Published in final edited form as:

J Consult Clin Psychol. 2006 June ; 74(3): 468–481. doi:10.1037/0022-006X.74.3.468.

Examining Clinical Judgment in an Adaptive Intervention Design: The Fast Track Program

Karen L. Bierman,

Department of Psychology, Pennsylvania State University

Robert L. Nix,

Prevention Research Center, Pennsylvania State University

Jerry J. Maples,

Department of Statistics and Methodology Center, Pennsylvania State University

Susan A. Murphy, and

Department of Statistics, University of Michigan

The Conduct Problems Prevention Research Group

Abstract

Although clinical judgment is often used in assessment and treatment planning, rarely has research examined its reliability, validity, or impact in practice settings. This study tailored the frequency of home visits in a prevention program for aggressive–disruptive children ($n = 410$; 56% minority) on the basis of 2 kinds of clinical judgment: ratings of parental functioning using a standardized multi-item scale and global assessments of family need for services. Stronger reliability and better concurrent and predictive validity emerged for the 1st kind of clinical judgment than for the 2nd. Exploratory analyses suggested that using ratings of parental functioning to tailor treatment recommendations improved the impact of the intervention by the end of 3rd grade but using more global assessments of family need did not.

Keywords

clinical judgment; prevention; behavior problems

Part science and part art, clinical judgment has been central to the practice of psychology since its inception. Clinical judgment can be descriptive or predictive and is often used in assessment and treatment planning.

Correspondence concerning this article should be addressed to Karen L. Bierman, Department of Psychology, 106 East Henderson, Pennsylvania State University, University Park, PA 16802. kb2@psu.edu. For additional information concerning Fast Track, see <http://www.fasttrackproject.org>.

Karen L. Bierman's fellow members of the Conduct Problems Prevention Research Group are, in alphabetical order, John D. Coie, Department of Psychology, Duke University; Kenneth A. Dodge, Department of Public Policy Studies, Duke University; B. Michael Foster, School of Public Health, University of North Carolina; Mark Greenberg, Department of Human Development and Family Services and Prevention Research Center, Pennsylvania State University; John E. Lochman, Department of Psychology, University of Alabama; Robert J. McMahon, Department of Psychology, University of Washington; and Ellen E. Pinderhughes, Eliot Pearson Department of Child Development, Tufts University.

Jerry J. Maples is now at the Statistical Research Division, United States Census Bureau.

The Conduct Problems Prevention Research Group is the developer of the Fast Track Curriculum and has a publishing agreement with Oxford University Press.

Despite the common use of clinical judgment, empirical studies have identified weaknesses in its application. Studies have demonstrated that clinical judgment can be inferior to actuarial procedures in diagnostic assessment (Dawes, Faust, & Meehl, 1989) and subject to numerous biases (see Arkes, 1981; Garb, 1998; and Goldberg, 1968, for reviews). Most of the studies examining clinical judgment involve simulations in which clinicians use vignettes to form opinions about unknown persons in decontextualized settings. Although useful for understanding cognitive processes, those simulations cannot capture all the complexities of real world decision-making (Rock, Bransford, Maisto, & Morey, 1987; Rock, Bransford, Morey, & Maisto, 1988). It is unclear whether such simulations provide good estimates of the potential value of clinical judgment in the context of significant, long-term therapeutic relationships (Beutler & Clarkin, 1990).

Studies of clinical judgment in practice settings suggest that its reliability and validity may depend on the type of judgment gathered and the way that information is used in assessment and treatment planning. For example, one type of clinical judgment involves global, categorical decision-making about diagnosis or treatment needs. Two recent studies found low levels of agreement when diagnoses assigned to adolescents by clinicians providing mental health services were compared with those derived from structured diagnostic interviews (Jensen & Weisz, 2002; Lewczyk, Garland, Hurlburt, Gearity, & Hough, 2003). Similarly, low levels of agreement characterized the decisions clinicians made about the intensity of treatment services needed (e.g., inpatient vs. outpatient) by youth with different symptom profiles (Bickman, Karver, & Schut, 1997). Particularly disturbing is evidence that factors like client race sometimes bias global clinical judgments regarding diagnoses and treatment, such as use of psychotropic medication (Pavkov, Lewis, & Lyons, 1989; Segal, Bola, & Watson, 1996).

In contrast, a second type of judgment involves asking clinicians to describe specific behaviors (Westen & Weinberger, 2004). Clinicians are considered informants with a unique perspective on client functioning, such as the client her or himself, a parent, a teacher, or a spouse. These descriptive judgments allow the clinician to quantify characteristics (“argues a lot,” “ability to use praise and positive attention appropriately”) that can be combined to represent multifaceted constructs such as psychiatric diagnosis or risk level. Conceptualized in this way, clinical judgment should be more valid because it represents the confluence of multiple judgments—one corresponding to each item on a measure—rather than a single judgment. It also reduces demands for inference and conjecture, as clinicians can rely on psychometrically valid measures with multiple items to structure and constrain their inferences rather than making global, categorical decisions about diagnosis or treatment needs (Westen & Weinberger, 2004). This kind of clinical judgment also makes optimal use of the training and experience clinicians receive, which emphasizes observation skills and judgments concerning the atypicality and clinical significance of various behaviors or characteristics (Westen & Weinberger, 2004). Applying clinical judgment in this way, Dutra, Campbell, and Westen (2004) found that therapists' ratings of youth behavior problems on a standardized scale showed reasonable reliability and validity, revealing meaningful associations with developmental and family history variables.

Adaptive Intervention Designs

One important kind of clinical judgment involves decision making regarding treatment planning: determining the intensity or type of intervention components needed by particular individuals or families to produce positive outcomes. Despite the fact that clinicians routinely construct individualized treatment plans (Beutler & Clarkin, 1990), only recently have researchers begun to evaluate the impact of adaptive intervention designs, in which interventions are varied systematically to address the heterogeneity that exists among individuals in their need for and response to intervention (Lavori & Dawson, 1998).

Prior research has linked adaptive intervention designs with improved intervention efficacy in areas of drug treatment and health promotion (Kreuter, Farrell, Olevitch, & Brennan, 2000; Lavori, Dawson, & Roth, 2000). Results have been mixed, however, when therapists have used clinical judgment to tailor mental health interventions. For example, Schulte, Kunzel, Pepping, and Schulte-Barhenberg (1992) compared the efficacy of three intervention conditions for treating phobias: (a) a standardized multicomponent intervention, (b) an adaptive intervention, in which therapists selected among components to create a tailored intervention, based on their assessment of each client's individual needs, and (c) an adaptive intervention, based on an assessment of a yoked partner's needs. Although clients in all three conditions showed improvement, clients who received the standardized intervention did significantly better than clients in the other two conditions. It is interesting that there was no difference in improvement for the clients who received an adapted intervention based on an assessment of their own or somebody else's needs. In another study, therapists provided maritally distressed couples with a standardized intervention or an adapted intervention that varied in the amount of time that was spent on each of six different components of treatment (Jacobson et al., 1989). In this case, there were no differences in outcomes across conditions at the end of treatment. When couples received the adaptive intervention, however, they were significantly more likely to maintain their gains 6 months later. Clearly, more research is needed to better understand the conditions under which clinical judgment provides a reliable and useful tool for making decisions about the type or intensity of intervention needed by various individuals or families.

The Fast Track Preventive Intervention Program

The present study provides an ecologically valid examination of clinical judgment applied in the context of Fast Track, a multisite, multicohort, multicomponent program designed to prevent conduct problems among high-risk children (Conduct Problems Prevention Research Group, 1992). As one of its components, Fast Track included home visits to promote parental functioning. To increase efficiency and impact, the Fast Track program used an adaptive design in which home visiting services were provided to families at different levels of intensity depending on clinical judgments of parental functioning and family need. Over a 2-year period, two kinds of clinical judgment were used: staff members completed a standardized rating scale assessing parental functioning (hereafter labeled *ratings of parental functioning*), and staff members made more global assessments of each family's need for services (hereafter labeled *global assessment of need*). Because these clinical judgments were made in the context of a longitudinal research study, it was possible to assess their concurrent and predictive validity by comparing them with a battery of multi-informant, multimethod, research-based measures. It also was possible to examine the degree to which each type of clinical judgment appeared vulnerable to nonsystematic or undesirable biases (e.g., biases associated with factors such as site or race). Finally, it was possible to use a recently developed statistical modeling technique (Murphy, van der Laan, Robins, & Conduct Problems Prevention Research Group, 2001) to estimate how the impact of the intervention was affected by relying on each kind of clinical judgment in making recommendations for the dose or frequency of home visits in this adaptive design.

The Fast Track preventive intervention program targeted multiple risk factors that have been linked to the development of serious antisocial behaviors, including child characteristics, parenting difficulties, community factors, and academic and social maladjustment (Kazdin, 1987; Patterson, Capaldi, & Bank, 1991). Children at-risk for serious conduct problems were identified at school entry and provided with multicomponent prevention services. Although we focus exclusively on home visits in this article, additional components of the program included a school-based curriculum delivered by teachers to promote social-emotional skills, group interventions (child social skill training and parent behavior management training), and individualized academic tutoring for children (see Bierman, Greenberg, & Conduct Problems

Prevention Research Group, 1996, and McMahon, Slough, & Conduct Problems Prevention Research Group, 1996, for a complete description of the intervention).

Rationale for Individualizing the Prescribed Dose of Fast Track Home Visiting

Designed to reinforce the behavior management skills taught in parent groups, home visits helped primary caregivers practice new skills and apply general principles to their individual family situations. Home visits also served to enhance parents' competence in solving more general life-management problems and providing an organized and supportive family environment for their children (Conduct Problems Prevention Research Group, 1992). During the 1st year of Fast Track, when children were in first grade, all families shared the standard recommendation of biweekly home visits. When children were in second and third grade, home visits were offered on a weekly, biweekly, or monthly basis, depending on clinical assessments of parental functioning and family need. By periodically adjusting the frequency of home visits, our goal was to recognize the heterogeneity that existed in the sample in terms of parents' skills, life situations, and intervention responsiveness, and to provide the optimal dose of intervention for each family across time (Lavori & Dawson, 1998). We believed that parents with greater skill deficits would require more frequent home visits to enhance improvements in parenting and promote positive child outcomes, whereas parents with higher skill levels and fewer problems would require less frequent home visits to produce the same outcomes. For these latter families, excessive home visits might elicit family resistance and increase attrition (McMahon et al., 1996).

Using Clinical Judgment to Tailor Home Visiting in the Fast Track Adaptive Prevention Design

Although adaptive interventions are appealing intuitively, researchers are just beginning to recognize the importance of evaluating the assessment and treatment planning processes used in these designs. These processes—in addition to the power of the intervention itself—affect the impact of the intervention on participant outcomes (Collins, Murphy, & Bierman, 2004). In some interventions, such as medical trials, the assessment of need can be based on single indicators that are measured with relative precision, such as blood pressure (Cooperative Research Group, 1988). In these interventions, recommended dose levels are informed by previously documented relations among participant characteristics, treatment needs, and outcomes (Collins et al., 2004). The challenge for individualizing dose in clinical intervention and prevention programs like Fast Track is that, theoretically, a family's need for services is affected by multiple aspects of parental functioning and family circumstances that may not be captured well in any single index. This is why clinical judgment is typically used as the basis for treatment planning in mental health practice, and why it was used to tailor the recommended dose for home visiting in Fast Track.

Although Fast Track collected an array of standardized measures of parent and child functioning during annual research assessments, these measures were not used to tailor the dose of home visiting for several reasons. First, it was not possible to conduct such comprehensive assessments more than once a year, and more frequent assessments were desired for treatment planning. Second, we questioned whether many of our measures would retain their validity if repeated more frequently. Finally, from a theoretical perspective, we believed that the clinicians, called family coordinators in Fast Track, who ran the parent groups and conducted home visits were in an excellent position to make sensitive and accurate judgments about parental functioning and family need. They had professional training—some with advanced degrees in social work or counseling psychology—and extensive experience in human service work. They reflected the ethnic diversity of the communities they worked in and, most important, knew their families best.

Family coordinators used clinical judgment in rating the dimensions of parental functioning that have been linked most closely to child conduct problems in the empirical literature. These included parenting skills, such as high levels of warmth and low levels of harsh physical punishment (Pettit, Harrist, Bates, & Dodge, 1991); parental well-being, such as depression and stressful life events (Luoma et al., 2001; Stanger, McConaughy, & Achenbach, 1992); and parent-school involvement (Reynolds, Weissberg, & Kaspro, 1992). Project guidelines specified that families should receive weekly, biweekly, or monthly home visits, depending on their scores from the ratings of parental functioning in those domains.

In addition to using clinical judgment to make those standardized ratings, Fast Track family coordinators also relied on their clinical judgment to make global assessments of family need for home visits and to adjust intervention dose recommendations accordingly. Prior research has shown that certain risk and protective factors can affect child conduct problems, sometimes without directly impairing parental functioning. For example, socioeconomic disadvantage, single parent families, the quality of the home environment, the degree of danger in the neighborhood, and child academic achievement each have been linked with child conduct problems (e.g., Costello, Compton, Keeler, & Angold, 2003; Leventhal & Brooks-Gunn, 2000; Lynam, Moffitt, & Stouthamer-Loeber, 1993). Hence, taking into account factors such as these that might impinge on parental functioning or children's success at school, family coordinators could use their clinical judgment about the impact of these extenuating circumstances to recommend more or less frequent home visits than the level prescribed on the basis of ratings alone.

The Present Study

Given its design, Fast Track provides one of the first opportunities to examine the use of clinical judgment in ongoing therapeutic relationships. Moreover, Fast Track enabled a comparison of two kinds of clinical judgment: ratings of parental functioning and global assessments of need.

The first goal of this study was to assess the reliability and validity of these two kinds of clinical judgment, by comparing them with research-based measures collected as part of the longitudinal study. On the basis of prior research, we hypothesized that the ratings of parental functioning would be reliable and valid (Dutra et al., 2004). We also hoped that global assessments of need for services would be reliable and valid, given that they were made by professional, supervised staff members in the context of long-term therapeutic relationships. At the same time, we recognized that clinical ratings and, especially, global assessments might be vulnerable to undesirable biases, such as drift across time and cohort, the establishment of idiosyncratic site-based norms, and differential perceptions of parental functioning by factors such as race. We wanted to evaluate this risk in the context of this well-specified intervention program. The second goal was to compare the ratings of parental functioning with global assessments of need. We wanted to understand why family coordinators made global assessments that indicated higher or lower levels of need for home visiting than their ratings of parental functioning would have indicated. We hypothesized that family risk and protective factors would explain that divergence, but also wanted to examine whether undesirable biases contributed to unreliable variations in global assessments of need. Finally, the third goal of this study was to evaluate how these two types of clinical judgment might have affected the impact of Fast Track on one of its primary outcomes, children's school behavior problems.

Method

Behavior Screening and Sample Selection

Participants were drawn from 54 elementary schools serving low-income neighborhoods in Durham, NC; Nashville, TN; Seattle, WA; and rural central Pennsylvania. Schools were

matched into sets at each site and were randomly assigned to intervention or control conditions. Using a multiple-gating screening procedure that combined teacher and parent ratings of disruptive behavior, we initially screened all 9,594 kindergarteners across three cohorts (1991–93) for classroom conduct problems by teachers, using the Teacher Observation of Child Adjustment-Revised (TOCA-R) Authority Acceptance Score. The parents of those children scoring in the top 40% within cohort and site then described the behavior problems of their children at home, using items from the Child Behavior Checklist and similar scales (91% agreed; $N = 3,274$). Children were selected for inclusion into the study on the basis of the sum of the standardized teacher and parent screening scores, moving from the highest score downward until desired sample sizes were reached within sites, cohorts, and conditions. Eight hundred ninety-one children ($Ns = 445$ for intervention and 446 for control) participated; 95% of the selected sample scored in the top 20% on both the parent and teacher screening measures, 76% of these children had teacher ratings on the Externalizing scale of the Child Behavior Checklist that were in the clinical range (t score of 60 or higher; mean t score = 66; see Lochman & Conduct Problems Prevention Research Group, 1995, for additional details on sample selection).

Participants

This study relied on data from 410 families of children in Fast Track's intervention group.¹ For the final outcome analyses only, the study also included data from 403 high-risk control group families. All of the children in the intervention group were about 6 years old when they began participating in Fast Track ($M = 6.50$ years, $SD = 0.49$), and they were 3 years older at the end of the present study. The intervention sample was 53% African American, 44% European American, and 3% Asian American, Latino, or American Indian. Seventy-one percent of the children were boys, and 29% were girls. Fifty percent of the intervention families included single parents, and 30% of the mothers had not completed high school. Overall, about 63% of these families were in the lowest two categories of socioeconomic status (SES; Hollingshead, 1975). By the end of this study, approximately 4% of those intervention families had moved out of town and 5% believed they no longer needed services.

Ratings of Parental Functioning and Global Assessments of Family Need

In August and January, Fast Track family coordinators completed assessments of all the families with whom they worked. The current measures were revised after the 140 children in Cohort 1 started second grade, but applied for the 270 families of children in Cohorts 2 and 3 in second grade and all families thereafter. Over the time period included in this study, 32 family coordinators (all women) provided ratings. On average, each family coordinator rated 8.60 ($SD = 4.03$) families with children in second grade and 11.80 ($SD = 4.16$) families with children in third grade.

Ratings of parental functioning measured three domains of parenting and well-being: (a) parenting skills were represented by three items reflecting quality of the parent–child relationship, positive parenting practices, and effective discipline; (b) parental well-being was represented by two items reflecting the parent's mental health (e.g., depression, substance use) and exposure to stressful life events; and (c) parent–teacher involvement was represented by one item reflecting positive engagement with the child's school and support for the child's education. Items were rated on a 4-point scale (1 = *severe problems* to 4 = *minor or no problems*); response categories were anchored by behavioral criteria. The six items were summed to create a total score.

¹The families of 35 intervention children who repeated the first grade were not included in this study because they entered the adaptive treatment phase of Fast Track on a different schedule.

Fast Track established project guidelines linking these ratings of parental functioning to three prescribed levels of home visiting: weekly home visits for families with scores lower than 9, biweekly home visits for families with scores between 9 and 16, and monthly home visits for families with scores greater than 16. On the basis of their global assessment of need for home visits, family coordinators could follow those guidelines or override them. In consultation with their supervisors, family coordinators could recommend more or less frequent home visits when warranted by special family circumstances. Combined training of family coordinators from all Fast Track sites, weekly individual and group supervision, team meetings, and regular cross-site conference calls were used to monitor consistency and preserve program fidelity.

Research-Based Measures of the Family and Child

During the spring of first grade and the spring of second grade, all teachers completed rating scales of children's behavior problems and parent–school involvement. In the summer before second grade and the summer before third grade, teams of two trained research assistants, naive concerning the intervention status of families, conducted standardized home interviews and observations of parent–child interaction. While one research assistant interviewed the parent, the other research assistant interviewed the child in a separate room.

Two sets of measures drawn from these annual assessments were used in this study: measures focused on the same dimensions of parental functioning as family coordinators' ratings, and measures of additional risk and protective factors that family coordinators might consider in making their clinical judgments. Basic information about each measure is presented below. Descriptive statistics are presented in Table 1; more extensive psychometric information, including lists of items in most scales, is available on our website: www.fasttrackproject.org,

Research-based measures of parental functioning

Research-based measures of parental functioning assessed parenting skills, parental well being, and parent–teacher involvement. Parenting skills were indexed by parental warmth displayed during a series of structured parent–child interaction tasks. Using the Interaction Rating Scale (Crnic & Greenberg, 1990), research assistants made six summary ratings, each on a 5-point scale ($\alpha = .90^2$). Interrater reliability on these six ratings (assessed with intraclass correlation coefficients) was .76 for families with children in second grade and .74 for families with children in third grade. Parenting skills also included a measure of physical punishment. This was operationalized as the number of times parents spontaneously mentioned using slapping, swatting, or spanking to address six common behavior problems, such as noncompliance ($\alpha = .46$).

Self-reports were used to assess parental well-being. On the Center for Epidemiological Studies Depression Scale (Radloff, 1977), parents used four response options to describe how often they had experienced 20 different symptoms of depression during the prior week ($\alpha = .88$). Stressful life events were assessed using a 16-item scale developed for Fast Track on which families reported the number and severity of stressful situations, such as moves or major medical problems, they had experienced during the preceding year ($\alpha = .61$).

Parent–teacher involvement was rated by teachers using a 21-item questionnaire developed for Fast Track. Items were based on a 5-point scale and reflected how much support the parent provided to promote children's success in school ($\alpha = .90$).

²Cronbach's coefficient alphas are reported for the 1st year of this study, when children were between the first and second grades, and are based on the 270 intervention families in Cohorts 2 and 3.

Risk and protective factors

Additional risk and protective factors affecting parental functioning and child outcomes included family SES, single-parent status, quality of the home environment, neighborhood danger, and child academic achievement. Family SES represented a weighted average of parents' education and occupational prestige, using the Hollingshead (1975) classification system. Single parent status was a dichotomous variable defined as not being married and not having a partner who had lived with the family for more than 1 year. Using the Post-Visit Inventory (Dodge, Bates, & Pettit, 1990), observers rated three features of the home environment—size, safety, and cleanliness—using a 4-point scale ($\alpha = .55$). Also using that inventory, observers rated four features of neighborhood danger using a 4-point scale ($\alpha = .73$). Children's academic achievement was assessed with standardized summary scores from the Diagnostic Reading Scales (Spache, 1981).

Child outcome

Child school behavior problems were assessed for the intervention and control group children with the authority acceptance subscale of the Teacher Observation of Classroom Adaptation – Revised (Werthamer-Larsson, Kellam, & Wheeler, 1991). When children were in kindergarten, prior to the beginning of any Fast Track services, and at the end of the third grade, teachers rated the frequency of 10 oppositional and aggressive behavior problems, such as stubborn, breaks rules, and fights, on a 6-point scale ($\alpha = .91$).

Results

Goal 1: Examining the Reliability and Validity of Clinical Ratings and Global Judgments

The first goal of this study was to examine the reliability and validity of the two kinds of clinical judgment used in Fast Track: ratings of functioning based on a standardized, 6-item scale and global assessments of need, reflected in recommendations for weekly, biweekly, or monthly home visits. We sought to examine whether two family coordinators made similar judgments; whether those judgments corresponded with other research-based measures of parental functioning, family need, and child outcome; and whether those judgments were prone to undesirable biases related to cohort, site, individual variation in the rating styles of family coordinators, child sex, and race.

Reliability of clinical judgment—The reliability of family coordinators' clinical judgment was assessed in three ways. First, Cronbach's coefficient alpha was computed for the six items of parental functioning. This measure of internal consistency ranged from .85 to .88 in each of the four time periods (beginning of second grade, middle of second grade, beginning of third grade, and middle of third grade) included in this study, with an average of .87.

Second, the stability of family coordinators' ratings and recommendations for home visits from one period to the next was examined. For the four periods included in this study, the correlation between the total score on the ratings of parental functioning in one period and the total score on the ratings of parental functioning in the next period ranged from .69 ($p < .001$) to .74 ($p < .001$), with an average of .71. The correlation for the global assessments of need—operationalized as the number of home visits actually recommended per month (1 = monthly, 2 = biweekly, and 4 = weekly)—across two successive periods ranged from .47 ($p < .001$) to .61 ($p < .001$), with an average of .53.

Third, the tendency of two family coordinators to view the same family similarly was assessed. Because each family was rated only by the family coordinator who was leading its parent group and conducting its home visits, it was not possible to assess interrater reliability in the traditional way. Family moves and turnover among family coordinators, however, led to

reassignments of family coordinators for about 10% of our intervention families. In these cases, we had two family coordinators assessing the same family during successive rating periods. When the total score on the ratings of parental functioning made by one family coordinator in one period was compared with the total score on the ratings of parental functioning made by a different family coordinator in the next period, the intraclass correlation coefficient was .61 ($n = 42$, $p < .001$). When the number of home visits recommended by one family coordinator for one period was compared with the number of home visits recommended by a different family coordinator in the next period, the intraclass correlation coefficient was .49 ($n = 36$, $p < .001$). Despite the fact that our scores of interrater reliability of clinical judgment were confounded by time, they were still comparable with scores from a simulation study (Allen, Coyne, & Logue, 1990) and would be considered good and fair, respectively, by common benchmarks (Cicchetti, 1994). Given that the stability of ratings and recommendations made by the same family coordinator over the same period were .72 and .53, respectively, the upper bound of interrater reliability was considerably lower than the 1.00 that would be traditional in such analyses. If our stability scores were used as the upper bound of our interrater reliability scores, then our estimates of interrater reliability would be much higher.

In general, it appears that family coordinators' clinical judgment was reasonably reliable. Consistent with recent research in clinical judgment (e.g., Dutra et al., 2004; Westen & Weinberger, 2004), family coordinators appeared to be more reliable in their ratings of parental functioning than in their global assessments of need for home visits.

Concurrent validity—To examine concurrent validity, we examined relations between family coordinators' clinical judgments and research-based measures of parental functioning and risk and protective factors. We focused on two rating periods, the fall of second grade and the fall of third grade, because family coordinators' clinical judgments for these periods were made in August, and the research-based measures were collected between June and August of the same year. In these analyses, the fact that each family coordinator worked with several families was considered a fixed effect and controlled for with a set of dummy variables. Preliminary tests to determine whether these relations were moderated by race—and should be considered separately by race—revealed no more significant or marginally significant differences than we would have expected by chance.

As shown in Table 2, family coordinators' ratings of functioning were significantly related to three of the research-based measures of parental functioning in second grade and all five of these measures in third grade. These ratings also were related to all five of the risk and protective factors in both grades, providing evidence of the concurrent validity of the clinical ratings. All correlations were of modest size, however, probably attenuated by the restricted range of functioning in the high-risk sample.

As expected, the two types of clinical judgments (family coordinator's ratings of parental functioning and their global assessments of family need for home visiting) were significantly correlated, with standardized coefficients of $-.44$ ($p < .001$) for families with children in second grade and $-.60$ ($p < .001$) for families with children in third grade. However, global assessments of need showed fewer significant correlations with the research-based measures than did the ratings, correlating significantly with 1–3 measures of parental functioning (in the second and third grades, respectively) and three of the five risk and protective factors in each grade.

In several cases, the magnitude of correlations with research-based measures also favored the ratings of parental functioning over the global assessments of needs. For example, correlations with observed parent–child warmth were significantly larger for ratings of parental functioning than for global assessment of need for home visits in both the second and third grades, $t(241) = 3.23$, $p < .001$, and $t(329) = 2.34$, $p < .05$, respectively. The same was true for relations to

maternal depression in both grades, $t(242) = 2.59, p < .01$, and $t(338) = 2.77, p < .01$; parent–school involvement in both grades, $t(240) = 2.38, p < .05$, and $t(339) = 2.97, p < .01$; family SES in second grade, $t(243) = 2.99, p < .01$; single-parent household in second grade, $t(243) = 2.83, p < .01$; quality of home environment in second grade, $t(231) = 2.53, p < .05$; and neighborhood danger in third grade, $t(321) = 2.26, p < .01$.

Desirable versus undesirable sources of influence—Prior research suggests that clinical judgment sometimes is vulnerable to nonsystematic or undesirable biases, such as those related to children's race (Payette & Clarizio, 1994; Segal et al., 1996). To determine whether such influences existed in Fast Track, hierarchical regressions were conducted to partition the variance in ratings of parental functioning and global assessments of family need. In these regressions, the first block of independent variables included the research-based measures of parental functioning: parental warmth, physical punishment, parental depression, stressful life events, and parent–teacher involvement. The second block of independent variables included the measures of risk and protective factors: family SES, single-parent status, home environment, neighborhood danger, and child academic achievement. Associations between those two blocks of measures and family coordinators' ratings of parental functioning and global assessment of need were considered desirable, replicable, and further evidence of the concurrent validity of clinical judgment. The third block of independent variables represented family participation in both parent groups and home visits, which—although not a measure of parental functioning, per se—could reasonably and reliably affect family coordinators' judgments regarding the need for additional parent skill training and home visits. The fourth, fifth, and sixth blocks of independent variables represented potential sources of undesirable influences: effects of cohort and site, effects of individual and idiosyncratic family coordinator rating styles, and effects of child sex and race.

As shown in Table 3, variables representing desirable and replicable sources of influence (research-based measures of parental functioning and family risk and protective factors) produced significant increments in the ability to explain family coordinators' ratings of parental functioning. In second grade, significant or marginally significant unique contributions were made by observer ratings of parent–child warmth and physical punishment, maternal depression, stressful life events, teacher-rated parent–school involvement, family SES, the quality of the home environment, and child academic achievement. Results were similar in Grade 3, with the exception of physical punishment and quality of the home environment. In addition, participation in parent groups made a significant unique contribution in predicting the ratings of parental functioning in third grade. Altogether, the replicable and desirable influences accounted for a total of 27% and 29% of the variance in ratings of parental functioning in the second and third grades, respectively.

In contrast, desirable sources of influence made smaller contributions to the prediction of clinician's global assessments of need for home visiting. For families with children in second grade, only the block of variables representing risk and protective factors produced a significant increment in the prediction of global assessments of need. For families with children in third grade, research-based measures of both parental functioning and risk and protective factors predicted global assessments of need. Significant or marginally significant unique contributions were made by the quality of the home environment and child academic achievement in both grades and, in addition, by parent–school involvement and neighborhood danger in third grade. Overall, replicable and desirable sources of influence accounted for only 8% and 11% of the variance in global assessments of need in the second and third grades, respectively.

Among the blocks of variables representing potential sources of bias, significant effects emerged for cohort and site in explaining both ratings of parental functioning and global

assessments of need in both years. These effects were relatively small, however, except when predicting global assessments of need in third grade. The effect of the family coordinators themselves was much larger, except in one case when the site effects were especially strong. Child sex and race did not account for significant variance in either kind of clinical judgment.

When the relative effects of replicable and desirable influences and potential sources of bias were considered together, replicable and desirable influences accounted for a much greater proportion of the explained variance in ratings of parental functioning (68% of the explained variance in second grade [.27/.40], and 64% of the explained variance in third grade), whereas potential sources of bias accounted for a much greater proportion of the variance in global assessments of need (81% of the explained variance in second grade, and 61% of the explained variance in third grade.)

Predictive validity—It could be that, in making their global assessments of need for home visiting, family coordinators tend to minimize current fluctuations in adaptation and focus on more stable strengths and weaknesses that have long-term implications for child well-being. If this is the case, global assessments of need might have greater predictive validity than ratings of parental functioning.

To examine that possibility, family coordinators' ratings of parental functioning and their global assessments of need for home visiting, made at the beginning of second and third grade, were related to children's behavior problems at school, measured at the end of third grade. Ratings of parental functioning in second and third grades were significantly related to children's future school behavior problems with standardized coefficients of $-.17$ ($p < .01$) and $-.21$ ($p < .001$), respectively. On the other hand, the global assessments of family need in second and third grade were hardly related to children's school behavior problems with standardized coefficients of $.06$ (ns) and $.09$ ($p < .10$), respectively. Although the magnitude of this difference is not statistically significant in families with children in second grade, it is in families with children in third grade, $t(332) = 2.10$, $p < .05$.

When hierarchical regressions were conducted that included all of the research (measures of parental functioning, family risk and protective factors, and demographic variables), family coordinators' ratings of family functioning produced a much larger increment in the ability to predict school behavior problems with a standardized coefficient of $-.13$ ($p = .11$) for children in second grade and $-.18$ ($p < .01$) for children in third grade. Family coordinators' global assessments of need, however, did not explain any additional variance in school behavior problems above and beyond all of those other variables and the ratings of functioning with standardized coefficients of $.00$ (ns) for families with children in second grade, and $.09$ (ns) for families with children in third grade. Overall, these findings suggest that clinical judgments based on rating scales describing functioning are more reliable, less vulnerable to undesirable bias, and have more concurrent and predictive validity than more global assessments of need.

Goal 2: Compare Clinical Ratings With Global Judgments

The next set of data analyses explored the deviations in clinical judgment between family coordinators' ratings of parental functioning and their global assessments of family need for home visiting. We wanted to better understand the factors that contributed to family coordinators' decisions to override project guidelines and recommend higher or lower levels of home visiting than those prescribed by their ratings of parental functioning.

On the basis of the total number of family coordinators' ratings of parental functioning in four time periods across 2 years, project guidelines specified that 62%, 34%, and 4% of the recommendations should have been for monthly, biweekly, and weekly home visits, respectively. The actual recommendations, incorporating clinical judgments to increase or

decrease the prescribed dose resulted in a similar overall distribution with 61%, 32%, and 7% of the recommendations being for monthly, biweekly, and weekly home visits, respectively. Overall, family coordinators followed project guidelines 82% of the time. They recommended more frequent home visits 14% of the time in which ratings of parental functioning were high and if families would have received monthly visits otherwise. They recommended less frequent home visits 26% of the time in which ratings of parental functioning were low and if families would have received weekly home visits otherwise. They also changed the recommended number of home visits 26% of the time in which ratings of parental functioning were moderate, sometimes increasing but usually decreasing the dose of home visits specified by project guidelines.³

Almost all family coordinators made some changes in home visit dose recommendations over the four time periods covered by this study; only two family coordinators always followed project guidelines. Nine family coordinators (28%) only increased dose recommendations, 16 family coordinators (50%) increased some dose recommendations and decreased others, and 6 family coordinators (19%) only decreased dose recommendations.

To identify factors related to family coordinators' decisions to increase or decrease dose recommendations, we conducted hierarchical logistic regressions. Similar to our previous analyses, we accounted for systematic and desirable influences on family coordinators' recommendations (e.g., research-based measures of parental functioning, risk and protective factors, and intervention participation) before we examined undesirable sources of influence (e.g., cohort and site; child sex and race). In addition, a variable indicating the level of home visiting prescribed by clinical ratings of parental functioning (e.g., weekly, bimonthly, or monthly) was entered before all sets of predictor variables. Because so few family coordinators chose to override project guidelines in a specific direction in a specific time period, the effects of family coordinators could not be estimated. The outcome variables for these analyses were coded such that scores of 1 indicated that family coordinators believed families needed more or less frequent home visits than their ratings of parental functioning would suggest and scores of 0 indicated agreement between the two kinds of clinical judgment.

Predicting staff judgments to increase home visiting recommendations—The first set of logistic regressions predicted family coordinators' decisions to recommend more frequent home visits than prescribed by their ratings of parental functioning. These analyses did not include families with low ratings of parental functioning because family coordinators could not recommend that home visits occur more frequently than once a week.

As shown in Table 4, in second grade, the decision to increase the frequency of home visits was predicted by the block of independent variables representing risk and protective factors only. Significant and unique coefficients within that block suggest that family coordinators were especially likely to recommend extra home visits if families lived in dangerous neighborhoods and if children showed poor academic achievement. In third grade, these decisions were predicted by the blocks of variables representing project guidelines, risk and protective factors, and cohort and site. Significant and unique coefficients with these blocks suggest that family coordinators were more likely to recommend extra home visits if their ratings of parental functioning specified monthly rather than biweekly contact, if parents were single, and if children showed poor academic achievement. Family coordinators at one urban site with especially distressed families were especially likely to recommend extra home visits.

³Interestingly, when family coordinators chose to modify dose recommendations, it was not because families just barely missed the threshold for the next level of frequency for home visits. For example, the families who were recommended for weekly rather than biweekly visits, or vice versa, were not likely to have total scores on the ratings of parental functioning near 8 or 9, and the families who were recommended for biweekly rather than monthly visits, or vice versa, were not likely to have scores near 16 or 17. In fact, families appeared less likely to have scores within two points of the thresholds than they were to have scores representing larger discrepancies.

As anticipated, these factors increased family coordinators' perceptions of child risk and a family's need for intervention in ways that were not fully captured by their ratings of parental functioning.

Predicting staff judgments to reduce home visiting recommendations—A similar set of hierarchical regression analyses examined factors explaining decisions to recommend less frequent home visits than specified by project guidelines. These analyses did not include families with high ratings of parental functioning, because family coordinators could not recommend that home visits occur less frequently than once a month.

Results of these analyses also are presented in Table 4. None of the blocks of variables representing systematic and desirable sources of influence were significantly related to these decisions. Site was the major factor predicting these decisions. In second grade, family coordinators from two sites were especially likely to recommend less frequent home visits than specified by project guidelines. In third grade, family coordinators from one of those sites continued to do so, whereas family coordinators from two other sites appeared especially unlikely to make such recommendations. In third grade, family coordinators were more likely to recommend fewer home visits than project guidelines specified if families were African American families, but this appeared to be due to the dynamics at one site only.

Goal 3: Evaluating the Effects of Clinical Judgment on Child Outcomes

In the final stage of data analyses, weighted regressions were used to address the question, “Would the effect of Fast Track on third-grade school behavior problems have changed if clinical judgment were used differently to tailor the intervention?” These analyses were based on recent advances in statistical theory (Murphy et al., 2001) and were designed to compare outcomes in the current trial with estimated outcomes in two simulated conditions: one in which no clinical judgment was used to determine the frequency of home visits, and one in which ratings of parental functioning, but not global assessments of need, were used to determine the frequency of home visits. To simulate these different implementation conditions, cases were differentially weighted in outcome analyses.

It should be stressed up front that these analyses only provide a basis for estimating how different levels of clinical judgment might have affected a primary outcome of the intervention; they were not designed to assess the actual effect of Fast Track at the end of third grade. Fast Track involved multiple intervention components in addition to home visits and examined multiple outcomes; a comprehensive report of third-grade outcomes is published elsewhere (Conduct Problems Prevention Research Group, 2002). The purpose of these analyses is to describe how choices made regarding the use of clinical judgment in an adaptive intervention design could affect outcomes, but they should not be considered as strong evidence of how they actually did affect outcomes.

Conceptual model guiding the comparisons—For this study, three alternative implementation conditions were compared. In the first condition, the intervention was modeled as if the research-based measures of parental functioning, rather than family coordinators' ratings, had been used in the determination of need and as if dose recommendations for home visits followed project guidelines with no exceptions. Theoretically, this condition represented what might have happened if Fast Track had not allowed any clinical judgment in assessment or treatment planning. In the second condition, the intervention was modeled as if home-visit dose recommendations were based only on family coordinators' ratings of parental functioning, without allowing any adjustments based on global assessment of family need. In the comparison condition—which was unweighted because it was identical to what actually occurred in Fast Track—dose recommendations for home visits were based on family

coordinators' ratings of parental functioning and adjusted according to global assessments of family need.

Derivation of the case weights—To estimate the size of the intervention effect under those different implementation conditions, child outcome scores were differentially weighted to represent the similarity between the number of home visits that would have been recommended under each alternative implementation condition and the number of home visits that actually were recommended in Fast Track. Numerators for the weights in the first alternative implementation condition were based on the relations between the research-based measures of parental functioning and recommendations for weekly, biweekly, or monthly home visits in each of the four time periods. Once those relations were found for the entire sample, the predicted probability that a specific family with its specific scores on those measures would have received the recommendation that it actually did was calculated. The numerators for the weights in the second alternative implementation condition were based on whether family coordinators used their ratings of parental functioning and followed project guidelines exactly in determining the frequency of home visits (numerator of 1) or whether they decided to use global assessments of need to override project guidelines (numerator of 0). The denominators for the weights in both hypothetical implementation conditions were based on the most parsimonious statistical model of dose recommendations in Fast Track: They represented a family's predicted probability, in each time period, of receiving the dose recommendation it actually did, given its specific ratings of parental functioning, its ratings of parental functioning in the previous time period, its recommendation for home visits in the previous time period, and its cohort and site.

Once the weights for each family were calculated for each alternative implementation condition, they were combined across time periods and used to regress children's school behavior problem scores at the end of third grade on their scores on the same measure in kindergarten before the intervention began and a dichotomous variable representing treatment status. Because the standard errors of regression coefficients based on estimated weights are smaller than they should be, conservative adjustments to these standard errors were computed as described in Murphy et al. (2001).

Assessment of impact on outcomes—Under the first alternative implementation condition, the treatment effect was nonsignificant with an effect size—calculated as the number of standard deviations between the mean of the intervention group and the mean of the control group—of .00. This suggests that children would have received no school benefit from being in Fast Track if we had relied exclusively on the research-based measures of parental functioning and used no clinical judgment at all to determine the frequency of home visits. Under the second alternative implementation condition, the treatment effect was statistically significant ($p < .05$) with an effect size of .23. This suggests that children would have improved at school as a result of being in Fast Track, if we had allowed clinical judgment in the ratings of parental functioning, but not in the global assessments of family need. In the unweighted comparison condition, which reflected what actually happened in Fast Track, the treatment effect was marginally statistically significant ($p < .10$) with an effect size of .14. The effect size from the second alternative implementation condition was more than 50% larger than this effect size. Together, these analyses suggest that it might have been better for our children and families if we had allowed clinical judgment in the ratings of parental functioning but not in the global assessment of family need.

Discussion

The present study represents an important first step in examining the use of clinical judgment in adaptive intervention designs. Despite the wide-spread use of clinical judgment as a basis

for making treatment recommendations in clinical practice, relatively little research is available to identify optimal ways to gather and use these judgments during ongoing interventions. As adaptive designs become more common in research and practice alike, developing techniques that can maximize the reliability and validity of clinical judgment as a tool for assessment and treatment planning becomes increasingly valuable (Collins et al., 2004). In this context, Fast Track offered a unique opportunity to understand and evaluate the use of two kinds of clinical judgment applied to tailor dose recommendations in an adaptive intervention design. Focusing on what happened in real therapeutic relationships over time, this study represents one of the most ecologically valid examinations of clinical judgment to date. It also is one of the first studies of clinical judgment to feature families and children rather than individual adults.

Reliability and Validity of Clinical Ratings and Global Assessments

In adaptive intervention designs like Fast Track, variations in intervention dose are planned to more effectively and efficiently meet the needs of participants. In such programs, the criteria identified and the measurement method used to assess the needs of participants and recommend different dose levels become an integral part of the program itself, affecting its replicability and impact (Collins et al., 2004). The empirically based developmental model that provided the foundation for the design of Fast Track suggested that variations in parental functioning (parenting skills, parental well-being, and parent-school involvement) would indicate differential need for home visits. The model also suggested that additional family risk and protective factors, such as single-parent status and child academic achievement, might affect the optimal dose of home visiting. Clinical judgment was selected as the strategy for assessing parental functioning and determining family need for home visits.

All of the analyses used to assess concurrent and predictive validity suggested that the ratings of parental functioning were more reliable and more valid than the global assessments of family need for home visiting. In the hierarchical regression analyses, systematic and desirable sources of influence accounted for 27% and 29% of the variance in clinical ratings of parental functioning in second and third grades, respectively, but they accounted for only 11% and 18% of the variance in the global assessments of need. Conversely, spurious effects (cohort, site, individual rater biases, child sex, and race) accounted for 13% and 16% of the variance in clinical ratings of parental functioning versus 17% and 35% of the variance in global assessment of need.

To some extent, the greater reliability and validity of the clinical ratings of functioning probably reflects the psychometric benefit of a scale composed of multiple items with more response options and a normal distribution. That is, compared with global assessments of need, which were assessed with one item ("What level of home visiting do you recommend for this family?") using three response options (weekly, biweekly, or monthly home visits), clinical ratings of functioning used 6 items each rated with a 4-point score, providing a much greater range of scores to differentiate families and a more normal distribution.

However, the higher levels of concurrent and predictive validity documented for ratings of parental functioning compared with global assessment of need is not likely due to the restricted range of scores alone. Indeed, there is some evidence that even single-item ratings of psychological characteristics can be reliable and valid (Robins, Hendin, & Trzesniewski, 2001; Startup, Jackson, & Bendix, 2002). In the present study, the two types of clinical assessment also differed substantially in the type of judgment they required. The results support the hypothesis that clinicians are relatively good at providing sensitive descriptions of client functioning when completing standardized rating scales in specific domains linked empirically with the development or remediation of a problem. They appear less good at synthesizing and optimally weighing multiple pieces of information to form one global, categorical assessment of treatment need. In other words, clinicians appear able to make valid and reliable inferences

in delimited domains in which they have received much training; they simply have more difficulty predicting how multiple factors will change and interact, leading to one specific outcome in the future (Dutra et al., 2004; Garb, 2005; Westen & Weinberger, 2004).

It is noteworthy that our family coordinators were not doctoral-level clinicians, and yet their judgments were still reliable and valid when made in the context of our standardized rating scale. Their ratings of family functioning provided incremental validity over research-based measures of the same constructs in predicting children's school behavior. Perhaps, if they had been doctoral-level clinicians using a better-validated measure of family functioning, their judgments would have been even better. It appears likely, however, that even doctoral-level clinicians would have had difficulty making good global assessments of family need for treatment (Garb, 2005; Westen & Weinberger, 2004).

If clinical ratings of parental functioning are to be used in treatment planning, they must be linked to a standard set of guidelines. In Fast Track, those guidelines were based on the best clinical judgment of the program developers. Different project guidelines might have produced different intervention effects. Ideally, the research base will grow so that empirical findings documenting optimal criteria and cutoffs can inform future guidelines (Collins et al., 2004).

It is interesting to note that the clinical judgments that involved increasing the prescribed level of home visits (beyond the level recommended by project guidelines) appeared more systematic and valid than the judgments in which the clinician decided to decrease home visit recommendations, despite evidence of family dysfunction. Regression analyses suggested that, as planned, family risk and protective factors predicted family coordinators' decisions to increase the frequency of home visits.

In contrast, decisions to reduce recommendations for home visits, providing fewer than recommended on the basis of ratings of parental functioning, were not significantly predicted by any of the replicable and desirable sources of influence, but were only affected by undesirable sources of potential bias. Although we hope not, family coordinators might have recommended fewer home visits for some difficult families to reduce their own workloads or to appear as if they were completing a higher percentage of their target number. Because their ratings of parental functioning indicated the presence of many problems detrimental to children's success, family coordinators understood that these families needed help. In staff meetings at the site where recommendations for home visits were most likely to be reduced, the family coordinators described these families as highly disorganized and somewhat resistant to treatment. The family coordinators believed that more frequent home visits were both unrealistic and counterproductive. This same dynamic appeal's to account for the finding that emerged in third grade, when African American families were more likely to receive recommendations for fewer home visits than project guidelines specified. This effect was driven by two African American family coordinators who worked almost exclusively with African American families at that site. They recommended fewer home visits than project guidelines specified for every one of their families. (The effect disappeared when these family coordinators and their families were deleted from the sample.) Anecdotal discussions revealed that these family coordinators were very concerned about their families and very protective of them. They recommended fewer home visits as a means of ensuring that the home visits the families did receive would be nonintrusive, welcome, positive, and helpful.

These findings are consistent with previous research describing therapists' work with challenging clients. It appears that therapists sometimes reduce efforts to alter maladaptive parenting practices when working with highly disorganized or resistant families (Patterson & Chamberlain, 1994; Patterson & Forgatch, 1985). Although understandable, this kind of extinction of attempts to promote behavior change is often countertherapeutic. In the study

comparing standardized and adaptive treatment of phobic disorders described in the introduction (Schulte et al., 1992), the presence or absence of in vivo exposure appeared to account for response to treatment. The authors speculate that, when given the option, therapists often attempted to minimize client resistance by not implementing this unpopular but especially powerful component of treatment. Likewise, our family coordinators' decisions to decrease the number of home visits as a means of reducing resistance and maintaining rapport may have reduced program impact for those families needing the most help.

The presence of these kinds of effects illustrates the subtle ways in which unintentional and idiosyncratic factors that affect decision making, such as personal beliefs and site-based norms, can influence more global assessments. Ratings that require more specific descriptions of defined domains appear less vulnerable to these sorts of biases.

Impact of Clinical Judgment on Outcomes

The results of the weighted regression analyses evaluating the effects of clinical judgment under different hypothetical conditions suggest that it was beneficial for Fast Track to use clinical judgment to tailor dose recommendations for home visiting. When the sample was weighted to simulate an implementation condition in which the research-based measures of parental functioning determined the frequency of home visits, the intervention effect on third-grade school behavior problems was nonexistent. When the sample was weighted to simulate an implementation condition using family coordinators' ratings of parental functioning to set dose recommendations, intervention had a significant effect reducing child school behavior problems. The results of the unweighted regression analyses, examining the impact of the intervention as conducted (e.g., setting the dose of home visits based on clinical ratings of parental functioning but allowing deviations based on global assessments of family need) demonstrated a reduced, marginally significant effect on that outcome.

The importance of these findings is that they illustrate the potential impact on intervention outcomes of using clinical judgment in different ways to inform treatment planning. They should not be interpreted as an evaluation of the impact of home visits on Fast Track outcomes. Fast Track included multiple components of treatment, only one of which was home visits. Most of the core components of Fast Track were standardized interventions delivered with fidelity to all children and families. It is noteworthy, however, that even within this context, the exploratory analyses that tested the simulated impact of using different forms of clinical judgment to adapt one component of the intervention still revealed differences in outcomes. These analyses suggest that further research evaluating the use of various kinds of clinical judgment for decision-making in adaptive designs is critical. True experiments in which intervention families are randomly assigned to different conditions that allowed different kinds of clinical judgment would provide a much stronger basis for conclusions. In the absence of such trials, however, the methodology used here represents a promising means of estimating what those effects might have been. This, in turn, can guide the design of future studies and, thus, accelerate the development of the critical empirical database. With this caveat in mind, the results of the weighted regressions are consistent with the other analyses undertaken in this study and, together, suggest several conclusions and future directions regarding the use of clinical judgment in adaptive intervention designs.

Conclusions and Future Directions

The results of this study suggest that clinical judgment may play a particularly useful role in the assessment of psychosocial characteristics and the description of client functioning, which can affect levels of optimal intervention dose. As in prior research (Dutra et al., 2004), clinicians in this study gave fairly reliable and valid ratings of parental functioning when provided with a rating scale that identified empirically validated and salient features of the construct and when

provided with concrete rating anchors to aid in their assessments. Future research focused on the development of reliable and valid clinical rating scales that could be used to tailor dose recommendations in adaptive interventions would be very useful. In addition, future research focused on the characteristics of the clinicians and their training that is related to an enhanced capacity to provide reliable ratings would be helpful.

Although adaptive interventions have proven more effective than standardized interventions in areas of drug treatment and health promotion (Kreuter et al., 2000; Lavori et al., 2000), studies to date examining adaptive mental health interventions reveal mixed findings. In some cases, adaptive interventions are less effective than standardized interventions (Chaffin et al., 2004; Schulte et al., 1992); in others, they convey no specific advantage (Project MATCH Research Group, 1997); whereas in others, the long-term effects are better for the adaptive than standardized intervention (Jacobson et al., 1989). Clearly, research that clarifies the conditions under which adaptive designs perform better than standardized interventions is needed.

This study suggests that clinical judgment can play a role in fostering effective adaptive interventions but primarily when used to assess client functioning. In contrast, the use of more global and categorical assessments in adaptive designs appears particularly vulnerable to undesirable biases and site-based effects. To avoid such influences, adaptive interventions may be best served when clinical judgment is based on empirically valid rating scales only. In this kind of model, clinicians are conceptualized as informants of adult, child, or family problems in the same way that parents, teachers, or the individuals themselves are informants of their status (Dutra et al., 2004; Westen & Weinberger, 2004). Clinical rating scales designed for intervention planning are guided and constrained by the specific questions of a psychometrically sound measure, and their effectiveness in guiding intervention recommendations may be evaluated empirically. Under these conditions, clinical judgment may provide a reliable, valid, and useful tool to apply in the assessment processes associated with adaptive designs and intervention planning.

Acknowledgments

We are grateful for the close collaboration of the Durham Public Schools, the Metropolitan Nashville Public Schools, the Bellefonte Area Schools, the Tyrone Area Schools, the Mifflin County Schools, the High-line Public Schools, and the Seattle Public Schools. We greatly appreciate the hard work and dedication of the many Fast Track staff members who implemented the project, collected the data, and assisted with data management and analyses.

This work was supported by National Institute of Mental Health (NIMH) Grants R18 MH48043, R18 MH50951, R18 MH50952, and R18 MH50953 and by National Institute on Drug Abuse Grant P50 DA10075 to the Center for Prevention Methodology. The Center for Substance Abuse Prevention and the National Institute for the Study of Drug Abuse also have provided support for Fast Track through a memorandum of agreement with the NIMH. In addition, this work was supported in part by Department of Education Grant S184U30002 and NIMH Grants K05MH00797 and K05MH01027.

References

- Allen JG, Coyne L, Logue AM. Do clinicians agree about who needs extended psychiatric hospitalization? *Comprehensive Psychiatry* 1990;31:355–362. [PubMed: 2387148]
- Arkes HR. Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology* 1981;49:323–330. [PubMed: 7276321]
- Beutler, LE.; Clarkin, JF. Systematic treatment selection: Toward targeted therapeutic interventions. Philadelphia: Brunner/Mazel; 1990.
- Bickman L, Karver MS, Schut JA. Clinician reliability and accuracy in judging appropriate level of care. *Journal of Consulting and Clinical Psychology* 1997;65:515–520. [PubMed: 9170776]
- Bierman, KL.; Greenberg, MT.; Conduct Problems Prevention Research Group. Social skills training in the Fast Track Program. In: Peters, RD.; McMahon, RJ., editors. Preventing childhood disorders, substance abuse, and delinquency. Thousand Oaks, CA: Sage; 1996. p. 65-89.

- Chaffin M, Silovsky JF, Funderburk B, Valle LA, Brestan EV, Balachova T, et al. Parent–child interaction therapy with physically abusive parents: Efficacy for reducing future abuse reports. *Journal of Consulting and Clinical Psychology* 2004;72:500–510. [PubMed: 15279533]
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 1994;6:284–290.
- Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. *Prevention Science* 2004;5:185–196. [PubMed: 15470938]
- Conduct Problems Prevention Research Group. A developmental and clinical model for the prevention of conduct disorders: The Fast Track program. *Development and Psychopathology* 1992;4:509–527.
- Conduct Problems Prevention Research Group. Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology* 2002;30:19–35. [PubMed: 11930969]
- Cooperative Research Group. Rationale and design of a randomized clinical trial on prevention of stroke in isolated systolic hypertension. *Journal of Clinical Epidemiology* 1988;41:1197–1208. [PubMed: 2905387]
- Costello EJ, Compton SN, Keeler G, Angold A. Relationships between poverty and psychopathology: A natural experiment. *Journal of the American Medical Association* 2003;290:2023–2029.
- Crnic KA, Greenberg MT. Minor parenting stresses with young children. *Child Development* 1990;61:1628–1637. [PubMed: 2245752]
- Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science* 1989 March 31;243:1668–1674. [PubMed: 2648573]
- Dodge KA, Bates JE, Pettit GS. Mechanisms in the cycle of violence. *Science* 1990 December 21;250:1678–1683. [PubMed: 2270481]
- Dutra L, Campbell L, Westen D. Quantifying clinical judgment in the assessment of adolescent psychopathology: Reliability, validity, and factor structure of the Child Behavior Checklist for Clinical Report. *Journal of Clinical Psychology* 2004;60:65–85. [PubMed: 14692010]
- Garb, HN. Studying the clinician: Judgment research and psychological assessment. Washington, DC: American Psychological Association; 1998.
- Garb HN. Clinical judgment and decision making. *Annual Review of Clinical Psychology* 2005;1:67–89.
- Goldberg LR. Simple models or simple processes? Some research on clinical judgments. *American Psychologist* 1968;23:483–496. [PubMed: 4385525]
- Hollingshead, AA. A four factor index of social status. Yale University; New Haven, CT: 1975. Unpublished manuscript
- Jacobson NS, Schmalings KB, Holtzworth-Munroe A, Katt JL, Wood LF, Follette VM. Research-structured vs. clinically flexible versions of social learning-based marital therapy. *Behaviour Research and Therapy* 1989;27:173–180. [PubMed: 2930443]
- Jensen AL, Weisz JR. Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *Journal of Consulting and Clinical Psychology* 2002;70:158–178. [PubMed: 11860042]
- Kazdin AE. Treatment of antisocial behavior in children: Current status and future directions. *Psychological Bulletin* 1987;102:187–203. [PubMed: 3310058]
- Kreuter, M.; Farrell, D.; Olevitch, L.; Brennan, L. Tailoring health messages, customizing communication with computer technology. Mahwah, NJ: Erlbaum; 2000.
- Lavori PW, Dawson R. Developing and comparing treatment strategies: An annotated portfolio of designs. *Psychopharmacology Bulletin* 1998;34:13–18. [PubMed: 9564193]
- Lavori PW, Dawson R, Roth AJ. Flexible treatment strategies in chronic disease: Clinical and research implications. *Biological Psychiatry* 2000;48:605–614. [PubMed: 11018231]
- Leventhal T, Brooks-Gunn J. The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin* 2000;126:309–337. [PubMed: 10748645]
- Lewczyk CM, Garland AF, Hurlburt MS, Gearity J, Hough RL. Comparing DISC-IV and clinician diagnoses among youths receiving public mental health services. *Journal of the American Academy of Child and Adolescent Psychiatry* 2003;42:349–356. [PubMed: 12595789]

- Lochman JE, Conduct Problems Prevention Research Group. Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology* 1995;6:549–559. [PubMed: 7673532]
- Luoma I, Tamminen T, Kaukonen P, Laippala P, Puura K, Salmelin R, Almqvist F. Longitudinal study of maternal depressive symptoms and child well-being. *Journal of the American Academy of Child and Adolescent Psychiatry* 2001;40:1367–1374. [PubMed: 11765281]
- Lynam D, Moffitt T, Stouthamer-Loeber M. Explaining the relation between IQ and delinquency: Class, race, test motivation, school failure, or self-control? *Journal of Abnormal Psychology* 1993;102:187–196. [PubMed: 8315131]
- McMahon, RJ.; Slough, NM.; Conduct Problems Prevention Research Group. Family-based intervention in the Fast Track Program. In: Peters, RD.; McMahon, RJ., editors. *Preventing childhood disorders, substance abuse, and delinquency*. Thousand Oaks, CA: Sage; 1996. p. 90-110.
- Murphy SA, van der Laan MJ, Robins JM, Conduct Problems Prevention Research Group. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association* 2001;96:1410–1423.
- Patterson, GR.; Capaldi, D.; Bank, L. An early starter model for predicting delinquency. In: Pepler, DJ.; Rubin, KH., editors. *The development and treatment of childhood aggression*. Hillsdale, NJ: Erlbaum; 1991. p. 139-168.
- Patterson GR, Chamberlain P. A functional analysis of resistance during parent training therapy. *Clinical Psychology: Science and Practice* 1994;1:53–70.
- Patterson GR, Forgatch MS. Therapist behavior as a determinant for client noncompliance: A paradox for the behavior modifier. *Journal of Consulting and Clinical Psychology* 1985;53:846–851. [PubMed: 4086685]
- Pavkov TW, Lewis DA, Lyons JS. Psychiatric diagnoses and racial bias: An empirical investigation. *Professional Psychology: Research and Practice* 1989;20:364–368.
- Payette KA, Clarizio HF. Discrepant team decisions: The effects of race, gender, achievement, and IQ on LD eligibility. *Psychology in the Schools* 1994;31:40–48.
- Pettit GS, Harrist AW, Bates JE, Dodge KA. Family interaction, social cognition, and children's subsequent relations with peers at kindergarten. *Journal of Social and Personal Relationships* 1991;8:383–402.
- Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol* 1997;58:7–29. [PubMed: 8979210]
- Radloff LA. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;1:385–401.
- Reynolds AJ, Weissberg RP, Kasrow WJ. Prediction of early social and academic adjustment of children from the inner city. *American Journal of Community Psychology* 1992;20:599–624. [PubMed: 1485613]
- Robins RW, Hendin HM, Trzesniewski KH. Measuring global self-esteem: Construct validation of a single item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin* 2001;27:151–161.
- Rock DL, Bransford JD, Maisto SA, Morey L. The study of clinical judgment: An ecological approach. *Clinical Psychology Review* 1987;7:645–661.
- Rock DL, Bransford JD, Morey LC, Maisto SA. The study of clinical judgment: Some clarifications. *Clinical Psychology Review* 1988;8:411–416.
- Schulte D, Kunzel R, Pepping G, Schulte-Bahrenberg T. Tailor-made versus standardized therapy of phobic patients. *Advances in Behavior Research and Therapy* 1992;14:67–92.
- Segal SP, Bola JR, Watson MA. Race, quality of care, and antipsychotic prescribing practices in psychiatric emergency services. *Psychiatric Services* 1996;47:282–286. [PubMed: 8820552]
- Spache, GD. DRS: Diagnostic Reading Scales. Examiner's Manual. Monterey, CA: McGraw-Hill; 1981.
- Stanger C, McConaughy SH, Achenbach TM. Three-year course of behavioral / emotional problems in a national sample of 4- to 16-year olds: II. Predictors of syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry* 1992;31:941–950. [PubMed: 1400129]

- Startup M, Jackson MC, Bendix S. The concurrent validity of the Global Assessment of Functioning (GAF). *British Journal of Clinical Psychology* 2002;41:417–422. [PubMed: 12437796]
- Werthamer-Larsson L, Kellam SG, Wheeler L. Effect of first grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology* 1991;19:585–602. [PubMed: 1755437]
- Westen D, Weinberger J. When clinical description becomes statistical prediction. *American Psychologist* 2004;59:595–613. [PubMed: 15491255]

Table 1
Means and Standard Deviations of Study Measures

Measure	Second grade			Third grade		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Ratings of parental functioning	250	17.20	4.13	354	17.22	4.23
Global assessment of need	250	1.65	0.95	354	1.49	0.77
Research-based measures of parental functioning						
Observed parent-child warmth	257	3.80	0.79	364	3.68	0.82
Physical punishment	259	0.13	0.16	383	0.11	0.14
Maternal depression	259	15.59	9.92	384	15.44	10.14
Stressful life events	259	5.21	4.03	384	5.40	4.07
Parent-school involvement	260	1.62	0.54	386	1.55	0.58
Risk and protective factors						
Family SES	260	25.80	12.83	386	25.25	12.41
Single-parent household	260	0.52	0.50	387	0.50	0.50
Quality of home environment	247	9.33	1.71	354	9.27	1.86
Neighborhood danger	248	8.88	2.22	359	8.93	2.42
Child academic achievement	259	0.07	0.92	377	0.13	0.93

Note. SES = socioeconomic status.

Table 2
Relations Between Clinical Judgments and Research-Based Measures

Measure	Second grade		Third grade	
	Ratings of functioning	Global assessment of need	Ratings of functioning	Global assessment of need
Ratings of parental functioning		-.44***		-.60***
Research-based measures of parental functioning				
Observed parent-child warmth	.29***	-.07	.27***	-.14**
Physical punishment	-.02	-.04	-.11*	.10
Maternal depression	-.25***	.07	-.29***	.14**
Stressful life events	-.09	.06	-.16**	.11
Parent-school involvement	.32***	-.16**	.39***	-.23***
Risk and protective factors				
Family SES	.36***	-.12*	.32***	-.25***
Single-parent household	-.20**	.00	-.16**	.07
Quality of home environment	.35***	-.18**	.21***	-.20***
Neighborhood danger	-.22***	.16**	-.23***	.10
Child academic achievement	.16*	-.12	.25***	-.22***

Note. All coefficients are standardized and comparable to bivariate correlations, except they control for the effects of the individual family coordinators. SES = socioeconomic status.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 3
Predicting Clinical Ratings of Parental Functioning and Global Assessment of Need

Sets of predictors	Second grade		Third grade	
	Ratings of functioning	Global assessment of need	Ratings of functioning	Global assessment of need
Replicable and desirable influences Research-based measures of parental functioning				
ΔF	11.42 ^{***}	1.93	18.25 ^{***}	4.30 ^{***}
df	5,215	5,215	5,292	5,292
Adjusted R^2	.19	.02	.23	.05
Risk and protective factors				
ΔF	4.99 ^{***}	2.95 [*]	4.31 ^{***}	3.98 ^{**}
df	5,210	5,210	5,287	5,287
Adjusted R^2	.26	.06	.27	.10
Intervention participation				
ΔF	2.15	2.41	5.66 ^{**}	2.40
df	2,208	2,208	2,285	2,285
Adjusted R^2	.27	.08	.29	.11
Potential sources of bias Cohort and site				
ΔF	3.27 [*]	3.59 ^{**}	4.09 ^{***}	12.86 ^{***}
df	4,204	4,204	5,280	5,280
Adjusted R^2	.30	.12	.33	.26
Family coordinator effects				
ΔF	2.36 ^{***}	5.51 ^{***}	3.56 ^{***}	1.25
df	25,179	25,179	26,254	26,254
Adjusted R^2	.40	.43	.46	.28
Child sex and race				
ΔF	1.56	0.04	0.18	1.47
df	2,177	2,177	2,252	2,252
Adjusted R^2	.40	.43	.45	.28

Note. Changes in the model fit statistics are listed in columns.

*
 $p < .05$.

**
 $p < .01$.

 $p < .001$.

Table 4
Predicting When Family Coordinators Deviated From Project Guidelines in Making Recommendations for Home Visits

Sets of predictors	Judgments of increased need for home visiting		Judgments of decreased need for home visiting	
	Second grade (<i>n</i> = 198)	Third grade (<i>n</i> = 263)	Second grade (<i>n</i> = 86)	Third grade (<i>n</i> = 108)
Project guidelines				
Δ-2LL	143.88	186.62 ^{**}	81.77	119.21
<i>df</i>	1	1	1	1
Rescaled <i>R</i> ²	.02	.06	.02	.00
Repiicable and desirable influences Research-based measures of parental functioning				
Δ-2LL	6.69	4.73	3.38	1.76
<i>df</i>	5	5	5	5
Rescaled <i>R</i> ²	.09	.09	.08	.02
Risk and protective factors				
Δ-2LL	22.95 ^{***}	19.59 ^{***}	4.47	5.40
<i>df</i>	5	5	5	5
Rescaled <i>R</i> ²	.29	.22	.16	.10
Intervention participation				
Δ-2LL	0.72	1.33	0.77	0.90
<i>df</i>	2	2	2	2
Rescaled <i>R</i> ²	.29	.23	.17	.11
Potential sources of bias Cohort and site				
Δ-2LL	6.43	23.86 ^{***}	51.04 ^{***}	52.11 ^{***}
<i>df</i>	4	5	3	5
Rescaled <i>R</i> ²	.34	.38	.82	.64
Child sex and race				
Δ-2LL	3.45	4.37	3.41	8.98 ^{**}
<i>df</i>	2	2	2	2
Rescaled <i>R</i> ²	.37	.40	.85	.71

Note. Changes in the model fit statistics are listed in columns. The rescaled *R*² divides the original *R*² by its upper bound to account for the fact that the dependent variable is discrete. LL = log likelihood.

^{**}
 $p < .01$.

^{***}
 $p < .001$