Power to detect a difference in correlations

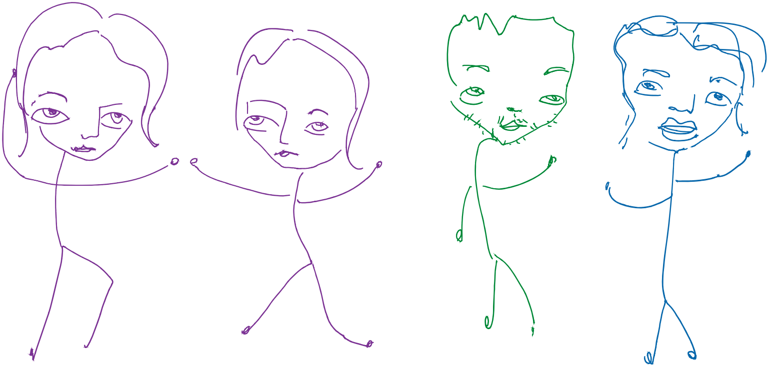
Applications for twin studies

Carl Higgs

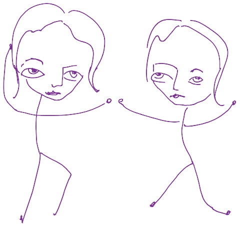
Centre for Epidemiology and Biostatistics
School of Population and Global Health
University of Melbourne

Supervisors: Enes Makalic, Elasma Milanzi and Katrina Scurrah

How much power do we have to detect a difference
in correlations between samples of
identical (monozygotic, MZ) and non-identical (dizygotic, DZ)
twin pairs?



Let x_{zjk} be the value of trait x in the kth member of the jth twin pair having zygosity $z \in \{\mathrm{MZ}, \mathrm{DZ}\}$



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

$$r_z = \hat{\rho}_z = \frac{Cov(x_{z\cdot 1}, x_{z\cdot 2})}{\hat{\sigma}_{x_{z\cdot 1}} \hat{\sigma}_{x_{z\cdot 2}}}$$

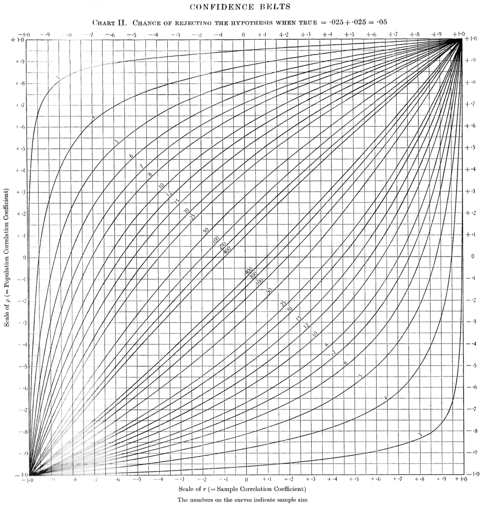
$$\hat{\delta}_r = r_{MZ} - r_{DZ}$$

$$\text{heritability} = 2\delta_\rho$$

Trait etiology	**Correlation in twin pairs**	
	Monozygotic (MZ)	**Dizogotic (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	$0 < r_{MZ} < 1$	$0 < r_{DZ} < 1$

Plan

- Review literature
  - Twins
  - Correlations
  - Power
  - Simulations
- Analysis plan
  - Efficiency
  - Data structure
  - Outputs
- Simulations
- Analysis
- Write up
- Dissemination



Power for difference in correlations

Frequentist NHST paradigm

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

Fisher's Z formula approach (David, 1938; Cohen 1988)

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

Simulation approach

- implement hypothesis tests for difference in correlations
- M times
  - draw from simulated bivariate MZ and DZ populations
  - run hypothesis tests for difference in sample correlations
- power given parameters is proportion of tests where $p < \alpha$

Simulations

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

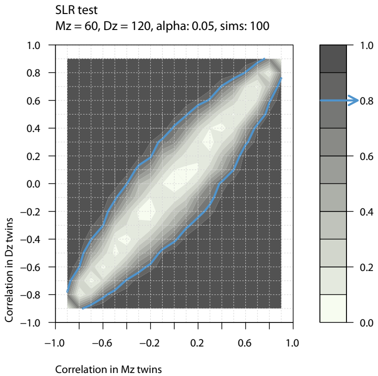
- Fisher's Z: $\frac{\text{arctanh}(r_{MZ}) - \text{arctanh}(r_{DZ})}{\sqrt{(n_{MZ}-3)^{-1} + (n_{DZ}-3)^{-1}}}$ (David, 1938)
- Zou's confidence interval (Zou, 2007)
- Generalized Variable Test (GVT; Krishnamoorthy & Lee 2014; Kazemi & Jafari, 2016)
- Signed log-likelihood ratio test (SLR; DiCiccio, 2001; Kazemi & Jafari, 2016)
- Permutation test (PT; Efron & Tibshirani, 1994)

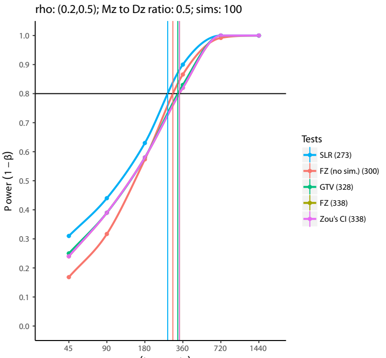
Plan 1000 simulations of all MZ DZ combinations of *rho* and *n* using three distinct distributions two correlation methods... 33 years? Fewer correlations, no PT test: 16 days

Main findings

- 530,670 power estimates comparing: tests, sample size, group size ratio, normality, correlation combinations, correlation methods
- Under approximate normality / mild skew, on average: SLR test $\sim$ 82% power; others $\leq$ 75%.

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

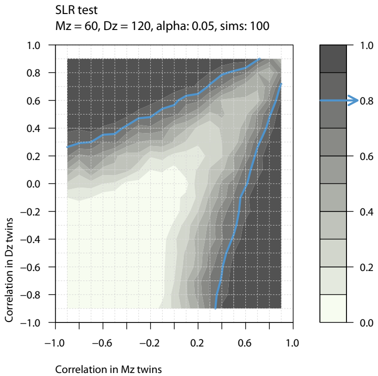


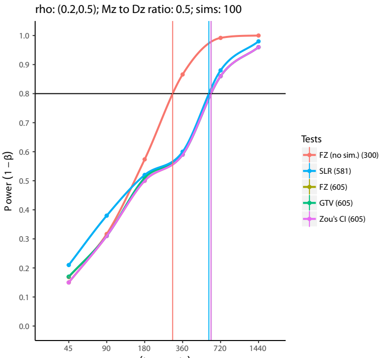


Main findings

- 530,670 power estimates comparing: tests, sample size, group size ratio, normality, correlation combinations, correlation methods
- Under extreme skew:
  Fisher's Z formula underestimates req. sample size by $\sim 50\%$

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





Strengths and Limitations

Achievements

- a flexible, extensible architecture
- solves future problems

But analysis is still in progress:

- 100 simulations for each parameter combinations
- 1000 is currently processing

Yet to be implemented:

- Intraclass correlations
- Partial correlations
  - incorporation of evaluation of correlation using mixed effects regression approaches could address above two issues
- Improve efficiency to allow higher resolution estimation

Sum up and next steps

We have developed an architecture for detecting differences in correlations and comparison of methods which is readily extensible for future applications (new tests, specific requests, new distributions, etc).