

## Participant-Centered Analysis in Complementary and Alternative Medicine Comparative Trials

MIKEL AICKIN, Ph.D.

### ABSTRACT

Participant-centered analysis involves applying the customary methods of statistical decision making at the level of the individual research participant. Consequently, each individual is declared a responder who benefited, a nonresponder, or possibly a responder who was harmed, using intensively collected data that were specific to that individual. There are several implications of the participant-centered approach. More data actually relevant to the important outcome need to be collected on individuals. The study results can be summarized in a simple table of responders/nonresponders by treatment group, and probabilities of true response can be estimated. The actual nature of the data collected and the statistical models used to analyze them drop into the background. Finally, production of individual-level decisions permits standard statistical approaches to be applied to the issue of which modality should be recommended for which person, instead of focusing on average effects and which modality should be recommended for everyone.

### INTRODUCTION

Consensus panels of the National Institutes of Health (NIH) have asserted that complementary and alternative medicine (CAM) research can be comfortably carried out using the methodologies that have been carefully identified over decades for allopathic biomedical research (Jonas, 1997; Levin et al., 1997; Vickers, 1996; Vickers et al., 1997). Indeed, attendees at several CAM research conferences that have been held over the past few years have consistently been told that their task is to fit CAM research into previous models, which have been so spectacularly successful in medical research (HMS/UCSF 2001; HMS 2002).

A less ideologic view might be that while well-developed research paradigms should be

used whenever they are appropriate, there remains a legitimate question about the existence of other possible paradigms, which might serve the CAM research community better. Asking this question raises important issues that need to be taken seriously by both CAM and allopathic researchers. On the one hand, pursuit of such new paradigms threatens to sever the connection between allopathic and CAM research, because if we vary the treatment (CAM versus allopathy) and the research method (conventional versus new approaches) at the same time, then how are we to know whether the final results are caused by differences in treatment or differences in research method? The other side of this dilemma is, however, that allopathic research methods have grown up historically in a symbiotic way with allopathic

---

Center for Health Research Kaiser Permanente Northwest Region, Portland, OR, and the Helfgott Research Institute at the National College of Naturopathic Medicine, Portland, OR.

medicine (drug trials are an example), and so it seems probable that the resulting methods have been naturally selected to favor the allopathic therapies, to the potential detriment of CAM alternatives. Allopathic researchers do not see this as a problem but perhaps some CAM researchers do.

The purpose of this paper is to offer a statistical rationale for a new, participant-oriented paradigm in CAM research. It is based on the observation that nearly all CAM modalities invest more time and intention in the patient-practitioner interaction than corresponding allopathic approaches do. And just as allopathic therapeutics call for less time with the patient, allopathic research involves collecting much less data on individuals than actually could be collected. Therefore, the first principle of participant-centered research is the imperative to collect relatively large amounts of data on each individual.

By itself, collection of more individual-level data does not solve any problems. In fact, this strategy actually creates additional difficulties for the conventional analysis of allopathic studies. Allopathic research focuses on the statistical significance of changes in means of measurements that were identified during the planning stage of the studies. The existence of measures on several different variables complicates the allopathic program, because it raises troubling questions about statistical adjustments for multiple testing (Hochberg and Tamhane, 1997). As a consequence, allopathic analyses are driven toward simplistic results partly as a consequence of the statistical methods that they regard as being valid.

Multiple measurements of a single variable on individuals create other problems for conventional analyses. It is widely regarded that measures on the same person over time should be regarded as being correlated. If one ignores this probable correlation in the conventional tests of group equality, then generally the statistical significance of the results will be overstated, which flies very forcefully in the face of conventional statistical wisdom. As a consequence, the collector of multiple measurements is compelled to use increasingly sophisticated software (Byrk and Raudenbush, 1992; Raudenbush et al., 2000),

which more and more researchers either distrust or do not understand.

This leads to the second principle of participant-centered research. The collection of large amounts of information at the individual level means that statistical inference should be carried out at the individual level. This tenet challenges the fundamental approach of statistical inference in biomedicine. The conventional approach is that the benefit (or lack thereof) of therapies should be judged by their effects on population parameters, such as means, proportions, or rates. The issue is, therefore, not whether any particular participant benefits or experiences harm from one or the other of the various treatments. The only issue is whether the policy of enforcing treatment A or treatment B across an entire population would result in a shift of the population mean (or other parameter) in the appropriate direction. The analysis of data at the individual level, on the contrary, means that benefit or harm will be assessed at the individual level, and that the assessment of the therapy will be judged by its statistically justified beneficial (or harmful) effect on each individual.

The research paradigm presented here consists of two parts. First, one must have a statistical decision-making rule that declares when a specific participant has been benefited (or harmed). The gross features of such a rule are well known, and depend only on its sensitivity and specificity. Using these ideas, we will see how to estimate the number of truly benefited (or harmed) individuals in each treatment group.

Second, one must fashion a decision rule out of the data that are collected on each individual. How this is done will depend on the nature of the data collected, but here the focus will be on the relatively simple case in which the effect is to be measured as a change in the mean of some measurement, where the term "mean" now refers to a population of measurements made on an individual. The application of statistical principles takes the population of measurements on an individual as the object of inference, instead of the population of measurements over many individuals.

Third, it is to be emphasized that these methods are put forward only in the case of con-

trolled trials. Although they can be applied at the individual level without a control group, nothing in the methodology removes the need for a valid comparison group.

## DECIDING WHO RESPONDS

Although there are several different ways to frame the issue of beneficial, harmful, or neutral responses, here we will simplify the discussion by focusing on the issue of benefit versus absence of benefit. Following classic principles, we first imagine that each person is actually benefited or not. Our fundamental problem is, however, that Nature is disinclined to reveal to us which of these alternatives happens to an individual participant. We must, therefore, rely on some decision-making rule, which is based on observations that we make on the specific person. This leads to a test for response. If the measurements form a certain pattern, we declare a response, and if not then we declare no response.

There are two fundamental quantities that characterize the usefulness of a decision making procedure. Departing somewhat from tradition, I will identify them as:

$\pi_1$  = probability of deciding benefit for a benefited person (sensitivity)

$\pi_0$  = probability of deciding benefit for a nonbenefited person (1-specificity)

The other two critical parameters are

$p_{\text{true}}$  = probability of being truly benefited

$p_{\text{obs}}$  = probability of observed decision of benefit

From a research study, we will be able to estimate  $p_{\text{obs}}$ , but what we want to know is  $p_{\text{true}}$ . These probabilities are linked through the following equation (so-called solution by cases):

$$p_{\text{obs}} = \pi_1 p_{\text{true}} + \pi_0 (1 - p_{\text{true}})$$

The important point here is that the values of  $\pi_1$  and  $\pi_0$  can be computed from param-

eters, and thus estimated. How they can be computed is discussed in the next section. Once they are computed or estimated, we can use

$$p_{\text{true}} = \frac{p_{\text{obs}} - \pi_0}{\pi_1 - \pi_0}$$

In point of fact, this is a theoretically correct equation that we intend to use for estimation; specifically, to convert an estimate of  $p_{\text{obs}}$  into an estimate of  $p_{\text{true}}$ . It is theoretically impossible for the population value of  $p_{\text{obs}}$  to fall below  $\pi_0$  or above  $\pi_1$ , but it is entirely possible for sample estimates of  $p_{\text{obs}}$  to fall outside of this range. Therefore, when  $p_{\text{obs}}$  falls below its theoretical lower bound we set our estimate of  $p_{\text{true}}$  to 0, and when  $p_{\text{obs}}$  falls above its theoretical upper bound, we set our estimate of  $p_{\text{true}}$  to 1. This is not illegitimate, since if everyone in the population were benefited, then  $p_{\text{true}}$  would be 1 and  $p_{\text{obs}}$  would be  $\pi_1$ , while if every one were not benefited then  $p_{\text{true}}$  would be 0 and  $p_{\text{obs}}$  would be  $\pi_0$ .

## CONSTRUCTING A DECISION RULE

Each data collection design leads to its own form of decision rule, but for purposes of exposition we will restrict ourselves here to a simple, but commonly occurring situation. We make a number of measurements (generically denoted  $x_0$ ) before the intervention period, and another collection (denoted  $x_1$ ) either during or after the intervention period. The usual model for such measurements states

$$x_{0i} = \xi + e_{0i}$$

$$x_{1i} = \xi + \delta + e_{1i}$$

Here,  $i$  is a subscript to track multiple measurements on the same person. Conventionally,  $\xi$  stands for the average measurement for this individual if he/she were put into the control group. The value  $\delta$  stands for the individual's response to the treatment they are given. For an individual in the control group,  $\delta = 0$  as a matter of definition. For one in the treatment group,  $\delta = 0$  means that the treatment is equivalent to the control, while conventionally  $\delta > 0$  indicates a benefit. Within-person models of

this sort have been discussed in the statistical literature on causal interpretations of linear regression (Pratt and Schlaifer, 1984). The  $e$ -terms in the model refer to effects caused by the fact that the measurements are made at different times, and perhaps under slightly different circumstances. They are intended to capture the accumulation of effects that are separate from the constitutional response of the individual.

We may also note that this framework is broad enough to include placebo effects (Guess et al., 1997; Harrington, 1997; Shapiro and Shapiro, 1997), in which case the control group might have essentially nothing (other than minimal observation) and the "treatment" group would be administered the placebo. There is no real difficulty in expanding this framework to include comparisons of "nothing," "placebo," and "active treatment" all in the same study.

In the conventional statistical model, the  $\xi$ 's and  $\delta$ 's are considered to be chance variables. This results from the natural assumption that they vary from person to person, and the perhaps rosy assumption that they are in some sense randomly sampled from an underlying population. The intercorrelation among measurements made on the same person is in fact a result of taking  $\xi$ 's and  $\delta$ 's to be chance variables. In the participant-centered approach, however, we are concerned not with the population of individuals, but rather with the population of potential measurements that could be made on the individual before us. Because  $\xi$  and  $\delta$  do not change as we make these measurements, they shift their roles from being chance variables to being person-specific parameters. Our decision-rule, therefore, takes the form of a conventional hypothesis test about the value of  $\delta$ , as a person-specific parameter.

In the simple situation we consider here, a test of  $\delta$  would be based on the difference between means  $\bar{x}_0$  and  $\bar{x}_1$  and measured on the specific person. The variance of this difference is:

$$v = \text{var}(\bar{x}_1 - \bar{x}_0) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}$$

where  $\sigma_0^2$  and  $\sigma_1^2$  are the variances of the  $e_{0i}$  and  $e_{1i}$  terms, and the  $n$ 's are the respective numbers of measurements, made pretreatment and

posttreatment. This quantity is estimated by plugging in the sample estimates for the  $\sigma$ 's on the right-hand side. These estimates and  $n$ 's all pertain to within-individual data. This formula assumes that the measurements are made far enough apart that the  $e$ 's are uncorrelated. When this assumption fails, there are methods for adjusting the variance estimates, but we do not pursue this issue here. (There are also other potential estimates, and degrees-of-freedom adjustments for cases where the two variances differ, but again for simplicity we do not include them here.)

Consequently, we can compute a conventional  $t$  statistic ( $t$ ) based on differences between means. We will take  $\delta = 0$  to signal the absence of a response, and positive values of  $\delta$  to measure a benefit. Taking the former as the null hypothesis, it is a textbook exercise to determine the critical value  $\tau$  such that the rule "reject  $\delta = 0$  when  $t \geq \tau$ " satisfies

$$\text{pr}(t \geq \tau: \delta = 0) = \pi_0$$

Recall from the preceding section that  $\pi_0$  is the probability of declaring a response when in fact it does not exist, and in conventional statistics it would be called the probability of a Type I error. Again following classic methods, we need to select a value  $\delta_{\text{resp}}$  such that when  $\delta \geq \delta_{\text{resp}}$  we have a response. In other words,  $\delta_{\text{resp}}$  is our minimal criterion for a response. It then follows once more in textbook fashion that

$$\begin{aligned} \text{pr}(t \geq \tau: \delta = \delta_{\text{resp}}) \\ = \text{pr}\left(t \geq \tau - \frac{\delta_{\text{resp}}}{\sqrt{v}}: \delta = 0\right) = \pi_1 \end{aligned}$$

Now  $\pi_1$  is the probability of declaring a response when it equals our minimal criterion, and in conventional statistics it would be called the "power" of the test. Since  $v$  is inversely related to both  $n_0$  and  $n_1$ , it follows that for a fixed  $\delta_{\text{resp}}$  the only way to achieve a prespecified  $\pi_1$  is by increasing the sample sizes.

So far, all we have done is slightly repackage one of the most elementary hypothesis test examples from beginning statistics texts. The only novelties are that we have taken the parameter of interest to pertain to the population of measurements from an individual (rather than a



population of individuals), and the rejection of the usual null hypothesis is interpreted as evidence of an individual-level response.

It should be clear at this point that designing the analysis involves setting  $\pi_1$  convincingly high and  $\pi_0$  convincingly low, so that the decision procedure actually distinguishes responders from nonresponders. It also involves selecting  $\delta_{\text{resp}}$  high enough to be considered by disciplinary specialists as a true response, but low enough to permit some actual response decisions, given the available within-person sample sizes. In the next section we consider in an actual data example how this can be done.

## LOWERING BLOOD PRESSURE BY DIET

The data we will use are from a study undertaken among moderate hypertensive individuals, to assess the effect of a diet high in fruits and vegetables (FV), and the effect of a diet additionally high in protein, low-fat dairy and fiber (Dietary Approaches to Stop Hypertension [DASH] diet), compared to the typical American diet (Appel, 1997). For our example, we use systolic blood pressure (SBP) measured during the first 2 weeks of the study (prior to the diet interventions), and then during the last 2 weeks (after 8–9 weeks on intervention). We will use as our measure of benefit the mean SBP at the start of the study minus the mean at the end, so that positive values indicate more lowering of SBP.

Keep in mind that there are four quantities involved in the crafting of the decision rule;  $\pi_0$ ,  $\pi_1$ ,  $\delta_{\text{resp}}$ , and  $v$ . In general, these values can vary from person to person, but we will want to make selections that yield interpretable results. We will always fix  $\delta_{\text{resp}} = 8$  mm Hg, which most physicians would regard as meaningful. Obviously, other choices could be made.

### *Constant variance approach*

In this method, we simplify the situation by using a common estimate of  $v$  for all participants, obtained by averaging their individual  $v$ 's. This departure from our within-individual perspective is purely for the convenience of having fixed values of  $\pi_0$  and  $\pi_1$ , since (as we

will see) using individual-level variances forces either  $\pi_0$  or  $\pi_1$  to vary between people. For the 459 study participants, we estimate  $v = 4.18$  (mm Hg). Further simplifying for ease of exposition, use the normal approximation to the  $t$  distribution, so that the rule “declare response if  $t \geq 0.84$ ” amounts to setting  $\pi_0 = 0.20$ . We can then compute  $\pi_1 = 0.86$  by the formula in the preceding section. (We could, of course, have fixed  $\pi_1$  in advance, and allowed  $\pi_0$  to assume whatever value it did.) The results are shown in Table 1. On the usual diet we estimate that probability of true improvement is 0.11, while it is 0.48 on FV and 0.59 on DASH. The actual numbers of responders were 42, 79, and 89. Each of these individuals was judged to have improved by a decision-rule with known properties.

### *Constant $\pi$ approach*

A more complex but more realistic approach is to estimate each individual's  $v$  separately. We can then fix either  $\pi_0$  or  $\pi_1$ . In the first case, using the same rule as before (“responder if  $t \geq 0.84$ ”) of course  $\pi_0 = 0.20$ . Now  $\pi_1$  varies from person to person, depending on their value for  $v$ . The results of the rule and the summary statistics for  $\pi_1$  are shown in Table 2.

For this table, we can estimate  $p_{\text{true}}$  by computing  $(y - \pi_0)/(\pi_1 - \pi_0)$  where  $y$  is 0 for a non-responder and 1 for a responder (and the  $\pi$ 's pertain to the individual), and then averaging these values over a treatment group.

The results are qualitatively similar to Table 1, with some additional separation between FV and DASH. Note that the mean (or median) of  $\pi_1$  was approximately 0.88, and that in every case it exceeded 0.5. This approach fixes the

TABLE 1. ANALYSIS USING PERSON-INDEPENDENT SD

Diet group	Number of patients	Responders	$p_{\text{true}}$
Usual	154	42 (0.27)	0.11
FV	154	79 (0.51)	0.48
DASH	151	89 (0.59)	0.59

Results for the analysis that uses a person-independent standard deviation (SD) and, therefore, fixed sensitivity = 0.86 and specificity = 0.80.

FV, fruits and vegetables; DASH, Dietary Approaches to Stop Hypertension diet.

TABLE 2. ANALYSIS USING PERSON-DEPENDENT SD WITH FIXED SPECIFICITY

Diet group		Number of patients			Responders			$p_{true}$
Usual		154			44 (0.29)			0.08
FV		154			78 (0.51)			0.43
DASH		151			99 (0.66)			0.68
Variable	n	Mean	SD	Minimum	0.25	Median	0.75	Maximum
$\pi_1$	459	0.87	0.11	0.51	0.81	0.89	0.96	1.00

Results for the analysis that uses a person-dependent standard deviation (SD) but fixes specificity at 0.80.  
FV, fruits and vegetables; DASH, Dietary Approaches to Stop Hypertension diet.

probability of a false response, and thus can only guarantee the probability of a true response through interpretation of the descriptive statistics for  $\pi_1$ . The other approach is to set  $\pi_1 = 0.80$  for everyone, which allows  $\pi_0$  to float. This leads to the rule “responder if  $t \geq 8/\sqrt{v} - 0.84$ .” The results are shown in Table 3, which are also similar to Table 1.

#### Allow ambiguous responses

The last method is a hybrid of the method of the previous subsection. To each individual we apply both the decision rule (of the form  $t \geq 0.84$  as above) which is engineered to have  $\pi_0 = 0.20$ , and also the rule ( $t \geq 8/\sqrt{v} - 0.84$  as above) which fixed  $\pi_1 = 0.80$ . This gives two decisions, by different rules, for each person. Our final decision is the average of the two. Because they are coded 0/1, an average of 0 means no response by either rule, a 1 means a response by both rules, and 0.5 means the rules conflict. Table 4 shows the results. Both the responders and nonresponders have been validated by two rules that have desirable properties. The fact that the ambiguities occur equally across the three groups provides support for

the idea that the responders can be compared between treatments, and likewise for the non-responders. Put differently, if one group had an excessive number of ambiguous cases, relative to the other groups, this would perhaps indicate less precision of the within-person measurements in that group, raising issues about the design or conduct of the study. This basic kind of approach has been advocated in conventional statistical decision-making, in order to simultaneously lower the two error probabilities, at the expense of producing some non-decisions (Mossman, 1999).

In order to compute the  $p_{true}$  column of Table 4, it is useful to describe the hybrid test in a slightly different way. Define  $\tau_{max}$  to be the larger of 0.84 and  $8/\sqrt{v} - 0.84$ ,  $\tau_{min}$  to be the smaller of the two, and observe that the decision rule is

responder if  $t \geq \tau_{max}$

nonresponder if  $t \leq \tau_{min}$

ambiguous if  $\tau_{min} < t < \tau_{max}$

From the first line it is possible to compute  $\pi_0$  and  $\pi_1$ , as we have done above, and now both

TABLE 3. ANALYSIS USING PERSON DEPENDENT SD WITH FIXED SENSITIVITY

Diet group		Number of patients			Responders			$p_{true}$
Usual		154			27 (0.18)			0.07
FV		154			60 (0.39)			0.41
DASH		151			78 (0.52)			0.59
Variable	n	Mean	SD	Minimum	0.25	Median	0.75	Maximum
$\pi_0$	459	0.13	0.11	0.00	0.04	0.11	0.19	0.49

Results for the analysis that uses a person-dependent standard deviation (SD) but fixes specificity at 0.80.  
FV, fruits and vegetables; DASH, Dietary Approaches to Stop Hypertension diet.

TABLE 4. ANALYSIS USING TWO CONSTANT  $\pi$  METHODS

<i>Diet group</i>	<i>Nonresponders</i>		<i>Ambiguous</i>		<i>Responders</i>		$p_{true}$	
Usual	107 (0.69)		23 (0.15)		26 (0.16)		0.07	
FV	72 (0.47)		26 (0.17)		56 (0.36)		0.39	
DASH	52 (0.34)		21 (0.14)		78 (0.52)		0.63	
<i>Variable</i>	<i>n</i>	<i>Mean</i>	<i>SD</i>	<i>Minimum</i>	<i>0.25</i>	<i>Median</i>	<i>0.75</i>	<i>Maximum</i>
$\pi_0$	459	0.11	0.07	0.00	0.04	0.11	0.20	0.20
$\pi_1$	459	0.78	0.05	0.51	0.80	0.80	0.80	0.80

Results when the two constant  $\pi$  methods are combined, allowing for disagreement between the two rules.  
FV, fruits and vegetables; DASH, Dietary Approaches to Stop Hypertension.

these values will vary between persons. The distributions are in Table 4, showing that no value of  $\pi_0$  exceeded 0.20, and while a few values of  $\pi_1$  fell below 0.80, the average was still 0.78. The advantage of opening a little window for ambiguous decisions is therefore to lower the probability of a false positive, consistent with conventional statistical attitudes toward the null hypothesis of no effect. We now proceed as in Tables 2 and 3, to form  $(y - \pi_0)/(\pi_1 - \pi_0)$  for each individual, and then average these within each treatment group to estimate the corresponding  $p_{true}$ . By this approach the response probability in the control group is a quite low 0.07, rising to 0.39 in the FV group and to 0.63 in the DASH group.

A similar exercise can be carried out for computing the true probability of nonresponse. We may be interested to note that this will show that something like 1 of 3 of the DASH participants failed to demonstrate a response, a fact of some importance in interpreting the significance of the results for individuals.

Since this concludes the blood pressure example, it may be worthwhile to point out that the selection of  $\pi_0$  near 0.20 was made precisely so that  $\pi_1$  would be near 0.80, reflecting the attitude that neither of sensitivity or specificity was more important than the other. How extreme one can make these two probabilities depends on the within-person variability of the results, and also on the number of within-person measurements. In this case the measurements were fixed by conventional design considerations, but with a participant-centered design one might want to raise both the number of premeasurements and postmeasurements.

## DISCUSSION

The purpose of this short paper has been to introduce the idea of participant-centered analyses of CAM comparative trials. It should be obvious that this is only a beginning, and that this approach offers many statistical and conceptual challenges. To note just one, raised by a referee, since the approach requires multiple measurements per participant, it does not appear immediately applicable to one-time definitive disease diagnoses, nor to death as an endpoint.

The example given here might seem to have been excessively complicated. The reason for this is, however, that it showed several different approaches, in order to explicate the decisions that need to be made in fashioning a participant-centered analysis. In practice, one would debate the issues among the scientific investigators, and come up with a single approach before the analysis. One of the intriguing features is that the final published results could then consist of a relatively simple table, similar to the ones shown above, with responder status and estimates of true responder probabilities. The actual form of the technical analysis, which might be based on means, on proportions, on rates, or on more elaborate statistical within-person models, would not be part of the results section in the publication, but rather part of the methods section. This might have the salutary effect of lifting methods sections out of the jejune ritual into which they have fallen.

Participant-centered analysis has an additional benefit, both for CAM and allopathic

clinical research. Virtually all approaches to biomedical studies these days apply to population parameters, and not to individual people. The attempt to find out who benefits and who does not leads to a much maligned subgroup analysis, with subsequent loss of statistical power and heightened suspicions of data-dredging. By defining the method of analysis so that it makes individual-level decisions about benefit (or harm), one creates a new variable that can then be used to try to explain their occurrence, using modern exploratory statistical methods. It goes one step further, however, by allowing detailed investigation of specific cases, in a search for unanticipated factors that might promote a favorable (or harmful) response. This promises a large step forward in doing CAM research focused on which modality should be recommended to an individual, instead of trying to make decisions about which modality should be recommended to everyone.

The dietary/blood pressure example given here was framed entirely in terms of benefit or no benefit. There are, however, no technical barriers to taking a wider decision-theory perspective, not only to allow harm as a possible decision, but also to permit decisions about various types of benefits or harms, or combinations thereof. Because many allopathic therapies offer benefits only at the risk of serious side effects, the participant-centered approach might be of special benefit in analyses that consider the personal and policy aspects of a proposed treatment, rather than concentrating obsessively on what might be a very small, though detectable, average effect.

An elaboration on this theme applies in cases where multiple different kinds of variables are measured. Several measures of various kinds of benefits could be acquired along with other measures of side-effects, harms, or costs. The issue then becomes one of deciding which patterns of benefits/harms/costs constitute an overall good response. Although it is traditional to form linear combinations of such measures in order to force the analysis back into the univariate mold, this is not actually necessary. The statistical problem then becomes to fashion within-person tests of good response

based on multivariate data, a challenging but not insuperable problem.

Another point, which may seem obvious, is that participant-centered analysis places a heavier measurement burden on individual participants. At the risk of injecting a personal note, I will say that many of the studies in which I have participated have spent excessive resources in collecting information that was at best peripheral to the purpose of the study, and that often these additional stores of information did not lead to subsequent publications. I would surmise that if those resources had been turned back to a more intensive set of measurements on truly relevant outcomes, the participant might have been better informed about his/her condition and progress, and we might have a better information base on the details of the fluctuations and progressions of disease processes, and their variations among ordinary people. Even in the absence of a participant-centered analysis, these might be good things, but with a participant-centered analysis there is a scientific rationale for them.

It is important to realize that the perspective put forth here is a fundamental departure from standard methods. Certainly one has a natural tendency to expect some kind of hypothesis test or confidence interval to accompany the data in the tables that were presented above, but this desire runs contrary to the spirit of participant-centered analysis. The issue is not to make a decision whether the therapy in question had a detectable effect in some overall sense, but to report how many people were benefited, as decided by a classic Neyman-Pearson hypothesis test applied at the individual level. The appropriate time to proceed to a population hypothesis test is either in a meta-analysis, or in the case of a truly large, definitive trial.

The converse is that a participant-centered analysis provides a new role for small, Phase I and Phase II studies. Even small studies, when they are well conducted, should find their place in the research literature, but conventional editorial and refereeing practice favors the trial with positive results, or with high power if the results are negative. It is in the nature of conventional analysis that small studies will seldom have either of these characteristics. This



applies pressure to conduct large, simple trials, in which the available resources frequently compel the researchers to restrict their study to two, or at most three comparative groups. It also channels research funding to large, well-equipped research facilities, or expensive multisite trials, leaving the less well-endowed research groups out in the cold. This illustrates how allopathic methods of analysis unintentionally drive health research policy in a direction that hinders CAM researchers.

As a cautionary note, participant-centered analysis should not be taken to mean a return to the biomedical methods of previous centuries, in which "cures" of individual patients were asserted casually and somewhat haphazardly in support of allopathic therapies that were, with some regularity, ultimately found to be worthless. Participant-centered analysis is also unrelated to the practice, widespread in cancer therapeutics, of declaring complete, partial, or nonresponders, on the basis of a single or small number of clinical measurements, with no underlying statistical decision procedure. It is, instead, a necessary piece of participant-centered analysis that it involves the collection of a sufficient amount of data on each individual to permit the application of traditional statistical decision-making at the individual level. Usually this entails multiple measurements of an underlying variable at a course of time points, or under different conditions. This is to ensure that the declaration of a response is always justified by a rule with known decision-theoretic properties, giving us some confidence that it represents what the therapy actually did for the participant.

## ACKNOWLEDGMENTS

I would like to thank my colleagues Mitch Haas and Cheryl Ritenbaugh for very helpful substantive comments on earlier drafts of this article, and also five referees for making useful comments leading to changes in the article.

This research was partially supported by a grant from the National Center for Complementary and Alternative Medicine, AT00956.

## REFERENCES

- Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, Lin PH, Karanja N. A clinical trial of the effects of dietary patterns on blood pressure. *N Engl J Med* 1997;336:1117-1124.
- Byrk AS, Raudenbush SW. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications, 1992.
- Guess HA, Kleinman A, Kusek JW, Engel LW, eds. *Science of the Placebo*. London, UK: BMJ Press, 2002.
- Harrington A, ed. *The Placebo Effect*. Cambridge MA: Harvard University Press, 1997.
- Hochberg Y, Tamhane AC. *Multiple Comparison Procedures*. New York NY: John Wiley & Sons, 1997.
- HMS/UCSF, International Scientific Conference on Complementary, Alternative & Integrative Medicine Research. University of California San Francisco, May 17-19, 2001.
- HMS, International Scientific Conference on Complementary, Alternative & Integrative Medicine Research, Harvard University, April 12-14, 2002.
- Jonas WB. Researching alternative medicine. *Nat Med* 1997;3:824-826.
- Levin JS, Glass TA, Kushi LH, Schuck JR, Steele L, Jonas WB. Quantitative methods in research on complementary and alternative medicine: A methodological manifesto. *Med Care* 1997;35:1079-1094.
- Mossman D. Three-way ROCs. *Med Decis Making* 1999; 19:78-89.
- Pratt JW, Schlaifer R. On the nature and discovery of structure. *Journal of the American Statistical Association* 1984;79:9-21, commentary 22-33.
- Raudenbush S, Byrk A, Cheong Y-F, Congdon R. *HLM5: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood IL: Scientific Software International, 2000.
- Shapiro A, Shapiro E. *The Powerful Placebo*. Baltimore, MD: The Johns Hopkins University Press, 1997.
- Vickers A. Methodological issues in complementary and alternative medicine research: a personal reflection on 10 years of debate in the United Kingdom. *J Altern Complement Med* 1996;2:515-524.
- Vickers A, Cassileth B, Ernst E, Fisher P, Goldman P, Jonas W, Kang S-K, Lewith G, Schultz, Silagy C. How should we research unconventional therapies? A panel report from the Conference on Complementary and Alternative Medicine Research Methodology, National Institutes of Health. *Int J Technol Assess Health Care* 1997;13:111-121.

Address reprint requests to:

*Mikel Aickin, Ph.D.*

*Center for Health Research*

*Kaiser Permanente Northwest Region*

*3800 N. Interstate Avenue*

*Portland, OR 97227-1098*

*E-mail: mikel.aickin@kpchr.org*

Copyright of Journal of Alternative & Complementary Medicine is the property of Mary Ann Liebert, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.