

Evidence in medicine: invited commentary

Olli S. Miettinen, MD, PhD

In the practice of medicine, evidence pertains to “gnosis” — diagnosis, etiognosis and prognosis. Specific, ad hoc evidence is constituted by the patient profile in known, gnosis-relevant respects, and this is interpreted — translated to gnostic probability — in the light of general, ideally scientific, evidence specific to the profile. The scientific evidence is about rates in defined domains and, ultimately, their profile-defining subdomains. Scientific evidence expressly relevant to diagnosis and etiognosis is almost nonexistent today, and intervention-prognostic research, while already appreciable in quantity, tends to lack relevance to practice mainly on account of relative inattention to the pivotal element in all practice-oriented research — express, principles-guided and detailed design of the object of study with a view to its relevance for practice, this in lieu of just an aprioristic and loose definition of it. Reporting of intervention-prognostic studies and syntheses of their results, too, remain underdeveloped. The current deficiencies in the scientific evidence for, and consequently in the scientific knowledge-base of, the practice of medicine constitute a challenge that clinical academia must, and increasingly will, meet — in partnership with editors of medical journals.

It speaks to the enlightenment of the editors of the *Canadian Medical Association Journal* that they, in their letter to me, have expressed concern about the prevailing understanding of the issues surrounding scientific evidence as an input to the practice of medicine. These issues are important beyond any question, richly deserving of a textbook. But as none yet exists, the editors asked for an orientational overview as a succedaneum, for now. I feel honoured by their directing such a request to me.

Specific versus general evidence

In the legal profession, conclusions are made about guilt (criminal) or liability (civil), and they are based on evidence *specific* to the case at issue. Each conclusion from such evidence rests on a judgement of probability. This judgement is made informally — intuitively, subjectively — and not on the basis of general evidence from “legal science” addressing the frequency (probability) of guilt or liability as it depends on the particulars of the evidence. The conclusion itself is implicit in the level of probability; it is of the categorical, yes–no type; and it, in and of itself, implies the decision about whether some punitive or compensatory action is to ensue.

The legal profession’s outlook meets medicine most notably in workers’ compensation decisions: if, in informal medical opinion, it is more probable than not that work was causal to the worker’s illness, then work was “liable” for the worker’s illness and full compensation is due, whereas otherwise no compensation is accorded. (An alternative worthy of consideration would be that if the probability of occupational etiology is P , then proportion P of the full compensation is allowed.)

In medical diagnosis, the ultimate question is whether a particular illness is at the root of the patient’s felt problems; and were the legal paradigm to be followed, the evidence would be taken to consist solely of “specific” evidence: “direct” evidence in the sense of symptoms, signs and test results, together with “circumstantial” evidence in the sense of the patient’s profile in respect to indicators of risk (cf. Fig. 1). Both the manifestational (direct) and risk (circumstantial) aspects of the specific evidence are relevant to the ultimate question only insofar as they are discriminating between the illness at issue and its differential-diagnostic alternatives as the possible “culprits” for the production of the patient’s predicament. This differential-diagnostic outlook is different from the legal one: in a court of law there is no line-up of all the possible culprits coupled with the challenge of evidence-based identification of the true culprit among them.

In the context of medical diagnosis, different from law, evidence takes on an added



Editorial

Éditorial

Dr. Miettinen is with the Departments of Epidemiology and Biostatistics and of Medicine, McGill University, Montreal, Que.

Can Med Assoc J 1998;158:215-21



meaning. When the diagnostician holds that in this patient at this time the probability of the presence of illness I is P , this is not supposed to represent an intuitive, subjective judgement. Instead, it is supposed to reflect *professional knowledge*: presumably, the physician knows that among instances like this (as for the patient profile) in general, the prevalence of I as the underlying illness is about P , and hence the probability P for its presence in this particular instance. And whence this knowledge? In modern medicine, ideally from medical (diagnostic) research, from *general evidence* in this sense (cf. Fig. 1). In this framework, then, for the inference that illness I is present with probability P , the evidence in one meaning consists of the available facts constituting the diagnostic profile for the case, and in another sense it consists of the evidence for the prevalence rate used as the diagnostic probability associated with the profile; specific (ad hoc) evidence is coupled with general evidence to arrive at the ad hoc judgement (Fig. 1).

In inferring that with probability P antecedent A was causal to this case of I (diagnosed with high probability), the evidence in the first sense — the specific evidence — consists of the fact that A had occurred together with the patient's profile bearing on the rate ratio expressing how-many-fold that history makes, causally, the current incidence of I — a matter principally of age and gender, say, together with history in respect to extraneous causes (Fig. 2). If this rate ratio is taken to have the value RR, then the corresponding etiognostic probability should be taken to be $(RR - 1)/RR$, expressing the "etiologic fraction," the proportion of cases of I with antecedent A and the rest of the profile that are caused by A. Thus, evidence in the second

sense — the general evidence — has to do with the magnitude of the causal RR, specific to the patient's etiognostic profile. Here, as always, the degree of ad hoc relevance of the general evidence depends on the degree to which it is specific to the type of situation that is at issue (cf. Fig. 2).

In respect to prognostic probabilities, the focus tends to be on the probability that a given intervention, if adopted in lieu of a particular alternative, would prevent (or cause) a particular health event or state in a particular period of prospective (prognostic) time. For setting such probabilities, the specific evidence generally consists of the patient profile in the sense of characteristics bearing on "background" risk, together with ones having more specifically to do with the patient's propensity to respond to the interventions. In epidemiologic jargon, at issue here, as in etiognosis, are patient characteristics in terms of "modifiers" of the magnitude of the effect parameter, here the risk difference (causal) between the 2 interventions (Fig. 3). For example, as for the risk of nonpatency of the occluded coronary artery in acute myocardial infarction when contrasting thrombolysis with "primary" angioplasty, time lag since the onset of symptoms is a major modifier of the risk difference. The general evidence, in turn, has to do with the magnitude of the intervention's effect, in terms of rate difference, specific to the patient profile. It derives from documented comparative experience, experimental perhaps, with the interventions in previous patients representing the same category — domain — as for indication and (or) freedom from contraindications; and to achieve the relevant specificity, the documentation addresses the way in which the magnitude

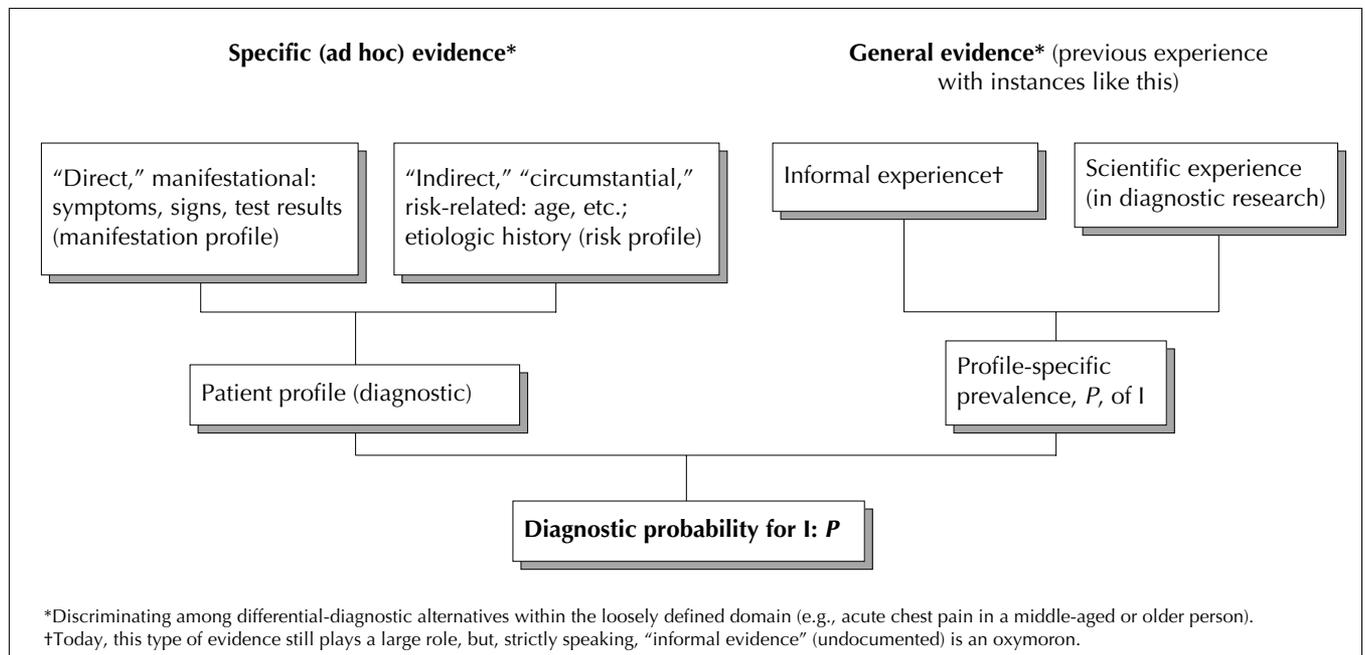


Fig. 1: Evidence for setting the diagnostic probability (P) for a particular illness (I).



of effect varies among subdomains of interest (cf. Fig. 3).

Thus, in medical inference (gnostic) in general, specific (ad hoc) evidence is coupled with suitably *specific general* evidence in arriving at the ad hoc probability. Somewhat more to the point, the ad hoc evidence is coupled with the practitioner's *belief* in respect to a general issue, with this belief ideally based on suitable *scientific* evidence.

These probabilities are not translated to conclusions of the yes–no type: gnosis is expressly *probabilistic* throughout, quantitative in respect to the probability. For it is in the essence of medicine that it is about the hidden, that it addresses unanswerable questions (gnostic, of the yes–no type) and that actions must be taken without express knowledge about the relevant truths about the case. Thus, one generally cannot — and, different from the spirit of law, one need not — “conclude” that illness I is present or not present, that antecedent A which was present caused it or did not cause it, or that a particular intervention (if used in lieu of a particular alternative) would or would not have a particular effect. Rather, one must merely infer the probability of this. What action one takes should be guided but not dictated by this probability. In particular, insofar as there are to be *guidelines* for practice, they should not be ones of decision but should, instead, define gnostic probabilities and their communication to the patient (“doctor” means teacher) — with maximal role for the patient in the decision-making about consequent actions (as for diagnostic testing, termination of etiologically suspect current medication use or whatever, or the choice of a prospective intervention).

Essence of scientific evidence

In the absence of generally accepted guidelines defining gnostic probabilities within the specialty and its practice environment, the practitioner needs to access the general — scientific — evidence at its source(s).

In its simplest form, the available scientific evidence consists of the published report of a single piece of original research. In it, the evidence is not the authors’ “conclusion,” and especially not when the authors “conclude,” as they commonly do, that in their study “there was no evidence that . . .” Underpinning this pseudo-conclusion — mere characterization of evidence — usually is nothing but a lack of “statistically significant” difference in the data, potentially explicable by paucity of subjects in the study. In the same vein, a clinician, too, is prone to say that “there is no evidence that . . .” when attempting to support some nihilistic presumption. Perhaps the very first and most commonly unlearned lesson about general evidence in medicine is that *absence of evidence is not evidence of absence* (of a differential effect, say).

The second thing to learn about scientific evidence in respect to a gnostic probability might well be the nature of study results pertaining to the probability. In its most elementary form a meaningful result is constituted by the empirical value of the parameter of gnostic interest — *empirical rate* per se (diagnostic prevalence rate, or descriptive–prognostic rate of incidence or prevalence), *empirical rate ratio* (etiologic) or *empirical rate difference* (interven-

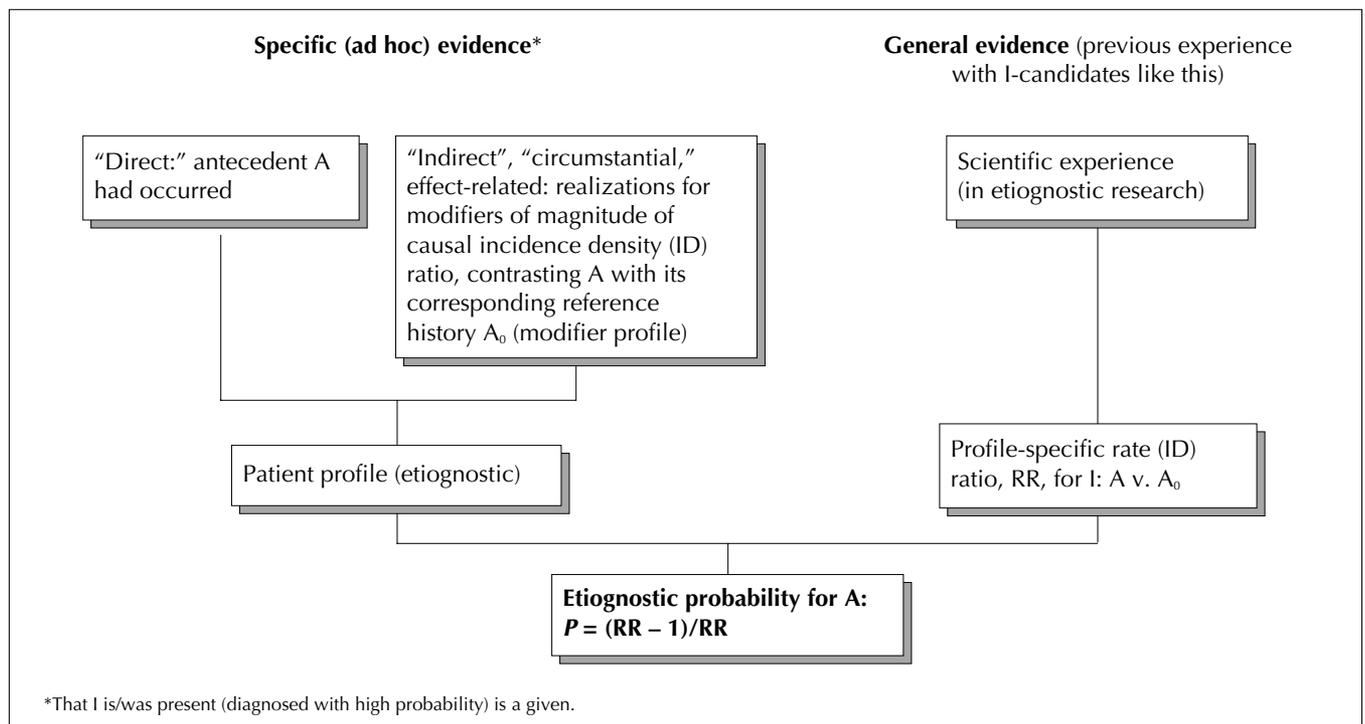


Fig. 2: Evidence for setting the etiologic probability (P) for a particular antecedent (A) as a cause of a case of illness (I).



tion–prognostic). How much bearing these quantitative descriptors of the data have on the magnitude of the probability of gnostic interest is expressed not by the associated *P* value but the associated measure of *imprecision*, either a standard error (implying confidence intervals) or a confidence interval per se. Insofar as “no evidence” has any true meaning, it is that of great imprecision, a very wide confidence interval; and insofar as evidence is taken to exist, and any worthwhile report does bear evidence, the evidence is *against* values inconsistent with the empirical value — usually taken to be values outside a 95% confidence interval. This lesson, too, remains often unlearned. Thus, in the report on a recent study on the relative effects of thrombolysis and primary angioplasty in acute myocardial infarction¹ and its associated commentaries,²⁻⁵ empirical differences of the sort that mortality rates were lower under thrombolysis got to be repeatedly confused with “reduction” in mortality (a matter of inference rather than evidence); the inherently theoretical concept of odds ratio was invoked in lieu of rate difference and confused with its empirical value; and *P* values were abundant in place of measures of imprecision for empirical differences in rates.

In the framework of an appropriate form for the object of study, the study result must address the way in which the magnitude of the empirical measure varies according to specifics in respect to various patient profiles within the study domain. In intervention research this is a matter of addressing differences in rate differences, not a matter of “subgroup analyses” specific to the subdomains and especially not of analysis specific to just one of the subdomains — as

was done in the recent study.¹ In that study, overall results were supplemented by those for a “high-risk” subdomain. It is illustrative to note what those “subgroup” results were: for mortality during hospitalization, just as for “long-term” (3-year) mortality, “there was no significant difference”; and for the former, the particulars that were reported (parenthetically) were these: “8.1 percent in the thrombolytic-therapy group vs. 8.7 percent in the primary-angioplasty group, *P* = 0.70.” Reviewers took this to be evidence that no high-risk patients “allegedly benefit from primary angioplasty.”⁴ Yet, assuming that the *P* value was derived from a 2-sided test (the Statistical Analysis and Results sections both failed to specify whether it was), the empirical rate difference (−0.6%) is associated with a 95% confidence interval (2-sided) the upper bound of which is 2.5% and favours angioplasty. Particularly notable is the fact that neither the study at issue¹ nor the associated commentaries²⁻⁵ paid any attention to how the relative effect depends on the time lag since the onset of symptoms.

The result, even if its form accords with that of a meaningful object, is only part of the evidence from a study. In and of itself it is, indeed, no evidence at all. It gains its meaning, if any, in the light of its documented way of having come about. The first element in this is the documented study plan, the study *protocol*, outlined in the Methods section of the study report. But this, in turn, is of little evidentiary meaning without documentation of the degree of success in the *execution* of the protocol as intended, a segment of evidence that also belongs in the Methods section of the report. Both of these bear on the *validity* of the study,

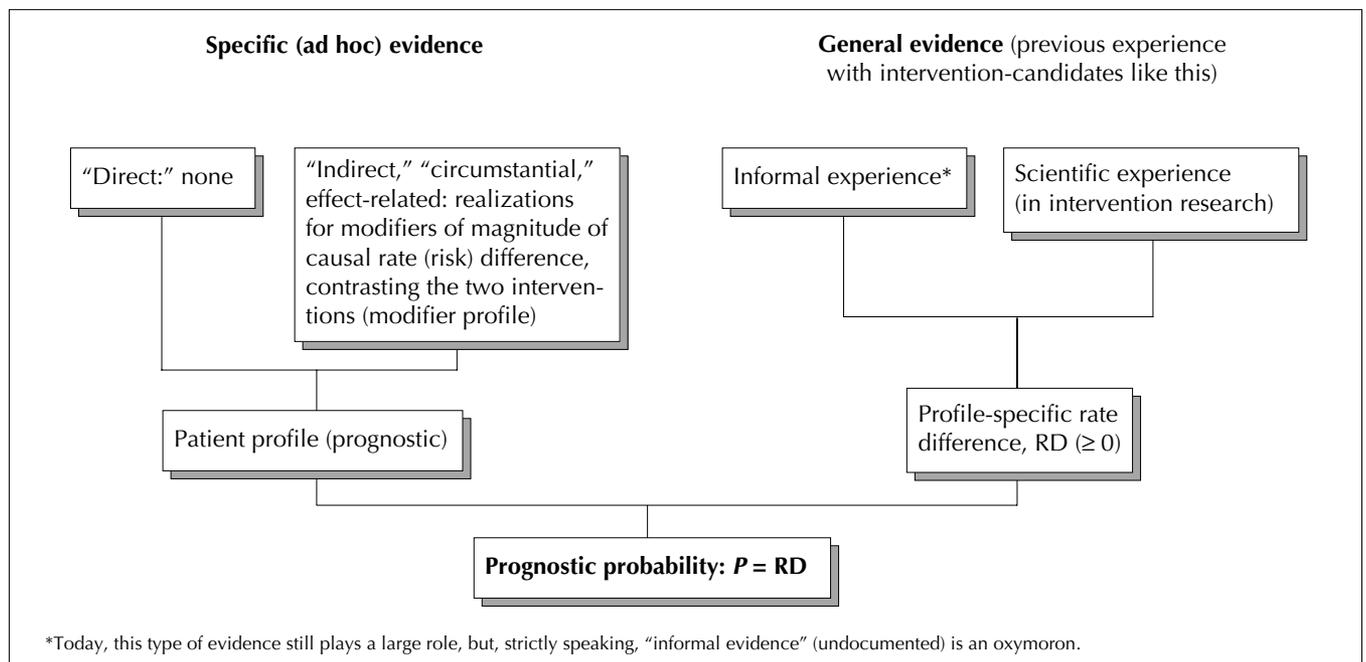


Fig. 3: Evidence for setting the prognostic probability (*P*) for a given intervention, if used in lieu of a particular alternative, preventing (or causing) a particular event or state in a particular period of prospective (prognostic) time.



the extent to which its empirical value(s) for the parameter(s) of interest is (are) free of bias; and if the validity is too wanting, the study report constitutes no evidence at all.

Even this aggregate of evidence is, however, for naught so long as the *object* of study remains misconstrued, as commonly it is. Consider, again, intervention research on the indication of acute myocardial infarction. The practitioner's concern is not merely the question of how to begin the intervention, notably as for the choice between thrombolysis and angioplasty, but also how to deal with its failure and complications as well as in-hospital reinfarction, what maintenance therapy to institute at hospital discharge and how to deal with subsequent coronary problems, reinfarction(s) included. In other words, the concern is with relative effects of candidates for the *algorithm* of choice, defined for a *sufficiently long period* of time — for the rest of time, really, in coronary artery disease. In particular, insofar as the evidence has to do with, say, 3-year survival after acute myocardial infarction, relating it to the initial choice between thrombolysis and angioplasty — as has been done^{1-5, i.a.} — rather than 3-year algorithms of intervention is empty of meaning. It is really striking to me how little attention is being given to express, principles-guided and detailed *design of the object* of study with a view to its conceptual meaningfulness and, ultimately, its *relevance* for the true concerns (gnostic) of practice. Indeed, I am unaware of any medical journal that, as a matter of editorial policy, expects suitable specification — and tenable justification — of the object of study in the report on a gnosic-oriented study. (The “objective” of whatever gnosic-oriented study is to provide evidence on its object.)

When more than one study has been conducted on the topic — the object of study, loosely defined — the very first challenge in respect to the evidence is to get hold of *all* of it, unpublished as well as published. For, as is well known, studies do not get to be published just because they are taken to provide valid evidence on a meaningfully conceptualized object of study but also because the result is “interesting” or, better yet, “provocative,” so that published evidence gets to be a biased subsegment of the entirety of evidence that has accrued. It is for this reason that, notably, the US drug-regulatory agency (the Food and Drug Administration), so different from medical journals of whatever degree of prestige, considers evidence only from studies that have been *preregistered* with it; it considers all such evidence, not just a select subsegment; and its efforts of quality assurance are exercised in the preregistration phase already.

With all of the studies identified, the next challenge is to narrow down to the truly meaningful ones among them, peer review being but an incomplete means of quality assurance in its limited domain of published studies; and finally the challenge is to *synthesize* the evidence from these.

The by-now-familiar term “meta-analysis” refers to formal — statistical — synthesis of the results from all valid studies on a defined object. Its “meta-result” (empirical value[s] for the parameter[s] of interest, together with measure[s] of imprecision) is, again, only part of the evidence: it can be given a meaningful interpretation only in the context of documentation of how the initial, full set of studies was identified, how the actually-synthesized studies among them were selected and how the synthesis itself was done.

Frustrations with scientific evidence

For today's practitioner of medicine wishing to make use of scientific evidence in setting gnostic probabilities, frustration abounds. Much of this has to do with the research itself in a very fundamental way. As for diagnosis, research still focuses on “sensitivity” (gross pseudonym for true positive rate) and “specificity” (gross pseudonym for true negative rate) for often pseudo-dichotomous tests or symptoms or signs, considered in isolation, rather than on the prevalence of the illness at issue as a joint descriptive function of an appropriate set of diagnostic indicators in a defined domain of presentation. Etiologic research is still quite exclusively motivated by the concern to identify opportunities for prevention rather than the needs of etiognosis, and for this reason it tends to lack the specificity that etiognosis requires as for the etiologic history and for the patient profile bearing on the ad hoc magnitude of the causal rate ratio; and besides, its understanding is still held back by commitment to the malformed methodologic concepts of “cohort,” “case-control” and “cross-sectional” study. As for the principal modality of intervention — pharmaco-intervention — the studies that provided the evidence for approval by the regulatory agency did not contrast the intervention algorithms among which a choice is to be made in practice, to say nothing about their failure to account for characteristics that bear on the likely magnitudes of their relative effects. Yet, on many practice-relevant topics of intervention there are a multitude of studies. Examples include such common concerns as the safety of NSAIDs in respect to gastrointestinal bleeding and the effects, both intended and unintended, of postmenopausal use of estrogen products. In respect to intervention research, much recent confusion has been brought about by the “outcomes movement,”⁶ notably its favourite yet highly malformed concepts of “effectiveness” and “generalizability,” and the methodologic aberrations that flow from this.

As for the synthesis of available scientific evidence of a meaningful form, a different type of frustration confronts the practitioner. Does he or she possess the competence for it? Does he or she have the time for this? Broadly speaking, no and no. Nor is it reasonable as a matter of efficiency that each practitioner separately engage in this. Someone com-



petent needs to take the time to do the synthesis for the colleagues concerned with the issue. But for this to be satisfactory, the practitioner needs to have faith in the synthesizer's competence, and diligence besides. As for competence, the consumer of a synthesis needs to judge whether the synthesizer sufficiently masters, for one, the relevant medical aspects of the topic and, for another, the principles of object formulation and validity assurance for research on it and, finally, the principles of synthesizing the results of valid studies on a properly formulated object of study. For by no means are different synthesizers of the original evidence on a given topic prone to produce essentially the same "meta-result," just as a patient may be subject to different diagnoses, etc., from different physicians — so long as detailed norms do not stipulate and guide the respective processes, as remains the case today. And even if the result be reproducible, it would not follow that it is tenable.

The art of synthesizing medical evidence, properly construed, is quite demanding and, consequently, judgement about its quality is frustratingly difficult for someone whose expertise is in patient care itself rather than synthesis of evidence for it. In fact, trusting a "meta-analyst" is more difficult for a practitioner than trusting a practitioner is for a patient. For, while the practitioner is licensed and board-certified to be trustworthy, a "meta-analyst" is generally self-appointed, even if the work is subjected to critical review. As for the Cochrane Reviews⁷ in particular, the reviewer groups are not constituted by persons who have laboured long and hard to master the understanding of the proper nature, valid production and valid synthesizing of scientific evidence in medicine at large and then in a particular topic area specifically. Instead, "The formation of a collaborative review group starts with individuals with a common interest in a particular health problem or group of health problems who come together to prepare systematic reviews on their topic."⁷ Not a word is there about any requisite qualifications. The groups are said to use "explicitly defined methods to reduce bias,"⁷ but the guidelines⁸ are in fact very far from being explicit, nor are they comprehensive in their coverage of the relevant topics. And as for the "peer" review, it is conducted by "individuals with expertise in meta-analysis and in the content area of the review (eg, diabetes) and by a potential user of the review (eg, a practicing clinician)."⁷ Yet in no way does this constitute expertise in understanding the very major controversies (and tragedies) surrounding the University Group Diabetes Program, for example, the evidence and its interpretation in respect to relative efficacies, and the possible cardiac side effects of tolbutamide use. And in particular, "expertise in meta-analysis" is but familiarity with some statistical trivia rather than mastery of principles of applied medical research in respect to both objects and methods, nor is the requisite mastery at all inherent in

even academic expertise in any given "content area." It would be interesting to know how often "experts" such as these have rejected a Cochrane Review, having already approved 276 of them (as of August 1997); for, if the rejections are uncommon, the *Cochrane Database of Systematic Reviews*⁷ is scarcely worthy of practitioners' trust.

Whoever harbours doubts about this gloomy depiction of the present state of scientific evidence in medicine does well to consider further the relative merits of thrombolysis and angioplasty in acute myocardial infarction. On the effects of these treatments for this indication, an extraordinarily large amount of research has been carried out and also subjected to "meta-analyses." Yet, in the face of the latest study report,¹ the editors of the journal were able to solicit 2 diametrically opposing conclusions, respectively titled "Thrombolysis — the preferred treatment"² and "Primary angioplasty — the strategy of choice."³ Moreover, both parties rebutted the other's piece without having learned anything at all from it,^{4,5} and the editors' consequent conclusion was that "Readers will have to judge for themselves."⁹ Enormous efforts and expenditures have, thus, resulted in evidence which guides presumed experts to highly divergent conclusions and leaves nonexpert readers to judge for themselves! The editors should have spared the readers and solicited an exposé on the principles of original research into the effects of interventions of interest in acute myocardial infarction, the principles of the synthesis of evidence from such research and the principles of interpretation of the aggregate of evidence — the prevailing ones, editorially sanctioned, and those that should substitute for these.

A vision of progress

Rather than dwelling further on today's practitioners' frustrations in respect to the scientific evidence that should guide the practice (gnosis in it), I offer a vision of the direction in which development should, and will, proceed. The root problem lies, I hold, with the system of values that still is possessing academic medicine. In particular, even a clinical professor's credentials are still defined mainly in terms of original research, and original basic research is still more highly valued than original applied research. (This reflects the Flexnerian tradition as for the concept of scientific medicine: medical science is laboratory science; as an expression of this, a physician still boasts a laboratory coat.) As a consequence, a clinical professor's scientific expertise still tends to be very narrow in scope, and off-focus even as such, in relation to the wide range of applied-science issues (of gnosis) that are faced in the practice of his or her specialty; the professor tends to be an authority on these without possessing the requisite expertise. But pressures toward a more rational set of values are mounting: patients and payers are ever less content with unsubstantiated authority, and *they* are ever



more familiar with evidence from applied medical research. The time will have to come, soon, when clinical professors come to grips with their true responsibility, that of being supreme authorities on the aggregate of applied-science evidence bearing on at least the most common challenges of practice in their respective specialties. When they do, this will mark the end of the present era of dilettantism in defining the burden of scientific evidence on any such topic. For this outlook will force each clinical professor to become a master of the theory of gnosis per se and then of gnosis-oriented original research, as well as that of synthesizing and interpreting the evidence from it; it will guide the professor away from the time-consuming travails of original gnosis-oriented research, to merely fostering it where needed; and above all, it will engender a devotion to the synthesis of original evidence and the dissemination of its results, personally obtained as well as those by other similar experts, to both residents and already-certified practitioners.

How fast such enlightenment will develop depends not only on leaders in faculties of medicine but also on editors of medical journals. Not only do editors decide on the extent to which quintessentially applied — gnosis-oriented — reports, original and synthetic, are presented, but they also guide the patterns of thought underpinning the production and presentation of the evidence. At the core of this are editorial policies in respect to the structure of the Abstract and that of the report proper. As it is today, many journals have explicit, though varying and still evolving, guidelines for the structure of the Abstract, but the body of the report is generally allowed to follow the routine of Introduction, (Subjects and) Methods, Results and Discussion. That the stipulated structure of the Abstract varies and consistently diverges from that of the report proper is prima facie evidence of need for further development. As for practice-guiding evidence in particular, a sensible structure, I suggest, would have the following elements: Introduction (to the object); *Object* (of study, of evidence), this being the pivotal element (recall the example above); Methods (of pursuing evidence on the object), “subjects,” “setting,” “context,” etc., being part of this, “type of study” being all of this, “design” being the genesis of this, protocol being the documented plan for this, and protocol execution being the essence of this; Results (the resulting numerical evidence), their form in conformity with that of the object, and their content as outlined above (without a single *P* value); and Discussion (of problems with the study or the synthesis, i.e., with the formulation of the object and [or] the design and execution of the methods). The outstanding structural deficiency today is, as I exemplified, the lack of attention to the pivotal element, the formulation of the object of study and the justification of this — the lack of an express section for this in both the report proper and its Abstract.

In conclusion, a word about the Conclusion section now commonly expected in the Abstract of a report. Objectivity is in the essence of evidence, but the interpretation of evidence is fundamentally subjective, as was illustrated by the example above. As science progresses, the subjective interpretations tend to converge toward shared understanding, toward *objective* interpretation in this sense, and indeed toward correct understanding. Only widely shared understanding among scientific experts qualifies as scientific *knowledge*, misunderstanding though it may represent. Authors’ reported conclusions are, thus, inconsistent with the nature of science, its “principle of *publicity*: only the outcomes of critical discussion within the scientific community may be tentatively accepted as the results of inquiry. In this sense, the real ‘subjects’ of scientific knowledge are the scientific communities rather than individual scientists.”¹⁰ As for how best to treat acute myocardial infarction and how to follow up on this, the implication of this is that, while scientific evidence and conclusions abound,¹⁻⁵ the requisite scientific knowledge is missing. But one day, upon further philosophic (values; essence of scientific medicine) and ontic (object-conceptualization for gnostic research) as well as corresponding epistemic (methodology in gnostic research) progress in clinical academia, through consequently more enlightened research and its more enlightened reporting and synthesis, and with solo conclusions from original studies replaced by critical discussion about the aggregate of evidence and its interpretation in ever-more-enlightened journals — true discussion in lieu of conclusions and rebuttals of these — yes, one day there will be not only evidence but also the thus-far-elusive knowledge.

References

1. Every NR, Parsons LS, Hlatky M, Martin JS, Weaver WD, for Myocardial Infarction Triage and Intervention Investigators. A comparison of thrombolytic therapy with primary coronary angioplasty for acute myocardial infarction. *N Engl J Med* 1996;335(17):1253-60.
2. Lange RD, Hillis LD. Thrombolysis — the preferred treatment [clinical debate]. *N Engl J Med* 1996;335(17):1311-2.
3. Grines CL. Primary angioplasty — the strategy of choice [clinical debate]. *N Engl J Med* 1996;335(17):1313-6.
4. Lange RD, Hillis LD. Rebuttal [clinical debate]. *N Engl J Med* 1996;335(17):1316.
5. Grines CL. Rebuttal [clinical debate]. *N Engl J Med* 1996;335(17):1317.
6. Epstein AM. The outcomes movement — Will it take us where we want to go? *N Engl J Med* 1990;323(4):266-70.
7. Bero L, Rennie D. The Cochrane Collaboration. Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *JAMA* 1995;274(24):1935-8.
8. *The Cochrane Collaboration handbook*. Oxford (UK): The Cochrane Collaboration; 1994.
9. Editors. Should thrombolysis or primary angioplasty be the treatment of choice for acute myocardial infarction? [clinical debate]. *N Engl J Med* 1996;335(17):1311.
10. Niiniluoto I. *Is science progressive?* Dordrecht: D. Reidel Publishing; 1984. p. 4.

Reprint requests to: Dr. Olli S. Miettinen, Department of Epidemiology and Biostatistics, McGill University, 1020 Pine Ave. W, Montreal QC H3A 1A2