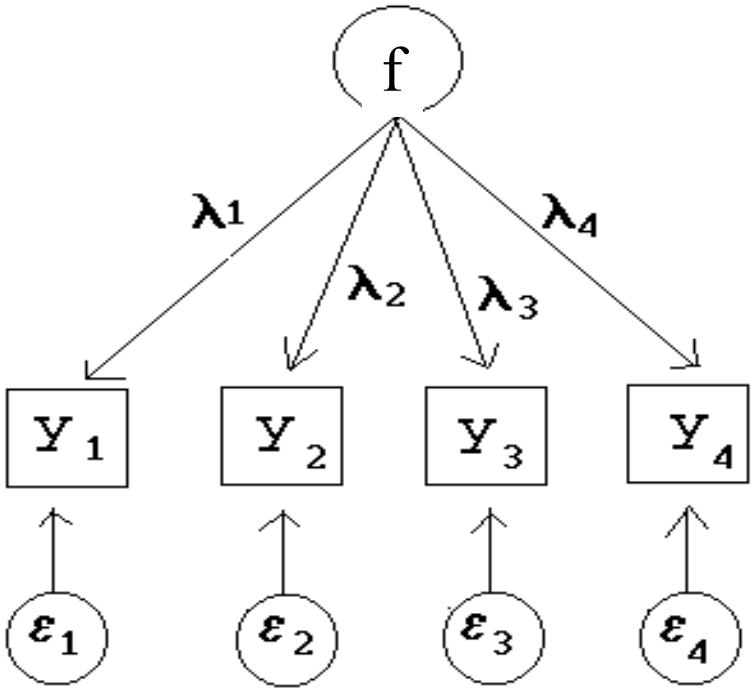
Development / longitudinal / repeated measures / time series / (not cross-sectional)

Dorret Boomsma
Eveline de Zeeuw
Conor Dolan
Netherlands Twin Register
Department of Biological Psychology
Vrije Univ., Amsterdam, The Netherlands

General multivariate model for 1 person: Regression of observed variables (y) on latent variables (f)



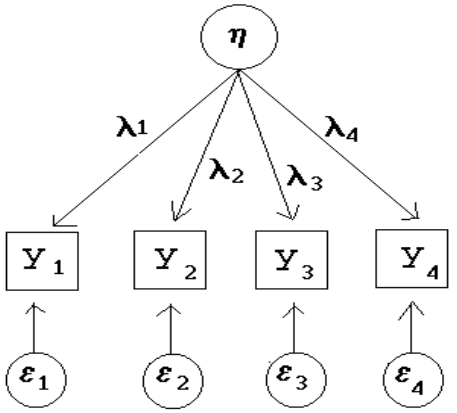
Covariance among variables is accounted for by a smaller number of latent factors.

Model: $y = \Lambda f + e$,

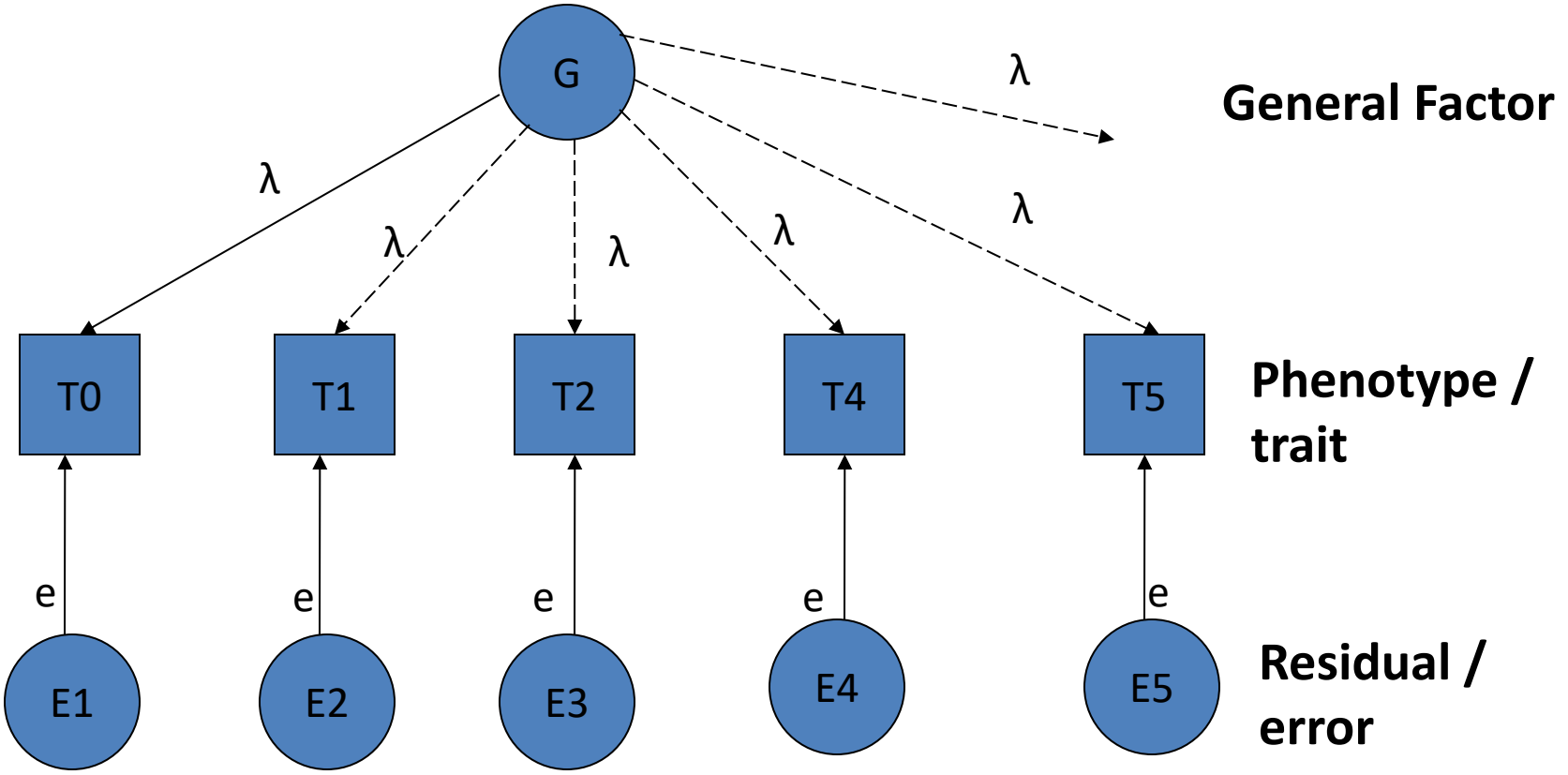
where y = observed variables
f = (latent) factor score(s)
e = unique factor / error
Λ = matrix of factor loadings

Structural equation models.

Sometimes x = Λ f + e is referred to as the measurement model, and relations among latent factors as the structural equation model.

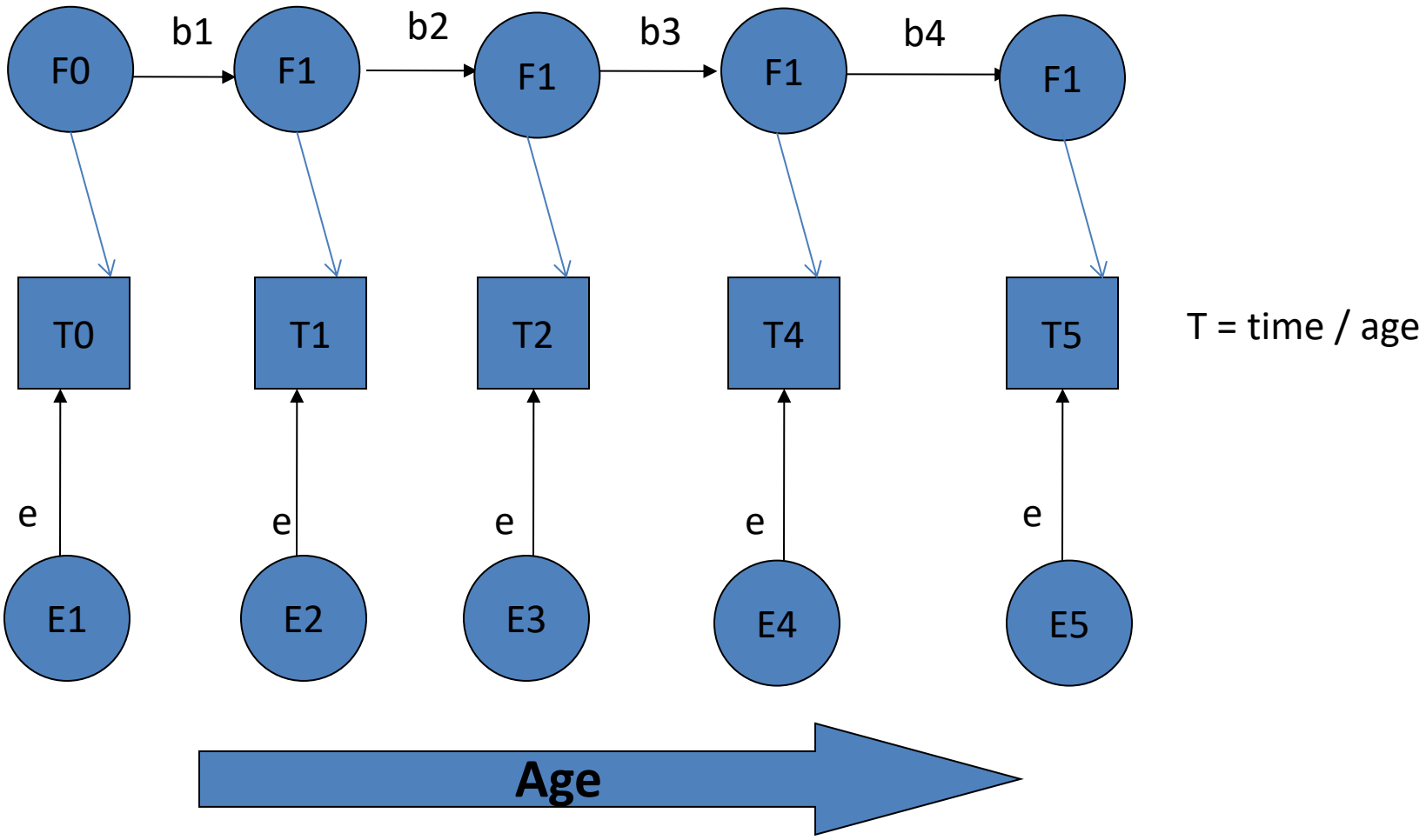


Multivariate model



If T= time and N of time points get large, what will happen?

Developmental model



Two general approaches to longitudinal modeling (not mutually exclusive)

Markov models:
(Vector) autoregressive models for continuous data
(Hidden) Markov transition models discrete data

Growth curve models:
Focus on linear and non-linear growth curves
Typically multilevel or random effects model

Which to use?
Use the model that fit the theory & hypotheses

Growth curve modeling ? If you're interested in growth trajectories. Linear or non-linear:

Twin Research (2000) 3, 165–177

www.nature.com/tr

Structured latent growth curves for twin data

Michael C Neale[1] and John J McArdle[2]

Autoregressive modeling ? If you're interested in e.g. the genetic stability of a trait (e.g combining data).

Psychological Medicine (2015), **45**, 1039–1049.
doi:10.1017/S003329171400213X

ORIGINAL ARTICLE

Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years

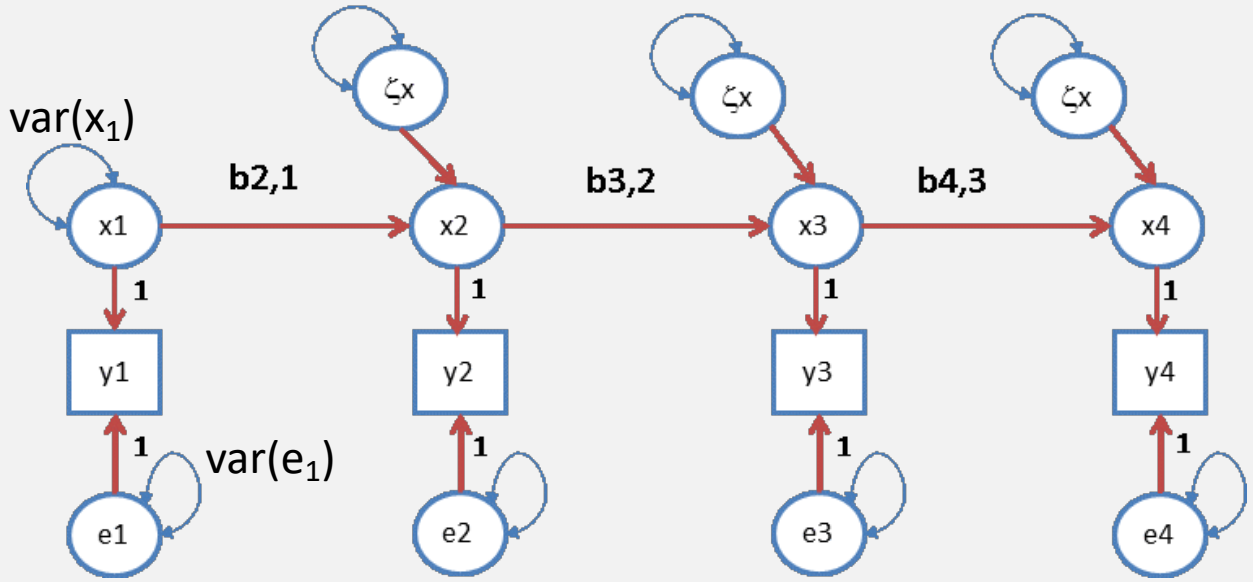
M. G. Nivard[1,2*], C. V. Dolan[1,3], K. S. Kendler[4], K.-J. Kan[1], G. Willemsen[1,5], C. E. M. van Beijsterveldt[1,5], R. J. L. Lindauer[6], J. H. D. A. van Beek[1,5], L. M. Geels[1,5], M. Bartels[1,5], C. M. Middeldorp[1,2,7†] and D. I. Boomsma[1,2,5†]

Do the Genetic or Environmental Determinants of Anxiety and Depression Change with Age? A Longitudinal Study of Australian Twins

Nathan A. Gillespie[1], Katherine M. Kirk[1], David M. Evans[1], Andrew C. Heath[2], Ian B. Hickie[3], and Nicholas G. Martin[1]

Can be combined (Nathan Gillespie)

First order autoregression model.



Innovation

True score

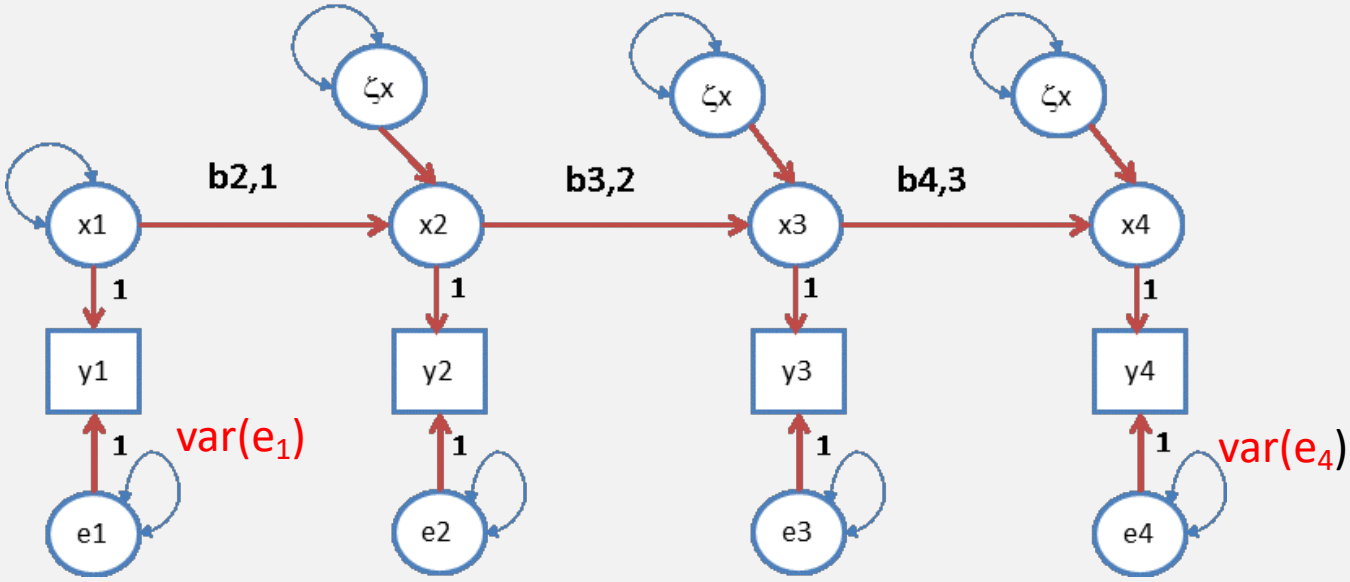
Observation

"error" / residual

$$y_{ti} = b_{0t} + x_{ti} + e_{ti} \quad \text{(measurement model)}$$

$$x_{ti} = b_{t-1,t}\, x_{t-1i} + \zeta x_{ti} \quad \text{(structural model)}$$

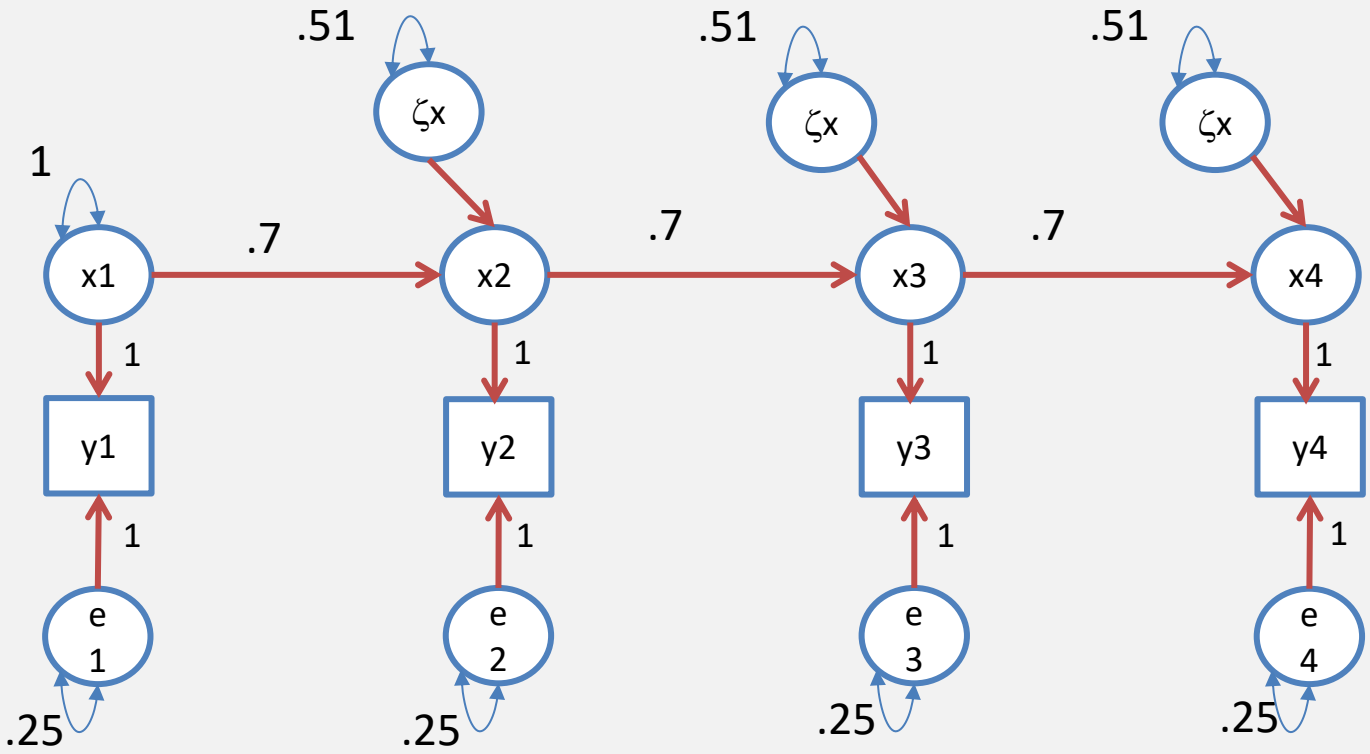
First order autoregression model.



Identification issue: var(e_1) and var(e_4) are not identified.

Solution set to zero, or equate :

var(e_1) = var(e_2) , var(e_3) = var(e_4)



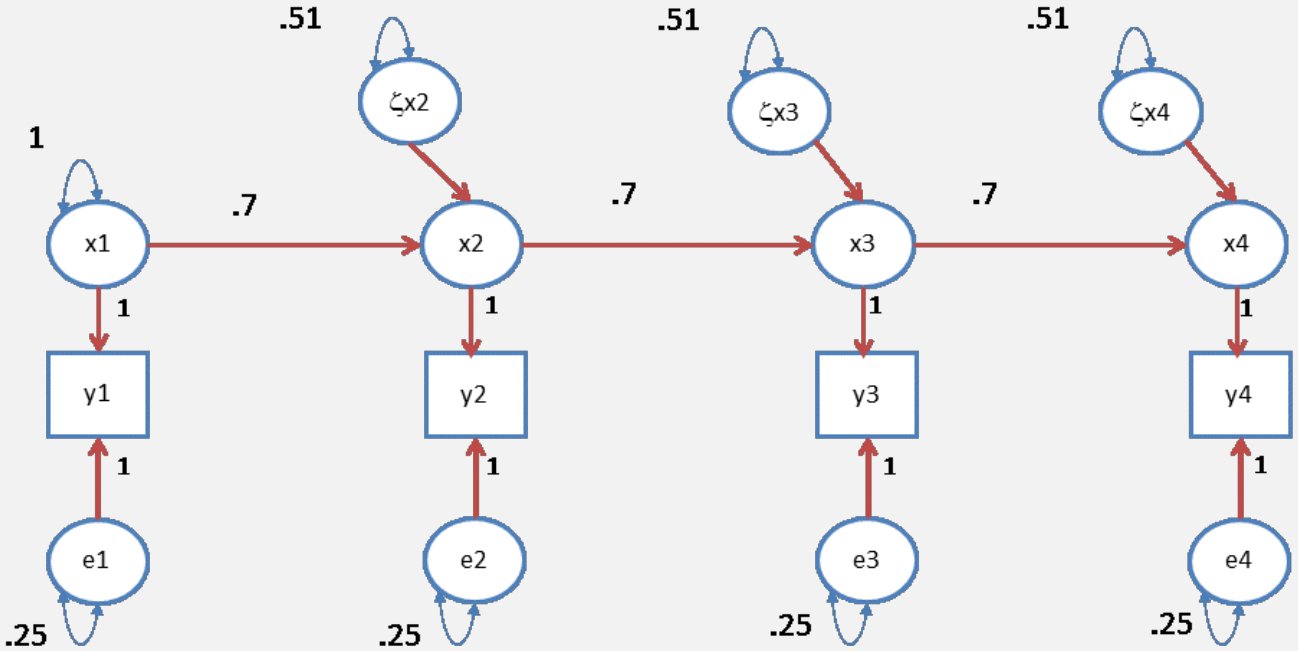
Covariance matrix

```
1.250 0.70 0.49 0.343
0.700 1.25 0.70 0.490
0.490 0.70 1.25 0.700
0.343 0.49 0.70 1.250
```

Correlation matrix

```
1.0000 0.560 0.392 0.2744
0.5600 1.000 0.560 0.3920
0.3920 0.560 1.000 0.5600
0.2744 0.392 0.560 1.0000
```

Cov y1,y2: (0.7*1 = 0.7; cov y1, y2 = 0.7*0.7*1 = 0.49, etc.



reliability: $rel(x_t) = var(x_t) / \{var(x_t) + var(e_t)\}$
$= 1/ (1+.25) = 1/1.25 = .8$

R^2: $b_{t-1,t}^2\, var(x_{t-1}) \;/ \{b_{t-1,t}^2\, var(x_{t-1}) + var(\zeta x_t)\}$
$= \{.7^2 * 1\} / (.7^2 * 1 + .51) = .49/1 = .49$

Correlation (t,t+1) : $b_{t-1,t}\, var(x_{t-1}) \;/ \{sd(y_{t-1}) * sd(y_t)\}$
$=\{.7 * 1\} / \{\sqrt{1.25} * \sqrt{1.25}\} = .56$

These models can be applied to phenoypic, genetic and environmental timeseries.

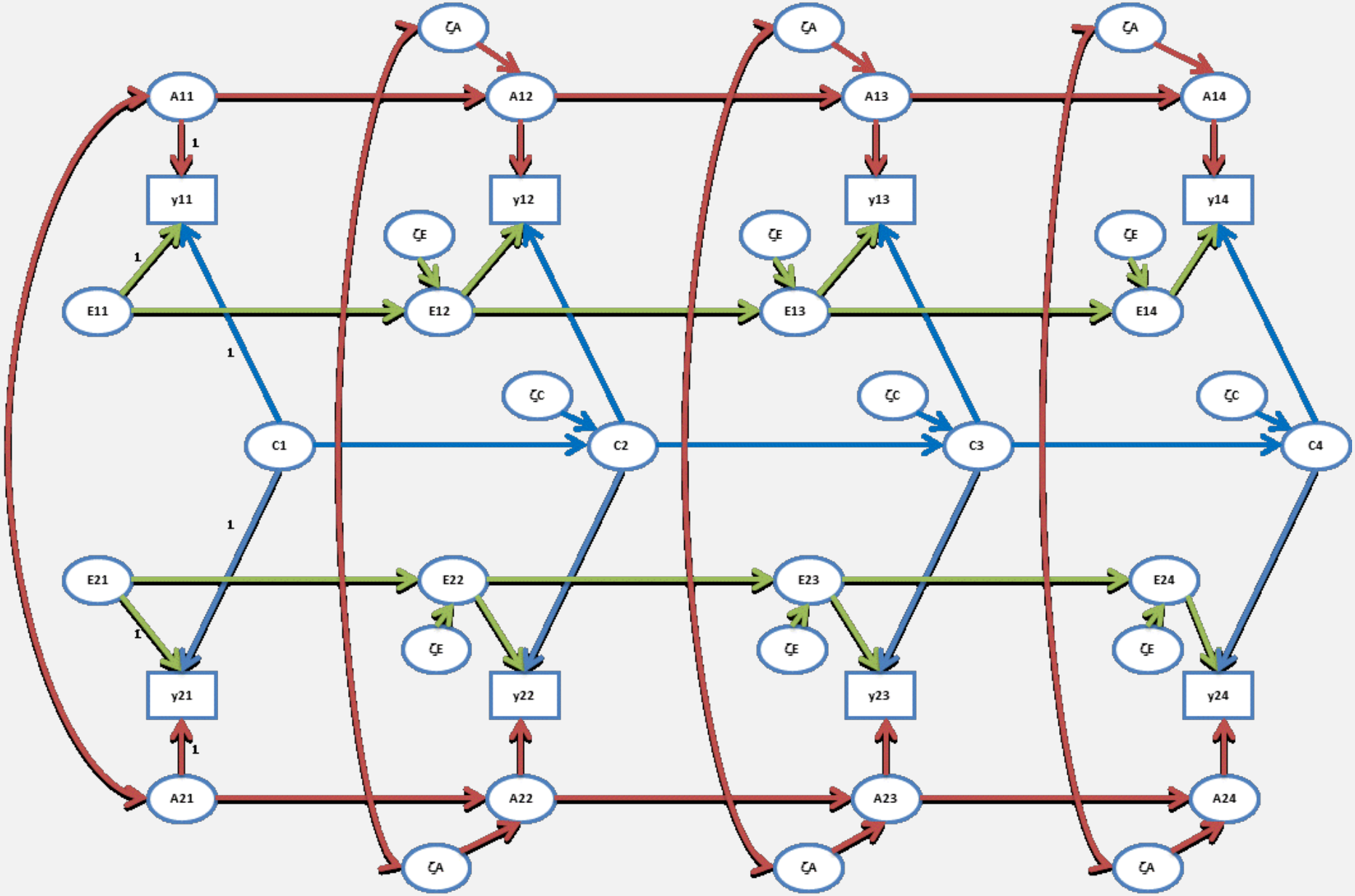
Estimation of genetic and environmental correlations

Animal / experimental studies

Bi/multivariate twin / family analyses

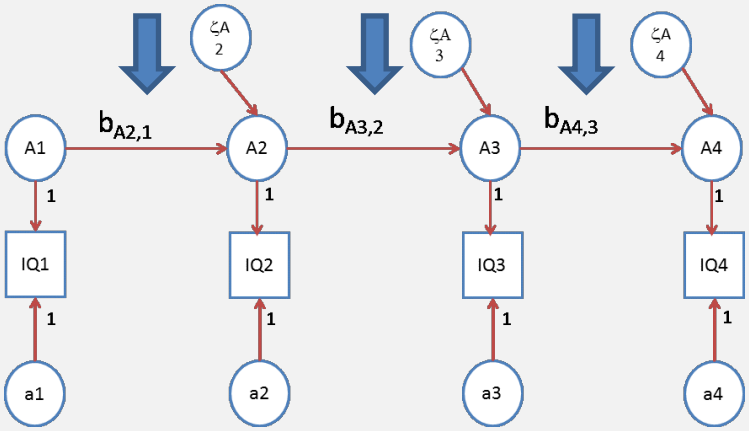
Bivariate SNP analyses

LD-score regression


ζA
ζA
ζA
A11
A12
A13
A14
1
y11
y12
y13
y14
ζE
ζE
ζE
1
E11
E12
E13
E14
1
ζC
ζC
ζC
C1
C2
C3
C4
1
E21
E22
E23
E24
1
ζE
ζE
ζE
y21
y22
y23
y24
1
A21
A22
A23
A24
ζA
ζA
ζA

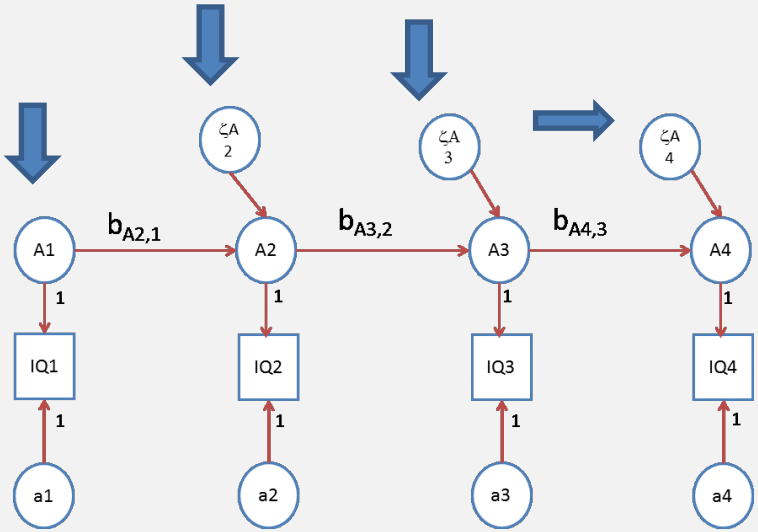
$$\Sigma_A = (I - B_A)^{-1} \Psi_A (I - B_A)^{-1\,t} + \Theta_A$$

$$B_A = \begin{matrix} 0 & 0 & 0 & 0 \\ b_{A21} & 0 & 0 & 0 \\ 0 & b_{A32} & 0 & 0 \\ 0 & 0 & b_{A43} & 0 \end{matrix}$$



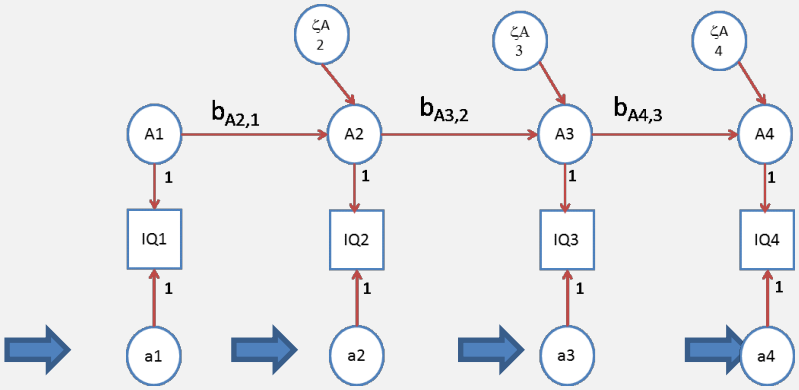
$$\Sigma_A = (I - B_A)^{-1} \Psi_A (I - B_A)^{-1\,t} + \Theta_A$$

$$\Psi_A = \begin{matrix} \text{var}(A_1) & 0 & 0 & 0 \\ 0 & \text{var}(\zeta_{A2}) & 0 & 0 \\ 0 & 0 & \text{var}(\zeta_{A3}) & 0 \\ 0 & 0 & 0 & \text{var}(\zeta_{A4}) \end{matrix}$$



$$\Sigma_A = (I - B_A)^{-1} \Psi_A (I - B_A)^{-1\tau} + \Theta_A$$

$$\Theta_A = \begin{matrix} \text{var}(a_1) & 0 & 0 & 0 \\ 0 & \text{var}(a_2) & 0 & 0 \\ 0 & 0 & \text{var}(a_3) & 0 \\ 0 & 0 & 0 & \text{var}(a_4) \end{matrix}$$



Constraint on residuals, e.g.:
var(a1)=var(a2)=var(a3)=var(a4)

Psychological Medicine (2015), **45**, 1039–1049. © Cambridge University Press 2014
doi:10.1017/S003329171400213X

ORIGINAL ARTICLE

Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years

M. G. Nivard[1,2*], C. V. Dolan[1,3], K. S. Kendler[4], K.-J. Kan[1], G. Willemsen[1,5], C. E. M. van Beijsterveldt[1,5], R. J. L. Lindauer[6], J. H. D. A. van Beek[1,5], L. M. Geels[1,5], M. Bartels[1,5], C. M. Middeldorp[1,2,7]† and D. I. Boomsma[1,2,5]†

Do the Genetic or Environmental Determinants of Anxiety and Depression Change with Age? A Longitudinal Study of Australian Twins

Nathan A. Gillespie[1], Katherine M. Kirk[1], David M. Evans[1], Andrew C. Heath[2], Ian B. Hickie[3], and Nicholas G. Martin[1]

Twin Research and Human Genetics
Volume 18 ▮ Number 6 ▮ pp. 746–754 ▮ © The Author(s) 2015 ▮ doi:10.1017/thg.2015.80

Genetic and Environmental Stability of Neuroticism From Adolescence to Adulthood

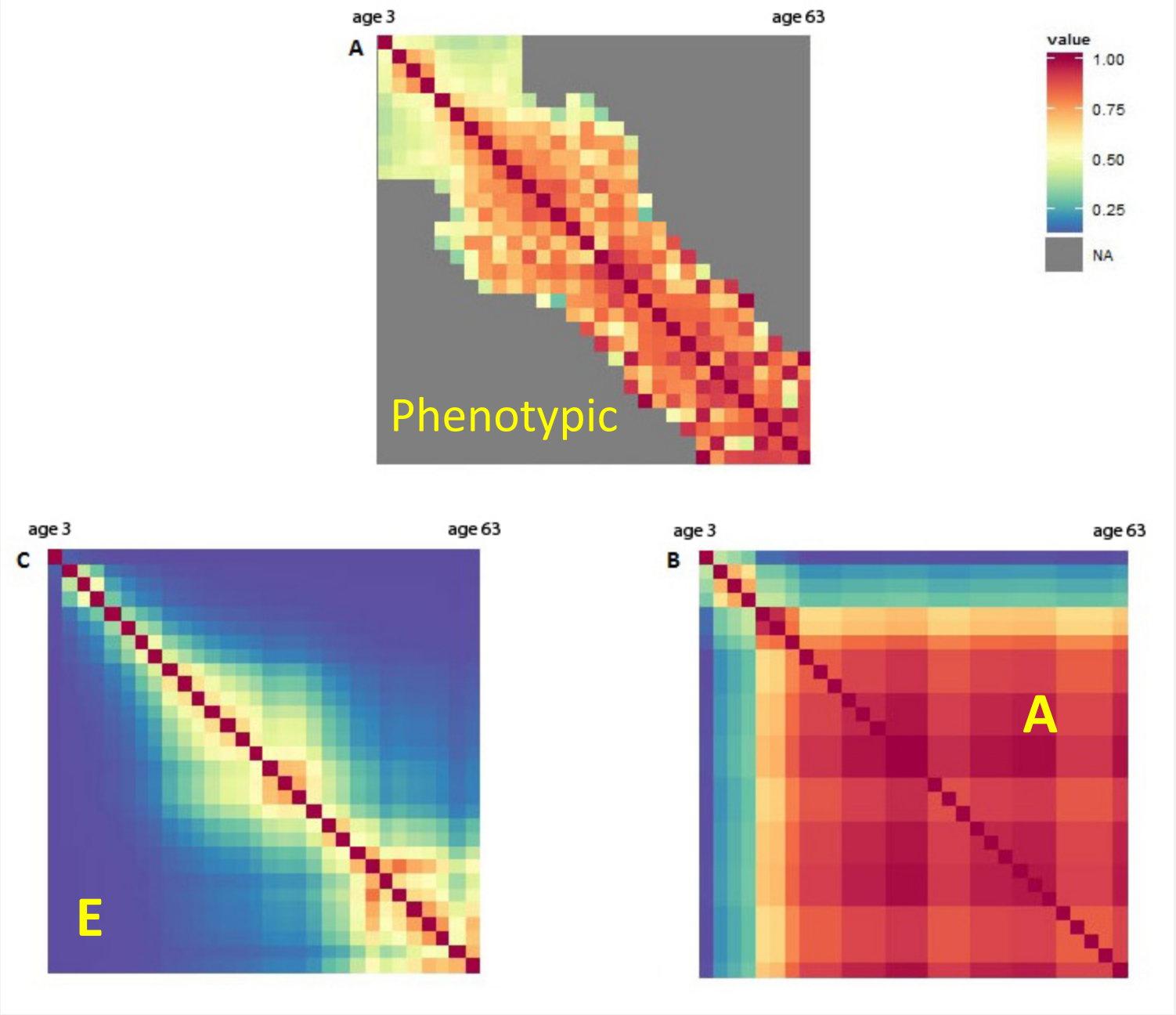
Michel G. Nivard,[1] Christel M. Middeldorp,[1,2] Conor V. Dolan,[1] and Dorret I. Boomsma[1,2,3]
[1]*Department of Biological Psychology, VU University Amsterdam, the Netherlands*
[2]*Neuroscience Campus Amsterdam, Amsterdam, the Netherlands*
[3]*EMGO+ Institute for Health and Care Research, Amsterdam, the Netherlands*

Genetic and Environmental Stability in Attention Problems Across the Lifespan: Evidence From the Netherlands Twin Register

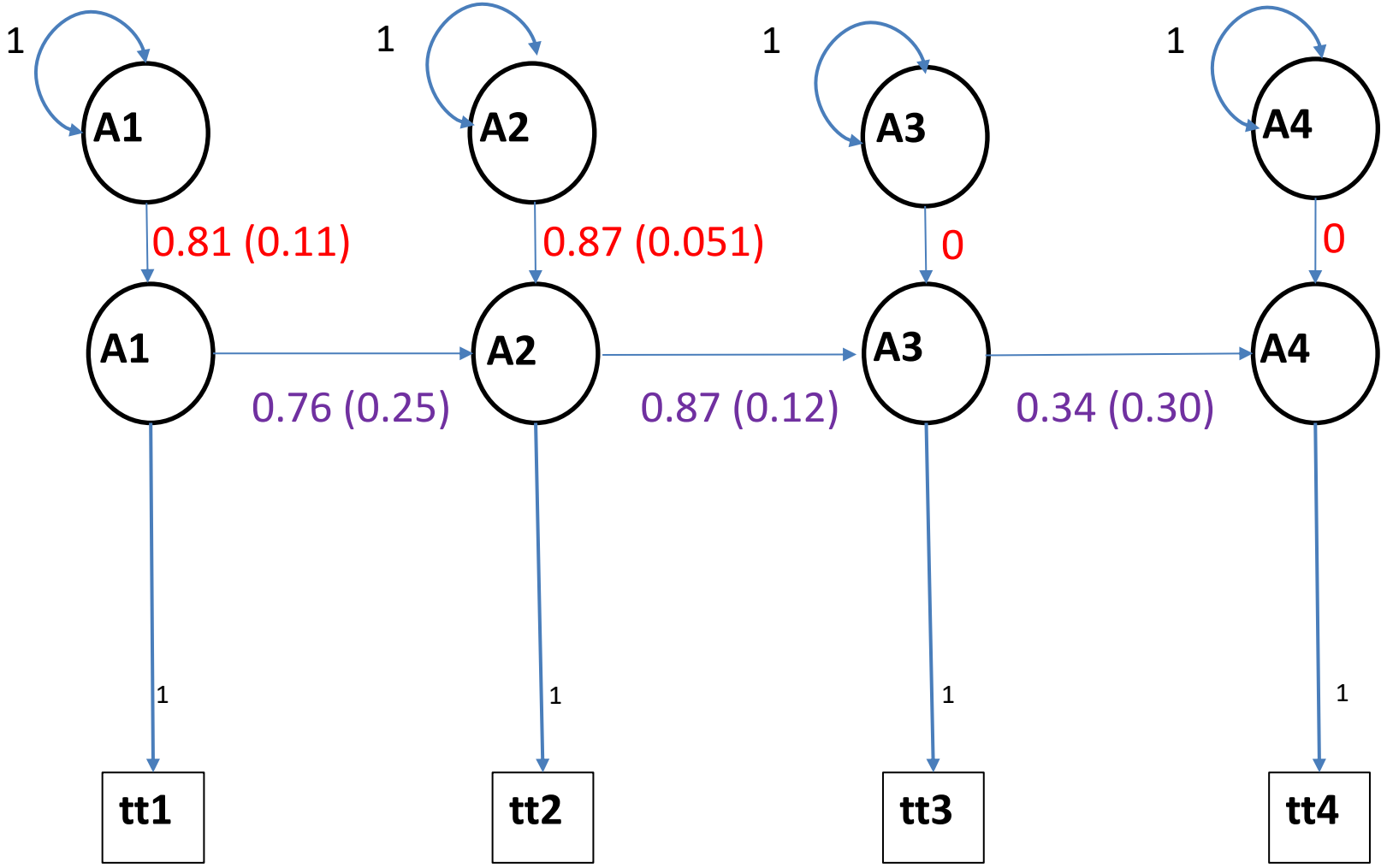
Kees-Jan Kan, Ph.D., Conor V. Dolan, Ph.D., Michel G. Nivard, M.Sc., Christel M. Middeldorp, Ph.D., Catharina E.M. van Beijsterveldt, Ph.D., Gonneke Willemsen, Ph.D., Dorret I. Boomsma, Ph.D.

This stability model can be applied at the genetic and non-genetic level.

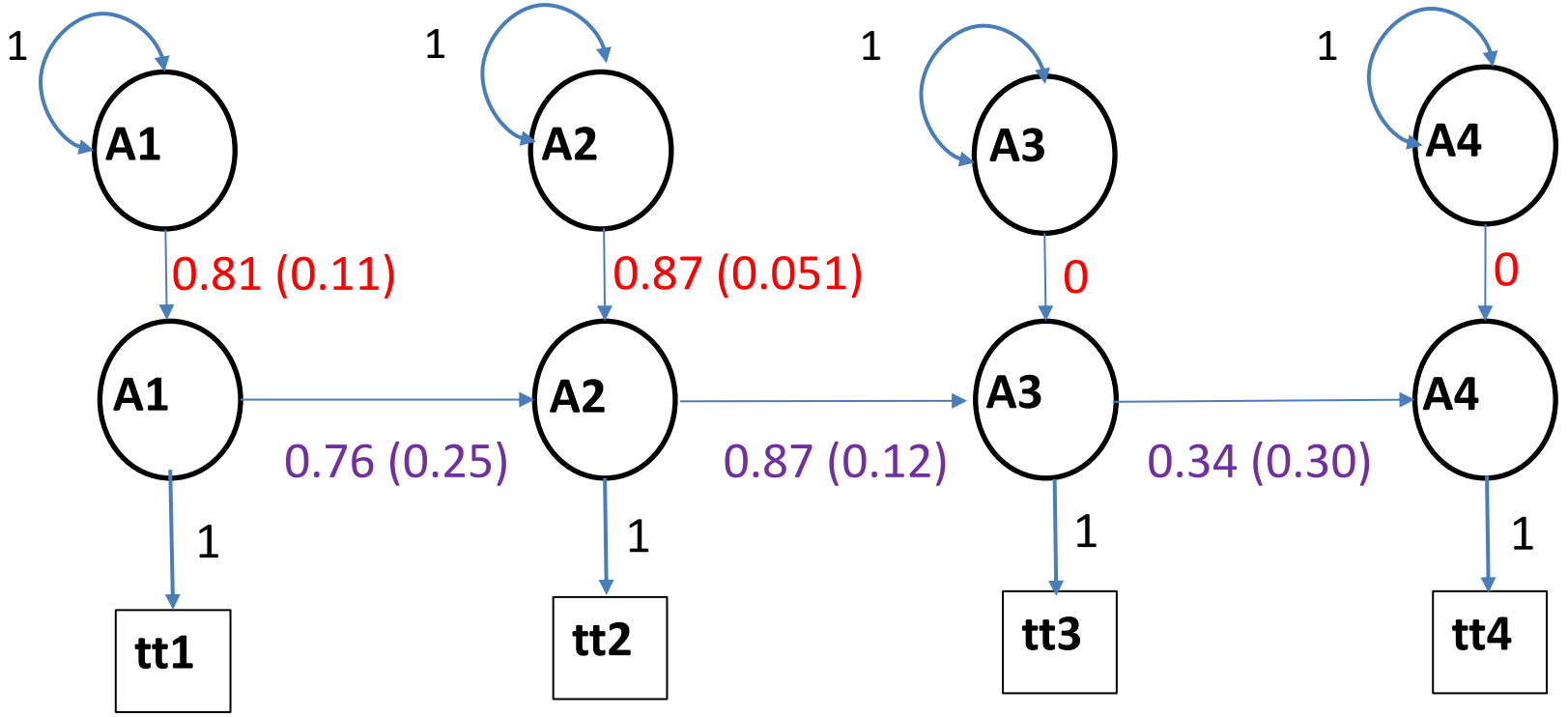
e.g. Nivard et al, 2014



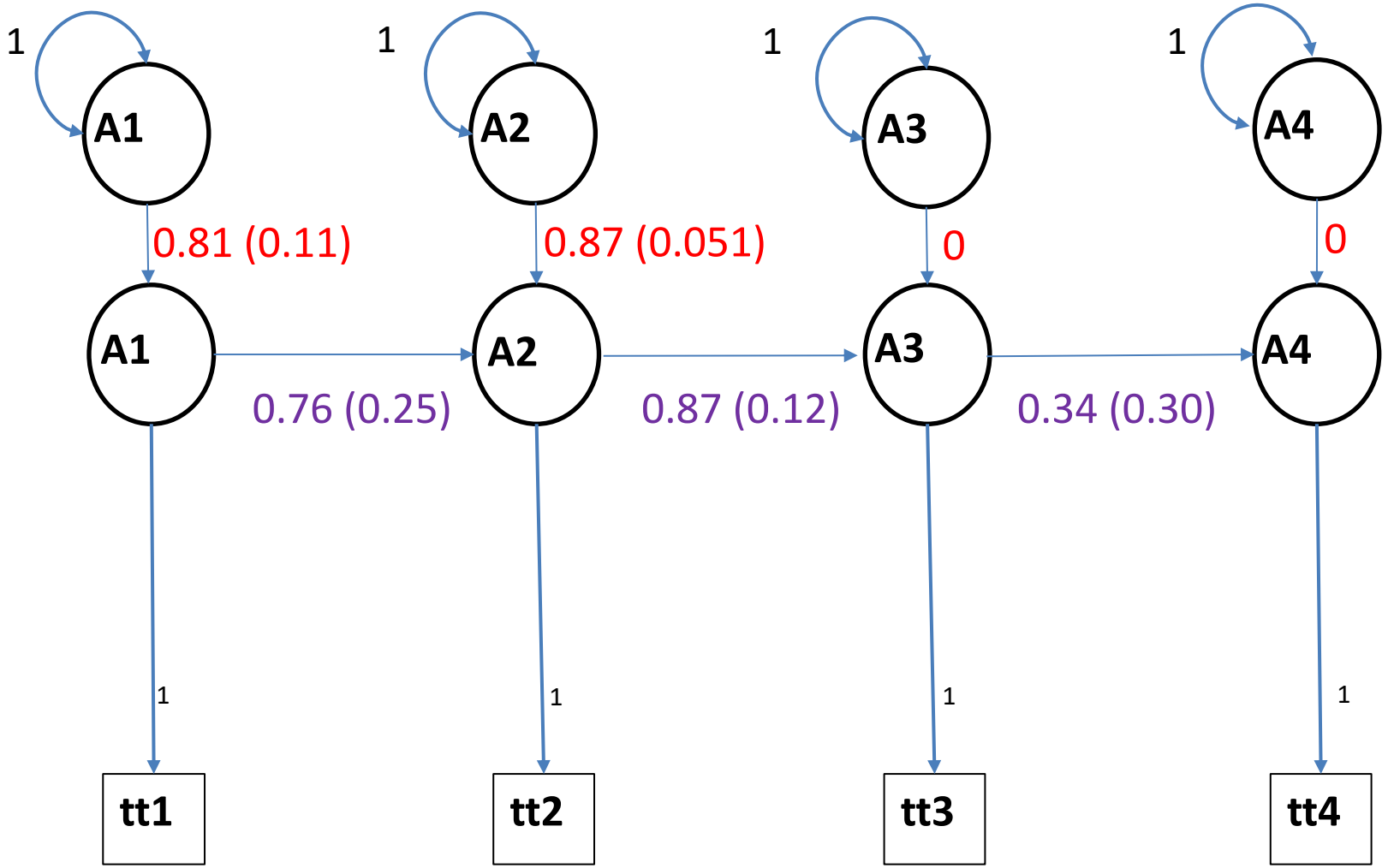
Anx/depression stability due to A and E from 3y to 63 years



Genetic Simplex (type D personality). Latent factors in circles, twin-time data in squares. Loadings from latent to observed traits and variation of innovations constrained at 1. Correlations between twins are specified at the level of the innovations (1 for MZ and 0.5 for DZ pairs).



Innovation genetic variance				
Variance due to Genetic Transmission				
Total V(A)				



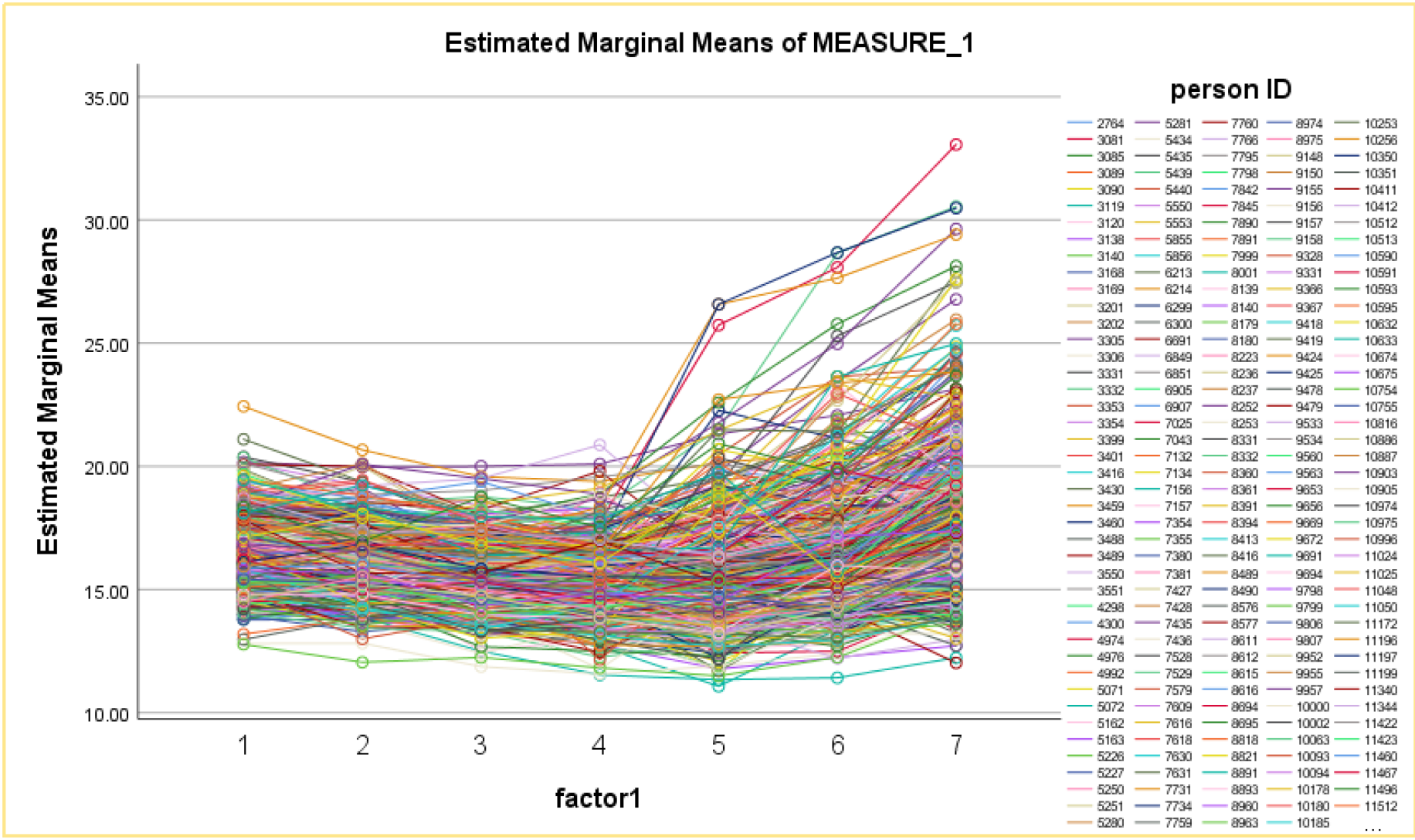
A Innovation	0.81^2=0.66	0.87^2=0.76	0	0
Transmission	-	0.66*0.76=0.50	1.26*0.87=1.10	1.10*0.34
Total V(A)	0.66	1.26	1.10	0.37

We will look at several approaches in an analysis of BMI data in young twins from the Netherlands Twin Register.

1) Saturated model
2) Cholesky ‘model’ (decomposition)
3) Simplex model

The BMI data set has data on 35,120 kids starting at birth.
We selected kids with complete data only and will analyze BMI at age 4, 7, 9 and 12 years.

For 7 ages: 1.5, 2, 3, 4 , 7, 10 and 12 years (N= 734): individual data



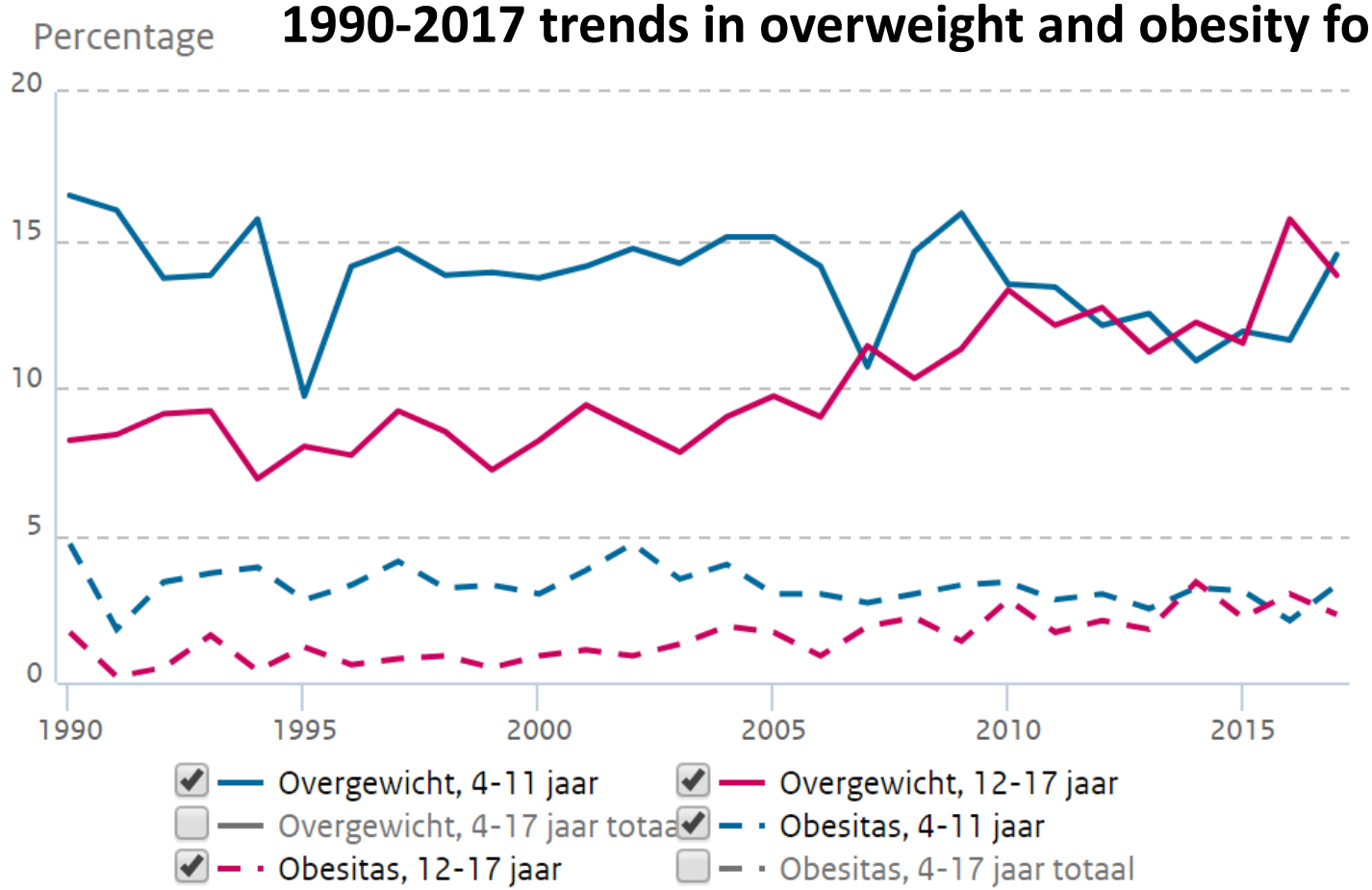
www.voedingscentrum.nl/professionals/kindervoeding-0-4-jaar/babyenkindervoeding/bmi-jongens-en-meisjes.aspx

Correlations

BMI across age (bold: all data / below diagonal : complete data)

BMI 4 yr	1	**.580**	**.493**	**.431**
BMI 7 yr	.590	1	**.729**	**.658**
BMI 10 yr	.515	.743	1	**.785**
BMI 12 yr	.442	.670	.803	1

1990-2017 trends in overweight and obesity for kids 4-18 year



Bron: CBS Gezondheidsenquête (tot en met 2013); daarna Gezondheidsenquête/Leefstijlmonitor CBS i.s.m. RIVM

- De gegevens zijn gestandaardiseerd naar de Nederlandse bevolking van 2017
- *De BMI-grenswaarden van overgewicht en obesitas van kinderen en jongeren wijken af van die van volwassenen, zie Definities.*

QUIZ

1) you have > 1 observation per person (from twin pairs) and fit a saturated model: what is the output / information ?

2) on the same data you fit a Cholesky decomposition: what is the information contained in the output?

3) if you aim for a model which recognizes time dependency in data what are the options?

QUIZ

1) > 1 observation per person (from twin pairs) and fit a saturated model: what is the output / information ?

a) ??????????????
b) ...
c) ...

QUIZ

1) > 1 observation per person (from twin pairs) and fit a saturated model: what is the output / information ?

a) ??????????????
b) An estimate for the mean value of the traits in twins
c) An estimate of the correlation for all traits in (MZ and DZ) pairs
d) An estimate of the phenotypic correlation between traits
e) An estimate of the (MZ and DZ) cross-correlations between traits
f) All of the above (except a))

QUIZ

2) on the same data you fit a Cholesky ‘model’ (decomposition): what is the information contained in the output?

a) The phenotypic correlations
b) The genetic correlations
c) The environmental correlations
d) b and c