Supplemental Information

CDP is a longitudinal study of children's development using a sample from 3 geographic areas: Knoxville, TN; Nashville, TN; and Bloomington, IN.[19] Parents were approached at random during kindergarten registration in 1987 and 1988 and asked to participate in a longitudinal study of child development; ~75% of those approached agreed to participate (n = 585). Kindergarteners were selected to represent a range of socioeconomic backgrounds in each geographic area.

Adulthood self-reports were available from 461 of the original participants (7 of the original CDP participants died before the adult interview time frame). Adulthood court records were available for 573 original participants, and adulthood peer reports were available for 371. Participants who had no adulthood self-report data available did not differ from the original participants on cohort, single parenthood, family stress, or child health measured at year 1 but did differ in sex, race and ethnicity, site, early abuse, SES, and mother's education at year 1.

FT is an intervention designed to decrease conduct problems for children at risk for aggression in kindergarten.[20] Fifty-five at-risk elementary schools were selected on the basis of neighborhood crime and poverty rates in Durham, NC; Nashville, TN; rural Pennsylvania; and Seattle, WA. Within each site, schools were matched on demographics, and one of each pair was randomly assigned to

intervention and the other to control. In addition to the intervention and control sample, FT includes a normative or community sample selected to represent the entire distribution of risk among kindergarteners in the control schools. In 1991, all kindergarteners' study schools (28 intervention and 27 control schools) were screened by teachers for conduct problems by using the Teacher Observation of Classroom Adaptation-Revised (TOCA-R) Authority Acceptance Scale.[36] Within each site, a stratified sample of 100 children from control schools was recruited on the basis of race and ethnicity, sex, and decile of the TOCA-R score to create the normative sample. Parents were asked to participate in person and over the phone (n = 387). In 1991, cohort 1 of the intervention and control sample for the FT intervention was also recruited by using a 2-stage screening process. In the first stage, researchers evaluated behavior problems at school using the TOCA-R scores. Parents of children scoring in the top 40% within site were solicited over the phone and in person to complete a 22-item instrument based on the Child Behavior Checklist[29] to capture behavior problems at home. Standardized teacher and parent scores were combined to create a severity-of-risk screen score. Children within site were rank ordered on the basis of this screen score, and study children were recruited in person and over the phone starting with the highest risk

until designated sample sizes were reached within site and condition. Ninety-one percent of recruited families consented, yielding a control sample of 155. By chance, 79 of these children were also recruited as members of the normative sample. This normative sample (ranging from low to high aggression) and the control participants (who were rated by teachers as being high in aggression but not randomized into the intervention group) from cohort 1 (n = 463) are used for this study.

Adulthood self-reports were available from 391 of the original participants (7 of the original FT participants died before the adult interview time frame). Adulthood court records were available for 395 original participants, and adulthood peer reports were available for 300. Participants who had no adulthood self-report data available did not differ from the original participants on race and ethnicity, single parenthood, early abuse, SES, family stress, child health, or mother's education in year 1 but did differ in sex and site.

Analyses are completed by using probability weights constructed as the inverse of the probability of being selected into the sample. The CDP sample was randomly selected from the population of kindergarteners enrolled in study schools at each site. Therefore, the within-cohort site probability weights equal the inverse of the sampling fraction (cohort-site sample size and cohort-site kindergarten population). For FT,

children at high risk for conduct problems are overrepresented in the sample because the high-risk control children are added to the normative community sample. Within sites, the population of kindergarteners from study schools is divided into 57 risk groups on the basis of teacher and parent screening scores capturing risk of conduct problems. Within each risk group, the probability weights are calculated as the inverse of the sampling fraction (risk-group sample size and risk-group kindergarten population). CDP and FT together yield a sample of 1048 participants.