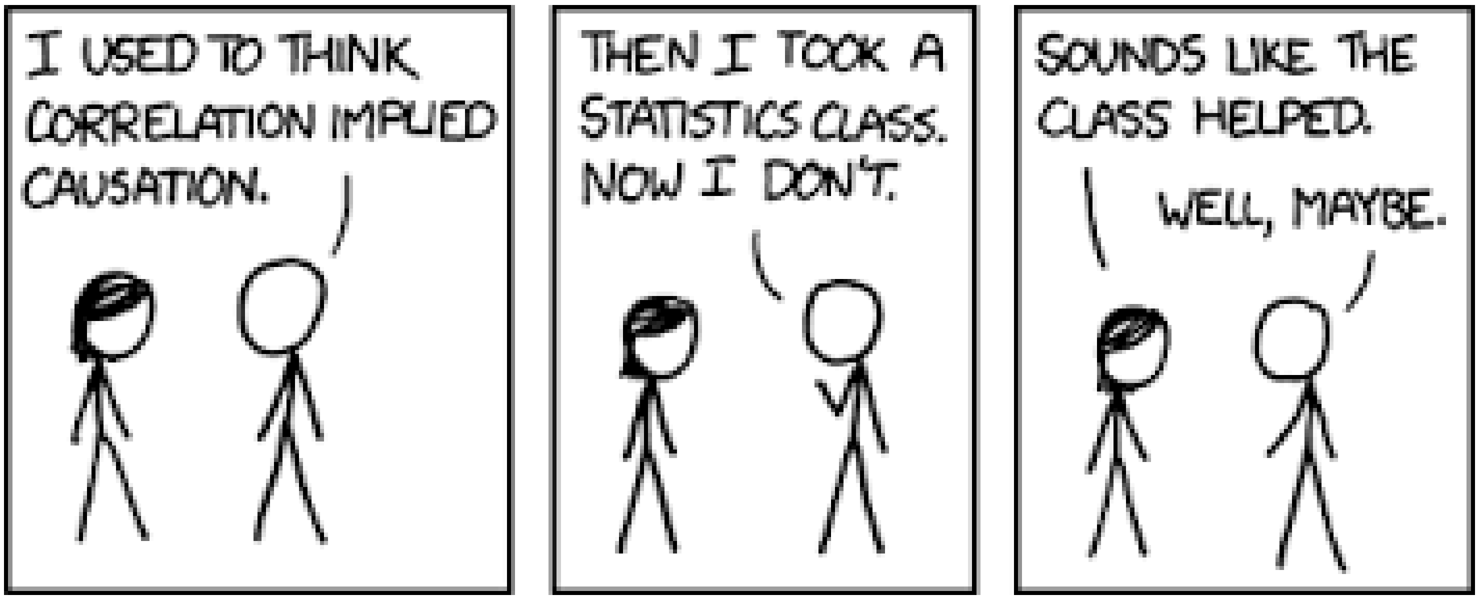
Testing *causality* in twin and family studies

Antti Latvala

26.5.2015


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Topics

Causality and confounding

How to test for causality in the presence of familial confounding:

1. Within-family analyses
2. Quantitative genetic modeling

Prolonged Breastfeeding Linked to Higher IQ and Wealth in Adulthood

TIME

- ✓ Several studies have linked longer duration of breastfeeding with better cognitive ability in the offspring
- ✓ Anderson ym. *Am J Clin Nutr,* 1999: Meta-analysis confirmed the association
- ✓ A dose-response relationship has been observed
- ✓ Breast milk contains long-chain saturated fatty acids that are essential for brain development

Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from Brazil

Cesar G Victora, Bernardo Lessa Horta, Christian Loret de Mola, Luciana Quevedo, Ricardo Tavares Pinheiro, Denise P Gigante, Helen Go Fernando C Barros

Lancet Glob Health 2015; 3: e199–205

Breast is best: human milk is the optimal food for brain development[1,2]

Ricardo Uauy and Patricio Peirano

Am J Clin Nutr 1999;70:433–4.

Breastfeeding leads to higher IQ, earnings later: study

AFP
MAR. 17, 2015, 8:57 PM | 635

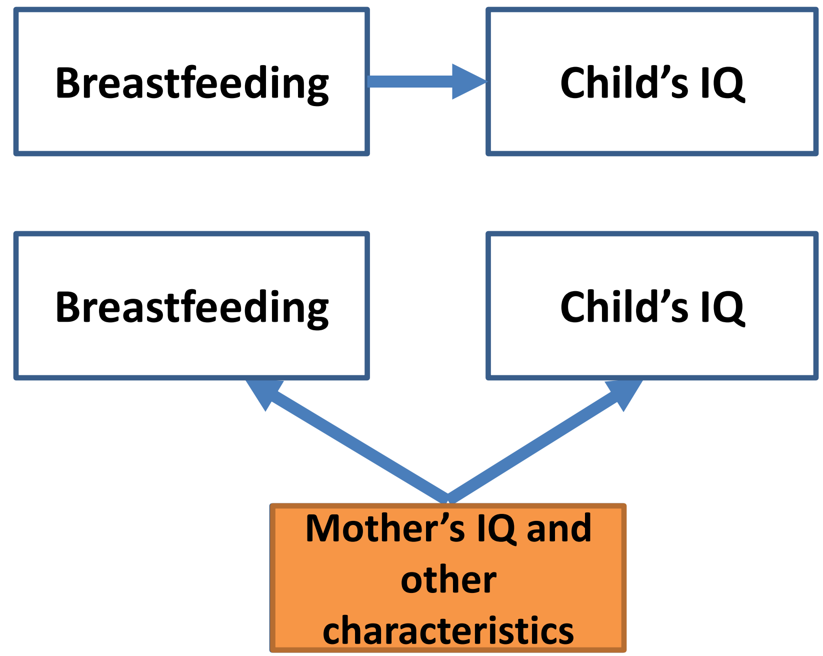
Causal or spurious?

A **confounder** is a factor which affects the outcome and correlates with the exposure (predictor)

Confounders can:

(1) Create a spurious association between exposure and outcome even though they do not have a real link

(2) Hide a real causal association



http://tylervigen.com/spurious-correlations

Causality and the experimental method

Causality is the relation between an event (**the cause**) and a second event (**the effect**), where the second event is understood as a consequence of the first

Cause: [from Latin *causa*: reason, purpose] The producer of an effect, result, or consequence

Causality can be understood through a **counterfactual model** :

Would the 'effect' have happened without the 'cause'?

To test causal hypotheses, we need to create approximations of the counterfactual situation

Experiment is the keystone of the empirical scientific method

- independent variables are manipulated to observe their effects on dependent variables
- researcher is in control of the treatment
- confounders are eliminated by random allocation of subjects into treatment and control groups
- randomized controlled trials (RCT) in medicine

L'esperienza della caduta dei gravi sul piano inclinato
G. Bezzuoli (1784-1855)

Causal relationships

According to a classic analysis formalized by John Stuart Mill (1806-1873), a causal relationship exists if:

1) *the cause was statistically related to the effect*
2) *the cause preceded the effect*
3) *we can find no plausible alternative explanation for the effect other than the cause*

So, to demonstrate that an association is causal, we should have

1) an association
2) the correct temporal order
3) **no plausible alternative explanations!**

Observational (=non-experimental) studies

Experiments are not feasible in

- *much of psychology and psychiatry*
- *sociology*
- *economics*
- *epidemiology*

THE TWO DISCIPLINES OF SCIENTIFIC PSYCHOLOGY[1]

LEE J. CRONBACH

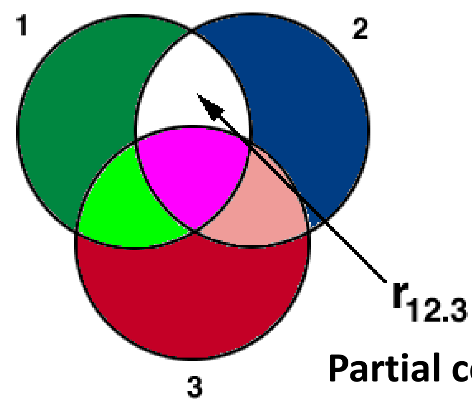
Cronbach 1957, *American Psychologist*

Epidemiology Faces Its Limits

SCIENCE • VOL. 269 • 14 JULY 1995

Statistical adjustment (using covariates) is often used in non-experimental studies to rule out confounding

Does breastfeeding predict children's intelligence when mothers' intelligence is controlled for?

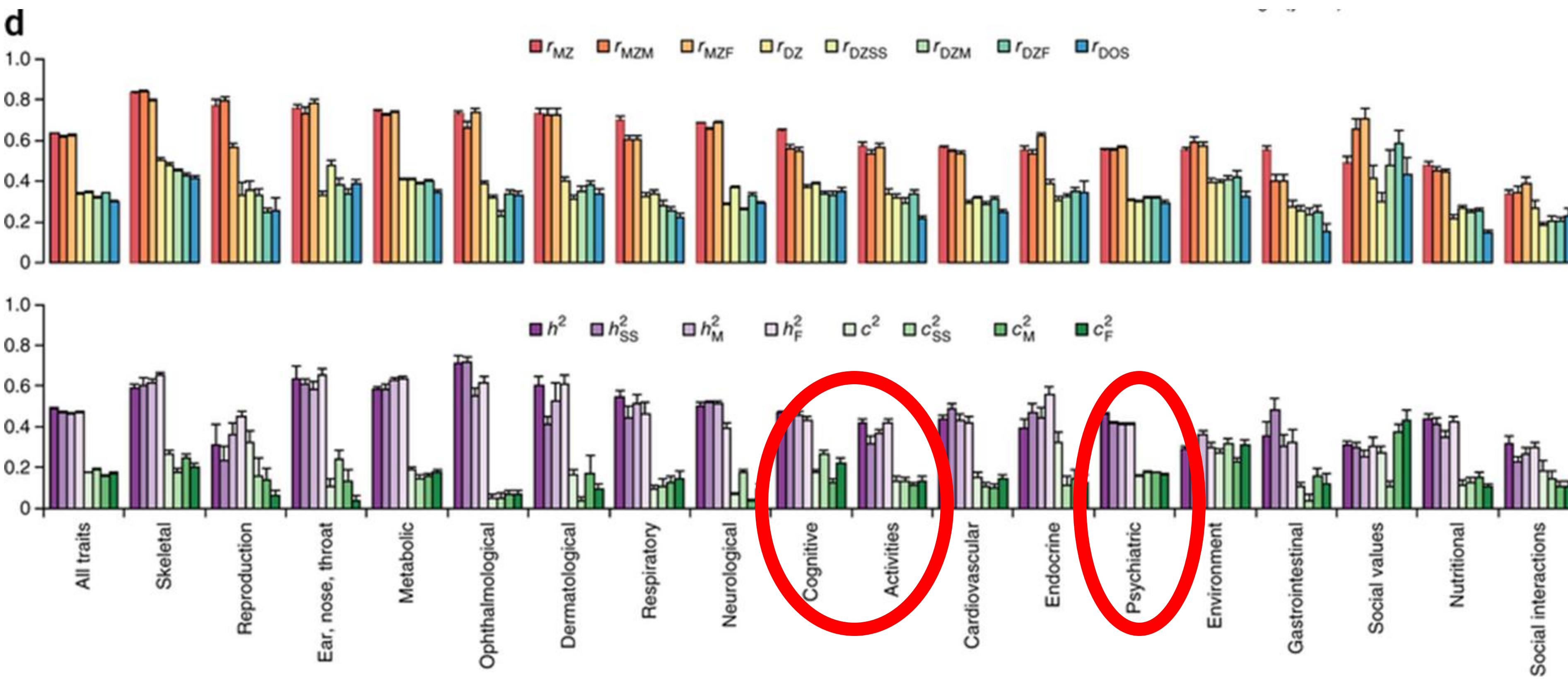


Partial correlation: correlation between 1 and 2 when correlations with 3 are removed

Adjustment problems

- *only measured variables can be adjusted for*
- *measurement error makes adjustment less efficient*
- *real causal relationships are assumed to be known*

Genetic differences explain ca. 50% of variance in behavioral and psychological traits



Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman[1,10], Beben Benyamin[2,10], Christiaan A de Leeuw[1,3], Patrick F Sullivan[4–6], Arjen van Bochoven[7], Peter M Visscher[2,8,11] & Danielle Posthuma[1,9,11]

NATURE GENETICS ADVANCE ONLINE PUBLICATION

Familial confounding

Genetic differences between individuals explain a large proportion of individual differences in behavior and psychological traits!

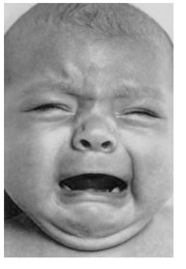
Same genetic variants can influence several traits (*pleiotropy)*

→ *Genetic background may be a confounder in many observed associations!*

Also shared environmental factors may be confounders (e.g. SES)

Gene–environment correlations

- × Passive: Parents provide both genes and rearing environment
- × Active / evocative: An individual's genetic makeup influences her characteristics which, in turn, influence exposure to different environments




Types of associations

The studied **predictor** variables (exposure, risk factor) can be:

1) External to the individual

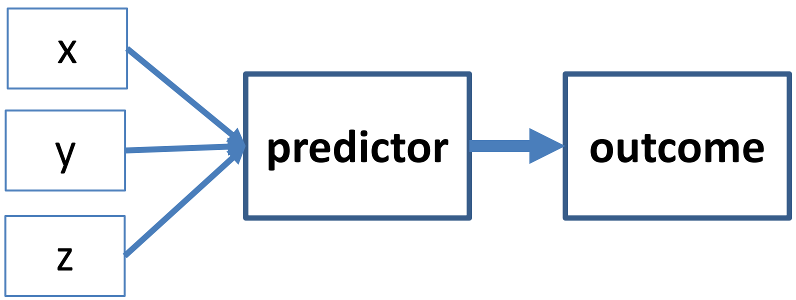
- *prenatal environment*
- *rearing environment, parental behavior*
- *larger environmental context*
- *chemical exposures etc.*

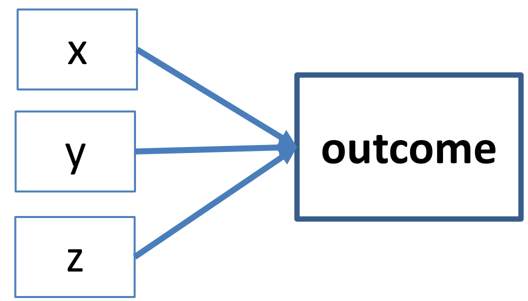
2) Characteristics of the individual

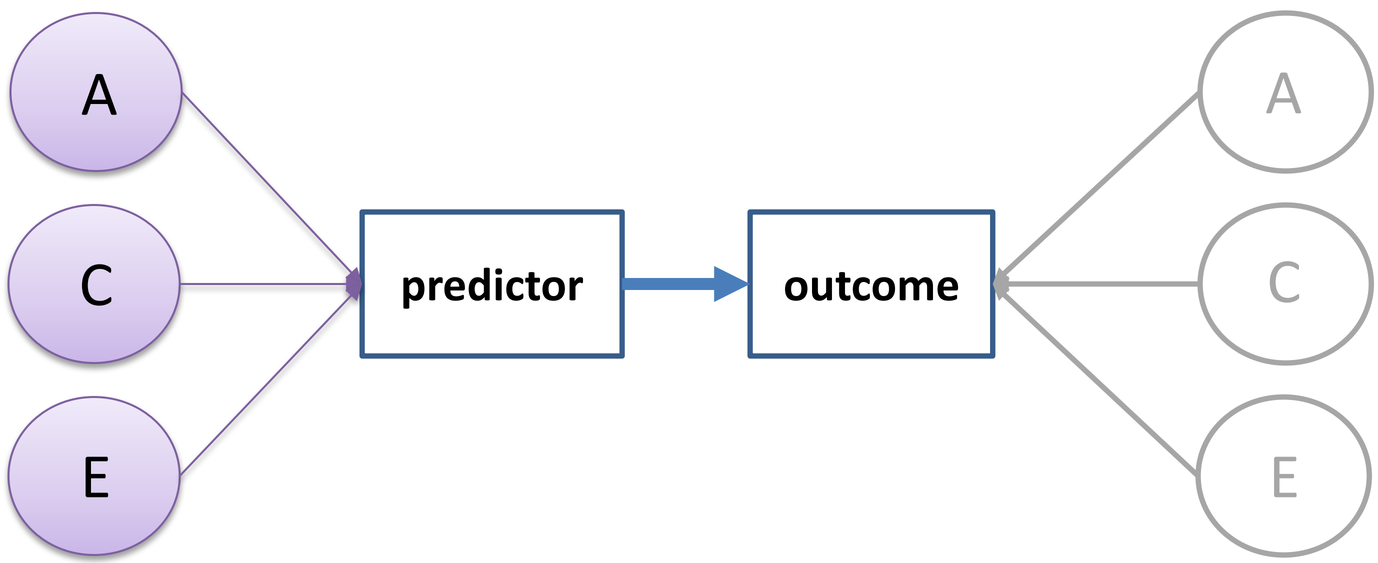
- *psychological traits (cognitive ability, personality)*
- *behavior (smoking, exercise)*
- *physical / physiological measures (obesity, blood pressure etc.)*

Causality in the context of ACE influences

If there is a causal association between the predictor and the outcome, all factors influencing the predictor should also have an effect on the outcome!







Topics

Causality and confounding

How to test for causality in the presence of familial confounding:

1. Within-family analyses
2. Quantitative genetic modeling

Co-twin control method

Mz twins are genetically identical

Differences in the outcome, associated with differences in exposure, are not confounded by genetic background

Co-twins reared together share environmental factors

How to compare:

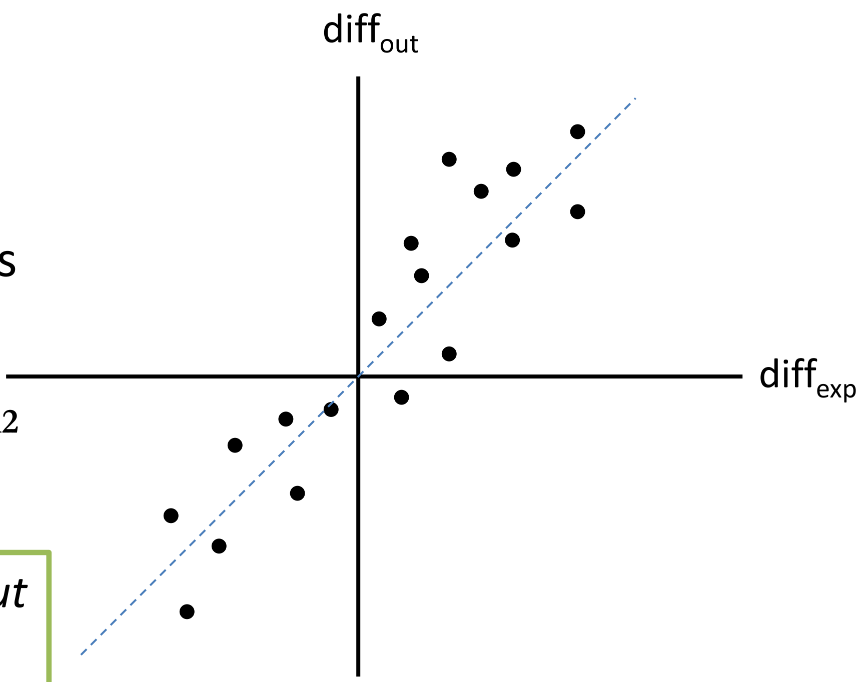
- Mean / prevalence in the exposed vs. unexposed co-twins within twin pairs
- Correlation / regression of difference scores of exposure and outcome within twin pairs

$$\mathbf{diff_{exp} = exposure_{twin1} - exposure_{twin2}}$$

$$\mathbf{diff_{out} = outcome_{twin1} - outcome_{twin2}}$$

Comparison of MZ co-twins has the most power but in practice DZ twins are often included to increase sample size

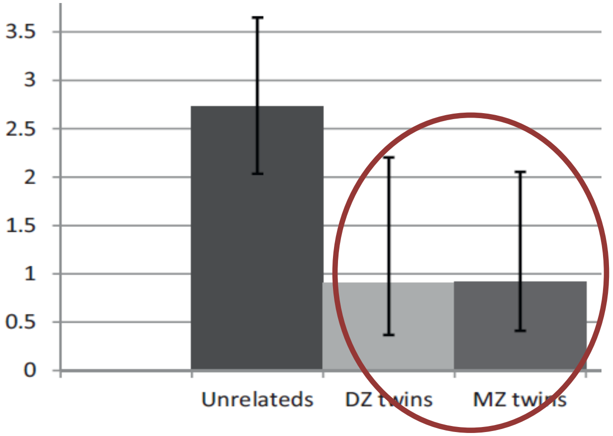
Japanese twin girls, Osaka 2014



		Is there a significant association between exposure and outcome?			
Study and outcome	Exposure	Individual level (*N*)	Within twin pairs (*N*)	Within MZ pairs (*N*)	Within DZ pairs (*N*)
Kujala et al. (2002)					
Mortality	Exercise	Yes (15,904)	No (658)	No (155)	No (475)
Mortality	Smoking	Yes (15,904)	Yes (809)	Yes (166)	Yes (591)
Mortality	Heavy alcohol	Yes (15,904)	Yes (507)	Yes (105)	Yes (376)
McGue et al. (2007)					
Physical function	Social activity	Yes (4,731)	NR	Yes (70)	NR
Cognitive	Social activity	Yes (4,731)	NR	Yes (70)	NR
Depression	Social activity	Yes (4,731)	NR	Yes (70)	NR

Early-onset cannabis use and risk of drop-out from education

Verweij et al. 2013

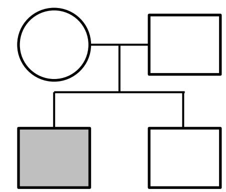


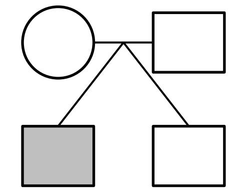
Low birth weight and autism (Losh et al. 2012)

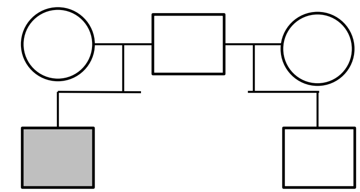
Pairs		ASD
Overall	No. discordant pairs *	34
	No. lighter child meeting cut-off	26
	OR (95% CI)	3.25‡ (1.47, 7.18)
MZ	No. discordant pairs *	7
	No. lighter child meeting cut-off	5
	OR (95% CI)	2.50 (0.49, 12.89)
DZ	No. discordant pairs *	19
	No. lighter child meeting cut-off	14
	OR (95% CI)	2.80† (1.01, 7.77)

In conditional logistic regression, a 100-gram increase in birth weight resulted in 13% reduced risk of ASD

Sibling comparisons







- ✓ Twins comprise ca. 2% of the population
- ✓ About one third of them are MZ
- ✓ The co-twin control method is a special case of sibling comparisons
- ✓ Comparing **full siblings** adjusts for 50% of genetic influences + all shared environmental influences
- ✓ **Half-siblings** share 25% of their genes and may also share environmental factors

- ➢ **Causality is supported by an association when familial influences are controlled for**
- ➢ **For genetic confounding the opposite is true:**

Population > Half-sibs > Full sibs > MZ twins

Comparisons can be made using:

- ○ Correlations of difference variables within pairs
- ○ ”Between-within” models, stratified models (e.g. linear mixed models, conditional logistic regression)

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij}$$

$$\boldsymbol{y_{ij} = \beta_0 + \beta_W(x_{ij} - \bar{x}_i) + \beta_B \bar{x}_i + \varepsilon_{ij}}$$

Within families effect

Between families effect

Effect of breastfeeding in siblings?

Effect of breast feeding on intelligence in children: prospective study, sibling pairs analysis, and meta-analysis

BMJ

October 2006

Geoff Der, G David Batty, Ian J Deary

	PIAT-total
	B (SE)
Unadjusted	4.69 (0.38)
Adjusted for:	
Mother's AFQT score	1.30 (0.36)
Mother's education	2.95 (0.37)

All significant at P<0.001

Breast feeding	PIAT-total	
	Difference (SE)	P value
Status†	–0.63 (0.94)	0.506
Duration‡	–0.13 (0.76)	0.866

Sibling Study Shows Little Difference Between Breast- and Bottle-Feeding

Alexandra Sifferlin @acsifferlin | Feb. 25, 2014

National Longitudinal Survey of Youth

- ✓ 3161 mothers, 5475 children
- ✓ 332 sib-pairs discordant for breastfeeding status
- ✓ 545 sib-pairs discordant for breastfeeding duration

Colen et a. 2014, *Social Science & Medicine*

- ✓ Extension of the previous study
- ✓ Breastfeeding predicted better outcome in 10/11 of the studied variables (e.g. school performance)
- ✓ The associations disappeared in sibling comparisons

Smoking during pregnancy and child development

Maternal smoking during pregnancy may cause:

Pre-term birth

Low birth weight and length

Child mortality

~~Poor cognitive development~~

~~ADHD~~

~~Behavior problems~~

~~Substance abuse~~

~~Obesity~~

~~Self-destructiveness~~

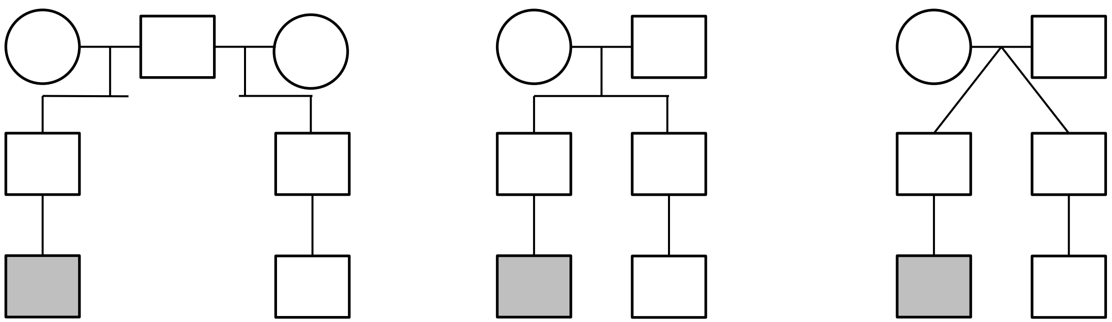
Johansson ym. 2009, *Epidemiology*
Lambe ym. 2006, *Epidemiology*
Lundberg ym. 2010, *Paediatr Perinat Epidemiol*
Iliadou ym. 2010, *Int J Epidemiol*
Agrawal ym. 2008, *Nicotine Tob Res*
Kuja-Halkola ym. 2010, *Int J Epidemiol*
D'Onofrio ym. 2008, *Dev Psychopathol*
D'Onofrio ym. 2010, *Child Dev*
Gilman ym. 2008, *Am J Epidemiol*
D'Onofrio ym. 2012, *Arch Gen Psychiatry*
Obel ym. 2011, *Int J Epidemiol*
D'Onofrio ym. 2010, *Arch Gen Psychiatry*
Cnattigius ym. 2011, *Eur J Epidemiol*

Cousin comparisons

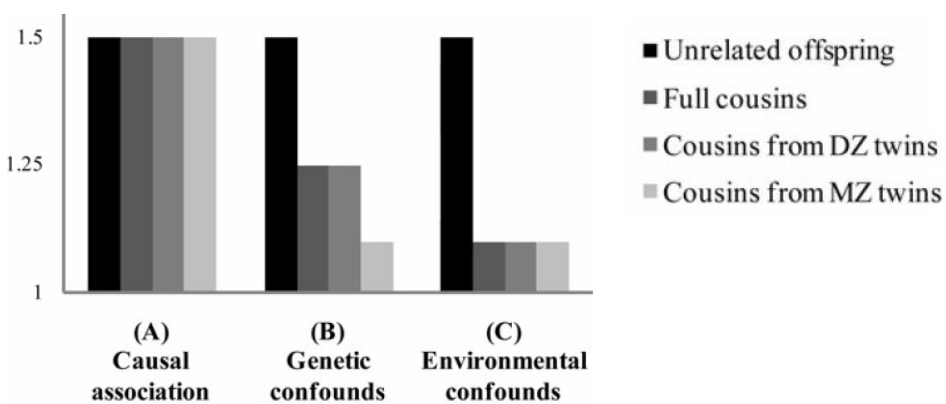
In sibling comparisons, the exposure has to differ between siblings

Many interesting exposures are shared by siblings, e.g. family SES, parental illness

By comparing cousins differing on the exposure, part of the genetic background can be controlled



6.25% shared genes **12.5% shared genes** **25% shared genes**



Cousin comparisons: Examples

Does parental **schizophrenia** affect offspring

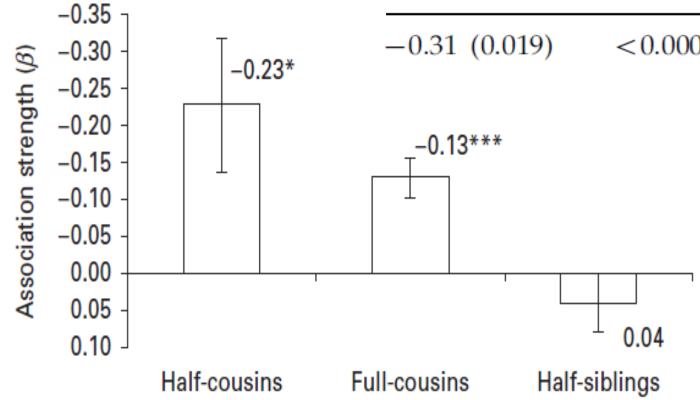
- School performance?
- Suicidality?

Parental relationship	Pairs	Odds ratio (95% CI)
Full-cousins	4285	1.96 (1.66–2.31)
Half-cousins	662	1.69 (1.17–2.44)

Odds ratio (95% CI)
Crude[a]
2.28 (2.10–2.47)

Ljung et al. 2013, *Psychol Med*

Estimate (standard error)	p
−0.31 (0.019)	<0.0001



Jundong et al. 2012, *Psychol Med*

Does parental **criminality** affect offspring

- Cognitive development?

full sample (N = 1,177,173)	Sons of half-siblings (n = 17,954)	Sons of full siblings (n = 204,226)	Sons of MZ twins (n = 1,348)
−0.53 [−0.54, −0.52]	−0.38 [−0.46, −0.29]	−0.22 [−0.25, −0.19]	0.14 [−0.18, 0.46]

Latvala et al. 2014, *Psychol Sci*

Assumptions and limitations of sibling and cousin comparisons (incl. co-twin control, children of twins)

Assumptions

- Generalizability to the population (e.g. also families without siblings/cousins, families with concordant exposures)
- No carry-over effects (e.g. exposure of sibling 1 should not affect outcome of sibling 2)

Limitations

- Do not rule out selection factors leading to differences between relatives (e.g. why are MZ co-twins discordant for X?)
- Sensitive to random measurement error which may reduce the within-pair associations even when there is no familial confounding
- Usually require large family datasets which can be time-consuming and expensive to collect

Topics

Causality and confounding

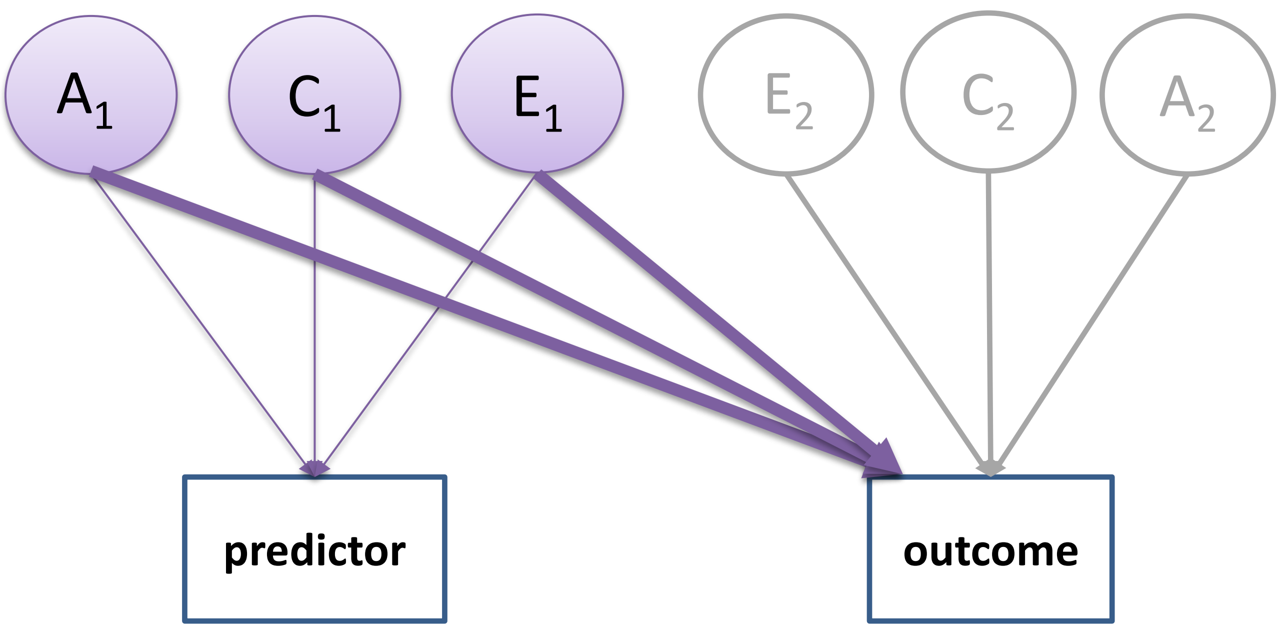
How to test for causality in the presence of familial confounding:

1. Within-family analyses

2. Quantitative genetic modeling

Bivariate Cholesky decomposition

- Assuming causality, causes of the predictor should cause changes in the outcome
- Cholesky decomposition can be used to test shared genetic and environmental variance
- The outcome may have specific ACE effects which are independent of the predictor

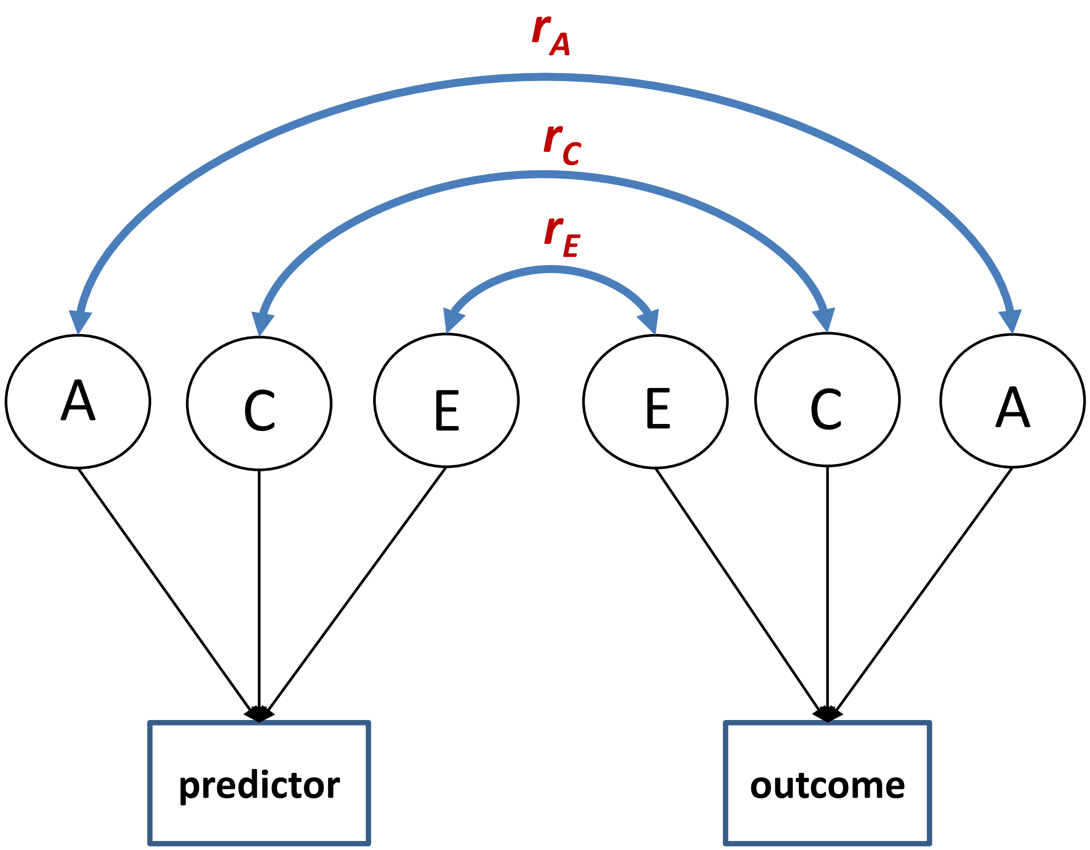


Cholesky decomposition: Genetic and environmental correlations

Overlap between genetic and environmental variance can be expressed as correlations between the latent variance components of the predictor and the outcome

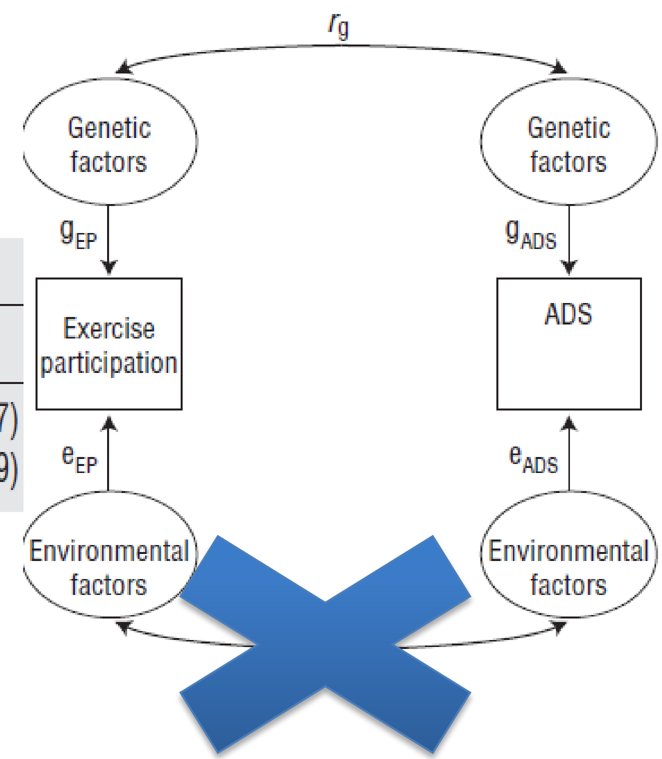
If the association is causal and the predictor has A, C, and E effects, there should be significant A, C, and E correlations as well

→ *Lack of* $\boldsymbol{r_E}$ *is often a sufficient indicator of a non-causal association*

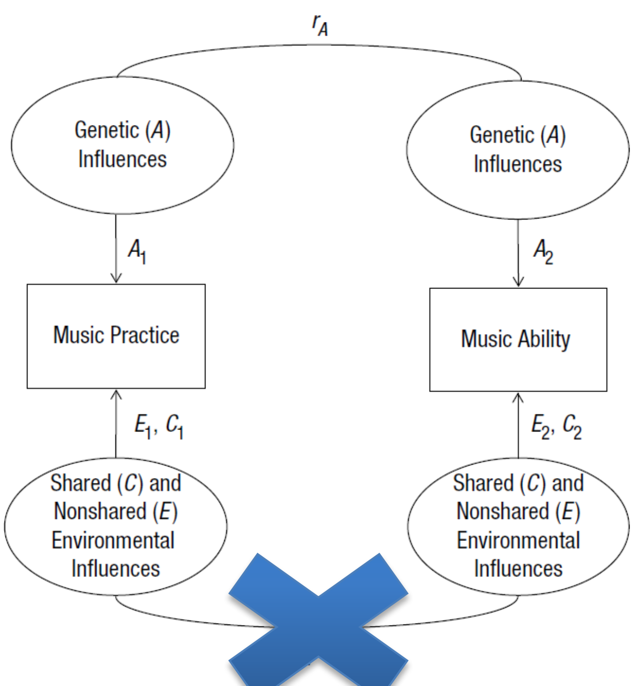


Physical exercise and symptoms of anxiety and depression (De Moor et al. 2008)

Genetic Correlation Estimate (95% CI) Between MET Hours and			
Depression	Anxiety	Somatic Anxiety	Neuroticism
−0.24 (−0.36 to −0.12)	−0.16 (−0.28 to −0.04)	−0.20 (−0.35 to −0.05)	−0.16 (−0.28 to −0.07)
−0.22 (−0.30 to −0.13)	−0.17 (−0.25 to −0.08)	−0.25 (−0.33 to −0.16)	−0.17 (−0.24 to −0.09)



Music practice and musical ability (Mosing et al. 2014)



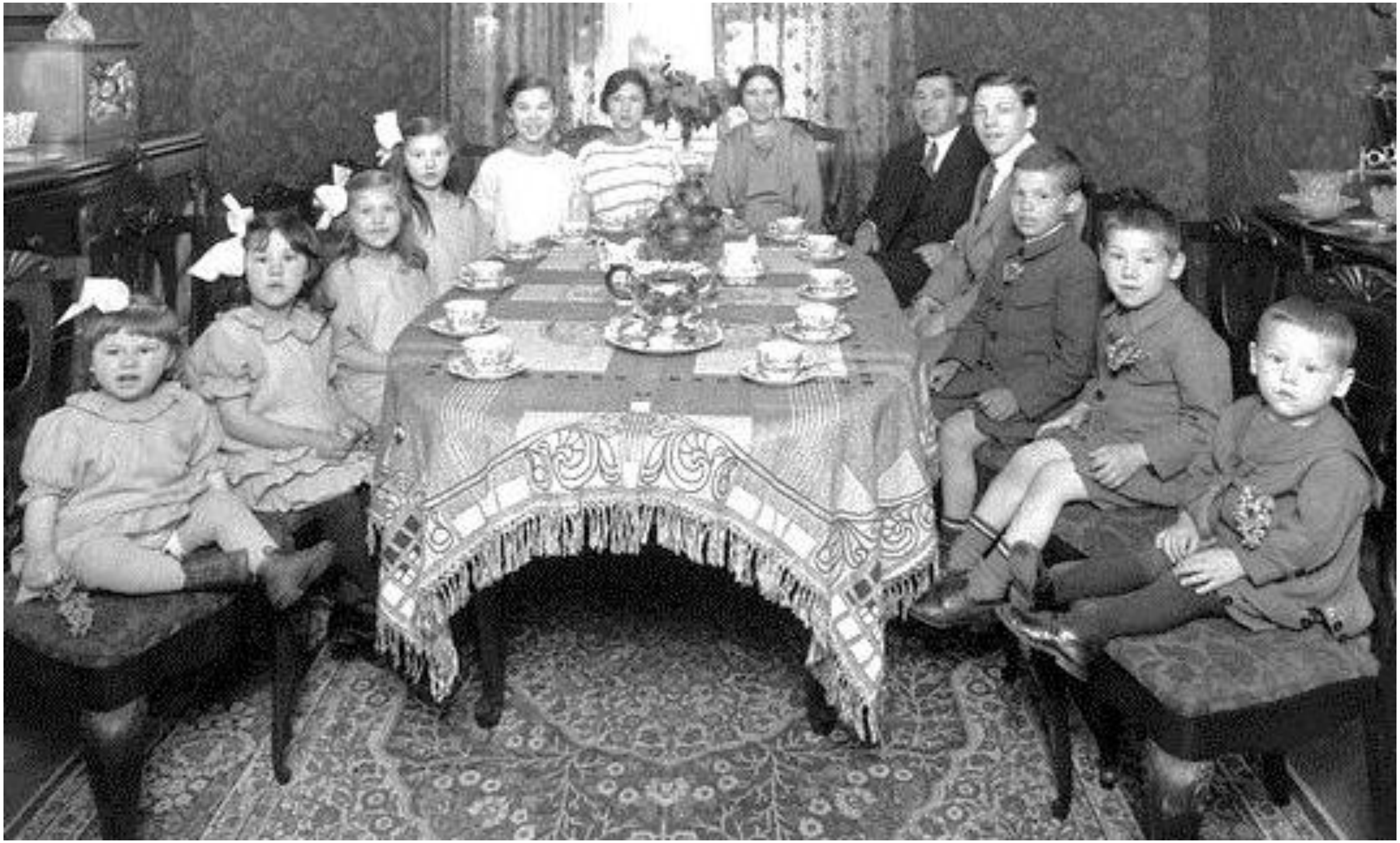
Sex and model	*df*	AIC	−2LL	Δ−2LL	*p*
		Melody			
Females					
Saturated	16,962	12,077.94	46,001.94	—	—
$r_C = 0$	16,963	12,077.50	46,003.50	1.56	.21
$r_E = 0$	16,963	12,075.94	46,001.94	0.00	1.00
$r_A = 0$	16,963	12,091.62	46,017.62	15.68	< .01
Males					
Saturated	16,962	12,077.94	46,001.94	—	—
$r_C = 0$	16,963	12,082.98	46,008.98	7.03	< .01
$r_E = 0$	16,963	12,076.55	46,002.55	0.61	.43
$r_A = 0$	16,963	12,093.15	46,019.15	17.21	< .01

Summary

- Importance of familial confounding
- Co-twin control is a powerful design but has also limitations
- Sibling and cousin comparisons are better than nothing!
- Quantitative genetic modeling can be used as another test of causality and it gives more comprehensive information

- All these methods are good at detecting familial confounding but they ***cannot prove causality****!*

Thank you for your attention!



Kauppias Ymär Abdrahim perheineen kotonaan Helsingissä v. 1925
Kuva: Eric Sundström

Further reading:

McGue et al. (2010) Causal inference and observational research: The utility of twins. *Perspect Psychol Sci* 5: 546-556.

Begg & Parides (2003) Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data. *Statist Med* 22: 2591-2602.

Carlin et al. (2005) Regression models for twin studies: a critical review. *Int J Epidemiol* 34: 1089-1099.

Frisell et al. (2012) Sibling comparison designs: Bias from non-shared confounders and measurement error. *Epidemiology* 23: 713-720.

De Moor et al. (2008) Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Arch Gen Psychiatry* 65: 897-905.

D'Onofrio et al. (2013) Critical need for family-based, quasi-experimental designs in integrating genetic and social science research. *Am J Public Health* 103: S46-S55.