

A Genetically Informed Study of the Association Between Harsh Punishment and Offspring Behavioral Problems

Stacy K. Lynch, Eric Turkheimer,
Brian M. D'Onofrio, Jane Mendle, and
Robert E. Emery
University of Virginia

Wendy S. Slutske
University of Missouri—Columbia

Nicholas G. Martin
Queensland Institute of Medical Research

Conclusions about the effects of harsh parenting on children have been limited by research designs that cannot control for genetic or shared environmental confounds. The present study used a sample of children of twins and a hierarchical linear modeling statistical approach to analyze the consequences of varying levels of punishment while controlling for many confounding influences. The sample of 887 twin pairs and 2,554 children came from the Australian Twin Registry. Although corporal punishment per se did not have significant associations with negative childhood outcomes, harsher forms of physical punishment did appear to have specific and significant effects. The observed association between harsh physical punishment and negative outcomes in children survived a relatively rigorous test of its causal status, thereby increasing the authors' conviction that harsh physical punishment is a serious risk factor for children.

Keywords: behavior genetics, children of twins, harsh punishment, corporal punishment, offspring outcomes

Because human children cannot be randomly assigned to rearing environments, research on the effects of parenting on childhood outcomes is complicated by a host of uncontrollable environmental and genetic factors (D'Onofrio et al., 2003). Traditional studies of punishment can usually do little more than observe correlations between variation in parenting styles and variation in child outcomes. Although numerous studies have identified associations between harsh punishment and negative outcomes in children (Emery & Laumann-Billings, 1998; Gershoff, 2002b; Parke, 1979; Smith & Farrington, 2004; Straus, 1994), it is unclear whether these associations are due to the punishment techniques of the parents, coexisting environmental factors, or genetic transmission of parental characteristics correlated with punishment style (Kazdin & Benjet, 2003).

Although there is general agreement regarding the asso-

ciation between harsh or abusive parenting and negative behavioral outcomes in offspring, most notably aggression, (Engeland, Jacobvitz, & Sroufe, 1988; McCord, 1979; Rutter, Quinton, & Liddle, 1983; Smith & Farrington, 2004), the consequences of nonabusive corporal punishment are more controversial. At least three recent literature reviews highlighted the lack of consensus regarding the use of corporal punishment (Gershoff, 2002a; Kazdin & Benjet, 2003; Larzelere, 2000). According to Gershoff's (2002b) meta-analysis, corporal punishment is associated with increased levels of aggression and higher rates of conduct disorder, lower levels of moral internalization, and poorer overall mental health.

Numerous methodological shortcomings of the punishment research have been raised in the literature. One problem pertains to the widespread variation in definitions of corporal or abusive punishment. A second is the ever-present difficulty in acquiring valid and reliable measures of punishment styles and outcome variables. A final issue involves uncontrolled genetic and environmental factors that can confound observed associations between parenting and child outcome. Each of these three themes will be reviewed below.

Inconsistent definitions of corporal punishment have contributed to contradictory findings (Holden, 2002). It may be that when narrowly defined, corporal punishment is not as strong a predictor of externalizing or internalizing behaviors in offspring as has been found in previous studies. For

Stacy K. Lynch, Eric Turkheimer, Brian M. D'Onofrio, Jane Mendle, and Robert E. Emery, Department of Psychology, University of Virginia; Wendy S. Slutske, Department of Psychological Sciences, University of Missouri—Columbia; Nicholas G. Martin, Genetic Epidemiology Group, Division of Population and Clinical Sciences, Queensland Institute of Medical Research, Brisbane, Queensland, Australia.

Correspondence concerning this article should be addressed to Stacy K. Lynch, Department of Psychology, University of Virginia, Box 400400, Charlottesville, VA 22902. E-mail: ak18t@virginia.edu

example, spanking may be unrelated to children's behavior problems when parents who routinely exercise harsher or even abusive forms of punishment are distinguished from those whose punishment style is limited to controlled spanking. Straus (2000) suggested that an appropriate and widely accepted definition of corporal punishment is the use of "physical force with the intention of causing the child to experience pain, but not injury, for the purpose of the correction of the child's behavior," and usually entails "spanking on the buttocks or slapping a child's hand for touching a forbidden or dangerous object" (Straus, 2000, p. 1110). Conflating corporal punishment with harsher forms of punishment or physical abuse makes it impossible to assess differences in the consequences of the two practices.

The second methodological challenge in parenting research involves the measurement of parental punishment style and outcomes in children. Information regarding the punishment techniques used by parents and negative outcomes experienced by children has usually been gathered from contemporary reports from parents or retrospective reports from children. The limitations of self-report in general, and retrospective self-report in particular, have been widely documented (Kendler et al., 2000). A recent review, however (Hardt & Rutter, 2004), suggested that retrospective self-report data may nevertheless be useful. Although Hardt and Rutter (2004) found that retrospective reports of abuse and neglect were often underestimated and biased in adulthood, the bias was less severe when measuring "serious, readily operationalized, adverse experiences in childhood" (p. 270). Gerlsma, Kramer, Scholing, and Emmelkamp (1994) found that errors of omission were often related to the emotional status of the reporter and that false positive reports were less frequent. Brewin, Andrews, and Gotlib (1993) demonstrated that, although there are certain limits to retrospective reports, claims regarding their unreliability are overestimated.

Most important, associations between punishment style and childhood outcome may be confounded by uncontrolled parental characteristics associated with punishment behavior that are transmitted to the offspring (Kazdin & Benjet, 2003). For example, if parents using harsh punishment are more aggressive as a matter of personality, and if aggressiveness is influenced by genetic factors (Rowe, 1994; Smith & Farrington, 2004), the children of these aggressive parents may be more aggressive in the absence of a direct causal effect of punishment. Confounds can also be environmental. If economically impoverished parents are more likely to use physical punishment, and their children are then more likely to abuse drugs and alcohol, this can likewise induce an association between punishment and child outcomes even if there is no causal link between them.

The current study uses a children of twins design that addresses some of these methodological concerns. Although twin models have sometimes been used to quantify genetic and environmental contributions to parenting behavior in adults or behavioral difficulties in children, such models are ill suited to the evaluation of the consequences of punishment style, however, because twins are usually concordant for parenting variables (D'Onofrio et al., 2003). Punishment

style is especially problematic, because parents tend to discipline children within the same family in the same manner.

The children of twins design circumvents some of these methodological difficulties (D'Onofrio et al., 2003; Gottesman & Bertelsen, 1989; Jacob et al., 2003; Mendle et al., 2004). Because the design is based on the twin status of parents rather than children, it is possible to parse variation in punishment into genetic and environmental sources and estimate separately their consequences for offspring outcomes. In the paradigmatic case, if one member of an identical twin pair uses harsh physical punishment and the other does not, a comparison of outcomes in their children provides a strong test of the putative causal relationship, because in a discordant monozygotic twin comparison most genetic and shared environmental confounding variables are controlled implicitly.

Although we are most concerned with using the children of twins design to control for possible confounding third variables, the current study also attempts to address some of the other methodological concerns we have mentioned. With regard to inconsistent and overly broad operational definitions of punishment, we will attempt to clearly define various disciplinary tactics such that corporal punishment is separated from other more severe forms of physical punishment. Although we will be able to address the question of measurement reliability by using average scores of multiple offspring to assign punishment scores to parents, measurement validity remains problematic, because we are constrained to assess harsh punishment with a retrospective item. Whenever retrospective measures from the same individual are used in both the predictor and outcome variables, findings should be viewed with some caution. Our aim is to understand whether children who reported more harsh punishment than their siblings also reported more problem behaviors and whether this relationship is consistent between and within twin families.

Method

Participants

The participating twins and their offspring came from a volunteer national twin panel registry in Australia, which was developed with support from the Australian National Health and Medical Research Council. The twins were born between 1918 and 1961. Participants were initially invited to participate in the study between 1980 and 1982. Although there are no ethnic minorities in the Australian Twin Registry, the sample's demographics largely reflect the non-Aboriginal population cohort of Australia at that time. A number of statistical tests for self-selection bias have been performed and have found no significant risks for abnormal behavior (Heath et al., 1997; Slutske et al., 1997). A questionnaire was mailed to the twins that requested such information as their smoking history, drug and alcohol history, sociodemographic variables, and other risk factors. The questionnaires were returned by 7,616 twin pairs and 567 singleton twins (Jardine & Martin, 1984). An 8-year follow-up survey was conducted in 1989. This survey included repeat assessments of the twins' alcohol use and smoking history of those twin pairs who responded to the 1981 survey (The Alcohol Cohort I Study). The response rate for this wave was

84.5% for the female twins and 80.5% for the male twins ($N = 6,327$; Heath & Martin, 1994). The third survey was a telephone interview using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA; Bucholz et al., 1994) conducted in 1992–1993 with a response rate of 86% ($N = 5,889$; Heath et al., 1997). The SSAGA measures alcohol abuse, Axis I disorders, childhood conduct disorder, and related diagnoses.

In 1999, the children of the twins began to participate in the study. The families from which the offspring were drawn were selected on the basis of either one or both of the twin parents reporting any of the following: divorce, alcohol dependence, conduct disorder, or depression; other families not meeting these criteria were selected as control subjects. These criteria were selected on the basis of the initial studies performed on the data. Of the 2,554 offspring, roughly half came from nuclear families in which the twin parent did not endorse a history of alcohol dependence, conduct disorder, depression, or divorce. Approximately a quarter of the offspring came from twin families in which neither twin reported a history of the disorders or divorce.

The offspring were administered a telephone interview assessing various aspects of behavior, personality, school performance, delinquency, and *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 1994) diagnostic items for oppositional defiant disorder, attention-deficit/hyperactivity disorder, and conduct disorder. Additional items addressed lifetime cigarette use, regular smoking, lifetime history of alcohol use, regular alcohol use, *DSM-IV* diagnostic criteria items for alcohol dependence and abuse, and illicit drug use. Items also included *DSM-IV* criteria for a major depressive episode, suicidal ideation, planning for a suicide attempt, suicide attempt, and self-injury. Items elicited information regarding the children's recollections of various aspects of their parents' style of discipline and other characteristics of their relationships with their parents. One item specifically asked what mode of punishment parents regularly used during childhood. At the time of this study, data were available on 2,554 children (born between 1960 and 1986) born to 889 twin pairs—431 monozygotic and 458 dizygotic; the number of children per family ranged from 0 to 8. Of the twin offspring, 51% were female and 49% were male. The mean age at the time of the interview was 25.1 years old, with a standard deviation of 5.6 years.

Measure of Punishment

The punishment item used in these analyses was as follows: "What was the usual way in which your [mother/father] punished or disciplined you?" The response choices were (a) nonphysical mild (scold, isolate, fine, remove privileges), (b) nonphysical harsh (lock in closet, deprive of food), (c) physical mild (slap, spank), (d) physical harsh (use weapon, punch, kick), and (e) did not punish. If a respondent answered with Choice 3, it was coded as corporal punishment. If a respondent answered with Choice 4, it was coded as harsh physical punishment. If the mother was the twin, then the maternal punishment item was considered the parental punishment item for the analyses. Likewise, if the father was the twin parent, then paternal punishment was considered parental punishment. We included the gender of the parent among other covariates in the analyses. A 1-year follow-up of 176 offspring resulted in kappa coefficients of .63 for the maternal punishment variable and .70 for the paternal punishment variable.

A punishment factor was created by using four items from the interview that addressed parental punishment. The factor had high internal consistency ($\alpha = .89$), and the aforementioned punishment item had a factor loading score of .75. The other three items

were as follows: "When you did something wrong, were you often smacked by your parent?" (.64); "Were you often punched or hit by your parent?" (.61); and "How often did your parent physically punish you so hard that it hurt the next day?" (.60).

Measures of Offspring Outcomes

An exploratory factor analysis (EFA) of dichotomous *DSM-IV* diagnostic items from the offspring SSAGA items was performed in Mplus, which permits consideration of the categorical nature of the items (Muthén & Muthén, 1998). The EFA resulted in a three-factor solution with interfactor correlations between .39 and .49. Each factor showed high internal consistency (Factor 1, $\alpha = .90$; Factor 2, $\alpha = .87$; Factor 3, $\alpha = .91$). Unit-weighted factor scores were obtained by summing the items that loaded on each factor. Complete details of the EFA are available upon request from Stacy K. Lynch or in D'Onofrio et al. (2005).

The first factor, Externalizing Behaviors, includes reports of oppositional defiant behaviors, attention problems, hyperactivity, conduct disorder items (excluding serious violations of rules), and report of recurrent legal problems due to alcohol use. The second factor, Drug and Alcohol Use, includes cigarette use, alcohol use, alcohol abuse, alcohol dependence, drug use, and problems associated with drug use. The third factor, Internalizing Behaviors, includes depressive episode criteria and suicidal items. A 1-year follow-up of 176 of the offspring of the twins indicated high reliability of the outcome factors. Kappa coefficients for the outcome behavioral factors were .78 for the measure of externalizing behavior, .89 for the drug and alcohol use measure, and .74 for the measure of internalizing behaviors.

Analyses

The data were first analyzed through group means comparisons of the five punishment groups (nonphysical mild, physical mild, nonphysical harsh, physical harsh, and no punishment) across the three offspring outcome factors determined by the EFA (Externalizing Behaviors, Drug and Alcohol Use, and Internalizing Behaviors). We used the results of this comparison to assess any associations between particular punishment style and the outcome pathology factors. Any significant associations were further explored with comparisons incorporating the avuncular (aunt or uncle) punishment status as a measure of genetic and shared environmental influences on the outcome factors.

Prior to running the next set of models, we regressed a number of the initial selection criteria out of the data (parental depression, parental drug use, and parental alcohol use). In other words, regression analysis was run on the data by using parental depression, parental drug use, and parental alcohol use as predictors. The residuals from this were used in the subsequent analyses.

Hierarchical linear modeling (HLM) was next used to provide appropriate controls for the differing number of children in the families. These regressions tested the effect of parental punishment style, avuncular punishment style, zygosity, and the interaction of zygosity and avuncular punishment on child outcomes. In each regression, age, the square of age, gender of the child, gender of the twin parent, socioeconomic status (SES) of the twin parent, and the divorce status of the twin parent were statistically controlled.

Results

Group Means Comparisons

The results of the preliminary mean comparisons can be found in Table 1. We dropped the nonphysical harsh group

Table 1
Offspring Outcomes by Parental Punishment Style (MZ and DZ Combined)

Outcome factor	Punishment status								Effect sizes		
	1		2		3		4				
	(Nonphysical mild; <i>n</i> = 1,640)		(Physical mild; <i>n</i> = 576)		(Physical harsh; <i>n</i> = 286)		(No punishment; <i>n</i> = 51)		1 vs. 2	1 vs. 3	2 vs. 3
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>			
Externalizing Behaviors	3.24	4.26	3.67	4.52	5.54	5.55	4.43	5.43	.099	.514	.384
Drug and Alcohol Use	9.88	5.64	10.16	5.93	12.99	6.85	10.49	6.01	.049	.533	.452
Internalizing Behaviors	3.94	4.31	3.67	4.33	4.51	4.73	4.47	4.72	.063	.203	.261

Note. MZ = monozygotic; DZ = dizygotic.

given that only 1 offspring fit that category. We computed effect sizes to determine which groups were different from the nonphysical mild group by using the following formula (Cohen, 1988):

$$d = \frac{\bar{X} - \mu_0}{\sigma} \quad (1)$$

Average scores across the three outcomes for the nonphysical mild group and physical mild group were quite similar. The comparisons between the nonphysical mild and the physical harsh punishment groups were more striking, as were the differences between the physical mild (corporal punishment) group and the physical harsh punishment group. According to Cohen's (1988) conventions, an effect size of .20 indicates a small effect, .50 indicates a medium effect, and .80 would point to a large effect. After examining which comparisons indicated a strong effect for punishment style, we retained for further analysis a mild punishment group consisting of the nonphysical mild offspring and the physical mild offspring and a second group consisting of physical harsh offspring. More detailed significance testing of these differences will be reserved for the HLM that follows.

The second means analysis, which can be found in Table 2, broke down the association between punishment methods and child outcomes according to the punishment style of the children's parents as well as the punishment status of their uncles or aunts (i.e., the children's parents' twins; we refer to these relations as avuncular regardless of gender or zygosity).

There are essentially three kinds of families in this breakdown: families in which the twin parents are concordant for

harsh physical punishment, families in which the twin parents are concordant for mild punishment, and those in which the parents are discordant. The children from the latter group can be further divided into those with a harsh-punishing parent and a mild-punishing aunt or uncle and those with a mild-punishing parent and a harsh-punishing aunt or uncle. Comparisons of outcomes in children born to twin parents discordant for punishment style are particularly telling because they control for numerous environmental and genetic confounds shared by the twin parents.

For example, for child externalizing behavior, having neither a parent nor an aunt or uncle using harsh physical punishment predicted a mean score of 3.53; having both a parent and an aunt or uncle who are harsh punishers predicted a score of 4.42. Much of this difference between twinships can also be observed within discordant twinships: Children with a harsh-punishing parent and a mild-punishing aunt or uncle had a mean externalizing score of 4.20, whereas children with a mild-punishing parent and a harsh-punishing aunt or uncle had a mean score of 3.54. The same pattern is observed for drug and alcohol use, but with respect to the internalizing behaviors factor, the trend is not as clear. This may be partly due to the fact that the original association between harsh punishment and this outcome was weaker. Significance testing for these results will be examined through the HLM analyses that follow.

HLM

As mentioned, prior to running the following HLM, we ran a regression using a number of the original selection criteria as predictors (parental depression, parental drug use, and parental alcohol). The residuals from this analysis were used in the means analyses and HLM that follow.

Table 2
Mean Offspring Outcome Scores for Parental Punishment and Avuncular Punishment

Factor	Parental mild/ avuncular mild			Parental mild/ avuncular harsh			Parental harsh/ avuncular harsh			Parental harsh/ avuncular mild		
	<i>M</i>	<i>SE</i>	<i>n</i>	<i>M</i>	<i>SE</i>	<i>n</i>	<i>M</i>	<i>SE</i>	<i>n</i>	<i>M</i>	<i>SE</i>	<i>n</i>
Externalizing Behaviors	3.53	4.52	1,712	3.54	4.27	243	4.20	5.08	507	4.42	5.06	91
Drug and Alcohol Use	10.09	5.77	1,712	10.91	6.19	243	11.09	6.34	508	12.30	6.59	91
Internalizing Behaviors	3.94	4.35	1,708	3.85	4.24	243	4.22	4.54	507	4.84	4.76	91

Family data of the kind we are analyzing can be understood in terms of a nested set of associations that can be elegantly described and analyzed by using HLM (Jenkins, Rasbash, & O'Connor, 2003). HLM is designed to account for data that are clustered in systematic but complex ways. In our design (see Figure 1), children were clustered into nuclear families of various sizes, and nuclear families were clustered within twinships (parents and their children, and parents' twins and their children).

In simple terms, HLM analyses comprise two kinds of effects. Fixed effects are the average effect of one variable on another, collapsed across the various levels of clustering; these are most similar to the effects in traditional regression models. So, for example, if a child outcome is regressed on a parenting practice, the resulting effect can be expressed as an average (or fixed) effect and a variance of the fixed effect across levels of clustering (the random effect).

The hierarchical structure of our data is as follows. At the first level are the multiple children born to the same twin parents. Among many other advantages, HLM is much more flexible than traditional regression with regard to differing numbers of experimental units per level, as is the case at the child level of our analysis, where different families include different numbers of children. At the second level of the analysis are the individual twin parents, and at the third level are the twin families that include the two twin parents and their children. Each twin family can be described as either monozygotic or dizygotic.

Any variable that varies at the level of the individual children—for example, a child's description of his or her parents' style of punishment—can be expressed as a set of deviations from a mean at a higher level of clustering. So,

for example, one can compute a mean punishment description of all children born to both twins in a twin family. The mean punishment description of children born to each twin parent can then be expressed as a deviation from the family (both twins) mean. Each child's individual description can then be expressed as a deviation from the twin parent mean. For any given child, the sum of the extended family mean, the twin family deviation, and the individual child deviation is equal to the child's original score. This presentation is a useful way to understand within- and between-family differences with respect to the relationship between punishment and outcomes in children. If the relationship (parameter) differs markedly at the twin family level of the cluster, it points to an uncontrolled genetic or shared environmental contribution to the association.

HLM usually proceeds as a sequence of increasingly complex models. Model 1 (see Table 3) included only an intercept and the covariates that will be included in all models: gender of the offspring, age of the offspring, the square of age, the gender of the twin parent, the divorce status of the twin parent, and the SES of the twin parent. The fixed value of the intercept, 2.36, represents the average within-family intercept when externalizing is predicted from gender, age, the square of age (which was entered to test for a nonlinear relationship between the age of the offspring and the outcome factor), gender of the twin parent, SES of the twin parent, and the divorce status of the twin parent. The random effects describe variation in the within-family intercept across nested levels of clustering. The random variance at the child level, 16.86, describes the variability in externalizing among children in the same family; the random effect at the twin level, 3.07, indicates

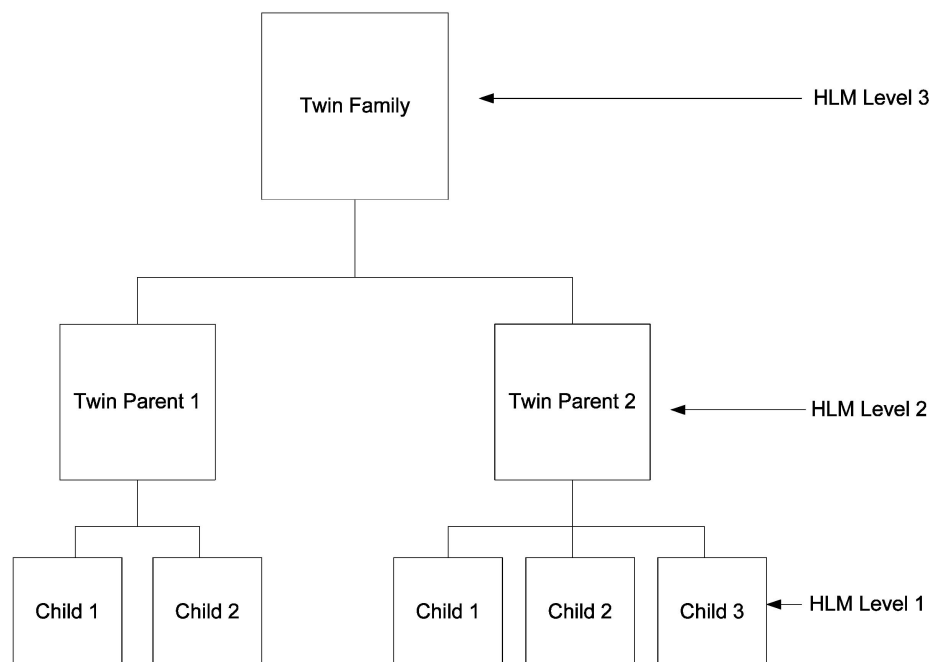


Figure 1. Nested levels in the children of twins design. HLM = hierarchical linear modeling.

Table 3
Hierarchical Linear Models of Harsh Physical Punishment and Externalizing Behavior

Parameter (<i>N</i> = 2,502)	Model 1	Model 2	Model 3	Model 4	Model 5
Random effects					
Families	−0.14* (.00)	−0.23* (.01)	−0.24* (.01)	−0.26* (.01)	−0.22* (.01)
Twins	3.07* (.53)	2.97* (.50)	2.93* (.49)	2.93* (.49)	3.00* (.51)
Children	16.86* (.65)	15.75* (.61)	15.65* (.61)	15.65* (.61)	15.14* (.63)
Fixed effects					
Intercept	2.36 (1.83)	2.02 (1.81)	2.11 (1.80)	2.06 (1.80)	1.98 (1.88)
Gender of offspring ^a	1.54* (.18)	1.38* (.18)	1.42* (.18)	1.42* (.18)	1.43* (.18)
Age of offspring	0.15 (.14)	0.16 (.14)	0.15 (.14)	0.15 (.14)	0.16 (.15)
Square of age	0.00 (.00)	0.00 (.00)	0.00 (.00)	0.00 (.00)	0.00 (.00)
Gender of twin parent	0.45* (.21)	0.39 (.20)	0.40 (.20)	0.41* (.20)	0.42* (.26)
SES of twin parent	−0.20* (.06)	−0.16* (.06)	−0.16* (.06)	−0.15* (.06)	−0.16* (.06)
Twin parent divorce ^b	0.99* (.23)	0.66* (.23)	0.64* (.23)	0.63* (.23)	0.61* (.24)
Harsh punishment					
Individual child report (1)		1.95* (.28)			
Individual child deviation (1)			1.92* (.34)	1.93* (.34)	1.68* (.35)
Twin family deviation (2)			1.99* (.38)	1.18* (.60)	1.37* (.61)
Extended family mean (3)				2.41* (.45)	2.16* (.60)
Zygosity of parent ^c					0.06 (.25)
Zygosity × Extended Family Mean					0.11 (.93)
AIC (smaller is better)		13,599.5	13,545.3	13,541.1	12,157.0

Note. All models control for parental depression, drug use, and alcohol use. Standard errors are given in parentheses. SES = socioeconomic status; AIC = Aikake's information criterion.

^a 0 = female, 1 = male. ^b 0 = married, 1 = divorced. ^c 0 = dizygotic, 1 = monozygotic.

* $p < .05$.

that considerably less variability in externalizing can be attributed to variation among the means of children raised in the same nuclear family. The negative variability at the family level suggests that there is less variability among the means of externalizing across extended families than would be expected on the basis of the variability within and between twin parents. This phenomenon can occur because of negative covariances among twin parents but is more likely simply a representation of sampling variation around a population value of zero. We estimated a second set of models without a random effect for clustering at the extended family level. Although eliminating the random effect produced small improvements in fit indices, the fixed effects did not change substantially, so the reduced models will not be reported here.

The relative magnitudes of the random variances are usually expressed in terms of intraclass correlations, which represent the percentage of the total variance of y that can be attributed to a particular level of the hierarchy. So, for example, the intraclass correlations of twin families is equal to $3.07/(3.07 + 16.86) = .15$, which suggests that about 15% of the variability in externalizing can be attributed to variation among twin families. The rest is explained by variation within families.

In Model 2, each child's description of his or her parent's punishment style was introduced as a predictor of the child's outcome. The resulting unstandardized regression coefficient of 1.95 ($p < .05$) describes the incremental difference in externalizing between a child born to harsh-punishing parents relative to mild-punishing parents. Because separate parameters were not included at higher levels of the hierarchy, the model assumes that the regression of externalizing on punishment is constant across levels.

In Model 3, the punishment variable was decomposed as described above, into a mean punishment score for each twin parent and a deviation of each child's punishment report from the nuclear family mean. The regression coefficients for these components are both significant and roughly equal (1.92 and 1.99, respectively), suggesting that the difference in average externalizing between harsh- and mild-punishing families is about the same as the difference between harshly and mildly punished children within the same family, as would be expected if harsh punishment is in fact a structural cause of externalizing. In Model 4, the punishment variable was further decomposed into three components: an extended twin family mean (2.41), a nuclear family deviation from the extended family mean (1.18), and an individual child deviation (1.93). In Model 5, we introduced as predictors the zygosity of the twin pairs and the interaction of zygosity with the extended family mean. The interaction could potentially allow us to decompose the between- and within-twin regressions into genetic and environmental components, but neither coefficient approached significance. With zygosity controlled, the main effect of harsh parenting remained significant and roughly equal at all three levels of the hierarchy.

In Table 4, we show the results of applying the same set of models to the relation between harsh parenting and drug and alcohol behavior in the offspring. We will not describe the entire sequence of models in the interest of brevity. With regard to the complete model, Model 5, we saw again a significant effect of reported harsh parenting, roughly equal at the three levels of the design (2.28, 1.75, and 2.38, respectively). We concluded that reports of harsh parenting are associated with greater drug and alcohol use among children raised together in a single family, that twins who

Table 4

Hierarchical Linear Models of Harsh Physical Punishment and Drug and Alcohol Use

Parameter (<i>N</i> = 2,502)	Model 1	Model 2	Model 3	Model 4	Model 5
Random effects					
Families	0.22* (.01)	0.03* (.00)	0.05* (.00)	0.04* (.00)	-0.24* (.01)
Twins	8.35* (.96)	8.57* (.94)	8.56* (.95)	8.55* (.95)	8.68* (1.02)
Children	22.60* (.91)	21.57* (.89)	21.62* (.89)	21.63* (.89)	21.44* (.94)
Fixed effects					
Intercept	-15.98 (2.27)	-16.51 (2.28)	-16.50 (2.29)	-16.53 (2.29)	-16.39 (2.42)
Gender of offspring ^a	2.58* (.22)	2.33* (.22)	2.35* (.22)	2.35* (.22)	2.38* (.23)
Age of offspring	2.02* (.18)	2.04* (.18)	2.04* (.18)	2.04* (.18)	2.03* (.19)
Square of age	-0.04* (.00)	-0.04* (.00)	-0.04* (.00)	-0.04* (.00)	-0.04* (.00)
Gender of twin parent	0.73* (.27)	0.70* (.28)	0.71* (.28)	0.71* (.28)	0.63 (.35)
SES of twin parent	-0.28* (.07)	-0.23* (.07)	-0.23* (.07)	-0.23* (.07)	-0.23* (.08)
Twin parent divorce ^b	1.83* (.30)	1.62* (.31)	1.63* (.31)	1.62* (.31)	1.54* (.32)
Harsh punishment					
Individual child report (1)		2.26* (.35)			
Individual child deviation (1)			2.36* (.41)	2.36* (.41)	2.28* (.42)
Twin family deviation (2)			2.10* (.50)	1.68* (.78)	1.75* (.80)
Extended family mean (3)				2.31* (.59)	2.38* (.80)
Zygosity of parent ^c					0.04 (.36)
Zygosity × Extended Family Mean					-0.32 (1.23)
AIC (smaller is better)		14,670.3	14,635.6	14,633.5	13,193.9

Note. All models control for parental depression, drug use, and alcohol use. Standard errors are given in parentheses. SES = socioeconomic status; AIC = Aikake's information criterion.

^a 0 = female, 1 = male. ^b 0 = married, 1 = divorced. ^c 0 = dizygotic, 1 = monozygotic.

* $p < .05$.

are reported to use more harsh parenting than their co-twins have children who report more alcohol use than those of their co-twin, and that twinships with more reported harsh parenting have children with more drug and alcohol use than twinships with less harsh punishment.

Table 5 repeats the model sequence for the outcome of internalizing behaviors. As with the models for drug and

alcohol use, we will discuss only the final model, Model 5. We saw a significant effect of reported harsh parenting at two levels of the design in this model, though all three levels were generally comparable (0.92, 0.51, and 1.96, respectively). From this model it seems that reported harsh punishment is associated with greater internalizing behaviors among children raised together in a single family, although

Table 5

Hierarchical Linear Models of Harsh Physical Punishment and Internalizing Behavior

Parameter (<i>N</i> = 2,502)	Model 1	Model 2	Model 3	Model 4	Model 5
Random effects					
Families	0.02* (.00)	0.03* (.00)	0.04* (.00)	0.02* (.00)	0.03* (.00)
Twins	2.25* (.50)	2.18* (.51)	2.16* (.51)	2.18* (.51)	1.93* (.54)
Children	15.84* (.62)	15.67* (.63)	15.67* (.63)	15.64* (.63)	15.91* (.67)
Fixed effects					
Intercept	0.50 (1.75)	0.42 (1.76)	0.47 (1.78)	0.42 (1.78)	0.88 (1.88)
Gender of offspring ^a	-1.55* (.17)	-1.61* (.17)	-1.62* (.17)	-1.62* (.17)	1.53* (.18)
Age of offspring	0.31* (.14)	0.31* (.14)	0.30* (.14)	0.30* (.14)	0.26 (.15)
Square of age	0.00 (.00)	0.00 (.00)	0.00 (.00)	0.00 (.00)	0.00 (.00)
Gender of twin parent	0.17 (.20)	0.11 (.20)	0.12 (.20)	0.13 (.20)	0.01 (.25)
SES of twin parent	-0.21* (.06)	-0.19* (.06)	-0.19* (.06)	-0.18* (.06)	-0.18* (.06)
Twin parent divorce ^b	1.26* (.22)	1.00* (.22)	1.06* (.22)	1.04* (.22)	0.87* (.23)
Harsh punishment					
Individual child report (1)		1.14* (.27)			
Individual child deviation (1)			1.01* (.34)	1.03* (.34)	0.92* (.35)
Twin family deviation (2)			1.34* (.38)	0.44 (.60)	0.51 (.61)
Extended family mean (3)				1.79* (.44)	1.96* (.59)
Zygosity of parent ^c					-0.06 (.26)
Zygosity × Extended Family Mean					-0.61 (.92)
AIC (smaller is better)		13,530.3	13,492.7	13,487.8	12,165.9

Note. All models control for parental depression, drug use, and alcohol use. Standard errors are given in parentheses. SES = socioeconomic status; AIC = Aikake's information criterion.

^a 0 = female, 1 = male. ^b 0 = married, 1 = divorced. ^c 0 = dizygotic, 1 = monozygotic.

* $p < .05$.

the equivalence of the relation across levels of the analysis is not as striking as it was for the other two factors.

Discussion

In the present study, we found that controlled corporal punishment (defined as slapping or spanking) has only a slight association with negative outcomes for children. These results differ from findings cited by Gershoff (2002b) but are in agreement with Baumrind, Larzelere, and Cowan's (2002) hypothesis that when appropriately defined, corporal punishment would not show significant associations with negative outcomes. Harsh physical punishment, however, did have significant associations with behavioral and emotional outcomes in the offspring. Our analyses indicated a moderate association between harsh physical punishment and two outcome domains in the children of twins: externalizing behaviors and drug and alcohol use. The association with internalizing behaviors was in the same direction but weaker. The associations between punishment style and child outcomes were maintained in comparisons of children born to discordant twin parents: In pairs of twin parents in which one parent used harsh physical punishment and the other did not, the children of the harsh-punishing parent displayed more externalizing and drug- and alcohol-related symptoms than the children of the nonharsh-punishing parent.

HLM allowed us to specify these effects with greater precision. The HLM analysis showed that the difference in outcome between harshly punished and mildly punished children was roughly the same at all three levels of our family design. That is, children in a family who report more harsh punishment than their siblings report more symptoms than their siblings; twins who are reported to use more harsh punishment than their co-twin have children with more symptoms than the children of their milder punishing co-twin; and twin pairs who are reported to use more harsh punishment have children with more symptoms than pairs who use milder punishment. This pattern of findings is precisely what one would expect if harsh punishment were indeed a cause of the outcomes. If, on the other hand, the association between harsh parenting and child outcomes was the result of an uncontrolled familial third variable that influenced both, the association would be observed at the level of twin families, but there would be no reason to observe a difference between discordant twin parents.

With this in mind, several limitations of the data and our analysis must be emphasized. First, the fact that the results are consistent with a causal hypothesis does not prove the hypothesis; in the absence of random assignment, such proof is probably impossible to obtain. All the twin design allows us to do is to eliminate genetic and shared environmental variables that might otherwise confound the relationship, which we nevertheless believe is a significant advance of simple comparisons of punishment styles across families. Second, of the three methodological issues that were mentioned in the introduction, we have done the least to remediate problems of measurement. Both punishment style and outcome were based on retrospective self-report,

which is certainly less than ideal. It would be very productive to collect children of twins data by using a variety of concurrent self-report and observational measures of parenting and outcome. Along these lines, it would be useful to have even more specific distinctions within the construct of harsh punishment.

Another limitation of our study is the lack of inclusion of the punishment methods and other characteristics of the spouses of the twins. The children of twins design controls for genetic and shared environmental confounds arising in the twin parents, and it would be reasonable to assume that a similar set of confounds exists in the non-twin spouses as well. The spouses, therefore, represent a significant source of remaining uncontrolled variability in the analysis.

A final limitation of our study pertains to the selection process through which this sample was obtained. Because a significant portion of the sample was selected for inclusion on the basis of either one or both of the twin parents reporting divorce, alcohol dependence, conduct disorder, or depression, the sample included in the analysis cannot be considered representative of the Australian population from which it was drawn. The limitation is especially pertinent with regard to the first means analysis in which the selection criteria were not controlled for. It must be realized that the relatively higher frequency of negative behaviors in this sample compared with other samples could hypothetically interfere with the magnitude of the results for corporal punishment. However, regressing much of the selection criteria out of the data and including still others as covariates in the HLM had little effect on the magnitude of our results.

On a practical level, our findings have two major implications for parents and developmental scientists. We found a clear distinction between mild physical punishment, such as controlled spanking, which was not associated with negative outcomes in children, and harsh physical punishment, which was associated with effects of considerable magnitude. This suggests that the common definitional practice of lumping together of varying degrees of physical punishment under the rubric of "corporal punishment" may be counterproductive. Moreover, the association we observed between harsh physical punishment and negative outcomes in children survived a relatively rigorous test of its causal status, thus increasing our conviction that harsh physical punishment is a serious risk factor for children. Genetically informative designs can be used to confirm the efficacy of parental behavior by isolating crucial comparisons between twins discordant for key parenting variables. Studies in which developmental processes are studied in a genetically informative context hold the key to a long-awaited integration of behavior genetics and developmental psychology.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Baumrind, D., Larzelere, R. E., & Cowan, P. A. (2002). Ordinary physical punishment: Is it harmful? *Psychological Bulletin*, 128, 580–589.

- Brewin, C. R., Andrews, A., & Gotlib, I. H. (1993). Psychopathology and early experience: A reappraisal of retrospective reports. *Psychological Bulletin*, 113, 82–98.
- Bucholz, K. K., Cadoret, R., Cloninger, C. R., Dinwiddie, S. H., Hesselbrock, V. M., Nurnberger, J. I., Jr., et al. (1994). A new semi-structured psychiatric inventory for use in genetic linkage studies: A report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, 55, 149–158.
- Cohen, J. (1988). *Statistical power for the behavioral sciences*. New York: Academic Press.
- D'Onofrio, B. M., Turkheimer, E., Eaves, L. J., Corey, L. A., Berg, K., Solaas, M. H., & Emery, R. E. (2003). The role of the children of twins design in elucidating causal relations between parent characteristics and child outcomes. *Journal of Child Psychology and Psychiatry*, 44, 1130–1144.
- D'Onofrio, B. M., Turkheimer, E., Emery, R. E., Slutske, W. S., Heath, A. C., Madden, P. A., & Martin, N. (2005). A genetically informed study of marital instability and the association with offspring psychopathology. *Journal of Abnormal Psychology*, 114, 570–586.
- Emery, R. E., & Laumann-Billings, L. (1998). An overview of the nature, causes, and consequences of abusive family relationships: Toward differentiating maltreatment and violence. *American Psychologist*, 53, 121–135.
- Engeland, B., Jacobvitz, D., & Sroufe, L. A. (1988). Breaking the cycle of abuse. *Child Development*, 59, 1080–1088.
- Gerlsma, C., Kramer, J. J., Scholing, A., & Emmelkamp, P. M. (1994). The influence of mood on memories of parental rearing practices. *British Journal of Clinical Psychology*, 33, 159–172.
- Gershoff, E. T. (2002a). Corporal punishment, physical abuse, and the burden of proof: A reply to Baumrind, Larzelere, and Cowan (2002), Holden (2002), and Parke (2002). *Psychological Bulletin*, 128, 602–611.
- Gershoff, E. T. (2002b). Parental corporal punishment and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, 128, 539–579.
- Gottesman, I. I., & Bertelsen, A. (1989). Confirming unexpressed genotypes for schizophrenia. *Archives of General Psychiatry*, 46, 867–872.
- Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry*, 45, 260–273.
- Heath, A. C., Bucholz, K. K., Madden, P. A. F., Dinwiddie, S. H., Slutske, W. S., Bierut, L. J., et al. (1997). Genetic and environmental contributions to alcohol dependence risk in a nationwide twin sample: Consistency of findings in women and men. *Psychological Medicine*, 27, 1381–1396.
- Heath, A. C., & Martin, N. G. (1994). Genetic influences on alcohol consumption patterns and problem drinking: Results from the Australian NH&MRC twin panel follow-up survey. In T. F. Babor, V. Hesselbrock, R. E. Meyer, & W. Shoemaker (Eds.), *Annals of the New York Academy of Sciences: Vol. 708. Types of alcoholics: Evidence from clinical, experimental, and genetic research* (pp. 72–85). New York: New York Academy of Sciences.
- Holden, G. W. (2002). Perspectives on the effects of corporal punishment: Comment on Gershoff. *Psychological Bulletin*, 128, 590–595.
- Jacob, T., Waterman, B., Heath, A., True, W., Bucholz, K. K., Haber, R., et al. (2003). Genetic and environmental effects on offspring alcoholism. *Archives of General Psychiatry*, 60, 1265–1272.
- Jardine, R., & Martin, N. G. (1984). Causes of variation in drinking habits in a large twin sample. *Acta Geneticae Medicae et Gemellologiae*, 33, 435–450.
- Jenkins, J. M., Rasbash, J., & O'Connor, T. G. (2003). The role of the shared family context in differential parenting. *Developmental Psychology*, 39, 99–113.
- Kazdin, A. E., & Benjet, C. (2003). Spanking children: Evidence and issues. *Clinical Child and Family Psychology Review*, 12, 99–103.
- Kendler, K. S., Bulick, C. M., Silberg, J., Hettema, J. M., Myers, J., & Prescott, C. A. (2000). Childhood sexual abuse and adult psychiatric and substance use disorders in women: An epidemiological and cotwin control analysis. *Archives of General Psychiatry*, 57, 953–959.
- Larzelere, R. E. (2000). Child outcomes of nonabusive and customary physical punishment by parents: An updated literature review. *Clinical Child and Family Psychology Review*, 3, 199–221.
- McCord, J. (1979). Some child rearing antecedents of criminal behavior in adult men. *Journal of Personality and Social Psychology*, 37, 1477–1486.
- Mendle, J., Turkheimer, E., D'Onofrio, B. M., Lynch, S. K., Emery, R. E., Slutske, W., & Martin, N. G. (2004). *Stepfather presence and age of menarche in children of twins*. Manuscript submitted for publication.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide*. Los Angeles: Muthén & Muthén.
- Parke, R. D. (1979). Punishment: Effects, side effects, and alternative strategies. In H. Hom & P. Robinson (Eds.), *Psychological processes in early education* (pp. 71–99). New York: Academic Press.
- Rowe, D. (1994). *The limitations of family influence: Genes, experience and behavior*. New York: Guilford Press.
- Rutter, M., Quinton, D., & Liddle, C. (1983). Parenting in two generations: Looking backwards and looking forwards. In N. Madge (Ed.), *Families at risk* (pp. 60–98). London: Heinemann.
- Slutske, W. S., Heath, A. C., Dinwiddie, S. H., Madden, P. A. F., Bucholz, K. K., Dunne, M. P., et al. (1997). Modeling genetic and environmental influences in the etiology of conduct disorder: A study of 2,682 adult twin pairs. *Journal of Abnormal Psychology*, 106, 266–279.
- Smith, C. A., & Farrington, D. P. (2004). Continuities in antisocial behavior and parenting across three generations. *Journal of Child Psychology and Psychiatry*, 45, 230–247.
- Straus, M. A. (1994). *Beating the devil out of them: Corporal punishment in American families*. New York: Lexington Books.
- Straus, M. A. (2000). Corporal punishment and primary prevention of physical abuse. *Child Abuse and Neglect*, 24, 1109–1114.

Received July 20, 2004

Revision received October 4, 2005

Accepted October 6, 2005 ■