THE JOURNAL OF ALTERNATIVE AND COMPLEMENTARY MEDICINE
Volume 10, Number 4, 2004, pp. 698–705

One Size Does Not Fit All: Aptitude *x* Treatment Interaction (ATI) as a Conceptual Framework for Complementary and Alternative Medicine Outcome Research. Part II—Research Designs and Their Applications

OPHER CASPI, M.D., Ph.D., and IRIS R. BELL, M.D., Ph.D.

ABSTRACT

When multiple treatment choices are available, the question is not just "which treatment is the best?" but more importantly "best or better for whom, when, and why?" Aptitude (or attribute) by treatment interaction (ATI) is a research paradigm that attempts to examine exactly that—how outcome depends on the match or mismatch between patients' specific characteristics and the treatments they receive. The purpose of this two-part paper is to introduce ATI methods as a conceptual framework into complementary and Alternative medicine/integrative medicine (CAM/IM) outcome research. Part I presented key concepts in ATI research. Part II presents ATI research designs and discusses their applications to the examination of the relationships between individuals and therapies, and the illumination of the mechanisms that make therapies differentially effective. Based on this examination, we conclude that ATI research offers invaluable insights into the multifaceted package of care typically delivered in contemporary medicine and therefore should be included in the portfolio of all CAM/IM outcome research.

INTRODUCTION

In Part I of this paper, we laid out the foundations for ATI research. We discussed what is meant by 'aptitude,' 'treatment,' and 'interaction,' and explained how the question "do interventions work?" needs to be reframed as "which interventions work, for whom, and under what conditions?" In Part II, we discuss various ATI research methods and their applications, and use the example of a recent meta-analysis of antidepressants and the placebo effect to illustrate the importance of ATI research within the context of outcome research. We conclude that including ATI research in the portfolio of all outcome research is imperative for the future of evidence-based medicine.

ATI RESEARCH SHOULD STRIVE TO BE CONFIRMATORY RATHER THAN MERELY EXPLORATORY

Correlational, rather than causal, analysis poses a significant threat to outcome research since patient outcomes may be correlated with treatments and aptitudes for a variety of reasons irrespective of the treatment given (Ridenour et al. 1999). Nonetheless, both outcome and ATI research have been largely correlational and exploratory thus far. In part, this might be because measuring and interpreting interaction effects is much more difficult than dealing with main effects (Rosnow and Rosenthal, 1989; 1991; 1995). For example, at least 175 different categories of patient aptitudes,

Program in Integrative Medicine, University of Arizona, Tucson, AZ.

40 different categories of provider characteristics, and 50 different technique categories have been proposed as mediators in psychotherapy research (Garfield and Bergin, 1986). Thus, there are nearly 1.5 million potential combinations of therapy, therapist, phase, and patient types that must be studied in order to rule out relevant differences among treatment types (Beutler, 1991). This is, of course, impossible. Instead, for ATI studies to be feasible and meaningful, one needs to hypothesize a-priori what it is about a patient that interacts with treatment(s) to differentially affect outcomes. Put another way, ATI research should be driven by plausible hypotheses closely tied to clinical theory rather than simply being a "hit-or-miss" "fishing" expedition fueled by spurious statistical associations.

But how can that be done? After all, individuals are complex amalgamations of interrelated and interacting characteristics. The same is true for multidimensional packages of care. This means that a-priori justifications need to be made in order to determine how many and which aptitudes and components of treatment to consider in ATI research. Studying one aptitude or therapy at a time while ignoring others may result in an unwarranted oversimplification. As Snow warned:

> Indeed, the recognition that aptitude variables, in particular, should not be considered only one at a time is an important first lesson. The world of person characteristics abounds in correlations, and it is unlikely that one aptitude effect is isolated from others. Every research design involves multiple aptitudes and higher order interactions whether it includes them formally or not" (Snow, 1991).

Yet, studying too many aptitudes or treatment components simultaneously can result in research findings that are almost impossible to interpret (Smith and Sechrest, 1991; Stiles et al., 1986). Furthermore, in the absence of theory that provides a compelling reason as to why a particular interaction should be searched for, serendipitous interactions remain exactly that, and there is little reason to expect them to be replicated (Shoam-Salomon and Hannah, 1991). As Beutler (Beutler, 1991) pointed out, the post-hoc methodology of looking for correlations further ensures that the patient dimensions that are explored as interaction variables are those that are convenient and available rather than those that have theoretical significance to understanding the conditions of differential efficacy.

Shoam-Salomon and Hannah suggested that, as a guiding principle, one would want to include as aptitudes those variables that, on the one hand, are most pervasive while, on the other hand, are differentially predictive of outcomes (Shoam-Salomon and Hannah, 1991). Following the principles of convergent and discriminant validity, this means that the aptitude should be strongly associated with the outcome for one intervention and not be associated with the outcome of other interventions (Campbell and Fiske, 1959). It is also desirable for each of the variables to correlate with an outcome specific to one treatment and not another and to be correlated as little as possible with one another (Bogden, 1951; Snow, 1991). However, this is more easily said than done since many of these variables are only proxies for more meaningful constructs and tend to influence and be influenced by therapeutic variables.

ATI RESEARCH DESIGNS

Several research and data-analysis methods have been identified as appropriate for ATI research, although none in particular are exclusive to ATI research. What makes them suitable for ATI research is the fact that they allow the researcher to explicitly test for the possibility that one or more aptitudes moderate or mediate one or more outcomes through an interaction with one or more treatments. In a previous paper, we discussed important aspects of complex systems multivariate outcome research that are vital to consider when designing ATI research (Bell et al., 2002). Hereby, we present concisely the three most commonly used methods for ATI research. A more detailed review of these and other methods is provided by Snow (Snow, 1991) and Ridenour et al. (Ridenour et al., 1999).

Standard experimental design

The most commonly used design within the ATI research paradigm is the simple randomized, controlled clinical trial in which the outcomes of two or more groups that received the same treatment are assessed with respect to different levels of an aptitude or set of aptitudes. Figure 1 depicts hypothetical results from such a study wherein regressions on outcome for each of three treatments, A, B, and C, differ by

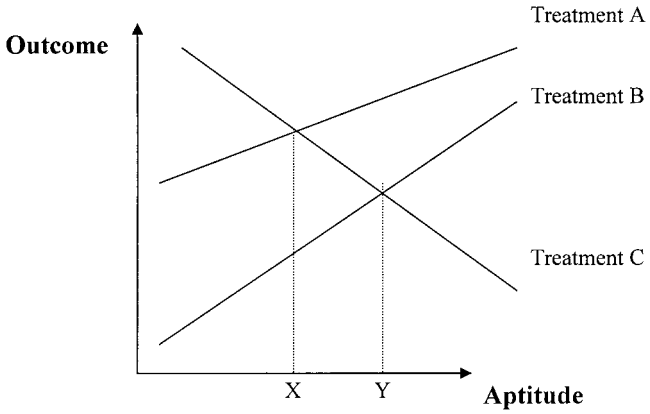

FIG. 1. Hypothetical example for an interaction between treatments and aptitude. (See text for explanation.) (Snow RE. Aptitude–treatment interaction as a framework for research on individual differences in psychotherapy. J Consult Clin Psychology 1991;59:205–216; used with permission).

level of aptitude. To optimize outcome in practice, one would apply treatment A to persons above point X and treatment C to persons below point X, on the aptitude continuum. If the treatments differ in cost or side-effects, decisions may change. For example, if treatment B is less expensive than treatment A, has fewer side-effects, and is similar in effect for persons above point X, then it might be used instead of treatment A; this changes the point of deciding which treatment to administer from X to Y. Formal methods exist for converting outcome variables to utility scales, thus leading to different conclusions in practice (Petitti, 2000). For example, if treatment A is the least expensive and less harmful option, the conversion shifts the treatment A regression line upward, perhaps making it the treatment of choice for all persons in the sample.

Regression discontinuity design

The regression discontinuity design (RD) is especially appropriate for ATI research when randomization is not feasible. In this design, patients are assigned to conditions on the basis of a cutoff score on an assignment variable. The assignment variable in ATI research is any aptitude measure taken prior to treatment. Subjects scoring on one side of the cutoff are assigned to one condition and those on the other side to another. Therefore, the assignment variable must have at least ordinal measurement characteristics, that is, be monotonically increasing such that A>B; true nominal variables such as ethnicity are specifically excluded.

RD designs are quite flexible. For example, if theory suggests a complex model, more than one assignment variable (e.g., more than one aptitude or one aptitude plus another variable, say, medical insurance status) can be used. Likewise, several individual treatments or packages of care can be administered simultaneously and more than one cutoff score can be used—both important features for ATI research what is also important is that RD designs can be combined with classical randomized and nonrandomized or quasiexperimental designs. For example, one can use multiple cutoff intervals with randomization in some intervals and RD in others or use a cutoff interval with RD outside the interval and self-selection within the interval. To avoid ethical concerns, one can use the basic RD design but, at the end of the study, give treatment to all the control participants. (For further discussion of RD design see Shadish et al., 2002).

Figure 2 depicts a hypothetical example of the results obtained from a RD study. A treatment effect will cause an upward or downward displacement in the regression line relating assignment to outcome—either a change in mean where outcomes scores on one side of the cutoff are increased by the amount of the mean effect or a change in slope where the regression line is steeper on one side of the cutoff than the other. This displacement of the mean or slope should occur at exactly the point on the assignment variable where the cutoff score defines the treatment contrast. It is this point-specific displacement (or discontinuity) of the regression line that gives the design its name.

Figure 2A shows the results of a hypothetical study that yielded a main effect for treatment and an interaction between treatment and aptitude. In addition to the discrepant regression lines at the point of the cutoff, the slope of the line to the right of the cutoff is steeper than the one to the left. All treatment participants benefit more than controls, but those with more aptitude do better than those with less aptitude regardless of treatment condition. Therefore, the size of the discontinuity is contingent on where it is measured on the assignment variable. The analysis will correctly yield different estimates of the effect depending on which score is subtracted from the aptitude variable.

Figure 2B, however, shows a shift in slope but no discontinuity at the cutoff score, which is produced by an in-

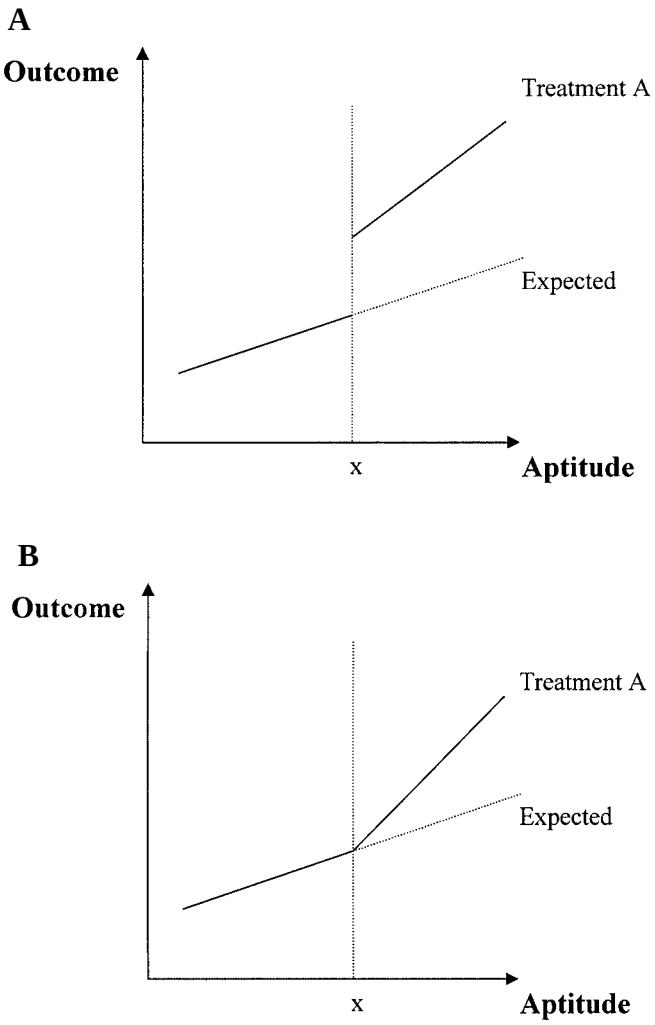

FIG. 2. Regression discontinuity design (modeled after Shadish et al. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin, 2002; with permission). **(A)** Treatment effect and interaction. **(B)** Interaction but no treatment effect.

teraction but no main effect of treatment. This is a complex situation to interpret. Some investigators would conclude that there is no effect because of the lack of a discontinuity at the cutoff, whereas others may conclude that this lack indicates a possible effect among subjects who are far from the cutoff. The latter interpretation has two problems (Shadish, 2002). One is that the logic of the RD design is partly tied to finding discontinuities at the cutoff because participants are most similar there. The other is that it is difficult to distinguish this situation from a nonlinear relationship generated by a quadratic function with no treatment effect. To understand the true nature of the effect in cases like this one better, it is especially beneficial to combine RD with other design features such as randomization (when possible) as described above.

Change curves (or "growth curves")

Change curve analysis refers to a way of representing data to illustrate change in an outcome variable over time. This approach offers a number of advantages that are desirable in the context of ATI research (Figueredo, 2002; Shoam & Rohrbaugh, 1995). First, growth curves, or trajectories of change preserve the data at the individual level. Second, unlike comparative clinical trial designs where suitable control conditions are crucial to demonstrating treatment effects, growth-curve analysis does not necessarily require them.

Instead, subjects are able to serve as their own controls. This allows examining how the shape or topography of change may vary with aptitudes and treatments so as to determine whether systematic individual differences exist in response to particular treatment. Third, it is not necessary to obtain measures on individuals at the same times nor even the same number of time points per individual, rendering the methodology relatively robust against problems involving missing data, a common challenge in much of outcome research. Fourth, it has been suggested that the change-curve methodology requires smaller sample sizes than between-groups controlled studies. Finally, the technique permits a meta-analytic approach in which the information from a whole set of individual change curves is combined so as to provide a conceptually simple and analytically powerful tool for dissecting the complexities of change over time. After each individual's raw outcome data have been plotted over time and modeled to a best-fitting curve, the researchers can group individual curves into change patterns or types according to any number of parameters, including curve shape, slope, intercept, and dispersion, depending on the specific data and questions in hand. (For a detailed review of the meta-analytic approach to change curves, see Figueredo et al., 2002).

As Shoahm and Rohrbaugh have pointed out, the change-curve methodology allows researchers to relate change types to meaningful interindividual differences by asking three basic ATI questions: (1) Do treatments differ from each other in the typical change curves they produce? (2) Is there a systematic difference in change curve among groups? (3) Do systematic differences covary with theoretically meaningful

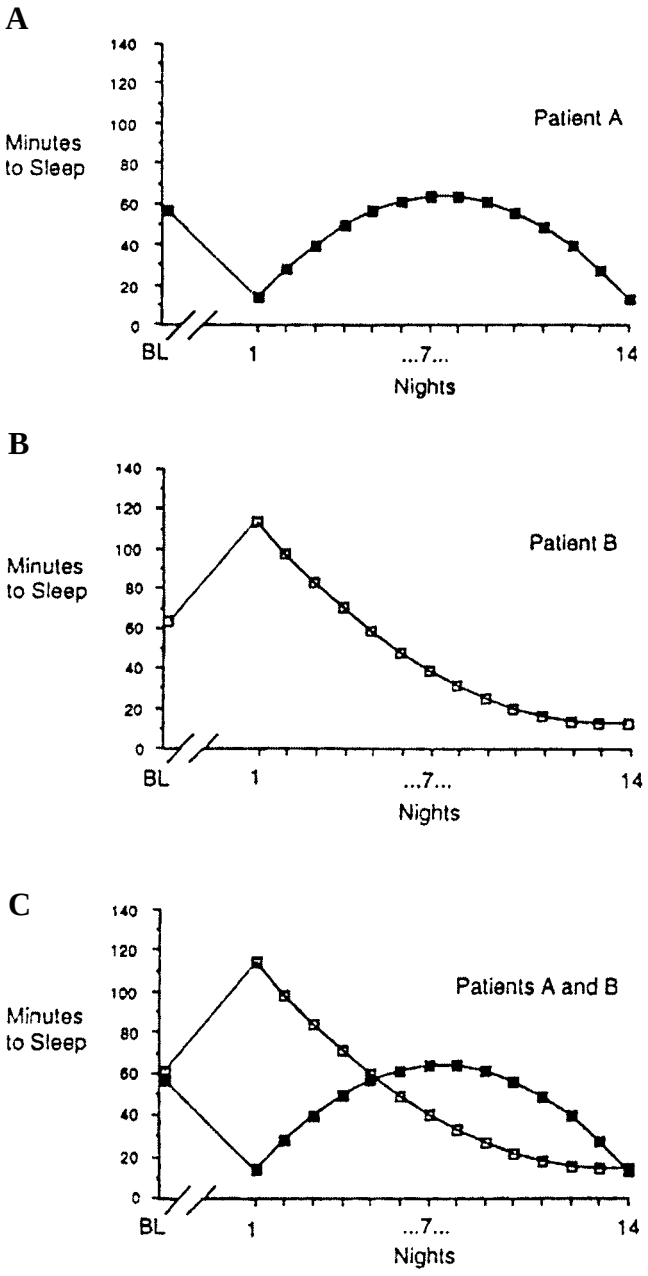

FIG. 3. An example for the heuristic value of change curves in aptitude by treatment interaction research. Change curves for two pilot patients receiving paradoxical intervention for insomnia. The first (BL) data point shows the mean sleep latency for 14 baseline nights; the remaining data points represent the best-fitting curve for sleep latencies on the 14 nights immediately following a paradoxical intervention (Shoam and Rohrbaugh, Aptitude x treatment interaction (ATI) research: Sharpening the focus, widening the lens. In: Aveline M, Shapiro DA, eds. Research Foundation for Psychotherapy Practice. Sussex, England John Wiley & Sons, 1995:73–95; used with permission).

moderating and mediating aptitude variables? (Shoam & Rohrbaugh, 1995). Accordingly, Figure 3 depicts a useful example for the heuristic value of the growth curve methodology in ATI research.

POTENTIAL SHORTCOMINGS OF ATI RESEARCH

Any attempt to study interactions among patients, providers, and treatment types is confronted with a number of major epistemological and methodological problems (Beutler, 1991; Shoam-Salomon & Hannah, 1991). As Cronbach put it: "Once we attend to interactions, we enter a hall of mirrors that extends to infinity. However far we carry our analysis—to third order or fifth order or any other untested interactions of a still higher order can be envisioned" (Cronbach, 1975).

One problem that we already addressed is the unwieldy number of patient and treatment variables that potentially interact with one another. We suggested that striving toward hypothesis-driven ATI research rather than using a merely exploratory approach should practically remit this problem. That is, an exploratory, correlational-based approach is only justified in the absence of theory and should always be followed by confirmatory research.

However, an equally disturbing problem is qualitative nature rather than quantitative. That is, many theoretically important aptitudes lack either consensual or consistent meaning (Beutler, 1991). As Anderson noted: "No uniform taxonomy is currently used to guide definition or description of personality variables in ATI research, despite the obvious need" (Anderson, 1998). This lack of clarity not only makes it increasingly difficult to interpret and integrate findings based on different constructs with unknown covariation with other constructs but also reduces the ability to merge studies that use different operations of the same constructs using techniques such as meta-analysis (Anderson, 1998). Therefore, aptitudes should be clearly defined and be explicated a-priori rather than explored in post hoc fashion.

The same vagueness is true with respect to what is meant by treatment within the framework of ATI research. We join other ATI researchers in warning against making assumptions such as (1) all treatments are alike, or (2) a treatment has been given, therefore the treatment's goal has been achieved. Rather, careful monitoring of treatment integrity and intensity are needed. Likewise, if one hypothesizes that a given treatment achieves its goal through the activation of processes such as expectancy, quality of the therapeutic alliance, et cetera, then how these mediating variables have been changed or manipulated as a result of the intervention needs to be measured directly (Aronson and Carlsmith, 1968; Smith and Sechrest, 1991).

Another important obstacle to the widespread use of ATI research are issues of statistical power and optimal sample size. Since interaction effects need to be shown to occur above and beyond the additive influence of main effects, some ATI designs require studies with large sample sizes and at least two different treatments, as explained above. Thus, compared to the search for main effects, research on interactions requires better measurement, more subjects, a wider variety of conditions, and specific a priori hypotheses (Smith and Sechrest, 1991). Careful selection of aptitude variables is crucial since each additional variable can exponentially increase the sample size needed to adequately control for Type II errors of all main effects and higher-order interactions that are being considered (Anderson, 1998; Edwards and Cronbach, 1952). (For a more detailed discussion on issues related to statistical power in ATI research see Aguinis, 1995; Cronbach and Snow, 1977; Dance and Neufeld, 1988; Gangestad and Snyder, 1985; Smith and Sechrest, 1991; Snow, 1991).

OUTCOME RESEARCH: FROM GROUP AVERAGES TO INDIVIDUAL DIFFERENCES

To illustrate the need for the changing face of outcome research that we advocate here, consider the recent analysis by Kirsch et al. on antidepressants and the placebo effect (Kirsch, et al. 2002). This analysis was based on efficacy data submitted to the U.S. Food and Drug Administration for approval of the six most widely prescribed antidepressants approved between 1987 and 1999. The researchers found that approximately 80% of the response to medication was duplicated in placebo control groups, and the mean difference between drug and placebo was approximately 2 points on the 17-item (50-point) and 21-item (62-point) Hamilton Depression Scale. These findings led the authors to conclude: "If drug and placebo effects are additive, the pharmacological effects of antidepressants are clinically negligible. If they are not additive, alternative experimental designs are needed for the evaluation of antidepressants."

We believe that ATI research is exactly what these "alternative experimental designs" need to be like in response to the inherent problem associated with all outcome research that is based on contrast of group averages. That is, different patients respond to different treatments and therefore "average" effects may underestimate the benefits derived by those patients who do respond to a given treatment. As Hollon et al. (2002), put it:

> Indices based on average effects presume that each member of a population receives an equal amount of benefit from each constituent component of the intervention. If only some members of a population benefit from a given component, then average effects that appear trivial could underestimate specific effects that are clinically meaningful for some groups of individuals. If that is the case, then differences between ac-

> tive medications and pill–placebo controls should be larger for more responsive patients and minimal or nonexistent for less responsive patients (an average effects model would predict constant differences of trivial magnitude across the whole range of patients) . . . Summary statistics based on the arithmetic mean may be prone to being misinterpreted when there is variability in differential response. Under such circumstances, the shape of the distributions will be different for different conditions, and effect sizes based on categorical response (probability) may be more informative than those based on average response (magnitude). Our sense is that such moderation is likely a consequence of individual differences between patients; however, it could also reflect variation in treatment implementation. Regardless of its source, such moderation needs to be taken into consideration when determining whether a given medication has a true pharmacological effect.

We could not agree more. Patients vary in their response to treatment, whether active or placebo. For example, Ribeiro et al. (1993), have shown that about half of severely depressed patients secrete excessive levels of cortisol as measured by the dexamethasone suppression test. These patients fare well with antidepressants (in fact, these patients do slightly better than suppressors do) but they have a singularly low placebo response rate (approximately 10%) whereas normal suppressors have a slightly higher than usual placebo response rate. More recently, Leuchter et al. (2002), used quantitative electroencephalography and cordance to measure differences in brain function between 51 depressed subjects receiving either active medication or placebo responders receiving either active medication or placebo. The researchers were able to show that placebo and active drug responders were indistinguishable clinically, yet, their brain function was markedly different. For example, placebo responders showed a significant increase in prefrontal cordance starting early in treatment, an effect that was neither seen in medication responders (who showed decreased cordance) nor in medication nonresponders or placebo nonresponders (who showed no significant change). These and other findings lend further support to our quest to shift the focus of outcome research from group averages to individual differences.

CONCLUSIONS AND IMPLICATIONS

A recent report by the Institute of Medicine, Washington, DC, begins by acknowledging that "between the health care that we now have and the health care that we could have lies not just a gap, but a chasm" (Institute of Medicine, 2001). The report lays out a number of rules for the redesign of health care among which are: "Providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit" and "[p]roviding care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions" (Institute of Medicine, 2001). While the former suggestion converges with the commonsense assertion that "different folks benefit from different strokes" (Shoam-Salomon, 1991), the latter mentions a number of aptitudes (i.e., preferences, needs, and values) that are likely to interact with treatment to moderate outcome(s), according to the ATI research paradigm. Could ATI, then, be one of the ways by which the health care system becomes more effective and efficient?

At the present time, this direction, we admit, seems quite speculative. If indeed medical outcomes depend on a complex matrix of patient *x* provider *x* treatment *x* problem *x* setting *x* outcome interactions, it is hard to imagine exactly how findings from ATI research would be integrated into an increasingly failing health care system. After all, the occurrence of interaction implies a limitation on generalizability of treatments effects, since it suggests that effectiveness is conditional (Cronbach, 1982; Smith and Sechrest, 1991). So, is it at all possible that patients, providers, and treatments can be matched in the near future so as to optimize outcome? We believe the answer depends, in part, on whether the health care system is ready to move to what we regard as the imperative future generation of evidence-based medicine (EBM) and outcome research.

EBM (which focuses on the need for rational, empirically proven health care decision making) is often thought to be at odds with the humanistic, narrative approach that strives to understand the illness experience by attending to the needs of patients (Jonas, 2001; Tonelli and Callahan, 2001). On the contrary, EBM does advocate the "compassionate use of individual patients' predicaments, rights, and preferences in making clinical decisions about their care" (Sackett et al., 1996). However, the main body of current research that generates evidence (i.e., randomized controlled trials) almost never takes into consideration any patients' aptitudes that we argue are likely to interact with the treatment that is being tested. In fact, with the exception of designs such as partially randomized patient preference studies (PRPP) (Lambert and Wood, 2000), the very terms "randomization" and "preferences" are, in a way, antithetical to one another. ATI research, however, offers a way to integrate aptitudes with evidence. After all, evidence does not make decisions, people do (Haynes et al., 2002).

Encouraged by what seems to be a complete theoretical fit between what the Institute of Medicine suggested and what EBM is about, we therefore suggest that hypothesis-driven ATI research should be a routine part of the outcome research portfolio. Doing so would allow us to examine the relationship between individuals and therapies with respect to outcome, and to illuminate the mechanisms and

processes that make therapies differentially effective. Evidence is only one of many important factors in effective decision making. As the late John Eisenberg, former Director of the Agency for Healthcare Research and Quality (AHRQ) put it, "worldwide access to evidence-based clinical decision making must coexist with respect for individual decision making shaped by local culture and circumstances. This is the balance between globalizing the evidence and localizing the decisions that will improve delivery of health care worldwide" (Eisenberg, 2002). Global evidence still needs localized decision making (Schneider and Eisenberg, 1998). ATI research can help us to achieve that.

REFERENCES

Aguinis H. Statistical power problems with moderated multiple regression in management research. J Manage 1995;21:1141–1158.

Anderson, KW. Utility of the five-factor model of personality in psychotherapy aptitude–treatment interaction research. Psychother Res 1998;8:54–70.

Aronson E, Carlsmith JM. Experimentation in social psychology. In: Lindzey G, Carlsmith E eds. Handbook of Social Psychology, vol. 2, 2nd ed. Reading; MA; Addison-Wesley, 1968:1–79.

Bell IR, Caspi O, Schwartz GRE, Grant KL, Rychener D, Gaudet TW, Maizes V, Weil AT. Integrative medicine and systematic outcomes research: Issues in the emergence of a new model for primary health care. Arch Intern Med 2002;162:133–140.

Beutler LE. Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. J Consult Clin Psychology 1991;59: 226–232.

Bogden HE. Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. Educ Psychological Meas 1951;11:173–196.

Campbell D, Fiske DW. Convergent and discriminant validation by the multitrait–multimethod matrix. Psychological Bulletin 1959;56:81–105.

Cronbach LJ. Beyond the two disciplines of scientific psychology. Am Psychologist 1975;30:116–126.

Cronbach LJ. Designing evaluations of educational and social programs. San Francisco: Jossey-Bass, 1982.

Cronbach LJ, Snow RE. Aptitude and Instructional Methods: A handbook for Research on Interactions. New York: Irvington, 1977.

Dance KA, Neufeld RWJ. Aptitude-treatment interaction research in the clinical setting: A review of attempts to dispel the "patient uniformity" myth. Psychological Bull 1988;104:192–213.

Edwards AL, Cronbach LJ. Experimental design for research in psychotherapy. J Clin Psychology 1952;8:51–59.

Eisenberg JM. Globalize the evidence, localize the decisions: Evidence-based medicine and the international diversity. Health Affairs 2002;21:166–168.

Figueredo AJ, Brooks A, Leff HS, Sechrest, L. A meta-analytic approach to growth curve analysis. Psychological Reports 2002; 87:441–465.

Gangestad S, Snyder M. "To carve nature at its joints": On the existence of discrete classes in personality. Psychological Rev 1985;92:317–349.

Garfield SL, Bergin AE, eds. Handbook of Psychotherapy and Behavior Change, 3rd ed. New York: John Wiley & Sons, 1986.

Haynes RB, Devereaux PJ, and Guyatt GH. Physicians' and patients' choices in evidence based practice: Evidence does not make decisions, people do. BMJ 2002;324:1350.

Hollon SD, DeRubeis RJ, Shelton RC, Weiss B. The emperor's new drugs: Effect size and moderation effects. Prevent Treat 2002;5:Article 28. Online documental: www.journals.apa.org/prevention/volume25/pre0050028c.html; accessed July 20, 2004.

Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press 2001.

Jonas WB. The evidence house: How to build an inclusive base for complementary medicine. Western J Med 2001;175:79–80.

Kirsch I, Moore TJ, Scoboria A, Nicholls SS. The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. Prev Treat 2002;5: Article 23. Online document at: www.journals.apa.org/prevention/volume25/pre-0050023a.html; accessed July 20, 2004.

Lambert MF and Wood J. Incorporating patient preferences into randomized trials. J Clin Epidemiol 2000;53:163–166.

Leuchter AF, Cook IA, Witte EA, Morgan M, Abrams M. Changes in brain function of depressed subjects during treatment with placebo. Am J Psychiatry 2002;159:122–129.

Petitti DB. Meta-Analysis, Decision Analysis and Cost Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine, 2nd ed. New York: Oxford University Press, 2000.

Ribeiro SC, Tandon R, Grunhaus L, Greden JF. The DST as a predictor of outcome in depression: A meta-analysis. Am J Psychiatry 1993;150:1618–1629.

Ridenour TA, Treloar JH, Biner PM, Henriksen LW, and Dean RS. Attribute–treatment interaction research using nonexperimental data and the allocation average. Int J Neurosci 1999;99:19–47.

Rosnow RL, Rosenthal R. Definition and interpretation of interaction effects. Psychological Bull 1989;105:143–146.

Rosnow RL, Rosenthal R. If you are looking at cell means, you're not looking at only the interaction (unless all main effects are zero). Psychological Bull 1991;10:574–576.

Rosnow RL, Rosenthal R. "Some things you learn aren't so": Cohen's paradox, Asch's paradigm, and the interpretation of interaction. Psychological Sci 1995;6:3–9.

Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. BMJ 1996;312:71–72.

Schneider EC, and Eisenberg JM. Strategies and methods for aligning current and best medical practices: The role of information technologies. Western J Med 1998;168:311–318.

Shadish WR, Cook TC, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton Mifflin, 2002.

Shoam V, Rohrbaugh M. Aptitude *x* treatment interaction (ATI) research: Sharpening the focus, widening the lens. In Aveline M, Shapiro DA, eds. Research Foundation for Psychotherapy Practice. Sussex, England: John Wiley & Sons, 1995:73–95.

Shoam-Salomon, V. Introduction to special section on client–therapy interaction research. J Consult Clin Psychology 1991;59: 203–204.

Shoam-Salomon V, Hannah, M.T. Client–treatment interaction in the study of differential change processes. J Consult Clin Psychology 1991;59:217–225.

Smith B, Sechrest, L. Treatment of aptitude *x* treatment interactions. J Consult Clin Psychology 1991;59:233–244.
Snow RE. Aptitude–treatment interaction as a framework for research on individual differences in psychotherapy. J Consult Clin Psychology 1991;59:205–216.
Stiles WB, Shapiro DA, Elliott R. Are all psychotherapies equivalent? Am Psychologist 1986;41:165–180.
Tonelli MR, Callahan TC. Why alternative medicine cannot be evidence-based. Academic Med 2001;767:1213–1220.

Address reprint requests to:
Opher Caspi, M.D., Ph.D.
Program in Integrative Medicine
The University of Arizona Health Sciences Center
1249 North Mountain Avenue
Tucson, AZ 85719

E-mail: ocaspi@ahsc.arizona.edu