Mental Health Services Research, Vol. 7, No. 1, March 2005 (© 2005)
DOI: 10.1007/s11020-005-1962-6

Identifying and Selecting the Common Elements of Evidence Based Interventions: A Distillation and Matching Model

Bruce F. Chorpita,[1,4] Eric L. Daleiden,[2] and John R. Weisz[3]

A model is proposed whereby the intervention literature can be empirically factored or *distilled* to derive profiles from evidence-based approaches. The profiles can then be matched to individual clients based on consideration of their target problems, as well as demographic and contextual factors. Application of the model is illustrated by an analysis of the youth treatment literature. Benefits of the model include its potential to facilitate improved understanding of similarities and differences among treatments, to guide treatment selection and matching to clients, to address gaps in the literature, and to point to possibilities for new interventions based on the current research base.

KEY WORDS: evidence-based; practice; common; elements; distillation; matching.

Efforts to summarize scientific knowledge regarding effective interventions have a long history in psychology. Although advances in the methods for organizing and interpreting the literature have seen great improvements (e.g., Chambless & Hollon, 1998), the use of interventions supported by research may be less common than the evidence warrants, given the many obstacles such interventions face in clinical settings (e.g., Addis, Wade, & Hatgis, 1999). In the hope of addressing some of these challenges, we outline here a model that appears to offer several proposed advantages over existing methods of summarizing evidence on psychological interventions. Consider the following two vignettes, which serve to illustrate the challenges in interpreting the evidence base on psychological interventions: *Vignette I*

[1]University of Hawaii, Honolulu, Hawaii.
[2]Child and Adolescent Mental Health Division, Hawaii Department of Health, Honolulu, Hawaii.
[3]University of California, Los Angeles, California.
[4]Correspondence should be directed to Bruce F. Chorpita, Department of Psychology, University of Hawai'i at Mānoa, 2430 Campus Road, Honolulu, Hawaii 96822; e-mail: chorpita@hawaii.edu.

Supervisor: I am planning to train my staff in an evidence-based treatment for depression, but I don't know what treatment to train them in., Administrator: Well, there seem to be plenty to choose from., Supervisor: True, according to the APA Div 12 criteria, there's Interpersonal Therapy and Cognitive Behavior Therapy . . .,

Administrator: Seems like there are a lot of good CBT manuals out there—maybe pick one of those?,

Supervisor: Well, the clinical trial of IPT has been replicated, but I don't think any of the CBT trials have been replicated.,

Administrator: Aren't there at least 5 good studies on CBT?,

Supervisor: Yes, but the manuals are all different.,

Administrator: CBT is CBT, if you ask me. That's replication as far as I am concerned.,

Supervisor: Well the Division 12 criteria are not entirely clear about that, but I see what you mean. If I go with CBT, which manual should I pick?

Administrator: There is a new version of the 1993 Smith and Jones protocol published this year, why not go with that? It's recent.,

Supervisor: But shouldn't I use the one that was tested in the trial?, Administrator: I think the '93 protocol is out of print. Besides, the new one is pretty similar, and it has some nice improvements. All the homework forms come on a CD ROM now.,

Supervisor: I guess that makes sense. I thought this would be easier. I just want to know what I should train my staff to do ....

Vignette II

Therapist A: I am starting with a new client next week, and I am trying to decide what approach to use.,

Therapist B: Tell me about your client.,

Therapist A: She is a Latino adolescent with aggression problems.,

Therapist B: Well, why don't you try one of the anger management protocols? I recall a positive study using a minority sample.,

Therapist A: Well, haven't they pretty much only been tested with boys?,

Therapist B: Yeah, but that shouldn't matter so much.,

Therapist A: And this girl is 13. Don't those manuals apply to ages 15 and up?,

Therapist B: Close enough, no? I mean, what else are you going to use?,

Therapist A: I guess I just wish I knew what factors were more important, her problem, her age, her background ...,

Therapist B: They're all important, I think.

These vignettes are intended to illustrate the real-world complexity of identifying and selecting interventions using the evidence base as a guide. Problems frequently encountered include issues regarding what approaches are considered similar, which findings can be aggregated, and how to maximize fit of the intervention with the individual's problem and context. To address such problems, we propose a *Distillation and Matching Model* (*DMM*). The model involves two basic processes—distillation and matching—each with specific advantages for addressing the challenge of intervention selection from the evidence base. The first process described in the model, *distillation*, is a method whereby interventions are conceptualized not as single units of analysis, but rather as composites of individual strategies, techniques, or components that can allow subsequent empirical grouping. This approach offers the following advantages:

1. *Technique identification*. The intervention literature can be factored by technique, using a structured methodology to determine which interventions are considered "cognitive behavioral," "behavioral parent training," etc. Individuals facing the task of treatment selection can thereby identify empirically sound classes of interventions. Current methods for interpreting the evidence base either restrict such analysis or force one to determine an intervention's membership in a larger group based on the protocol's name (e.g., is "cognitive" in the title?) or an informal review of its contents (e.g., does it appear to be mainly behavioral and parent-focused?).
2. *Evidence accumulation*. This ability to determine whether manuals differ significantly in their content also allows for the accumulation of evidence across revisions. For example, whether a study represents a replication trial (thus adding to the empirical weight of a particular protocol) or a test of a new approach can be determined using this model. In this manner, clinical tests contribute to an accumulation of information in methodologically determined classes of interventions (e.g., a trial on protocol A might add to the existing literature on "exposure therapy" or it might add to a new literature, depending on whether distillation shows that protocol A includes exposure therapy). Also, a new edition of a previously tested manual can be scored for its similarity to the approach used in the clinical research trial, so there is less guesswork about whether a new manual is "substantially different."

The second aspect of the model, *matching*, is a method for summarizing client, setting, or other factors that might be relevant considerations for selecting an intervention. The advantages of this approach are as follows:

1. *Gauging association between treatment content and study characteristics*. The association of client, setting, or other study characteristics with intervention content can be empirically determined. Currently, it is unclear which factors are more commonly associated with successful demonstrations of interventions and which are not. Thus, in selecting an intervention, one might consider age or gender as important variables, but the literature demonstrates no obvious association between these variables and protocol content.

When a specific association does exist, the model identifies it.

2. *Determining which study characteristics matter most*. These client, setting, or other study characteristics can be ranked in order of their association with intervention content. Currently, prevailing models for treatment selection assume diagnosis as the primary selection factor. The DMM would allow for the determination of whether diagnosis is in fact the most appropriate first variable to consider. Depending on the literature reviewed, the model could as easily identify that gender is the most appropriate variable to consider when selecting an intervention. For example, preschool boys might all benefit from a similar class of interventions (e.g., "rewards and consequences"), thus implying a public health strategy for intervention or dissemination.

The operation of the DMM will be described in more detail below, but first we begin with a review of prior efforts to summarize the intervention literature and their different levels of analysis.

BACKGROUND

In some of the first attempts to draw broad inferences about therapy effects, several early reviews summarized observations that all therapies were more or less equally effective (e.g., Luborsky, Singer, & Luborsky, 1975; Smith, Glass, & Miller, 1980) or in at least one case, equally ineffective (Eysenck, 1952). Partly as a result of such conclusions, these early reviews spawned considerable controversy regarding level of analysis. Critics of the reviews pointed out that considering multiple forms of psychotherapy the same, multiple client types the same, and minimizing possible interactions of interventions and client factors raised numerous problems (e.g., see Beutler, 1991, 2002; Chambless, 2002; Chambless & Ollendick, 2001; King & Ollendick, 1998; Luborsky et al., 2002; Rounsaville & Carroll, 2002).

Along those lines, an accumulation of subsequent reviews has suggested that in fact some differences might exist between interventions conceptualized at the level of their theoretical background. For example, Weisz and colleagues (Weiss & Weisz, 1995; Weisz, Donenberg, Han, & Weiss, 1995; Weisz, Weiss, Alicke, & Klotz, 1987; Weisz, Weiss, Han, Granger, & Morton, 1995) found that for child and adolescent populations, there were clear advantages of behavioral over nonbehavioral methods. Such observations highlight the importance of examining treatment by client combinations (e.g., the observed advantage of behavior therapy might not have been detected in a sample collapsing across the lifespan). A logical extension of this idea is that there might still be findings that are undetected by current approaches, because of methodologies that (a) collapse levels of a factor together based on superficial similarities (e.g., "dialectical behavior therapy" and "systematic desensitization," two rather different types of behavior therapy, potentially being lumped together as "behavior therapy") or (b) impose only a small number of rationally determined factors to consider (e.g., considering therapy type and client age, but not ethnicity, gender, treatment setting, therapist background, etc.). In other words, there is almost certainly critical information in the existing accumulation of research findings that has yet to be identified and summarized (e.g., Beutler, 2002; Rounsaville & Carroll, 2002).

This notion of data potential is perhaps best considered in the context of Kiesler's (1966) paper on patient-treatment interaction (see also Paul, 1967), echoed more recently in the context of child psychotherapy research by Kazdin, Bass, Ayers, and Rodgers (1990). Collectively, these authors articulated what has become a standard refrain for inference about psychotherapy effects. That is, it is important to know *what works for whom and under what conditions*. Such inferences can presumably be drawn in a number of ways from the weight of the empirical literature on psychological intervention outcomes, but few models for arriving at those answers have been outlined.

CURRENT STATE OF THE FIELD

One of the most significant efforts to systematically define the manner in which treatments should be collectively evaluated arose from the American Psychological Association (APA) Task Force on Psychological Intervention Guidelines and the Division 12 Task Force on Promotion and Dissemination of Psychological Procedures (Task Force on Promotion and Dissemination of Psychological Procedures, 1995; Task Force on Psychological Intervention Guidelines, 1995). The APA Task Force developed a template for evaluating the degree of internal and external validity of outcome studies and concluded that the strongest designs were group designs involving active control groups (which best rule out alternative explanations of effects) and random assignment to conditions. The randomized clinical trial was thus considered to be a "gold standard" for psychotherapy research.

Following these efforts, the APA Task Force on Promotion and Dissemination of Psychological Procedures outlined a more detailed definition of efficacy in order to develop an official list of those "empirically supported treatments" (ESTs) with the best empirical support. The APA Clinical Division Task Force outlined two levels of support. The first, "well-established" essentially required at least two randomized clinical trials with active controls or a large series of single case design experiments using active treatment comparisons. In addition, independent replication and manualization of the treatments was required. The second level of support, "probably efficacious," required at least two randomized clinical trials with waitlist controls, a single randomized clinical trial with an active control, or a small series of single case design experiments using active treatment comparisons. These guidelines have been a major influence in the review and evaluations of interventions for the past 8 years, and to that end, APA's Clinical Child and Adolescent Division has followed with similar reviews (Lonigan, Elbert, & Bennett Johnson, 1998).

CURRENT IMPLICATIONS AND CHALLENGES

The EST approach offered distinct advantages over previous approaches, in that the rules for knowledge accumulation were made clear, thus establishing replicable standards for the review process. However, a possibly unintended consequence is that the level of analysis shifted to specific manuals: "There are many interventions that fall under a rubric . . . and the brand names [e.g., "brief dynamic therapy," "cognitive behavior therapy"] are not the critical identifiers. The manuals are." (Chambless et al., 1996, p. 7). For example, now it is less easy to speak of the support for behavior therapy than to speak of a specific manual that has satisfied particular empirical criteria.

On the other hand, exceptions to this approach have emerged in subsequent reviews (e.g., Lonigan et al., 1998), and apparent exceptions appear in the summaries of the APA Clinical Division Task Force. These exceptions summarized interventions in general classes based on global judgments about the observed similarities among manuals. Thus, it may be most accurate to say that the current systems based on the adult and child reviews of the evidence base seem to involve *two* approaches, each of which is subject to some limitations: (a) identifying evidence based treatments based on specific manuals, and (b) identifying evidence based treatments based on global judgments about whether various manuals share enough similarities to constitute a single generic treatment. The DMM outlines a systematic method to organize the evidence that does not fall prey to the weaknesses of either a or b. That said, the use of manuals as a method of defining treatment is the most systematic at present, and thus further discussion is warranted regarding the limitations of manuals as a means of summarizing the information in the literature.

By defining interventions at the level of manuals, research and knowledge accumulation begin anew with each significant change in a manual, and many manuals are changed significantly with each new clinical trial. When applying the strictest rules of scientific inference, such changes effectively erase past credit for efficacy. For example, a change to the final session of a behavior therapy protocol could require an investigator to return to pilot trials for evidence of apparent benefit before conducting larger tests of efficacy or effectiveness. One could argue that the change was insignificant; however, the determination of whether such a change was substantial enough to warrant a new series of tests is currently a subjective matter. Difficult decisions aside, the strict consideration of interventions at the level of manuals places a formidable empirical burden on intervention developers, confronting them with the alternatives of (a) potentially endless testing and retesting or (b) rigid adherence to a specific protocol even when adaptations might improve it.

Such issues have been discussed in the context of treatment manual "flexibility" and comparisons between different models for development and testing of interventions. As an example, Weisz (2004) has contrasted a rather traditional approach to intervention testing that grows out of the medical tradition with an alternative model that may warrant attention in the development of psychosocial treatments. In the generic *medical-pharmaceutical (MP) model* (see Greenwald & Cullen, 1984; National Institutes of Health, 1994; Weisz, 2004), experimental treatments and their protocols are first developed in the laboratory, then held constant and tested via a series of laboratory *efficacy* trials, and then, late in the testing process, brought into community clinical settings and tested with clinicians and real patients "to measure the public health impact" (Greenwald & Cullen, 1984). While the MP model may work well for biological interventions, Weisz suggests that an alternative *deployment-focused model* (DFM) may be better suited to the production of clinic-worthy psychosocial treatments. In the DFM, initial evidence of benefit in an efficacy trial is followed by a series of adaptations and effectiveness trials that increasingly engage the clients, providers, settings, and conditions toward which a treatment is ultimately aimed; the idea is to maximize fit between the intervention and its eventual context.

Different models of intervention development and testing are apt to have different implications for the speed of innovation, and they may confront rather different risks. A model that discourages adaptation may, across years of testing, produce replicated knowledge about a specific intervention, but that intervention may eventually fail in the clinical practice context for which it was ultimately intended. In contrast, a model that proposes adaptation will foster more rapid innovation but may risk uninterpretability unless it specifies boundary conditions or a generalizability framework for making inferences about the adaptation. How much change is too much before an intervention must be considered new and in need of a new series of clinical tests? This question is not easily answered without considering interventions at a different level of specificity than simply their manuals.

In addition, assuming that the MP and DFM models are both workable frameworks for treatment development, each with certain strengths and risks, neither speaks explicitly about strategies for selection of approaches at the level of an individual client. In this context, a critical question is how the clinician, faced with a particular client in a particular context, is to choose the most appropriate psychotherapy approach? For therapists who want to use research evidence as the basis for such a decision, the dominant zeitgeist since the Clinical Division Task Force may be a "main effect" model of choosing a manual from the list of empirically supported treatments, based on a client's primary diagnosis. To account for additional client and environmental characteristics, a "main effects in context" model (Dawes, 1992) might be used in which the set of relevant manuals is identified based on primary diagnosis or problem area, and the manual that has been tested in a context most similar to the therapy context would be selected next. But what should be done if no studies exist to guide this context fitting process?

Even when there are multiple studies employing a similar treatment model, a conundrum may be encountered. For example, Weisz and Hawley (1998) pointed out that although there were six randomized clinical trials supporting cognitive behavior therapy for depressed youth, eight different manuals were involved, each supported by only a single study. This circumstance not only suggests relatively limited confidence in any one protocol, but also affords no guidance to a clinician seeking to select the best manual from among the eight (Chorpita et al., 2002).

A similar dilemma confronted our research team working on a MacArthur Foundation-supported initiative on youth mental health care (Weisz et al., 2003). Our effort to find a beneficial intervention for depressed children looked promising at first, as we identified numerous randomized trials of child depression treatment. But we were unable to identify a qualifying, supportive replication for any single intervention manual for the target population of youth ages 8–13. Further, when we narrowed the candidate manuals down to a smaller number based on child age, the treatments that emerged had only been tested in group therapy; we were seeking an approach to use in individual therapy. When a group of researchers and treatment developers themselves cannot easily resolve such challenges, the implications are ominous for a practicing clinician seeking to select an intervention.

Other effects of using manuals as the primary level of analysis may also be considered. For example, overspecification of interventions (defining each manual as completely different from each other) can lead to unnecessary duplication among technologies, creating impediments to training and dissemination. For example, three cognitive behavioral treatment manuals for depressed adults might contain largely the same approaches in terms of content, but would be considered three different interventions each with a different research base. Particularly given the trend of increasing specificity in classifying problems, it is possible that for each age group and for every possible problem, a different manualized approach will be developed. According to a strict interpretation of the MP model, each manual would be used on its own for its specific target populations and problems. This has already led to numerous "empirical silos" or independent research tracks that do not readily build cumulatively on one another. The result is a proliferation of evidence-based approaches with minimal guidelines for how to aggregate or select among them. For example, the Substance Abuse and Mental Health Services Administration's National Registry of Effective Programs cited 53 model programs as of September 2003, many of which do not appear on the APA Clinical Division list of evidence-based treatments.

Altogether, these problems with treatment selection, research accumulation, and protocol adaptation are unexpected side effects of using manuals to define treatments. This is not to discourage the use of manuals, per se. Psychotherapy manuals help clearly specify therapy content, operationalize therapeutic procedures, specify a sequence to the operations, and provide a reliable dimension on which to aggregate across studies. However, some of these benefits may be preserved without accepting the manual as the only appropriate unit of analysis. Specifically, many of the benefits of manuals accrue from the fact that they codify procedures in a reliable fashion and support adherence checks. When one separates the procedural codification function of manuals from the definition of manuals as the unit of analysis (i.e., a protocol-ordered sequence of codified procedures with highly specified content), other avenues for expanding the reach of the extant literature may become apparent. Our subsequent discussion highlights an application of analysis at the level of practice techniques rather than manuals. For the sake of our illustration, we assume that such codification will roughly but not finely specify therapy content, support adherence checks, and provide a reliable dimension along which to aggregate across studies, but will ignore sequencing, duration, and other parameters of the intervention delivery (e.g., Bickman, Vol. 7(1), pp. 1–4). Our intention is to sketch one alternative to the use of intact manuals for mapping the accumulation of research data, understanding data relations, and extracting prescriptive heuristics to apply to novel situations. If only at the simplest level, such methods might begin to speak to Kiesler's (1966) concerns.

A DISTILLATION AND MATCHING MODEL

In many ways, the distillation and matching model is essentially an adaptation of *data mining* applied to the clinical research base. In information sciences, the field of Knowledge Discovery and Data Mining has offered a variety of strategies for identifying patterns in data that yield practical information which serve as a useful conceptual framework (Brodley, Lane, & Stough, 1999; Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Brodley et al., outlined knowledge discovery as being characterized by six stages: (1) development of an understanding of the domain of interest, (2) creation of a target data set, (3) correction and preparation of that data set, (4) application of data reduction algorithms, (5) application of data mining algorithms, and (6) interpretation of the mined patterns by domain experts. For a variety of reasons, knowledge discovery and data mining are not wholly applicable to the data from clinical trials, owing mainly to the incomplete nature of the literature. For example, if Intervention A has been tried only with aggressive boys, but not with aggressive girls or nonaggressive boys, one cannot determine whether the association between gender or treatment target is more important. Such confounds are evident throughout the psychological intervention literature. Nevertheless, the concepts from knowledge discovery and data mining, and the steps of the process described by Brodley et al. (1999), serve as a useful framework within which to describe the distillation and matching approach.

Step 1. Development of an Understanding of the Domain of Interest

As reviewed above, efforts have been ongoing within psychology to produce reviews and to develop strategies to interpret the summary of research findings in the intervention literature (e.g., Chambless et al., 1996; Eysenck, 1952; Task Force on Promotion and Dissemination of Psychological Procedures, 1995; Task Force on Psychological Intervention Guidelines, 1995). Nevertheless, one of the more undeveloped aspects of the intervention literature involves the taxonomy for coding the specific treatment techniques (e.g., "relaxation training," "time out") that are combined within full treatment manuals. Although manuals have been developed to codify groups of treatment techniques when combined, a standard lexicon of distinct techniques has not been developed in a way that can be applied across treatment manuals and used to identify the components of each.

We are not asserting that common terms for treatment operations are absent in the research literature (e.g., relaxation training, overcorrection, etc.), but rather we are highlighting that a standard taxonomy and lexicon of psychosocial treatment operations, similar to the Diagnostic and Statistical Manual-IV codes for diagnosis or the Current Procedural Terminology-IV codes for general medical procedures, are not available to clearly and consistently codify practices. The ability to reliably define and discriminate treatment operations (i.e., defining the independent variable) is a central part of accumulating knowledge across studies (Chambless & Hollon, 1998). As previously noted, the present zeitgeist is to classify interventions by the set of manual titles (e.g., "Coping Cat," Kendall, 1990) as providing the lexicon for discussing treatment approaches. The use of manual titles, which are highly specified, extensively operationally defined terms, as the sole lexicon has precluded alternative levels of analysis of the intervention literature.

Some notable exceptions have been directed to the development of therapist reported therapy procedures checklists (e.g., Weersing, Weisz, & Donenberg, 2002), which have been found to reliably factor into theoretically based groupings (e.g., Psychodynamic, Cognitive, and Behavioral) and meaningfully relate to child characteristics (e.g., greater use of behavioral techniques with younger children). The constructs associated with such technique checklists are also relevant to measuring codified treatment procedures, but we are not familiar with any such application.

As a foundation for the model, it is therefore important to propose that such a taxonomy be developed. Because the current paper is merely intended to illustrate the feasibility of the distillation and matching concepts, a definitive taxonomy will not be presented here. Rather, a general strategy for the development and application of such an approach is outlined.

Of foremost importance in this proposed taxonomy is the term "practice element." We define a practice element as a discrete clinical technique or strategy (e.g., "time out," "relaxation") used as part of a larger intervention plan (e.g., a manualized treatment program for youth depression), based on the assumptions that (a) practice elements can be explicitly defined (e.g., using a definition or coding manual), (b) their presence within various interventions can be reliably coded, and (c) different treatments may share practice elements in common. Practice elements, as we construe them, may be congruent with sessions in some parts of some manualized interventions (e.g., a session may be devoted entirely to relaxation training), but in other manuals there is little or no congruence (e.g., some manuals might include multiple practice elements in each session). Practice elements are defined by their content, not by their duration, periodicity, or location within a manual. Practice elements may be delivered simultaneously (e.g., "exposure" and "response prevention"), concatenated ("cognitive restructuring" followed by "relaxation"), used in a single session (e.g., an entire session of "psychoeducation"), or addressed in multiple sessions (repeated sessions involving "parent training in the use of rewards"). See Table I for examples.

Table 1. Hypothetical Examples of Practice Elements and Definitions

Activity scheduling	The assignment or request that a child identify or list activities the child finds pleasant or mood-elevating, and that the child schedule and participate in some of those activities, with the goal of promoting or maintaining involvement in satisfying and enriching experiences.
Cognitive/coping	Any techniques designed to alter interpretation of events through examination of the child's reported thoughts, typically through the generation and rehearsal of alternative counter-statements. This can sometimes be accompanied by exercises designed to comparatively test the validity of the original thoughts and the alternative thoughts through the gathering or review of relevant information.
Parent praise	The training of parents or others involved in the social ecology of the child in the administration of social rewards to promote desired behaviors. This can involve praise, encouragement, affection, or physical proximity.
Problem solving	Techniques, discussions, or activities designed to bring about solutions to targeted problems, usually with the intention of imparting a skill for how to approach and solve future problems in a similar manner.
Self-monitoring	The repeated measurement of some target index by the child.
Tangible rewards	The training of parents or others involved in the social ecology of the child in the administration of tangible rewards to promote desired behaviors. This can involve tokens, charts, or record keeping, in addition to first-order reinforcers.
Time out	The training of or the direct use of a technique involving removing the youth from all reinforcement for a specified period of time following the performance of an identified, unwanted behavior.

This approach of specifying interventions based on their components requires that several decisions be made in advance–for example, (a) whether the coding scheme will encompass the entire set of possible practice elements, or focus mainly on those likely to show up in the target data set (i.e., evidence based approaches only, as determined by selected criteria), (b) at what level of specificity to code interventions (e.g., code the presence or absence of exposure, or the presence or absence of exposure's many constituent steps), and (c) whether to define practice elements entirely in terms of techniques, or whether to code other aspects of interventions such as parameters of therapeutic alliance, amount of homework assigned, etc.

For the present illustration, we have chosen to limit the set of practice elements to those appearing commonly in the target literature (i.e., evidence based protocols per APA Clinical Division criteria), as to do otherwise would result in a nearly endless list of possible practice elements. We have chosen this to rather than coding specific steps. The rationale is that coding specific steps as individual elements is unlikely to lead to a greater understanding of the domain, given that these steps typically only occur when organized as the entire technique itself. For example, taking fear ratings each minute would not typically occur outside the context of exposure, so there may be little incremental benefit for coding at that level of specificity.

Finally, for this illustration we have chosen to focus solely on clinical techniques as practice elements, rather than on other aspects of therapy. This is not to say that other dimensions of interventions are not important or could not be selected and used in the model. Initial choices regarding codes to produce data for the model are rational choices to be made by clinical experts. Once these codes are chosen, the DMM empirically organizes the domain according to those selected dimensions. One could, for example, code for the presence or absence of such elements as high therapeutic alliance, the assignment of homework, the charging of a fee, or the presence of behavioral rehearsal in therapy sessions. To the extent that data produced by such codes showed substantive association with contextual variables of interest (e.g., gender, ethnicity, therapist background, diagnosis), the model would identify these associations. To the extent that such variables were truly "nonspecific" (i.e., universally associated with positive outcomes; e.g., therapist competence), the model would show that those variables were simply present in all contexts. Inclusion of truly nonspecific factors would be helpful to remind users of the final profiles to either apply or avoid applying these factors (e.g., always be warm), but they would not affect decisions about matching treatment selections to contextual characteristics. Thus, our decision to focus on specific techniques was a rational choice based on both practical and theoretical considerations that, if modified, could affect the output of the model but would not alter its fundamental methodology. The DMM is not intended to imply a certain set of codes nor to aid in their initial construction.

Step 2 and 3: Creation, Correction, and Preparation of the Target Data Set(s)

Two data sets are required to implement distillation and matching procedures. The first we refer to as the *study data set*. It consists of the set of empirical studies of treatment efficacy and effectiveness, with each study coded with respect to population (e.g., child characteristics, selection criteria), intervention procedures (e.g., manuals), efficacy (e.g., yes-no judgments of whether the study provides empirical support for the intervention on the dependent variables, effect sizes for the dependent variables), and effectiveness (e.g., therapist variables, client variables, setting). Information regarding efficacy is necessary to determine inclusion in the data set. In other words, if there is no use of a manual, or if the intervention was not superior to a control condition, the intervention would be excluded from analysis. Information about effectiveness or context (e.g., client age range, treatment setting) provides the set of variables for matching interventions to contexts.

The second data set we refer to as the *procedures data set*. It consists of the set of treatment protocols (e.g., manuals when available) as cases and includes codes for practice elements selected for examination. Thus, each protocol is rated yes/no on the presence of every practice element of interest (e.g., "rewards," "time out," "relaxation").

Our present illustration used a subset of studies included in the Hawaii Evidence Based Services Committee review (Chorpita et al., 2002), as the study database and created a procedures data set by coding the relevant manuals with a preliminary practice element codebook developed for this illustration. The Hawaii database included several hundred group-design random-assignment clinical trials testing psychotherapies for youths younger than age 18. The studies were published in the years 1968–2002, inclusive. The problem domains included in the database included conduct problems (encompassing oppositional defiant disorder and conduct disorder as well as disruptive or antisocial conduct), ADHD (encompassing ADHD and ADD diagnoses as well as related problems of attention, overactivity, or impulsivity), depression (encompassing diagnoses of major depressive disorder, minor depressive disorder, and dysthymic disorder as well as elevated depressive symptoms), and anxiety (encompassing anxiety disorders, phobias, and elevated anxiety symptoms).

The practice element codebook for this illustration included definitions for 26 practice elements, defined as outlined in the example codes in Table I. These 26 were selected as the most common elements to appear in the target data set from a larger pool of items nominated by several panels of practitioners, intervention developers, and other domain experts. Preliminary examination of inter-rater agreement on the 26 codes performed by the first two authors coding manuals for anxiety (Coping Cat; Kendall, 1990), depression (Primary and Secondary Control Enhancement Training; Weisz, Weersing, Valeri, & McCarty, 1999), and disruptive behavior (Defiant Children; Barkley, 1997) provided initial evidence that the coding system can reliably detect well defined treatment operations (mean $\kappa = .76$).

Step 4: Application of Data Reduction Algorithms

Distillation involves the reduction of complexity in data to a smaller set of meaningful units—in this case, the reduction of the many features of intervention packages to a set of common practice elements. An analogy from the data mining literature illustrates the value of distillation as a data reduction strategy. One of several promising algorithms for understanding patterns in complex data involves what is known as *spoiling*. Spoiling is typically used to identify prototypes in large data sets, which, once identified, are subject to further analysis. Brodley et al. (1999) outlined an example of how spoiling was used to identify volcanoes on the surface of Venus from a data set involving a set of images totaling nearly 31 billion pixels. The goal was to define prototype images against which raw pixilated data could be compared. The images of known volcanoes could not be grouped into prototypes until idiosyncratic details of their form could be removed. To accomplish this, detailed images were spoiled by averaging neighboring pixels and reducing a 15×15 image to a 2×3 image. The effect greatly simplifies the image with the effect of blurring data together. Having lost much of their unique detail, spoiled images are then more easily clustered into groups using a second algorithm. Detection strategies that attempted to match volcanoes against images that were spoiled far outperformed algorithms that compared data to original images.

This approach offers a strategy by which to average together information by blurring away the idiosyncratic features. To return to an example from child mental health, several manuals for depression might have different cartoon character guides, use different acronyms, or play different games, but if their underlying strategies were essentially the same, this common information would be important to distill together. The distillation method of coding practice elements, analogous to spoiling, is designed to extract only the desired information while filtering out additional features of the manuals.

This approach is designed to do more than just combine manuals from within a particular framework (e.g., Behavioral Parent Training). Unlike previous efforts that include theoretical orientation as grouping factor (e.g., Chorpita et al., 2002), distillation at the level of practice elements “spoils away” information about underlying rationale at this step and reintroduces such information only in the final step of analysis. Therefore, distillation may aggregate theoretically disparate but pragmatically similar approaches or disaggregate theoretically similar but pragmatically disparate approaches. For example, Interpersonal Therapy and CBT for depression both appear to involve behavioral rehearsal strategies, problem solving, and activity scheduling, whose collective effects on depression could be similar enough to suggest a unitary factor of depression interventions. On the other hand, CBT for anxiety and CBT for depression each have many elements not shared by the other, which could suggest that CBT is not a unitary factor. These relations are important to determine, as they may be relevant for ultimately mapping the relations of interventions to problems.

Step 5: Application of Data Mining Algorithms

With the domain of interest distilled as outlined above, it then becomes possible to group practice elements into *practice element profiles*. These profiles represent relative frequency counts for the use of each practice element in a given context. An example

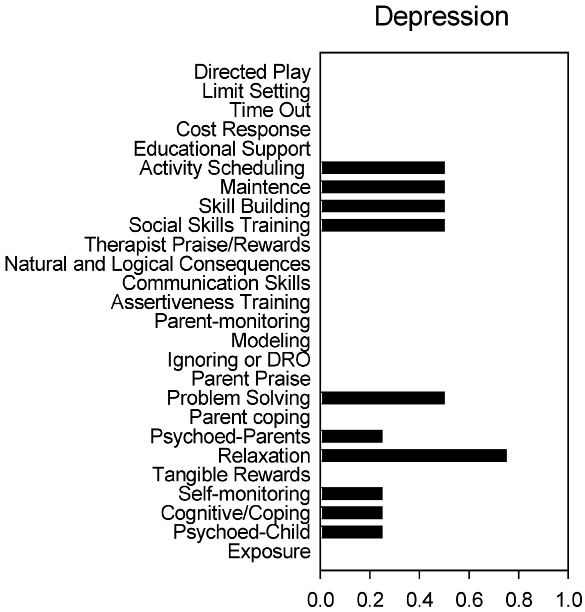

Fig. 1. Hypothetical practice element profile for treatments for depression in youth.

of a practice element profile appears in Fig. 1. The figure shows the relative frequency of practice elements among a set of interventions found to be efficacious for depression in adolescents. In this example, the most common technique among efficacious treatments for depression appears to be relaxation, but a variety of other practice elements appear to be relatively common as well.

In order to determine how to build these profiles, it is necessary to determine the associations between practice elements and a variety of treatment factors (e.g., treatment targets, client characteristics, setting characteristics). For example, one might wish to ask what is effective for depressed mood, for substance use, or for truancy. In this model, this aspect of building a practice element profile from the clinical research data is called "matching," as it involves the determination of which parameters of a target client in context match the occurrence of certain patterns of practice elements in the literature. For example, if the literature suggests that exposure is more commonly paired with rewards among Japanese children with specific phobias, the corresponding profile (rewards and exposure) would be suggested for that client.

A variety of algorithms exist by which to determine these associations, with major strategies involving variations of associative analyses and decision trees. For purposes of illustration, a decision tree approach will be outlined here, using an interaction detector algorithm. Such algorithms typically use a selected test statistic to identify points of similarity and difference among cells in the problem space, building a decision tree that outlines important interactions (e.g., Chi Square Automated Interaction Detector; Kass, 1980). An example might be that gender and diagnosis interact such that different practice element profiles emerge for depressed males and females, while "main effect" practice element profiles for males and females do not differ significantly. This would be the identification of a problem area × gender interaction on the frequency of practice elements employed in research trials, suggesting that treatments for males and females do not look different in general, but do for depression.

To implement this procedure, the procedures data set (involving coded practice elements) is merged with the study data set (involving the coded study factors of interest, e.g., setting, client age, gender, and ethnicity, therapist training, etc.). This allows for the creation of a factor space, which crosses each value of the variables within the study data set factors with every other. For simplification, continuous dimensions such as age may be reduced to ordinal categories (e.g., children, adolescents). Next, all possible pairs of cells within a factor (e.g., males and females within the factor gender) are evaluated using a chosen test statistic (e.g., Chi-square, ANOVA/ANCOVA, Intraclass correlation). For example, if diagnosis were to be included as a factor from the study data set, the practice element profile corresponding to each diagnosis would be compared with the practice element profile for every other diagnosis. The pair of profiles with the greatest similarity is considered, and in the event that the resulting test statistic exceeds a chosen acceptable-to-merge value (e.g., $p < .05$), those profiles are collapsed together or aggregated. For example, the profile for anxiety disorders might be so similar to the profile for phobias that merging would be dictated. The process repeats within other diagnoses, until no more merging is possible. What results is a clustered organization of that factor into levels that are reasonably different according to the statistical criterion. This entire process is then repeated within every other factor of interest, so that all factors are reduced to a minimum number of statistically different levels—with each representing a distinct practice element profile.

The next step is to determine the differences across levels within a factor (e.g., levels 5–12, 13–18, 18–25, 25–45, 45–65, and over 65 within the factor age). The same test statistic is calculated to determine which factor yields the greatest differentiation among the patterns of practice elements. Provided it meets an acceptable-to-split criterion, that factor is chosen as the first branching of the decision tree. For example, if the chosen test statistic was maximized for the difference between boys and girls, gender would be the first split in the tree. Likewise, if the differences among *k* clusters of disorders maximized that statistic, then the first split in the tree would be for diagnosis (splitting into *k* branches). The first branching of the tree is thus one that points to interventions (as defined by practice element profiles) that are maximally different.

Once the first split in the tree is established, we return to the very first step within each branch of the tree. That is, all factors are again examined at the greatest level of specificity, with all within-factor pairs examined to determine the possibility of merging. For example, if the first branching of the tree splits into Anxiety, Depression, and Disruptive Behavior, then we would go back to all cells within Anxiety, attempt to re-cluster them, then select the factor that produces the maximum value of the test statistic when splitting on that factor for the next branch. This would mean that differences between boys and girls that had been "washed out" in the first step of the analysis might now emerge and allow for further branching within Anxiety. Once the next split is achieved, we start over again to see if we can cluster (e.g., within anxious girls) on all remaining dimensions (e.g., age, ethnicity). Again we test for the strongest factor for branching, and continue until (a) the remaining factors all merge into a single unit, (b) no differentiated factors produce values that exceed the criterion for an acceptable split, or (c) there are no factors left to examine (i.e., we are at the last possible node in the tree, e.g., African American girls with anxiety ages 7–11).

Step 6: Interpretation of the Mined Patterns by Domain Experts

At some point, the resulting simplification of the data needs to be reviewed by experts in the field. This is because the purely empirical approach to matching does not discriminate between chance or artifactual patterns in the data and those consistent with theory. In the decision tree example, it is possible for branches to appear that have minimal heuristic value. Brodley et al. refer to this process of editing the results as "pruning the decision tree." In larger data sets, a variety of cross validation approaches could be employed; however, the relatively limited size of the intervention outcome literature does not lend itself to such procedures. Through a rational review, then, only branches that are consistent with theory or clinical knowledge would be maintained.

APPLICATIONS

A distillation and matching analysis of the outcome literature would yield a decision tree for matching clients to treatments, along with a profile of practice elements representing the average intervention for each final node in the tree. The question then becomes how to use such information. One approach is to use the leaf node profile to point to "best candidate" manuals. For example, if selecting an intervention for a 9-year-old Caucasian boy with disruptive behavior, one might inspect the resulting distilled profile matching that child's problem and characteristics, and attempt to identify a manual whose contents most closely match the practice element profile. This approach most differs from current strategies under circumstances in which (a) there might be no manual tested for that context and (b) there are multiple manuals supported for that context. In the former case, the distillation and matching approach would yield the closest match based on the entire literature. In the latter case, there would be an average of all the supported treatment approaches, which would likely point to the most reliably tested set of practice elements (those that show up in the greatest number of positive studies). Under circumstances in which a single manual represents a given context (e.g., Multisystemic Therapy for African American adolescent males with delinquent behavior), the profile would exactly reproduce the codes for that manual.

Another approach that could be employed involves the "modular" reassembly of the practice elements. In other words, the techniques themselves could be matched to specific targets according to a variety of algorithms drawn from the evidence based approaches or the broader literature. Although distillation need not lead to this type of modular application, modularity has some logical advantages in certain contexts, particularly those for which manualized interventions face implementation barriers. The concept of modularity, its application to the intervention literature, and its relation to distillation are beyond the scope of the present paper, and will be discussed in detail elsewhere (Chorpita, Daleiden, & Weisz, 2003; Chorpita, Taylor, Francis, Moffitt, & Austin, 2004).

A WORKING EXAMPLE

As an illustration of the model, we can take as an example the approach of applying a decision tree algorithm. This example uses preliminary data from analyses involved in the development of a coding protocol, applied to selected data from a large review of parameters in the child treatment outcome literature (Chorpita et al., 2002). The example uses codes for a limited number of practice elements (26) across a limited number of protocols demonstrating efficacy in controlled research ($n = 49$). Thus, the following is not a definitive review or analysis. Detailed comparative analyses using a larger target data set and comparing multiple data mining algorithms will be presented elsewhere.

We examined the following four matching factors: (1) diagnosis (Attention Deficit Hyperactivity Disorder, Opppositional Defiant Disorder, Conduct Disorder, Specific Phobia, Overanxious Disorder, Separation Anxiety Disorder, Social Phobia, Depression), (2) age (2–6, 7–11, 12–17), (3) ethnicity (Asian, Black, Hispanic, Native American, White), and (4) gender (Female, Male). All 49 interventions were coded for the presence or absence of all 26 practice elements specified in the model. The algorithm used was a variant of an automated interaction detector approach, with intraclass correlations (ICC) being used to determine differences among patterns of practice elements. The ICC model estimates the similarity of ratings of k targets among j judges. In our example, practice elements were the targets ($k = 26$), and each grouping of the literature (e.g., anxiety versus depression) was analogous to a set of judges. A significant ICC value signified that the patterns of practice elements were highly correlated across groups of studies being compared.

In the first step of the matching analyses, all pairwise comparisons were conducted within each factor. All comparisons within age and gender yielded nonsignificant ICC values, so all levels were collapsed together within each factor (i.e., boys and girls were combined and age groups were combined at this level of the tree). Profiles for ethnicity collapsed partially, such that Asian grouped separately from all others, which formed a single second group. The seven diagnoses collapsed to four groupings: (1) Disruptive Behavior (Attention Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, Conduct Disorder), (2) Anxiety (Overanxious Disorder, Separation Anxiety Disorder, Social Phobia), (3) Specific Phobia, and (4) Depression. This represented the first reduction of levels within factors.

The next step involved the selection of the factor to represent the first branching of the tree. This was done by selecting the factor that yields the greatest differentiation among patterns of practice elements. Age and gender were not eligible, as each factor had failed to discriminate practice element profiles in the previous step. Ethnicity yielded a higher ICC value than diagnosis, and so diagnosis was selected as the first branch in the tree. Figure 2 shows the observed branching of the tree, and Fig. 3 provides examples of the resulting practice element profiles for the four diagnostic clusters.

Within each branch of the tree, the entire matching process was repeated with the remaining factors (i.e., gender, ethnicity, age). Levels of age and gender

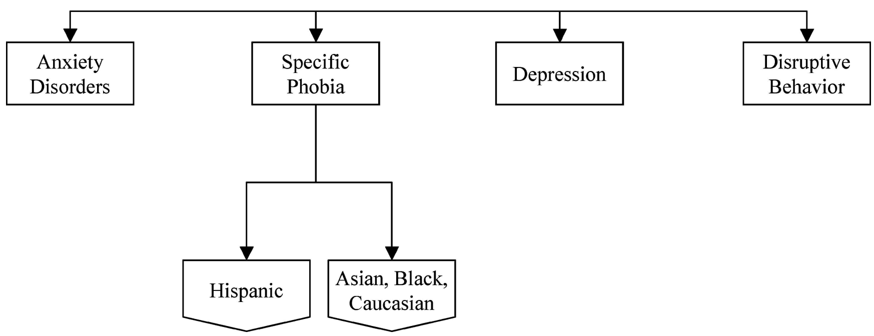

Fig. 2. Decision tree from example data.

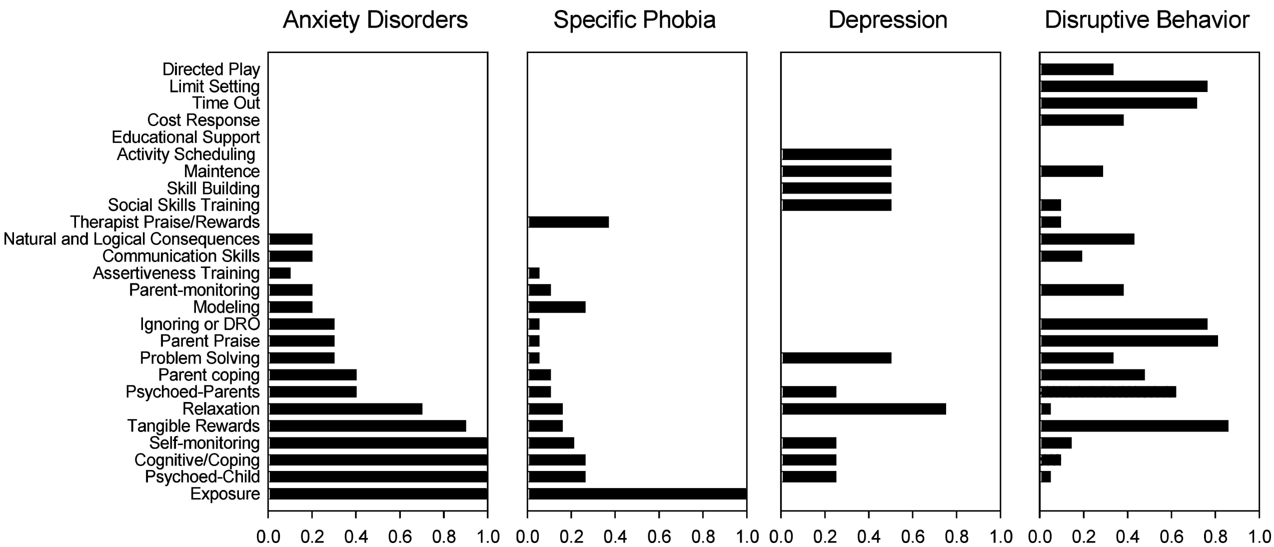

Fig. 3. Practice element profiles corresponding to first split in the decision tree.

merged again. Ethnicity merged completely when examined within depression, disruptive behavior, and anxiety, but yielded a split for ethnicity under specific phobia (see Fig. 4). Within each branch (specific phobia for Hispanics, specific phobia for non-Hispanics), the matching process was repeated, with age levels and gender merging once again. These now formed terminal points or "leaf nodes" in the decision tree. No more branching was possible with the information given (see Fig. 2).

The final step typically involves expert review of the tree, as there can be a tendency for artifactual

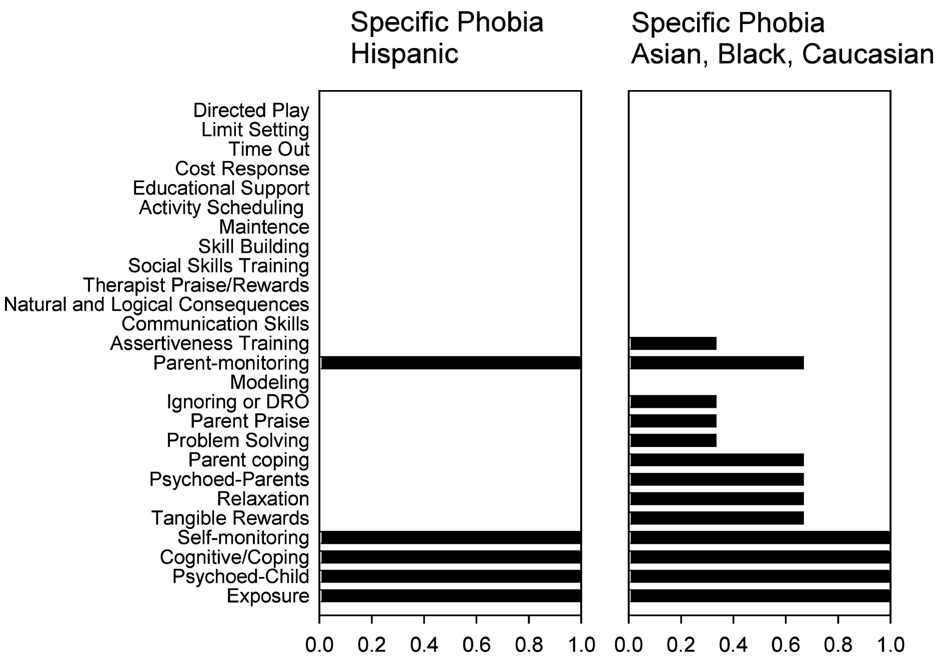

Fig. 4. Practice element profiles corresponding to final nodes in the decision tree.

branching to evolve that is not strongly supported by theory or logic. For example, in the anxiety disorders branch, it is possible that ethnicity might have yielded a split for Caucasian American and Australian National, given that CBT for anxiety has been tested in Australia and the U.S. in slightly different forms. An expert reviewer might find that such a split is confounded by the fact that different manuals were tested in different countries. In such a case, the tree could be "pruned" and the branching at that level would be eliminated, with the practice element profiles averaged at that juncture. In this same way, dozens of other branches have the potential to emerge, differentiate the practice element profiles, and be inspected, confirmed, or disconfirmed by domain experts. Because of the limited nature of the sample data and a criterion designed to minimize oversplitting in our illustration, no further review was necessary.

IMPLICATIONS AND FUTURE DIRECTIONS

The distillation and matching model outlines an approach with potentially high utility in service delivery systems and training programs. To date, no other model systematically classifies interventions according to content and allows for empirical identification of associations among client and contextual variables in the research literature. The implications for decision support in both clinical training and practice point to increased precision in clinical decision-making.

A number of strengths of the model should be noted. First, the selection of variables by the user allows for a highly flexible system of analysis. For example, although this analysis was performed with 26 codes corresponding to clinical techniques, it could just as easily involve any element that are a part of clinical practice (e.g., nontraditional clinical techniques, therapeutic alliance variables), whose association with study variables (e.g., client characteristics, setting variables) is of interest. We view the coding of other factors such as alliance, competence, and assignment of homework, etc. as potentially fruitful avenues for future applications of the DMM. The limits of the model are defined by the scope of interest of the user and by how well such factors are specified in treatment protocols.

Second, the same range of possibilities exists for study variables. For example, we coded age, gender, and ethnicity for the preliminary analysis, but decisions could be made using any other variables of interest. For example, such variables as therapist background, client socioeconomic status, group or individual therapy format, rural or urban setting, among others can be examined for their associations with practice content. Simply put, any information recorded in the literature that can be coded can be used to build a decision tree.

Finally, another notable strength of the model is its dynamic nature. That is, the decision tree of interest can be updated continually with each new study that emerges. Determination of whether a treatment is "new" or finds membership in an existing class is determined through the analysis. As the literature develops, the branching of the decision tree should become more and more highly articulated, to the extent that different treatment content is found to be successful across various contexts.

Despite some of the positive aspects of the model, there are a number of issues that require clarification. The distillation and matching approach works by aggregating information across successful clinical research trials, and thus in essence represents frequency counts of the relative occurrence of practice elements in various contexts. While these frequency counts may be informative, as they point to sets of practice elements that demonstrated positive effects in the context of interest, there is no way to tell from the literature whether certain practice elements are necessary or sufficient for clinical change. For example, a practice element profile for anxiety might show certain practice elements occurring with high frequency in successful trials for anxiety; however, there is no evidence to show that those elements are therapeutically active ingredients of the intervention. All that is known is that aggregating across all successful studies, there is an observed pattern of what techniques are more or less common. Causal inferences cannot be made without specific experimental demonstration.

Along similar lines, the frequency of the occurrence of techniques in the current demonstration of the model regarded each study as an equal contributor to the calculation of a practice element profile. Thus, a study with 100 participants counts the same as one with 25 participants, which could lead to biased estimates of frequency. It is, however, possible to weight the frequency estimates by sample size, such that larger studies contribute more to the final practice element profile. Similarly, effect sizes on outcome variables can be used to weight frequency estimates, such that more effective interventions contribute more heavily to the calculation of a profile. Such methodological variations were beyond the scope of the initial data set used for this model demonstration, but the effects of such variations on the decision model should be evaluated in future investigations with this model.

Another point of concern exists regarding the large amount of missing data in the literature. For example, considering the example decision tree in Fig. 2, it is not immediately apparent what decision would be made regarding a Native American youth with Specific Phobia. One solution that can be employed without making any initial assumptions about the factor of interest (in this case, ethnicity) would be to revert to the higher node. Thus, with this “higher node” approach, selection of an intervention for a Native American youth with Specific Phobia would be guided by overall practice element profile for specific phobia in general. As mentioned earlier, prescriptive use of the decision model leaves one to decide whether (a) to craft new intervention based on the aggregate practice element profile or (b) to select from among the existing protocols that which corresponds most closely to the aggregate profile. In the former case, the youth would be treated with an intervention representing the most common elements across treatments for specific phobia for all youth. In the latter case, one must select a specific manual or protocol, and in essence, will have to select from a lower-order node in the tree. Presumably, the intervention selected would have a profile looking most like the higher node average.

A second way to handle the problem of incomplete data in the literature is to impose a hierarchical structure on factors of interest. For example, if there are no clinical trials for intermittent explosive disorder, the “higher node” approach would suggest in Fig. 2 that one revert to the root node (not shown), which represents the average across all therapy approaches for all disorders. However, imposing a structure on the diagnosis factor, one might choose to classify intermittent explosive disorder as a disruptive behavior disorder. Such decisions are only as good as the assumptions that underlie them, but in some cases they may lead to better choices. Future investigations of the model might benefit from comparative analysis of the “higher node” and “imposed structure” approaches to detail the implications of each type of decision with real data.

As implied in the data mining (e.g., spoiling) examples above, such problems would be more easily handled given the presence of enormous amounts of clinical outcome data, with each factor of interest fully crossed with all others across the literature. Such improved information availability would be expected to benefit the DMM as well as all competing algorithms for making the same decisions (e.g., clinical judgment, unstructured literature review, quantitative meta-analysis, etc.). Alternatively, the limited state of the current knowledge base limits the quality of decisions regardless of the data combination algorithm used. The treatment outcome knowledge base is unlikely to ever support this ideal and strategies for managing gaps in the literature are relevant to all of these approaches. Given the limited state of the knowledge base, additional loss of information through data preparation steps (e.g., spoiling) may be of concern. The decision about what is relevant (i.e., how to code content) is a rational choice and affects all data combination algorithms. Once data are defined, the DMM provides a structured framework for harnessing all available data and systematically organizes choices for informed decision-making. The DMM model is similar to meta-analysis in this regard and represents an improvement over other strategies (e.g., clinical judgment, unstructured literature review).

In summary, the model is intended to allow for a more detailed examination of intervention content and its association with other variables of interest than is currently possible. The implications appear positive for improved understanding of information in the evidence base and implementation practices relating to the summary of available data. Application of this model to reliable data sets involving study information and intervention content is recommended, in the hope that it will yield useful insights for research, training, and practice.

ACKNOWLEDGMENTS

This project was supported through a grant from the Hawaii Department of Health and carried out in conjunction with the Research Network on Youth Mental Health, sponsored by the John D. and Catherine T. MacArthur Foundation.

REFERENCES

Addis, M. E., Wade, W. A., & Hatgis, C. (1999). Barriers to dissemination of evidence-based practices: Addressing practitioners’ concerns about manual-based psychotherapies. *Clinical Psychology: Science and Practice*, *6,* 430–441.

Barkley, R. A. (1997). *Defiant children: A clinician's manual for assessment and parent training* (2nd ed.). New York: Guilford.

Beutler, L. E. (1991). Have all won and must all have prizes? Revisiting Luborsky et al.'s verdict. *Journal of Consulting and Clinical Psychology*, *59,* 226–232.

Beutler, L. E. (2002). The dodo bird is extinct. *Clinical Psychology: Science and Practice*, *9,* 30–34.

Brodley, C. E., Lane, T., & Stough, T. M. (1999). Knowledge discovery and data mining. *American Scientist*, *87,* 54–61.

Chambless, D. L. (2002). Beware the dodo bird: The dangers of overgeneralization. *Clinical Psychology: Science and Practice*, *9,* 13–16.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66,* 7–18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, *52,* 685–716.

Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist*, *49,* 5–18.

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2003). *Modularity in the design and application of therapeutic interventions*. Manuscript submitted for publication.

Chorpita, B. F., Taylor, A. A., Francis, S. E., Moffitt, C. E., & Austin, A. A. (2004). Efficacy of modular cognitive behavior therapy for childhood anxiety disorders. *Behavior Therapy*, *35,* 263–287.

Chorpita, B. F., Yim, L. M., Donkervoet, J. C., Arensdorf, A., Amundsen, M. J., McGee, C., et al. (2002). Toward large-scale implementation of empirically supported treatments for children: A review and observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology: Science and Practice*, *9,* 165–190.

Dawes, R. M. (1992). Quandary: Correlation coefficients and contexts. In L. Montada, S. Filipp, et al. (Eds.), *Life crises and experiences of loss in adulthood* (pp. 521–528). Hillsdale, NJ: Erlbaum.

Eysenck, H. J. (1952). The effects of psychotherapy. *Quarterly Bulletin of the British Psychological Society*, *3,* 41.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Data mining and knowledge discovery in databases: An overview. *Communications of the ACM*, *39,* 27–34.

Greenwald, P., & Cullen, J. W. (1984). The scientific approach to cancer control. *CA – A Cancer Journal for Clinicians*, *34,* 328–332.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29,* 119–127.

Kazdin, A. E., Bass, D., Ayers, W. A., & Rodgers, A. (1990). Empirical and clinical focus of child and adolescent psychotherapy research. *Journal of Consulting and Clinical Psychology*, *58,* 729–740.

Kendall, P. C. (1990). *Cognitive-behavioral therapy for anxious children: Therapist manual*. Ardmore, PA: Workbook.

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin*, *65,* 110–136.

King, N. J., & Ollendick, T. H. (1998). Empirically validated treatments in clinical psychology. *Australian Psychologist*, *33,* 89–95.

Lonigan, C. J., Elbert, J. C., & Bennett Johnson, S. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology*, *27,* 138–145.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice*, *9,* 2–12.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies: Is it true that "everyone has won and all must have prizes"? *Archives of General Psychiatry*, *32,* 995–1008.

National Institutes of Health. (1994). Behavioral therapies development program. *NIH Guide, 22*(26). The URL is http://www.drugabuse.gov/PAs/PA-94-078.html

Paul, G. L. (1967). Outcome research in psychotherapy. *Journal of Consulting and Clinical Psychology*, *31,* 109–118.

Rounsaville, B. J., & Carroll, K. M. (2002). Commentary on dodo bird revisited: Why aren't we dodos yet? *Clinical Psychology: Science and Practice*, *9,* 17–20.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy*. Baltimore: John Hopkins University Press.

Task Force on Promotion and Dissemination of Psychological Procedures, Division of Clinical Psychology, American Psychological Association. (1995). Training in and dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, *48,* 3–23.

Task Force on Psychological Intervention Guidelines, American Psychological Association. (1995). *Template for developing guidelines: Interventions for mental disorders and psychosocial aspects of physical disorders*. Washington, DC: American Psychological Association.

Weersing, V. R., Weisz, J. R., & Donenberg, G. R. (2002). Development of the Therapy Procedures Checklist: A therapist-report measure of technique use in child and adolescent treatment. *Journal of Clinical Child and Adolescent Psychology*, *31,* 168–180.

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral and nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology*, *63,* 317–320.

Weisz, J. R. (2004). *Psychotherapy for children and adolescents: Evidence-based treatments and case examples*. Cambridge, UK: Cambridge University Press.

Weisz, J. R., Chorpita, B. F., Duan, N., Glisson, C., Green, E. P., Hoagwood, K. H., et al. (2003). *Research network on youth mental health: Evidence-based practice for children and adolescents.* Los Angeles: University of California.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, *63,* 688–701.

Weisz, J. R., & Hawley, K. M. (1998). Finding, evaluating, refining, and applying empirically supported treatments for children and adolescents. *Journal of Clinical Child Psychology*, *27,* 206–216.

Weisz, J. R., Weersing, V. R., Valeri, S. M., & McCarty, C. A. (1999). *Therapist's Manual PASCET: Primary and secondary control enhancement training program*. Los Angeles: University of California.

Weisz, J. R., Weiss, B., Alicke, M. D., & Klotz, M. L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, *55*(4), 542–549.

Weisz, J. R., Weiss, B., Han, S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117,* 450–468.