

SYMPOSIUM ON EVIDENCE BASED MEDICINE

Evidence-based medicine and ethics: a practical approach

P Vineis

J Med Ethics 2004;30:126–130. doi: 10.1136/jme.2003.007211

Correspondence to:
P Vineis, University of
Torino and ISI Foundation,
Torino; Paolo.vineis@
unito.it

Revised version received
21 November 2003
Accepted for publication
27 November 2003

The clinical decision is supposed to be based on evidence. In fact, what counts as evidence is far from being established. Some definition of “proof” is needed to distinguish between scientific medicine and charlatanism. My thesis is that unfortunately a clear-cut boundary between evidence and lack of evidence cannot be found, for several reasons that I summarise in the paper. Evidence in medicine very often has fuzzy boundaries, and dichotomising fuzziness and uncertainty can have serious consequences. Physicians and patients should accept the irreducible fuzziness of many of the concepts they use when dealing with health and disease.

INTRODUCTION

The clinical decision is supposed to be based on evidence. In fact, what counts as evidence is far from being established. According to supporters of the evidence-based medicine movement, the best empirical basis is contributed by randomised controlled trials (RCTs), or at least these contribute the best evidence when they are relevant. On the other side, critiques claim that in many cases RCTs cannot be conducted, and often they are not relevant, so that the standard of proof should be relaxed. It should be clear that some definition of “proof” is needed to distinguish between scientific medicine and charlatanism. Physiopathology—that is, the reference to a mechanism to support the introduction of a new drug—is a criterion that has failed several times in the past: for example, the widespread practice of phlebotomy in eighteenth-century medicine had some “physiopathological” basis but no effectiveness at all. Even if we accept that RCTs are not relevant in some fields, we still need to define what we accept as a proof; this is clearly a problem of transparency of medical practice.

My thesis is that, unfortunately, a clear-cut boundary between evidence and lack of evidence cannot be found, for several reasons:

- for many clinical practices, even if we have well-conducted RCTs, all we can achieve is a “weight of evidence” overall evaluation, because we face conflicting results from RCTs (example 1 in appendix);
- in other instances RCTs are not available simply because they have not been conducted, and we only have access to observational investigations;
- or the quality of the RCTs is poor, so that a meta-analysis is not easily interpretable (examples 2 and 3 in appendix);
- or RCTs cannot be easily conducted for practical or ethical reasons (example 4 in appendix).

Of course, we also have clear instances in which meta-analyses contribute in an unequivocal way to the adoption or banning of a treatment (examples 5–8 in appendix).

In addition, we need to integrate the scientific evidence with the patient’s preferences, with economic constraints, with the healthcare organisation, with ethical obligations ...; this kind of integration is the object of *clinical guidelines*, in which, ideally, evidence is a necessary but insufficient component.

EVIDENCE AND CLINICAL DECISION: INUS AND FUZZY SETS

The model I propose is taken from a different field, causality, and has been suggested by the philosopher John Mackie. Mackie claims that causality cannot be reduced to single necessary and sufficient causes, but rather should be described in terms of elements that he calls INUS (Insufficient Non-redundant component of an Unnecessary Sufficient complex).¹ In his example, Why did the house burn?, the causal complex is formed by the association of fire in the fireplace, a strong wind, a defect in the alarm system, and the fact that the house is wooden. If we analyse each component, none of them is a single sufficient cause, but only their conjunction gives origin to an overall sufficient complex. However the complex is not necessary, because the house could burn in many different ways (for example, because I put it on fire). According to Mackie, although none of the elements is sufficient, at least one is necessary (non-redundant)—that is, in its absence the complex would be ineffective (in the example: eliminating the fire in the fireplace would make the whole complex ineffective). Let us try to apply this same reasoning to medical decision. The physician has to integrate several elements into a decisional complex. Consider, for example, the prescription of ovariectomy in young women (below the age of 50) with a diagnosis of breast cancer. According to the systematic review in the *Cochrane Library*,² there are 12 RCTs on this topic (example 9 in appendix). Most of them showed some advantage associated with ovariectomy in that particular category of patients, but none of them reached statistical significance. However, when a meta-analysis was done, a statistically significant odds ratio (OR) of 0.72 was obtained, indicating a 18% reduction of mortality associated with treatment.

Should the oncologist decide to prescribe ovariectomy, based on this statistical evidence, with a rather weak “mechanistic” (biological) basis? Instead of being an exception, the ovariectomy example is the rule: in very few instances is an RCT (or a meta-analysis) or a biological explanation so strong as to be considered “definitive”. In other words, we have to face a large “grey” area that seems to

Abbreviations: CI, confidence intervals; INUS, Insufficient Non-redundant component of an Unnecessary Sufficient complex; OR, odds ratio; RCT, randomised controlled trial

restrict the expectations of both supporters and detractors of the RCT. The practical oncologist might decide that, based on the Cochrane meta-analysis, the weight of evidence is quite strong because there was good a priori evidence on the hormone sensitivity of breast cancer in young women. Thus, by weighting the empirical evidence coming from trials with the mechanistic background, he or she can decide to prescribe ovariectomy. But one can reason exactly in the opposite way: considering the relatively small advantage (18% reduction in mortality), particularly in women who have received chemotherapy, and the important side effects, including reproductive problems, the weight of the empirical evidence could be reduced.

The clinical decision is a kind of INUS, a complex that includes insufficient components (ethical, economic, scientific ...). It is an unnecessary complex because either the oncologist might come to the same decision based on a different mixture of components or they might come to a different decision; and none of the components, clearly, is sufficient. Let us expand the latter concept: the tendency to overview evidence in favour of the patient's choice (which becomes an isolated "sufficient" cause for decision) is at the roots of recent episodes of malpractice, for example the "Di Bella" case in Italy; or, using only evidence as the sole criterion can cause irrational choices, for example in terms of resource allocation. However, it is important to perceive correctly one feature of Mackie's definition of INUS—that is, that at least one component is necessary (non-redundant); I believe this component is evidence: without evidence there will never be good clinical decision. Of the other criteria, one might argue that none is really necessary—for example, the respect for the patient's decisional autonomy does not apply when they are in a coma (in fact paternalism is justified when it is aimed at reintegrating the patient's decisional autonomy).

If we accept that evidence is a necessary component, still how to weight the evidence depends on the definition of effectiveness we adopt: according to an *objective* definition we will give little weight to the patient's preferences and much weight to the reduction in mortality plus mechanistic evidence. My point of view is that effectiveness, like disease, is a "family resemblance"—that is, a "fuzzy" concept. Concepts are almost never sharp (defined on the basis of a single (monothetic) property), but they are "polythetic"—that is, they are like a "long rope twisted together out of many shorter fibres".³ In particular, the concept of effectiveness cannot be defined on the basis of a singular property (reducing mortality), but of several properties that are partially overlapping in the actual instances: for some people effectiveness is mainly subjective, for others it is mainly objective, and no single definition is the right one. In summary, we have to face that effectiveness is a "fuzzy" concept. This means that we cannot use the results of trials (meta-analyses) as the only source of information and decisions about care: the work of the physician consists just in integrating different kinds of knowledge, although evidence is a necessary component.

In general, my viewpoint is that medicine is made up of several different "fuzzy sets". In fact, a clear-cut dichotomy between what should be included in the field of medicine and what should not is impossible: on one side we have clearly organic conditions, such as tuberculosis or smallpox and on the other side we have "behavioural" patterns for which medicine is not competent, but the borders between the two categories are obviously uncertain and fuzzy. Fuzziness applies to both disease definition and to the identification of causal patterns (table 1). While for smallpox we have both a clear-cut definition of the disease and a clear understanding of the causal agent, in the case of diseases like schizophrenia,

bulimia, or anorexia, both disease and causal agents have blurred boundaries. The disease cannot be easily distinguished from other similar symptomatological constellations (for example, bulimia from other conditions characterised by obesity and "binge eating"), and the causal complexes are rather ill-defined and vague. In other words, for such diseases we shift from the classic scientific paradigm of "explanation" to a more evasive and slippery paradigm as those used by psychosocial sciences. There is a continuum between the categories shown in table 1, and diseases like smallpox are only one extreme of the spectrum, the other extreme being represented, for example, by several psychic disorders.

EXAMPLES

The appendix shows summaries from the *Cochrane Library*.² I have chosen nine examples that can be considered typical of a few categories that I have mentioned above. To be clear, not all the troubles in medical decision arise from the intrinsic "fuzziness" of such discipline. Some of the examples below show that fuzziness is a problem indeed, while others (like example 3) just show that good quality data are not available. Effectiveness can be defined in a very clear way (think of orthopaedic interventions), but nevertheless the available information can be unsatisfactory.

In the first example, the evidence is rather sparse (only 453 patients) and the results are conflicting (fig 1), with a non-statistically significant trial showing protection and one significant trial showing an excess of deaths in the corticosteroid arm. The second example, treatment of giardiasis, is paradigmatic of the lack of good trials, at least in some fields; in this case 34 trials were identified, but only one was methodologically acceptable. The third example is more complex, since the information available was not enough to evaluate the efficacy of the treatment. If all missing data (dropouts) are attributed to disease progression (worst case scenario) then treatment is associated with a slight adverse effect. Lack of data of good quality is the main problem in this example. In the fourth case, the subject itself is difficult (giving information to children and adolescents on their cancers) and ethically sensible. One might ask: Is it ethical to start a randomised trial that implies that one arm does not receive information or receives information that is considered a priori to be worse than for the other arm? Is the randomised trial an adequate tool for this research subject? What is the best way to ascertain effectiveness? It is not surprising that trials are extremely heterogeneous in this example. One wonders, however, whether heterogeneity could be really overcome, or it is not inherent in the subject.

The fifth, sixth and seventh examples show how useful meta-analyses can be. In all three cases, individual trials were equivocal, but the overall consideration of their results showed (i) that postoperative radiotherapy causes damage

Table 1 Classification of diseases and their aetiological agents, with examples

		Disease	
		Monothetic definition*	Polythetic definition*
Agent (s)	Monothetic definition	Smallpox (group 1)	Tuberculosis (group 2)
	Polythetic definition	Lung cancer (group 3)	Bulimia and anorexia (group 4)

*On the basis of clinical signs and symptoms only.

Review: Corticosteroids for acute ischaemic stroke
 Comparison: 01 corticosteroids vs. placebo
 Outcome: 01 All deaths

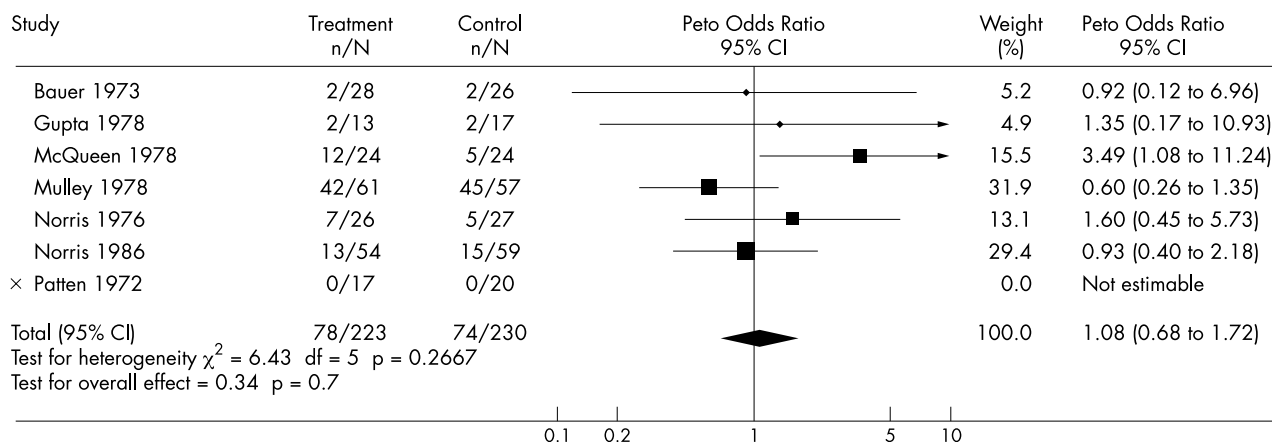


Figure 1 Meta-analysis of corticosteroids versus placebo in acute ischaemic stroke. © Cochrane Library, reproduced with permission. Source: Qizilbash N, Lewington SL, Lopez-Arrieta JM. Corticosteroids for acute ischaemic stroke (Cochrane Review, most recent amendment February 2002). In: *The cochrane Library*, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd.

to patients, (ii) that aminophylline only causes side effects in patients with asthma treated with β -agonists, and (iii) that anticoagulants do more harm than benefit in acute ischaemic stroke. On the other hand, the eighth example (warfarin in atrial fibrillation) is paradigmatic of a situation in which a meta-analysis clearly reveals—more than single trials—that the benefits are considerable and the treatment should be transferred into practice. Finally, the ninth example is commented upon in the previous section.

CONCLUSIONS

Dichotomising fuzziness and uncertainty can have serious consequences. Let us imagine that 54% of the voters in a referendum are in favour and 46% are against. Let us imagine that those who voted “yes” are only partially convinced of that choice (for those who know Italian politics this is clearly a realistic assumption)—for example, they have a 75% propensity to the “yes” and 25% propensity to the “no”. If we weight the vote for the degree of uncertainty (abandoning a dichotomous approach) then we have that, in fact, 50.5% were for the “no” and 49.5% for a positive vote, a result that is opposite to the one based on an unweighted dichotomy. The example shows that what counts is not only the distribution of the voters in the two categories, but also the “degree of overlapping” of the categories themselves (a concept typical of the “fuzzy” logic—that is, a function of uncertainty). I believe that both physicians and patients should accept the irreducible fuzziness of many of the concepts they use when dealing with health and disease.

APPENDIX: SUMMARIES FROM THE COCHRANE LIBRARY¹

1. Corticosteroids in ischaemic stroke

Seven trials involving 453 people were included. Details of trial quality that may relate to bias were not available for most trials. No difference was shown in the odds of death within one year (OR 1.08, 95% confidence interval (CI) 0.68 to 1.72). Treatment did not appear to improve functional outcome in survivors. Six trials reported neurological impairment but pooling of data was impossible because no common scale or time interval was used. The results were inconsistent between individual trials (see fig 1). The only adverse effects reported were small numbers of

gastrointestinal bleeds, infections, and deterioration of hyperglycaemia across both groups.

2. Treatment of giardiasis

Of the 34 trials that were included, only one trial was without serious methodological flaws. Compared with placebo, drug treatment was associated with an improved cure rate (OR 11.5, 95% CI 2.3 to 58). Metronidazole treatment longer than three days had a better parasitological cure rate than other long treatment courses (OR 2.4, 95% CI 1.3 to 4.4), but there was significant heterogeneity between the trials. Available evidence has not detected a difference in cure between single dose therapy and longer treatment courses (OR 0.33, 95% CI 0.08 to 1.34). Within the single dose regimens, the available evidence did not demonstrate a difference in parasitological cure rate between tinidazole and other short therapies (OR 3.4, 95% CI 0.95 to 12), but had a higher clinical cure rate (OR 5.3, 95% CI 2.7 to 10.7).

Suggested reading

On EBM

Evidence-based Medicine Working Group. Evidence based medicine: a new approach to the teaching of medicine. *JAMA* 1992;**268**:2420–5.

Sackett DL. Clinical epidemiology. What, who, and whither. *J Clin Epidemiol* 2002;**55**:1161–6.

Vandenbroucke IP. Evidence based medicine and “Medicine d’Observation”. *J Clin Epidemiol* 1996;**49**:1335–8.

On “fuzziness” of medicine

Vineis P. Definition and classification of cancer: monothetic or polythetic? *Theor Med* 1993;**14**:249–56.

On philosophical problems of medical diagnosis and therapy

Murphy EA. *The logic of medicine*, 2nd ed. Baltimore: Johns Hopkins University Press, 1997.

On effectiveness

Ashcroft R. What is clinical effectiveness? *Stud Hist Philos Biol Biomed Sci* 2002;**33**:219–33.

3. Interferon and multiple sclerosis

Although 1215 patients were included in this review, only 919 (76%) contributed to the results concerning exacerbations and progression of the disease at two years. Specifically, interferon significantly reduced the occurrence of exacerbations (relative risk (RR) 0.80, 95% CI 0.73 to 0.88, $p < 0.001$) and progression of the disease (RR 0.69, 95% CI 0.55 to 0.87, $p = 0.002$) two years after randomisation. However, the correct assignment of dropouts was essential to the demonstration of efficacy, most conspicuously concerning the effect of the drug on disease progression. If interferon-treated patients who dropped out were deemed to have progressed (worst case scenario) the significance of these effects was lost (RR 1.31, CI 0.60 to 2.89, $p = 0.5$). The evolution in magnetic resonance imaging (MRI) technology in the decade in which these trials were performed and different reporting of data among trials made it impossible to carry out a quantitative analysis of the MRI results. Both clinical and laboratory side effects reported in the trials were more frequent in the treated patients than in the controls. No information was available regarding side effects and adverse events after two years of follow up. The impact of interferon treatment (and its side effects) on the quality of life of patients was not reported in any trial included in this review. The reviewers' conclusions were: the efficacy of interferon on exacerbations and disease progression in patients with relapsing remitting MS was modest after one and two years of treatment. It was not possible to conduct a quantitative analysis beyond two years. Longer follow up and more uniform reporting of clinical and MRI outcomes among these trials might have allowed for a more convincing conclusion.

4. Communicating with children and adolescents about their cancer

Six studies met the criteria for inclusion. They were diverse in terms of the interventions evaluated, study designs used, types of people who participated and the outcomes measured. One study of a computer-assisted education programme reported improvements in knowledge and understanding about blood counts and cancer symptoms. Both studies of school reintegration programmes reported improvements in some aspects of psychosocial wellbeing (one in anxiety and one in depression), social wellbeing (two in social competence and one in social support) and behavioural problems; and one reported improvements in physical competence. The reviewers' conclusions were: interventions to enhance communication involving children and adolescents with cancer have not been widely or rigorously assessed. The weak evidence that exists suggests that some children and adolescents with cancer may derive some benefit from specific information-giving programmes and from interventions that aim to facilitate their reintegration into school and social activities. More research is needed to investigate the effects of these and other related interventions.

5. Postoperative radiotherapy (PORT) in non-small cell lung cancer

There were nine trials with 2128 patients (median follow up 3.9 years). The results show a significant adverse effect of PORT on survival with a hazard ratio of 1.21 or 21% relative increase in the risk of death. This is equivalent to an absolute detriment of 7% at two years (95% CI 3% to 11%) reducing overall survival from 55% to 48%. Exploratory subgroup analyses suggested that this detrimental effect was most pronounced for patients with stage I/II, N0–N1 disease, whereas for stage III, N2 patients there was no clear evidence of an adverse effect.

6. Aminophylline in acute asthma

Fifteen studies were included. Overall, the quality of the studies was only moderate; concealment of allocation was assessed as clearly adequate in only seven (45%) of the trials.

The doses of aminophylline and other medications and the severity of asthma varied between the studies. There was no statistically significant effect of aminophylline on airflow outcomes at any time period. The aminophylline treated group had higher values of peak expiratory flow rate (PEFR) at 12 (PEFR 8 l/min or 2.3%) and 24 hours (PEFR 22 l/min or 6.4%), but these were not significant ($p > 0.05$). Two subgroup analyses were performed by grouping studies according to mean baseline airflow limitation ($n = 11$ studies) and the use of any steroids ($n = 9$ studies). There was no relation between baseline airflow limitation nor the use of steroids on the effect of aminophylline. Aminophylline treated patients reported more palpitations/arrhythmias (OR 2.9, 95% CI 1.5 to 5.7) and vomiting (OR 4.2, 95% CI 2.4 to 7.4), but no difference was found in tremor or hospital admissions. The reviewers' conclusions were: in acute asthma, the use of intravenous aminophylline did not result in any additional bronchodilation compared with standard care with β -agonists. The frequency of adverse effects was higher with aminophylline. No subgroups in which aminophylline might be more effective could be identified. These results should be added to consensus statements and guidelines.

7. Anticoagulants in ischaemic stroke

Twenty one trials involving 23 427 patients were included. The quality of the trials varied considerably. The anticoagulants tested were standard unfractionated heparin, low molecularweight heparins, heparinoids, oral anticoagulants, and thrombin inhibitors. Based on eight trials (22 450 patients) there was no evidence that anticoagulant therapy reduced the odds of death from all causes (OR 1.05, 95% CI 0.98 to 1.12). Similarly, based on five trials (21 846 patients), there was no evidence that anticoagulants reduced the odds of being dead or dependent at the end of follow up (OR 0.99, 95% CI 0.94 to 1.05). Although anticoagulant therapy was associated with about nine fewer recurrent ischaemic strokes per 1000 patients treated, it was also associated with a similar sized (nine per 1000) increase in symptomatic intracranial haemorrhages. Similarly, anticoagulants avoided about four pulmonary emboli per 1000, but this benefit was offset by an extra nine major extracranial haemorrhages per 1000. Sensitivity analyses did not identify a particular type of anticoagulant regimen or patient characteristic associated with net benefit.

8. Warfarin in patients with atrial fibrillation

Fourteen articles were included in this review. Warfarin was more efficacious than placebo for primary stroke prevention (aggregate OR of stroke 0.30, 95% CI 0.19 to 0.48), with moderate evidence of more major bleeding (OR 1.90, 95% CI 0.89 to 4.04). Aspirin was inconclusively more efficacious than placebo for stroke prevention (OR 0.68, 95% CI 0.29 to 1.57), with inconclusive evidence regarding more major bleeds (OR 0.81, 95% CI 0.37 to 1.78). For primary prevention, assuming a baseline risk of 45 strokes per 1000 patient years, warfarin could prevent 30 strokes at the expense of only six additional major bleeds. Aspirin could prevent 17 strokes, without increasing major haemorrhage. In direct comparison, there was moderate evidence for fewer strokes among patients on warfarin than on aspirin (aggregate OR 0.64, 95% CI 0.43 to 0.96), with only suggestive evidence for more major haemorrhage (OR 1.58, 95% CI 0.76 to 3.27). However, in younger patients, with a mean age of 65 years, the absolute reduction in stroke rate with warfarin compared with aspirin was low (5.5 per 1000 person years) compared with an older group (15 per 1000 person years). Low dose warfarin or low dose warfarin with aspirin was less efficacious for stroke prevention than adjusted dose warfarin. The reviewers' conclusions were: evidence strongly supports warfarin in atrial fibrillation for

patients at average or greater risk of stroke, although clearly there is a risk of haemorrhage. Although not definitively supported by the evidence, aspirin may prove to be useful for stroke prevention in subgroups with a low risk of stroke, with less risk of haemorrhage than with warfarin. Further studies are needed of low molecular weight heparin and aspirin in lower risk patients.

9. Ovariectomy in breast cancer

Among 2102 women aged under 50 when randomised, most of whom would have been premenopausal at diagnosis, 1130 deaths and an additional 153 recurrences were reported. The 15 year survival was highly significantly improved among those allocated ovarian ablation (52.4 v 46.1%, 6.3 (SD 2.3) fewer deaths per 100 women, log rank $2p = 0.001$), as was recurrence free survival (45.0 v 39.0%, $2p = 0.0007$). The numbers of events was too small for any subgroup analyses to be reliable. The benefit was, however, significant both for those with ("node positive") and for those without ("node

negative") axillary spread when diagnosed. In the trials of ablation plus cytotoxic chemotherapy versus the same chemotherapy alone, the benefit appeared smaller (even for women with oestrogen receptors detected on the primary tumour) than in the trials in the absence of chemotherapy (where the observed survival improvements were about six per 100 node negative women and 12 per 100 node positive women). Among 1354 women aged 50 or over when randomised, most of whom would have been perimenopausal or postmenopausal, there was only a non-significant improvement in survival and recurrence free survival.

REFERENCES

- 1 **Mackie JL.** *Logic and Knowledge: Selected Papers of JL Mackie*, Vol 1. Oxford: Oxford University Press, 1995.
- 2 *Cochrane Library*, Issue 2. Oxford: Update Software, 2002; www.update-software.com/Cochrane (accessed March 2003).
- 3 **Wittgenstein L.** *Philosophical Investigations*. New York: Macmillan, 1953.