Twin Research and Human Genetics (2020), **23**, 265–270
doi:10.1017/thg.2020.76

CAMBRIDGE UNIVERSITY PRESS

Article

The Power to Detect Cultural Transmission in the Nuclear Twin Family Design With and Without Polygenic Risk Scores and in the Transmitted–Nontransmitted (Alleles) Design

Susanne Bruins[1,2], Conor V. Dolan[1,2] and Dorret I. Boomsma[1,2]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands and [2]Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

Abstract

We compare the power of two different approaches to detect passive genotype–environment (GE) covariance originating from cultural and genetic transmission operating simultaneously. In the traditional nuclear twin family (NTF) design, cultural transmission is estimated from the phenotypic covariance matrices of the mono- and dizygotic twins and their parents. Here, phenotyping is required in all family members. A more recent method is the transmitted–nontransmitted (T–NT) allele design, which exploits measured genetic variants in parents and offspring to test for effects of nontransmitted alleles from parents. This design requires two-generation genome-wide data and a powerful genome-wide association study (GWAS) for the phenotype in addition to phenotyping in offspring. We compared the power of both designs. Using exact data simulation, we demonstrate three points: how the power of the T–NT design depends on the predictive power of polygenic risk scores (PRSs); that when the NTF design can be applied, its power to detect cultural transmission and GE covariance is high relative to T–NT; and that, given effect sizes from contemporary GWAS, adding PRSs to the NTF design does not yield an appreciable increase in the power to detect cultural transmission. However, it may be difficult to collect phenotypes of parents and the possible importance of gene × age interaction, and secular generational effects can cause complications for many important phenotypes. The T–NT design avoids these complications.

Keywords: cultural transmission; nontransmitted alleles; genetic nurture; nuclear twin family design; power

(Received 3 September 2020; accepted 16 September 2020; First Published online 16 October 2020)

This study aimed to compare the power of the nuclear twin family (NTF) design and the transmitted–nontransmitted (T–NT) alleles design to detect genotype–environment (GE) covariance due to cultural transmission. The classical NTF design uses the implied phenotypic covariance matrices of monozygotic (MZ) and dizygotic (DZ) twins and their parents, and the T–NT design exploits measured genetic variants (GVs) in parents and offspring, to estimate genetic and cultural transmission.

Simultaneous genetic and cultural transmission leads to passive GE covariance (Plomin et al., 1977). *Passive GE covariance* occurs when parental genotypes influence the rearing environment of their offspring, which is sometimes referred to as *cultural transmission* (e.g., Cavalli-Sforza & Feldman, 1973; Eaves, 1976a, 1976b; Eaves et al., 1977; Fulker, 1988; Maes et al., 2006). Because the offspring inherits half of each parent's alleles, and the offspring is subject to influences of the rearing environment shaped indirectly by parents' genotypes, a covariance arises between genotypic and environmental influences. In addition to passive GE covariance (on which we focus in this article), *evocative and active GE covariance* are also distinguished (Plomin et al., 1977). The latter arises when an individual's behavior and preferences are influenced by the individual's genotype, and the individual actively chooses and creates environments that suit their behavior and preferences. The former occurs when an individual's actions, influenced by the individual's genotype, systematically evoked certain responses from the individual's environment.

GE covariance is of substantive interest as it is thought to be important in cognitive development (Cheesman et al., 2020; Scarr & McCartney, 1983; Zavala et al., 2018) and in the development of behavioral problems (Bornovalova et al., 2014; Harold et al., 2013; Jaffee & Price, 2007; Rutter & Silberg, 2002). It is of statistical interest, as the correct interpretation of model parameters in the classical twin design hinges on the assumption of no GE covariance (Keller et al., 2010). For instance, (unmodeled) covariance between A (additive genetic variable) and C (shared environmental variable) biases the estimate of the shared environmental variance (σ_C) in the classical twin ACE model (Eaves et al., 1977; Keller et al., 2010; Keller et al., 2009; Purcell, 2002).

While there are various designs and models that allow for the estimation of GE covariance (e.g., Carey, 1986; Dolan et al., 2014; Dolan et al., 2020; Eaves et al., 1977), we focus on two designs here

Author for correspondence: Conor V. Dolan, Email: c.v.dolan@vu.nl

Cite this article: Bruins S, Dolan CV, and Boomsma DI. (2020) The Power to Detect Cultural Transmission in the Nuclear Twin Family Design With and Without Polygenic Risk Scores and in the Transmitted–Nontransmitted (Alleles) Design. *Twin Research and Human Genetics* **23**: 265–270, https://doi.org/10.1017/thg.2020.76

© The Author(s) 2020. Published by Cambridge University Press This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

https://doi.org/10.1017/thg.2020.76 Published online by Cambridge University Press

that specifically assess cultural transmission. The NTF design (Keller et al., 2009) extends the classical twin design by including the parents of the twins. In the NTF design, the family environment is defined as an environment shared between all family members that arises due to cultural transmission. Additionally, one can estimate the variance due to sibling shared environment (environment shared between members of a twin pair or sibling pair), variance due to nonadditive genetic effects, unshared environment (all variance due to influences unique to the individual, including measurement error), genotype–family environment covariance, and phenotypic assortative mating.

A more recent design to detect GE covariance stemming from cultural transmission is the T–NT alleles design. In this design, cultural transmission is manifested in the effect on the offspring phenotype of parents' nontransmitted alleles (Bates et al., 2018; Kong et al., 2018). This method can be applied to parent–offspring trios but can easily be extended to include multiple offspring, including twins. As parents pass on half of their alleles to their offspring, the polygenic risk score (PRS) of the offspring is a function of the *transmitted alleles* from both parents. For the other nontransmitted half of the parental alleles, a *nontransmitted PRS* can be calculated. In a linear regression, if genetic transmission is the only pathway of transmission from parent to offspring, the regression coefficient in the regression of offspring phenotype on the nontransmitted PRS should be zero. Rejection of the (null) hypothesis that the regression coefficient is zero suggests that the parental genotype has an indirect effect on the offspring phenotype, that is, that cultural transmission is present.

Our aim is to compare the power of the NTF design and the T–NT design to detect cultural transmission and to assess whether the addition of PRS to the NTF design improves the power to detect cultural transmission in this design. The article is organized as follows. First, we present the classical NTF design, the T–NT design, and the NTF design including PRS. Second, we outline our strategy with respect to simulation and power analysis. Third, we present the results of our simulation studies and discuss the implications.

The NTF Design

The relationship between the total additive genetic variance inferred in the classical NTF design and the variance explained by the PRS is as follows. Suppose that there are M GVs contributing to the variance of phenotype Ph, and that we have measured all M relevant GVs. For convenience (but without loss of generality) assume also that the GVs are in gametic phase (linkage) equilibrium, we have, for individual i:

$$\begin{aligned} Ph_i &= b_0 + b_1 GV_{1i} + b_2 GV_{2i} + b_3 GV_{3i} + \ldots + b_M GV_{Mi} + e_i \\ \sigma^2_{Ph} &= b_1^2\sigma^2_{GV1} + b_2^2\sigma^2_{GV2} + b_3^2\sigma^2_{GV3} + \ldots + b_M^2\sigma^2_{GVM} + \sigma^2_E \\ \sigma^2_{Ph} &= g^2_1 + g^2_2 + g^2_3 + \ldots + g^2_M + \sigma^2_E = \sum_{m=1}^{M} g^2_m + \sigma^2_E, \end{aligned}$$

where $g^2_m = b^2_m var(GV_m)$, that is, the additive genetic variance due to the m'th genetic variant (GV_m), $\sum_{m=1}^{M} g^2_m$ is the total additive genetic variance, and σ^2_E is the residual (e) variance. The additive genetic variance as estimated in the NTF design (or the classical twin design) is an estimate of $\sum_{m=1}^{M} g^2_m$. This estimate of genetic variance may be biased if assumptions of the design are violated. Since the PRS is based on a subset (T) of GVs, the total additive genetic variance (σ^2_A) is the sum of the variance captured by the PRS based on these T measured alleles (σ^2_{PRS}), and the remaining latent additive genetic variance (σ^2_{AL}):

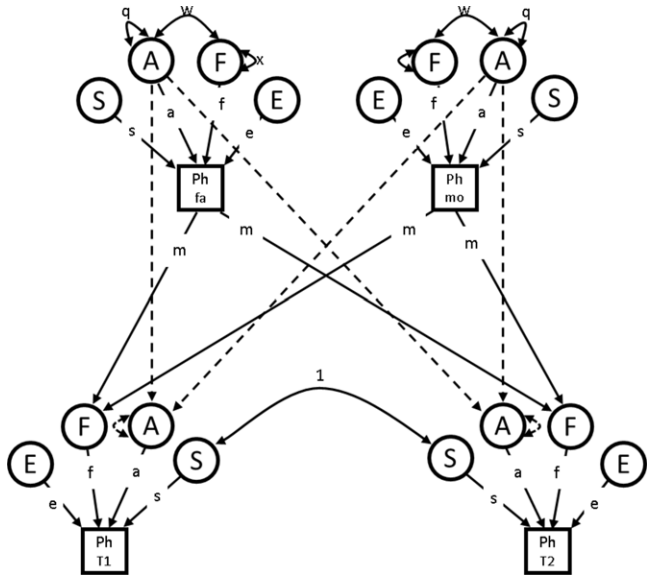

Fig. 1. Path diagram of the classical nuclear twin family (NTF) design given random mating (in the parameterization and notation of Keller et al., 2009). The circles denote latent variables; the squares are observed/measured phenotypic values. Single-headed arrows are paths; double-headed arrows indicate covariances. Solid lines are free parameters; dashed lines are fixed parameters. Dashed paths between parents–offspring A are fixed to .5. *Note*: A, additive genetic; F, family environment due to cultural transmission; S, sibling environment shared between twins; E, unshared environment; Ph, phenotype; *m*, cultural transmission; w, covariance between family environment and additive genetic variable. Constraints include $\sigma^2_A = 1, \sigma^2_S = 1, \sigma^2_E = 1, f = 1$.

$$\sigma^2_{PRS} = \sum_{t=1}^{T} g^2_t, \sigma^2_A = \sum_{m=1}^{M} g^2_m = \sigma^2_{AL} + \sum_{t=1}^{T} g^2_t,$$

and the proportion of explained additive genetic variance is ${R_A}^2 = \frac{\sigma^2_{PRS}}{\sigma^2_A}$. Since the total additive genetic variance σ^2_A is identified by the implied covariance between MZ and DZ twins and their parents, σ^2_{AL} and σ^2_{PRS} are $\sigma^2_{AL} = \sigma^2_A \times (1 - R^2)$ and $\sigma^2_{PRS} = \sigma^2_A \times R^2$, respectively. Note that this decomposition of the additive genetic variance into observed and latent components assumes that the observed genetic variance is not inflated by noise in the PRS.

The path diagrammatic representation of the classical NTF design, given random mating, is given in Figure 1. Using path tracing (Wright, 1920), we can deduct the model-implied variances and covariances. In the NTF design, we then have

$$\sigma^2_{Ph} = a^2\sigma^2_A + s^2\sigma^2_S + e^2\sigma^2_E + f^2\sigma^2_F + 2afw,$$

where the q and x in Figure 1 equal σ^2_A and σ^2_F, respectively. While w is the covariance between the latent variables A and F, the term $2afw$ is the total contribution of the covariance between genotype A and family environment F to the phenotypic variance. Given the scaling in Figure 1 (based on Keller et al., 2009), we have

$\sigma^2_S = 1, \sigma^2_E = 1, \sigma^2_A = 1, f = 1$, so the equation can be rewritten as

$$\sigma^2_{Ph} = a^2 + s^2 + e^2 + \sigma^2_F + 2aw$$

We find the variance of the family environment σ^2_F and the genotype–family environment covariance w by $\sigma^2_F = 2(m^2\sigma^2)$

https://doi.org/10.1017/thg.2020.76 Published online by Cambridge University Press

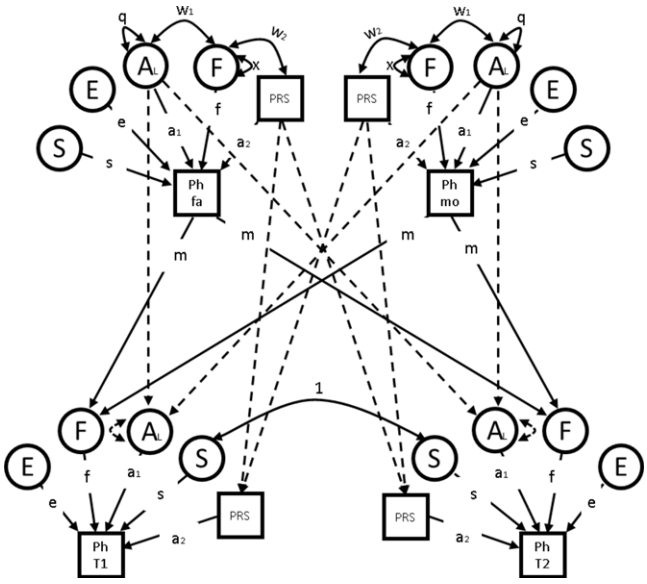

Fig. 2. Path diagram of the nuclear twin family design including PRS, again assuming random mating in the latent model. The circles denote latent variables; the squares are observed/measured values. Single-headed arrows are paths; double-headed arrows indicate covariances. Solid lines are free parameters; dashed lines are fixed parameters. Dashed paths between parents–offspring A_L and PRS are fixed to .5. *Note*: A_L, latent additive genetic; PRS, observed (transmitted) additive genetic; F, family environment due to cultural transmission; S, sibling environment shared between twins; E, unshared environment; Ph, phenotype; *m*, cultural transmission path; w1 and w2, covariance between family environment and latent and observed additive genetic variables. Constraints include $\sigma^2_{AL} = 1, \sigma^2_{PRS} = 1, \sigma^2_S = 1, \sigma^2_E = 1,\ f = 1$.

and $w = m(a\sigma^2_A + wf) = \frac{a\sigma^2_A m}{1-fm} = \frac{am}{1-m}$, given $\sigma^2_A = 1$ and $f = 1$. Note that $w \neq 0$ and $\sigma^2_F > 0$ only if there is cultural transmission, that is, $m \neq 0$.

The NTF design, including PRSs, is depicted in Figure 2. Using path tracing and given the same scaling and constraints as in the classical NTF design, the complete model-implied covariance structure is

$$\sum = \begin{bmatrix} \sigma^2_{fa} & \sigma_{mofa} & \sigma_{T1fa} & \sigma_{T2fa} \\ \sigma_{famo} & \sigma^2_{mo} & \sigma_{T1mo} & \sigma_{T2mo} \\ \sigma_{faT1} & \sigma_{moT1} & \sigma^2_{T1} & \sigma_{T2T1} \\ \sigma_{faT2} & \sigma_{moT2} & \sigma_{T1T2} & \sigma^2_{T2} \end{bmatrix},$$

where *fa*, *mo*, *T*1 and *T*2 stand for the phenotypic (co)variance of father, mother, twin 1 and twin 2, respectively. Variances are assumed to be equal, and (genotypic as well as cultural) transmission is assumed to be equal for both parents, such that

$$\sum_{MZ} = \begin{bmatrix} \sigma^2_{Ph} & \sigma_{P,P} & \sigma_{P,MZ} & \sigma_{P,MZ} \\ \sigma_{P,P} & \sigma^2_{Ph} & \sigma_{P,MZ} & \sigma_{P,MZ} \\ \sigma_{P,MZ} & \sigma_{P,MZ} & \sigma^2_{Ph} & \sigma_{MZ,MZ} \\ \sigma_{P,MZ} & \sigma_{P,MZ} & \sigma_{MZ,MZ} & \sigma^2_{Ph} \end{bmatrix}, \sum_{DZ} = \begin{bmatrix} \sigma^2_{Ph} & \sigma_{P,P} & \sigma_{P,DZ} & \sigma_{P,DZ} \\ \sigma_{P,P} & \sigma^2_{Ph} & \sigma_{P,DZ} & \sigma_{P,DZ} \\ \sigma_{P,DZ} & \sigma_{P,DZ} & \sigma^2_{Ph} & \sigma_{DZ,DZ} \\ \sigma_{P,DZ} & \sigma_{P,DZ} & \sigma_{DZ,DZ} & \sigma^2_{Ph} \end{bmatrix},$$

where the total phenotypic variance is $\sigma^2_{Ph} = a^2_1 + a^2_2 + x + 2a_1w_1 + 2a_2w_2 + s^2 + e^2$, the covariance between parents is $\sigma_{P,P} = x$, the covariance between parent and (twin) offspring is $\sigma_{P,T} = \frac{1}{2}(a_1 + w_1) + \frac{1}{2}(a_2 + w_2) + m\sigma^2$, the covariance between MZ twins is $\sigma_{MZ,MZ} = a^2_1 + a^2_2 + x + 2a_1w_1 + 2a_2w_2 + s^2$ and the covariance between DZ twins is $\sigma_{DZ,DZ} = \frac{1}{2}a^2_1 + \frac{1}{2}a^2_2 + x + 2a_1w_1 + 2a_2w_2 + s^2$ (equations and matrices adapted from Keller et al., 2009).

The T–NT PRS Design

In the T–NT design, the phenotype is regressed on the transmitted and nontransmitted PRS, such that for individual *i*, we have

$$Ph^2 = b0 + b1 \times PRS_{Ti} + b2 \times PRS_{NTi} + \epsilon_i,$$

where PRS_{Ti} is the PRS that was transmitted from parents to individual *i*, and PRS_{NTi} is the nontransmitted PRS of individual *i*, based on the alleles that were not transmitted to individual *i*. While the transmitted PRS in the T–NT design is equal to the PRS in the NTF design, the relation between the nontransmitted PRS in the T–NT design and the cultural transmission effects in the NTF design is more complicated. In the NTF design, cultural transmission processes are captured by the family environment, the genotype–family environment covariance, and the cultural transmission from the parental phenotype to the family environment itself. In the T–NT design, however, cultural transmission is solely represented by the regression of the offspring phenotype on the nontransmitted parental PRS.

Power

Model identification of the NTF is well established, and the addition of PRS does not pose an identification problem. However, within an identified model, the power in tests concerning the parameters is an open question. We conducted power analysis in the NTF models using exact data simulation (van der Sluis et al., 2008). Specifically, the power to detect cultural transmission was calculated as the power to reject a misspecified model in which $m = 0$, when in truth, *m* took values of $m = .05$, $m = .10$, $m = .15$ and $m = .20$. Data were simulated for various parameter settings of a^2, s^2 and e^2 (and $m = .05 - .20$). Detailed model parameters and variance decomposition per scenario are given in Table 1. The PRSs were simulated such that the PRS explained 10% of the additive genetic variance (i.e., ${R_A}^2 = .10$). Since the noncentrality parameter is linearly related to sample size, the NCP was weighted for the number of families (*N*), and power was calculated for $N = 100 - 10.000$ (with an MZ/DZ family ratio of 1).

Given these parameter settings, we compared the power to detect cultural transmission in the NTF design and T–NT design, given identical parameter settings and sample sizes, using ordinary (i.e., not exact) data simulation. To compare power of the T–NT design with that of the NTF design, we must ensure that the sample sizes are comparable. To do so, we determined the number of families for which, in the NTF design + PRS model, the power to reject $m = 0$ was .80, given specific parameter settings. We performed a general estimation equations (GEEs; e.g., Minică et al., 2014) analysis on the phenotype and (transmitted and nontransmitted) PRS data of both members of twin pairs. GEE automatically adjusts the standard errors and test statistic for the dependency. Since the twins are related, the effective *N* is defined as $= \frac{2N}{1+\rho}$, where ρ is the twin correlation. For instance, if we find that the NTF design + PRS requires 1000 twin families (i.e., 2000 individual twins) to reject $m = 0$ with a power of .80, we use $N = 2000$ for the GEE analysis. If then, for example, $\rho_{MZ} = .6$ and $\rho_{DZ} = .4$, we effectively have $NE = \frac{1000}{1.6} + \frac{1000}{1.4} \approx 1339.3$ unrelated individuals. Therefore, the NCP from the GEE was corrected for the effective number of unrelated individuals.

https://doi.org/10.1017/thg.2020.76 Published online by Cambridge University Press

Table 1. Parameter settings and variance components for 12 data simulations

	Par. set.		Variance components							
Sim	a^2	m	σ^2_{AL}	σ^2_{PRS}	σ^2_F	σ^2_S	σ^2_E	$2\sigma_{AL,F}$	$2\sigma_{PRS,F}$	σ^2_{Ph}
1	0.8	.05	.72 (.45)	.08 (.05)	.01 (.01)	.20 (.13)	.50 (.31)	.08 (.05)	.01 (.01)	1.59
2	1.0	.05	.90 (.50)	.10 (.06)	.01 (.01)	.20 (.11)	.50 (.28)	.10 (.05)	.01 (.01)	1.81
3	3.0	.05	2.70 (.67)	.30 (.07)	.02 (.01)	.20 (.05)	.50 (.12)	.28 (.07)	.03 (.01)	4.04
4	0.8	.10	.72 (.42)	.08 (.05)	.04 (.02)	.20 (.12)	.50 (.29)	.16 (.09)	.02 (.01)	1.71
5	1.0	.10	.90 (.46)	.10 (.05)	.04 (.02)	.20 (.10)	.50 (.26)	.20 (.10)	.02 (01)	1.96
6	3.0	.10	2.70 (.61)	.30 (.07)	.09 (.02)	.20 (.05)	.50 (.11)	.60 (.14)	.07 (.02)	4.46
7	0.8	.15	.72 (.39)	.08 (.04)	.09 (.05)	.20 (.11)	.50 (.27)	.25 (.14)	.03 (.02)	1.87
8	1.0	.15	.80 (.42)	.10 (.05)	.10 (.05)	.20 (.09)	.50 (.23)	.32 (.15)	.04 (.02)	2.15
9	3.0	.15	2.70 (.54)	.30 (.06)	.22 (.05)	.20 (.04)	.50 (.10)	.95 (.19)	.11 (.02)	4.98
10	0.8	.20	.72 (.35)	.08 (.04)	.17 (.08)	.20 (.10)	.50 (.24)	.36 (.17)	.04 (.02)	2.07
11	1.0	.20	.90 (.38)	.10 (.04)	.19 (.08)	.20 (.08)	.50 (.21)	.45 (.19)	.05 (.02)	2.39
12	3.0	.20	2.70 (.48)	.30 (.05)	.45 (.08)	.20 (.04)	.50 (.09)	1.35 (.24)	.15 (.03)	5.65

Note: σ^2_{Ph} is the total phenotypic variance given specified parameter settings. σ^2_A, σ^2_F, σ^2_S, σ^2_E and $\sigma_{A,F}$ indicate the phenotypic variance that is explained by genotype, family environment, sibling environment, unshared environment and genotype–environment covariance, respectively. This decomposition is based on $\sigma^2_{Ph} = a^2 + s^2 + e^2 + \sigma^2_F + 2aw$. Standardized variance components are in parentheses (proportions of phenotypic variance). Sim, simulation scenario; Par. set., Parameter settings. Parameter settings represent unstandardized input parameters, where $e^2 = .50$ and $s^2 = .20$ over all simulations.

For the NTF model, we used the NCP to calculate the power to detect cultural transmission. Given exact data simulation, the NCP equals the value of the log-likelihood ratio test statistic in the test of $m = 0$. For the simulations involving the T–NT model, we used ordinary simulation (not exact), so that the NCP is the average test statistic of the regression coefficient of the nontransmitted PRS, minus the degrees of freedom, which was obtained by running 5000 replications. The test statistic in the GEE is a robust z statistic. Since asymptotically, $z^2 = \chi^2$, the squared test statistics follows a (central) χ^2 distribution (1 *df*), under the null hypothesis. Power and required sample sizes are reported given $\alpha = .05$. Analyses were conducted in R (version 3.5.1; R Development Core Team, 2018). Structural equation modeling was performed in OpenMx (Neale et al., 2016), using the NPSOL optimizer. For the GEE modeling, the R-package *gee* was used (Carey et al., 2012). R-scripts for the simulations and model fitting are provided in the Supplementary Materials on the Cambridge Core website.

Results

The transmitted and nontransmitted PRS are uncorrelated, and the regression coefficients of the nontransmitted PRS increased with increasing values of m, indicating that the simulated PRS indeed captured the cultural transmission effects. As expected, the regression coefficient of the transmitted PRS also increases with increasing m, reflecting the presence of cultural transmission. Model information and power calculations for the three models are given in Table 2. As can be seen from Table 2, the power to reject $m = 0$ does not differ substantially between the classical NTF design and the NTF design with PRS, but the NTF designs have greater power to reject $m = 0$ than the T–NT design. Figure 3 displays the power to detect the cultural transmission effects of scenario 7 over a range of sample sizes ($N = 100-10{,}000$).

Since the power of the T–NT design is expected to increase as the additive genetic variance explained by the PRS increases, we tested a model with $a^2 = .8$, $s^2 = .2$, $e^2 = .5$ and $m = .15$, in which the PRS explained either 50% or 100% of the additive genetic variance. We recognize that such high R^2 values are unrealistic; hence, this scenario was designed to inform us as to power contributions as genome-wide association study power increases. As can be seen from Table 3, when the R^2 of the PRS increases, the power to detect cultural transmission also increases, in both the NTF design + PRS and the T–NT design.

Discussion

The present aim was to compare the power of the NTF and T–NT designs to detect cultural transmission. In addition, we tested the benefits of incorporating PRS in the NTF design. The classical NTF design is well powered to detect cultural transmission in sample sizes common in twin studies. Unless the PRS explains a large portion of the additive genetic variance, inclusion of PRS in the design did not result in an appreciable improvement in power. Compared to the T–NT design, the NTF design has greater power, given comparable sample sizes and parameter settings, at the cost, of course, of phenotyping the parents, issues of cultural change across phenotyping, and so forth. The T–NT design requires much larger samples to detect cultural transmission effects. The difference in power is due to the fact that, by definition, the T–NT design only captures part of the transmission effects due to the regression of the nontransmitted PRS being an imperfect representation of the total cultural transmission effect. First, cultural transmission effects are also captured in the transmitted PRS. Second, currently most PRSs only explain a (relatively) small proportion of the total additive genetic variance (e.g., Baselmans et al., 2020). Therefore, the extent to which GE covariance can be captured by PRSs is proportional to the amount of genetic variance captured by the PRSs. For example, if the total passive GE covariance accounts for 27% of the phenotypic variance, a PRS explaining 10% of the additive genetic variance will capture only 3% of the phenotypic variance due to the passive GE covariance (Table 2, simulation 12).

The T–NT design is an ingenious addition to the designs suitable to detect passive GE covariance. At present, work is underway to specify T–NT design as a structural equation model (e.g.,

https://doi.org/10.1017/thg.2020.76 Published online by Cambridge University Press

Table 2. Required sample sizes and power for NTF design, NTF design + PRS and T–NT design

		NTF design						T–NT design								
		NTF design			NTF design + PRS				Transmitted			Nontransmitted				
Sim	nfam	−2LL	*p*	pow	−2LL	*p*	pow	NE	*b* (*SE*)	*Z*	*p*	*b* (*SE*)	*Z*	*p*	pow	.80
1	3500	7.4	.006	.78	7.9	.005	.80	4515	.30 (.02)	17.65	<.001	.01 (.02)	0.85	.397	.11	50,369
2	3100	7.3	.007	.77	7.8	.005	.80	3937	.33 (.02)	17.34	<.001	.02 (.02)	0.88	.395	.14	40,100
3	2000	7.1	.008	.76	7.8	.005	.80	2395	.58 (.04)	16.03	<.001	.03 (.04)	0.81	.403	.13	27,924
4	800	7.2	.007	.77	7.8	.005	.80	1011	.31 (.04)	8.59	<.001	.03 (.04)	0.85	.400	.12	10,359
5	720	7.2	.007	.77	7.8	.005	.80	896	.35 (.04)	8.52	<.001	.04 (.04)	0.85	.396	.14	9403
6	480	7.3	.007	.77	7.9	.005	.80	564	.61 (.08)	7.95	<.001	.06 (.08)	0.79	.413	.12	7100
7	330	7.2	.007	.77	7.8	.005	.80	408	.33 (.06)	5.65	<.001	.05 (.06)	0.86	.398	.12	4209
8	300	7.2	.007	.77	7.8	.005	.80	365	.37 (.07)	5.63	<.001	.06 (.07)	0.86	.397	.14	3715
9	180	7.3	.007	.77	7.8	.005	.80	207	.64 (.13)	4.97	.001	.10 (.13)	0.76	.419	.13	2460
10	170	7.4	.006	.78	7.9	.005	.80	205	.35 (.09)	4.13	.005	.07 (.09)	0.83	.399	.13	2044
11	150	7.4	.006	.78	7.9	.005	.80	179	.39 (.10)	4.03	.006	.08 (.10)	0.81	.401	.14	1868
12	90	7.5	.006	.78	7.9	.005	.80	102	.68 (.19)	3.58	.015	.13 (.19)	0.70	.424	.12	1278

Note: NTF, nuclear twin family; T–NT, transmitted–nontransmitted; Sim, simulation scenario; pow, power; −2LL, −2 log-likelihood; *Z*, *z* statistic of *b*; nfam, number of families; NE, effective number of unrelated individuals (rounded to the nearest integer); .80 is the effective number of unrelated individuals required for a power of .80 (rounded to the nearest integer).

Table 3. Required sample sizes and power given unstandardized input parameters $a^2 = .8$, $s^2 = .2$, $e^2 = .5$, $m = .15$ *and* $R^2 = .5$ *(sim. A) or* $R^2 = 1$ *(sim B)*

		NTF design						T–NT design								
		NTF design			NTF design + PRS				Transmitted			Nontransmitted				
Sim	nfam	−2LL	*p*	pow	−2LL	*p*	pow	NE	*b* (*SE*)	*Z*	*p*	*b* (*SE*)	*Z*	*p*	pow	.80
A	236	5.2	.023	.62	7.8	.005	.80	304	0.74 (.06)	12.67	<.001	0.11 (.06)	1.87	.168	.40	650
B	122	2.7	.103	.37	7.8	.005	.80	157	1.05 (.08)	13.56	<.001	0.16 (.08)	2.02	.146	.69	230

Note: NTF, nuclear twin family; T–NT, transmitted–nontransmitted; Sim, simulation scenario; pow, power; −2LL, −2 log-likelihood; *Z*, *z* statistic of *b*; nfam is the number of families, NE, effective number of unrelated individuals (rounded to the nearest integer), .80, effective number of unrelated individuals required for a power of .80 (rounded to the nearest integer). For simulation A, parameters are $\sigma^2_{AL} = .40\ (.21)$, $\sigma^2_{PRS} = .40\ (.21)$, $\sigma^2_F = .08\ (.05)$, $\sigma^2_S = .20\ (.11)$, $\sigma^2_E = .50\ (.27)$, $2\sigma_{AL,F} = .14\ (.08)$, $2\sigma_{PRS,F} = .14\ (.08)$, $\sigma^2_{Ph} = 1.87$. For simulation B, parameters are $\sigma^2_{AL} = 0$, $\sigma^2_{PRS} = .80\ (.43)$, $\sigma^2_F = .08\ (.05)$, $\sigma^2_S = .20\ (.11)$, $\sigma^2_E = .50\ (.27)$, $2\sigma_{AL,F} = 0$, $2\sigma_{PRS,F} = .28\ (.15)$, $\sigma^2_{Ph} = 1.87$.

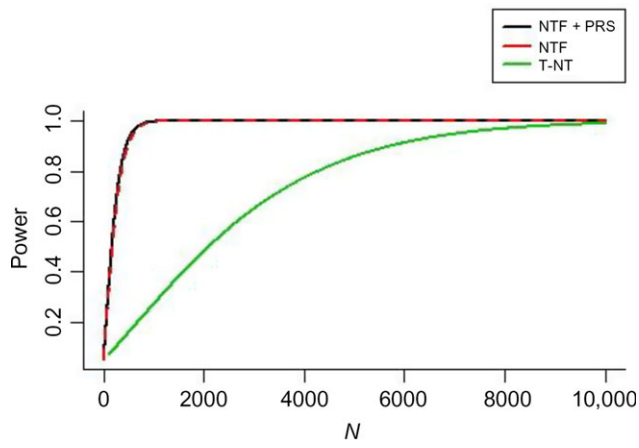

Fig. 3. Power plot of the power to detect genotype–environment correlation due to cultural transmission, in three models. This represent scenario 7, where genotype–environment correlation due to cultural transmission explains 15% of the phenotypic variance ($2\sigma_{AL,F} = .14$ and $2\sigma_{PRS,F} = .02$). In the nuclear twin family (NTF) design, *N* stands for number of families, while in the transmitted–nontransmitted (T–NT) design, *N* is the effective number of unrelated individuals. A power of .80 is achieved at $N = 330$ in the NTF + PRS design and at $N = 4209$ in the T–NT design.

Balbona et al., 2020; Bates et al., 2018; Kim et al., 2020), which will increase its flexibility and scope. For instance, Balbona et al. (2020) and Kim et al. (2020) proposed extensions of the T–NT design in which family environment was modeled as a latent variable. In addition, Kim et al. (2020) proposed that existing bias in the model might be due to unmodeled latent additive genetic variance.

In conclusion, when the PRSs explain a relatively small part of the total additive genetic variance, the incorporation of PRS in the NTF design does not provide any benefits to the power to detect cultural transmission. The classical NTF design is itself relatively well powered to detect cultural transmission, and a decent sample of nuclear twin families is currently more informative with respect to cultural transmission than is measured genotypic information. However, it must be borne in mind that our results are based on phenotypes that are subject to random mating. This assumption may not hold for many phenotypes. While the NTF design can accommodate primary phenotypic assortative mating by the addition of a co-path between parents, the T–NT design does not require random mating, given that assortative mating only occurred in the parental (but not grandparental) generation. In addition, it is often difficult to collect phenotypes of parents, and even when we can, the possible importance of gene × age

https://doi.org/10.1017/thg.2020.76 Published online by Cambridge University Press

interaction and secular generational effects can cause complications for many important phenotypes. The T–NT design avoids these complications.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/thg.2020.76.

Acknowledgments. DIB acknowledges Royal Netherlands Academy of Science Professor Award (PAH/6635).

References

Balbona, J. V., Kim, Y., & Keller, M. C. (2020). Estimation of parental effects using polygenic scores. *BioRvix*. http://doi.org/10.1101/2020.08.11.247049.

Baselmans, B. M., Yengo, L., van Rheenen, W., & Wray, N. R. (2020). Risk in relatives, heritability, SNP-based heritability and genetic correlations in psychiatric disorders: A review. *Biological Psychiatry*. http://doi.org/10.1016/j.biopsych.2020.05.034

Bates, T. C., Maher, B. S., Medland, S. E., McAloney, K., Wright, M. J., Hansell, N. K. & Gillespie, N. A. (2018). The nature of nurture: Using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Research and Human Genetics*, *21*, 73–83.

Bornovalova, M., Cummings, J., Hunt, E., Blazei, R., Malone, S., & Iacono, W. (2014). Understanding the relative contributions of direct environmental effects and passive genotype–environment correlations in the association between familial risk factors and child disruptive behavior disorders. *Psychological Medicine*, *44*, 831–844.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics*, *16*, 319–341.

Carey, V. J., Lumley, T., & Ripley, B. D. (2012). gee: Generalized Estimation Equation Solver, R package version 4.13-18. http://CRAN.R-project.org/package=gee

Cavalli-Sforza, L. L., & Feldman, M. W. (1973). Cultural versus biological inheritance: phenotypic transmission from parents to children. (A theory of the effect of parental phenotypes on children's phenotypes). *American Journal of Human Genetics*, *25*, 618.

Cheesman, R., Hunjan, A., Coleman, J. R., Ahmadzadeh, Y., Plomin, R., McAdams, T. A., Eley, T. C., & Breen, G. (2020). Comparison of adopted and nonadopted individuals reveals gene–environment interplay for education in the UK Biobank. *Psychological Science*, *31*, 582–591.

Dolan, C. V, Huijskens, R. C. A., Minică, C. C., Neale, M. C., & Boomsma, D. I. (2020). *Incorporating polygenic risk scores in the ACE twin model to estimate A-C covariance*. Manuscript submitted for publication.

Dolan, C. V., de Kort, J. M., van Beijsterveldt, T. C., Bartels, M., & Boomsma, D. I. (2014). GE covariance through phenotype to environment transmission: An assessment in longitudinal twin data and application to childhood anxiety. *Behavior Genetics*, *44*, 240–253.

Eaves, L. (1976a). A model for sibling effects in man. *Heredity*, *36*, 205–214.

Eaves, L. (1976b). The effect of cultural transmission on continuous variation. *Heredity*, *37*, 41–57.

Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. (1977). A progressive approach to non-additivity and genotype–environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology*, *30*, 1–42.

Fulker, D. W. (1988). Genetic and cultural transmission in human behavior. In B. S. Weir, E. J. Eisen, M. M. Goodman, & G. Namkoong (Eds.), *Proceedings of the second international conference on quantitative genetics* (pp. 318–340). Sinauer Associates.

Harold, G. T., Leve, L. D., Elam, K. K., Thapar, A., Neiderhiser, J. M., Natsuaki, M. N., Shaw, D. S. & Reiss, D. (2013). The nature of nurture: Disentangling passive genotype–environment correlation from family relationship influences on children's externalizing problems. *Journal of Family Psychology*, *27*, 12–21.

Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, *12*, 432–442.

Keller, M. C., Medland, S. E., & Duncan, L. E. (2010). Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behavior Genetics*, *40*, 377–393.

Keller, M. C., Medland, S. E., Duncan, L. E., Hatemi, P. K., Neale, M. C., Maes, H. H., & Eaves, L. J. (2009). Modeling extended twin family data I: Description of the Cascade model. *Twin Research and Human Genetics*, *12*, 8–18.

Kim, Y., Balbona, J.V., & Keller, M.C. (2020). Nature of nurture analysis via structural equation modeling. Paper presented at the meeting of Behavior Genetics Association.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*, 424–428.

Maes, H. H., Neale, M. C., Kendler, K. S., Martin, N. G., Heath, A. C., & Eaves, L. J. (2006). Genetic and cultural transmission of smoking initiation: An extended twin kinship model. *Behavior Genetics*, *36*, 795–808.

Minică, C. C., Boomsma, D. I., Vink, J. M., & Dolan, C. V. (2014). MZ twin pairs or MZ singletons in population family-based GWAS? More power in pairs. *Molecular Psychiatry*, *19*, 1154–1155.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T., Maes, H., & Boker, S. M. (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika*, *81*, 535–549.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322.

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research and Human Genetics*, *5*, 554–571.

R Development Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org

Rutter, M., & Silberg, J. (2002). Gene–environment interplay in relation to emotional and behavioral disturbance. *Annual Review of Psychology*, *53*, 463–490.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype→ environment effects. *Child Development*, 424–435.

van der Sluis, S., Dolan, C. V., Neale, M. C., & Posthuma, D. (2008). Power calculations using exact data simulation: A useful tool for genetic study designs. *Behavior Genetics*, *38*, 202–211.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences of the United States of America*, *6*, 320–332.

Zavala, C., Beam, C. R., Finch, B. K., Gatz, M., Johnson, W., Kremen, W. S., Neiderhiser, J. M., Pedersen, N. L., & Reynolds, C. A. (2018). Attained SES as a moderator of adult cognitive performance: Testing gene–environment interaction in various cognitive domains. *Developmental Psychology*, *54*, 2356–2370.

https://doi.org/10.1017/thg.2020.76 Published online by Cambridge University Press