ABSTRACT – Guidelines for medical management are now part of medical life. A fool – loosely defined as someone who does not know much about a particular area of medicine – will do well to follow guidelines when treating patients, but a wise man (again, loosely defined as someone who does know about the disease in question) might do better not to follow them slavishly. The problem is that the evidence on which guidelines are based is seldom very good. Clinical trials have a variety of problems which often make their relevance to ‘real world’ medicine dubious. The interpretation of trial results depends heavily on opinion, and a guideline that purports to be evidence based is actually often opinion based. A guideline will depend on the opinions of those who wrote it, and the wise man will use his judgement and give due weight to his own opinions and expertise.

KEY WORDS: clinical trials, evidence-based medicine, guidelines

Fools and wise men

There cannot be many areas of medicine not covered by guidelines. We begin to feel that unless we follow guidelines we are in some way providing inferior treatment, and we even wonder if someone may call our management negligent. But guidelines inappropriately applied are the antithesis of the concept that a patient should be treated as an individual. We therefore must ask ourselves whether a particular guideline makes sense in general, and sense for each particular patient – but once we allow that degree of ‘clinical freedom’ we seem to be back in the bad old days when Doctor Knew Best. If guidelines were infallible there would be no problem, but of course they are not; the difficulty is to know whether a particular set of guidelines is sensible or not.

Here I need to define my particular use of the words ‘fool’ and ‘wise man’. For the purposes of my argument I shall use ‘fool’ to mean ignorant: someone treating a patient with a disease of which he has no special knowledge would clearly do well to follow local guidelines to the letter. Alternatively, he might look up management in a textbook, which would come to much the same thing, though presumably the information might be somewhat out of date. Anyone who knows himself to be ignorant of some field would indeed be a fool to ignore appropriate guidelines. But I use the term ‘wise man’ to include those who do know a reasonable amount about the disease they are treating, and who are wise enough to realise that guidelines can, and should, be questioned. I shall develop my arguments using cardiovascular examples, for to use others would show me to be a fool. Those who are wise in other fields will be able to use their own examples to make the same points that I shall make here.

Evidence-based guidelines

In 1983, carried away with the enthusiasm generated by the series of clinical trials of beta blockers in myocardial infarction, I suggested that ‘clinical freedom’ was dead.1 Doctors could no longer treat their patients as they, as individuals, saw fit. They had to base treatment on published trials which provided the only way of knowing whether treatments were valuable or not. An individual doctor could never treat enough patients with a single disease to learn whether or not a treatment worked. I did not coin the phrase ‘evidence-based medicine’, but this is what I meant and the term was defined by Sackett as

\[ \text{the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual} \]

Trial acronyms

- COBALT: Continuous infusion versus double-blind administration of ALTeplase
- CURE: Clopidogrel in Unstable angina to prevent Recurrent Events
- DIG: Digitalis Investigators Group
- GUSTO-1V ACS: Global Utilisation of Strategies To Open coronary arteries – 4th study, Acute Coronary Syndromes
- INJECT: International Joint Efficacy Comparison of Thrombolytics
- ISIS-1, -3, and -4: International Studies of Infarct Survival (1st, 3rd and 4th studies)
- RITA-1: Randomised Interventional Treatment of Angina (1st study)
- TRENT: Trial of Early Nifedipine Treatment
patients… [It] means integrating clinical expertise with the best available evidence from systematic research.²

In this definition we can already see the problem of evidence-based medicine, which has to allow for ‘clinical expertise’. By that, Sackett meant

the proficiency and judgement that individual clinicians acquire through clinical experience and clinical practice.

A guideline must obviously be evidence-based. Any guideline worth its salt must be based on systematic research, but what room is left for clinical expertise? Who is to judge the expertise? Perhaps the greater the expertise the more the systematic research is to be questioned, and that would fit with my suggestion that guidelines are for the guidance of wise men. It might help if we were to accept that anything to do with a definition of evidence-based medicine that allows the incorporation of ‘clinical expertise’ will actually be opinion-based as much as it is evidence-based. This leads to the obvious question, ‘Whose opinion?’, and then authorship of guidelines becomes as relevant as their contents.

Systematic evidence and clinical expertise

Within living memory great doctors were the repository of clinical expertise, and patients always wished – perhaps still do wish – to be seen by an individual doctor with a high reputation. Patients do not queue to see a doctor whose main claim to fame is that he unfailingly and without question applies guidelines to all the patients he sees. Everyone can remember his or her favourite grand old clinician. However, history shows that while clinical expertise is important, it must at least be tempered by the results of systematic research. The digitalis story provides the classic example of the interdependence of the two.

In 1785 Withering published his Account of the foxglove.³ He had been using it for the treatment of dropsy for ten years, and in the introduction he wrote

The use of the Foxglove is getting abroad and it is better the world should derive some instruction, however imperfect, from my experience, than that the lives of men should be hazarded by its unguarded exhibition, or that a medicine of so much efficacy should be condemned and rejected as dangerous and unmanageable … Experience and cautious attention gradually taught me how to use it. For the past two years I have not had occasion to alter the modes of management; but I am still far from thinking them perfect.

In short, Withering’s dramatic discovery depended entirely on clinical expertise. He described a long series of cases, and concluded that his treatment was most effective ‘if the pulse be feeble or intermittent’ – clearly atrial fibrillation. He described how the foxglove was ineffective, and indeed potentially dangerous, when ascites was present in ‘a hard drinker’, and he found that it was equally ineffective in ‘ovarium dropsy: this species of encysted dropsy is not without difficulty distinguishable from an ascites, yet it is necessary to distinguish them because the two diseases require different treatment’. Some physicians of the day, and afterwards, did not heed Withering’s warnings and digitalis began to get a bad name as a drug that caused as much harm as good. The argument about the value of digoxin in patients with heart failure but whose hearts were in sinus rhythm was only settled by the 7,000 patient DIG trial published in 1997.⁴ This study showed that digoxin treatment had no effect on mortality, because a reduction in deaths from heart failure was balanced by an increased number of deaths from (presumed) arrhythmias. The frequency of worsening of heart failure was, however, reduced by digoxin treatment.

Thus with digoxin, clinical expertise came first and systematic research second. The DIG trial could not have been carried out safely or sensibly without prior clinical observation, because the wrong sort of patients might well have been included or excluded. Without Withering’s clinical expertise, the value of digoxin would probably have been overlooked. Whether digoxin should be included in guidelines for the treatment of heart failure becomes a matter of opinion: it can be argued that as it has no effect on mortality (at least when the heart is in sinus rhythm) it is not worth taking, or alternatively that it reduces hospital admission so it is.

History provides other examples where clinical observation was enough for a treatment to be accepted without anything that would now be recognised as a clinical trial (for example, penicillin) or where the trial preceded any useful observations, as with James Lind’s elegant study of apples and oranges for the treatment of scurvy.

Of digoxin, penicillin, and apples and oranges, probably only apples and oranges would get into a guideline on the basis of systematic research. Digoxin and penicillin might well be rejected by a guideline committee that insisted on favourable results from randomised trials, especially if they thought Withering’s expertise was a bit out of date. But the wise man would probably use all three of these treatments if he thought them appropriate to his patient.

The limitations of clinical trials

The randomised, double-blind clinical trial now occupies centre stage in evidence-based medicine and the writing of guidelines, and will presumably continue to do so until molecular biology gives us techniques that show who will respond to what drug. Until then we are stuck with clinical trials, but we must not accept them as all-powerful simply because we have at the moment no better way of assessing the efficacy of a treatment.

Of the many worries about clinical trials, perhaps the most important is their relevance to clinical practice – to ‘real world’ medicine. A comparison with the overall fatality rate (in treated and untreated groups) in a published clinical trial with what ‘feels right’ in practice gives the first pointer to trouble. If we look at the trials of thrombolysis in acute myocardial infarction, either the early trials where active treatment was compared with placebo, or the later ones where two thrombolytic agents were compared, we find that the 30-day fatality rate was 8 or 9%.

The true ‘real world’ fatality rate after myocardial infarction, if all patients admitted to hospital are included, is probably at least
15%. Much of this difference stems from the fact that the elderly, in whom myocardial infarction is most often fatal, do not figure prominently in the trials. A long-term follow-up of infarct patients in Nottingham found that among those who were given a thrombolytic agent as part of a clinical trial, the fatality rate was much what all the trials showed. The fatality rate of patients who were given a thrombolytic outside the confines of a trial was 50% higher, and that of patients not given a thrombolytic at all (for whatever reason) was more than twice as high. Clearly the thrombolytics trials selected low-risk patients, leaving the question of their value in the ‘real world’ largely unanswered. It is difficult to write guidelines when evidence is in short supply.

In some of the trials of beta blockers after myocardial infarction the fatality rate was also low. For example, in the ISIS-1 trial, which provided the only evidence for the use of atenolol, the fatality rate was about 4%. This is so low that the result cannot be assumed to apply to the generality of patients with infarcts.

The heart failure trials tell the same story. Heart failure is essentially a disease of the elderly, and of elderly women at that. Few of the trials of the many drugs shown to be effective in treating heart failure included many elderly women, so we cannot be certain that the trial results apply to them. We can suspect that even if the drugs ‘work’ in a physiological/pharmacological sense they may not be effective. Elderly patients admitted to hospital with heart failure always have other diseases, and all are taking drugs – sometimes several drugs – for non-cardiovascular problems. The potential for drug interactions in these circumstances is enormous, the most obvious being between angiotensin-converting enzyme inhibitors and non-steroids, which are so commonly used in the elderly. The morbidity of a drug interaction can easily outweigh any benefit from a particular anti-failure drug, but it is clearly going to be impossible for a guideline to cover all possible drug combinations.

Treatment is, in these circumstances, a matter of judgement – or perhaps, expertise.

The importance of disease registers

The only systematic way of showing the relationship of a trial result to the real world is with a trial register. A complete record needs to be kept of all the patients who were considered for the trial, so that at the very least it is possible to see what proportion of them was included. Better, the reasons for exclusion need to be logged, and ideally the excluded patients are followed in exactly the same way as those who were included in the trial. This is obviously time-consuming and therefore expensive and seldom done. When registers have been kept and all the patients followed up, the findings have been as one would suspect: the patients included in the trial were those already at low risk.

Thus in the TREAT trial of nifedipine for the treatment of early myocardial infarction, the fatality rate was about 6%, nifedipine showing no advantage over placebo. But the register of all infarct patients admitted to the hospitals participating in the study showed that the death rate among excluded patients was 27%. Similarly, in the TRACE study of trandolopril in patients with impaired left ventricular function after acute myocardial infarction, the one-year fatality rate for those included in the trial was about 20%, but was three times that in patients who were excluded.

Registers that simply show the number of patients included in a trial, compared with the total who might have been considered, inevitably raise doubts about the relevance of the trial result. The fact that the doubt is unquantifiable leads to a further loss of confidence. Thus in the RITA-1 trial that compared coronary angioplasty and bypass grafting for the treatment of angina, the register showed that only about 3% of the patients undergoing angiography for angina in the participating centres were included in the study. The trial showed that angioplasty and bypass grafting produced similar results in terms of death and myocardial infarction, and allowed the two procedures to be compared for cost effectiveness. But should treatment of a patient with angina really be based on the result of a trial that may not have included ‘typical’ patients? A guideline would have to take RITA and the other similar studies into account, but a wise man might have doubts about the wisdom of following such a guideline in any particular patient.

The importance of post-trial surveillance

Because of the doubts and problems I have described, one might have expected that the effects of any drug shown to be effective in a clinical trial would immediately be checked in routine practice. Because of the general love affair with the double blind trial this is seldom done. The RALES study showed that spironolactone dramatically reduced fatality in patients with severe heart failure. This seemed an excellent study, and as soon as it was published we incorporated its results into our routine practice and we audited the first 50 patients so treated (unpublished). To our

---

**Key Points**

- Someone not an expert in a particular field (here described as a fool) will do well to follow guidelines when treating an individual patient, but someone who knows his subject (a wise man) will know that guidelines are fallible.
- The main problem is the interpretation of clinical trials. Trials usually include only ‘low risk’ patients, and their relevance to the ‘real world’ is often doubtful.
- Post-trial observational data are seldom available but when they are they may suggest that the trial results do not readily translate into clinical practice.
- The interpretation of the results of clinical trials is always a matter of opinion, and guidelines are therefore as much ‘opinion based’ as ‘evidence based’.
- Medical treatment by a wise man who depends on an assessment of each individual patient and the application of expertise, judgement and common sense, will usually be preferable to a slavish application of guidelines.
dismay – but not necessarily surprise – we found that the drug was less easy to use than the published trial had suggested. Whereas in the trial only 2% of patients developed an unacceptably high plasma potassium level, we observed this in 14%, and treatment had to be discontinued for a variety of reasons in 22%, compared with 0.6% in the trial.

How can such observations be included in guidelines? Our patients differed from those in the trial – the mean age was 75 compared with 65 in the trial and 46%, compared with 27%, were women. The real world and the trial world are, quite simply, different.

Clinical trials and opinion

The extent to which guidelines depend on opinion is disturbing for anyone who believes they should be evidence-based. Whether they think about it or not, clinicians do apply their own opinions to the implications of clinical trial results. This is clearly seen in the use of beta-blockers for the secondary prevention of myocardial infarction, where the clinical body at large uses atenolol, a beta-blocker for which there is not much evidence. It is obvious that most people believe that one beta-blocker is equivalent to another. This has to be a matter of opinion, for no post-infarction comparative studies of beta-blockers has been published.

It is possible to design comparative drug trials to demonstrate equivalence, as was done in the INJECT trial\textsuperscript{12} that compared streptokinase and reteplase in the treatment of patients with acute myocardial infarction. But it has to be appreciated that equivalence can only be demonstrated within quite wide limits if the number of patients to be included in the trial is to be within a practical range. Attempts to prove equivalence within limits that are too strict can lead to curious results, such as in the COBALT trial\textsuperscript{13} which compared two methods of administering alteplase. The statistical result was that a fatality rate of 7.98% seen when alteplase was given by one means was not equivalent to the rate of 7.53% when another regimen was used. Few would think that these two results were not equivalent, but that would be a matter of opinion taking precedence over the statistical basis of that trial. There is nothing wrong with the application of opinion, but one needs to be clear that that is what is happening.

Opinion also plays a major role in the systematic combination of multiple clinical trials by the statistical technique of meta-analysis, which has somehow become elevated to a position above that of even a good single trial in the minds of many guideline writers. This is the point where the wise man will inevitably part company with the guideline.

Meta-analysis may, with a little luck, give some indication of whether a drug or group of drugs has an effect on some disease. What the technique can never do is tell a doctor how to treat an individual patient, and thus its place in guideline writing must be questioned.

A classic example of a meta-analysis which is scientifically interesting but therapeutically useless is the ‘anti-platelet’ trialists analysis of the effect of drugs that modify platelet behaviour in the treatment of vascular disease.\textsuperscript{14} The conclusion was that anti-platelet drugs lead to a significant improvement in outcome, and this is important because for many years there was doubt about the clinical relevance of tests of platelet function. From the therapeutic point of view, however, the meta-analysis was unhelpful because a variety of ‘anti-platelet’ drugs, including those not now used, like sulphipyrazone, were combined to study the effect of ‘anti-platelet’ treatment. An individual doctor treating an individual patient – or someone writing a guideline – needs to prescribe or suggest a particular drug in a particular dose for a particular type of patient, and this is something meta-analysis can never provide.

Meta-analysis of the results of a group of trials that were individually inconclusive is, however, quite likely to be misleading. For example, a series of small studies had suggested that intravenous magnesium reduced mortality after myocardial infarction. Meta-analysis of these trials suggested that the effect was real, and magnesium treatment might well have been incorporated into guidelines for the management of all patients with an acute infarction. However, before this could happen the ISIS-4 trial,\textsuperscript{15} which was much bigger and better-designed than its predecessors, showed that magnesium had no effect. Opinion about the different trials becomes as important as the results themselves: the fool will be forced to follow whatever the next lot of guidelines happen to say, but the wise man will keep an open mind about the wisdom of prescribing magnesium.

The wise man will, in fact, take results of any meta-analysis with a pinch of salt. When Le Lorier \textit{et al.}\textsuperscript{16} reviewed 19 meta-analyses with the results of 12 large relevant trials published later, they found that without the large trials, meta-analysis would have caused the adoption of an inefficient treatment in a third and the rejection of a useful treatment in another third.

From evidence to guidelines

Guidelines are written from evidence filtered through opinion. The opinion is crucial – but whose opinion should it be? The wise man, who knows the field, will know what a guideline contains as soon as he reads the list of authors.

The National Service Framework for Coronary Heart Disease exemplifies the problems.\textsuperscript{17,18} Here we have a massive, ‘official’ document which some would take to be a managers’ handbook but when its hundreds of bullet points are stripped away it can be seen to be a guideline for patient management. The treatments advised would be suitable for a first-year medical student. Resuscitation is advised for cardiac arrest, pain relief is important in myocardial infarction, hospital admission is a good idea, aspirin and heparin are good for unstable angina, diuretics are useful in heart failure and so on. Treatment has been reduced to a simple level with which none could disagree. Presumably the authorship had something to do with it – the writing committee comprised 41 people, three of whom were cardiologists and another eight were clinicians with other interests. The remainder were non-clinical. Guidelines that say nothing of use, written by a committee out of touch with the subject, are unlikely to find favour with clinicians and indeed bring the whole of the guideline concept into disrepute.
Even when a guideline is produced by a more clinically relevant body such as the National Institute for Clinical Excellence (NICE), the same principles apply. The NICE committee is made up of a variety of experts in different disciplines who take specific advice from a small number of specialists in the relevant field. These specialists may or may not hold an opinion widely shared by their (equally expert) colleagues. Thus the original NICE guidelines on the use of glycoprotein 2b 3a receptor antagonists recommended their use in patients with high-risk acute coronary syndromes at a time when the cardiovascular fraternity was still in some doubt about the strength of the evidence. The trial for which all had been waiting, GUSTO-IV ACS showed that these drugs do not have the beneficial effect that had been hoped for, and the guidelines had to be re-written.

The final problem with guidelines is the cost of their implementation. Many centres did not follow the original NICE guidelines on glycoprotein 2b 3a antagonists because of their cost, but supposing the advice had been good, where would a hospital trust that prevented clinicians from using this type of drug have stood if a patient complained of lack of recommended treatment? It could only be a matter of time before the matter was tested in the courts. The current guidelines for the management of acute coronary syndromes published by the European Society of Cardiology suggest, among many other things, that on the basis of the CURE trial patients with unstable angina should be treated with the anti-platelet drug clopidogrel as well as aspirin. In my trust, and I suspect many others, clopidogrel has not been permitted on cost grounds. Should cost be taken into account by guideline committees? Does the UK have to follow European guidelines? Does every hospital have to produce its own set of guidelines, taking its budget into account? In short, whose guidelines should we follow, and must we follow any?

The wise man’s attitude to guidelines

The fundamental aim of a guideline is to get away from individualised treatment. The enthusiast will claim ‘levelling up’ and the cynic will – as always – fear ‘levelling down’. A guideline intends to play down the role of clinical expertise and emphasise the importance of systematic research, but the results of the systematic research – which at the moment means clinical trials – are far from certain and always open to interpretation by opinion. Even the original definition of evidence-based medicine had to pay lip service to the expertise of the individual clinician.

The wise man will know the limits of the evidence, and also the limits of his own expertise. The wise clinician will treat his patient as an individual, even if his therapy is derived from a statistical analysis of the treatment of many patients. But what if his advice to a particular patient differs from a guideline? Should he should paraphrase Churchill’s view of statistics and misquote him by saying “The only guidelines I believe are the ones I have written myself”?

Perhaps the last word should go to Lady Thatcher, who summed up her attitude to guidelines in the Inquiry into the ‘Arms to Iraq’ affair. The transcript of her evidence includes the following:

Lady Thatcher: Guidelines are exactly what they say they are. They are guidelines. They are not the law.

QC: But do they have to be followed?

Lady Thatcher: Of course they have to be followed, but they are not strict law. That is why they are guidelines and not the law and, of course, they have to be applied according to circumstances.

References

15 ISIS-4 (Fourth International Study of Infarct Survival) Collaborative


