



Boosting statistical power by modelling phenotypic measurement



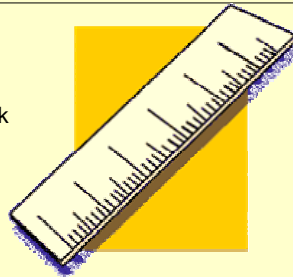
Stéphanie van den Berg¹, Cees Glas² & Dorret Boomsma¹ (1Vrije Universiteit Amsterdam, 2Twente University)

e-mail: SM.van.den.Berg@psy.vu.nl

BACKGROUND

Observation

Behavioural traits, as opposed to physical traits, often lack a standard unit of measurement and are consequently suboptimally measured, using behavioural counts or a limited set of items.



Problem

- Behaviour is often best described in probabilistic terms: individuals with a high true latent score generally have a high probability of exhibiting a particular behaviour.
- When this probabilistic aspect is ignored when analysing observed raw scores, correlations may be seriously attenuated.

Consequences

Correlations are underestimated, leading to poor statistical power and wrong conclusions.

Not including a measurement model may lead to

- Underestimated sibling correlations, with implications for heritability estimates and linkage results
- Underestimated genotype-behaviour correlations, with implications for association results

SOLUTION

To correct for attenuation, include a probabilistic measurement model of how observed data relates to an unobserved trait. Estimate this measurement model *concurrently* with the actual structural model (e.g., variance decomposition, association model, etc.) to obtain parameter estimates corrected for attenuation.

For example, use a **Rasch measurement model** to model binary responses (yes, no) to particular items that may differ in prevalence e.g. "Exhibits behaviour A" (yes, no), "Exhibits behaviour B", (yes, no), etc.

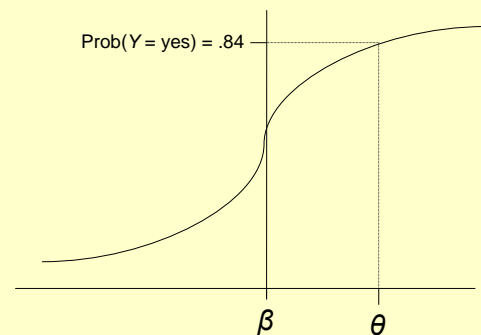
The Rasch model assumes that every individual j has a score on a latent variable, θ_j . Whether a behavioural item is scored yes or no, depends on a subject parameter θ and a parameter for the particular behaviour β :

$$\text{Prob}(Y_{ij} = \text{'yes'} | \theta_i, \beta_j) = \exp(\theta_i - \beta_j) / (1 + \exp(\theta_i - \beta_j))$$

The higher the subject parameter, and the lower the behaviour parameter, the more likely the subject shows this particular behaviour.

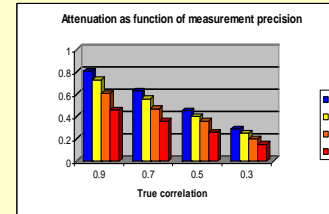
The Rasch model can be extended to include graded responses ('not at all', 'a little', 'a lot'), differential factor loadings, covariates, repeated measures, hierarchical structures (e.g., family data), modelling of missing data, and multidimensionality.

RASCH (1960) MODEL



SIMULATION OF ATTENUATION EFFECTS

Continuous bivariate normally distributed individual differences were simulated and behavioural sum scores are generated using a probabilistic model with varying degree of precision. For example, counting the number of times that genetically related mice press a lever with a max. of 5, 10, 20, or 50 opportunities (average probability of a response is 50% in the population). Or counting the number of different types of aggressive behaviours in identical twins during the preceding six months, from a total of either 5, 10, 20 or 50 descriptions of medium aggressive behaviour (average probability of showing a particular behaviour is 50% in the population).



True correlation	Number of opportunities/ behaviours			
	50	20	10	5
.90	.81	.73	.61	.46
.70	.63	.56	.47	.36
.50	.45	.40	.36	.26
.30	.29	.25	.20	.15

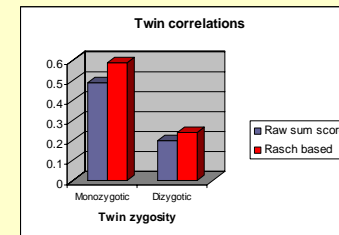
APPLICATION WITH DEPRESSION DATA



Method

Depressive symptoms were assessed thru self-reports in a non-clinical sample of 1370 identical (monozygotic) twin pairs and 1517 fraternal (dizygotic) twin pairs from the Netherlands Twin Register. We used the 17-item Anxious Depression scale from the Young Adult Self-Report (Achenbach, 1997).

Results



Based on raw scores, MZ twin correlation was **.49**, DZ twin correlation **.20**, suggesting a heritability estimate of **48%**

When however liability θ was modelled using a Rasch model, correlations on the latent trait were estimated at **.59** and **.24**, respectively, suggesting a heritability of **58%**

Variance of θ was concurrently decomposed into additive genetic variance and non-shared environmental variance while correcting for a sex effect on the latent trait. Inference is based on the posterior marginal distributions using an MCMC algorithm implemented in WinBUGS. Heritability was estimated at 55% (48-61%).

Conclusion: When you include a measurement model for your behavioural phenotype in the genetic modelling, you remedy attenuation effects and you boost your statistical power. This is even the case with 17 polytomous items. The fewer the items (i.e. the less precise your measurement) the greater the effect of employing a measurement model.