2009

Research Design and Statistical Applications

Grayson Holmbeck
Loyola University Chicago, gholmbe@luc.edu

Kathy Zebracki

Katie McGoron

CHAPTER 4

Research Design and Statistical Applications

GRAYSON N. HOLMBECK
KATHY ZEBRACKI
KATIE McGORON

What is the role of research *in* the field of pediatric psychology? To answer this question, it is useful to imagine what clinical practice would be like if we had no research foundation for our work. Without such a foundation, practitioners would have no basis for suggesting specific interventions or understanding why some interventions are successful and why others fail. Similarly, without a research foundation, assessments conducted with children would be based on unstandardized assessment methods, and no normative data would be available. Clearly, most of us would agree that scientific research is the foundation of pediatric psychology, including all activities in which pediatric psychologists are engaged (Noll, 2002; Roberts & Ilardi, 2003).

The purpose of this chapter is to review research designs and methods in the field of pediatric psychology. We begin with a focus on the importance of *theory* as a basis for conducting pediatric psychology research, and then move on to a discussion of research questions often posed by pediatric psychologists. Next, we provide an overview of research designs commonly used in pediatric psychology, including a review of challenges faced by pediatric psychologists who conduct research in pediatric settings. Moreover, we discuss several methodological and statistical issues that are important to consider in designing research and conducting data analyses. We conclude with a look to the future, discussing recommendations for research in the field of pediatric psychology.

The Importance of Theory in Pediatric Psychology Research

A conceptual model or theoretical framework facilitates the development of a program of research (as opposed to a set of unrelated studies) and drives all aspects of the research

endeavor (Riekert & Drotar, 2000; Thompson & Gustafson, 1996). Influential theories in the field of pediatric psychology tend to share many features: (1) a clarity of focus; (2) a developmental emphasis; (3) the ability to address limitations of previous research; (4) specification of predictors (i.e., independent variables) and outcomes (i.e., dependent variables), with a clear rationale for each; (5) a clear articulation of links between predictors and outcomes (which sometimes involves specification of mediational and moderational effects), with accompanying testable hypotheses; and (6) clear implications for interventions.

Types of Research Questions

After articulating the theory, framework, or model that will be the basis for their investigations, researchers express their research interests in the form of research questions and hypotheses. Kazdin (1999) has outlined several general types of research questions from the field of clinical psychology, and these are the focus of this section.

What Is the Relationship between the Variables of Interest?

Although the first question may be the simplest type of research question, it is also a very common one that has been employed in a variety of research areas. This type of research question incorporates most cross-sectional and longitudinal correlational designs. Although the designs and data analyses used to answer such questions can be quite sophisticated, the *correlation* is the basis for all of these research questions. In some cases (i.e., cross-sectional designs), one can merely document a statistical association between two variables; in other cases (i.e., longitudinal designs), one may be able to determine which variables temporally precede the onset of other variables or changes over time.

What Factors Influence the Magnitude of the Relationship between the Variables?

Variables that have an impact on the association between two or more other variables are typically referred to as "moderator" variables (Baron & Kenny, 1986; Holmbeck, 1997, 2002). A moderator is a variable that influences the strength or the direction of a relationship between a predictor variable and a criterion variable (Figure 4.1). Sup-

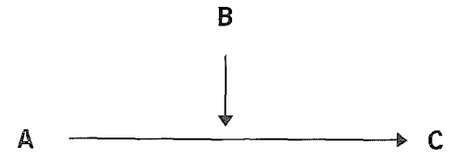

FIGURE 4.1. Moderated relationship among variables (A, predictor; B, moderator; C, criterion/outcome). From Rose, Holmbeck, Coakley, and Franks (2004). Copyright 2004 by Lippincott Williams and Wilkins. Reprinted by permission.

pose a researcher is interested in examining whether the relationship between familial stress and child adjustment to a chronic condition depends on the level of uncertainty that characterizes a child's condition. That is, a significant association between stress and adjustment may emerge *only* when there is considerable uncertainty regarding the child's illness status. By testing "level of uncertainty" as a moderator of the relationship between stress and outcome, the researcher can specify certain conditions under which family stress predicts child adjustment.

Pediatric psychologists often posit moderational processes when conducting studies of risk, protective, and resilience factors (Rose, Holmbeck, Coakley, & Franks, 2004). "Resilience" refers to the process by which children successfully navigate stressful situations or adversity and attain developmentally relevant competencies (Masten, 2001). A "protective" factor either ameliorates negative outcomes or promotes adaptive functioning. The protective factor serves its protective role only in the context of adversity; it does not operate in low-adversity conditions. Protective factors are contrasted with "resource" factors, which have a positive impact regardless of the presence or absence of a stressor (Rutter, 1990; see Figure 4.2). It is also important to note that a protective factor represents a moderational effect (i.e., a statistically significant interaction effect), whereas a resource factor represents an additive effect (i.e., two main effects; Figure 4.2). Risk and vulnerability factors operate in much the same way as resource and protective factors, but in the opposite direction (Figure 4.3). A "vulnerability" factor is a moderator that increases the chances for maladaptive outcomes in the presence of adversity (Rutter, 1990) and only operates in the context of adversity. By contrast, a variable that negatively influences an outcome regardless of the presence or absence of adversity is a "risk" factor (Rutter, 1990; see Figure 4.3).

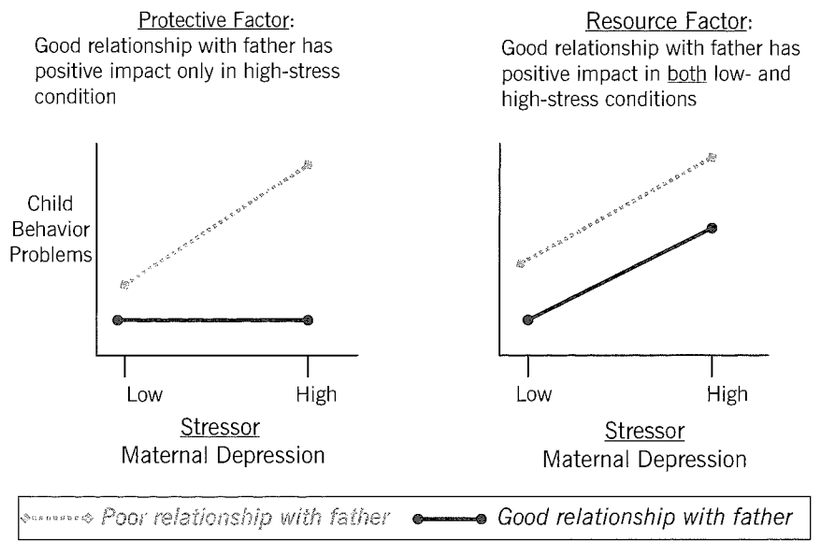

FIGURE 4.2. Protective and resource factors. From Rose, Holmbeck, Coakley, and Franks (2004). Copyright 2004 by Lippincott Williams and Wilkins. Reprinted by permission.

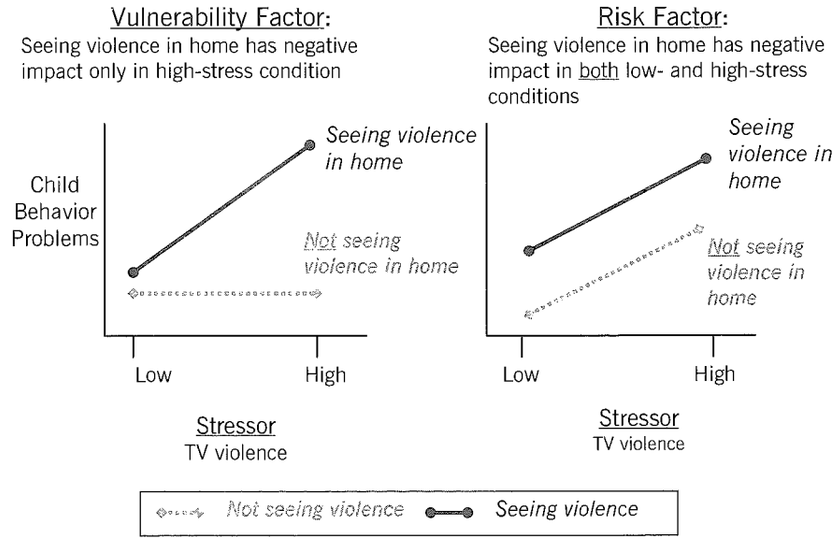

FIGURE 4.3. Vulnerability and risk factors. From Rose, Holmbeck, Coakley, and Franks (2004). Copyright 2004 by Lippincott Williams and Wilkins. Reprinted by permission.

What Mechanism Explains the Relationship between the Variables?

A mechanism that explains "why" two or more variables are associated is often referred to as a "mediator" variable. Often a mediator variable is conceptualized as the mechanism through which one variable (i.e., the predictor) influences another variable (i.e., the criterion; Baron & Kenny, 1986; Holmbeck, 1997, 2002; MacKinnon, 2008; see Figure 4.4). Suppose a researcher finds that parental intrusive behavior is negatively associated with child adherence to a medical regimen. Given these findings, a researcher could explore whether a third variable (e.g., child independence) might account for or explain the relationship between these variables. In this case, parental intrusiveness would have a negative impact on level of child independence, which in turn would contribute to poor medical adherence (Holmbeck, Johnson, et al., 2002; see Figure 4.4). Although the logic underlying meditational models is quite straightforward, several rather complex mediational models have recently been proposed (e. g., see Bauer, Preacher, & Gil's

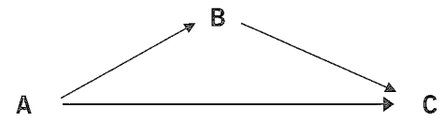

FIGURE 4.4. Mediated relationship among variables (A, predictor; B, mediator; C, criterion/outcome). From Rose, Holmbeck, Coakley, and Franks (2004). Copyright 2004 by Lippincott Williams and Wilkins. Reprinted by permission.

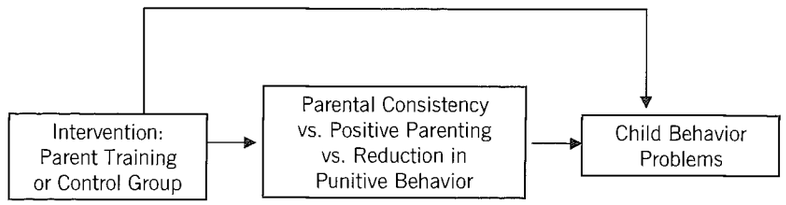

FIGURE 4.5. Mediators in intervention research: Parenting behaviors as mediators of the relationship between parent training (intervention) and child behavior (outcome). From Rose, Holmbeck, Coakley, and Franks (2004). Copyright 2004 by Lippincott Williams and Wilkins. Reprinted by permission.

[2006] discussion of mediation within the context of multilevel modeling, or Rose et al.'s [2004] discussion of mediated moderation and moderated mediation).

A research design that includes random assignment to treatment and control conditions provides a particularly powerful design for drawing conclusions about causal mediational relationships (i.e., "why" an intervention works; Kraemer, Wilson, Fairburn, & Agras, 2002; Weersing & Weisz, 2002). Such intervention/mediation models allow a researcher not only to test potential mediators within an experimental design, but also to examine the differential utility of several mediational variables. In other words, the researcher can determine which mediator best accounts for the effectiveness of a given treatment (e.g., see Figure 4.5; Forgatch & DeGarmo, 1999).

Are There Differences between Groups?

Research questions often focus on differences between groups—for example, whether children with a chronic condition have the same number of peer friendships as children without a chronic condition. Such research can be very complex, including questions of whether groups differ in adjustment trajectories over time. Although this type of research question is a variation on the correlational question posed above, group differences research tends to focus on the hypothesized differentiation of discrete groups, rather than on associations between two or more continuous variables. Perhaps the most compelling group differences research involves randomized controlled trials, where the investigators are interested in whether outcomes differ between a treatment group and a control condition after participants have been randomly assigned to the different groups. We turn to this type of research in the next section.

Research Designs in Pediatric Psychology

In this section, several types of designs and research strategies are discussed: (1) experimental and treatment outcome research, (2) quasi-experimental designs, (3) observational research designs, (4) single-participant designs, and (5) meta-analytic techniques. To conclude this section, we discuss several challenges that are specific to conducting research with pediatric populations.

Experimental and Treatment Outcome Designs

Well-designed and well-implemented randomized controlled trials (sometimes referred to as randomized clinical trials or RCTs) are considered the "gold standard" in evaluating the efficacy of behavioral interventions and ensuring unbiased comparisons across groups (Altman et al., 2001). Most importantly, they are the basis for determining whether an intervention can be classified as "empirically supported" (i.e., evidence-based; see Nelson & Steele, Chapter 7, this volume; see also Beale, 2006; Chambless & Ollendick, 2001; Kazdin & Weisz, 1998; Spirito, 1999). The most noteworthy feature of an RCT is that participants are randomly assigned to conditions—a design feature that addresses most of the threats to the internal validity of the study. RCTs, however, are not flawless; inadequate methodologies can lead to biased results, which misinform clinical practice and decision making in health care policy (Moher et al., 1998; Schulz, Chalmers, Hayes, & Altman, 1995). Moreover, given that participants are randomly *assigned* to conditions, such designs do not advance our knowledge of how individuals select, enter into, and engage in treatment.

Internal and External Validity

One of the strengths of RCTs is that they directly address issues related to the internal validity of the study. These threats to the validity of the findings have been discussed in detail in several texts that focus on research methodology (e.g., Kazdin, 2003). Briefly, the degree to which an experiment has internal validity relates to whether group differences (i.e., treatment vs. control) can be attributed to the intervention rather than to other extraneous factors (Kazdin, 1999, 2003). Another way to put it is that the investigators are interested in ruling out alternative explanations for their findings by eliminating all differences between the groups other than the intervention manipulation. Indeed, there are several types of confounds (or factors) that may operate differentially across groups (e.g., historical factors, the effects of assessment on the outcomes of interest, differential attrition; Kazdin, 1999). Threats to the external validity of the study focus on the degree to which the findings of the study can be generalized to circumstances that may differ from the experimental conditions characterizing a given study (Kazdin, 1999, 2003).

Control Groups

A critical decision in designing an RCT is the choice of a control condition (Kendall, Flannery-Schroeder, & Ford, 1999). If one is working in a relatively new area of research, one may ask whether an intervention is more effective than the absence of any form of intervention. In this case, one may be interested in including a no-treatment control group. A useful alternative to the no-treatment control condition is to include either an attention placebo control group or a standard care control condition (Kendall et al., 1999). These types of control groups address concerns related to "attention" from the interventionist. In the case of the attention placebo control group, the participants who have been randomly assigned to the control condition are exposed to a "treatment," which is expected to be ineffective in producing significant change in the outcome of interest. Standard care control groups can be employed when the popula-

tion of interest is already exposed to some level of treatment because of a condition inherent to the population (e.g., standard clinic care in children with Type 1 diabetes). If a treatment has already been shown to be effective in prior work, investigators may choose to employ a waiting-list control condition, whereby the control group will receive the treatment *after* the study is completed (Kendall et al., 1999). There are two advantages of this strategy: (1) All participants in the study will eventually be given the opportunity to receive the treatment; and (2) the waiting-list condition can be assessed for treatment effects after they have been exposed to the intervention, thus providing a cross-validation of findings. Finally, in the case where there is already sufficient evidence that a treatment condition "works" better than no treatment, the treatment of interest can be compared to an alternative treatment that has been shown to be effective in past research.

Intent-to-Treat Analyses

In any longitudinal research, it is rare that all participants who begin a study complete all components of the study over time. The same could be said for an RCT, which is a type of longitudinal study (given the use of pretesting, posttesting, and follow-up assessments). In many studies, there are differences between participants who complete the study and those who do not, which can undermine the external validity of the investigation.

This issue of attrition takes on added importance in RCTs. In an RCT, if one examines treatment effects only for those who completed the study, such effects may be exaggerated (or biased), because those who were not benefiting from the treatment may be the same participants who dropped out of the study. Those who conduct RCTs have developed a method for managing this problem—namely, intent-to-treat analyses (Hollis & Campbell, 1999; LaValley, 2003). When conducting data analyses, an investigator includes all participants from the groups to which they were randomized, regardless of whether they dropped out of the study. Several approaches to intent-to-treat analyses have been employed (Hollis & Campbell, 1999). Some use the last-observation-carried-forward (LOCF) strategy to manage missing values in the context of a longitudinal study (including RCTs) (LaValley, 2003; Streiner, 2002). With this approach, the last value reported for a respondent who has dropped out of the study is carried forward and is used for all subsequent "missing" data points. As suggested by Streiner (2002), multiple-imputation analyses or growth curve analyses will be less biased than the LOCF approach. With multiple-imputation analyses, missing values are replaced with values that have been "imputed" (or estimated) from data provided by other participants in the data set (Little & Rubin, 2002). With growth curve analyses, missing values are not imputed; instead, all data from the participants are utilized, and a curve is generated for each participant based on all available data (Singer & Willett, 2003).

Clinical Significance

When conducting an RCT, one may find statistical differences between the groups at posttesting; however, if the sample sizes are quite large, the actual differences between

the groups may be very slight. As discussed by Kendall and colleagues (1999), the clinical significance of an intervention is important to assess as an adjunct to an evaluation of statistical significance. With clinical significance, one is assessing the degree to which the participants no longer suffer from the condition that made them eligible for the RCT. Several strategies can be used to document clinical significance (e.g., the number of participants whose scores on the outcome of interest have moved into the normative range, or whether participants continue to meet diagnostic criteria for the condition of interest; Kazdin, 2003).

The CONSORT Criteria

Reporting findings from an RCT in a clear and comprehensive manner is essential for determining the internal and external validity of the intervention. The Consolidated Standards of Reporting Trials (CONSORT) statement, published in 1996 (Begg et al., 1996) and revised in 2001 (Altman et al., 2001), was designed to facilitate critical review and understanding of RCTs by guiding authors on how to report trials and guiding reviewers on how to systematically evaluate the findings of RCTs. The CONSORT statement includes a 22-item checklist (Figure 4.6) and flow diagram (Figure 4.7) of essential data to be included when reporting on an RCT. Readers are referred to *www.consort-statement.org* for the full statement and a detailed explanation of the checklist items.

The CONSORT statement was initially developed for use with a two-group, parallel-design medical intervention trial; however, modifications and extensions for use with other designs, types of interventions, and data have been made (Moher, Altman, Schulz, & Elbourne, 2004). Stinson, McGrath, and Yamada (2003) found that CONSORT items are applicable to psychological interventions; however, Drotar (2002) found that most reports of pediatric RCTs failed to provide the information necessary to assess the studies' validity and to apply the interventions in clinical practice. Most recently, the CONSORT Group developed an extension for trials assessing nonpharmacological treatments, such as behavioral interventions (Boutron, Moder, Altman, Schulz, & Ravaud, 2008). Moreover, the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement was developed to provide guidelines for nonrandomized designs similar to those that CONSORT provides for RCTs (Des Jarlais, Lyles, Crepaz, & TREND Group, 2004). Readers are referred to *www.trend-statement.org* for a copy of the TREND checklist.

In addition to the standard CONSORT checklist and flowsheet, Davidson and colleagues (2003) suggest that investigators report on the five following items when conducting RCTs in behavioral medicine: (1) background training and professional credentials of the treatment providers; (2) type, duration, and form of supervision of the treatment providers; (3) treatment preference or allegiance of the treatment providers and patients; (4) manner of testing and treatment delivery; and (5) treatment fidelity. Furthermore, Wysocki (2008) recommends that the following additional elements be considered by those submitting manuscripts reporting RCTs to the *Journal of Pediatric Psychology*: (1) attention to ethical issues, (2) verification of treatment integrity, (3) attention to cost effectiveness and dissemination of the intervention, and (4) registration of the clinical trial (e.g., *www.clinicaltrials.gov*).

PAPER SECTION and topic	**Item**	**Descriptor**	**Reported on page #**
TITLE & ABSTRACT	1	How participants were allocated to interventions (e.g., "random allocation," "randomized," or "randomly assigned").	
INTRODUCTION Background	2	Scientific background and explanation of rationale.	
METHODS Participants	3	Eligibility criteria for participants, and the settings and locations where the data were collected.	
Interventions	4	Precise details of the interventions intended for each group, and how and when they were actually administered.	
Objectives	5	Specific objectives and hypotheses.	
Outcomes	6	Clearly defined primary and secondary outcome measures, and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	
Sample size	7	How sample size was determined, and, when applicable, explanation of any interim analyses and stopping rules.	
Randomization—Sequence generation	8	Method used to generate the random allocation sequence, including details of any restrictions (e.g., blocking, stratification)	
Randomization—Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	
Randomization—Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment. If done, how the success of blinding was evaluated.	
Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); methods for additional analyses, such as subgroup analyses and adjusted analyses.	
RESULTS Participant flow	13	Flow of participants through each stage (a diagram is strongly recommended). Specifically, for each group, report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome. Describe protocol deviations from study as planned, together with reasons.	
Recruitment	14	Dates defining the periods of recruitment and follow-up.	
Baseline data	15	Baseline demographic and clinical characteristics of each group.	

(cont.)

FIGURE 4.6. CONSORT statement checklist: Items to include in reporting an RCT. From *www.consort-statement.org*. Copyright by The CONSORT Group. Reprinted by permission. The CONSORT Statement is a document that is periodically updated to account for the evolving nature of the research that supports it. It is currently being updated, with an anticipated publication date of late 2009. Upon publication of this next revision, the CONSORT 2001 checklist and flow diagram being used in this chapter will become outdated. Please refer to *www.consort-statement.org* to ensure that you are always using the most updated version of the CONSORT Statement.

PAPER SECTION and topic	Item	Descriptor	Reported on page #
Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention to treat." State the results in absolute numbers when feasible (e.g., 10/20, not 50%).	
Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (e.g., 95% confidence interval).	
Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those prespecified and those exploratory.	
Adverse events	19	All important adverse events or side effects in each intervention group.	
DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with multiplicity of analyses and outcomes.	
Generalizability	21	Generalizability (external validity) of the trial findings.	
Overall evidence	22	General interpretation of the results in the context of current evidence.	

FIGURE 4.6. *(cont.)*

Quasi-Experimental Designs

As discussed by Greenhoot (2003), the primary difference between experimental and quasi-experimental intervention designs is that the former designs involve random assignment of participants to levels of the independent variable (e.g., intervention vs. control in an RCT), whereas the latter do not involve random assignment. Quasi-experimental designs are often the method of choice when random assignment to conditions is not possible. The most common quasi-experimental design is the nonequivalent control group design. For example, suppose one is interested in comparing outcomes of two camp programs for children with attention-deficit/hyperactivity disorder, and random assignment to camps is not feasible. Of course, the potential limitation of this type of design is that there are selection differences between the camp programs (e.g., there may be demographic differences between the children who select one program vs. the other program). The use of a pretest is an important feature of this design, and demographic differences between groups can be controlled as covariates.

Observational Research Designs

Most research in pediatric psychology employs observational research designs and methods. Kazdin (2003) and Mann (2003) have reviewed different types of designs that fall into this category, including (1) cohort studies and (2) case–control studies. Cohort studies are used to examine variables that precede the development of some outcome. They can also be used to determine the "incidence" of a condition (i.e., the number of new cases of a condition over time within a specified population of interest). For exam-

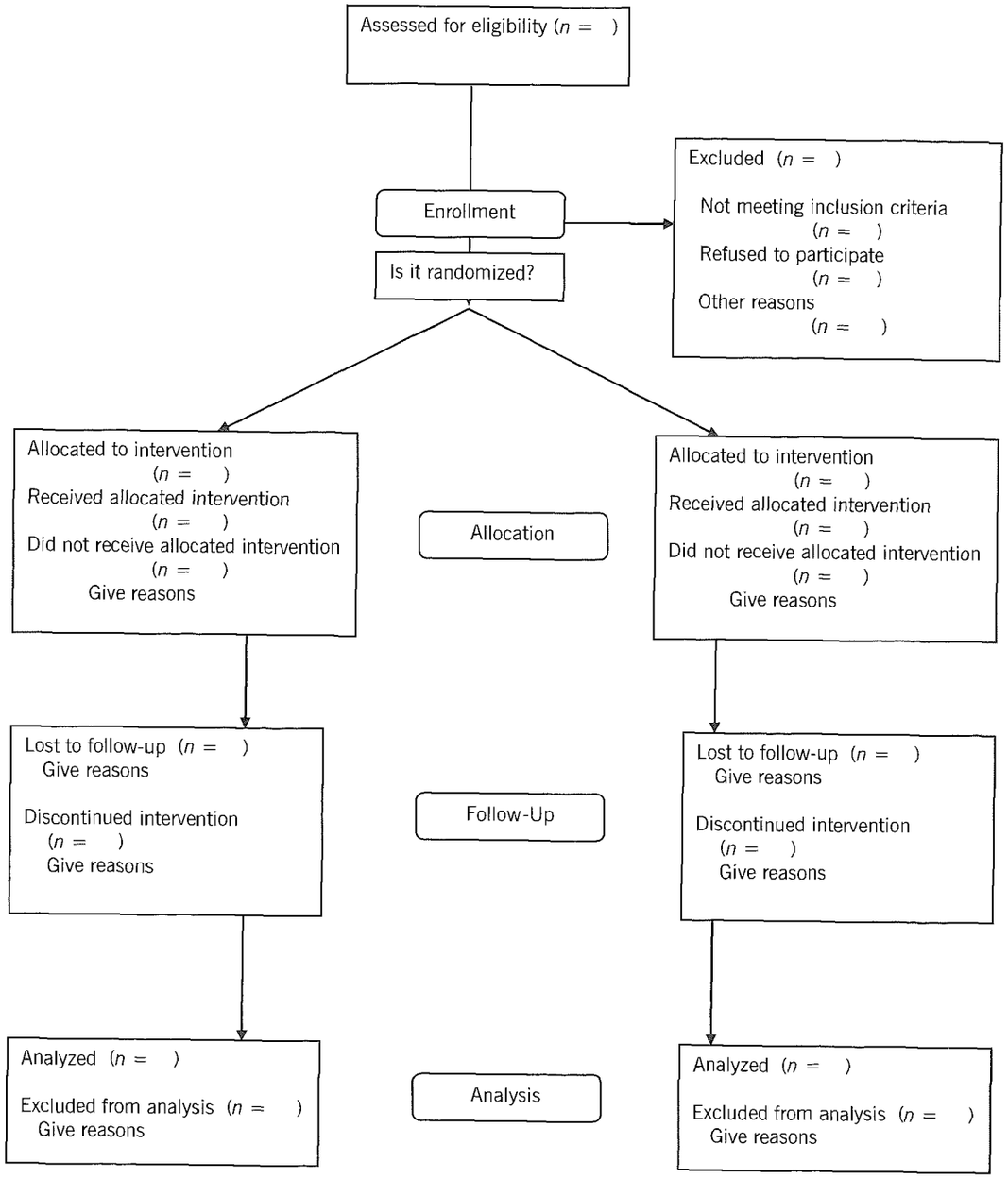

FIGURE 4.7. The CONSORT flowchart. From *www.consort-statement.org*. Copyright by The CONSORT Group. Reprinted by permission. The CONSORT Statement is a document that is periodically updated to account for the evolving nature of the research that supports it. It is currently being updated, with an anticipated publication date of late 2009. Upon publication of this next revision, the CONSORT 2001 checklist and flow diagram being used in this chapter will become outdated. Please refer to *www.consort-statement.org* to ensure that you are always using the most updated version of the CONSORT Statement.

ple, one might examine a cohort of individuals over time to determine what variables are associated with the occurrence of lung cancer or a heart attack (Mann, 2003). Or one might conduct a longitudinal study of children exposed to a hurricane to determine what variables are associated prospectively with the onset of posttraumatic stress disorder symptoms (Kazdin, 2003). The advantage of cohort designs is that they allow one to establish a time line that precedes the outcome of interest with predictors that are not biased by the occurrence of the outcome (Kazdin, 2003). In a case–control study, the investigator identifies samples that do or do not exhibit the outcome of interest (e.g., depression, divorce, a traumatic brain injury). Such a design can also be used to determine the "prevalence" of a condition (i.e., the frequency of a condition's occurrence at a certain point in time). An important difference between case–control studies and cohort studies is that cohort studies follow a group of participants who have not yet exhibited the outcome of interest to determine who will and who will not exhibit the outcome of interest (Kazdin, 2003; Mann, 2003). In case–control studies, those who already have the outcome are compared with those who do not. The most common case–control design is cross-sectional, in which two groups are compared on variables of interest.

Single-Participant Designs

Single-participant designs have long been used in measuring intervention effects at the individual level (Barlow & Hersen, 1984; Kazdin, 1982) and have significantly contributed to our knowledge base in pediatric psychology (Rapoff & Stark, 2008). Single-participant designs are fundamentally similar to group comparison approaches; however, participants are used as their own controls. Although there are several design options for single-participant techniques, all designs share at least four common characteristics: (1) objective data/baseline assessment, (2) continuous assessment, (3) change in only one variable at a time, and (4) replication across individuals or dependent variables.

There are several single-participant design options, with the most common being (1) A-B designs, (2) reversal designs, (3) multiple-baseline designs, and (4) changing-criterion designs (Barlow & Hersen, 1984; Kazdin, 1982). The simplest method, the A-B design, allows comparison of baseline behavior (i.e., "A," usual care or no treatment) and behavior after an intervention or treatment (i.e., "B"); it is most suitable for use when a return to a baseline condition is unethical, impractical, or undesired. Reversal designs, also known as A-B-A or A-B-A-B, are extensions of the A-B design with baseline and intervention phases repeated. One strength of such designs (relative to the A-B design) is the ability to show a functional relationship between the intervention and outcome over time. A multiple-baseline design consists of a series of A-B designs that can be implemented within the same individual across different behaviors, within the same individual across different settings, or within the same behavior across different individuals. Finally, a changing-criterion design is an A-B design involving multiple interventions following an initial baseline, with the criterion for successful outcomes becoming more stringent over time.

Single-participant designs have several advantages. First, they allow for examination of interparticipant and intraparticipant variability in outcomes. Second, single-participant designs can accommodate small sample sizes, such as those in studies of rare conditions, and can be used when withholding treatment is unpractical or unethical. Third, these designs may enhance clinical practice by allowing clinicians to monitor

and assess real-time change and to modify interventions accordingly. Finally, single-participant studies can serve as an initial step in developing empirically validated treatments and evidence-based practices. There are also limitations to using single-participant designs. Lack of generalizability is the most prominent threat to external validity; however, this can be addressed by replication (e.g., repeating the same procedures with several additional patients). There are also several threats to internal validity that need to be considered, such as the impact of extraneous events, maturation effects, carryover effects, and multiple-intervention inference.

Meta-Analytic Techniques

Meta-analysis is a technique used to summarize and pool results from multiple studies to produce aggregated outcomes (Durlak, 1999; Lipsey & Wilson, 2001). Because a major obstacle in conducting research in the field of pediatric psychology is the recruitment of large samples, meta-analysis may have a higher level of utility in this field (by aggregating data across multiple small-sample studies). Within the literature on intervention, meta-analysis can highlight successful treatments as well as promising new directions. At the most general level, a meta-analysis is conducted as follows: (1) A research question is formulated, and hypotheses are clearly stated; (2) a comprehensive sample of studies is obtained (i.e., one conducts a thorough literature review of both published and unpublished studies and selects studies based on explicit inclusionary criteria); (3) information from individual research reports is coded; (4) analyses are conducted with statistics specially designed for meta-analyses; and (5) conclusions are drawn, and recommendations for future research are provided.

Researchers employ measures of effect size to convey results in meta-analysis. Although different studies may make use of different measures of effect size, one common index is Cohen's *d*. In the context of an RCT, this effect size is calculated by subtracting the mean of the control group from the mean of the target group, divided by a pooled standard deviation (other statistics often used in meta-analysis are the product–moment correlation [*r*] and odds ratios). Simply put, effect sizes express the magnitude of difference between two groups in standard deviation units, which allows results across studies to be compared and pooled.

Challenges in Conducting Research with Pediatric Populations

Several research issues pertain specifically to the study of pediatric populations. First, it is important to determine the setting in which the data will be collected. Because many pediatric populations regularly attend hospital clinics, clinic-based data collections may be a relatively efficient strategy. On the other hand, there are certain drawbacks to this strategy: (1) Children and/or parents may be particularly stressed during clinic visits; (2) a child is often accompanied by only one caregiver, making it difficult to assess all family members; and (3) clinic settings are busy environments, which may be distracting to research participants. Data collections from children with a chronic condition may also be complicated if there are cognitive impairments accompanying the condition or if there is a temporary exacerbation of the condition. In a longitudinal study with a pediatric population, a researcher is studying a physical condition that may change over time (with respect to presentation or severity). Moreover, treatments for the condition may also change over time, which could have an impact on the severity of the condition. In such work, whether cross-sectional or longitudinal, sample sizes are another very important concern. Studies in pediatric populations are often underpowered (even when there is a low level of attrition); thus multisite trials are common.

Methodological and Statistical Issues

In this section, we discuss several issues relevant to data collection and statistical analyses in the field of pediatric psychology.

Multisource, Multimethod Data in Pediatric Settings

Once a researcher has formulated a particular research question, decisions need to be made concerning the research design, including the nature of the data to be collected. For instance, what sources or informants will provide the data? And what methods will be used to collect the data? Answers to these questions are critical, because they will have an impact on the ability to rule out alternative explanations for the findings (see Holmbeck, Li, Schurman, Friedman, & Coakley, 2002, for an extended discussion of issues related to the collection and management of multisource, multimethod data; see also Palermo & Wilson, Chapter 15, this volume, for information on methods of collecting data electronically).

Strategies for Managing Attrition and Retention of Participants

In conducting an RCT or any type of longitudinal study, attending to issues of attrition and retention is critical. Several strategies are available to reduce attrition. First, it is helpful to foster the participants' commitment to the study. This can be accomplished by sending project newsletters to participants, although it is critical that the primary hypotheses of the study not be revealed in such newsletters. Second, it is important to develop a tracking system to keep participants' contact information current. Third, at each data collection point, it is important to gather all current contact information (including email addresses), as well as contact information for individuals who will always know the whereabouts of a given participant. Finally, if researchers have funds to compensate participants for their work, they can increase the compensation at each data collection point, with a "bonus" provided to those who complete all data collections (although researchers should avoid making such inducements coercive).

Cleaning Data

Using strategies to ensure the integrity of data is critical (Farrell, 1999). For example, after data have been entered, it is important to run frequency analyses on all variables to check for out-of-range values. Moreover, it is useful to employ double-data-entry procedures to detect errors in data entry. It is beneficial to enter data at the item level, rather than at the scale level, so that psychometrics can be examined (e.g., alpha coef-

ficients). One also needs to be attentive to when items need to be recoded (in cases where the item is keyed in a direction opposite to that of the scale of which it is a part). Moreover, one also has to make decisions about how to handle missing values (Farrell, 1999; Little & Rubin, 2002). Once the data have been cleaned and decisions have been made about missing values, it is useful to examine the data for "univariate outliers" (i.e., values that fall outside the typical range for one's sample), as well as for "multivariate outliers" (i.e., unusual combinations of scores across variables for given participants) (see Tabachnick & Fidell, 2007). If a variable is significantly skewed, it is useful to consider data transformations (e.g., log transformations) (Farrell, 1999; Tabachnick & Fidell, 2007).

Cultural and Ethnic Factors

The field of pediatric psychology has witnessed a shift of emphasis to multiculturalism and diversity (Clay, Mordhurst, & Lehn, 2002). Prevalence rates of many diseases vary by race and ethnicity (e.g., obesity, sickle cell disease, spina bifida, Tay–Sachs disease) (Clay et al., 2002), and treatment success is often moderated by cultural and ethnic variables (Clay et al., 2002). Interestingly, Clay and colleagues (2002) conducted a review of 71 empirically supported treatments in pediatric psychology (the reports were published in 1965–1997), and found that only 27% of the studies reported the racial or ethnic composition of the sample and only 18% reported the socioeconomic status (SES) of the sample. These authors recommended that investigators take the following issues into consideration when conducting culturally oriented research in pediatric psychology: (1) the influence of culturally relevant family constructs; (2) the degree to which health care beliefs, practices, and utilization may be influenced by culture; (3) ways in which treatments can address the unique barriers faced by low-SES families and those from underrepresented groups; (4) the independent and interactional effects of health and minority status; (5) ways in which some cultural variables may be protective; (6) the cultural appropriateness of assessment measures; and (7) the degree to which cultural issues are considered in interpreting research results.

Power, Effect Sizes, and Confidence Intervals

As of 2007, the *Journal of Pediatric Psychology* has required that investigators include effect sizes and confidence intervals in their submitted manuscripts, when appropriate (see also Wilkinson & Task Force on Statistical Inference, 1999). Given that studies in the field of pediatric psychology usually have small sample sizes, these recommendations are particularly relevant. If investigators were to focus only on statistical significance, a correlation coefficient of .30, for example, might be significant in one sample but nonsignificant in another sample, depending on the sample size. But the effect size for an r of .30 would be identical across the two studies (in fact, r is a measure of effect size). Several papers have appeared that demonstrate methods for computing effect sizes (e.g., Rosenthal, 1994). To determine the sample size necessary to detect an effect of a given size, one typically conducts a power analysis prior to collecting data (Wilkinson & Task Force, 1999). Finally, confidence intervals provide "margins of error" around a statistical value; in other words, it is a measure of the precision of a statistical value. For example, one might compute confidence intervals around a mean, which is computationally a function of (but not equivalent to) the standard error (Cumming & Finch, 2005). One then plots one's findings with confidence intervals (or error bars).

Suggestions for Conducting Data Analyses and Protecting Statistical Conclusion Validity

Perhaps the best advice that can be given about data analyses is the following: "Although complex designs and state-of-the-art methods are sometimes necessary to address research questions effectively, simpler classical approaches can often provide elegant and sufficient answers to important questions. Do not choose an analytic method to impress your readers or to deflect criticism" (Wilkinson & Task Force, 1999, p. 598). More generally, Kazdin (2003) discusses several possible threats to statistical conclusion validity, or the statistical evaluation component of the study, that have an impact on the quality of the study's conclusions: (1) low statistical power, (2) violated assumptions of statistical tests, (3) a lack of reliability for some or all of the measures, (4) running large numbers of analyses, and (5) random heterogeneity in the respondents.

Conclusions and Recommendations: The State of the Art and a Look to the Future

The purpose of this chapter has been to highlight issues for investigators to consider when designing research in the field of pediatric psychology. We have intentionally begun our discussion with a focus on theory, because we believe that the process of theory generation drives all other aspects of the research endeavor. We now offer several directions for future research in the field of pediatric psychology, based on our review. First, we recommend that more research be longitudinal and developmentally oriented (Holmbeck, Bruno, & Jandasek, 2006). Second, we recommend that researchers go beyond examining bivariate associations between predictors and outcomes in single pediatric samples. Third, we recommend that scholars attempt to specify and examine the influence of moderator variables; in this way, they should be able to determine to whom the effects apply or do not apply. Fourth, for findings that have considerable support in the literature, we suggest that researchers begin to theorize about variables that may explain (or mediate) such associations. Fifth, we recommend careful attention to issues of internal and external validity in designing a study, to rule out alternative explanations for the findings. Finally, we suggest that researchers take good care of their data by minimizing missing data, cleaning the data prior to conducting data analyses, attending to issues of data integrity (data distributions, outliers), and attempting to increase retention. With advances in research on pediatric populations, we will understand better the impact of chronic conditions as these conditions unfold over time. With such understanding, we will be able to design developmentally relevant intervention strategies for such youths and their families.

Acknowledgment

Completion of this chapter was supported by a research grant from the National Institute of Child Health and Human Development (No. R01-HD048629).

References

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, *134*, 663–694.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs* (2nd ed.). New York: Pergamon Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163.

Beale, I. L. (2006). Scholarly literature review: Efficacy of psychological interventions for pediatric chronic illnesses. *Journal of Pediatric Psychology*, *31*, 437–451.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reporting randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*, *276*, 637–639.

Boutron, I., Moher, D., Altman, D., Schulz, K., & Ravaud, P. (2008). Extending the CONSORT statement for randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine*, *148*, 295–309.

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

Clay, D. L., Mordhurst, M. J., & Lehn, L. (2002). Empirically supported treatments in pediatric psychology: Where is the diversity? *Journal of Pediatric Psychology*, *27*, 325–337.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist*, *60*, 170–180.

Davidson, K., Goldstein, M., Kaplan, R., Kaufmann, P., Knatterud, G., Orleans, C. T., et al. (2003). Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine*, *26*, 161–171.

Des Jarlais, D., Lyles, C., Crepaz, N., & TREND Group. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, *94*, 361–366.

Drotar, D. (2002). Enhancing reviews of psychological treatments with pediatric populations: Thoughts on next steps. *Journal of Pediatric Psychology*, *27*, 167–176.

Durlak, J. A. (1999). Meta-analytic research methods. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 419–428). New York: Wiley.

Farrell, A. D. (1999). Statistical methods in clinical research. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 72–106). New York: Wiley.

Forgatch, M. S., & DeGarmo, D. S. (1999). Parenting through change: An effective prevention program for single mothers. *Journal of Consulting and Clinical Psychology*, *67*, 711–724.

Greenhoot, A. F. (2003). Design and analysis of experimental and quasi-experimental investigations. In M. C. Roberts & S. S. Ilardi (Eds.), *Handbook of research methods in clinical psychology* (pp. 92–114). Oxford, UK: Blackwell.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis?: Survey of published randomized controlled trials. *British Medical Journal*, *319*, 670–674.

Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, *65*, 599–610.

Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology*, *27*, 87–96.

Holmbeck, G. N., Bruno, L. F., & Jandasek, B. (2006). Longitudinal research in pediatric psychology: An introduction to the special issue. *Journal of Pediatric Psychology*, *31*, 995–1001.

Holmbeck, G. N, Johnson, S. Z., Wills, K. E., McKernon, W., Rose, B., Erklin, S., et al. (2002). Observed and perceived parental overprotection in relation to psychosocial adjustment in preadolescents with a physical disability: The mediational role of behavioral autonomy. *Journal of Consulting and Clinical Psychology*, *70*, 96–110.

Holmbeck, G. N., Li, S. T., Schurman, J. V., Friedman, D., & Coakley, R. M. (2002). Collecting and managing multisource and multimethod data in studies of pediatric populations. *Journal of Pediatric Psychology*, *27*, 5–18.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.

Kazdin, A. E. (1999). Overview of research design issues in clinical psychology. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp.3–30). New York: Wiley.

Kazdin, A. E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Kazdin, A. E., & Weisz, J. R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*, *66*, 19–36.

Kendall, P. C., Flannery-Schroeder, E. C., & Ford, J. D. (1999). Therapy outcome research methods. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 330–363). New York: Wiley.

Kraemer, H. C., Wilson, T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, *59*, 877–883.

LaValley, M. P. (2003, October). *Intent-to-treat analysis of randomized clinical trials*. Paper presented at the meetings of the American College of Rheumatology and the Association of Rheumatology Health Professionals, Orlando, FL.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

Mann, C. J. (2003). Observational research methods. Research design II: Cohort, cross-sectional, and case control studies. *Emergency Medicine Journal*, *20*, 54–60.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist*, *56*, 227–238.

Moher, D., Altman, D. G., Schulz, K. F., & Elbourne, D. (2004). Opportunities and challenges for improving the quality of reporting clinical research: CONSORT and beyond. *Canadian Medical Association Journal*, *171*, 349–350.

Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., et al. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, *352*, 609–613.

Noll, R. B. (2002). How to build a better mousetrap: Introduction to the special issue on methodology and design. *Journal of Pediatric Psychology*, *27*, 1–3.

Rapoff, M., & Stark, L. (2008). Editorial: *Journal of Pediatric Psychology* statement of purpose: Section on single-subject studies. *Journal of Pediatric Psychology*, *33*, 16–21.

Riekert, K. A., & Drotar, D. (2000). Adherence to medical treatment in pediatric chronic illness: Critical issues and answered questions. In D. Drotar (Ed.), *Promoting adherence to medical treatment in chronic childhood illness* (pp. 3–32). Mahwah, NJ: Erlbaum.

Roberts, M. C., & Ilardi, S. S. (Eds.). (2003). *Handbook of research methods in clinical psychology*. Oxford, UK: Blackwell.

Rose, B. M., Holmbeck, G. N., Coakley, R. M., & Franks, E. A. (2004). Mediator and moderator effects in developmental and behavioral pediatric research. *Journal of Developmental and Behavioral Pediatrics, 25*, 58–67.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Sage.

Rutter, M. (1990). Psychosocial resilience and protective mechanisms. In J. Rolf, A. S. Masten, D. Cicchetti, K. H. Nuechterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 181–214). New York: Cambridge University Press.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association, 273*, 408–412.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.

Spirito, A. (1999). Introduction [to special series of papers on empirically supported treatments in pediatric psychology]. *Journal of Pediatric Psychology, 24*, 87–89.

Stinson, J., McGrath, P., & Yamada, J. (2003). Clinical trials in the *Journal of Pediatric Psychology*: Applying the CONSORT statement. *Journal of Pediatric Psychology, 28*, 159–167.

Streiner, D. L. (2002). The case of the missing data: Methods of dealing with dropouts and other research vagaries. *Candian Journal of Psychiatry, 47*, 68–75.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Thompson, R. J., & Gustafson, K. E. (1996). *Adaptation to chronic childhood illness*. Washington, DC: American Psychological Association.

Weersing, V. R., & Weisz, J. R. (2002). Mechanisms of action in youth psychotherapy. *Journal of Child Psychology and Psychiatry, 43*, 3–29.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.

Wysocki, T. (2008). Editorial: *Journal of Pediatric Psychology*—Statement of purpose section on randomized trials. *Journal of Pediatric Psychology, 33*, 12–15.

CHAPTER 5

Health Insurance and Pediatric Psychology Services

WILLIAM DOUGLAS TYNAN
MEREDITH LUTZ STEHL
JENNIFER SHROFF PENDLEY

Mental health insurance and reimbursement of services are of utmost concern to pediatric psychology and its practice. Reviewing the history of private and federal insurance makes it possible to gain a better understanding of the system that guides insurance and reimbursement models that function today. Moreover, the everyday practice of pediatric psychology requires an understanding of managed care, as well as of how to use mental health codes and health and behavior (H & B) codes. Finally, it is important to attend to more specific insurance-related issues that have an impact on pediatric psychology, including integrated care, the use of bundling, and employee assistance programs.

A Brief History of Health Insurance

Payment for health care by government and private insurers has a history that dates back to the 19th century in both the United States and Europe, with the earliest plans emphasizing secondary costs (e.g., loss of patient income, social costs, indirect costs to society) rather than those of direct care. Whereas Europe focused on national systems for compulsory sickness insurance, the United States relied on other means, due to a decentralized federal government and a vast rural population (Starr, 1982). Unions, lodges or societies based on national origin, and other benevolent societies filled the local needs. During this period, Americans also bought accident and life insurance to cover indirect costs for injured workers. Today in the United States, the government and employers are the largest insurers for health—providing coverage to specific populations, but still leaving nearly 46 million people without coverage (Assistant Secretary for Planning and Evaluation, 2005). Employers in this country have a history of furnishing