | |
| 44 | 5.5 | 7.5 | 8.6 | | | | | | | | | | |
| 38 | 8.0 | | | | | | | | | | | | |
| 36 | 9.7 | | | | | | | | | | | | |
| ≤35 | | | | | | | | | | | | | |

Sheffield table for primary prevention of CHD

A patient whose value falls in the area with no entries has an estimated risk of coronary events of less than 3.0% per year.

Notes on use of this table

- Do not use for decisions on secondary prevention: patients with myocardial infarct, angina, peripheral vascular disease, or symptomatic carotid disease already have high CHD risk.
- At this CHD risk (3% events per year) treatment with a statin (but not necessarily other drug classes) is justifiable.
- The value of cholesterol reduction at <5.5 mmol/L or age >70 years is not established.
- Use table after appropriate advice on smoking, diet, and control of systolic blood pressure to ≤160 mm Hg.
- Use the average of two or more cholesterol concentrations.
- Consider individual factors (eg. low high-density-lipoprotein cholesterol or bad family history), in final treatment decisions.
- The table is valid for UK, Northern European, and North American populations – Southern European and Far Eastern populations have lower CHD risk in relation to the standard risk factors.

Instructions

- Choose the table for men or women.
- Identify the correct column for smoking, hypertension, diabetes, and left ventricular hypertrophy (LVH) by ECG.
- Identify the row showing the age of the subject.
- Read off the cholesterol concentration at the intersection of the appropriate column and row:
If there is no entry, cholesterol need not be measured.
If there is an entry, measure serum cholesterol.
If the average cholesterol on repeated measurement is at or above the level shown, the CHD event risk is ≥3.0% per year – consider treatment.
The table can be used to look forward to possible need for measurement or treatment at an older age.

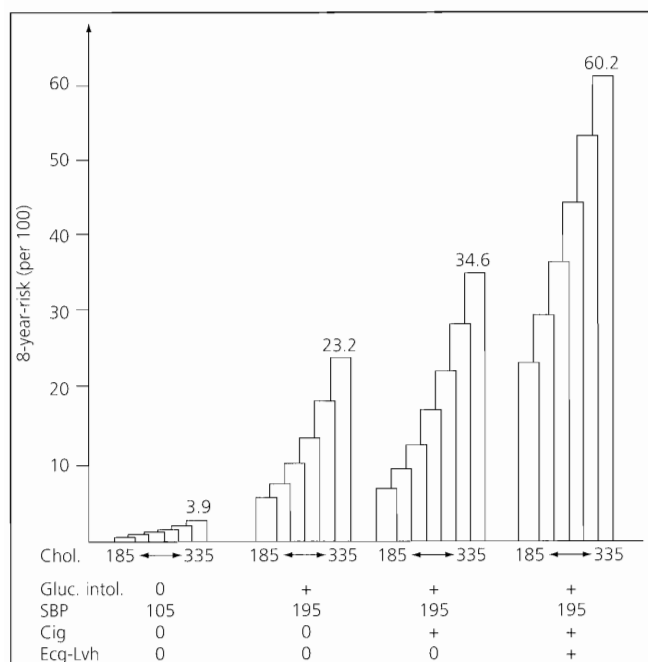


Figure 1. Framington data grouped into risk groups.

esterol levels do not matter. High-risk individuals are a problem, reduction of their cholesterol should be considered even if cholesterol values are 'normal'.

The number-needed-to-treat has enormous practical consequences because a high number means side effects to a high number of people. The same efficacious treatment that is beneficial in high-risk patients, therefore, can become detrimental in low-risk individuals. Indeed, Davey Smith et al. showed in their meta-analysis, in which they grouped studies according to the initial coronary risk of the study population, that lowering cholesterol in high-risk patients with established heart disease reduced overall mortality, whereas lowering cholesterol in low-risk individuals with no or only a few other risk factors increased overall mortality.⁹

How should GPs be guided?

Evidence-based guidelines, consequently, should not simply recommend treatments for a certain condition, but rather lead to the question whether the treating doctor has come to an appropriate understanding of the number-needed-to-treat involved and has discussed the treatment option with the patient accordingly. Therefore, these guidelines should read like this: 'Have you assessed the risk difference your patient may gain from the treatment and have you discussed whether your patient considers this reduction to be important enough to supersede side effects or other inconveniences?' And, of course, we should not only tell patients about risks, but also about the chances of staying healthy. 'Your cholesterol level would usually be considered high. Your risk of getting a heart attack is 5% in the next five years and it can be reduced to 3% through treatment. This means that the chance you will remain

healthy is 95%; if we treat you, this chance will increase slightly to 97%.' And we should perhaps add that long-term effects of treatment beyond this time have not been studied.

This is how GPs should guide themselves and how 'statistics' make sense. The fact is, however, that general medicine has readily accepted guidelines based on fallacious generalisations and exaggerations from specialist disciplines without even seeing the problems that come with them. 'Quality assurance' in general practice, moreover, is often restricted to the question as to how well doctors are compliant in following inadequate guidelines. Much more important than the ever increasing number of guidelines, therefore, is that every GP knows the simple distinction between specificity and positive predictive value (post-test likelihood) and between relative and absolute risks (number-needed-to-treat concept).

Limited evidence and informed judgement in general practice

A group of GPs that attended the 3rd UK workshop on teaching evidence-based medicine in Oxford stress in their report that general practitioners must incorporate the basic methods of evidence-based medicine into their work.¹⁰ According to Lipman et al. an expert group is, of course, in a much better position than an individual to be confident that it has thoroughly explored and appraised a topic before producing evidence-based guidelines. But it should also be said that each set of evidence-based practice guidelines may, in practice, reflect only a small part of the clinical workload that requires evidence. They identified ten important characteristics and effects thought to apply to the practice of evidence-based medicine in primary care, including empowering, enabling, and encouraging: 'Words like "empowering", "enabling", and "encouraging" suggest that the practice of evidence-based medicine is more than a mere technical process. We believe that if general practitioners and other primary care clinicians adopt this approach to clinical problem solving there will be profound professional, cultural and even political consequences for primary care.'

Lipman et al. then refer to the notion also formulated by David Sackett that 'evidence-based medicine is based on the principle of deriving structured, answerable questions directly from the problems presented by each patient and not just those with a problem for which evidence has been provided by guidelines. Only when the question is clear may the GP go on to the process of finding, appraising and applying evidence.'

This important message from these GPs as well as Sackett's clarification about the nature of evidence-based medicine underline the notion that for general practice (and most medicine in general) the primary question is what evidence we need and not what evidence we have. This is too easily overlooked and perhaps also not yet quite understood in its revolutionary consequences. This challenge for general practice appears to be such a formidable task because

we must learn to forget about much of what we have learnt in our medical training. It is a platitude to say that we have been used to treating conditions and ‘pathologies’ instead of the patient’s problems. Today, however, clinical epidemiology provides good tools to assess whether a pathological result or finding has a positive predictive value high enough to become important for the patient (and not only for the doctor). It provides good tools to assess whether a treatment option for which we have ‘good evidence’ that it works goes along with a number-needed-to-treat low enough to compete with another option which usually works well in our practice, but for which good external evidence is lacking. ‘Quality circles’ often focus on how to improve on hypertension or diabetes management where it seems easy to define indicators of performance. If good screening performance is achieved this is called ‘quality’. But what about the number-needed-to-treat concept in hypertension?¹¹ Do the guidelines of hypertension and diabetes make sense to the more pragmatic reasoning and needs of general practice? What about the predictive value of blood pressure in the context of the pre-measurement risk? How many so-called hypertensives do I need to treat to prevent one adverse event, or with what likelihood will my patient ever experience a benefit and what is the likelihood that he will suffer labelling effects etc.?

Guidelines should not tell doctors what to do; they should guide doctors to ask all these questions before reaching an informed decision about whether a treatment option truly does more good than harm and whether it makes sense to the patient. In our critical appraisal courses we often see that participants become quite enthusiastic about the new intellectual challenge of how to think about the quality of practice and how to define it in a well thought-out manner. So-called basic sciences are perhaps the basis for specialist medicine. The basic science for general practice, however, is clinical epidemiology, the science of informed decision-making as well as of the construction of clinical research relevant to practice.¹²

The general practitioners at the 6th UK workshop on teaching evidence-based medicine found that combining critical appraisal skills with a humane understanding of the

patient’s situation and his often unrealistic hopes seems to be the key to effective counselling. The patient needs to understand why a possible, expensive intervention, for example, would not yield a true benefit. According to the GPs, the patient’s process of replacing unrealistic hopes with a more mature acceptance of the inevitable was only possible because the GPs themselves had understood on the basis of their own critical appraisal skills why the treatment was unlikely to benefit. They did not merely tell the patient that the guidelines did not recommend the treatment.¹⁰

General practice is about labelling effects, utility analysis, and non-specific effects

Using a diagnostic test in general practice makes sense only if the pre-test likelihood of the disease being confirmed is already high. In a patient with signs of ischaemia on an exercise ECG, for example, who is a healthy middle-aged individual coming for a check-up, this ‘pathological’ result constitutes nothing but a false positive result in well over 90% of cases. False positive labelling and its iatrogenic consequences is a particular research problem of general practice. Doing a test just to be on the safe side does not provide true certainty because in such situations the post-test likelihood hardly differs from the pre-test likelihood. General practice needs to develop reassurance practices, which are less harmful and less costly. What difference does it really make, for example, if the pre-test likelihood of no disease is 99.5% without a screening mammography but 99.8% with a negative screening mammography if the test is done?¹³

Mammography is also a good example to illustrate the importance of using the number-needed-to-treat concept in assessing benefit. It is well known that mammography screening can reduce breast cancer mortality, but does this lead to a true benefit for women? In terms of absolute risks, less than 1 in 1000 women over a period of ten years will be the lucky one to profit from a prevented or delayed death due to screening. Roughly ten times more often a woman will have to face the bad luck of overdiagnosed breast cancer, which would have remained silent without screening. In about 15% or even 50% women have to face the anxiety and inconvenience of a positive screening result (table 2).^{13,14} This is about 250 to 500 times more often than the chance of experiencing a delayed or prevented breast cancer death. This perspective resembles gambling and the question of whether such screening constitutes a net benefit is not a medical one. Some people like gambling and some women may want to use screening mammography, but would we trust the economic profession if it said that going to a casino regularly prevents and solves the problem of poverty? Women should be properly informed about what they can and have to expect and they should be able to decide according to their own values. General practice, indeed, should ask what the true benefit of a medical intervention is before accepting specialist guidelines; these are traditionally based on the

Table 2. Early detection in practice.

| Effect of mammographic screening | Per 100,000 women-years | Per prevented cancer death |
|----------------------------------|-------------------------|----------------------------|
| Prevented cancer death | 6.2 | - |
| Increase in overt cancer | 52 | 8.4 |
| Lengthening of disease period | 180 | 30 |
| Positive mammography finding | 1500 | 250 |
| Screening mammographies | 39,000 | 6300 |

Schmidt JG. *J Clin Epidemiol* 1990;**43**:215-25.

outdated concept that improving disease statistics is equivalent to improving the situation of the patients concerned. The research that matters to GPs and their patients is on intangible costs and quality of life aspects (utility analysis).

Eventually, we need to understand why clinical research has been based on placebo-controlled trials. The placebo-controlled trial was developed to test the efficacy of pharmacological agents (the internal validity of pharmacological theories). As a consequence, the placebo-controlled trial does not evaluate treatment as it works in practice. Before the era of evidence-based medicine, it was believed that the study of pharmacological efficacy was sufficient and patient benefit was an extrapolation from it. In general practice, however, the entire therapeutic setting including perhaps placebo components is what determines the intervention effect. The big step general practice has to take, therefore, is to dare to talk about matters such as placebo and other non-specific effects which are said to be 'just in the head'. Pure effects are a rather unimportant research question. Academic centres, even if they teach evidence-based medicine, will continue to be so preoccupied with what is tolerated as being 'scientific' that they will not help general practice out of the troubles. We have to do it ourselves by studying basic research methodology and by using it for the very questions that arise from our particular challenges in a complex practice environment. This is also important because it helps us to understand what we have been doing and what we could do for the benefit of our patients in general practice. We can design pragmatic clinical trials, for example, in which the doctor-patient relationship is not separated, but studied or even randomised as a unit.³ ■

References

- 1 Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology; A basic Science for Clinical Medicine*. Boston/Toronto: Little, Brown and Company, 1991.
- 2 Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making*. Butterworth-Heinemann: Boston, 1988.
- 3 Schmidt JG, ed. Placebo - Valuable if it helps the patient? Methods and design of a pragmatic clinical research oriented towards patient benefit. *Forsch Komplementärmed* 1998;5:suppl. 1.
- 4 McCormick JS. The place of judgement in clinical practice. In: Schmidt JG, Steele RE, eds. *A critical view of medical reasoning - steps toward timely medical practice*. Mainz: Verlag Kirchheim, 1994:227-32.
- 5 Feinstein AR, Horwitz RI. Problems in the 'Evidence' of 'Evidence-based Medicine'. *Am J Med* 1997;103:529-35.
- 6 Kannel WB, Castelli WP, Gordon T. Cholesterol in the prediction of arteriosclerotic heart disease: New perspectives based on the Framingham Study. *Ann Int Med* 1979;90:1985-91.
- 7 Schmidt JG. Cholesterinscreening: Die Irrationalität von Grenzwerten und die Berücksichtigung des Gesamtrisikos für eine rationale Therapie. In: Kochen MM. *Rational Pharmacotherapy in General Practice*. Berlin Heidelberg: Springer-Verlag, 1991:67-81.
- 8 Sheffield table for primary prevention of CHD. *Lancet* 1996;348:1352.
- 9 Davey Smith G, Song F, Sheldon TA. Cholesterol lowering and mortality: The importance of considering initial level of risk. *Br Med J* 1993;306:1367-73.
- 10 Lipman, Rogers, Jones Elwyn. Evidence-based medicine in primary care: some views from the 3rd workshop on teaching evidence-based medicine. *Evidence-Based Med* 1997;2:133-4.
- 11 Jackson RT, Sackett DL. Guidelines for managing raised blood pressure: Evidence based or evidence burdened? *Br Med J* 1996;313:64-5.
- 12 Feinstein AR. Why do we need clinical epidemiology? A practice-oriented clinical science. In: Schmidt JG, Steele RE, eds. *A critical view of medical reasoning - steps toward timely medical practice*. Mainz: Verlag Kirchheim, 1994:233-43.
- 13 Schmidt JG. The epidemiology of mass breast cancer screening: A plea for a valid measure of benefit. *J Clin Epidemiol* 1990;43:215-25.
- 14 Elmore JG, Barton MB, Mocerri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338:1089-96.