International Journal of Technology Assessment in Health Care, **19:3** (2003), 435–445.

ASSESSING ALTERNATIVE MEDICINE: METHODOLOGICAL AND RESEARCH POLICY CONCERNS

Stephen H. Linder
The University of Texas, School of Public Health

Abstract

Objectives: Setting funding priorities among research projects for complementary and alternative therapies is especially vulnerable to arbitrary, partisan criteria and opportunistic readings of controversial evidence. This study develops a procedural approach to characterizing, in a transparent and even-handed way, the available evidence on these treatments and demonstrates a simple analytical scheme for prioritizing competing, but typically incommensurable, research projects for public support.
Methods: A simple five-step scheme for categorizing therapies is developed and justified, based on a taxonomy of the study designs used to generate research evidence. Once identified, candidate therapies are assigned to ordered categories along these design criteria and effectively grouped into equivalence classes by type of evidence. Finally, a partial ordering on the therapies is formed within each class by means of secondary screening criteria.
Results: Twenty-five candidate therapies are assigned to equivalence classes. The intent, in effect, is to restrict comparisons to those therapies that fall within a particular class of similar study designs. Within-class orderings avoid the problem of having better-known or better-supported therapies crowd out lesser known ones when it comes to allocating dollars for more research. A set of criteria and procedure for prioritizing spending for further research is demonstrated.
Conclusions: Relying on an open, formal procedure for comparing unconventional therapies offers protection against prejudgment in setting funding priorities, especially when weak clinical evidence relates more to a low investment in research than lack of efficacy.

Keywords: Decision support techniques, Research support, Complementary therapies

The establishment of the Office of Alternative Medicine (OAM) in 1992, as a component of the U.S. National Institutes of Health, extended official recognition to the potential clinical significance of unconventional forms of health care and treatment. The Office was charged, in part, with funding and promoting research to assess this significance. The motivating problem, stated at the time, was that relatively few unconventional therapies had been successfully investigated by conventional research methods. Although skeptics within the biomedical community (6) held up the randomized trial as the only method capable of generating objective data on the clinical value of any therapy, conventional or otherwise,

Special thanks are due to Mary Ann Richardson, DrPH, formerly of the UT School of Public Health, currently with the National Center for Complementary and Alternative Medicine, National Institutes of Health, Bethesda, Maryland. Dr. Richardson introduced me to this problem and graciously provided information on several unconventional therapies. Partial funding was provided by UT's Center for Alternative Medicine through a contract with the Office of Alternative Medicine, U.S. National Institutes of Health.

proponents of these therapies (1) countered that neither trials nor other conventional methods could capture the holistic character of many unconventional treatments. The OAM adopted a middle course between the two.

Although systematic research serves as a necessary basis for judging clinical practice regardless of the therapy at issue, OAM contended, the clinical trial was not the only way to go about it. Other methods needed to be included, and more preliminary questions had to be answered well before clinical trials entered the picture (17). Three assumptions were critical to OAM's position. First, research questions about clinical practice were to be addressed in a certain order, that is, some questions had to be postponed until the answers to others became known. Second, the answers to these temporally sequenced questions formed a hierarchy of evidence; as more questions were answered, each subsequent piece of evidence would move judgment closer to conclusiveness. Although not all types of evidence had the same probative weight, each type on a given level in this hierarchy had some role to play in supporting, what amounted to, cumulative judgments of clinical significance. Finally, the choice of research method at any given level in the hierarchy would be dictated by the question being asked and not by the nature of the therapy or the imperative for greater rigor. In other words, pick the method to fit the question, and save the hardest questions for last.

From this perspective, all unconventional therapies are amenable to some systematic assessment and cannot be deemed "unscientific" until the proper questions have been asked and answered. And although relatively few therapies appeared ripe for trials, most were supported by some kind of evidence that could be assessed, and new questions posed (2). Once the research door was officially opened, however, it became clear that not all therapies could come through at once. Priorities still needed to be set. Funding remained a key issue. Again, there were conflicting positions to contend with, and the political forces that had forged the OAM (and prompted the resignation of its first director) would not easily be placated (12).

In part, as testimony to the political success of OAM's via media strategy and under pressure from the growing popularity of unconventional therapies, the Congress upgraded the OAM to full Center status in 1998, creating the National Center for Complementary and Alternative Medicine (P.L. 105–277). The first director would be drawn from within the NIH laboratory system, however, a specialist in virology and infectious diseases with no ties to the alternative medicine community—in sharp contrast to the last director of the OAM, whose sympathies for alternative therapies were well known (10). The new National Center quickly assembled advisory panels and peer reviewers, and adapted the prevailing NIH system of evidence-based reviews to help insulate its decisions about priorities and investments from external pressures.

In the meantime, these careful first steps were upstaged by the formation of a White House Commission on Complementary and Alternative Medicine Policy (Executive Order 13147). Its membership included two former OAM directors and an array of alternative-therapy advocates and practitioners. Their charge was to recommend policy on, among other things, how best to integrate alternative therapies into mainstream medicine and training and on what government's role should be in licensing, reimbursement, and regulation. To use a homely metaphor, the policy cart was firmly planted out in front of the evidentiary horse.

In contrast to most disease-related clinical research, the investigation of alternative medicine therapies takes place in a polarized climate of advocacy and resistance, potentially rife with politics and stakeholder interests (12). This circumstance puts the National Center for Complementary and Alternative Medicine in a paradoxical position: any special treatment afforded alternative therapies can suggest agency "capture" to skeptics, while honing to the usual, scientific-research line within NIH suggests, to alternative-medicine advocates at least, cooptation by the establishment orthodoxy. Responding to this paradox,

the via media now offers outreach and training to the alternative medicine community and a conventional clinical research emphasis (at home in a several other NIH Centers) to ensure scientific legitimacy. And although OAM's hierarchy of evidence appears in the Center's strategic plan (14), the central problem of setting research and investment priorities across therapies is said to be handled in two ways, neither of which pays the paradox its due.

The first way, familiar from OAM, is to include in a post hoc manner a variety of market and feasibility criteria, such as popularity and ripeness. The second way draws on evidence-based medicine principles and assumes that the data from study reviews will be decisive, even though the evidence for unconventional therapies may be, in a word, unconventional. Even if these reviews can overcome their conventionality bias, results will still need to be assigned a priority for further investigation and investment. We are once again led back to the possibility of politicized choices.

The purpose of this article is to offer a tentative solution to the problem of post hoc criteria and ad hoc prioritizing. The formal procedure proposed here (i) relies on minimal assumptions, (ii) accommodates a broad range of evidence, (iii) establishes an analytical basis for comparing therapies, and (iv) takes advantage of OAM's hierarchy of evidence notion to avoid prematurely excluding underscrutinized therapies. The product is a partial ordering of therapies within classes that represent particular levels on the evidence hierarchy. The intent, in effect, is to restrict comparisons to those therapies that fall within a particular class of similar study designs. Within-class orderings avoid the problem of having better-known or better-supported therapies crowd out lesser known ones when it comes to allocating dollars for more research. The method proposed here structures discretion over priority-setting in a way that openly balances the enthusiasm of advocates against the caution of skeptics.

Assigning therapies to classes based on the status of their evidence effectively stratifies for the single most important confounder in any therapy-to-therapy comparisons–the relative conclusiveness of what is known about them. Instances where the evidence is partial or rudimentary in scientific terms might then receive the same consideration for research support as those ripe for trials, unless there are compelling practical reasons for not doing so. In this context, grouping candidate therapies helps to ensure more equal treatment in the face of widely disparate evidence. To be sure, pitting anecdotal results against trial outcomes in comparing two distinct therapies is not only comparing apples to oranges but immature to mature ones at that.

The next section builds a scheme for categorizing therapies based on a taxonomy of the designs used to generate research evidence. Attention then turns to the assignment of twenty-five candidate therapies and their grouping into comparable evidence classes. Finally, a set of criteria and procedure for prioritizing therapies within classes will be demonstrated. A discussion of policy implications follows.

PROCEDURE FOR PRIORITIZING RESEARCH

The principal advantage of a formal procedure for setting research priorities is its transparency. Assumptions must be explicit and each of the steps presented in practical detail. Furthermore, each element of the taxonomic model, introduced to impose an ordering across the collection of candidate therapies, can be scrutinized one-at-a-time. This transparency disciplines how the construction proceeds: it should be kept as simple as possible, and with a minimum of arbitrary elements that elude justification. A related advantage is its impartiality. The assessments operate exclusively along the stated criteria and stress equivalent treatment for each therapy. Although biases may intrude on the selection of criteria, such intrusions must be justified and do not act directly on any given therapy's relative standing.

Table 1. Steps in Setting Research Priorities

1. Identify candidate therapies
   - Consult official and unofficial sources
   - Search the published literature for research studies
2. Develop a taxonomy of study designs
   - Fashion criteria for comparing designs
   - Designate ordered categories along each criterion
3. Categorize therapies by their study designs
   - Assign therapies to ordered categories
   - Compute evidence profiles based on category assignments
4. Classify therapies based on their evidence profiles
   - Cluster therapies by similarity across evidence profiles
   - Form equivalence classes based on cluster membership
5. Assign priority to therapies within each class
   - Select external criteria for judging potential significance
   - Tally the results of pairwise comparisions along these criteria

As with most algorithms, the deployment and operation of this procedure is typically more iterative than sequential, with tasks spanning portions of several steps at the same time. Nonetheless, the clearest depiction of the procedure remains a set of discrete, sequential steps; these are summarized, along with their associated tasks, in Table 1.

Step 1. Identify Candidate Therapies

The prioritization process begins with decisions about scope and focus. There are not only widely diverse forms of unconventional therapy in different modalities, ranging from reflexology and Reiki to homeopathic regimens, but many different kinds of relief or enhancements being sought (15). The focus here is on cancer for several reasons. First, unconventional therapies have made more inroads into biomedicine for this class of conventional diseases than for most others, in part, because of oncologists' aggressive pursuit of adjunctive therapies (11). The perceived failure of conventional anticancer therapy also leads many patients, otherwise without hope, to search for unconventional cures. Second, the pathology of metastatic diseases offers clear markers for treatment effect, such as changes in tumor size, once clinical trials become relevant. Within the cancer area, the adjunctive modalities thought by conventional clinicians to hold the most promise are the biopharmacological agents (4;13). These include herbs, plant extracts, enzymes and minerals, as well as synthetic compounds.

To develop a short list of candidate therapies, a wide range of sources were relied upon, paying attention to official recognition by health agencies here and abroad, ongoing clinical practice and research, and the claims experts recognized within either the conventional or unconventional communities. Twenty-five therapeutic agents with putative anticancer properties were identified. The next task was to inventory the relevant research studies through an exhaustive search of the published and unpublished literature, paying special attention to study design.

Step 2. Develop a Taxonomy of Study Designs

Because no schemes for classifying so diverse a set of study designs existed, one needed to be developed that would permit all plausible forms of evidence to be accommodated but, at the same time, was as parsimonious as possible. The basic taxonomic requirement is to devise a minimal set of criteria for differentiating among studies. These criteria should be conceptually independent, if not mutually exclusive, and sufficiently meaningful to support

reproducible assignments and the formation of at least an ordinal level of differentiation among the items to be classified on each criterion.

Using a simple spatial analogy, the basic structure of my taxonomic scheme forms a three-dimensional space, organized by three perpendicular axes. Each axis corresponds to a criterion for judging a basic methodological feature of study design. The substantive categories along each criterion appear as ordered segments along each axis. To preserve the ordering of each criterion's categories, the corresponding axis segment is assigned a rank number. As we move from the origin along any axis, the rank numbers increase, as does the methodological rigor of studies assigned to the corresponding categories. As an aid to interpretation, we also label each segment with a term meant to summarize a particular design feature.

Using labeled, ordinal segments rather than numerical coordinates vastly simplifies the classification and keeps the abstract but arbitrary assumptions behind the spatial model to a minimum. The rank numbers together with the axes are sufficient to ensure the meaningfulness of a relative position within the space. The relative methodological standing of any given study design, then, can be modeled as a location where three segments intersect rather than as a unique point. A randomized clinical trial, a case-control study, and an anecdotal report would each occupy distinct locations, for example: the report would fall closest to the origin, the trial would be the furthest away, and the case-control study would fall somewhere between these two. Given this construction, however, there would be no way to differentiate among studies occupying the same segment, without imposing some additional weighting principle from the outside. As we note below, the *number of studies completed* will be introduced as such a principle, once we shift in Step 3 from classifying types of studies to classifying therapies.

The first axis corresponds to a design characteristic we refer to as Treatment. A systematic concern for treatment first appeared in the definitive work of Fisher on experimental design (8) and was refined by the efforts of Campbell and Stanley (3). To attribute causal efficacy to a particular experimental factor, or treatment, one must be able to manipulate it in a controlled way and observe its effects on selected units – the basis of a "true" experiment. In the absence of direct manipulation, however, the introduction of a factor can be observed as a discrete event, and its effects characterized in some way that limits error, preferably with ex ante and ex post observations. Campbell and Stanley called these studies "quasi-experimental". Finally, with only ex post observation, a factor can still be treated as an attribute of certain units bearing some relation to putative effects; these are widely known as observational studies and have been fully characterized by Cochran (5). Along the Treatment axis in my spatial rendering, observational studies occupy the segment nearest the origin, quasi-experimental studies lie along the middle, and experimental studies are at the upper end.

The second axis corresponds to a methodological characteristic called Selection. Selection was also discussed by Fisher but elaborated in later work on sample surveys by Yates (18). At issue is how the observational units subject to experimental manipulation are selected. According to Fisher, units must be randomly assigned to any controlled variation in the experimental factors, otherwise the effects one observes could easily be due to nonexperimental or unsuspected causes. At the other extreme, when the observational units are persons motivated to experience certain effects, they may select themselves for, or subject themselves to, various treatments. The difficulty lies in untangling the workings of the motivational factors from the experimental ones, when it comes to understanding observed or reported effects. Nevertheless, from the experimenter's point of view, selection is about sampling for assignment. The segment closest to the origin, along the Selection axis, accommodates self-selection; the segment furthest away includes random sampling or assignment. The segment in-between includes all of those design characteristics intended

to establish artificial equivalence among units as a means of addressing nonexperimental effects. The most common are stratification and blocking schemes.

The third axis, labeled Comparison, captures those aspects of study design that define a benchmark for determining the relative magnitude of any observed effect. Within the domain of clinical studies, the benchmark typically references persons as observational units. If all were exposed to the same treatment, for example, their different characteristics may become very important in sorting out variation in any observed effects. If, on the other hand, not all were exposed, then the unexposed can serve as a benchmark for any effects observed from exposing the rest. Finally, there may be multiple groups of unexposed, each with certain characteristics that can help to isolate changes in the exposed groups. The segment near the origin of the Comparison axis contains designs with a single self-referencing group. The middle segment involves designs with at least two differentially exposed groups, the simplest being an exposed group matched to an unexposed group on some characteristics thought relevant to the treatment. The farthest segment from the origin is reserved for complex designs using multiple groups.

Step 3. Categorize Therapies by Study Design

The next step initiates the classification process by characterizing the study information assembled in Step 1 in terms of the three criteria from Step 2. The spatial model for arraying study designs based on their methodological characteristics is now applied to the categorization of the twenty-five candidate therapies. Because most of the candidate therapies appear in more than one published study and have been assessed with multiple designs, we need to accommodate the number as well as the types of studies. The alternative to accommodating all of the studies in some kind of weighting scheme would be to select a representative design for each therapy; this selection might be the design most frequently used or perhaps the most rigorous one among those used. Both involve a narrow screening of study information that in one way or another discounts evidence. Using only the modal design ignores all research experience with any more rigorous designs, unless of course, the most rigorous design was also the most frequently used–a rare event that did not occur for any of the therapies in this sample. Using only the most rigorous design, on the other hand, ignores the probative importance of replication and corroboration that multiple studies afford. It would fail to differentiate the therapy with only a single attempt at a complex design, from those with an extensive record of such attempts. Instead, we opt for a portfolio approach, where number of studies is treated as a weight for study type on each of the three criteria.

Each study of a particular candidate therapy is assigned to one of the three segments along each of the three axes in the spatial model. This effectively produces a frequency distribution of human studies for twenty-five therapies across nine labeled categories. The next task is to weigh the number of studies by the type of design; here, the rank numbers for each segment are used to weigh the counts of studies for each therapy, yielding a composite number on each axis. Because the number of studies vary widely from therapy to therapy, ranging from none to eighty-seven, the final task is to normalize each therapy's composite numbers by its total number of studies. This rescales the composite numbers for each therapy onto a 0 to 3 range. As we move from the lower to the upper end of this range, the mode of the frequency distribution of studies shifts from the first segment on a given axis toward the third segment, that is, from less to more complex designs. The greater the magnitude of a therapy's composite number, in other words, the more rigorous is the preponderance of study evidence addressing it. Thus, each therapy can now be represented by a unique profile of its evidence in the form of its three composite numbers. The next step will treat these profiles of composite numbers as coordinates for grouping therapies based on the similarity of their evidence.

Step 4. Classify Therapies by Evidence Profile

Once we extend the spatial model a step further and treat these evidence profiles as coordinates, we can develop a taxonomy of therapy classes, inductively, by attending to the spatial patterns that result (16). In spatial terms, similar evidence profiles will be located close together and dissimilar ones, far apart. Where these profiles group together in a well-defined cluster, we have empirical support for treating them as members of a homogeneous design class. The question then is, how can we ensure against arbitrary boundaries for cluster membership? There will typically be a few points located in-between the obvious clusters. If we draw the cluster boundaries by hand, the assignment of these few is idiosyncratic. Because priorities will be set within classes, the addition of an extra therapy to one class or another can affect the ordering. A clustering algorithm that forms these classes on a strict numerical basis offers a resolution less prone to bias than ad hoc visual methods.

To identify classes, a clustering routine was used to group evidence profiles based on a pairwise measure of similarity. Several routines were used, because their technical assumptions vary, and the intent was to find a stable solution insensitive to the differences across these routines. The appropriate number of clusters was determined by both inspection and comparison across several clustering solutions. Eight distinct clusters emerge from this analysis and are taken to constitute empirically defined equivalence classes. The spatial display of these results appears in Figure 1 and completes the classification phase of the priority setting. The typical way to describe clusters is by their centroids, formed by averaging the profiles of each cluster member. The plot shown in Figure 1 locates each of the eight clusters by its centroid along the three axes of the spatial model. The classes that these clusters represent are labeled by one of their member therapies; those chosen as class labels also happen to assume the highest priority in their respective classes, as will be shown in the step that immediately follows. The membership list for each of the 8 classes appears in Table 2, the focus of the next step.

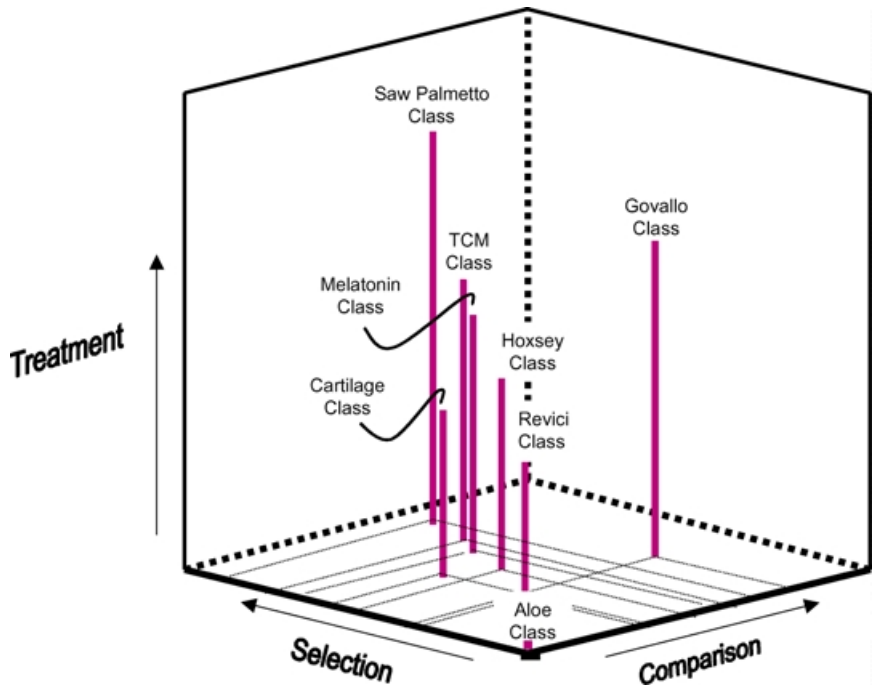

Figure 1. The relative ordering of therapy classes.

https://doi.org/10.1017/S0266462303000382 Published online by Cambridge University Press

Table 2. Candidate Therapies Listed by Class Membership

Class	Therapy
Saw palmetto	Saw palmetto
	Selenium
TCM	TCM
	Mistletoe
	Green tea
	Garlic
	Hydrazine sulfate
	MTH-68
Melatonin	Melatonin
	Coley
Govallo	Govallo
Cartilage	Cartilage
	Antineoplastons
	Cat's claw
Hoxsey	Hoxsey
	Livingston
	Gerson
	Macrobiotic
Revici	Revici
	714X
	Essiac
	IAT
	Homeopathy
Aloe	Aloe
	Citrus pectin

Before considering the assignment of priorities, however, it is useful to review Figure 1 in greater detail, because it provides a summary picture of the relative ordering of the therapy classes by their evidence profiles. The Aloe Class resides at the origin of all three methodological characteristics and had the fewest studies of any kind. Following an arc that moves back to the left and upward in a stair-step manner, we reach the Saw Palmetto Class distinguished by the disproportionate rigor of its studies; whereas a third of the therapies generated more studies, none except the therapies in this class had a majority of its studies concentrated at the higher end of the axes. Although the classes appear to be dispersed throughout the space, there are some interesting patterns to note.

If we posit the Hoxsey Class as a benchmark, roughly defining the midpoint along all three axes, we can take advantage of the customary descriptions for location in three dimensions. Along the Comparison Axis, the Aloe, Revici, and Cartilage Classes are in front of the Hoxsey Class, and the remainder are behind it. Along the Selection Axis, the Aloe, Revici, and Govallo Classes are to the right of it and the remainder are on the left. For Treatment, again, the Aloe, Revici, and Cartilage Classes fall below it and the rest are above. First, note that all of the classes behind the Hoxsey benchmark are also above it; similarly, those in front are also below it. In other words, there appears to be an empirical relationship between the use of more rather than fewer groups for comparison and the controlled rather than natural variation in one's treatment. This finding comports with common sense about the way advanced designs are constructed. And yet, the pattern seems to hold on both the left and right of the benchmark class, that is, across different positions on the Selection Axis.

In short, having at least three axes, given the various design combinations they support, is important. If we had relied on the four or five typical designations of study types, for

example, anecdotal, observational, case-control, and randomized trial, we may have been able to capture a limited frequency distribution of study designs, but the complex combinations represented in the taxonomy would have been lost. Therapies with conceptually different and practically distinct designs would have been grouped together, undermining the reason for an evidentiary classification in the first place.

Second, although my intent here is to establish priorities for further studies and not to pass judgment on the efficacy of one therapy over another, we can draw some substantive inferences on the relative probity of the evidence available to do this. The three classes located behind, to the left, and above the benchmark, specifically, the Melatonin, Traditional Chinese Medicine, and Saw Palmetto Classes, tend to have more studies of a more rigorous kind than do the others. Furthermore, their evidence profiles, both the counts and types of studies, are roughly alike in comparison to the classes located elsewhere in the space. On the whole, these three classes have generated substantially more, probative evidence to support judgment about their member therapies than have the others. Consequently, they may be closer to some more definitive determination of efficacy with a few careful replications. Decisions about research investment, then, are likely to differ in this region of the space from those in most other locations, where the evidence profiles are generally less probative.

Step 5. Assign Priority to Therapies within Each Class

In the final step, external criteria, unrelated to the evidence profiles underlying the taxonomy, are introduced to prioritize therapies within each class. Here, certain clinical and nonclinical features of these therapies, such as, their relative cost, availability, reliability, potential toxicity, and side effects, can be brought to bear on how best to invest limited research resources. Setting priorities in this context involves forming a partial ordering of therapies. The ordering itself might be based on composite ratings, where each therapy receives a performance score on each criterion, and the summation of these reflects relative priority. Unfortunately, few of the criteria at issue are sufficiently well-developed to permit reliable, global ratings across the therapies without comparing one with another. Instead, the ordering can be derived from simple comparisons between therapies judged in pairs. Judgments are strictly relative and focus on which one of the two is the more desirable or the better one on a particular criterion of interest. We then cumulate the number of times a given therapy is judged superior to its peers in pairwise contests, and the cumulative score yields the rank ordering.

The screening criteria used for the rankings include the relative prevalence of use, drawn either from surveys or market data on purchases, and several clinical factors of interest. First, how extensive is the nonhuman research? More laboratory research was assumed to be superior to less. Second, is there a known mechanism of action? A known mechanism was favored over an unknown one. And third, how severe are the known side-effects? Less severe side-effects were deemed more desirable. These are especially important when considering adjunctive therapies for cancer treatment.

The final criterion relates more to the complexity of the treatment regimen than to the clinical properties of the therapy. If a therapy is considered holistic, in the sense of requiring multiple modalities of treatment and typically invoking mind-body interactions, it likely demands special skills beyond the training of most conventional practitioners. In this context, a less holistic regimen of treatment, more easily standardized across sites and practitioners, was favored over a more holistic and, thus, more complex one.

The final prioritization of therapies appears in Table 2 as a set of eight within-class orderings. The commitment to even-handedness in the assessment of these very different therapies dictates that funding allocations should reflect not just the ordering within a given class but also the presence of multiple classes, each with its own top contenders. In the

absence of any systematic, nonarbitrary way to order the classes themselves, the best we can do is to take advantage of the ordering within them and somehow spread the funds across classes. On the surface, however, there is no reason to believe that each class should get an equal share of the available support. The relative share might well be proportional to the costs of the requisite study designs. To be sure, the designs called for to advance the evidence in some classes will be much more expensive to field than those for other classes. Full-blown trials, for example, will be more costly than retrospective, case studies. In any event, whatever allocation does occur will be insulated from wide disparities in the state of the science.

POLICY IMPLICATIONS

When it comes to allocating funds, there are at least three ways to proceed. The first relies on a partisan resolution. Advocacy groups press their claims for funding, and some administrative unit judges the merits of these and assigns the money accordingly. This is a market model that is subject to serious distortions whenever the requisite information is missing or shaped by enthusiastic marketeers. Here, the efficacy of the promotional campaign can displace the efficacy of the therapy itself, because little definitive knowledge exists about most unconventional therapies.

As mentioned earlier, the second way relies on biomedicine's request-for-proposals process and expert peer review. The problem with this approach is that it depends heavily on the existence of an infrastructure of well-defined fields of inquiry, standardized, systematic methods, and an invisible college of experienced investigators who share certain norms and commitments. In the absence such an infrastructure, the process is highly vulnerable to the intrusion of extra-scientific criteria for judging research potential. Absent much agreement on the existing evidence, the outcome would be highly sensitive to the composition of the reviewing panel. Research priorities, then, could be effectively set by those selecting the participants, rather than by the deliberation of the participants themselves. Countering the potential disadvantages faced by unconventional therapies in such a context is one of the chief reasons that OAM was formed; it also motivates current efforts to establish a separate peer review process consisting principally of investigators outside conventional biomedical circles. Until the requisite infrastructure can be built, the normal peer review process seems ill-suited to judging unconventional therapies.

The third way is to devise an impartial procedure and stick to it until an adequate infrastructure emerges to support peer review. The procedure described above is intended to serve this need. Much of its detail arises from efforts to minimize the intrusion of arbitrary or ad hoc judgments. What complexity there is, comes primarily from the diversity of study designs and evidence and not from any effort to make the unsystematic appear systematic or to obscure the underlying assumptions. The central concept is the within-class comparison. Without it, judgments would be procrustian, collapsing different kinds of evidence onto a single dimension or overarching criterion. Impartiality with respect to the diversity of evidence depends upon within-class comparisons. Once it is assured, priorities can be set by a variety of criteria, depending upon the treatment focus.

Although the procedure for prioritization works, it is important to take note of what this means. Priorities relate to the external criteria chosen and not to any claims of efficacy, many of which are scientifically premature. Furthermore, the classes formed to structure the priority setting reflect the form and probity of the evidence but not its content. The published studies are not of uniform quality nor are their results necessarily supportive of therapeutic claims. The basic idea is to differentiate among therapies based on the rigor of their evidence to make plans for further study. There is no effort to rule out further study altogether or to pass judgment on efficacy at the current state of available knowledge. The

emphasis here is exclusively prospective. Where should more research effort be invested to advance the state of the science on each of these candidate therapies? It is the certainty of limited funds that makes prioritization a necessary task.

From an investment perspective, the relative likelihood of a payoff for government funders of any particular research project in this domain seems to hinge on the expected conclusiveness of its findings; either claims need to be put to rest, or the therapy earns a valid place in conventional treatment regimens. Inconclusive findings would suggest that a competing project with possibly greater potential for conclusiveness was overlooked in error. Obvious parallels can be found in other areas of unconventional technologies, where the evidence is partial, subjective, and difficult to compare across proposed research projects–and there is organized skepticism over any chance of proven effectiveness. We are left then with speculative judgment as the basis for unconventional project selection and the risk of inconclusiveness in research findings. Although there are many subtle elements that can be accommodated in less formal judgment processes, the need for accountability in public funding tends to favor more systematic and reviewable processes for R&D decisions. The procedure presented earlier suggests a plausible avenue for dealing with the unconventional in research project selection.

REFERENCES

1. Bird C. Gaston Naessens vs. scientific medicine: Brilliant French-born Canadian biologist continues his struggle against the cancer establishment. *Townsend Letter for Doctors and Patients.* 1991;94:313-320.
2. Boik J. *Cancer and natural medicine: A textbook of basic science and clinical research.* Princeton, MN: Oregon Medical Press; 1995.
3. Campbell D, Stanley J. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally; 1963.
4. Cassileth B, Chapman C. Alternative and complementary cancer therapies. *Cancer.* 1996;77:1026-1034.
5. Cochran WG. *Planning and analysis of observational studies.* New York: Wiley; 1983.
6. Ember L. Alternative medicine slammed. *Chem Eng News.* 1997;75:17-18.
7. Evers M. Testimony Submitted to the U.S. Senate Labor and Human Resource, Subcommittee on Public Health and Safety, 105th Congress. Washington, DC: October 15, 1997.
8. Fisher RA. *The design of experiments.* New York: Hafner; 1935.
9. Hawkins M, Friedman M. National Cancer Institute's evaluation of unconventional cancer treatments. *J Natl Cancer Inst.* 1992;84:1699-1702.
10. Jonas W, Jacobs J, Dossey L. *Healing with homeopathy: The doctor's guide.* New York: Warner Books; 1998.
11. Lerner M. *Choices in healing: Integrating the best of conventional and complementary approaches to cancer.* Cambridge, MA: MIT Press; 1994.
12. Marshall E. The politics of alternative medicine. *Science.* 1994;265:2000-2002.
13. Moss RW. *Cancer therapy: The independent consumer's guide to non-toxic treatment and prevention*. New York: Equinox Press; 1992.
14. National Center for Complementary and Alternative Medicine. *Expanding horizons of healthcare, strategic plan, 2000-2005*. Available at: www.nccam.nih.gov/strategic. Accessed
15. Office of Technology Assessment. *Unconventional cancer treatments.* Washington, DC: U.S. Government Printing Office; 1990.
16. Sokal R, Sneath P. 1963. *Principles of numerical taxonomy.* San Francisco: W.H. Freeman.
17. Vickers A, Cassileth B, Ernst E, et al. How should we research unconventional therapies? *Int J Technol Assess Health Care.* 1997;13:111-121.
18. Yates F. *Sampling methods for census and surveys.* New York: Hafner; 1960.