

Research question

Power to detect a difference in correlations between samples of identical (monozygotic) and non-identical (dizygotic) twin pairs



Background

Let's assume that...

- ▶ each pair of twin siblings share the same environmental history
- ▶ zygosity doesn't influence environmental interactions
- ▶ trait is normally distributed
- ▶ accurate classification of zygosity
- ▶ accurate measurement of trait
- ▶ No dominant genetic variance (where one allele reliably masks the expression of the other when the two co-occur at a particular locus)
- ▶ mz and dz twins have been sourced from equivalent populations

Background

Let x_{zjk} be the value of trait x in the k th member of the j th twin pair having zygosity $z \in \{\text{mz, dz}\}$

$$(x_{z.1}, x_{z.2}) \sim \mathcal{N}(\mu, \Sigma)$$

$$r_z = \hat{r}_z = \frac{\text{Cov}(x_{z.1}, x_{z.2})}{\hat{\sigma}_{x_{z.1}} \hat{\sigma}_{x_{z.2}}}$$

$$\hat{\delta} = r_{mz} - r_{dz}$$

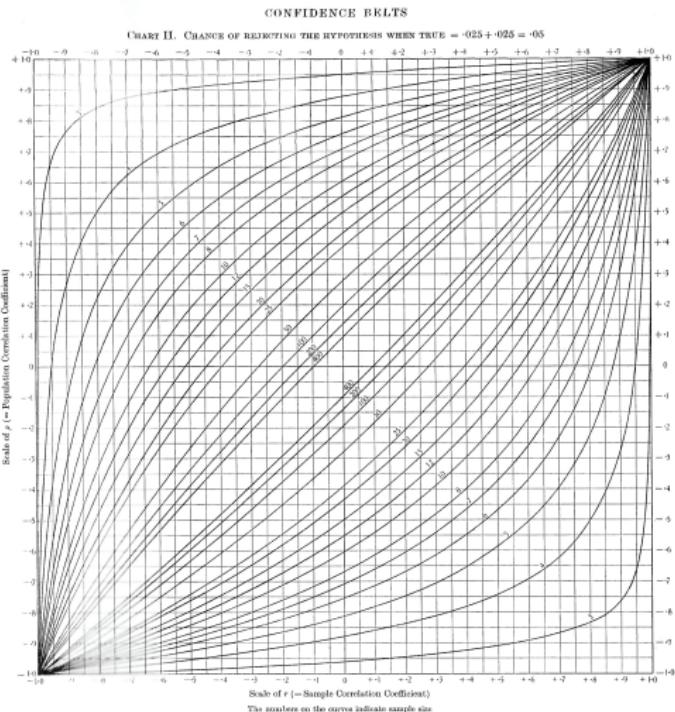
$$\text{heritability} = 2\delta$$



Trait etiology	Correlation in twin pairs	
	Monozygotic (mz)	Dizygotic (dz)
genetic	$r_{mz} = 1$	$r_{dz} = 0.5$
shared env.	$r_{mz} = 1$	$r_{dz} = 1$
individual env.	$r_{mz} = 0$	$r_{dz} = 0$
combination	$0 < r_{mz} < 1$	$0 < r_{dz} < 1$

A skip through the literature

- ▶ Galton, F. 1888
- ▶ Pearson, K. 1895
- ▶ Fisher, RA. 1915
- ▶ David, FN. 1938
- ▶ Falconer, DS. 1960
- ▶ Neale, M. et al. 1994
- ▶ Visscher, PM. 2004
- ▶ Verhulst, B. 2017



Power for difference in correlations¹

Frequentist NHST paradigm

α (Type I error) and β (Type II error); Power is $1 - \beta$

Fisher's Z formula approach

$$\text{power} = 1 - \Phi\left(\Phi_{\alpha/2}^{-1} - \text{abs}\left(\frac{\text{arctanh}(r_{mz}) - \text{arctanh}(r_{dz})}{\sqrt{(n_{mz}-3)^{-1} + (n_{dz}-3)^{-1}}}\right)\right)$$

Simulation approach

- ▶ implement hypothesis test(s) for difference in correlations
- ▶ take m draws from simulated bivariate mz and dz populations
- ▶ run hypothesis test(s) on each pair of draws
- ▶ power given parameters is proportion of tests where $p < \alpha$

¹Cohen1988.

Simulations

Using r_{mz} and r_{dz} derived from simulated bivariate normal and non-normal populations with covariance $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, the following hypothesis tests were run:

- ▶ Fisher's Z:
$$\frac{\operatorname{arctanh}(r_{mz}) - \operatorname{arctanh}(r_{dz})}{\sqrt{(n_{mz}-3)^{-1} + (n_{dz}-3)^{-1}}}$$
- ▶ Generalized Variable Test (GVT)²
- ▶ Signed log-likelihood ratio test³
- ▶ Permutation test⁴
- ▶ Zou's confidence interval⁵

Plan 1000 simulations of all mz dz combinations of ρ and n using three distinct distributions and re-run using Spearman correlation (rank based, non-parametric)...

²Krishnamoorthy2014, KazemiJafari2016.

³DiCiccio2001, KazemiJafari2016.

⁴Efron1994.

⁵Zou2007.

Preliminary results: simulation times

$1000 \times 39^2 \times 7^2 \times 6 \times 3 \times 2 = 2,234,400,000$ simulations...

Optimisation approach⁶

- ▶ hand coded
- ▶ functions compiled using either JIT compiler or RCCP

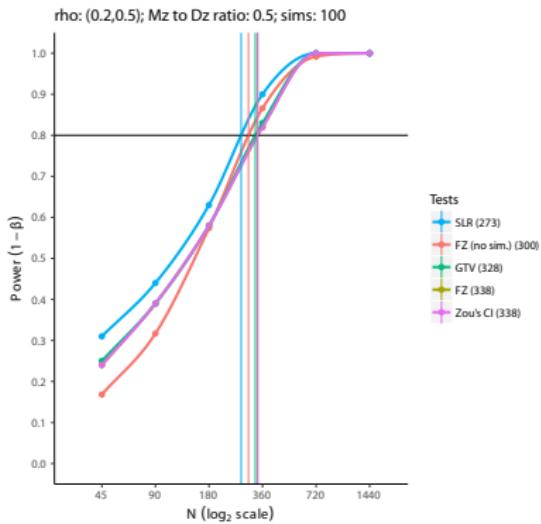
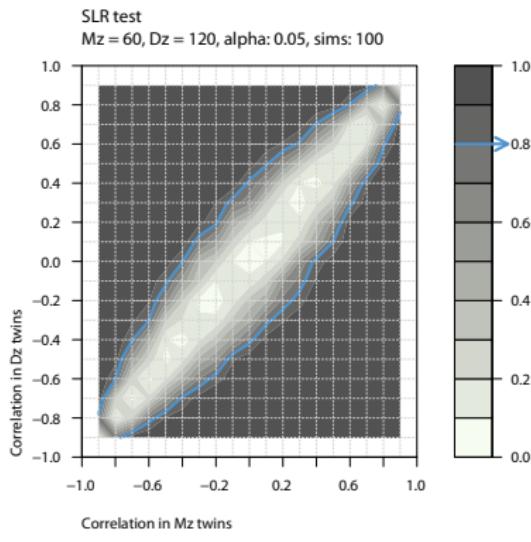
Test	Time/1000 runs (secs)	
	as is	compiled
Fisher's z (no sim)	0.03	0.02
Fisher's Z	0.56	0.36
GTV (r)	18.65	16.55
GTV (R C++)		12.08
Permutation (PT)	2473.18	2341.62
SLR	0.49	0.33
Zou's CI	0.38	0.28
Total (sans r-compiled GTV)	2493.29	2354.69
$\times 1520 \times 49 \times 3 \times 2 / 60 / 60 / 24 / 365$	35 years	33 years
$- PT : \times 361 \times 49 \times 3 \times 2 / 60 / 60 / 24$	25 days	16 days

Main findings

106,134 sets of power estimates comparing 5 tests (wide form)

Example

Power to detect $\hat{\delta}_\rho$ in Mz and Dz twins $\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$

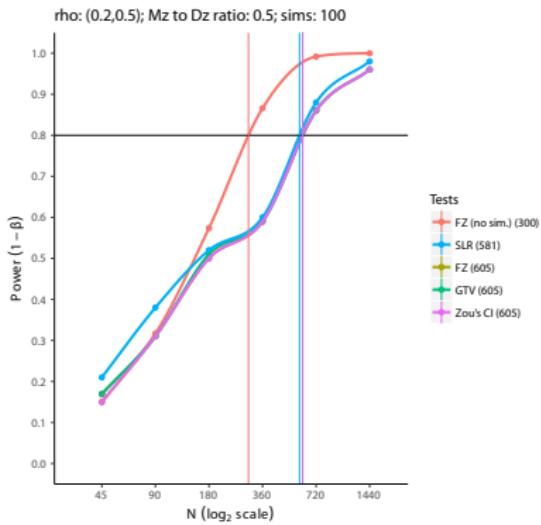
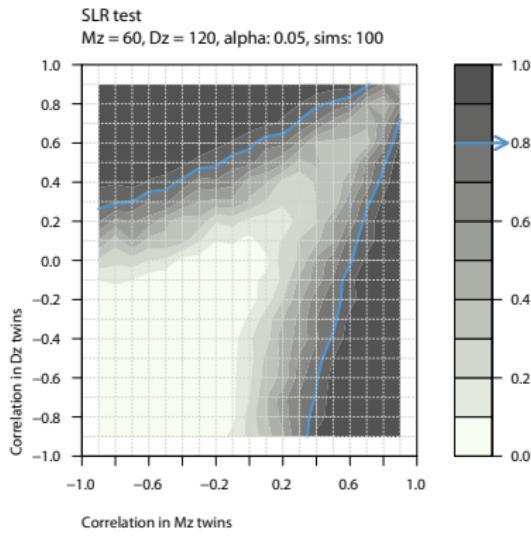


Main findings

106,134 sets of power estimates comparing 5 tests (wide form)

Example

Power to detect $\hat{\delta}_\rho$ in Mz and Dz twins $\sim \mathcal{G}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 & \rho \\ \rho & 5 \end{bmatrix}\right)$



Limitations

- ▶ Analysis not yet complete
- ▶ 100 simulations for each parameter combinations
 - ▶ 1000 is currently processing
- ▶ ICC not currently implemented
- ▶ Partial correlations are important, and not yet covered
 - ▶ incorporation of evaluation of correlation using mixed effects regression approaches could address above two issues
- ▶ Included tests are not exhaustive
- ▶ Code could be more efficient, allowing higher resolution estimation
 - ▶ e.g. original intent of -0.95 through 0.95 in 0.05 increments

Sum up and next steps

Secure | https://anaestheteteck.shinyapps.io/corr_power_app/

Apps LMS Databases Data analysis Careers Health Uni Misc play CIS - Current Index Catalogue : Library Statistical Analysis WolframAlpha: Com

Power to detect a difference in Pearson correlations in Mz and Dz twins

Scenario

Maximum N:

Ratio of Mz to Dz twin:

Corr(x,y) in Mz twins:

Corr(x,y) in Dz twins:

