How Does the Inclusion of Twins Conceived via Fertility Treatments Influence the Results of Twin Studies?

S. Alexandra Burt and Kelly L. Klump
Department of Psychology, Michigan State University, East Lansing, MI, USA

Rates of twinning have risen dramatically over the last 30 years, from 1 in 53 births in 1980 to 1 in 30 births in 2009 (Martin et al. (January 2012). *Three decades of twin births in the United States, 1980–2009*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Health Statistics). This increase is largely attributable to increases in the use of fertility treatments (i.e., ovulation induction and in vitro fertilization) combined with delays in parenthood. Although this increase means that more twins are available for recruitment into twin studies, it also has potential consequences for the heritability estimates obtained in these studies. This study sought to evaluate this possibility, making use of the ongoing Michigan Twins Project ($N = 7,261$ families with twins aged 3–17 years), an arm of the Michigan State University Twin Registry. Results revealed that, on average, twins conceived via fertility treatments had lower rates of behavior problems than those conceived naturally, although these behavioral differences could be explained largely by demographic and socio-economic differences across the two types of twin families. Twin similarity did not meaningfully differ across fertility treatment status. We thus conclude that estimates of genetic and environmental influences obtained from twin studies over the last 10–15 years are more or less unaffected by the inclusion of twins conceived via fertility treatments in their samples.

■ **Keywords:** twins, assisted reproductive technology, heritability, fertility treatment

Rates of twinning have changed dramatically over the last 30 years. In 1980, only 1 in every 53 babies born in the United States was a twin, whereas 1 in every 30 babies was a twin in 2009 (Martin et al., 2012). This represents a 76% increase in the twinning birth rate. Most of this increase is directly attributable to the increasing use of fertility drugs and assisted reproductive technologies (Martin et al., 2012). Twin registries in the United States today will thus necessarily include a large proportion of twins conceived via fertility treatments (FERT). How do these twins and their families compare with naturally conceived twins? Does the inclusion of twins conceived via FERT change the representativeness of twin registry samples and, thus, our ability to generalize our findings to the broader population? And, finally, does the inclusion of twins conceived via FERT change the heritability estimates obtained from twin studies? Given the prominence of twin studies in etiologic research, it would be critically important to answer these questions.

One clear difference between the two types of conceptions is in their resulting zygosities: FERT specifically increase the rate of dizygotic (DZ; fraternal) twins compared with monozygotic (MZ; identical) twins (Hall, 2003). In particular, because fertility medications work by increasing gonadotropins and, thereby, stimulating the production of more than one egg for fertilization, the vast majority of twins conceived via FERT are DZ (either same sex or opposite sex). Although twins conceived via FERT do not generally have more congenital abnormalities than those conceived naturally (once correcting for the preponderance of multiple births and parental factors), they do have higher rates of cerebral palsy, lower birth weights, and are born 3.5 days before, on average, as compared to those conceived naturally (as representative publications, see Davies et al., 2012; Lambalk & van Hooff, 2001).

The parents of twins conceived via FERT also differ, at least at a mean level, from those of naturally conceived twins (as reported in Davies et al., 2012; van Beijsterveldt

et al., 2011). The former appear to be older, better educated, and better off financially (Davies et al., 2012), in part not only because delayed childbearing would presumably increase the need for FERT but also because many forms of FERT are not covered by health insurance in the United States and can be quite expensive (which would act here as a form of selection). Building on the possibility of socio-economic differences across the two family types, we might also expect twins conceived via FERT to be better adjusted psychologically (on average) than those conceived naturally, because higher socio-economic status is known to have positive downstream consequences for child psychological health (Leventhal & Brooks-Gunn, 2000). Moreover, some studies have found that parents who have twins following FERT evidence lower levels of parental stress and higher levels of warmth and emotional involvement with their children (Golombok & MacCallum, 2003). In short, twin families in which the twins were conceived via FERT may differ in meaningful ways from those who did not.

There are a small handful of studies (Goody et al., 2005; Tully et al., 2003; van Beijsterveldt et al., 2011) that have evaluated the possibility of twin and parent differences across FERT status. Tully et al. (2003) and van Beijsterveldt et al. (2011) both matched DZ twins conceived via FERT to DZ pairs conceived naturally on a number of child and family variables (e.g., ethnicity, parental income/education, twin birth weight, gestational age, and maternal age at birth) and found no meaningful differences in parental adjustment, parenting behavior, or child psychological and behavioral problems, as assessed cross-sectionally (Tully et al., 2003) or over time (van Beijsterveldt et al., 2011). Such findings indicate that, once you control for family characteristics, there are no differences between twins conceived via FERT and those conceived naturally. Although certainly reassuring in some ways, such results tell us little about how the inclusion of twins conceived via FERT might shape the characteristics of twin registries more generally. Goody et al. (2005) sought to do just this, comparing 101 DZ twin pairs conceived via FERT and 1,073 (unmatched) naturally conceived DZ twin pairs. They found evidence that parents who had conceived their twins via FERT were older and better off financially than those who had not. Similarly, twins conceived via FERT had lower levels of teacher-reported attention problems (but not conduct problems or internalizing symptoms), although this difference did not persist to parental informant-reports.

Although the latter results are interesting, the small sample of twins conceived via FERT, combined with the inconsistent results for twin behavior problems, renders their study somewhat less conclusive than one would like. There is thus a need for a study that examines the possibility of mean differences by FERT status across twins and their parents using a larger sample size. Another, arguably more important, issue relates to heritability estimates. Specifically, do heritability estimates change with the inclusion of

twins conceived via FERT? Goody et al. (2005) looked at DZ correlations and found that they were generally lower for twins conceived via FERT than those conceived naturally. However, they did not directly evaluate the possibility of changes in heritability estimates. There is thus a very clear need for a study to compute heritability estimates with and without such twins to explicitly evaluate the possibility that their inclusion alters heritability estimates. The importance of such analyses is further bolstered by prior work (Stoolmiller, 1998), suggesting that the range restriction in adoptive families may distort etiologic influences on child outcomes. Similarly, it is possible that there are some epigenetic alterations specific to the use of FERT, which could also act to influence twin similarity; indeed, Goody et al. (2005) argued that the stochastic nature of these events would serve to suppress twin similarity. In short, there is good reason to examine whether the inclusion of twins conceived via FERT in twin registries alters heritability estimates.

The goals of this study were threefold. First, prior research (Goody et al., 2005) had indicated that family characteristics varied (on average) across families with twins conceived naturally (NAT) and families with twins conceived via the use of FERT. This study sought to confirm these results using a far larger sample of FERT twins ($N =$ 3,073 NAT twin pairs and 1,871 FERT twin pairs). Second, we sought to confirm the presence of mean differences in child emotional and behavior problems by FERT status, and to clarify whether any such differences were in fact a function of differences in the family characteristics of FERT and NAT twins. Finally, because the influence of FERT twins on heritability estimates has not yet been formally examined, this study sought to do this as well.

Methods

Participants

The Michigan State University Twin Registry (MSUTR) includes several independent twin projects (Klump & Burt, 2006). The 7,261 families included in the current study were assessed as part of the ongoing Michigan Twins Project (MTP) within the MSUTR. The primary aim of the MTP is to collect health data on a large sample of twins that can be used both for data analysis and to select twin families for follow-up research. The twins were 49.9% female, and ranged in age from 3 to 17 years (mean age 9.06 years, *SD* 4.4 years) at the time of their assessment, although a few pairs ($n = 12$) had turned 18 by the time their assessment was completed.

Families were recruited via state of Michigan birth records, in collaboration with the Michigan Department of Community Health. The Michigan Department of Community Health manages birth records and can identify all twins born in Michigan. Birth records are confidential in Michigan; thus, the following recruitment procedures were designed to ensure anonymity of families until they indicated an interest in participating. The Michigan

Department of Community Health identified twins in our age range who lived in Michigan and made use of the Michigan Bureau of Integration, Information, and Planning Services database to locate each family's current address through parents' drivers license information. The Michigan Department of Community Health then mailed premade packets to parents. Families interested in participating simply mailed the completed questionnaire back to study investigators in a prepaid, addressed envelope, or participated online. Parents who did not respond to the first mailing were sent additional mailings approximately 1 month apart until either a reply was received or up to four packets had been mailed. Response rates for MSUTR projects range from 55% to 86%, depending on the target twin population. These rates are similar to or better than those of other twin registries that use similar types of anonymous recruitment mailings (Baker et al., 2002; Hay et al., 2002). The representativeness of our sample is reported later.

Zygosity was established using physical similarity questionnaires administered to the twins' primary caregiver (Peeters et al., 1998). On average, the physical similarity questionnaires used by the MSUTR have accuracy rates of 95% or better. In these data, 28.4% of the twin pairs ($n = 2,060$) were MZ, 35.5% of the twin pairs ($n = 2,576$) were same-sex DZ, and 36.2% of the twin pairs ($n = 2,625$) were opposite-sex DZ.

Measures

FERT status. A single yes–no item assessed whether or not the twins were conceived via FERT: 'Were the twins conceived with the aid of FERT or medications?' In these data, 68.6% of pairs ($n = 4,979$; 61.7% DZ) were conceived naturally and 27.0% ($n = 1,962$; 95.5% DZ) were conceived via FERT. A total of 320 families (or 4.4% of the sample) did not answer this question. These families were omitted from subsequent analyses. Because virtually all twins conceived via FERT (95.4%) were DZ, the bulk of our analyses (with the exception of the formal calculation of heritability estimates) were restricted to DZ twin families (total $N = 3,073$ NAT families and 1,871 FERT families), following the approach of Goody et al. (2005). Note that twin sex did not vary across FERT status and thus was not considered further.

Child behavioral and emotional problems. We made use of the Strengths and Difficulties Questionnaire (SDQ; Goodman & Scott, 1999) to assess child behavioral and emotional problems. We specifically focused on the Conduct Problems scale (i.e., stealing, hot temper, and physical fights; five items, $\alpha = 0.64$), the hyperactivity/inattention scale (i.e., restlessness, overactivity, and distractibility; five items, $\alpha = 0.82$), and the Emotional Problems scale (i.e., anxious/depressive symptoms, including sad mood, worrying, and nervousness; five items, $\alpha = 0.62$). The SDQ is highly correlated with other measures of psychopathology (e.g., the Child Behavior Checklist) and demonstrates good

predictive validity for related diagnoses (Goodman & Scott, 1999). Only 4.5% of twins had missing SDQ data on any scale. To adjust for positive skew, all three variables were log-transformed prior to analysis (skews before and after transformation ranged from 0.93 to 1.53 and –0.21 to 0.32, respectively).

Family/twin characteristics. A number of twin and twin family characteristics were also assessed in the MTP, typically via a single item. These included twin race/ethnicity, twin birth weight, number of siblings, maternal age at twin birth, parental education (averaged across parents here, when information on both were available), approximate annual household income, and the presence or absence of maternal smoking.

Analyses

We first compared the families of NAT and FERT twins on twin family characteristics. We next sought to compare rates of behavior problems across naturally conceived versus FERT-conceived twins. Analyses were conducted using Hierarchical Linear Modeling (HLM) to account for the non-independence of observations within families while maximizing statistical power. HLM also allows us to compute and compare estimated marginal means across FERT status. We next evaluated whether these mean differences (should they be present) persisted once we also regressed the aforementioned twin family characteristics onto the SDQ scale in HLM.

For our final set of analyses, we evaluated whether and how univariate heritability estimates for the SDQ scales might vary with and without FERT-conceived twins. To accomplish this, we first computed intraclass correlations using a saturated model in Mx, a structural-equation modeling program (Neale et al., 2003), and statistically compared these correlations across fertility status using the Fisher r-to-z transformation. In keeping with both the intraclass correlations computed here and by prior meta-analytic work (Burt, 2009), we fitted the ACE model (defined in Table 4) to the conduct problems and emotional symptoms data and the ADE model (also defined in Table 4) to the hyperactivity data. We then added FERT-conceived twin data to the NAT data and recomputed these estimates (this approach was chosen in place of computing heritability estimates separately for NAT and FERT twins, given the near-total absence of MZ FERT twins). To evaluate whether the addition of FERT twins served to meaningfully alter the heritability estimates, we constrained the two sets of estimates to be equal to one another. A significant change in model fit (as described later) would imply that the estimates cannot be constrained and that the inclusion of FERT twins does indeed serve to meaningfully alter heritability estimates. A non-significant change in model fit, by contrast, would imply that the heritability estimates are robust

TABLE 1

Mean Differences in Twin and Twin Family Characteristics Across FERT Status

	Census data	NAT families (cell Ns range from 2,253 to 3,066)	FERT families (cell Ns range from 1,211 to 1,863)
Caucasian (%)	80.2	81.0*	92.7*
Graduate or professional degrees (%)	9.6	10.3*	19.9*
Mean family income	$73,373	$75,940*	$104,090*
Mothers who smoke (%)	–	24*	7*
Number of siblings	–	2.23 (1.17)*	1.65 (0.98)*
Maternal age at twin birth (years)	–	30.2 (5.1)*	32.6 (4.4)*
Twin age (years)	–	9.6 (4.5)*	8.2 (4.1)*

Note: NAT and FERT refer to naturally conceived twins and twins conceived via assisted reproductive technologies, respectively. Census data refers to the 2008–2010 American Community Survey estimate for the state of Michigan, with the exception of that for ethnicity, which refers to state of Michigan Census estimates at the time the twins were born. N = number of families. These vary across cells due to the presence of missing data. *Indicates that mean is significantly different across NAT and FERT families, at $p < .001$.

to the inclusion of FERT-conceived twins (i.e., the two sets of heritability estimates are equivalent to one another).

Because of the small amount of missing data, we made use of full-information maximum-likelihood raw data techniques, which produce less biased and more efficient and consistent estimates than pairwise or listwise deletion in the face of missing data. When fitting models to raw data, variances, covariances, and means are first freely estimated to get a baseline index of fit (minus twice the log-likelihood; –2 lnL). Model fit for the more restrictive biometric models was then evaluated using four information theoretic indices that balance overall fit with model parsimony: the Akaike's Information Criterion (AIC; Akaike, 1987), the Bayesian Information Criteria (BIC; Raftery, 1995), the sample-size-adjusted Bayesian Information Criterion (SABIC; Sclove, 1987), and the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). The lowest or most negative AIC, BIC, SABIC, and DIC among a series of nested models is considered best. As fit indices do not always agree (because they place different values on parsimony, among other things), we reasoned that the best-fitting model should yield lower or more negative values for at least three of the four fit indices.

Results

Mean differences in family characteristics between NAT and FERT twins are presented in Table 1. As seen there, NAT twin families appear to be more or less similar to the general population in the state of Michigan, at least in terms of ethnic and socio-economic indicators. FERT families, by contrast, were significantly more likely to be White (Cohen's *d* effect size [ES] = 0.36), to have a graduate or professional degree (ES = 0.57), and to have higher mean family incomes (ES = 0.55). NAT twins were also significantly older (ES = 0.33) and had more siblings (ES = 0.54) than FERT twins. Similarly, compared with FERT mothers, NAT mothers were significantly younger than when the twins were born (ES = –0.50) and were more likely to identify as a 'smoker' (ES = 0.48). Twin birth weights also differed across FERT and NAT twins (86.2 ounces vs. 88.4 ounces; ES = –0.11,

$p < .01$). In short, FERT twins and twin families look quite differently economically and demographically, at least on average, compared with both NAT families and residents of the state of Michigan more generally.

Do Mean Levels of Twin Emotional and Behavioral Problems Vary Across FERT Status?

We next sought to evaluate whether twins conceived via FERT evidenced more or fewer behavioral and emotional problems than NAT twins. Analyses were done via HLM to account for the independence of twins within families. Fixed-effect estimates of the differences between NAT and FERT twins, which correspond to the differences in their respective estimated marginal means, are presented in Table 2. As seen there, FERT twins evidenced significantly lower levels of conduct problems and hyperactivity/inattention (i.e., standardized differences were –0.15 and –0.10, respectively), but equivalent levels of emotional symptoms, as compared to NAT twins. We next sought to clarify whether the mean differences in externalizing behaviors exhibited by NAT and FERT twins were a function of the very different demographic profiles of NAT and FERT families. We thus evaluated whether these mean differences persisted once we regressed the twin family characteristic variables from Table 1 onto each of the SDQ variables (also done in HLM).[1] As seen in Table 2, the effects of FERT status on SDQ conduct problems and hyperactivity/inattention fully dissipated once we controlled for the aforementioned differences in NAT and FERT family characteristics. Conduct problems and hyperactivity/inattention were instead independently predicted by lower parental education, lower family income, a non-majority twin ethnicity (albeit less so for hyperactivity/inattention), younger ages of the twins, and the presence of maternal smoking.

Do Intraclass Correlations Vary Across FERT Status?

We next calculated intraclass correlations separately by FERT status, thereby allowing us to evaluate whether DZ twin similarity varied across FERT status. Results are presented in Table 3. As seen there, none of the same-sex correlations varied significantly across FERT status (despite

TABLE 2

Unstandardized HLM Fixed-Effect Estimates (SE)

	Conduct problems		Hyperactivity/inattention		Emotional symptoms	
FERT status (0 = NAT, 1 = FERT)	**−0.150 (0.024)****	−0.012 (0.030)	**−0.102 (0.022)****	−0.015 (0.028)	0.014 (0.024)	0.047 (0.031)
Parental education	–	**−0.091 (0.017)****	–	**−0.089 (0.016)***	–	−0.011 (0.017)
Family income	–	**−0.097 (0.017)****	–	**−0.063 (0.016)****	–	**−0.078 (0.017)****
Twin ethnicity	–	**0.124 (0.040)****	–	0.073 (0.037)*	–	−0.007 (0.040)
Twin age	–	**−0.173 (0.013)****	–	**−0.151 (0.012)****	–	**−0.100 (0.014)****
Maternal age at twin birth	–	−0.022 (0.015)	–	−0.010 (0.012)	–	**−0.042 (0.015)****
Maternal smoking	–	**0.278 (0.038)****	–	**0.152 (0.035)****	–	**0.111 (0.038)****

Note: NAT and FERT refer to naturally conceived twins and twins conceived via assisted reproductive technologies, respectively. Like FERT status, maternal smoking (0 = no, 1 = yes) and twin ethnicity (0 = Caucasian, 1 = non-Caucasian) were dummy-coded prior to analysis. Annual family income, twin age at assessment, maternal age at twin birth, and parental education were standardized prior analysis to facilitate interpretation of the unstandardized fixed-effect estimates. Given our sample size, $p < .01$ was used as the criterion for statistical significance (indicated by bold and a double asterisk). Marginally significant predictors (in this case, those with significance values of $p < .05$, two tailed) are indicated by a single asterisk.

TABLE 3

Intraclass Correlations for Dizygotic (DZ) Twin Pairs

	NAT Same-sex DZ (N = 1,458–1,460)	FERT Same-sex DZ (N = 863)	NAT Opposite-sex DZ (N = 1,443–1,444)	FERT Opposite-sex DZ (N = 937)
Conduct problems	**0.29** (0.243, 0.337)	**0.30** (0.237, 0.359)	**0.30** (0.248, 0.342)	**0.25** (0.184, 0.305)
Hyperactivity/inattention	**0.09** (0.040, 0.142)	**0.08** (0.016, 0.149)	**0.11*** (0.061, 0.163)	0.00* (−0.062, 0.066)
Emotional symptoms	**0.23** (0.179, 0.276)	**0.21** (0.147, 0.275)	**0.24** (0.192, 0.289)	**0.21** (0.152, 0.275)

Note: NAT and FERT refer to naturally conceived twins and twins conceived via assisted reproductive technologies, respectively. N = number of families. Bold font indicates that the correlation is significantly greater than zero. 95% confidence intervals are presented below the correlations in parentheses.
*Indicates that intraclass correlation is significantly different across FERT status, at $p < .05$.

the very large sample sizes). There was also relatively little evidence of variation across FERT status among opposite-sex twin pairs. The only exception was observed for hyperactivity/inattention, for which the NAT correlation was significantly larger than the FERT correlation (albeit minimally so).

Heritability Estimates

We next evaluated whether the heritability of twin emotional and behavioral problems varied across FERT status. As MZ twin pairs are necessary to compute heritabilities,

they were included in these analyses. Model fit statistics are reported in Table 4. As seen there, the constrained model uniformly provided the better fit to the data, indicating that the inclusion of FERT twins did not meaningfully alter estimates of genetic and environmental influences. Estimates of genetic and environmental influences on conduct problems, for example, were identical with and without FERT twins. For hyperactivity/inattention, broad genetic influences (i.e., both additive and non-additive) were estimated at 63% and 62% of the variance, with and without FERT twins, respectively. Similarly, additive genetic influences on

TABLE 4

Model Fit Statistics

SDQ scale	Model	−2 lnL	df	AIC	BIC	SABIC	DIC
Conduct problems	Unconstrained ACE model	62,391.33	22,810	16,771.33	−75,399.53	−39,155.81	−54,438.54
	Constrained ACE model	**62,392.23**	**22,813**	**16,766.23**	**−75,413.10**	**−39,164.62**	**−54,449.35**
Hyperactivity/inattention	Unconstrained ADE model	62,895.01	22,809	17,277.01	−75,147.99	−38,905.86	−54,187.92
	Constrained ADE model	**62,895.63**	**22,812**	**17,271.63**	**−75,161.70**	**−38,914.81**	**−54,198.88**
Emotional symptoms	Unconstrained ACE model	63,053.30	22,813	17,427.30	−75,083.56	−38,835.08	−54,119.81
	Constrained ACE model	**63,053.78**	**22,816**	**17,421.78**	**−75,097.34**	**−38,844.09**	**−54,130.83**

Note: The ACE model estimates additive genetic, shared, and non-shared environmental influences. The ADE model estimates additive genetic, non-additive genetic, and non-shared environmental influences Heritability estimates were computed with and without the FERT-conceived twins respectively. In the unconstrained model, these two sets of estimates are allowed to vary. In the constrained model, these heritability estimates are constrained to equal one another. Should the constrained model provide the better fit to the data, it would imply that the inclusion of FERT-conceived twins does not meaningfully alter the heritability estimates for that scale. The best-fitting model for each scale (as indicated by the lowest AIC, BIC, SABIC, and DIC values for at least three of the four fit indices) is highlighted in bold.
AIC = Akaike's Information Criterion; BIC = Bayesian Information Criteria; SABIC = sample-size-adjusted Bayesian Information Criterion; DIC = Deviance Information Criterion; SDQ = Strengths and Difficulties Questionnaire.

emotional symptoms were estimated at 53% and 54% of the variance, with and without FERT twins, respectively.

As a final check on these results, we reran these analyses on a random subsample of 500 families (a typical twin study size) to ensure that they were not influenced in any way by our very large sample. Proportions of MZ, NAT DZ, and FERT DZ were maintained (i.e., 139 MZ, 221 NAT DZ, and 140 FERT DZ). Results again indicated that heritability estimates did not vary with the inclusion of FERT twins ($\Delta X^2 = 0.019$ on 3 degrees of freedom).

Discussion

Consistent with prior research, family characteristics such as ethnicity, parental education, family income, maternal age at twin birth, and proportion of mothers who smoked were found to vary significantly across NAT and FERT families (absolute Cohen's *d* effect sizes ranged from 0.33 to 0.57). The FERT twins also evidenced lower rates of externalizing problems than NAT twins (–0.15 for conduct problems and –0.10 for hyperactivity/inattention), but not internalizing. Perhaps not surprisingly, the aforementioned socio-economic and demographic differences between families appeared to fully account for these mean differences in twin behavior. Such findings strongly suggest that twin researchers examining mean-level processes should attend to either FERT status or the socio-economic and demographic variables that differ across FERT and NAT families.

Twin similarity, by contrast, did not meaningfully differ across FERT and NAT twins, suggesting that although the inclusion of FERT families in twin study samples may serve to suppress mean levels of externalizing in those samples, they do not alter the corresponding heritability estimates. Constraint analyses confirmed this impression. We thus conclude that estimates of genetic and environmental influences obtained from twin studies over the last 10–15 years are more or less unaffected by the inclusion of FERT twins in their samples.

Although the above findings of mean differences are consistent with those of prior research, our finding that twin correlations did not differ across FERT status differs from that of Goody et al. (2005). They found evidence that DZ correlations were lower for FERT than for NAT twins, although it is worth noting that these differences were inconsistently significant across phenotype and informant (likely reflecting their rather small sample of FERT twin pairs, $N = 101$). Visual inspection of the current data also revealed that the DZ correlations were slightly lower for FERT twins than for NAT twins, but these differences were not significant, with one exception (for hyperactivity/inattention). As noted, however, this very small difference did not translate into differences in our heritability estimates across FERT status.

There are limitations of this study that should be considered. First and foremost, although our use of a partic-

ularly large sample of twin families was, in many ways, a strength of this study, large samples are generally characterized by briefer and less comprehensive phenotype definition (a necessary trade-off given resource constraints). This limitation of large survey samples applies here as well: most family characteristics were assessed with a single item. Even our core behavioral phenotypes were assessed with only five-item scales (albeit scales with reasonable psychometric properties and acceptable validity; see Goodman & Scott, 1999). Second, and building on the above point, all measures were assessed via the same informant (almost always the twins' mother), leading to some concern regarding shared method variance. The fact that our mean-level findings replicated those from a more finely characterized sample with multiple informant reports (Goody et al., 2005) allays this concern to some extent. Nevertheless, future research should continue to explore these questions using more varied and in-depth phenotypic assessments. Third, we were not able to evaluate whether results differed according to the specific type of fertility treatment used, as this information was not collected. As prior work has suggested that there may be some differences in outcome across the various forms of treatment (e.g., Davies et al., 2012), future work should seek to evaluate the role of treatment type in twin similarity.

Conclusion

The current results suggest that although twin studies of mean-level effects on externalizing psychopathology should attend to FERT status in their analyses, studies of genetic and environmental influences on psychopathology need not do so. Such findings are reassuring in the sense that the estimates of genetic and environmental influences obtained from twin studies over the last 10–15 years are likely to be more or less unaffected by the probable inclusion of FERT twins in their samples. Another, more speculative conclusion concerns the presence of epigenetic effects unique to FERT twins, as prior research has argued that such effects would manifest by reducing the similarity of FERT compared with NAT twins. The absence of this pattern, either in the DZ twin correlations or in the corresponding heritability estimates, argues against the presence of systematic differences in epigenetic effects in FERT compared with NAT twins.

Endnote

1 Because the number of siblings is not a known independent predictor of externalizing behaviors (a result confirmed when adding this variable to the regressions here; results not shown), this variable was omitted from the HLM analyses.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Baker, L. A., Barton, M., & Raine, A. (2002). The Southern California Twin Register at the University of Southern California. *Twin Research, 5*, 456–459.

Burt, S. A. (2009). Rethinking environmental contributions to child and adolescent psychopathology: A meta-analysis of shared environmental influences. *Psychological Bulletin, 135*, 608–637.

Davies, M. J., Moore, V. M., Wilson, K. J., Van Essen, P., Priest, K., Scott, H., Haan, E. A., & Chan, A. (2012). Reproductive technologies and the risk of birth defects. *New England Journal of Medicine, 366*, 1803–1813.

Golombok, S., & MacCallum, F. (2003). Practitioner review: Outcomes for parents and children following nontraditional conception: What do clinicians need to know? *Journal of Child Psychology and Psychiatry, 44*, 303–315.

Goodman, R., & Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: Is small beautiful? *Journal of Abnormal Child Psychology, 27*, 17–24.

Goody, A., Rice, F., Boivin, J., Harold, G. T., Hay, D. F., & Thapar, A. (2005). Twins born following fertility treatment: Implications for quantitative genetic studies. *Twin Research and Human Genetics, 8*, 337–345.

Hall, J. G. (2003). Twinning. *The Lancet, 362*, 735–743.

Hay, D. A., McStephen, M., Levy, F., & Pearsall-Jones, J. (2002). Recruitment and attrition in twin register studies of childhood behavior: The example of the Australian Twin ADHD Project. *Twin Research, 5*, 324–328.

Klump, K. L., & Burt, S. A. (2006). The Michigan State University Twin Registry (MSUTR): Genetic, environmental, and neurobiological influences on behavior across development. *Twin Research and Human Genetics, 9*, 971–977.

Lambalk, C. B. & van Hooff, M. (2001). Natural versus induced twinning and pregnancy outcome: A Dutch nationwide survey of primiparous dizygotic twin deliveries. *Fertility & Sterility, 75*, 731–736.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*, 309–337.

Martin, J. A., Hamilton, B. E., & Osterman, M. J. K. (January 2012). *Three decades of twin births in the United States, 1980–2009*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *MX: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.

Peeters, H., Van Gestel, S., Vlietinck, R., Derom, C., & Derom, R. (1998). Validation of a telephone zygosity questionnaire in twins of known zygosity. *Behavior Genetics, 28*, 159–161.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.

Sclove, L. S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 53*, 333–343.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, 64*, 583–639.

Stoolmiller, M. (1998). Correcting estimates of shared environmental variance for range restriction in adoption studies using a truncated multivariate normal model. *Behavior Genetics, 28*, 429–441.

Tully, L. A., Moffitt, T. E., & Caspi, A. (2003). Maternal adjustment, parenting and child behavior in families of school-aged twins conceived after IVF and ovulation induction. *Journal of Child Psychology and Psychiatry, 44*, 316–325.

van Beijsterveldt, C. E. M., Bartels, M., & Boomsma, D. I. (2011). Comparison of naturally conceived and IVF-DZ twins in the Netherlands twin registry: A developmental study. *Journal of Pregnancy*, Article ID 517614 (1–9). doi:10.1155/2011/517614.