Continuous heterogeneity

Danielle Dick & Sarah Medland

Boulder Twin Workshop
March 2006

Ways to Model Heterogeneity in Twin Data

- Multiple Group Models
  - Sex Effects
  - Young/Old cohorts

Problem:

- Many variables of interest do not fall into groups
  - Age
  - Socioeconomic status
  - Regional alcohol sales
  - Parental warmth
  - Parental monitoring

    - Grouping these variables into high/low categories loses a lot of information

‘Definition variables’ in Mx

- General definition: Definition variables are variables that may vary per subject and that are not dependent variables
- In Mx: The specific value of the def var for a specific individual is read into a matrix in Mx when analyzing the data of that particular individual

‘Definition variables’ in Mx

create dynamic var/cov structure

- Common uses:

1. As covariates/effects on the means (e.g. age and sex)

2. To model changes in variance components as function of some variable (e.g., age, SES, etc)

Cautionary note about definition variables

- Def var should not be missing if dependent is not missing
- Def var should not have the same missing value as dependent variable (e.g., use -2.00 for def var, -1.00 for dep var)

Definition variables used as covariates

General model with age and sex as covariates:

$$y_i = \alpha + \beta_1(\mathrm{age_i}) + \beta_2\,(\mathrm{sex_i}) + \varepsilon$$

Where $\boldsymbol{y_i}$ is the observed score of individual *i*, $\boldsymbol{\alpha}$ is the intercept or grand mean, $\boldsymbol{\beta_1}$ is the regression weight of age, **age$_\mathbf{i}$** is the age of individual *i*, $\boldsymbol{\beta_2}$ is the deviation of males (if sex is coded 0= female; 1=male), **sex$_\mathbf{i}$** is the sex of individual *i*, and $\boldsymbol{\varepsilon}$ is the residual that is not explained by the covariates (and can be decomposed further into ACE etc).

Standard model

- Means vector

$$\begin{pmatrix} m & m \end{pmatrix}$$

- Covariance matrix

$$\begin{pmatrix} a^2 + c^2 + e^2 & \\ Za^2 + c^2 & a^2 + c^2 + e^2 \end{pmatrix}$$

Allowing for a main effect of *X*

- Means vector

$$\begin{pmatrix} m+\beta X_{1i} & m+\beta X_{2i} \end{pmatrix}$$

- Covariance matrix

$$\begin{pmatrix} a^2+c^2+e^2 & \\ Za^2+c^2 & a^2+c^2+e^2 \end{pmatrix}$$

```
! Basic model + main effect of a definition variable

G1: Define Matrices
Data Calc NGroups=3
Begin Matrices;
X full 1 1 free     !genetic influences
Y full 1 1 free     !common environmental influences
Z full 1 1 free     !unique environmental influences
M full 1 1 free     ! grand mean
B full 1 1 free     ! moderator-linked means model
H full 1 1          !coefficient for DZ genetic relatedness
R full 1 1    ! twin 1 moderator (definition variable)
S full 1 1    ! twin 2 moderator (definition variable)
End Matrices;
Ma M 0
Ma B 0
Ma X 1
Ma Y 1
Ma Z 1
Matrix H .5
Options NO_Output
End
```

```
G2: MZ
Data NInput_vars=6 NObservations=0
Missing =-999
RE File=f1.dat
Labels id zyg p1 p2 m1 m2
Select if zyg = 1 /
Select p1 p2 m1 m2 /
Definition m1 m2 /
Matrices = Group 1
Means M + B*R | M + B*S /
Covariance
X*X' + Y*Y' + Z*Z' | X*X' + Y*Y' _
X*X' + Y*Y' | X*X' + Y*Y' + Z*Z' /

!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
End
```

Select phenoTw1, phenoTw2, modTw1, modTw2

```
G2: MZ
Data NInput_vars=6 NObservations=0
Missing =-999
RE File=f1.dat
Labels id zyg p1 p2 m1 m2
Select if zyg = 1 /
Select p1 p2 m1 m2 /
Definition m1 m2 /
Matrices = Group 1
Means M + B*R | M + B*S /
Covariance
X*X' + Y*Y' + Z*Z' | X*X' + Y*Y' _
X*X' + Y*Y' | X*X' + Y*Y' + Z*Z' /

!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
End
```

Tell MX modTw1, modTw2 are Definition Variables

```
G2: MZ
Data NInput_vars=6 NObservations=0
Missing =-999
RE File=f1.dat
Labels id zyg p1 p2 m1 m2
Select if zyg = 1 /
Select p1 p2 m1 m2 /
Definition m1 m2 /
Matrices = Group 1
Means M + B*R | M + B*S /
Covariance
X*X' + Y*Y' + Z*Z' | X*X' + Y*Y' _
X*X' + Y*Y' | X*X' + Y*Y' + Z*Z' /

!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
End
```

```
G3: DZ
Data NInput_vars=6 NObservations=0
Missing =-999
RE File=f1.dat
Labels id zyg p1 p2 m1 m2
Select if zyg = 2 /
Select p1 p2 m1 m2 /
Definition m1 m2 /
Matrices = Group 1
Means M + B*R | M + B*S /
Covariance
X*X' + Y*Y' + Z*Z' | H@X*X' + Y*Y' _
H@X*X' + Y*Y' | X*X' + Y*Y' + Z*Z' /

!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
End
```

```
 MATRIX X
This is a FULL matrix of order    1 by    1
            1
1      1.3228
```

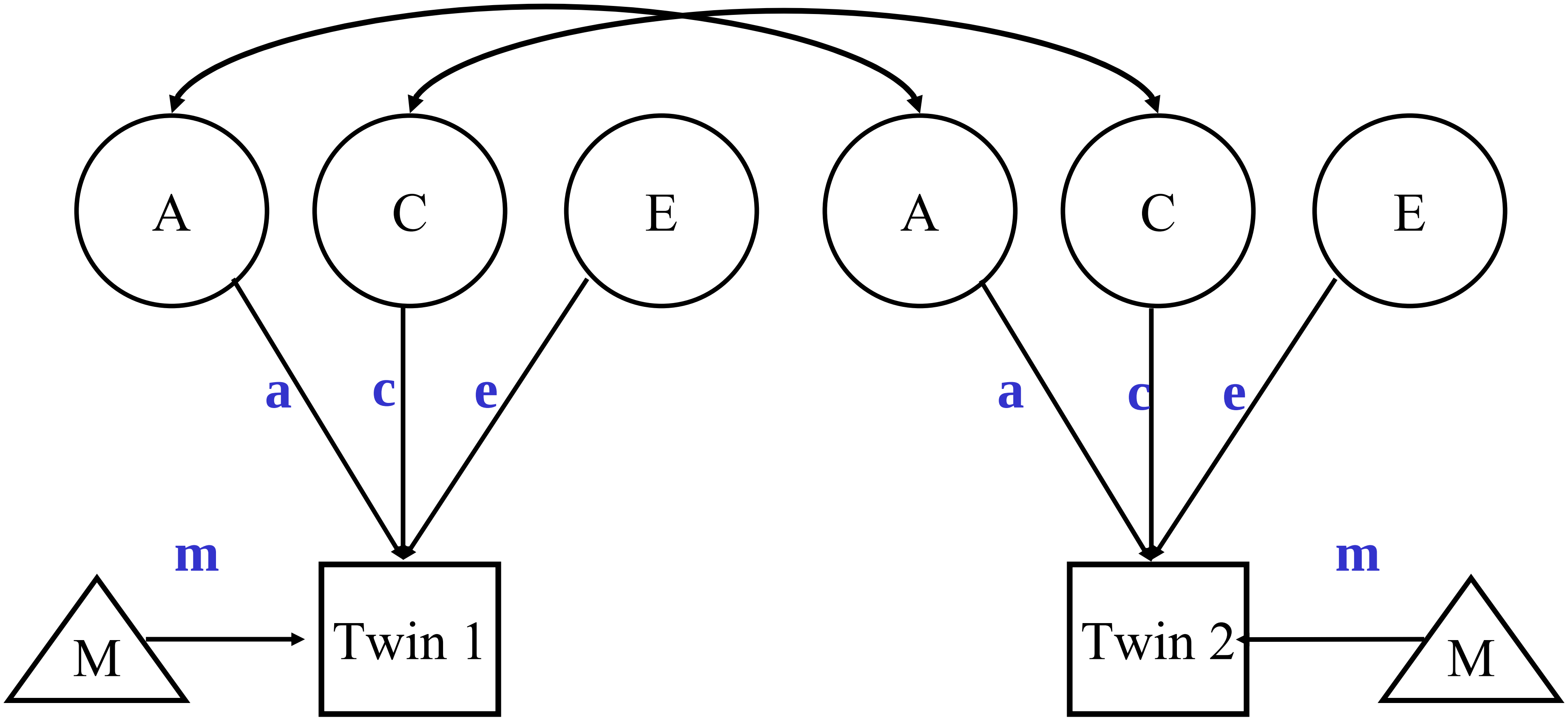
MATRIX B

```
This is a FULL matrix of order    1 by    1
         1
1   0.3381
```

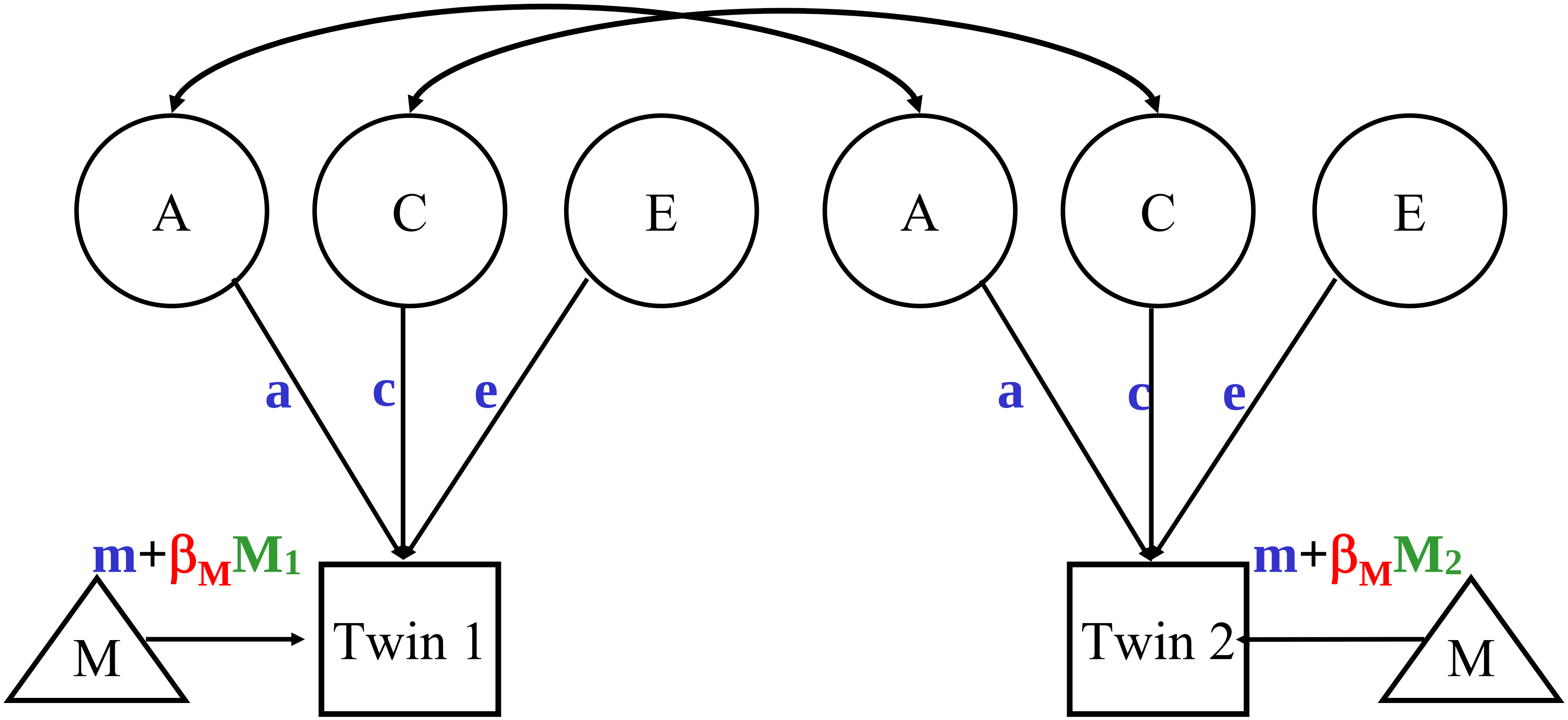
```
 MATRIX Y
This is a FULL matrix of order    1 by    1
            1
1      1.1051

 MATRIX Z
This is a FULL matrix of order    1 by    1
         1
1   0.9728

 MATRIX M
This is a FULL matrix of order    1 by    1
         1
1   0.1035

Your model has    5 estimated parameters and    800 Observed
  statistics

-2 times log-likelihood of data >>>  3123.925
Degrees of freedom >>>>>>>>>>>>>>>>       795
```

```
MATRIX X
This is a FULL matrix of order    1 by    1
            1
1      1.3078

 MATRIX B
This is a FULL matrix of order    1 by    1
         1
1   0.0000

 MATRIX Y
This is a FULL matrix of order    1 by    1
            1
1      1.1733

 MATRIX Z
This is a FULL matrix of order    1 by    1
         1
1   0.9749

 MATRIX M
This is a FULL matrix of order    1 by    1
         1
1   0.1069

Your model has    4 estimated parameters and    800 Observed
  statistics

-2 times log-likelihood of data >>>  3138.157
Degrees of freedom >>>>>>>>>>>>>>>>       796
```

Model-fitting approach to GxE



Adding Covariates to Means Model



‘Definition variables’ in Mx

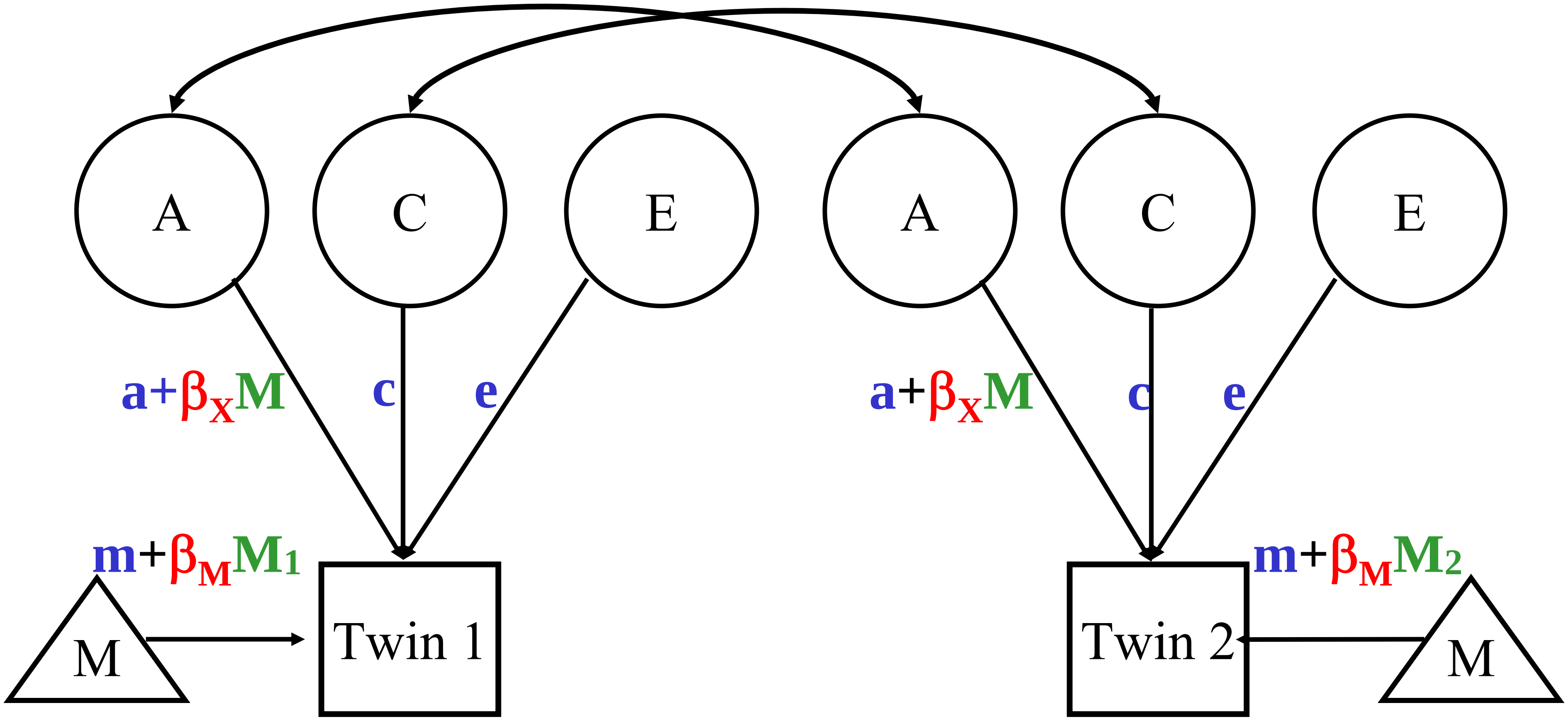
create dynamic var/cov structure

- Common uses:

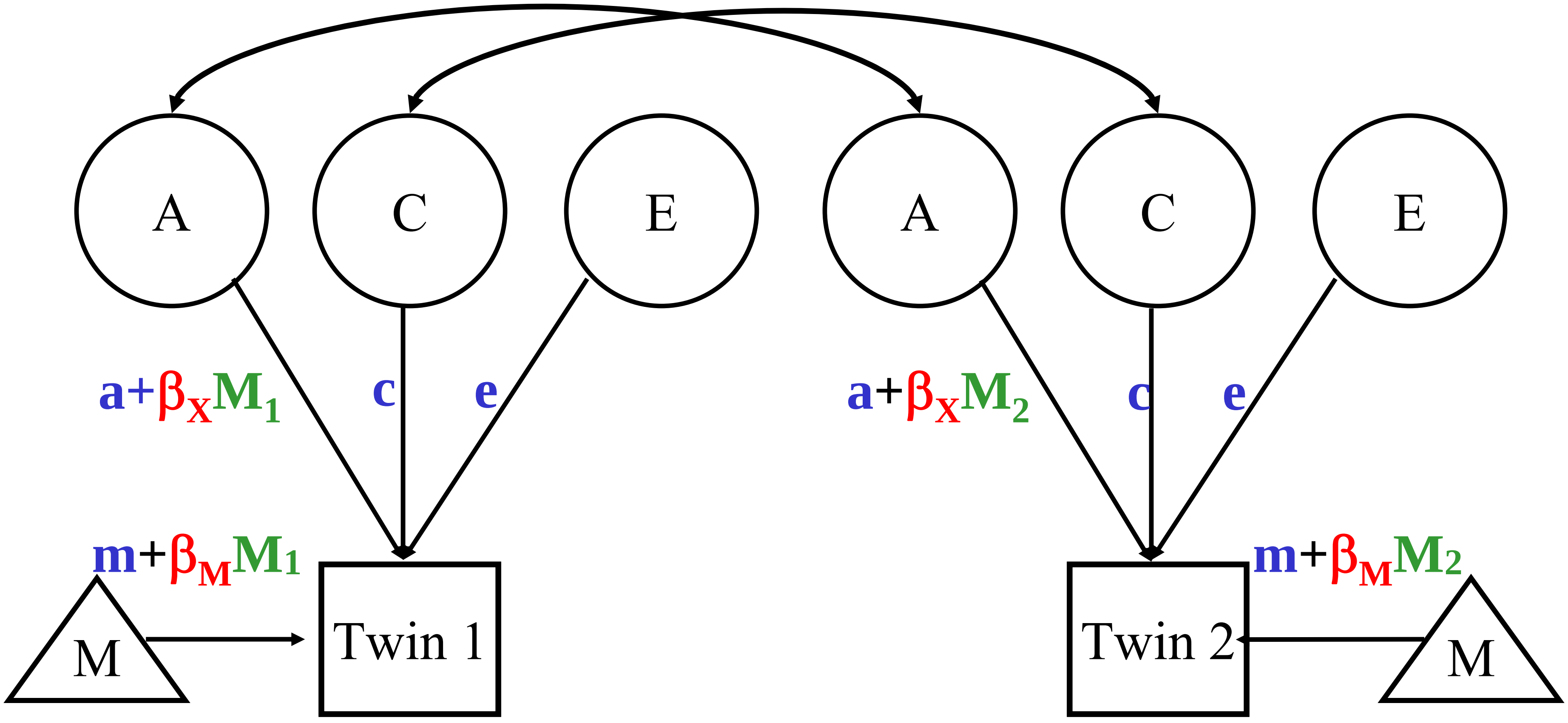
1. As covariates/effects on the means (e.g. age and sex)

2. To model changes in variance components as function of some variable (e.g., age, SES, etc)

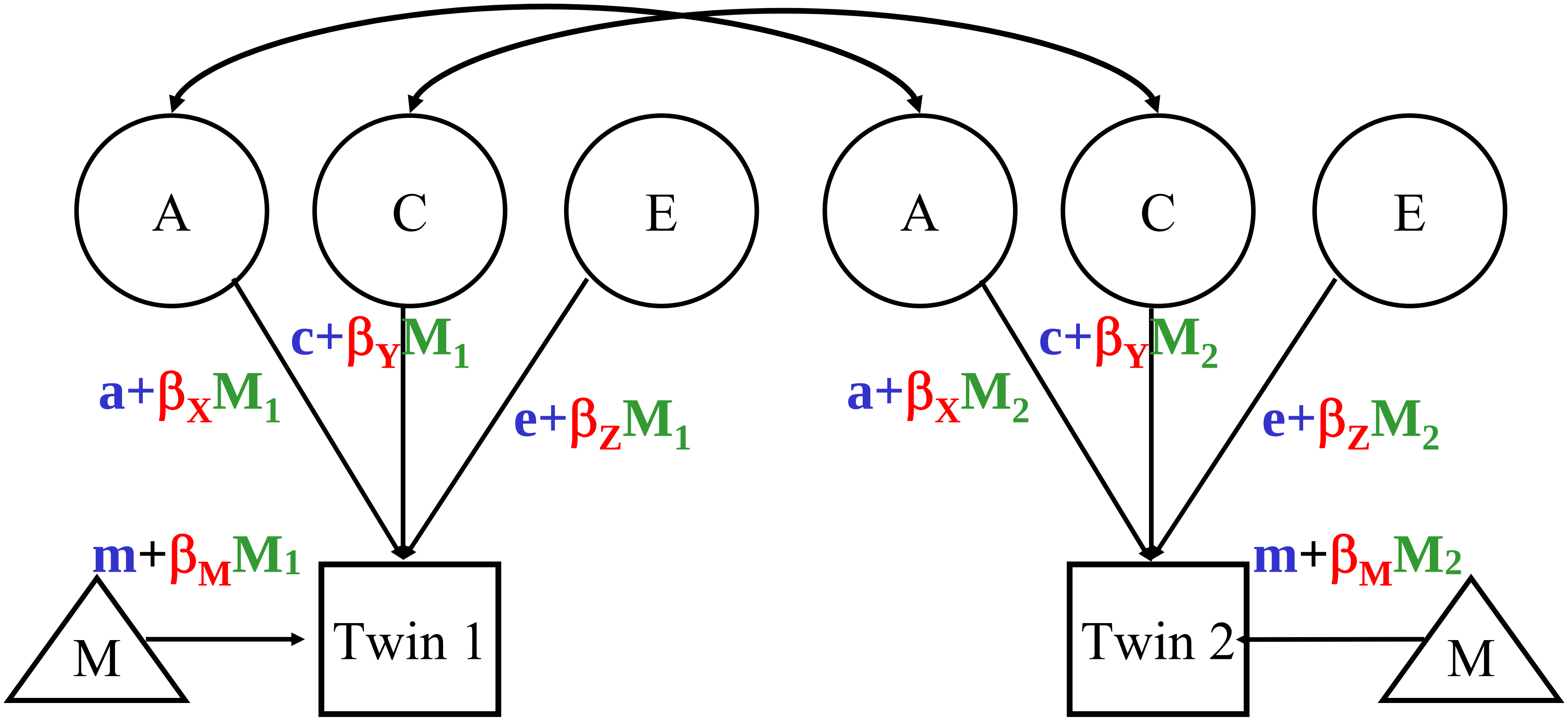
Model-fitting approach to GxE



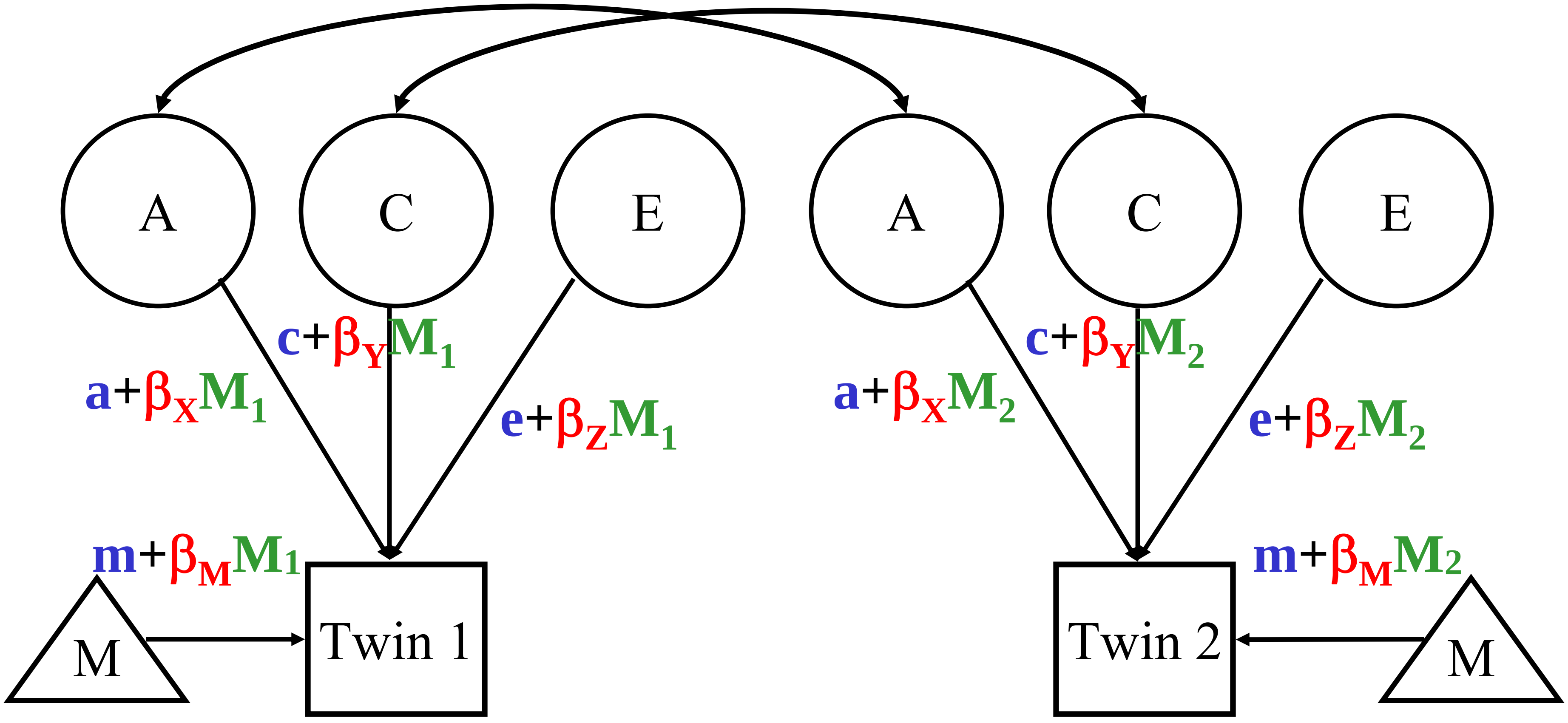
Individual specific moderators



E x E interactions

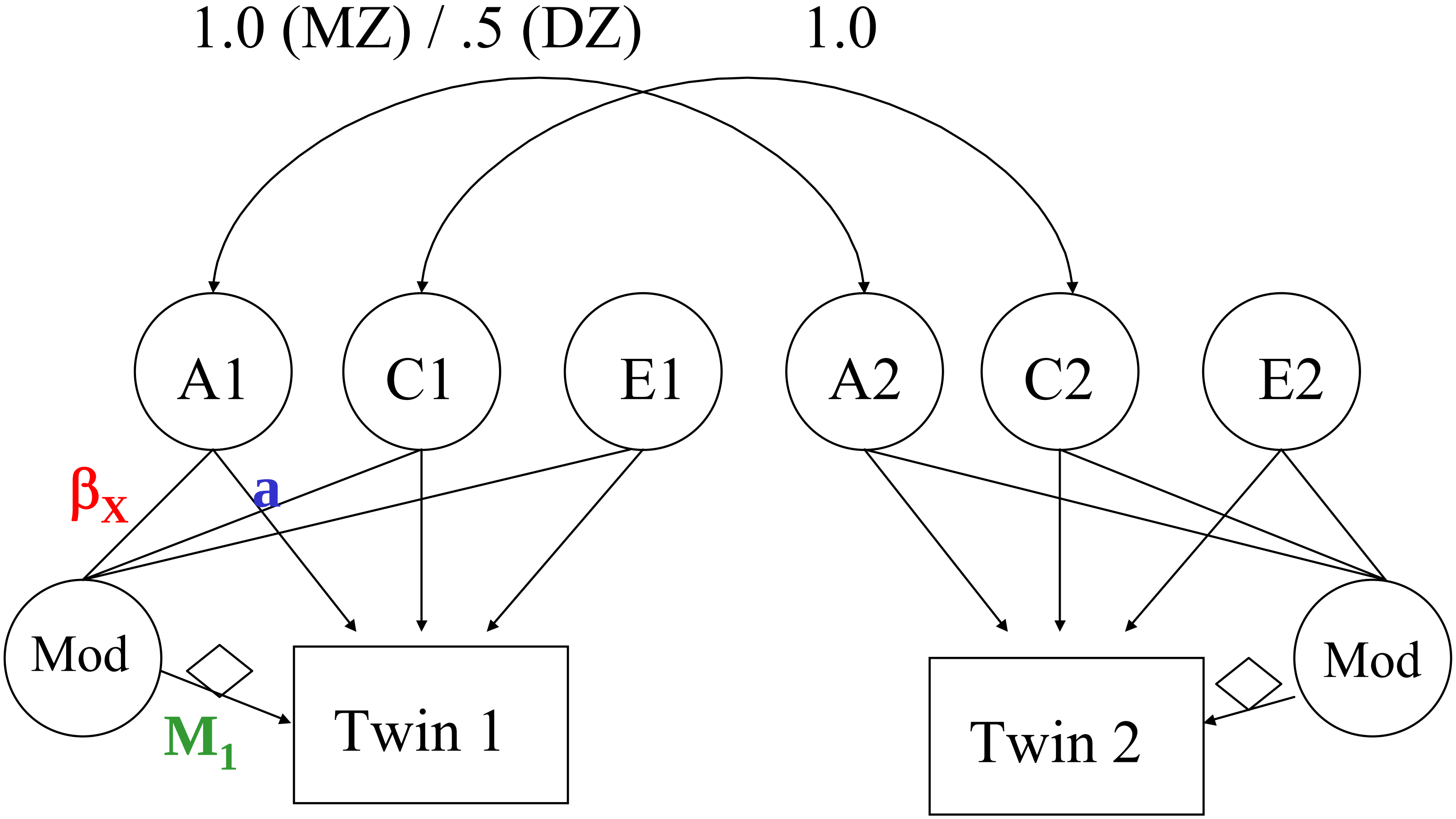


ACE - XYZ - M



Main effects and moderating effects

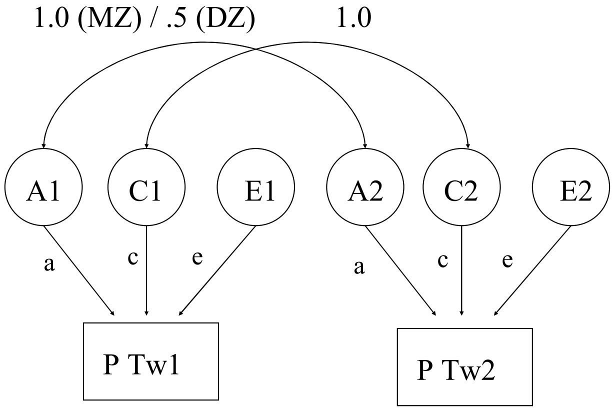
Definition Variables in Mx GUI



Dick et al., 2001

- Classic Twin Model:

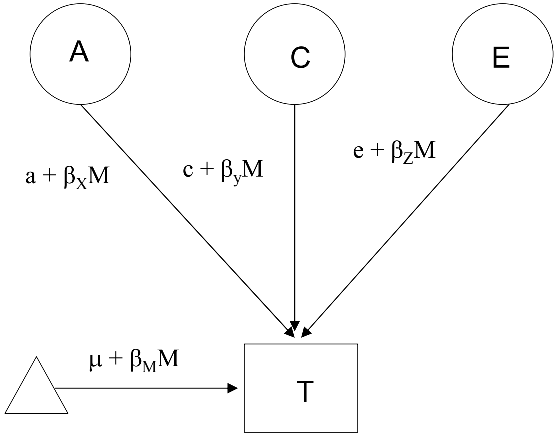
  $\mathrm{Var}\,(P) = a^2 + c^2 + e^2$



- Moderation Model:

  $\mathrm{Var}\,(P) =$

  $(a + \beta_X M)^2 + (c + \beta_Y M)^2 + (e + \beta_Z M)^2$



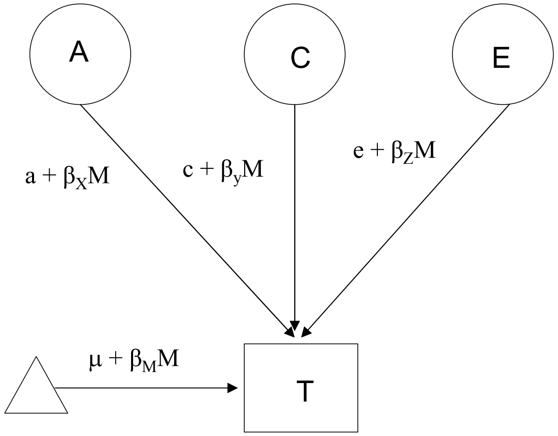
Purcell 2002,
Twin Research

$$\text{Var (T)} = (a + \beta_X M)^2 + (c + \beta_Y M)^2 (e + \beta_Z M)^2$$

Where M is the value of the moderator and

Significance of β_X indicates genetic moderation
Significance of β_Y indicates common environmental moderation
Significance of β_Z indicates unique environmental moderation

B_M indicates a main effect of the moderator on the mean

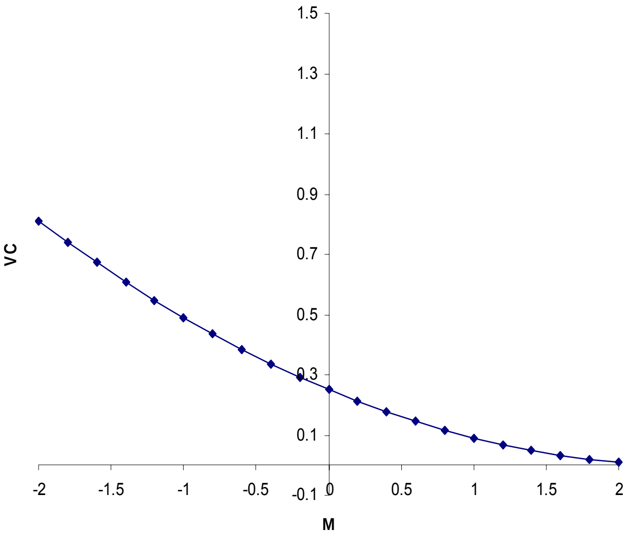


Plotting VCs as Function of Moderator

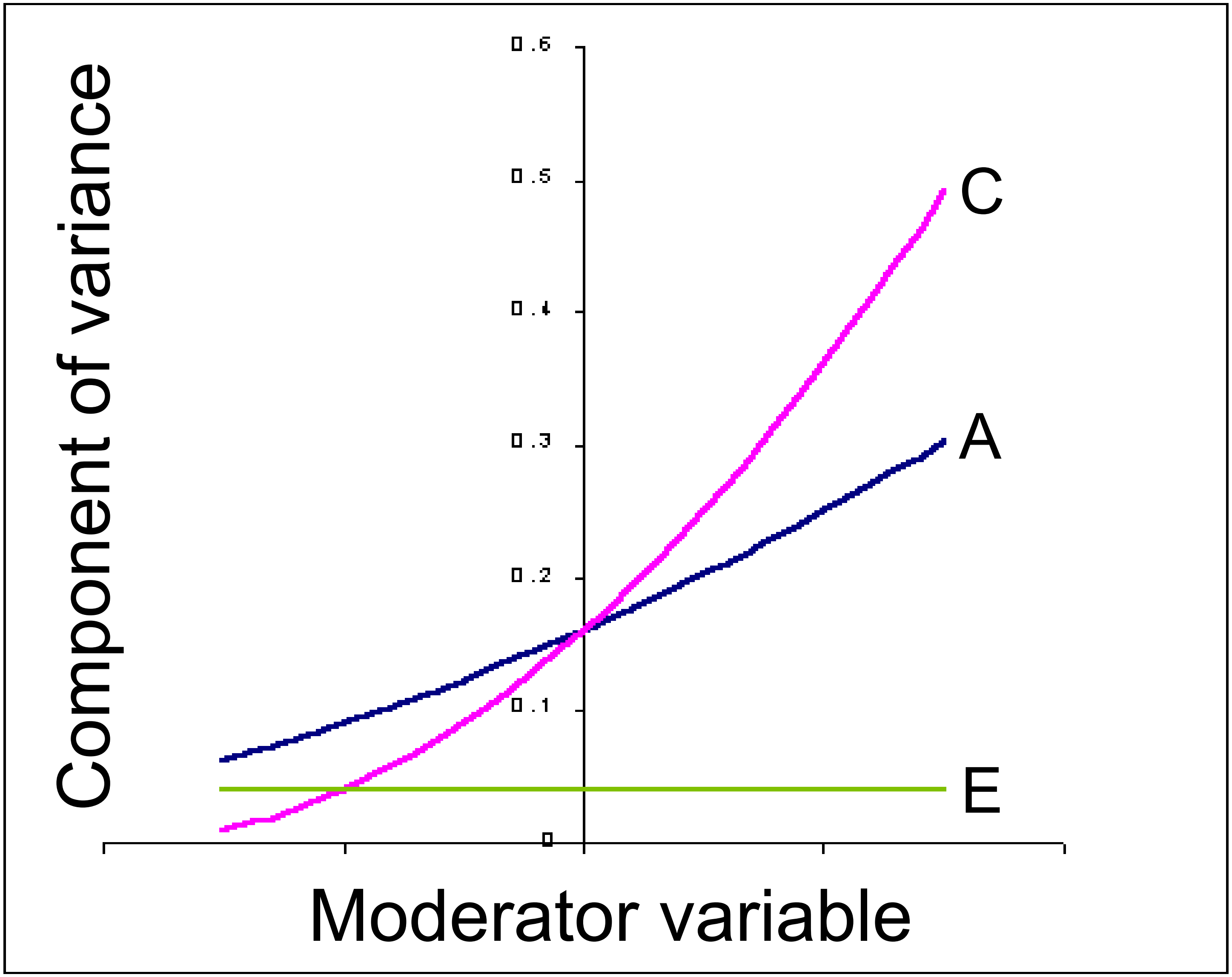
- For the additive genetic VC, for example
  - Given a, β (estimated in Mx model) and a range of values for the moderator variable
- For example,

$a = 0.5$, $\beta = -0.2$ and M ranges from -2 to +2

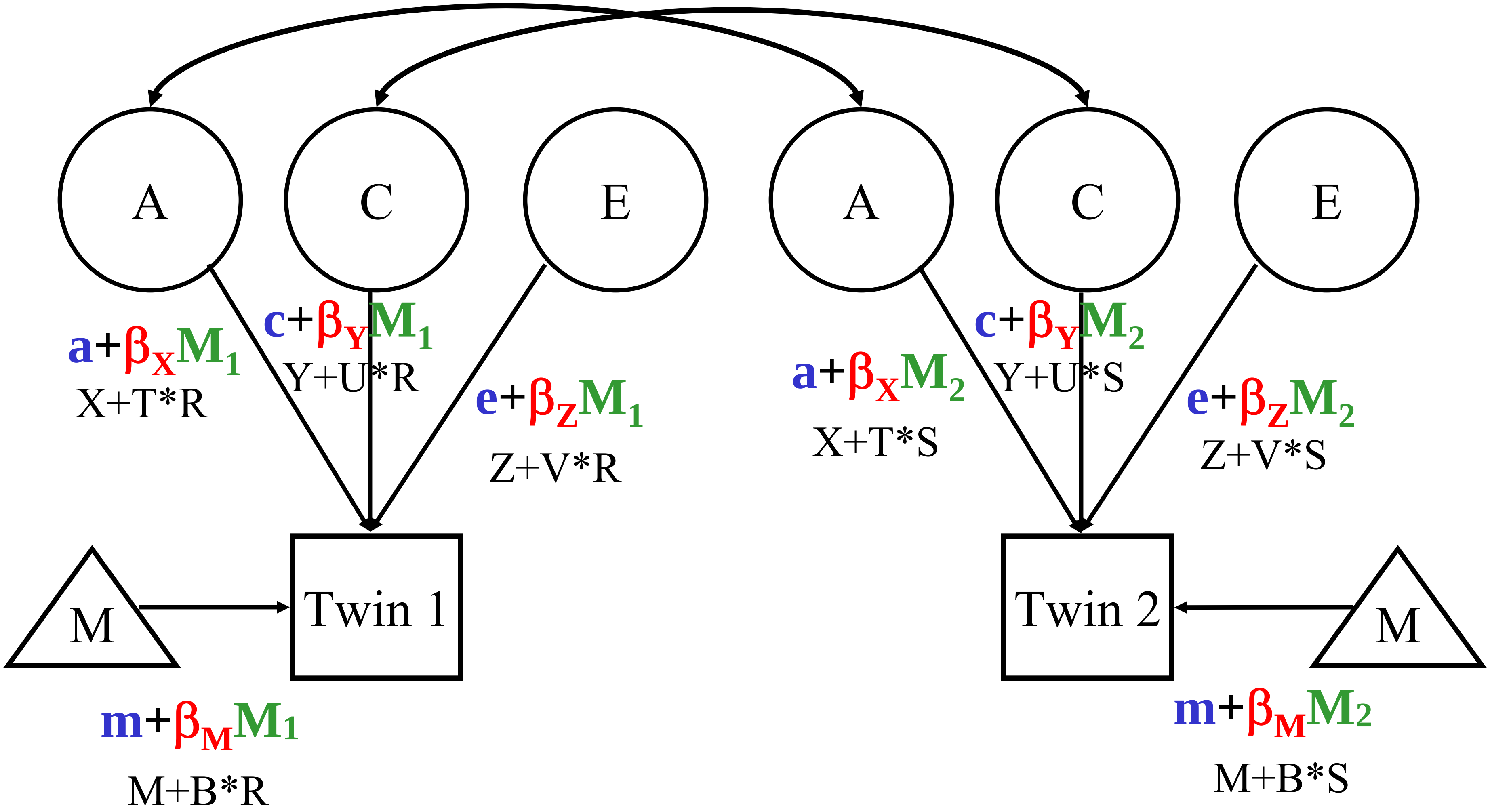
M	$(a+\beta M)^2$	$(a+\beta M)^2$
-2	$(0.5+(-0.2\times-2))^2$	0.81
-1.5	$(0.5+(-0.2\times-1.5))^2$	0.73
…		
+2	$(0.5+(-0.2\times2))^2$	0.01



Model-fitting approach to GxE



Matrix Letters as Specified in Mx Script



```
! GxE - Basic model
G1: Define Matrices
Data Calc NGroups=3
Begin Matrices;
X full 1 1 free
Y full 1 1 free
Z full 1 1 free
T full 1 1 free        ! moderator-linked A component
U full 1 1 free        ! moderator-linked C component
V full 1 1 free        ! moderator-linked E component
M full 1 1 free        ! grand mean
B full 1 1 free        ! moderator-linked means model
H full 1 1
R full 1 1     ! twin 1 moderator (definition variable)
S full 1 1     ! twin 2 moderator (definition variable)
End Matrices;
Ma T 0
Ma U 0
Ma V 0
Ma M 0
Ma B 0
Ma X 1
Ma Y 1
Ma Z 1
Matrix H .5
Options NO_Output
End
```

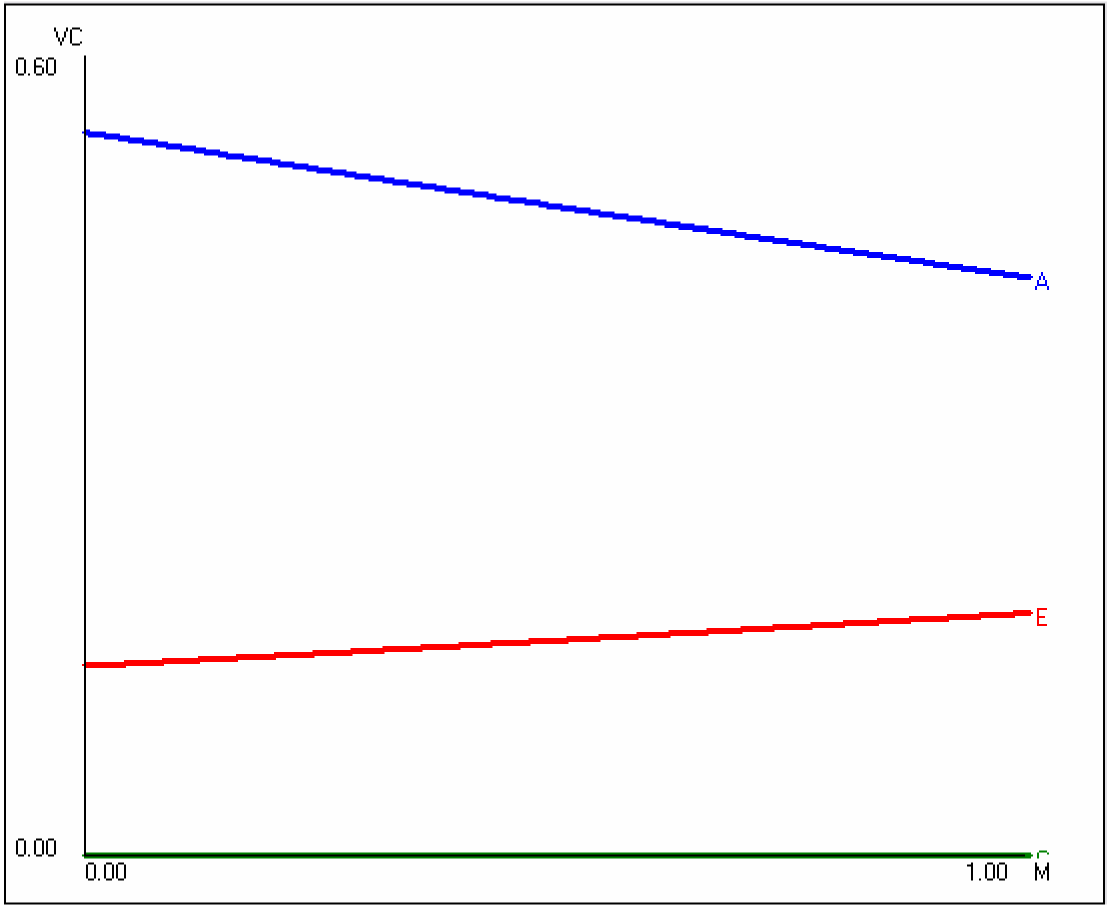
```
G2: MZ
Data NInput_vars=6 NObservations=0
Missing =-999
RE File=f1.dat
Labels id zyg p1 p2 m1 m2
Select if zyg = 1 /
Select p1 p2 m1 m2 /
Definition m1 m2 /
Matrices = Group 1
Means M + B*R | M + B*S /
Covariance
(X+T*R)*(X+T*R) + (Y+U*R)*(Y+U*R) + (Z+V*R)*(Z+V*R) |
(X+T*R)*(X+T*S) + (Y+U*R)*(Y+U*S) _
(X+T*S)*(X+T*R) + (Y+U*S)*(Y+U*S) |
(X+T*S)*(X+T*S) + (Y+U*S)*(Y+U*S) + (Z+V*S)*(Z+V*S) /

!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
Options NO_Output
End
```

```
G2: DZ
Data NInput_vars=6 NObservations=0
Missing =-999
RE File=f1.dat
Labels id zyg p1 p2 m1 m2
Select if zyg = 1 /
Select p1 p2 m1 m2 /
Definition m1 m2 /
Matrices = Group 1
Means M + B*R | M + B*S /
Covariance
  (X+T*R)*(X+T*R) + (Y+U*R)*(Y+U*R) + (Z+V*R)*(Z+V*R) |
H@(X+T*R)*(X+T*S) + (Y+U*R)*(Y+U*S) _
H@(X+T*S)*(X+T*R) + (Y+U*S)*(Y+U*S) |
  (X+T*S)*(X+T*S) + (Y+U*S)*(Y+U*S) + (Z+V*S)*(Z+V*S) /

!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
Options NO_Output
End
```

Practical

- Cohort (young/old) model using definition variables (coded 0/1)
- Extension to continuous age

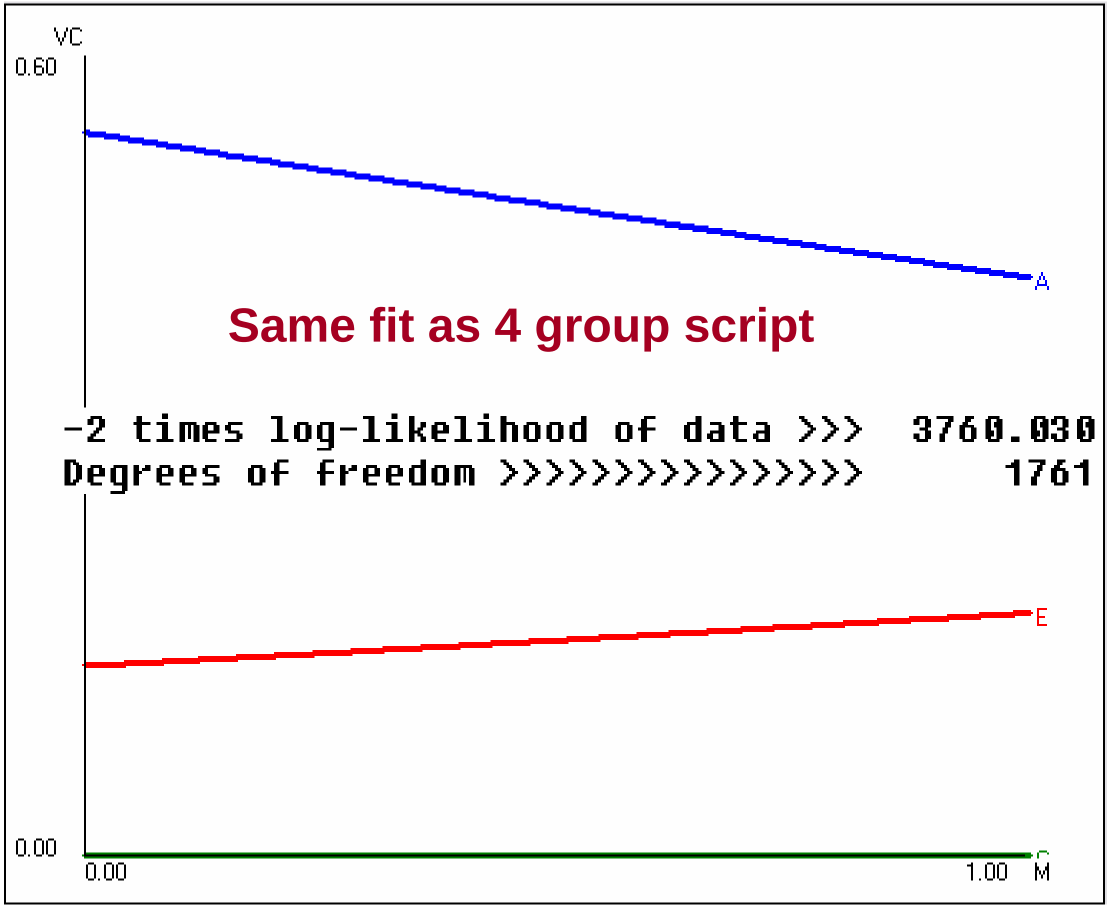
Cohort Moderation



Younger Cohort

Older Cohort

Cohort Moderation



Same fit as 4 group script

```
-2 times log-likelihood of data >>>   3760.030
Degrees of freedom >>>>>>>>>>>>>>>>       1761
```

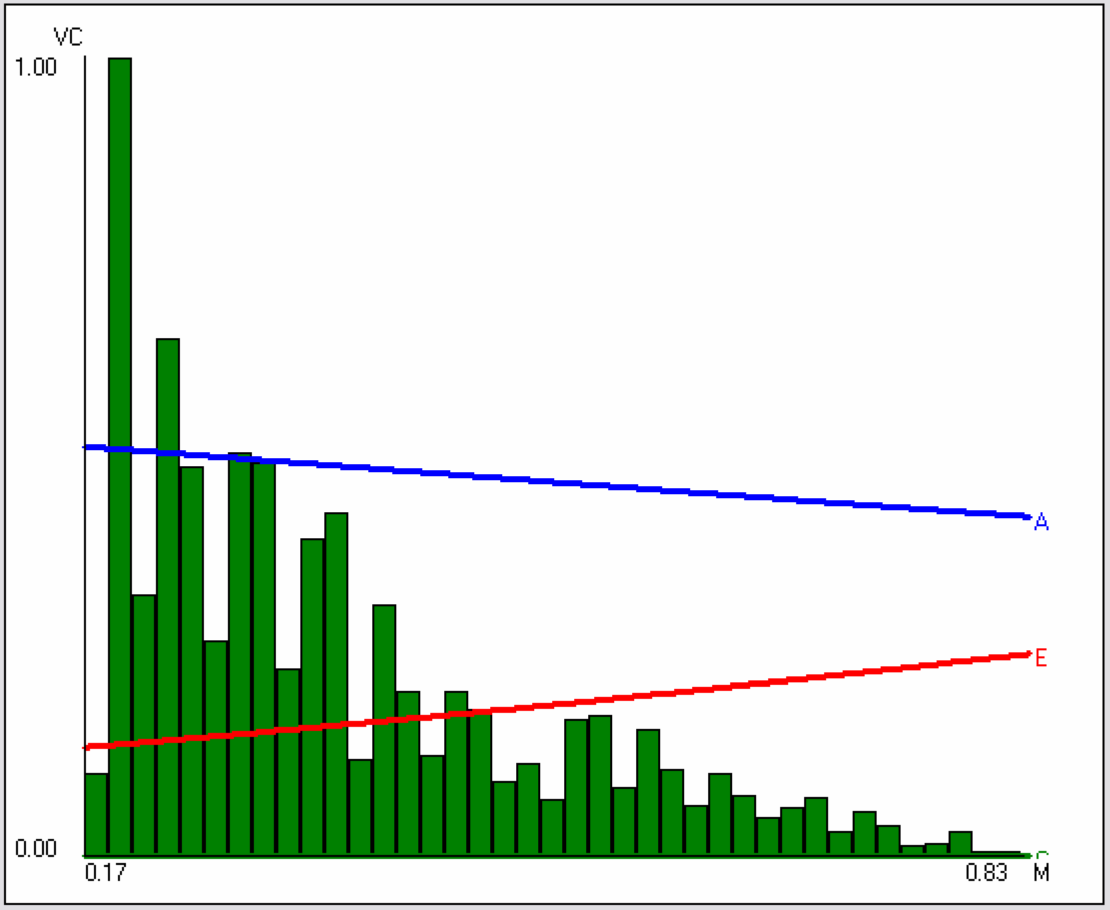
Younger Cohort

Older Cohort

Your task

- Add tests to age_mod.mx to test
  - the significant of age moderation on A
  - the significant of age moderation on E
  - the significant of age moderation on both A and E jointly

Age Moderation



17 years old 83 years old

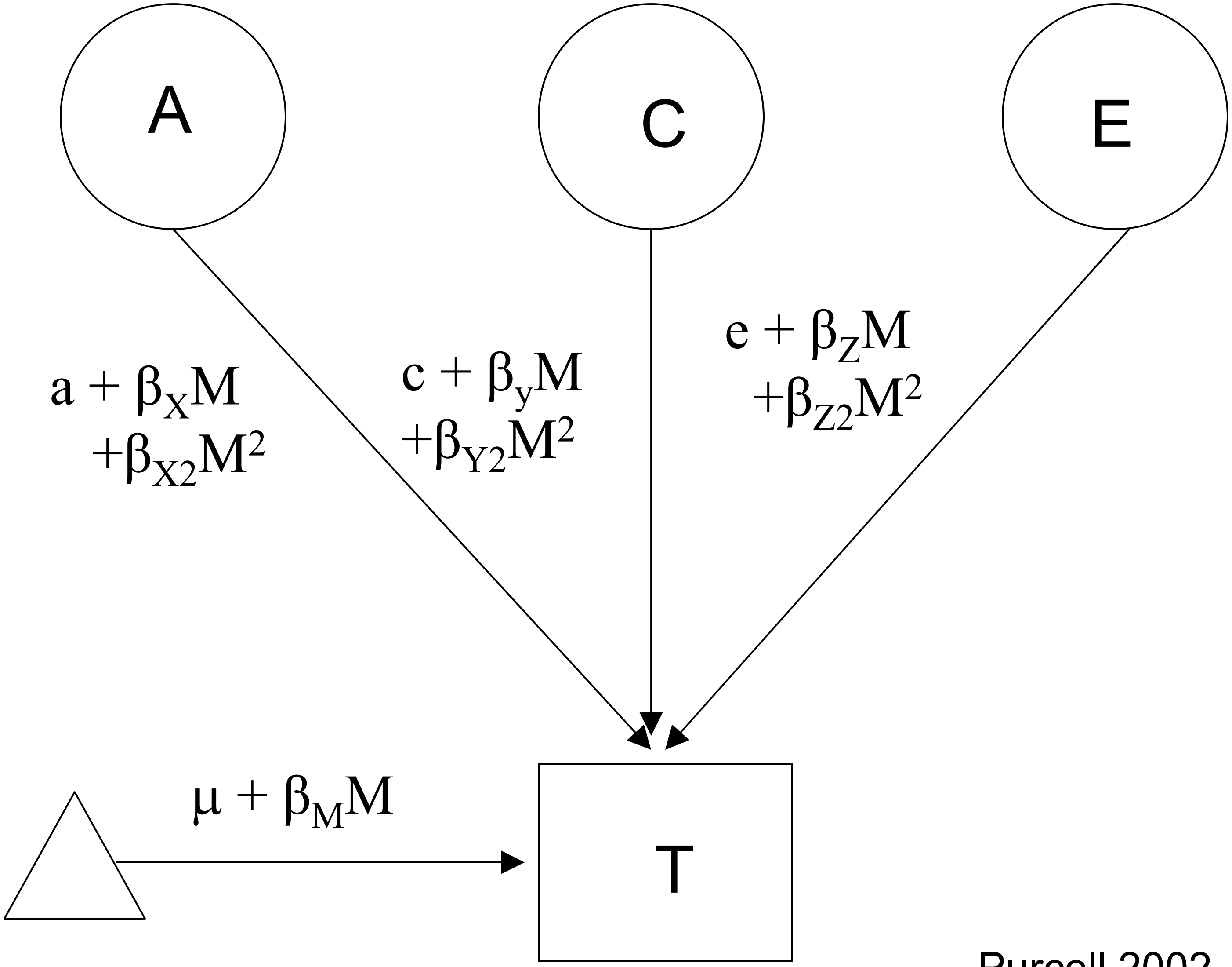
Comparing Results

Model	-2LL	df	chi-sq	p
4 group cohort	3760.030	1761		
Cohort mod.	3760.030	1761		
Age mod.	3764.448	1761		
Drop A mod.	3764.873	1762	0.426	0.514
Drop E. mod.	3768.636	1762	4.189	0.041
No mod.	3768.680	1763	4.232	0.120

Why is the fit worse using the continuous moderator?

- Artefact – was the GxE due to the arbitrary cut-point?
- Confound – is there a second modifier involved?
- Non-linear – would we expect the effect of age on BMI in adults to be linear?

Nonlinear Moderation can be modeled with the addition of a quadratic term



A

C

E

$a + \beta_X M$
$+\beta_{X2} M^2$

$c + \beta_y M$
$+\beta_{Y2} M^2$

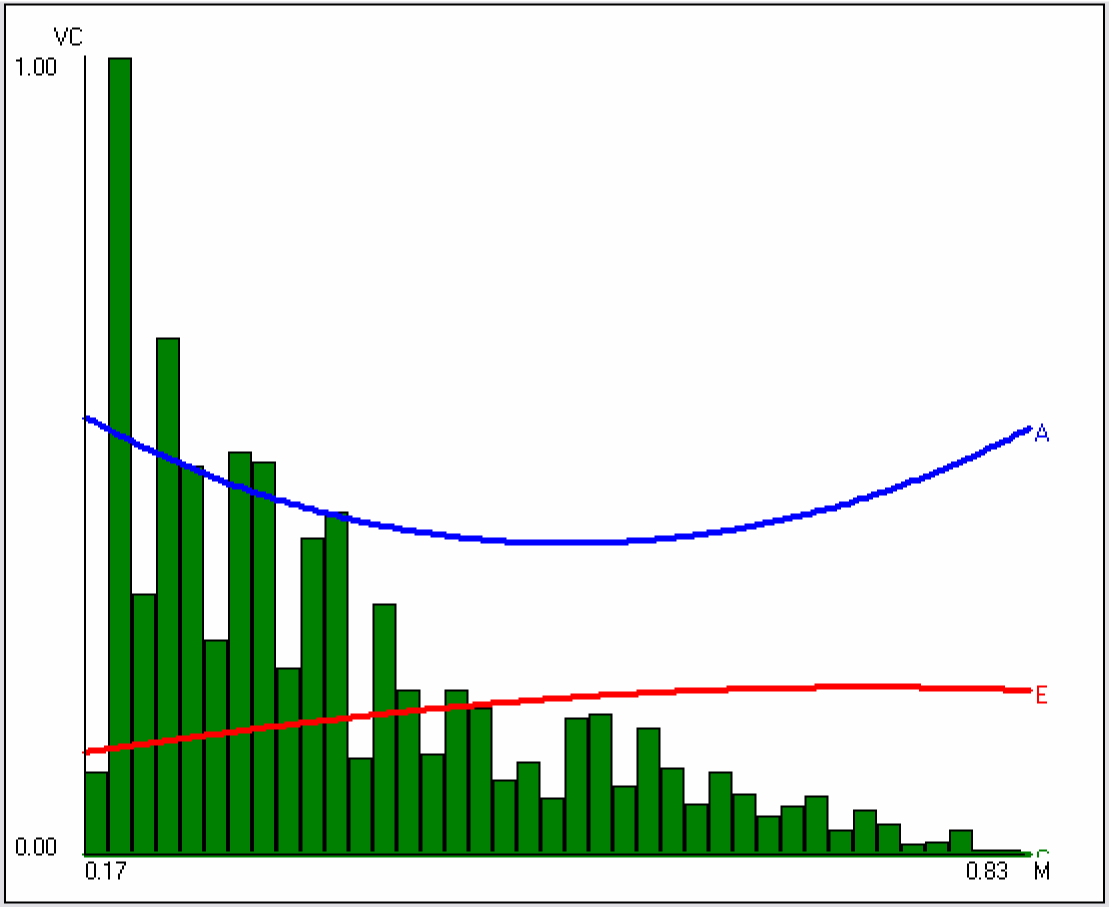
$e + \beta_Z M$
$+\beta_{Z2} M^2$

$\mu + \beta_M M$

T

Purcell 2002

Non-linear Age Moderation



17 years old

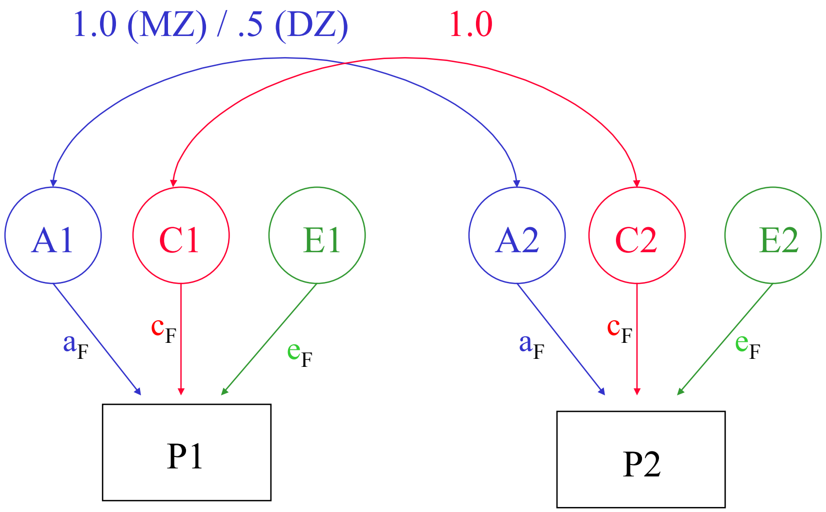
83 years old

Comparing Results

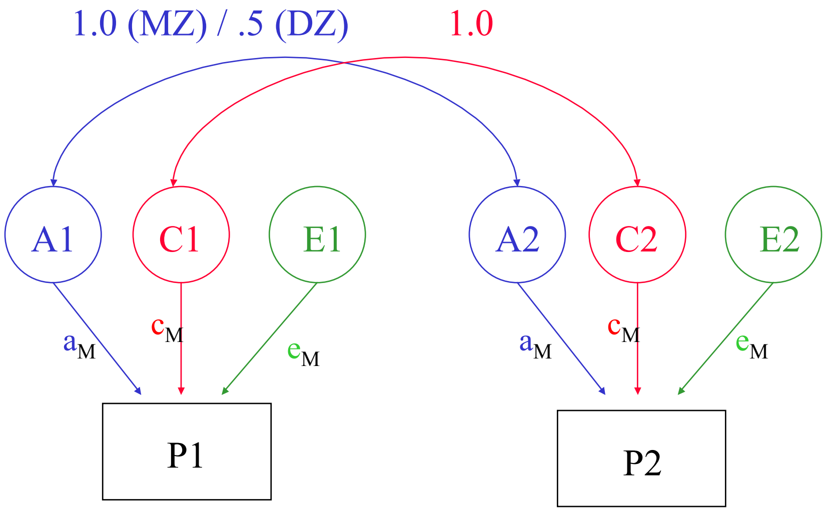
Model	-2LL	df	chi-sq	p
Age + Age^2 mod.	3698.365	1758		
Drop both A mod.	3701.449	1760	3.084	0.214
Drop both E mod.	3702.907	1760	4.543	0.103
Age mod. No Age^2 terms.	3764.448	1761	66.083	.00000 etc
No mod.	3768.680	1763	70.315	.00000 etc

Age Effects

Young

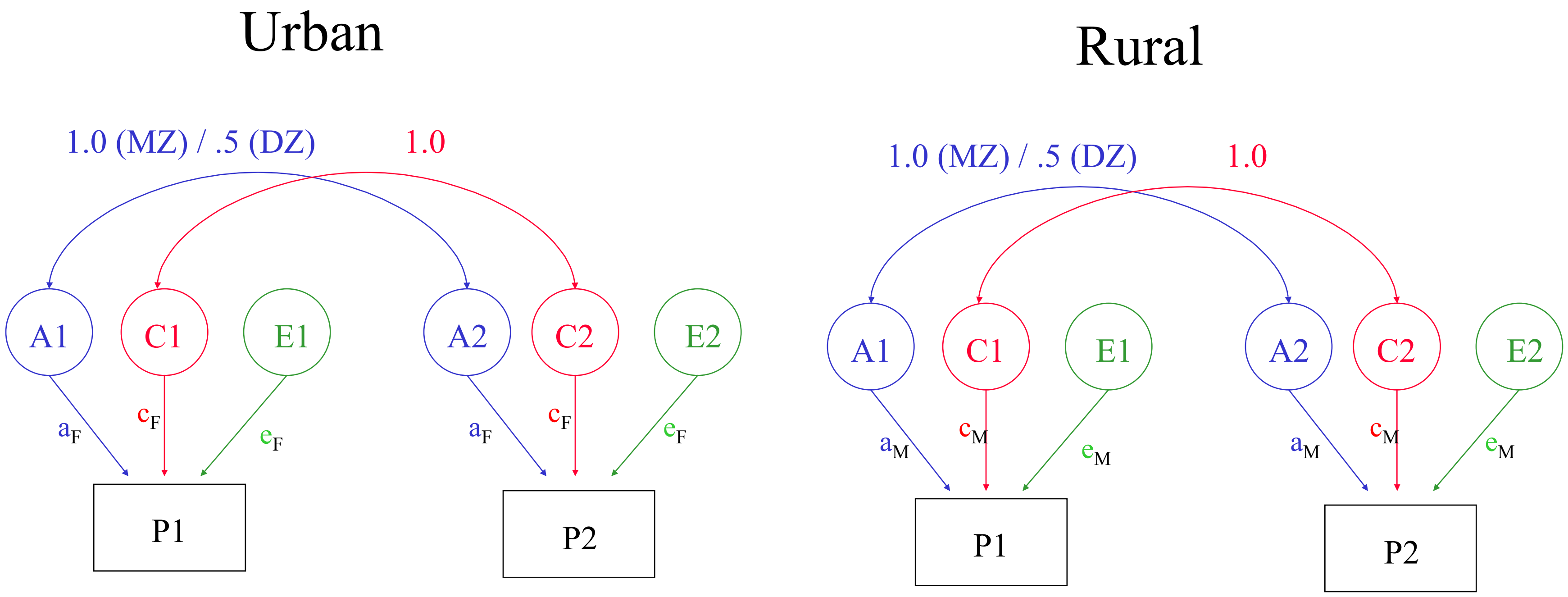


Old



$a_Y = a_O$? $c_Y = c_O$? $e_Y = e_O$?

GxE Effects



$a_U = a_R$? $c_U = c_R$? $e_U = e_R$?

Gene-Environment Interaction

- *Genetic control of sensitivity to the environment*
- *Environmental control of gene expression*

- Bottom line: nature of genetic effects differs among environments

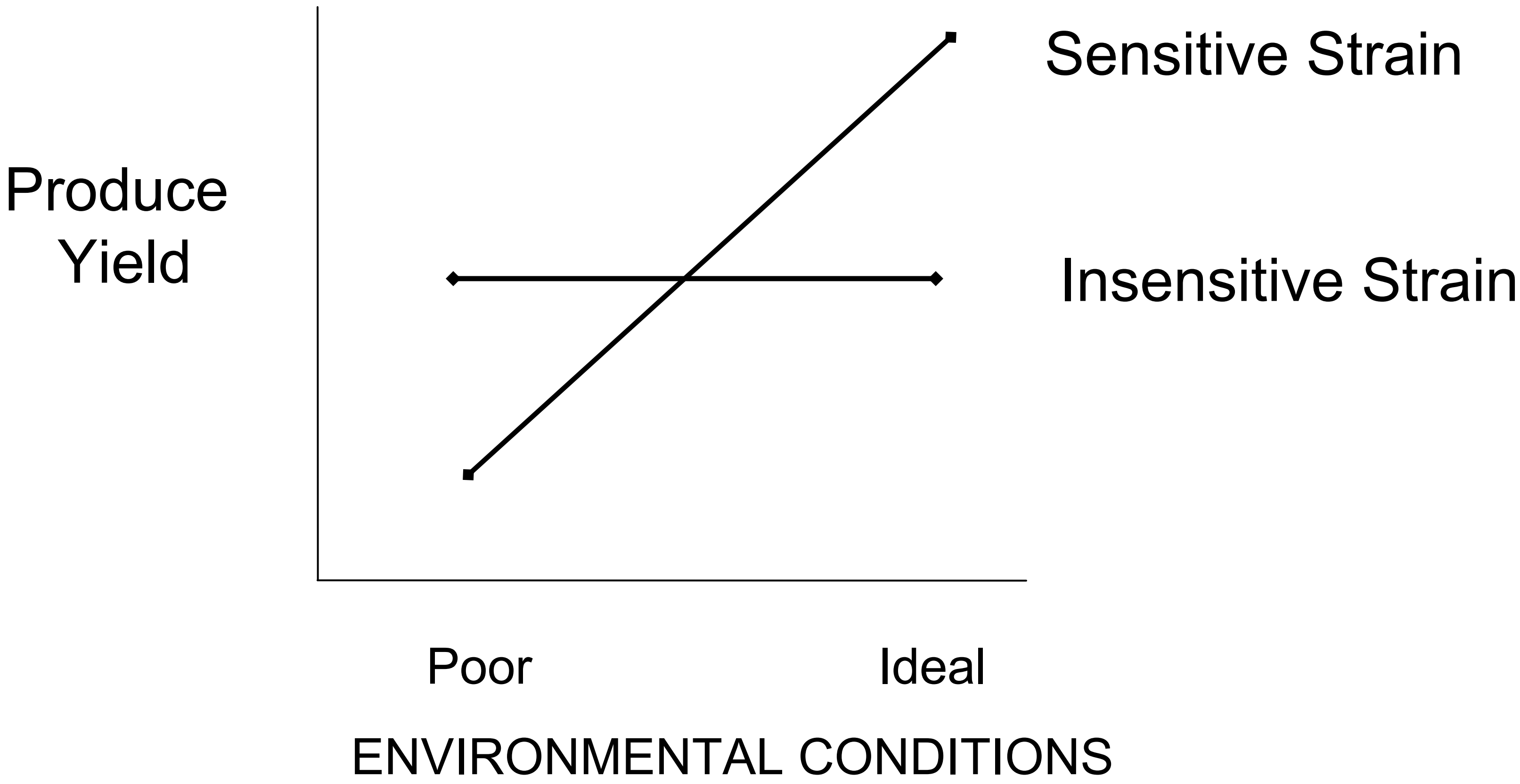
Gene-Environment Interaction

- *First observed by plant breeders:*
  - Sensitive strains – did great under ideal conditions (soil type, sunlight, rainfall),

    but very poorly under less than ideal circumstances

  - Insensitive strains – did OK regardless of the condition; did worse under ideal conditions but better under poor conditions

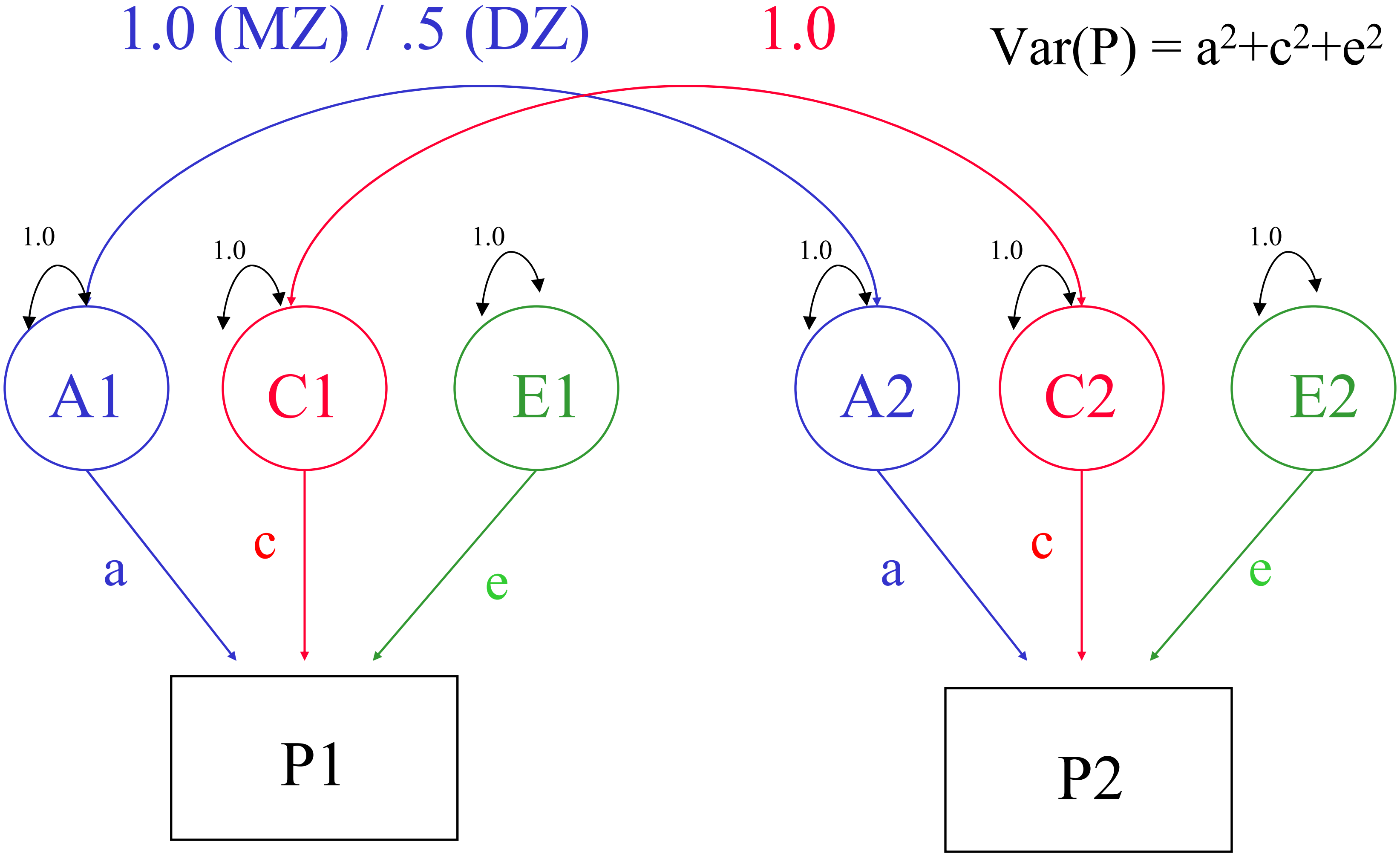
Conceptualizing Gene-Environment Interaction



Standard Univariate Model

$P = A + C + E$

$\mathrm{Var}(P) = a^2+c^2+e^2$



Contributions of Genetic, Shared Environment, Genotype x Environment Interaction Effects to Twin/Sib Resemblance

	Shared Environment	Additive Genetic Effects	Genotype x Shared Environment Interaction
MZ Pairs	1	1	1 x 1 = 1
DZ Pairs/Full Sibs	1	½	1 x ½ = ½

Contributions of Genetic, Shared Environment, Genotype x Environment Interaction Effects to Twin/Sib Resemblance

	Shared Environment	Additive Genetic Effects	Genotype x Shared Environment Interaction
MZ Pairs	1	1	1 x 1 = 1
DZ Pairs/Full Sibs	1	½	1 x ½ = ½

In other words—if gene-(shared) environment interaction is not explicitly modeled, it will be subsumed into the A term in the classic twin model.

Contributions of Genetic, Unshared Environment, Genotype x Unshared Environment Interaction Effects to Twin/Sib Resemblance

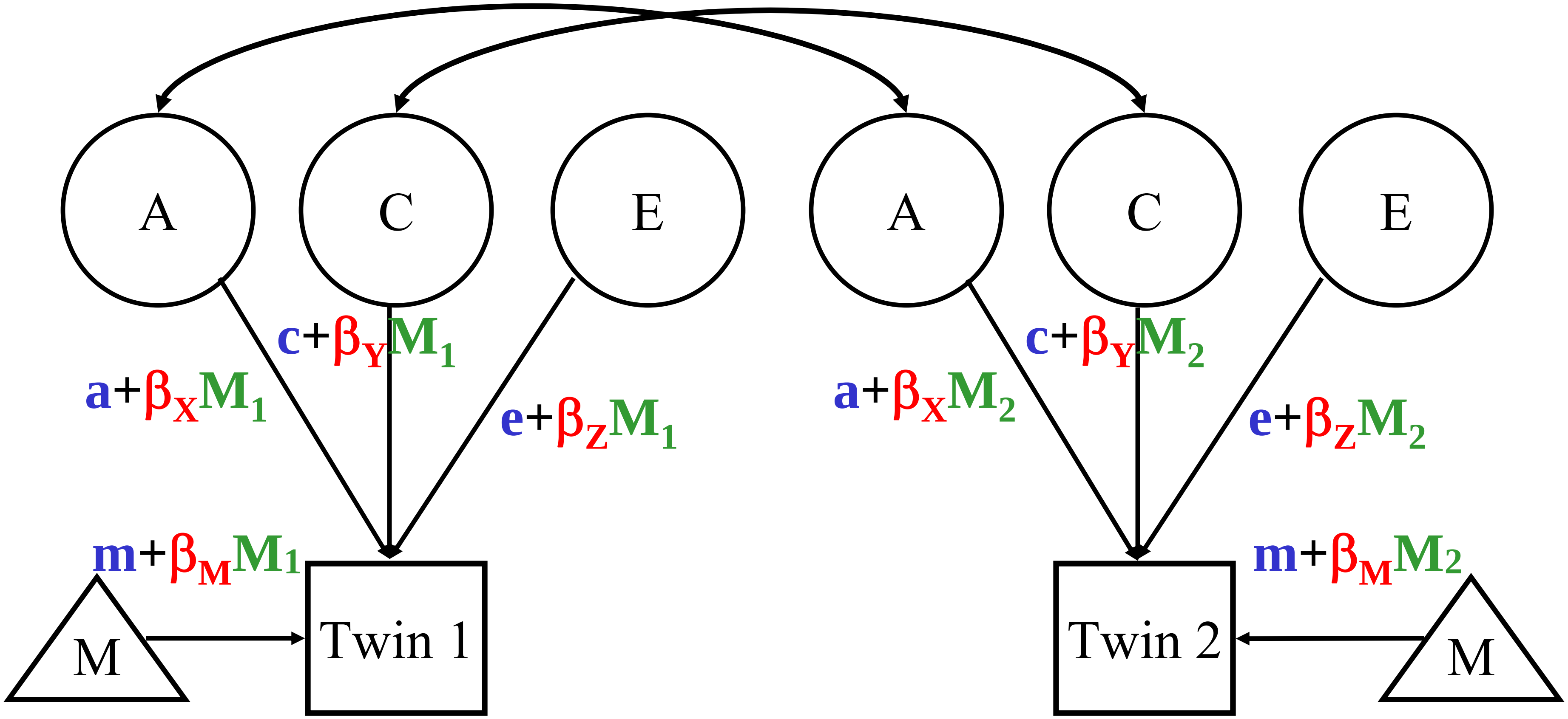
	Unshared (Unique) Environment	Additive Genetic Effects	Genotype x Unshared Environment Interaction
MZ Pairs	0	1	0 x 1 = 0
DZ Pairs/Full Sibs	0	½	0 x ½ = 0

Contributions of Genetic, Unshared Environment, Genotype x Unshared Environment Interaction Effects to Twin/Sib Resemblance

	Unshared (Unique) Environment	Additive Genetic Effects	Genotype x Unshared Environment Interaction
MZ Pairs	0	1	0 x 1 = 0
DZ Pairs/Full Sibs	0	½	0 x ½ = 0

If gene-(unshared) environment interaction is not explicitly modeled, it will be subsumed into the E term in the classic twin model.

ACE - XYZ - M



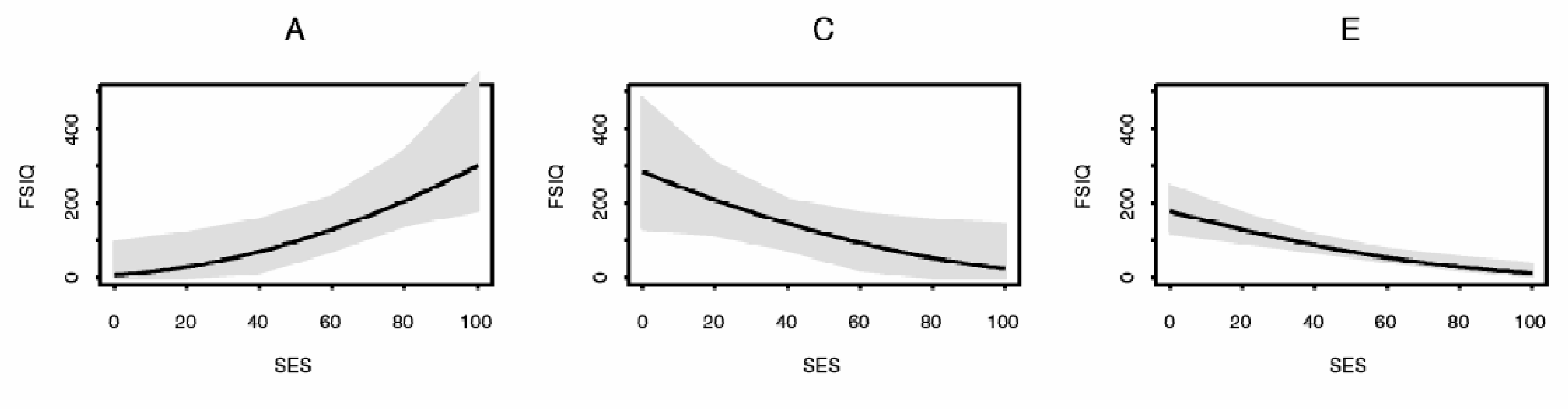
Main effects and moderating effects

Gene x (Quasi-Continuous) Environment Models

	a^2	c^2	e^2
Fewest 20-24 yr olds	0.11	0.57	0.32
Most 20-24 yr olds	0.61	0.01	0.38
Lowest migration	0.17	0.48	0.35
Highest migration	0.64	0.01	0.35

Dick et al., 2001

Turkheimer *et al* (2003)



Testing for GxE Moderation

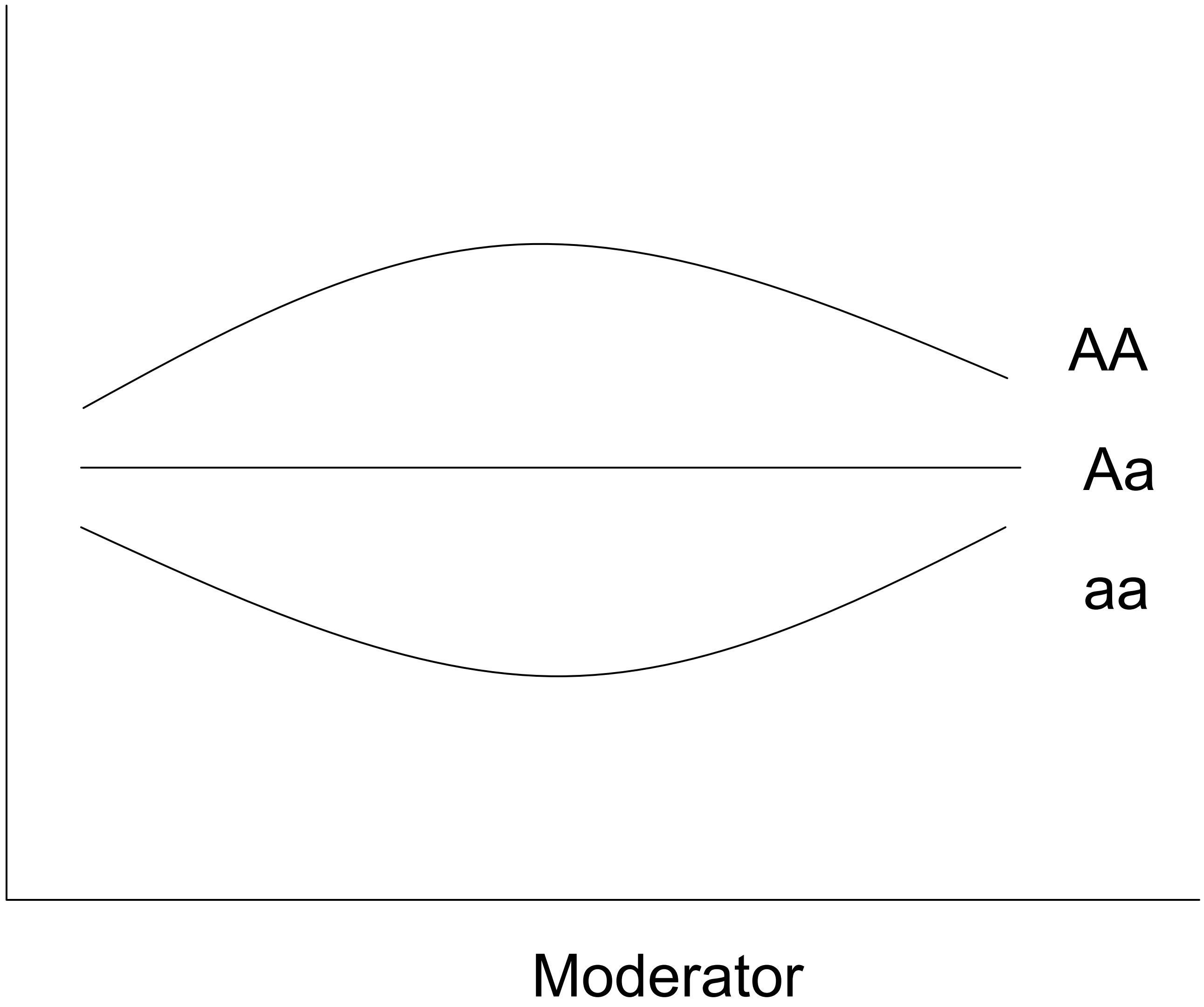
- Continuous data
  - Moderation of means and variance

- Ordinal data
  - Moderation of thresholds and variance

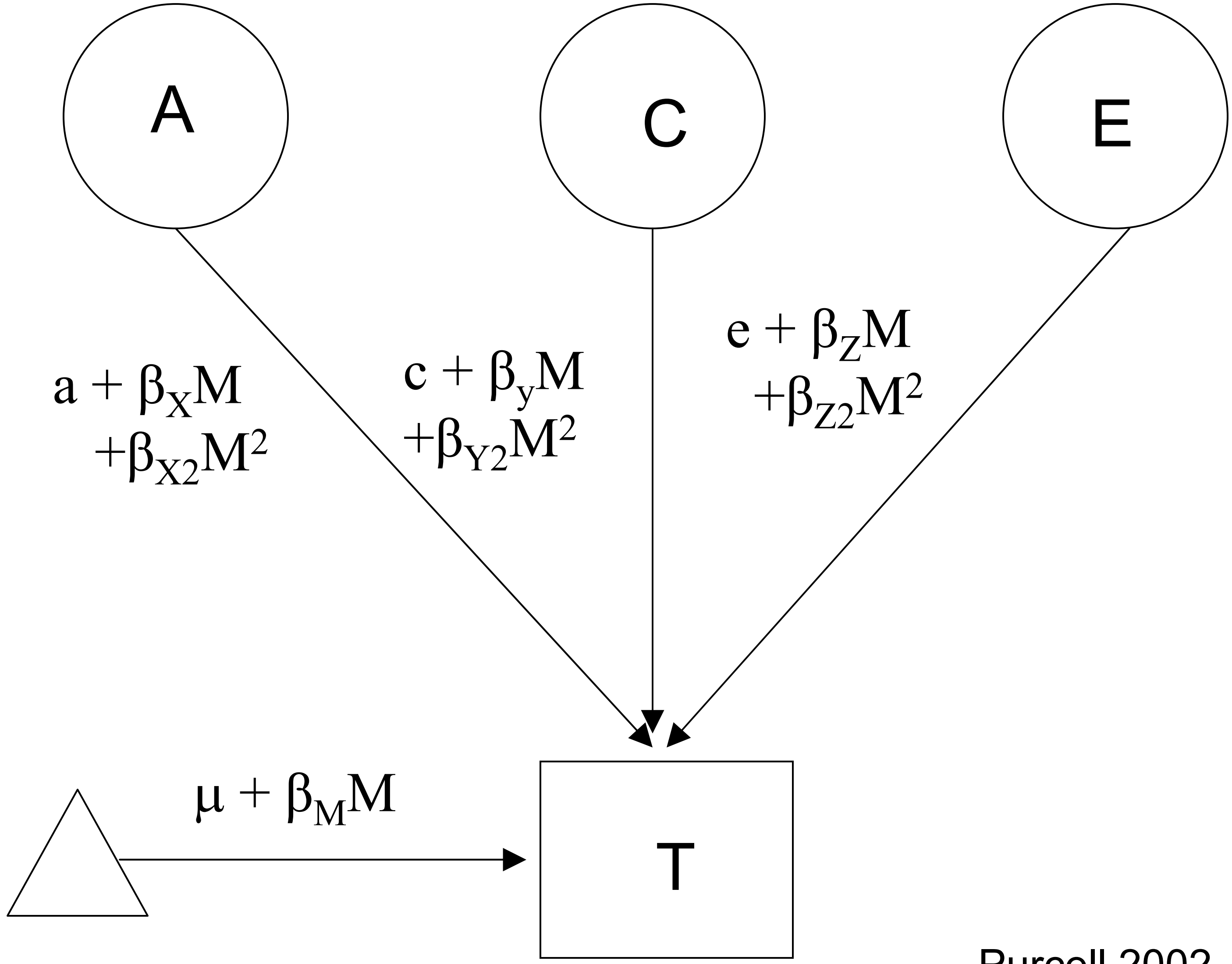
Expanding the Model

- Nonlinear Interaction
- Gene-Environment Correlation

Nonlinear Moderation



Nonlinear Moderation can be modeled with the addition of a quadratic term


A

C

E

T

$a + \beta_X M + \beta_{X2} M^2$

$c + \beta_y M + \beta_{Y2} M^2$

$e + \beta_Z M + \beta_{Z2} M^2$

$\mu + \beta_M M$

Purcell 2002

Gene-environment Interaction

- *Genetic control of sensitivity to the environment*
- *Environmental control of gene expression*

Gene-environment Correlation

- *Genetic control of exposure to the environment*
- *Environmental control of gene frequency*

This complicates interpretation of GxE effects

- If there is a correlation between the moderator (environment) of interest and the outcome,

  and you find a GxE effect, it's not clear if:

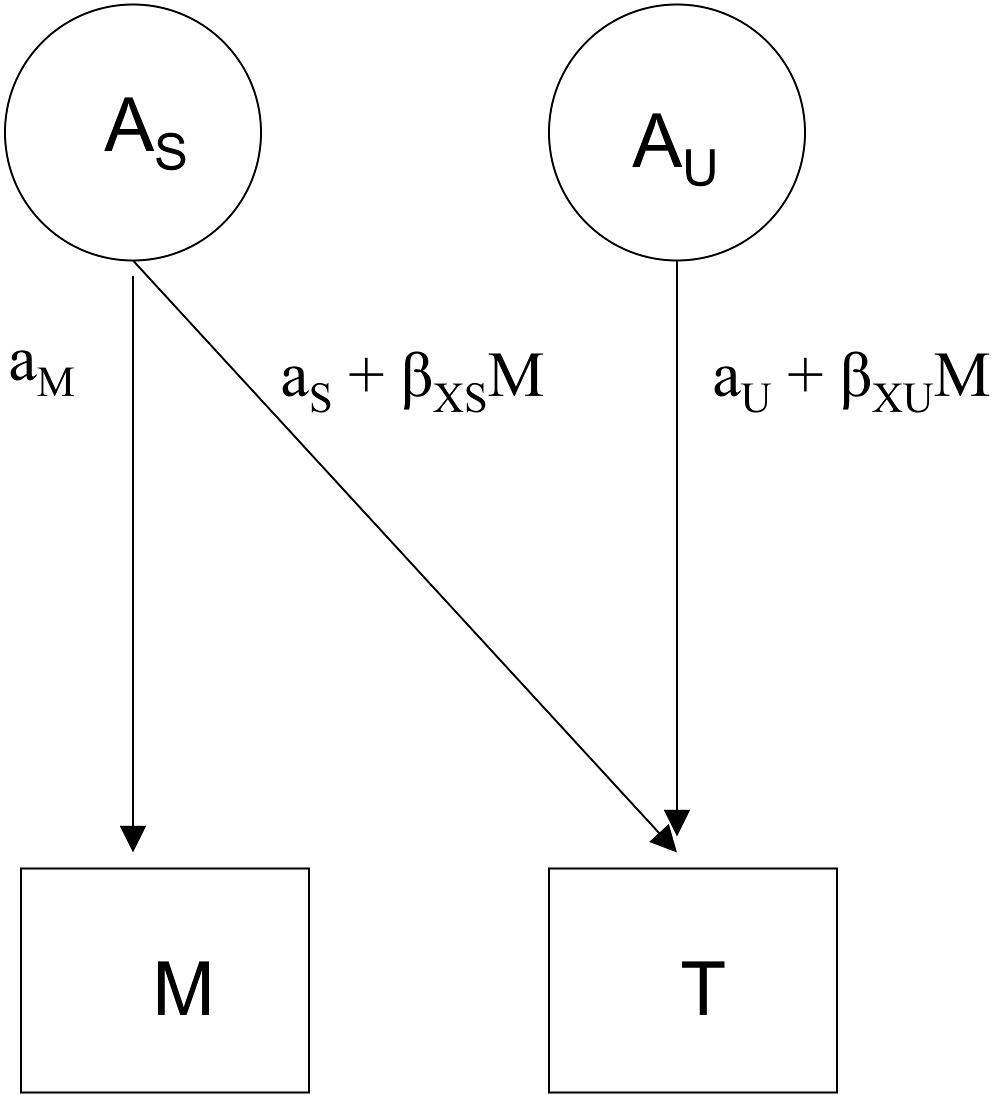
  - The environment is moderating the effects of genes

    or

  - Trait-influencing genes are simply more likely to be present in that environment

Ways to deal with rGE

- Limit study to moderators that aren't correlated with outcome
  - Pro: easy
  - Con: not very satisfying
- Moderator in means model will remove from the covariance genetic effects shared by trait and moderator
  - Pro: Any interaction detected will be moderation of the trait specific genetic effects
  - Con: Will fail to detect GxE interaction if the moderated genetic component is shared by the outcome and moderator
- Explicitly model rGE using a bivariate framework



β_{XS} indicates moderation of shared genetic effects

B_{XU} indicates moderation of unique genetic effects on trait of interest

Ways to deal with rGE

- Limit study to moderators that aren't correlated with outcome
  - Pro: easy
  - Con: not very satisfying
- Moderator in means model will remove from the covariance genetic effects shared by trait and moderator
  - Pro: Any interaction detected will be moderation of the trait specific genetic effects
  - Con: Will fail to detect GxE interaction if the moderated genetic component is shared by the outcome and moderator
- Explicitly model rGE using a bivariate framework
  - Pro: explicitly models rGE
  - Con: Power to detect B_{XU} decreases with increasing rGE; difficulty converging
  - NOTE: THIS MODEL IS NOT INFORMATIVE FOR FAMILY-LEVEL VARIABLES (E.G., SES, PARENTING, ETC)

Final Things to Consider

Unstandardized versus standardized effects

	ENVIRONMENT 1		ENVIRONMENT 2	
	Unstandardized Variance	Standardized Variance	Unstandardized Variance	Standardized Variance
Genetic	60	0.60	60	0.30
Common environmental	35	0.35	70	0.35
Unique environmental	5	0.05	70	0.05
Total variance	100		200	

Final Things to Consider

- Unstandardized versus standardized effects
- Don’t forget about theory!

Final Things to Consider

- Unstandardized versus standardized effects
- Don’t forget about theory!
  - “Moderation in all things….including moderation”

    -Mike Neale

http://pngu.mgh.harvard.edu/%7Epurcell/gxe/

Confidence intervals

- Easy to get CIs for individual parameters

- Additionally, CIs on the moderated VCs are useful for interpretation

- e.g. a 95% CI for $(a+\beta M)^2$, for a specific M

- Define two extra vectors in Group 1

```
P full 1 13
O Unit 1 13
Matrix P -3 -2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5 3
```

- Add a 4th group to calculate the CIs

```
CIs
Calc
Matrices = Group 1
Begin Algebra;
 F= ( A@O + T@P ) . ( A@O + T@P ) /
 G= ( C@O + U@P ) . ( C@O + U@P ) /
 I= ( E@O + V@P ) . ( E@O + V@P ) /
End Algebra;
Interval @ 95 F 1 1 to F 1 13
Interval @ 95 G 1 1 to G 1 13
Interval @ 95 I 1 1 to I 1 13
End;
```

Calculation of CIs

```
F= ( A@O + T@P ) . ( A@O + T@P ) /
```

- E.g. if P were $\begin{pmatrix} 0 & 1 & 2 \end{pmatrix}$

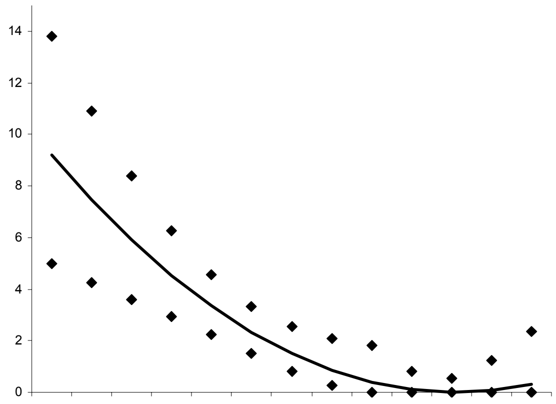
then `( A@O + T@P )` equals $\left(a \otimes \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} + x \otimes \begin{pmatrix} 0 & 1 & 2 \end{pmatrix}\right)$

or $\begin{pmatrix} a & a & a \end{pmatrix} + \begin{pmatrix} 0 & x & 2x \end{pmatrix}$

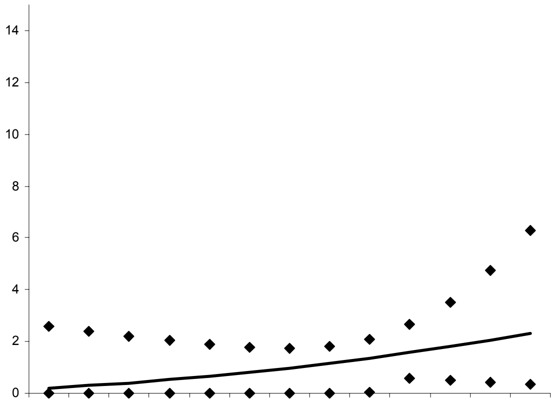
or $\begin{pmatrix} a & a+x & a+2x \end{pmatrix}$

Finally, the dot-product squares all elements to give $\begin{pmatrix} a^2 & (a+x)^2 & (a+2x)^2 \end{pmatrix}$

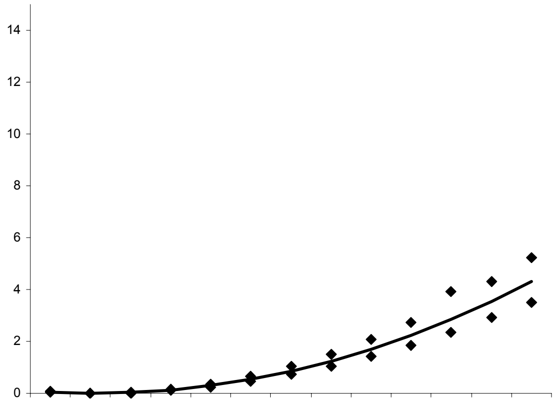
Confidence intervals on VCs



A



C



E

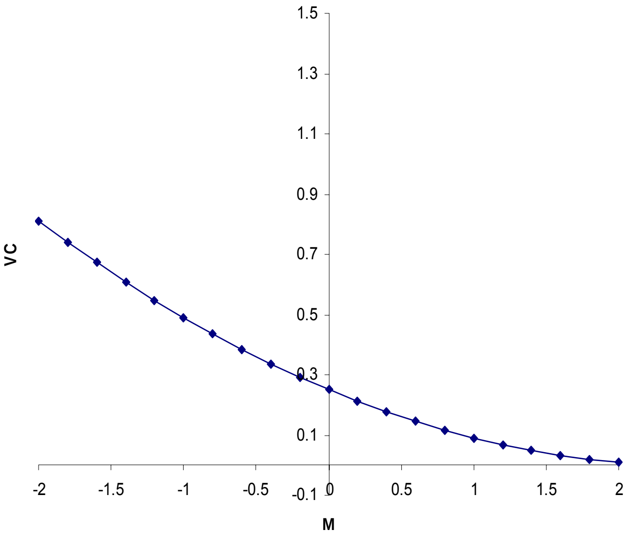
Plotting VCs

- For the additive genetic VC, for example
  - Given *a*, *β* and a range of values for the moderator variable

- For example,

  $a = 0.5$, $\beta = -0.2$ and M ranges from -2 to +2

M	$(a+\beta M)^2$	$(a+\beta M)^2$
-2	$(0.5+(-0.2\times-2))^2$	0.81
-1.5	$(0.5+(-0.2\times-1.5))^2$	0.73
…		
+2	$(0.5+(-0.2\times2))^2$	0.01



Sample Gene-Environment Interaction