



The Course and Malleability of Aggressive Behavior from Early First Grade into Middle School: Results of a Developmental Epidemiologically-based Preventive Trial

Sheppard G. Kellam, George W. Rebok, Nicholas Ialongo and Lawrence S. Mayer

Abstract—The course and malleability of aggressive behavior from beginning elementary school through transition into middle school was investigated. In a developmental epidemiologically defined population of 1000 urban first graders, a two year classroom-based randomized preventive trial was aimed at reducing aggressive behavior, an antecedent of delinquency, violent behavior, and heavy drug use in adolescence and adulthood. Earlier we reported impact in first grade on teacher and peer ratings and on classroom observations. We report here on the course and on sixth grade teacher ratings of aggressive behavior. Improvement was observable during transition times, in first grade and in middle school, among the males in the preventive intervention who were more aggressive in first grade.

Keywords: Developmental epidemiology, preventive trials, aggression, childhood and adolescence, malleability

The importance of aggressive behavior early in childhood has been amply documented with respect to the manifestation of a variety of problem outcomes later in the life course. The context of the school and classroom has been particularly prominent in these findings. Aggressive behavior in the form of tardiness, breaking rules and fighting has repeatedly been found to predict antisocial behavior, violence, criminality and heavy substance use, including I.V. drug use, through adolescence and well into adulthood especially in males (Anthony, 1985; Ensminger, Kellam & Rubin, 1983; Farrington, Gallagher, Morley, St Ledger & West, 1988; Kellam, Brown, Rubin & Ensminger, 1983; McCord, 1988; Robins, 1978; Tomas, Vlahov & Anthony, 1990). Our Woodlawn/Chicago studies beginning in the 1960s and still continuing documented these relationships prospectively in epidemiologically defined urban

Accepted manuscript received 30 April 1993

Francis Scott Key Medical Center, Baltimore, MD, U.S.A

Requests for reprints to: Professor S. G. Kellam, Prevention Research Center, Department of Mental Hygiene, School of Hygiene and Public Health, Mason F. Lord Building, Suite 500, Francis Scott Key Medical Center, 4940 Eastern Avenue, Baltimore, MD 21224, U.S.A.

populations as early as first grade, using teacher ratings as a core measure in a semi-structured interview periodically conducted in first and third grades (Ensminger *et al.*, 1983; Kellam, Branch, Agrawal & Ensminger, 1975; Kellam *et al.*, 1983). Block and colleagues have reported these important developmental relationships from age 3 to 14 (Block, Block & Keyes, 1988; Shedler & Block, 1990); White, Moffitt and Silva (1989) reported similar relationships from age 3 to 15, although with somewhat less specificity as to outcome.

Eron (1990) has suggested that aggressive antisocial behavior crystallizes around age 8. Although aggression in early childhood can be highly stable, possibly as many as half the children who exhibit aggressive behavior early in life course do not go on to become criminals or drug users as adolescents or young adults (Ensminger *et al.*, 1983; Kellam *et al.*, 1983; Robins, 1978). Understanding the processes underlying continuity and discontinuity in aggression is important to both theory-building and prevention. Over the course of first grade in Woodlawn, some children responded aggressively to the teacher's social task demands and persisted with this behavioral response through the school year, while others began later. Results from the 10 year follow-up show that the predictive strength of this maladaptive behavioral response was evident as early as the end of the first-grade year (Kellam, Simon & Ensminger, 1983). The process of social adaptation evolved over the year with children differentiating into those who adapted to the social tasks and those who did not or could not.

This paper reports on variation in the course of malleability of aggression over an important stage of life—the transition into first grade through the transition into middle school. We report on the results of an intervention directed at the highly interactive process of demand/response between teacher and child in the context of the first and second grade classroom. This process is termed *social adaptation*, and the resultant ratings of adequacy of behavioral performance *social adaptational status* (SAS). We hypothesized that the social adaptational process was malleable and that the results of improving the precision of defining the tasks and involving the classmates in the administering of the reinforcements would result in improved behavior of the aggressive child, leading to improved developmental trajectories. Testing the malleability of the developmental course of aggressive behavior is a separate issue from the question of the origins of aggressive behavior. Regardless of whether the individual is predisposed to aggressive responses biologically, and/or has learned them through prior life experience, and/or whether societal influences exacerbate or inhibit the behavior—the potential for change toward less aggressive behavioral responses remains unknown until tested through experimental preventive intervention trials.

Findings from several preventive trials have suggested the importance of targeting antisocial behavior, particularly aggressive behavior in boys during the early elementary grades, for preventing later delinquency and drug abuse (Hawkins, Catalano, Morrison, O'Donnell, Abbott & Day, 1992; Hawkins, Von Cleve & Catalano, 1991; Kolvin, Miller, Fleeting & Kolvin, 1988; Tremblay, Masse, Perron, Leblanc, Schwartzman & Ledingham, 1992; Tremblay *et al.*, 1991; Tremblay *et al.*, 1992). The Seattle Social Development Project (Hawkins *et al.*, 1991, 1992) involved four years of experimental intervention with teachers and parents of a multiethnic panel of children as they passed from grades 1 through 4 in eight schools of an urban

school district. The intervention targeted family behavior management practices and teacher instructional strategies with the aim of increasing opportunities for children's social involvement, skills for successful participation in these units, and social reinforcements for prosocial participation. The boys in the experimental group were rated by teachers as significantly less aggressive and experimental group girls were rated as significantly less self-destructive than the control group at the end of two years of intervention. A longitudinal-experimental study conducted in Montreal (Tremblay *et al.*, 1991, 1992) targeted antisocial behavior of boys who were disruptive in the classroom, using parent training and children's social skills training to prevent delinquency. Results from three years of follow-up after two years of treatment (at ages 7 and 8) indicated that the treated disruptive boys were less physically aggressive in school, were more often in an age-appropriate regular classroom, had less serious school adjustment problems, and reported fewer delinquent behaviors.

We studied the course and malleability of aggression, along with other maladaptive behavioral responses to the social task demands of the classroom, through the application of two precisely defined preventive intervention trials in a parallel design aimed at specific but correlated predictors. Achievement and aggressive behavior are correlated; and both are important predictors of long-term outcomes (Kellam *et al.*, 1983; Rutter, Tizard & Whitmore, 1970; Shaffer *et al.*, 1979). One intervention, Mastery Learning (ML), was aimed at poor achievement while the other, the Good Behaviour Game (GBG), was aimed at aggressive behavior. Using the two trials in parallel, we studied the developmental interrelationships among the two target antecedents and their more distal outcomes.

In this paper we focus on the GBG intervention. The detailed results at the end of first grade are reported in a different manuscript (Dolan *et al.*, 1993). We describe here the course of impact through the transition into middle school. This paper is concerned primarily with the school context due to space limits, although we will present results on the relationship of teacher ratings of aggression to conduct disorder as identified on the Diagnostic Interview Schedule for Children (DISC), in order to provide evidence of the link between teacher ratings and conduct disorder.

The children who received the ML intervention are considered here as an active control for the amount of attention the GBG classrooms and children received. The specificity of the results of the GBG can be reasonably assured by the use of the two separate but theoretically related preventive interventions in a parallel design. We made the decision to use ML as an active control since it had impact on early achievement, but had no direct effect on aggression. ML did have an influence on aggressive behavior mediated through achievement, however, and this is reported in Kellam, Mayer, Rebok, and Hawkins (1993). The GBG had short-term impact on aggression but no direct or indirect impact upon achievement. Each intervention thus appears to be specific to its own proximal target, as revealed by the absence of direct effects on the proximal target of the other intervention. This lack of crossover effect implies a different etiologic source of early learning problems and behavioral problems (Kellam, Mayer, Rebok & Hawkins, 1993).

The research we report here on the course and malleability of aggressive behavior is developmental epidemiological, derived from integrating orientations and methods from life-course development and community epidemiology (Kellam, 1991; Kellam

& Ensminger, 1980). Our approach is based on community epidemiology which is concerned with the non-random distribution of a health problem or related factor in a fairly small population in the context of its environment such as neighborhood, school or classroom. General characteristics such as poverty and ethnicity are held constant. The focus is on variation in risk among subgroups of individuals and why some people in some circumstances are at greater risk. Generalization from beyond the community under study is done by replication in similar and/or different communities. The classic *host/agent/environment* orientation to the modeling stems from this epidemiological perspective (Morris, 1975). This orientation refers to a way of conceptualizing cause or etiology as involving vulnerability in the person (the host), illness producing conditions in the environment, and a causal process of interaction (the agent) between the individual and environmental risk conditions.

The integration with life course development allows the study of variation in developmental antecedents and paths in a defined population in a defined ecological context. Antecedents along the paths can be specifically and precisely targeted, their frequencies determined, and the variation in the function of the antecedents assessed with markedly reduced bias in the subject selection. The preventive trial can be aimed at the question of the malleability of the antecedent and of its relationship to outcomes. Such trials not only allow the exploration of the variation in developmental paths, but also the differences among the responders and the non-responders to the preventive intervention. The trial has the dual purposes of experimentally testing the developmental functions of the target antecedent and of developing effective preventive programs (Kellam & Rebok, 1992; Kellam & Werthamer-Larsson, 1986).

Developmental epidemiologically-based preventive trials such as the one we will describe should have considerable ecological validity, because they are carried out on representative populations in main social fields such as the communities' classrooms, families and peer groups. We have elaborated the developmental epidemiological orientation with the concepts of social adaptation and SAS mentioned above. Within this framework are the social task demands faced by individuals in each of the main social fields relevant to each stage of life within a defined community. These include in many societies the family of procreation, the school and classroom, the peer group, the intimate/marital social field, and the work social field (Kellam *et al.*, 1975). Within each social field specific persons (termed *natural raters*) have the role of defining the social tasks and rating the adequacy of performance or behavioral responses of the individuals in that field. Teachers have such a role in the classroom, as do parents in the family, spouses in the marital field, and supervisors at work. SAS measures are sometimes formal as in the case of grades and sometimes informal such as parental valuations that the child is bad or good. In addition, SAS is not an unbiased or objective measure of the child. Chance, idiosyncrasy of the natural rater, and the fit of the child with the classmates may all play a role in the teacher's ratings.

The developmental epidemiological framework further distinguishes between SAS, a social status measure regarding degrees of passing or failing in a specific social field at a specific stage of life, and *psychological well-being* (PWB) (Kellam, 1990; Kellam & Ensminger, 1980; Kellam *et al.*, 1975). PWB refers to internal states such as psychiatric distress in the level of symptoms such as depression, anxiety, bizarre feelings and thoughts, mental disorders, and neuropsychological or neurophysiological status.

PWB can influence the individual's social adaptational performance and/or be affected by the resultant SAS. Although the role of PWB, as antecedent and/or consequence of poor SAS, is not the subject of this paper, it is the next step in analyses of the impact of the two Baltimore trials.

In the present report the classroom is our focus and the teacher is the natural rater. Some classroom social task demands with their maladaptive responses include adapting to rules and authority (aggressive behavior; shy behavior; social participation) and concentration (attention or concentration problems) (Kellam *et al.*, 1975, 1991; Werthamer-Larsson, Kellam & Wheeler, 1991). Poor reading achievement is another example of poor SAS, while the depression that may have preceded and/or followed it is an example of PWB. Throughout the Woodlawn/Baltimore research our focus has been on the developmental course of relationships among SAS in specific social fields and PWB. Among whom, in what social field, and on what tasks does SAS lead to PWB, or PWB lead to SAS, or a reciprocal relationship occur in which each influences the other over the course of development?

The theoretical importance of SAS in contrast to the internalizing/externalizing dichotomy is that it enlarges and contextualizes the idea of externalizing symptoms and places these in a dynamic relationship to internalizing symptoms. The internalizing/externalizing framework seems to place the locus of attention on the individual rather than also including the way the individual's behavior and condition fit with the context and with the natural rater's view of acceptable behavior. SAS focusses attention on the specific social field and stage of life and on the interactive and demand/response qualities of the processes leading to ratings of adaptation or maladaptation. The basic thesis of SAS is that succeeding or failing at specific social tasks in specific main social fields such as classroom, peer group, and/or family can have an important developmental influence on later specific outcomes.

The GBG is aimed at the social adaptational process in the classroom related to rules and authority. The child, the teacher and the classmates, as well as the structure of the classroom itself are involved. We chose this setting for the experimental intervention to test the hypothesis that social maladaptive aggressive behavioral responses were malleable through changing the social adaptational process in the classroom; that the changes would remain in the child's coping responses to later social task demands concerned with rules and authority; and that these developmental effects would improve the risk of the long-term outcomes cited above.

The subject population was epidemiologically defined to insure against the biases inherent in the use of volunteer or clinic samples (Greenley & Mechanic, 1976; Greenley, Mechanic & Cleary, 1987; Kellam, 1990). Those who volunteer or come for help to a clinic may be quite different from those who do not and, as such, the developmental models that are derived from the volunteer or clinic samples may not generalize to the population of interest as a whole.

The results on the short-term impact of the two preventive trials in Baltimore support the hypothesis that early social maladaptive behaviors are malleable (Dolan *et al.*, 1993). For both males and females, the GBG had a significant impact on aggressive behavior as rated by teachers. Classmates' nominations of aggressive behavior among males (but not females) also were significantly reduced. By examining scatter-plots of aggressive behavior ratings in the fall and the spring, along with the regression

slopes, it appeared that the more severe end of aggressive behavior was affected by GBG for both genders.

Hypotheses and research questions

The major goals of the present paper were: (a) to examine the course of aggression among male and female children from entrance to first grade through six years of elementary and early middle school, and (b) to evaluate the longer-term impact of the GBG on this developmental course. We hypothesized that the course of aggression is malleable, in that the behavior of children receiving the GBG intervention will improve. More specifically, we hypothesized that GBG children who are more aggressive in fall of first grade will exhibit reduced aggressive behavior from early elementary school through middle school than children in control conditions or those receiving the ML intervention. GBG children who are not aggressive in fall of first grade are hypothesized to exhibit lower incidence of aggressive behavior over the course of the six years than initially non-aggressive children in the control classrooms or ML classrooms.

Method

Subjects

The total sample for the first phase of the intervention research consisted of 1,197 children who entered the first grades of 19 Baltimore public schools during the 1985-86 academic year and were present at first report card time when the first measures were done prior to intervention. The sample was 49% males. Sixty-five percent of the sample was African American, 31% was Caucasian and 4% represented other ethnic groups. At first grade, the children ranged in age from 5.0 to 9.7 years (mean age = 6.6 years, $SD = 0.49$). Of the 1,197 children available for participation, 1,084 were assessed. Of the remaining children, 48 either had transferred out of the participating schools prior to consent being requested, or could not be reached for response to the consent request. About 5.0% of the parents of the 1,197 children refused to allow their children to participate. Chi-square analyses revealed that refusal rates, although fairly low, varied as a function of geographic area, $\chi^2(16) = 45.59, p < .0001$. The highest rates of refusals were in Areas 1 and 4, which were made up primarily of middle income, two-parent families, living in well-maintained row or detached homes.

Annual follow-up and replicate sampling

Detailed tracking and assessment procedures were developed to follow up the children from first grade through middle school. A replicate sampling procedure was used for each annual follow-up period. The replicate sampling plan was devised to ensure representativeness of the follow-up samples despite unprecedented budget cuts that reduced the extent of sample follow-up. The total population of children for follow-up was divided into three or four replicate samples that were representative of the total population with regard to levels of severity and geographical area. This strategy included assessment of the difficult and geographically distant cases in each replicate as the follow-up proceeded rather than confounding the data by focussing on the easily located or geographically nearest ones. The small differences reported below between the children assessed and those lost to follow-up provide evidence that our replicate sampling plan probably helped.

At the time of the six-year follow-up, 170 children were found to have repeated at least one grade or were in lower grades than sixth grade and are included in these results. Although their outcomes are not distinguished from the results reported here, they will be separated for specific study in later reports. A total of 693 children participated in the same intervention or control condition for two years. Of those 693, 590 children were assessed at the six-year follow-up. The characteristics of the children

at pre-intervention who had been assessed at each of the follow-ups (spring 1986, fall 1986, spring 1987, spring 1988, 1989, 1990, 1991) were compared with those of children who were not assessed each time. These characteristics included first-grade teacher-rated aggressive behavior, shy behavior and concentration problems; self-reported depressive symptoms; and first-grade California Achievement Test reading scores. Significant differences in baseline scores between the two groups were found for teacher-rated aggressive behavior at each follow-up assessment except one (spring, 1989), although the effect sizes for each difference were trivial ($R^2 = .00-.02$). For each comparison, the children not assessed had been rated slightly higher on aggression in fall of first grade by their teacher than the children who were assessed at follow-up. We plan to try to assess the children lost to follow-up to determine if they affect the impact results. Over the entire follow-up, four of the comparisons for concentration problems, three for reading achievement, two for shy behavior, and one for depression reached statistical significance ($p < .05$), but the effect sizes again were trivial ($R^2 = .00-.02$ for all but two comparisons where $R^2 = .05$).

Design

The Baltimore City Public Schools and the investigators designated a widely varied set of elementary schools in the eastern half of Baltimore City, where five different urban areas were selected. The choices were made without restrictions other than neighborhood and school characteristics. Most of the children in this community attend neighborhood schools in the elementary grades. The five urban areas represented five different social contexts with respect to the ethnic makeup of neighborhoods, their socioeconomic status (SES), type of housing, family structure and stability. Some of the areas exhibited many of the characteristics of community decay and poverty that are associated with high risk of problem behavior. Some were more lower to middle-class areas and had many characteristics of community organization and access to resources associated with lower risk. Each urban area was served by at least three public elementary schools that could be matched, each with two or more first-grade classrooms.

Before starting the preventive trials, we examined the schools and the areas they served in relation to a profile of characteristics. Matching characteristics included students' achievement levels, family SES and ethnicity. From a total of 19 schools, the three or four most similar schools were identified within each of five urban areas for assignment to one of three conditions: one school to GBG, one to ML and the remaining school(s) to an *external control* condition with no experimental intervention.

The interventions included two cohorts that were direct replications in the same schools. In each cohort, classroom membership remained stable over the course of the two years of intervention except for children who moved to other schools and new children who replaced them in the classrooms. This paper will be concerned solely with results of Cohort 1. Analyses of Cohort 2 will be used to test developmental models constructed from Cohort 1.

Both interventions were administered at the classroom level. In order to avoid confounding intervention status with school effects, we divided classrooms in intervention schools (GBG and ML schools) into two types: intervention classes (that is, GBG or ML) and *internal control* classes (classes in the same school which did not receive either intervention). This design also provided for an estimate of leakage or spillover effects that might happen if all or part of the intervention strategies were adopted in the comparison classrooms. Individual first-grade classrooms were randomly assigned to intervention or internal control within the intervention schools. Prior to beginning the intervention work, the school principals agreed to assign the entering first-grade children to classrooms using a random method. Randomizing backed up by corrections in assignments as needed produced very similar classrooms of children at least at the beginning of the first-grade year. Gender, pre-school and kindergarten level, and conduct grades were assessed to insure good balance. The intervention and assessment framework used for the preventive trial design from fall of first grade through the end of sixth grade is shown in Table 1.

Measures

The core construct of aggression described in this paper is based on an SAS measure from fall of first grade periodically through spring of sixth grade. This measure consists of a semi-structured interview with the teacher to obtain ratings of each student's performance of the main social tasks required. In the Woodlawn studies, teacher ratings of early aggressive behavior in males (not females) predicted increased teenage delinquency and drug, alcohol and cigarette use at 16 or 17 (Kellam *et al.*, 1975;

Table 1. Intervention and assessment framework for preventive trials in 19 public elementary schools

	No. of classrooms	No. of students	85-86 Grade 1 F/S	86-87 Grade 2 F/S	87-88 Grade 3 F/S	88-89 Grade 4 F/S	89-90 Grade 5 F/S	90-91 Grade 6 F/S
Cohort I								
Good Behavior Game	8	153	ROXO	XO	O	O	O	O
Mastery Learning	9	163	ROXO	XO	O	O	O	O
Internal and external controls	24	377	RO O	O	O	O	O	O

F = Fall.

S = Spring.

O = Assessment.

X = Intervention.

R = Random assignment within intervention schools and matched schools as external controls.

Kellam, 1990). In addition to the Woodlawn findings, teacher ratings of early aggressive behavior in the delinquency and substance abuse literature are consistently related to both delinquency and later heavy substance use (Farrington & Hawkins, 1991).

The SAS measure described here represents a first-stage measure (that is, data gathered on the whole population) in a multistage sampling and assessment strategy. Strategies involving second-stage measures of smaller, stratified probability samples for more microanalytic studies are discussed in Kellam (1990) and Kellam and Rebok (1992). In the current paper a diagnostic interview for conduct disorder as a second-stage measure was administered to all children who reported three or more DSM-III-R conduct disorder behaviors in a first-stage interview and to a representative sample of the total population.

Aggressive behavior. The primary measure of aggressive behavior used in this paper is the *Teacher Observation of Classroom Adaptation-Revised (TOCA-R)*. Teacher ratings on this structured interview were obtained in the fall and spring of first and second grades, and spring of third, fourth, fifth and sixth grades. Children who had fallen behind their proper grade were assessed in their actual grade. The TOCA-R instrument was originally developed and used in the Woodlawn studies (Kellam *et al.*, 1975), and after modification, has been used as a major periodic assessment instrument for the prevention trials (Werthamer-Larsson *et al.*, 1991). The interview consists of an engagement section to develop rapport, a rating section to obtain behavioral assessments of each child in the class, and a termination section to obtain teachers' feedback about the interview. Teachers were interviewed by trained Prevention Center staff members during the school day in a private place. During the interview the teachers made ratings of each child in their class on three factors from the TOCA-R that represent the classroom social task demands in addition to the achievement demand. The tasks with the maladaptating behavioral responses were: Authority Acceptance (aggressive behavior); Social Contact (shy behavior); and Concentration (concentration problems). The aggressive behavior scale includes items such as *breaks rules, fights and harms property* (see Table 2). The shy behavior scale includes items such as *plays with classmates and initiates interactions*. The concentration scale includes items such as *pays attention, works well alone and completes assignments*.

The interview items were rated using a 6-point frequency-based scale, with items ranging from *almost never* to *almost always*. Positive items (such as *plays with others, self-reliant, completes assignments and pays attention*) were reversed to allow a consistent level of adaptation among items. After reversal, 1 reflected positive items occurring almost always and 6 reflected negative items occurring almost never.

The psychometric properties of the TOCA-R were investigated in a sample of 1,043 first-grade children in 19 East Baltimore elementary schools (Werthamer-Larsson *et al.*, 1991). The alphas for the three factors of Authority Acceptance, Social Contact and Concentration were .92, .85 and .96, respectively. Test-retest reliability over a 2-week period was investigated in a subsample of 361 children. Spearman correlations and Pearson correlations between Time 1 and Time 2 behavioral factors were all significant, with coefficients ranging from .74 to .94. The behavioral factors were significantly associated with a

Table 2. Aggressive behavior items on the TOCA-R

Lies	Stubborn
Breaks rules	Takes others' property
Yells at others	Teases classmates
Breaks things	Harms property/damages property on purpose
Fights	Harms others/hurts others physically
Argues with adults	Skips school
Uses a weapon in a fight	Runs away from home overnight
Starts fights with classmates	Sets fires
Has broken into someone else's house, building or car	Loses temper

set of concurrent measures including spring standardized achievement scores, fourth quarter report card grades and work habits, and absence during first grade. In addition, the behavioral factors were significantly associated with a set of antecedent variables including gender, kindergarten standardized achievement scores, kindergarten report card grades and kindergarten work habits, whether the child repeats first grade, preschool experience and changing schools between kindergarten and first grade.

Conduct disorder. The Diagnostic Interview Schedule for Children (DISC 2.25C; Shaffer, Fisher, Piacentini, Schwab-Stone & Wicks, 1991) was used to identify conduct disorder at the sixth grade follow-up. The Conduct Disorder Module of the DISC 2.25C was administered to a second-stage sample of 184 children consisting of a randomly selected sample of 27 children and 157 children who had screened positive on a conduct problems checklist, which included items reflecting the 13 Diagnostic and Statistical Manual of Mental Disorders III-Revised (DSM-III-R, American Psychiatric Association (APA), 1987) criteria for conduct disorder. The screener had 11 items because the DSM-III-R item of *forcing sexual acts* was not used due to community objections and the two DSM-III-R items regarding *taking of property of others, in a confrontational manner or a covert manner* were collapsed into one item. To be deemed positive on the screener, the child needed to report engaging in three or more DSM-III-R conduct problems in the last six months. The entire DISC conduct module was administered if the child screened positive.

The DISC is a fully structured interview that specifies the exact wording and sequence of questions and provides a complete set of categories for classifying respondents' replies. The highly structured format is intended to minimize clinical judgement in eliciting diagnostic information and recording responses. It is designed to be administered by lay interviewers, trained to precisely follow the interview schedule. The DISC 2.25C generates DSM-III-R diagnoses (and provisional DSM-IV diagnoses) as well as the number of diagnostic criteria met and symptom counts for discrete diagnostic entities.

The DISC was originally developed under contract from National Institute of Mental Health (Costello, Edelbrock, Kalas, Dulcan & Klaric, 1984), but has undergone a number of revisions aimed at improving its reliability and validity. The most recent work in this regard has been carried out by Shaffer as part of the NIMH sponsored Methodology for the Epidemiology of Children and Adolescents (MECA) Study (Shaffer *et al.*, 1991). A major objective of this latest revision was to lower false positive responses to individual diagnostic criteria by adding qualifying adverbs (e.g. "very", "almost always") or additional rule out questions.

Intervention

Good Behavior Game. The GBG is a classroom team-based behavior management strategy that promotes Good Behavior by rewarding teams that do not exceed maladaptive behavior standards. After baseline measurement of precisely defined behaviors, the teacher assigned each child in the class to one of three heterogeneous teams ensuring that teams contained equally aggressive/disruptive children. Each time a child was disruptive the team was given a point. The goal of the strategy was to encourage students to manage their own and their teammates' behavior through the process of the group reinforcement and through mutual self-interest. Details of the training and the intervention strategy are provided in the *Good Behavior Game Training Manual* (Dolan, Turkkan, Werthamer-Larsson & Kellam, 1989). The eight teachers who participated received 40 hours of training.

While the GBG was in progress in the classroom, the teacher assigned a checkmark on the blackboard next to the name of a team whenever one of its members displayed one of the specified disruptive behaviors. Disruptive behaviors were *verbal disruption*, *physical disruption*, *out-of-seat without permission* and *noncompliance*. A team could win the game if the total number of team checkmarks did not exceed four at the end of the game period. Thus, all teams could win during a particular game period. Initially, children on the winning teams received a tangible reward (stickers, erasers) and later they engaged in a rewarding activity (extra recess, class privileges). In addition, all teams who won the most games during the week were termed the Weekly Winners and received a special reward on Friday.

During the first weeks of the intervention, the GBG was played three times each week for a period of 10 minutes. Over successive weeks, duration increased approximately 10 minutes per game period every three weeks, up to a maximum of three hours, although the checkmark criterion for winning the game remained at four. Initially, game periods were announced and the rewards were delivered immediately after the game. Later, the teacher initiated the game period without announcement and the rewards were delayed until the end of the school day or week. Over time the game was played at different times of the day, during different activities, and even in different locations (such as in the hallway walking to the cafeteria). In this manner, the GBG evolved from a procedure that was highly predictable and visible with a number of immediate reward props, to a procedure with unpredictable occurrences and locations, and with deferred rewards. The original description of the GBG can be found in Barrish, Saunders and Wolfe (1969). A full description of the ML intervention, which is included here as an active control for the GBG, is given in Dolan *et al.* (1993).

Results

The following analyses focus on the results of the 1985–86 cohort of students (Cohort 1) exposed to either the GBG or ML; students in GBG and ML internal control classrooms; and students in external control classrooms. All students had to remain in the same intervention condition over two years, and then remain in the School District for the follow-up period through the six years from pre-intervention into middle school, or in lower grades in the case of those who repeated one or more grades.

The population we report here can be considered representative of the elementary public school children in the eastern part of Baltimore, who remained in their same school for two years and remained in the public schools anywhere in the city for the six years. Most of those who were lost to follow-up had moved out of the system and a small number of others refused to participate. Analyses are now planned to compare the study children to those who moved out of the School District.

A total of 693 students were in the same intervention condition for two years. The sample for the GBG condition was 153 from 8 classrooms; the sample for the GBG internal control condition was 86 from 6 classrooms; and the sample for the external control condition was 157 from 11 classrooms. The sample for the ML condition was 163 from 9 classrooms. There were 134 students from 7 classrooms in the ML internal control condition.

The developmental epidemiology of aggression from fall of first grade to six years later

The course of aggression in the study population for males and females over the six years was examined for each assessment point for the combined internal and external control groups (see Figs 1 and 2). As shown in Fig. 1, the course of aggression among males appears fairly flat, but the variances over time fluctuate, with the greatest individual variation among the children occurring in third grade. Major gender

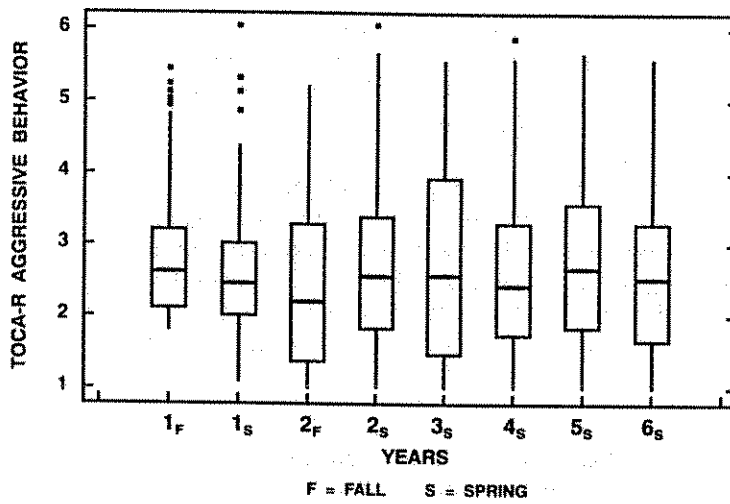


Fig. 1. Developmental epidemiology of male aggressive behavior from early first grade over six years: all control males.

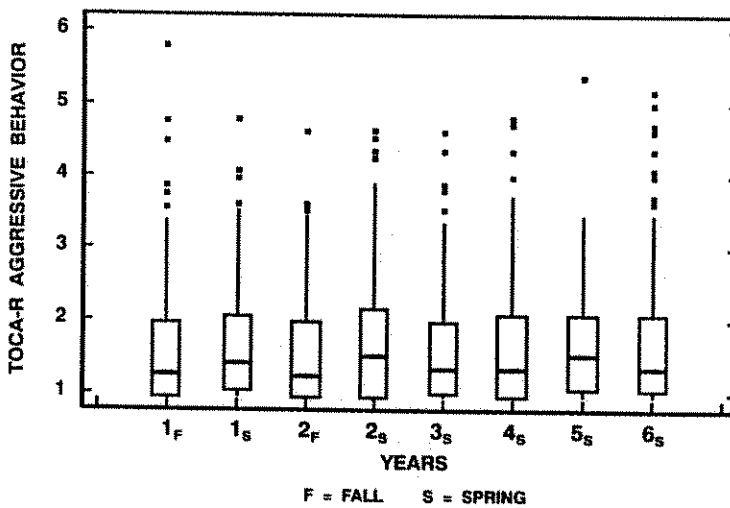


Fig. 2. Developmental epidemiology of female aggressive behavior from early first grade over six years: all control females.

differences can be seen in Figs 1 and 2, with TOCA-R teacher ratings revealing much higher levels of aggression over time for males, with much lower levels of aggression for females throughout the six years. Males also evidence greater variability in aggression over time, although note that there are many highly aggressive outliers among females at every measurement point.

GBG impact on aggression from fall of first grade through six years later

The impact of GBG was tested separately by gender by analyses of covariance (ANCOVAs), first comparing the total male children and then the total female children who received GBG to the GBG and ML internal controls, the external controls, and the ML children. The results for males can be seen in Table 3. No main effects of GBG were found compared to the combined control group, including ML, for the total population of males. There also were no main effects of the GBG for all females.

The subpopulations of male and female children at different levels of aggression in the fall of first grade were tested separately to examine the variation in impact as a function of the level of pre-existing aggression. The results for males are shown in Table 3. There were increasing and significant effects of GBG at sixth grade for the higher levels of aggression at fall of first grade. The GBG had increasing effect as the level of aggression rose in the fall of first grade, but only among males, and only among males at or above the median on aggression in first grade.

Table 3. Analyses of covariance results for Good Behavior Game (GBG) impact in spring of year 6 for the total population of males and males at different levels of baseline aggression

<i>Source</i>	<i>df</i>	<i>SS</i>	<i>F</i>	<i>p</i>
<i>TOCA-R for All Males</i>				
Fall aggressive behaviour—year 1	1	32.76	27.25	.0001
GBG—controls	4	6.68	1.39	.2377
Error	269	323.43		
<i>TOCA-R > 1.7</i>				
Fall aggressive behavior—year 1	1	16.36	12.06	.0007
GBG—controls	4	12.20	2.25	.0673
Error	132	179.03		
<i>TOCA-R > 2.0</i>				
Fall aggressive behavior—year 1	1	11.33	9.40	.0028
GBG—controls	4	12.36	2.56	.0430
Error	98	118.10		
<i>TOCA-R > 2.2</i>				
Fall aggressive behavior—year 1	1	16.83	15.83	.0002
GBG—controls	4	23.53	5.53	.0006
Error	77	81.84		
<i>TOCA-R > 2.4</i>				
Fall aggressive behavior—year 1	1	14.34	13.16	.0005
GBG—controls	4	18.22	4.18	.0044
Error	68	74.07		

GBG impact from fall of first grade through six years later on the more aggressive children

The evolving course of aggression in those males who were at or above the median in first grade in the GBG and the GBG internal control conditions is presented in the form of box plots in Figs 3 and 4. Here the medians and variances are clearly portrayed indicating possibly a slight rise in aggression in spring of third grade.

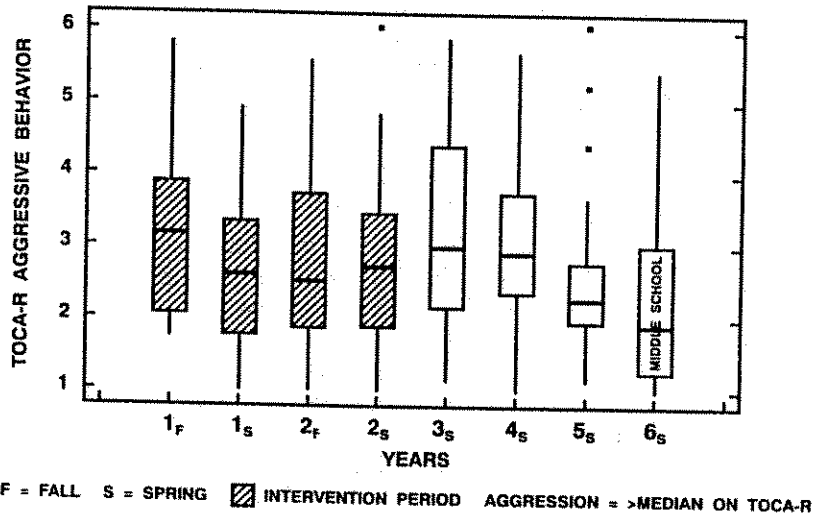


Fig. 3. Course of aggressive males from early first grade into middle school: Good Behavior Game males.

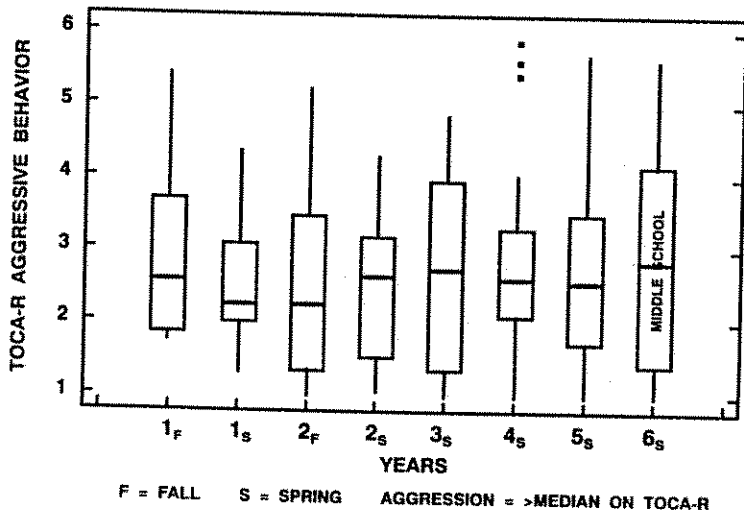


Fig. 4. Course of aggressive males from early first grade into middle school: Good Behavior Game internal control males.

Although there is narrower variation, GBG internal controls generally continue at that level, while the aggression drops after third grade among GBG males steadily from the fourth and fifth grades and into middle school. Data are similar for higher aggressive males in the external control and ML intervention and ML internal control conditions. The drop in aggression for the GBG condition was not seen for females or for males who were below the median in aggression in fall of first grade.

GBG impact on slopes, variances and correlations from fall of first grade to six years later

Here we make use of regression analyses. Shown in Fig. 5 are the scatterplots and regression lines for GBG males and GBG internal control males. The regression lines for GBG and GBG internal controls, overlayed on scatterplots from fall of first grade to spring of year six demonstrate the changing slopes and variances and continuity of aggression as a function of the GBG. Note the markedly reduced variances, slopes and correlations among GBG males compared to the GBG internal control males. The effects of GBG varied as a function of severity of aggression. As can be seen in the lower right part of the plot, almost all males who began with fairly high levels of aggression and who showed marked reduction of aggression were in GBG. The scatterplot shows these children distributed along the X axis. Note also that the GBG does not appear to protect the children who were not aggressive at the start from becoming aggressive. This can be seen in the left side of the scatterplot by the GBG children and control children distributed along the Y axis.

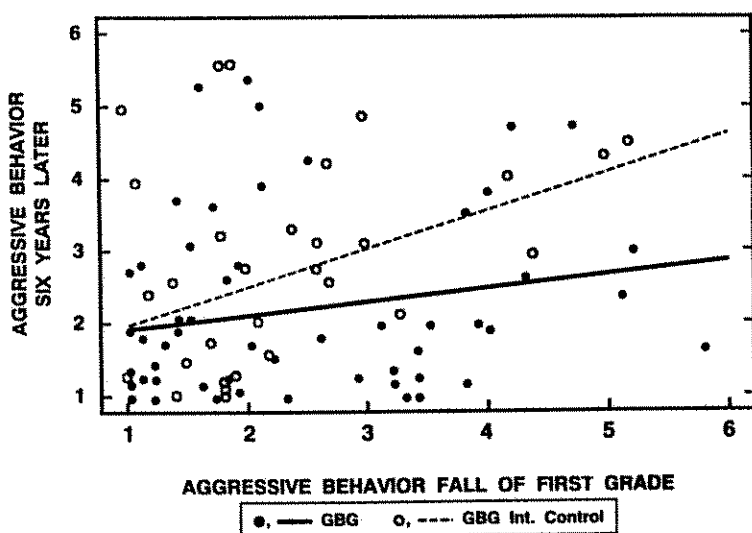


Fig. 5. Impact of Good Behavior Game on aggressive behavior in males from fall of first grade through six years later.

Teachers' aggression ratings and conduct disorder symptoms on the DISC

The different levels of teacher-rated aggression on the TOCA-R by the number of conduct disorder symptoms on the DISC for males and females in the combined control groups at middle school follow-up are shown as boxplots in Figs 6 and 7.

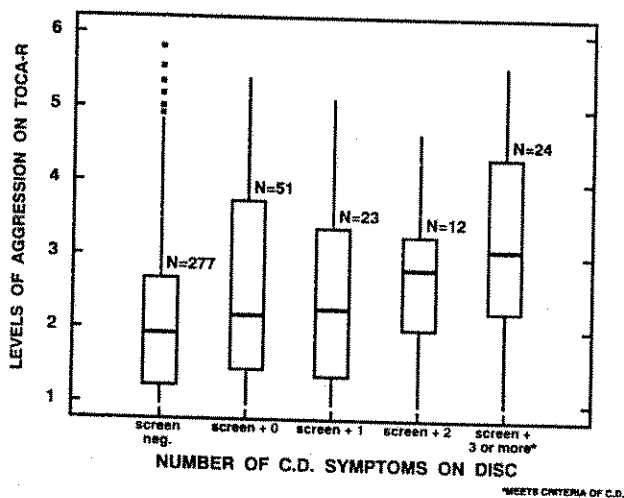


Fig. 6. Teacher's aggression ratings and levels of C.D. symptoms among all control males at middle school follow-up.

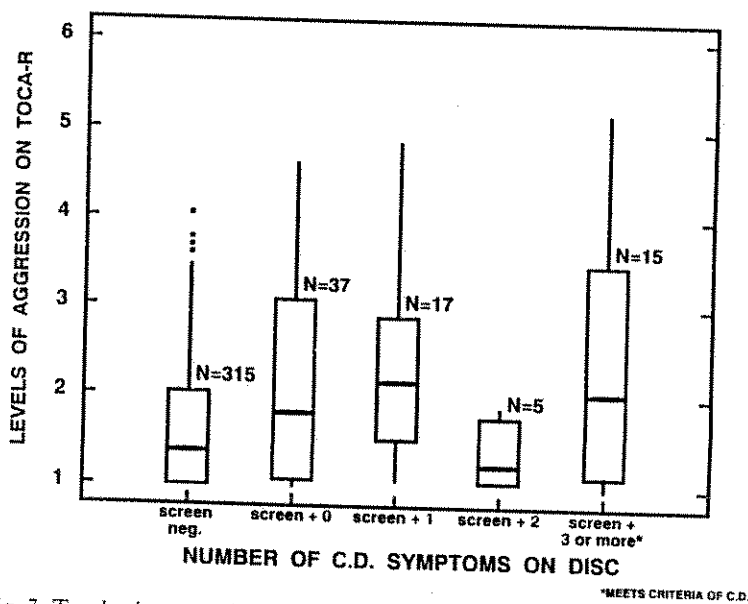


Fig. 7. Teacher's aggression ratings and levels of C.D. symptoms among all control females at middle school follow-up.

For males the medians for TOCA-R aggression show a steady rise as the number of symptoms on the DISC increases. The greatest individual variation in TOCA-R aggression can be seen for males who screened negative and those who screened positive but had no symptoms on the DISC, compared to those who screened positive and had three or more symptoms. Although the levels of teacher-rated aggression are much lower for females than for males, there is a similar increase in aggression as the number of conduct disorder symptoms increases, with the exception of the very few females with two symptoms. The variances in TOCA-R aggression show fluctuation around the median for females, with the greatest variation in the same two groups as males, that is, females who screened negative or those who screened positive with no DISC symptoms and those who screened positive with three or more symptoms. As shown in Figs 6 and 7, the numbers of females and males who screened positive are small, and this limited further analyses of these data until later assessments are completed.

Discussion

The course of aggression among males from early in first grade over the following six years is generally stable with possibly a modest increase around third grade in but are reasonably representative of the total population of males who were stable in their intervention status in first and second grades and were still in Baltimore public schools for the remaining period. The course of aggression for females also was stable but at a much lower level than that of the males. There was no evidence of a rise in aggression in third grade among males. The variance among males fluctuated considerably over time, with more variation around third grade along with the rise in the median. Although third grade appears worth investigation as a period of increased variance, it is interesting to note that there was less treatment effect at that time than in first grade or at middle school. The meaning, if any, of the third grade broader variance and slight rise in the median among males remains a question to be addressed.

These data provide a developmental epidemiological baseline against which the impact of the GBG can be measured (Kellam *et al.*, 1991). Kellam *et al.* (1991) argue that baseline in developmental epidemiologically-based preventive trials should be viewed as the evolving course of a developmental model, with the intervention addressing a hypothesized mediator along the developmental course. In this paper we are focussing on a very simplified model based on the hypothesis that early aggression over the course of elementary school is the path leading to later aggression and to a set of problem outcomes. In subsequent papers we will address hypothesized mediators and moderators of the course of aggression and of the responses of aggressive children to the GBG.

The association of the impact of the GBG with the transition in first grade and later with the transition into middle school supports their importance as times of greater malleability and variation in the developmental paths of the children. Over the course of first grade children are expected to take on the role of student. Among many others the role involves separating to some extent from the parents and engaging with a

new natural rater, one who is legally empowered beyond pre-school to define social task demands and rate the adequacy of the child's performance. The child's tasks include understanding the teacher's task demands and responding adequately in the context of the classroom amidst his or her classmates' diverse responses and often diversionary behavior. Later on, the middle school transition involves the movement from one teacher to multiple teachers, multiple subjects, greater and more complex academic demands, greater demands for self-monitoring and reliance, and greater influence of peers. Each of these transitions presents developmental challenges that draw upon previously developed social adaptational skills and strategies. If development in these areas has been delayed or maladaptive, the evidence of this will be seen in the child's social maladaptive responses to these new developmental challenges.

Children who had been in GBG in first and second grade tended to have benefitted from an intervention that was aimed at the social adaptational process between the child and the natural rater (the teacher), that specified the desirable behaviors precisely, and integrated the teacher, classmates and reinforcers into a coherent integrated social context. More of the former GBG children apparently were successful in their behavioral responses to the new social task demands in middle school, providing evidence that their earlier experience with the GBG left them with a broader armamentarium of behavioral responses to the new transition into middle school.

Caspi and Moffitt (1991) have argued that during times of life course transition, when more innovative coping responses are required, more variation in successful or unsuccessful social adaptation will become apparent. The Baltimore results support this hypothesis and further suggest that during times of transition such as first grade there is important malleability in the child's social adaptational responses. These responses can be influenced through the social adaptational demand/response process with the natural rater and can have lasting impact that may emerge in the next life transition. We see this social adaptational process as the basic interface between the individual and the immediate social field. It is a process that apparently leads to long-lasting impact *for better or worse* on the future SAS of the individual, and possibly on his or her future PWB as well.

The sleeper effect or non-continuous effect of GBG that the data seem to show may be a manifestation of the greater individual variation associated with life transitions. Sleeper effects in general may reflect the tendency for greater individual variation at such times. Using this hypothesis, we can design preventive trials to include life transition times in the follow-up strategy. In addition, periodic assessments of subject populations are needed to assess the course of impact in order to detect not only the non-responders, but those whose response drops off and those who persist in their response. Epidemiologically defining the study population is necessary to reduce selection bias and allow for studying the variation in developmental courses and responses to the social adaptational demands of the new stage of life.

We hypothesized that the children who participated in the GBG who were more aggressive in fall of first grade would exhibit a reduction in aggressive behavior from early elementary school through middle school. This hypothesis appears to have been confirmed; but we also hypothesized that GBG participants who were not aggressive in fall of first grade would exhibit a lower incidence of aggressive behavior over the course of the six years than non-aggressive children in either the GBG or ML internal

control classrooms or in the active control ML classrooms. This did not occur. Boys who were not aggressive in first grade were not protected by GBG from demonstrating later aggressive behavior. The boys who improved had exhibited aggressive behavior in order to benefit. The gender effects also remain unclarified as to their determinants, other than the fact that girls were not nearly as aggressive as boys early in school.

Higher severity of aggressive behavior was not indicative of non-responsiveness to intervention. The question remains as to the relevance of the results to conduct disorder as a diagnosis. We examined the relationship of TOCA-R to conduct disorder and found a clear link. Children who met criteria for conduct disorder had much higher ratings on TOCA-R than those who did not; and the relationship was linear particularly in the case of males, with increasing levels of aggression on TOCA-R as the number of conduct disorder symptoms increased. We tried to examine the distribution of conduct disorder in relation to the GBG and the various control conditions. The number of cases was very small; the prevalence for conduct disorder was about 6%. Since most of the children are only 12–13 years old, the data do not include those cases not yet diagnosable.

The scatterplot from the regression model suggests that the more aggressive first graders were influenced by GBG. It is worth noting that the developmental epidemiological studies done on the long-term prediction of early aggression to later anti-social and drug problems indicate that teacher ratings of aggressive behavior at the moderate to severe levels are very important predictors (see for example, Anthony, 1985; Ensminger *et al.*, 1983; Kellam *et al.*, 1983; Tomas *et al.*, 1990). This level of aggression involved 20% or more of Woodlawn first graders. These arguments are made to suggest that the levels of aggression apparently affected by GBG are not trivial and have potential importance to public mental health, heavy drug use, delinquency, criminality and prevention.

Diagnostic criteria in the DSM nosology entail cutpoints that are often arbitrary in behavioral or mental disorders. Valid tests of the cutpoint should include the nature of its relationships to its antecedents and to its consequences. In the case of conduct disorder, the cutpoint may need to be more inclusive of children with a history and course of aggressive behavior with seriously increased risk of the problem outcomes already noted.

In this paper we have used box plots to show the entire variation with quartiles marking the more or less severe. This method allows flexibility for defining caseness or disorder. The quartiles allow the study of where and how the risk is distributed in criminal, violent, self-destructive or disordered conditions. These data easily lend themselves to the study of antecedents and consequences. We propose that this way of conceptualizing incidence and prevalence—that is, keeping the actual variance in view—is appropriate for many conditions that are not like infectious diseases, including certain physiological and psychological ones, in addition to conduct disorder.

The box plots do not inform us about the continuity of individual children. Rather, they tell us about the population or subpopulations of the children. Individual growth curves are needed to reveal whether the same or different children are in the top quartile of aggression from one time to the next, for example. This next step in the analyses is currently being developed in parallel with studies of growth curve models comparing

subgroups of children defined by hypothesized mediators and moderators (see Bryk & Raudenbush, 1987; Raudenbush & Bryk, 1988; Raudenbush & Chan, 1991).

The box plot representing the last assessment period clearly indicates children who were improved by the GBG in the bottom quartile and with those who were not helped by the GBG in the top quartile. Those who did not respond well to GBG can be compared to those who did by including in the analyses family, peer group, classroom, school and community characteristics. Such analyses will inform theory regarding developmental malleability and the design of the next set of preventive trials and are next steps in the Baltimore analyses (Kellam & Rebok, 1992).

The search for good responses in comparison to non-responses or poor responses is a centrally important function of preventive trials and is at the core of the potential such trials offer for theory-building. Beyond the child's classroom SAS are his/her SAS in other social fields and the characteristics of the fields themselves. The growing importance of the intimate social field and sexuality, along with the increasing salience of SAS in the peer group, are potential moderators of response to GBG (Caspi & Moffitt, 1991). Family processes such as the adequacy of behavior management and monitoring also are important to model (Patterson, Reid & Dishion, 1991). Ecological measures of neighborhood are another part of the database to be assessed. Critical moderators could be the GBG team itself, the teacher and his/her quality of implementation, and school and classroom effects (Kellam, 1990; Rutter *et al.*, 1970).

The data on the impact of GBG are fairly compelling in regard to teacher ratings of aggression, but three questions remain: First, are these results replicable? Second, is there generalization to other social fields than the school/classroom? And third, are the children most severely in need responding to the GBG or do the poor or non-responders require a back-up intervention, and is a third level of intervention beyond the second needed for those requiring treatment?

To answer the first question, the second cohort is a direct replicate of the first and will be used to test developmental models from the first cohort. The second question about generalization will be addressed in the next stages of analysis, with peer and family data, police records and longer term follow-up.

The third question on intervention can be answered now without further research. The GBG is the first level of a three-tiered system in our public mental health view. The aggressive children who do not respond well are in need of more intervention, and they are in the upper whisker of the GBG box plot. They may not be merely those with the most severe aggressive ratings in first grade, nor merely the ones at the very top of the upper whisker. They may be those with various kinds of comorbidity such as attentional problems or depression who did not respond well to the GBG. They may be those whose families do not monitor or reinforce prosocial behavior. The next stage of research at the universal level will address the impact of combining the achievement intervention with the GBG, and will investigate the efficacy of a family intervention for all children involving strengthening family behavior management and the learning environment at home, along with closer parent/teacher partnership.

The three-tiered system we propose puts emphasis on the first, universal level and is experienced by all children. The universal level is a core arena for prevention

research. It is most ecologically integrated within the socialization structures and processes; and it holds promise of taking advantage of malleability of developmental paths and social adaptational processes. It is less mental health professional, more economical, and entails less labeling. The second level should be immediately behind the first and available for children not responding well to the universal level. The third level entails more mental health or other professional services and should be available for the poor or non-responders from the first two levels. This three-level system would represent a public mental health response system—one that emphasizes health and socialization rather than repair. The evidence from the GBG universal intervention is that this economical, theory-based intervention brought about improvement in the aggressive level of the more aggressive males. Equally important, it differentiated the good responders from the poor or non-responders, thus identifying those children who need the second or third levels.

The great benefit of this stage of prevention research is that it brings the assessment of the population's needs together with the needs of the developmental sciences and the creation of public mental systems to the benefit of all three enterprises. This research results from integrating three scientific perspectives: life-course development, community epidemiology and preventive intervention trials. This combination of perspectives is highly promising for the next stage of research on development, variation in development among people, and the possibility of improving their destiny.

Acknowledgements—The contributions of the City of Baltimore, its families and children, and the administration of the Baltimore City Public Schools have been and continue to be enormous. The Prevention Program is a collaboration between the Baltimore City Public Schools and the Prevention Research Center of the Department of Mental Hygiene, The Johns Hopkins University School of Hygiene and Public Health. This work of the Prevention Program would not have been possible without the participation and support of the leadership, faculty and staff of the School District. Our work has been based on the search for mutual shared interests in the assessments and the interventions, and has been carried out under the aegis of the Board of School Commissioners and the Superintendent.

Dr Walter Amprey, Superintendent of Baltimore City Public Schools, Deputy Superintendent Dr Lillian Gonzales, and Director Dr Juanita Lewis have overseen and guided the current work on follow-up and the new work on the next stages of intervention development and evaluation. Originally it was Ms Alice Pinderhughes, former Superintendent of the Baltimore City Schools who worked out with the Prevention Research Center the collaborative arrangements. The assistance and support of the following have been vital to our program: Dr Leonard Wheeler, former Assistant Chief Superintendent for Elementary Schools who was for many years our major coordinator and School District liaison. Dr Carla Ford, former Supervisor, Office of Early Childhood Education, made continuing contributions to the intervention design and implementation monitoring.

The original search for a behavioral intervention to address the social adaptational process between teacher and children was led by Dr Alan Harris, and the first implementor and supervisor was Dr Jaylin Turkkan. Dr Lawrence Dolan was overall field supervisor the first two years, followed by Dr Lisa Werthamer-Larsson, who is on the Department of Mental Hygiene faculty and continues to play a key leadership role in the Prevention Research Center. Dr James C. Anthony is P.I. of the drug abuse aspects of the Center's work and a leader in the original design of the two intervention trials. Dr C. Hendricks Brown played a key biostatistical and methodological role in this work and continues to consult with us as well as to analyze the direct observation data.

The leadership of Ms Elva Edwards and her staff in community base-building and crisis back-up are essential aspects of the Center's work and are gratefully acknowledged. Our editor, Ms Fionnuala Regan and our manuscript producer Ms Alice Brogden did their usual fine job in spite of the difficulties of scientific jargon and the complex styles of the authors.

The studies on which this paper is based have been supported by the following grants, with supplements from the National Institute on Drug Abuse: National Institute of Mental Health, NIMH Grant Number P50 MH38725, Epidemiologic Prevention Center for Early Risk Behavior; NIMH Grant Number 1R01 MH42968, Periodic Outcome of Two Preventive Trials; and NIMH Grant Number 1R01 MH40859, Statistical Methods for Mental Health Preventive Trials; and for the follow-up NIDA Grant DA-00787.

References

- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd Edn, rev.). Washington, DC: Author.
- Anthony, J. C. (1985). Young adult marijuana use in relation to antecedent childhood misbehaviors. In L. S. Harris (Ed.), *Problems of drug dependence, 1984* (pp. 238-244). Rockville, MD: NIDA Research Monograph.
- Barrish, H. H., Saunders, M. & Wolfe, M. D. (1969). Good behavior game. Effects of individual contingencies for group consequences and disruptive behavior in a classroom. *Journal of Applied Behavioral Analysis*, *2*, 119-124.
- Block, J., Block, J. H. & Keyes, S. (1988). Longitudinally foretelling drug usage in adolescence: early childhood personality and environmental precursors. *Child Development*, *59*, 336-355.
- Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147-158.
- Caspi, A. & Moffitt, T. E. (1991). Individual differences are accentuated during periods of social change: the sample case of girls at puberty. *Journal of Personality and Social Psychology*, *61*, 157-168.
- Caspi, A. & Moffitt, T. E. (in press). Continuity amidst change: a paradoxical theory of personality coherence. *Psychological Inquiry*.
- Costello, A. J., Edelbrock, C. S., Kalas, R., Dulcan, M. K. & Klaric, S. H. (1984). *Development and testing of the NIMH Diagnostic Interview Schedules for Children in a clinical population* (Contract No. RFP-81-0027). Rockville, MD: Center for Epidemiological Studies, National Institute of Mental Health.
- Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larsson, L., Rebok, G. W., Mayer, L. S., Laudoff, J., Turkkan, J., Ford, C. & Wheeler, L. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology*, *14*, 317-345.
- Dolan, L., Turkkan, J., Werthamer-Larsson, L. & Kellam, S. (1989). *The good behavior game training manual*. Baltimore, MD: The Johns Hopkins Prevention Research Center.
- Ensminger, M. E., Kellam, S. G. & Rubin, B. R. (1983). School and family origins of delinquency: comparisons by sex. In K. T. Van Dusen & S. A. Mednick (Eds), *Prospective studies of crime and delinquency* (pp. 73-97). Boston: Kluwer-Nijhoff.
- Eron, L. D. (1990). Understanding aggression. *Bulletin of the International Society for Research on Aggression*, *12*, 5-9.
- Farrington, D. P., Gallagher, B., Morley, L., St Ledger, R. J. & West, D. J. (1988). Are there successful men from criminogenic backgrounds? *Psychiatry*, *51*, 116-130.
- Farrington, D. P. & Hawkins, J. D. (1991). Predicting participation, early onset and later persistence in officially recorded offending. *Criminal Behavior and Mental Health*, *1*, 1-33.
- Greenley, J. R. & Mechanic, D. (1976). Social selection in seeking help for psychological problems. *Journal of Health and Social Behavior*, *17*, 249-262.
- Greenley, J. R., Mechanic, D. & Cleary, P. (1987). Seeking help for psychological problems. A replication and an extension. *Medical Care*, *25*, 1113-1128.
- Hawkins, J. D., Catalano, R. F., Morrison, D. M., O'Donnell, J., Abbott, R. D. & Day, L. E. (1992). The Seattle Social Development Project: effects of the first four years on protective factors and problem behaviors. In J. McCord & R. E. Tremblay (Eds), *Preventing antisocial behavior: interventions from birth through adolescence* (pp. 139-161). New York: Guilford Press.
- Hawkins, J. D., Von Cleve, E. & Catalano, R. F. (1991). Reducing early childhood aggression: Results of a primary prevention program. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 208-217.

- Kellam, S. G. (1990). Developmental epidemiologic framework for family research on depression and aggression. In G. R. Patterson (Ed.), *Depression and aggression in family interaction* (pp. 11-48). Hillsdale, NJ: Lawrence Erlbaum.
- Kellam, S. G. (1991). Developmental epidemiological and prevention research on early risk behaviors. In L. P. Lipsitt & L. L. Mitnick (Eds), *Self-regulatory behavior and risk taking: causes and consequences* (pp. 51-70). Norwood, NJ: Ablex.
- Kellam, S. G., Branch, J. D., Agrawal, K. C. & Ensminger, M. E. (1975). *Mental health and going to school: the Woodlawn program of assessment, early intervention, and evaluation*. Chicago: University of Chicago Press.
- Kellam, S. G., Brown, C. H., Rubin, B. R. & Ensminger, M. E. (1983). Paths leading to teenage psychiatric symptoms and substance use: developmental epidemiological studies in Woodlawn. In S. B. Guze, F. J. Earls & J. E. Barrett (Eds), *Childhood psychopathology and development* (pp. 17-55). New York: Raven Press.
- Kellam, S. G. & Ensminger, M. E. (1980). Theory and method in child psychiatric epidemiology. In F. Earls (Ed.), *Studies of children* (pp. 145-180). New York: Prodist.
- Kellam, S. G. & Rebok, G. W. (1992). Building developmental and etiological theory through epidemiologically based preventive intervention trials. In J. McCord & R. E. Tremblay (Eds), *Preventing antisocial behavior: interventions from birth through adolescence* (pp. 162-195). New York: Guilford Press.
- Kellam, S. G., Mayer, L. S., Rebok, G. W. & Hawkins, W. E. (1993). The effect of improving achievement on aggressive behavior and of improving aggressive behavior on achievement through two preventive interventions: an investigation of etiological roles. Manuscript submitted for publication.
- Kellam, S. G., Simon, M. B. & Ensminger, M. E. (1983). Antecedents in first grade of teenage substance use and psychological well-being: a ten-year community-wide prospective study. In D. F. Ricks & B. S. Dohrenwend (Eds), *Origins of psychopathology* (pp. 17-42). Cambridge: Cambridge University Press.
- Kellam, S. G. & Werthamer-Larsson, L. (1986). Developmental epidemiology: a basis for prevention. In M. Kessler & S. E. Goldston (Eds), *A decade of progress in primary prevention* (pp. 154-180). Hanover, NH: University Press of New England.
- Kellam, S. G., Werthamer-Larsson, L., Dolan, L. J., Brown, C. H., Mayer, L. S., Rebok, G. W., Anthony, J. C., Laudolff, J., Edelsohn, G. & Wheeler, L. (1991). Developmental epidemiologically-based preventive trials: baseline modeling of early target behaviors and depressive symptoms. *American Journal of Community Psychology*, 19, 563-584.
- Kolvin, I., Miller, F. J. W., Fleeting, M. & Kolvin, P. A. (1988). Social and parenting factors affecting criminal-offence rates: findings from the Newcastle Thousand Family Study (1947-1980). *British Journal of Psychiatry*, 152, 80-90.
- McCord, J. (1988). Parental behavior in the cycle of aggression. *Psychiatry*, 51, 14-23.
- Morris, J. N. (1975). *Uses of epidemiology* (3rd Edn). New York: Churchill Livingstone.
- Patterson, G. R., Reid, J. & Dishion, T. (1991). *Antisocial boys*. Eugene, OR: Castalia.
- Raudenbush, S. W. & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and teachers on student learning. *Review of Research in Education*, 16, 423-475.
- Raudenbush, S. W. & Chan, W. S. (1991). *Growth curve analysis in accelerated longitudinal designs with application to the National Youth Survey*. Unpublished manuscript.
- Robins, L. N. (1978). Sturdy childhood predictors of adults antisocial behavior: replications from longitudinal studies. *Psychological Medicine*, 8, 611-622.
- Rutter, M. L., Tizard, J. & Whitmore, K. (Eds) (1970). *Education, health and behavior: psychological and medical study of childhood development*. New York: Wiley.
- Shaffer, D., Fisher, P., Piacentini, J., Schwab-Stone, M. & Wicks, J. (1991). *Diagnostic Interview Schedule for Children (DISC 2.25)*. New York State Psychiatric Institute.
- Schaffer, D., Stokman, C., O'Connor, P. A., Shafer, S., Barmack, J. E., Hess, S. & Spaulden, D. (1979, November). *Early soft neurological signs and later psychopathological development*. Paper presented at the meeting of the Society for Life History Research in Psychopathology and Society for the Study of Social Biology, New York.
- Shedler, J. & Block, J. (1990). Adolescent drug use and psychological health: a longitudinal inquiry. *American Psychologist*, 45, 612-630.

- Tomas, J. M., Vlahov, D. & Anthony, J. C. (1990). Association between intravenous drug use and early misbehavior. *Drug and Alcohol Dependence*, *25*, 79-89.
- Tremblay, R. E., Masse, B., Perron, D., Leblanc, M., Schwartzmann, A. E. & Ledingham, J. E. (1992). Early disruptive behavior, poor school achievement, delinquent behavior, and delinquent personality: longitudinal analyses. *Journal of Consulting and Clinical Psychology*, *60*, 64-72.
- Tremblay, R. E., McCord, J., Boileau, H., Charlesbois, P., Gagnon, C., LeBlanc, M. & Larivee, S. (1991). Can disruptive boys be helped to become competent? *Psychiatry*, *54*, 148-161.
- Tremblay, R. E., Vitaro, F., Bertrand, L., LeBlanc, M., Beauchesne, H., Boileau, H. & David, L. (1992). Parent and child training to prevent early onset of delinquency: the Montreal longitudinal-experimental study. In J. McCord & P. E. Tremblay (Eds), *Preventing antisocial behavior: interventions from birth through adolescence* (pp. 117-138). New York: Guilford Press.
- Werthamer-Larsson, L., Kellam, S. G. & Wheeler, L. (1991). Effect of first-grade classroom environment on child shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology*, *19*, 585-602.
- White, J. L., Moffitt, T. E. & Silva, P. A. (1989). A prospective replication of the protective effects of IQ in subjects at high risk for juvenile delinquency. *Journal of Consulting and Clinical Psychology*, *57*, 719-724.

