Evidence based medicine overviews, bulletins, guidelines, and the new consensus

Competent clinical practice is necessarily based on good scientific evidence of effectiveness. Few would argue with this justification of “evidence based medicine” (EBM) and the best, most appropriate clinical interventions and therapies have probably always been based on the most relevant available evidence. So, we may ask, why is EBM a new idea. The “syndrome” has a multiple epidemiology. Explanations lie with the profusion of evidence, rate of accrual of new evidence (including contradictory evidence), changes in means of storing, retrieving, and disseminating evidence, and the increasing pressure on healthcare professionals from both government and the public to ensure that clinical practice is based on evidence. The present paper reviews this rapidly expanding “industry” of secondary research, which summarises the findings of primary studies and trials and packages the synopses for easy assimilation by busy practitioners and managers, and discusses some of its advantages, disadvantages, and possible consequences.

Need for evidence and a “hierarchy of evidence”
Evidence to support a clinical action or decision may be drawn from many sources. Epidemiologists have described for a number of years a “hierarchy of evidence”, ranging from clinical experience, case report or anecdote through structured observational studies, like the case-control study, to randomised controlled trials. Some have recently added the statistical overview, or “meta-analysis”. It would be naive to assume that “experienced clinical opinion” is based wholly on a clinician’s own experience; it will invariably incorporate a wider knowledge drawn from medical school (the influence of a charismatic teacher may last a lifetime), relevant texts, and from selected reading of more recent literature and may well reflect a current “consensus view”. It would be equally naive to regard one clinical trial, even a large well designed trial, as providing the definitive answer on its own, because a trial is often restricted to a selected subset of patients under atypically controlled circumstances and with insufficient numbers to evaluate side effects. The trial results need to be interpreted in context and the guideline for good clinical practice lies in the synthesis of results of repeated trials, findings of relevant observational studies, and knowledge of the underlying biological, biochemical, or pharmacological mechanisms. That synthesis was previously to be found in textbooks, but books are increasingly being replaced by reviews, bulletins, and guidelines.

Rationale for EBM, overviews, bulletins, and guidelines
The principal justification for EBM is the quest for relevant and reliable evidence to demonstrate effectiveness, or relative effectiveness, so that practitioners offer only effective, or most effective therapies. Search for effective treatments and evidence of effectiveness is so logical it cannot be new to clinical medicine: practitioners have sought more effective treatments for generations. What is new, or perceived to be new, is that knowledge advances so quickly that textbooks are out of date and therefore a new means of generating and disseminating a synthesis is needed. A new modus operandi may be indicated, since it is said that there are more scientists practising now than in the whole of previous history. With more clinical and relevant basic science research being undertaken and more journals reporting the findings (two million articles annually in 20 000 biomedical journals), there is a plethora of potential evidence. Other factors have contributed to the recent emergence of EBM. One is probably the institutionalisation of medical care and a change from the two party doctor-patient relationship to a three party employer-doctor-patient interaction. We no longer have independent clinicians, offering treatment or care to patients, but major industries, employing professionals from a wide variety of disciplines. Another is probably the economic pressure or lack of political willingness to provide all, that is sought by patients, or all, that is offered by clinicians. Rationing by one name or another is endemic in most health services. While Cochrane, before the founding of the NHS, marched behind a banner proclaiming “all effective treatments should be free”, Hine in launching the Clinical Effectiveness Initiative in Wales suggested a rephrasing “all free treatment should be effective” (D Hine, personal communication, 1996). The combined effects of
institutionalised provision, budget limitations, an ever expanding armory of potentially effective clinical interventions, a possibly expanding expressed demand for health care and a rapidly expanding supply of information, provides more than sufficient justification for systematic appraisal of the accumulating evidence.

Recognition of the need for reliable up to date synthesis of relevant information of effectiveness has led to development of systematic reviews, statistical overviews, bulletins, and guidelines. Systematic reviews seek to summarise the findings in the extensive and often extending primary literature and to emphasise trends, if there are trends, and explain differences, if there are differences. Systematic reviews may or may not include statistical overviews (or “meta-analysis”): they often do, since the most objective information on effectiveness or relative effectiveness is often in the form of numerical comparisons. Guidelines replace textbooks on clinical practice, to provide concise summaries of recommended actions on the basis of evidence of effectiveness, and bulletins, which can be either systematic reviews or guidelines or contain elements of both, ensure the rapid and widespread dissemination of the synthesised evidence.

Evolution of EBM

The recent “discovery” of EBM has been heralded as a paradigm shift but early antecedents should not be forgotten: all to often citations in the medical literature are limited to the most recent 10 (or even five) years of Medline, suggesting that “medical memory” is quite short. Examples of search for evidence in previous centuries might include the classic experiment of treating scurvy with citrus fruits, experiments with cow pox vaccination, and rice and beri beri. Recent developments are generally attributed to Cochrane and Sackett. Cochrane repeatedly challenged clinical colleagues on the absence of good experimental evidence and, armed with observational data of variation, provoked them with “if practice varies so widely, you cannot all be right” (A L Cochrane, personal communication, 1971). An example here might be the prescription of antibiotics for respiratory tract infections in general practice. The observational study of Howie et al, which showed variation from 25% to 96% (of patients with comparable symptoms being prescribed antibiotic), provided a suitable introduction for a randomised controlled trial. Sackett’s contribution was more firmly centred on ward round teaching, challenging students to quote not “chapter and verse” from leading textbooks but original references of “evidence”, to support a proposed clinical action.

One of the leading developments in systematic reviewing, the international Cochrane Collaboration, which evolved from 10 years of reviewing clinical trials in the field of pregnancy and childbirth is named in Cochrane’s memory. The UK NHS research and development programme supports a number of Cochrane Collaboration groups and has also established a number of other systematic review or effectiveness bulletin writing groups, which prepare and distribute bulletins for the benefit of practising clinicians and managers, too busy to review the primary literature themselves.

Advantages and disadvantages of systematic reviews and statistical overviews

The main advantages of systematic reviews lie with their rationale and scientific rigour, evaluating all the relevant evidence and not simply a convenient sample. The disadvantages arise mainly from their practice, that they are not sufficiently scientific and may not all be sufficiently rigorous, and their consequences. The pros and cons will be discussed briefly in the following sections.

Repeatability

One of Hill’s “criteria for causality” was that an association should be replicable in different samples, populations, places, and times. When a systematic review finds an effect consistently, it clearly adds importantly to the evidence but more often than not some trials (studies) show an effect, while others do not. The systematic review puts emphasis on seeking out and including all relevant and competent trials (studies), whether or not they report on effect. The “bottom line” summary of such a review may be that x/y trials (studies) show the effect but many factors can profoundly influence such a finding. To illustrate this point, two reviews, that appeared in leading medical journals within weeks of each other, examined the role of low molecular weight heparin compared with unfractionated heparin in surgery and, despite accessing the same literature, presented very different summaries (table 1).

Sample size and power

One of the strongest cases for the statistical overview (or meta-analysis), an important component of many systematic reviews, lies with the limited power of most primary trials, owing to small patient numbers. A small trial has only a small chance of correctly detecting a difference (or no difference) at conventional levels of significance. For example, if survival following treatment by a new therapy were to improve from 80% to 90%, no mean improvement, a trial of 200 patients (100 new v 100 standard) would have only an evens chance of demonstrating this improvement at p<0.05 (smaller trials would be unable to report a “significant” improvement of this magnitude: the improvement would need to be greater to achieve p<0.05). Yet when a number of trials evaluate the same treatment, follow a common protocol and compare the same outcome, their findings can be pooled. The effective increase in sample size and hence power allows the null hypothesis (no difference between new therapy and standard) to be rejected with greater confidence or, put the way it is more readily comprehended by most clinicians, allows the new treatment to be shown to be a statistically significant improvement over the standard.

Table 1 Comparison of two statistical overviews of low molecular weight heparin compared with unfractionated heparin in general and orthopaedic surgery

<table>
<thead>
<tr>
<th>Overview</th>
<th>Trials included</th>
<th>Deep vein thrombosis RR (95% CI)</th>
<th>Pulmonary embolism RR (95% CI)</th>
<th>Major bleeding RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leizorowicz et al</td>
<td>25</td>
<td>0.86 (0.72 to 1.04)</td>
<td>0.62 (0.33 to 1.15)</td>
<td>1.02 (0.90 to 1.16)</td>
</tr>
<tr>
<td>Nurmoshad et al</td>
<td>17</td>
<td>0.70 (0.65 to 0.95)</td>
<td>0.44 (0.21 to 0.95)</td>
<td>1.01 (0.70 to 1.48)</td>
</tr>
<tr>
<td></td>
<td>8*</td>
<td>0.91 (0.68 to 1.23)</td>
<td>0.62 (0.21 to 1.87)</td>
<td>1.32 (0.69 to 2.56)</td>
</tr>
<tr>
<td>Orthopaedic surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leizorowicz et al</td>
<td>14</td>
<td>0.83 (0.68 to 1.02)</td>
<td>0.53 (0.27 to 1.03)</td>
<td>1.09 (0.76 to 1.58)</td>
</tr>
<tr>
<td>Nurmoshad et al</td>
<td>6</td>
<td>0.68 (0.54 to 0.86)</td>
<td>0.43 (0.22 to 0.82)</td>
<td>0.75 (0.26 to 2.14)</td>
</tr>
<tr>
<td></td>
<td>5*</td>
<td>0.75 (0.56 to 0.99)</td>
<td>0.76 (0.42 to 1.41)</td>
<td>1.19 (0.36 to 3.90)</td>
</tr>
</tbody>
</table>

*Stricter criteria of trial methodology.

RR = relative risk; CI = confidence interval.
There are many examples, mainly from cancer chemotherapy trials and pharmaceutical trials in cardiology, where combining the data from several small trials, most of which individually showed no benefit, yielded highly significant results. The example of thrombolysis in acute myocardial infarction has been cited many times as evidence of the power of the statistical overview. By 1980 only three of 19 trials showed benefit but in combination (all 19 combined) the benefit was significant. In 1982 an overview selected eight to show significantly reduced risk of early mortality p<0.001 (fig 1), well before the two really large trials, that are regularly cited nowadays as evidence of effectiveness of early thrombolysis.

Medical statisticians encourage the use of estimations with 95% confidence intervals, rather than hypothesis testing with p values. The present example can be used to illustrate the point. When each of the early trials of thrombolysis is reported as relative risk of early mortality (risk of death among patients given thrombolysis/risk of death among controls), it becomes clear that three of the “non-significant” trials also estimated a risk below 1 (fig 1).

Homogeneity and heterogeneity

The principle underlying pooling of similar trials, evaluating a new therapy, is that there is clinical homogeneity; in other words that the trials are truly similar, include similar patients, follow common protocols, administer the same therapy, control for the same potential confounders, and measure the same end points. In practice few trials are so similar: investigators by design or by accident introduce individuality. In some trials, particularly those of new pharmacological trials with registered trials of the same therapy and pharmacological trials in cardiology, variations may be relatively minor and therapeutic effectiveness may be affected little by, for example, introducing more older patients or failing to control for season of year. However it is important to recognise that this is not true for all poolings of trials and that differences in, for example, patient population, diagnosis, co-morbidity, dose or timing, or control for potential confounders may strongly influence outcomes. The potential problems of heterogeneity become much greater in poolings of non-pharmacological therapies and health care evaluations. An example that illustrates diversity is the evaluation of psychological rehabilitation after myocardial infarction: among 12 trials of “outpatient” programmes the interventions differed between programmes in for example profession of therapist, mode of therapy, duration of sessions and number of visits, and between trials in patient inclusion criteria, frequency and duration of follow up, and outcome measures (table 2). Individual therapists and trialsists (when they were different) were enjoying their autonomy but in seeking the “bottom line” effect of psychological interventions after myocardial infarction by pooling these trials, it is necessary to make some quite broad assumptions about what is common to these programmes and their evaluations. There are statistical tests for heterogeneity and these help to ensure that there is some degree of similarity in trial results before they are pooled. Alternatively heterogeneity can be investigated in further post hoc analysis to seek its explanation. In the above example of “clinical heterogeneity”, “statistical homogeneity” would imply that programme content had little effect on outcome.

Publication bias (primary research)

A bias in the selective publication of trials and studies has probably existed as long as the publication of medical science but has only received widespread recognition since the introduction of statistical overviews. Although popularly attributed to the preferences of editors for papers with results or stories to tell, a study of “rejections” revealed that more were by authors than by editors. Many trials or studies with null findings are discontinued or left incomplete in preference for more rewarding lines of research, lines that offer better prospects of significant findings. Negative trials may also be delayed and “successful” trials republished, sometimes unrecognisably under different authorship. Early in the evolution of the statistical overview, its advocates recommended systematic search for unpublished trials to provide an unbiased summary in the pooling. Comparisons of overviews of published trials with registered trials of the same therapy demonstrate the effect of bias, in some examples the former suggest significant benefit while the latter do not.

Unpublished pharmacological trials may be relatively easy

<table>
<thead>
<tr>
<th>Table 2 Randomised trials of psychological rehabilitation during outpatient phase (phase III) after myocardial infarction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial</strong></td>
</tr>
<tr>
<td>Ibrahim (1975)</td>
</tr>
<tr>
<td>Rahe (1979)</td>
</tr>
<tr>
<td>Naismith (1979)</td>
</tr>
<tr>
<td>Mayou (1983)</td>
</tr>
<tr>
<td>Stern (1983)</td>
</tr>
<tr>
<td>Horlick (1984)</td>
</tr>
<tr>
<td>Friedman (1984)</td>
</tr>
<tr>
<td>Fruen Smith (1985)</td>
</tr>
<tr>
<td>Burgess (1987)</td>
</tr>
<tr>
<td>P Re Co (1991)</td>
</tr>
<tr>
<td>Lewis (1992)</td>
</tr>
<tr>
<td>Jones (1996)</td>
</tr>
</tbody>
</table>

*Patient assessed by telephonist; nurse only visited those in need; †take home material; ‡10 minute introduction to material.

Figure 1 Trials of thrombolysis in myocardial infarction, as overviewed in 1982 (relative risk is a log scale) (after Stamper et al).
to locate, since sources of supply are limited, but it is nearly impossible to locate all trials of non-pharmacological therapies, although registers can help. However tracking down unpublished trials, whether or not assisted by pharmaceutical supplier or register, will only partially overcome the effect of publication bias, as long as non-publication includes non-completion and non-submission. It has also been argued that inclusion of accessible unpublished trials introduces bias, since they are not peer reviewed. There are proposals that all grant funded trials (research) should provide summaries of their principal findings, however negative, in a form that may be accessed, understood, and included in overviews. This should help reduce the size of the unpublished problem but is unlikely to eliminate it (see secondary researcher bias).

Post hoc trial selection: secondary researcher bias
An example from the evolving history of cholesterol lowering trials helps to illustrate the potential for secondary researcher bias. When there were some 20 trials in the published literature, one review selected seven (on grounds of quality), which nearly showed a statistically significant mortality reduction, while another selected six, which nearly showed the converse, an increase in mortality. Regression analysis of a fuller list of 19 trials showed a relationship with amount of cholesterol reduction, and more recently a review of 35 trials showed that people (or patients) at an initially high risk of death from cardiovascular disease benefited (significantly), while those at initially low risk suffered (significantly)

One important principle of scientific inquiry is neutrality or objectivity in the testing of hypotheses. This is possible in primary clinical research if one follows the dictum: first the hypothesis then the test of the hypothesis (for example new drug A is more effective than the current standard B and then a randomised clinical trial with prior determined outcome measures), although in practice much primary research follows promising leads (drug A is likely to be more effective, because of its chemical structure or its efficacy in animal experiments). It is easier when the roles of theoretician and experimentalist are separated as in some of the basis sciences (for example, physics). By contrast the review does not follow this important sequence: first the hypothesis then the test. A review is invariably post hoc and a reviewer starts with the findings of (various trials) and selects from these findings, those that suit his purpose. The potential for secondary researcher bias is inseparable from post hoc reviewing and underpins the need for protocol and rigour in the systematic review. Protocol and thoroughness may reduce the effects of secondary researcher bias but all users of evidence, derived in this way, should be aware of the potential.

Good protocol recommends that the trial inclusion/exclusion criteria are decided “before” conducting the review but these decisions remain post hoc, in that the trials have been completed and published and reviewers are often familiar with many trials before deciding to undertake the review at all, let alone deciding on details like inclusion/exclusion criteria. Inclusion or exclusion of individual trials is often based on trial “quality” but assessment of quality in published reports is subjective and variable: studies suggest that reviewers can disagree on “facts” reported in papers let alone on whether the papers describe well designed and well conducted research.

It has also been recommended that statistical overviews search out important missing data, as well as unpublished trials. This acknowledges that many trial reports are incomplete, having suffered under the editor’s knife and that some information relevant to the post hoc pooling exercise may be available directly from primary researchers. This can work but it is somewhat idealistic. Secondary researchers are frustrated when their inquiries elicit “sorry I cannot help” replies but, from the primary researcher’s perspective, the answer to a request is often not readily available (in a “working table” or a table edited out of the final paper) and requires extensive search through original records and reanalysis, long after a grant has closed and assistants moved to other jobs.

It is important that secondary researchers acknowledge the probability or indeed inevitability of prior bias in conducting a review. A scientific training helps to control bias but does not eliminate it. The recommendations of reviewing systematically embody many of the essential scientific principles and help to make the findings fairly objective, excepting that the whole exercise remains post hoc. Secondary researcher bias should be better recognised, because of the weight of evidence that now appears to be accorded to overviews.

Quality and value of systematic reviews and statistical overviews
Concise summaries of the relevant evidence from selected competent trials or studies clearly provide busy practitioners, managers and planners with answers, without their having to engage in time consuming searches of the primary trials literature themselves. Furthermore, when reviews are updated electronically or bulletins are revised at perhaps two yearly intervals, practitioners should feel confident to act secure in the knowledge that the evidence is up to date. However this assumes two competencies: first that the review is a competent synthesis of the relevant primary trials and secondly that the user (practitioner, manager, or planner) competently interprets the review. The first point (covered in the previous section) requires that the review be thoroughly and systematically undertaken and bias kept to a minimum but, in the current climate of rapidly increasing popularity of reviews and overviews, we cannot assume that all are thorough, systematic, and unbiased. It is necessary for end users to develop critical appraisal skills for reading of reviews in just the same way as for the primary trials literature.

Comprehension is enhanced when reviews are clearly written and present their findings in an orderly manner. Reviews and journals can help with more accurately worded abstracts. There have been misleading examples of reviews or statistical overviews (meta-analysis) published under the heading “original research” and of

![Figure 2](http://pmj.bmj.com/PostgradMedJ.1st published as 10.1136/pmj.76.897.383 on 1 July 2000. Downloaded from http://pmj.bmj.com/ on March 11, 2021 by guest. Protected by copyright.)
formatted abstracts in which "setting, population (or patients), and results" have given the impression that "this study randomised 20,000 patients . . . . . . in general practice . . . and found a highly significant risk reduction" when it should more truthfully have indicated "that this (office-based) study selected 10 trials from a possible 20 . . . . and found that seven of 10 showed benefit". It is good form to indicate inclusion/exclusion criteria and list excluded trials and reasons for excluding the excluded. However, just as in reports of primary research, typical reviews are often overly abbreviated (to 3-4 pages) and brevity can allow sins of omission, for example giving reasons for excluding trials without giving all the reasons. Some reviews may concede sloppy science behind fashionable jargon (for example, the current computer database searching terminology). Regrettably there are examples in the literature of reviews that have (i) incorrectly entered the findings of primary trials, (ii) double counted trials (with favourable outcomes), (iii) failed to seek important missing data from authors of published trials, and (iv) failed to inquire after the findings of completed but unpublished trials. Critical appraisal skills can spot inconsistencies or inaccuracies in reviews but often fuller knowledge of the subject and careful rereading of the original literature is necessary to identify omissions.

Generalisability of evidence from systematic review
Reviews, even when well conducted, are only as good as the primary research upon which they are based. Trials that follow similar protocols may give rise to tidy reviews with little heterogeneity. However it is not uncommon for tightly defined trials to provide answers that are not generalisable to typical clinical situations, because they include only a small proportion of the total population, for example exclusive diagnostic definitions, sex (m) and age (<60 years). A pooling of trials may represent a larger proportion and therefore be more generalisable, if there is some intertrial heterogeneity, although often not, because individual trials tend to employ similar patient selection criteria. For example before 1995 published trials in cardiac rehabilitation included fewer than 5% of men over age 65 and fewer than 5% of women, although nearly half of cardiac patients are elderly and nearly one third are women (table 2).50 In the past context setting and cautious informed extrapolation of trial findings to more typical clinical scenarios were included in discussion sections of original papers or were covered in editorials. These appear to be getting less prominence or even being overlooked in modern reviews, that concentrate on the "statistics".

Utility of "bottom line" summary
Systematic reviews tend to address one question by evaluating a therapy or intervention on one outcome measure. In pharmacological overviews this may be because the agent was designed to achieve one outcome and the trials used a common protocol and focused on this one outcome. In health care evaluation overviews it may be pragmatic to find common ground between inherently heterogeneous trials and select one outcome measure in common. For example, in cardiac rehabilitation, reviews have focused on mortality, not as a prime objective of rehabilitation, but because many of the published trials reported mortality. Within the published literature there is considerable variety in other outcomes reported, including for example, exercise tolerance, quality of life score, health service utilisation rates, medication usage, and further cardiac events and few of these are common throughout the trials.

Another point to consider in the single outcome summary is that it may create an artificial homogeneity of naturally heterogeneous trials, for example in many trials mortality is reported over widely varying times possibly from weeks to years, yet typical statistical overviews simply estimate a weighted average "standardised effect", without inquiring of original investigators whether they might have mortality at a common time, say 12 months. This takes on trust that relative differences in the shorter and longer term are constant and also accepts perhaps naively that original reports did not select one (favourable) mortality outcome to report. This issue was recognised early in the development of meta-analysis and may be partially addressed by poolings of data rather than poolings of trials, but regrettably many secondary researchers do not seek a prior standard common outcome from original investigators and simply accept published reports at face value.

There may be an attractive simplicity in the single "bottom line" approach for the busy end user (practitioner, manager, or planner)., if the merits of a drug, therapy, or service can be fairly summarised in one line. However acceptance of one outcome measure over all others and over possible side effects implies a dominant weighting of the one chosen measure. Recipients of the therapy or service may not accord the same weight as those choosing on their behalf. For example in cancer chemotherapy, one patient may value improved five year survival, another maintaining a singing voice, and another minimal side effects during treatment. A clinician, who advocates one treatment on the basis of survival results alone may not be making the most appropriate choice for all his patients. The Cochrane Collaboration has done much to develop and improve methodology in estimation of the average therapeutic effect but recognises that much work has yet to be done in estimation of rarer but nevertheless important side effects.

An additional consequence of the search for concise single line summaries, whether of truly homogenous trials or of a common outcome of naturally heterogeneous studies, is that secondary researchers prefer to address the more easily answerable questions. The prospect of completing a publishable review/overview in a limited period, say one academic year with a postgraduate student to assist, is greatly enhanced if there is a good supply of similar trials/studies reporting a similar finding. This may arise from the issue covered earlier about post hoc selection of "hypothesis" but it also means that some of the more challenging clinical questions are being left out of reviews.

Dissemination of evidence: bulletins and guidelines
Leading medical journals, that formerly sought (and competed) to publish primary trials or studies often with accompanying editorials, have been quick to give prominence to reviews and overviews. Systematic reviews may appear to include more evidence than a single trial but they do not provide new evidence. While "newness" used to be one of the prime considerations for acceptance of manuscripts in leading journals, it now appears to be "weight of evidence" that counts.

A new abstracting journal, focused specifically on EBM, promised it readers "la creme de la creme", culled from the world's leading medical journals, with each selected article précised neatly into a single page.

This journal clearly linked to the evidence based movement, soon included abstracts of reviews. A review is already a précis of primary trials and as such averages out individual variations. A précis of a précis inevitably looses more detail and by loosing detail risks misreporting the key message. Furthermore citation bias may contribute to more significant studies being abstracted, since papers with positive (significant) findings receive more citations than those without (or with non-significant) findings, in the primary literature. We
have yet to see whether or not this will follow in citations of the secondary research literature but it would come as no surprise, if reviews that report effects receive wider publicity than those that do not.

A number of organisations have been established with the prime purpose of publishing and disseminating reviews; examples in the UK include the NHS Centre for Reviews and Dissemination at Leeds and Bandolier at Oxford.51 The bulletins are distributed widely to practitioners and managers at little or no cost to recipients, as costs are subsidised by the NHS centrally. While some bulletins may be based on reviews published in refereed journals, many are compiled by groups specially convened for the purpose. Some of these new bulletins are explicit in their use of reviews (secondary research) and not the primary trial literature and some are prepared by relatively inexperienced reviewers on unrealistically tight schedules. This allows the possibility of replication of flawed reviews and furthermore, by rephrasing and representing material in bulletins, protects flawed reviews from critical appraisal by their readers. Such replication in bulletins provides a means by which poor reviews may become “respectable” simply through their wide circulation. Readers should be aware of the possibility that evidence in some bulletins may not be so sound and the reviews, on which the bulletins are based, may not be as systematic and thorough as we could reasonably expect of “official” documents. It is not yet clear how the editorial process and rejection rates compare with established journals: limited experience would suggest that they are rather different.

Bulletins from an employer, the NHS, despite legal disclaimers, have an inherent stamp of authority. Although acknowledging that they may remain in date for no more than 18 months, if new evidence comes to light, they imply that they embody the last word in evidence. Some bulletins may be “out of date” before publication. They are so new that we have yet to see how thoroughly they may be revised, when their “sell by” dates arrive. Revisions or corrections, when due, may not be commissioned because the subject has not been “done” and central NHS is more concerned about a new set of issues and diseases. Unlike journals, bulletins allow for no “correspondence pages”, which serve a useful or even essential purpose in scientific advancement. If bulletins distributed by the employer purport to contain the latest evidence, do not use external referees in the same way as established journals and have no provision for readers, including many experienced practitioners and researchers, to question, discuss and debate the “evidence”, it goes unchallenged. This heralds a doctrinaire approach to clinical practice.

Guidelines have evolved in parallel with EBM and bulletins and with a common genesis, that science is now advancing so quickly and new evidence emerging so fast that the textbooks, which served previous generations so well, cannot keep up.52 Guideline writing groups also tend to cite evidence of reviews, although not as exclusively as bulletins. A possible reason is conciseness, a guideline based on reviews might cover 4–6 pages and a dozen references, which is a complete order of magnitude less than would be needed, if based on the original research literature. Guidelines differ from bulletins in a number of important respects. They are mostly written by expert groups, often including representatives of primary trials and mostly under the auspices of professional bodies. Many have been debated and discussed over several years through specialist committees, professional bodies, national and international conferences and most are reported in academic journals, where correspondence allows further discussion and debate. Some, however, may show signs of having been put together with undue haste. Guidelines and bulletins have common features, both tend to be based on reviews, which in turn are summaries of original research, and there remains the possibility that some important details get lost in making a précis of a précis.

Clinical standards, protocols, a new consensus, and the end of research?

A certain idealism and logic underpin the current promulgation of clinical standards and protocols, based on systematic reviews of relevant evidence. In theory these should help clinicians to improve their practice, by providing concise and up to date guidance for discontinuing ineffective practices and treatments and offering their patients only the most effective. If only it were so simple. But as standards and protocols, particularly when distributed by employers, they give more than guidance; they introduce regulation and many practitioners perceive therein a challenge to professional autonomy. Professional self regulation (with guidance to best practice) may be preferable.

To stay with the theme of evidence, this final and perhaps, from the population perspective, logical step in collation, dissemination, and implementation inevitably introduces a conformity of practice. The very objectives of clinical standards and audit of these standards are to minimise variation and reduce deviation, with the ultimate goal of eliminating the ineffective and promoting the effective. Such conformity necessarily impedes innovation and experimentation—or research, which is the basis of the evidence on which reviews, bulletins, guidelines, and standards depend. Is there a risk that in our haste to review, process, and disseminate best evidence to encourage (impose?) best practice, we may squeeze out the very source of evidence for best practice—primary research?

The author thanks Janet Hill for typing the manuscript and Michael Jones for drawing the figures.

R R WEST
University of Wales College of Medicine,
Health Park, Cardiff CF4 4XW, UK

7 Jenner E. An enquiry into the causes and effects of the variola vaccine. London: Law, 1798.


49. West RR, Jones DA. Publication bias in statistical overviews of trials: example of physiological rehabilitation following myocardial infarction. *Heart* 1997;77(suppl 11).
