

Heterogeneity

Danielle Dick, Hermine Maes,
Sarah Medland, Danielle Posthuma, et al!

Boulder Twin Workshop
March 2008

Heterogeneity Questions I

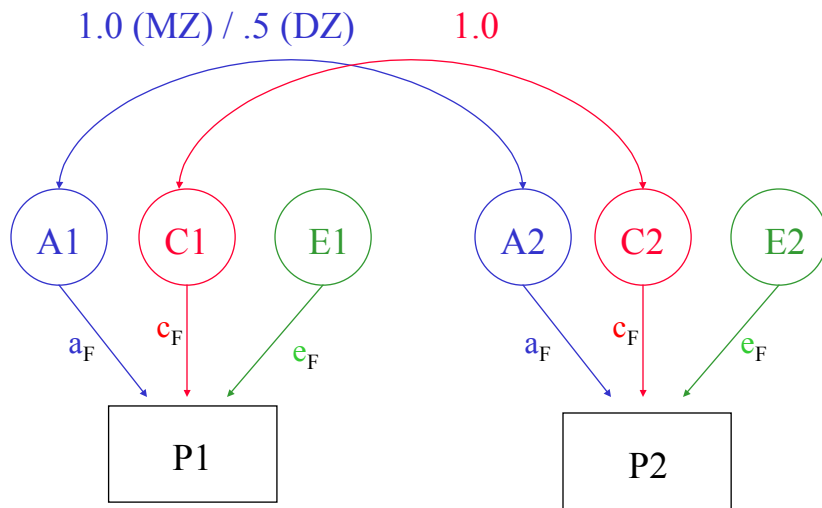
- Univariate Analysis: What are the contributions of additive genetic, dominance/shared environmental and unique environmental factors to the variance?
- Are the contributions of genetic and environmental factors equal for different groups, such as sex, race, ethnicity, SES, environmental exposure, etc.?

Ways to Model Heterogeneity in Twin Data

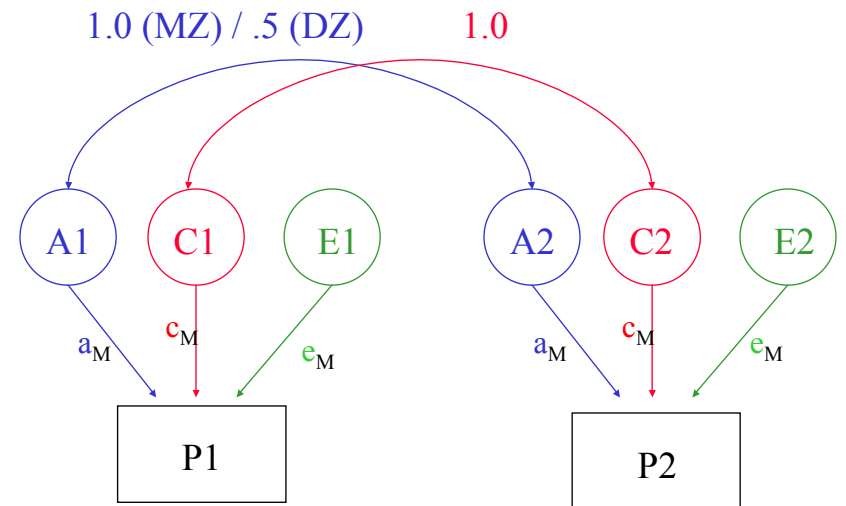
- Multiple Group Models
 - Sex Effects
 - Young/Old cohorts
 - Urban/Rural residency

Sex Effects

Females



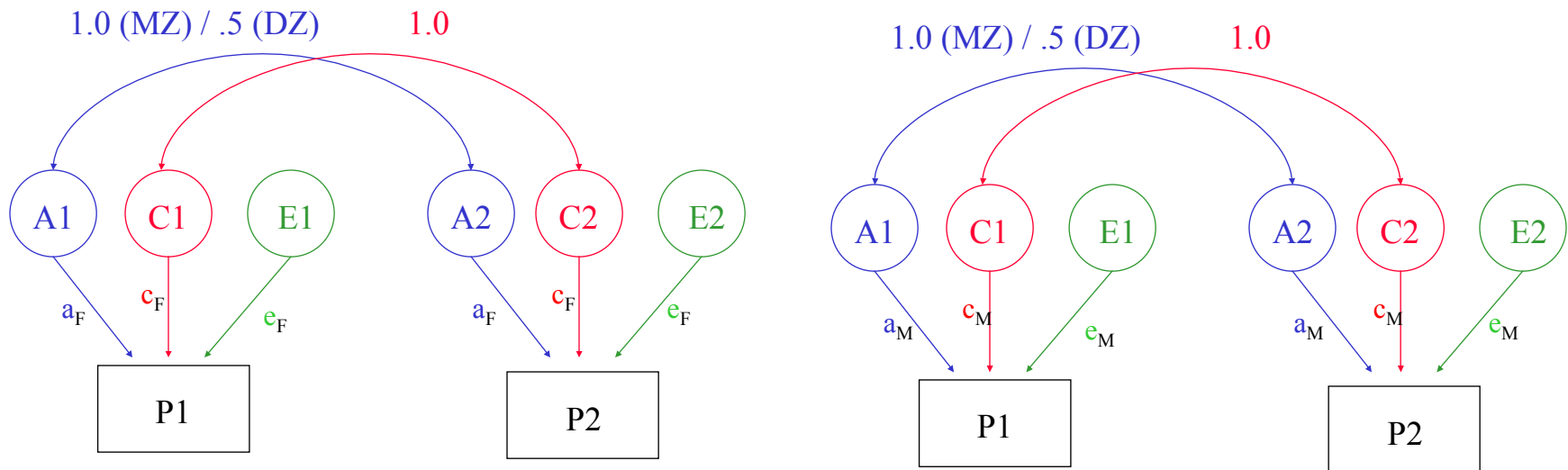
Males



Sex Effects

Females

Males



$$a_F = a_M ?$$

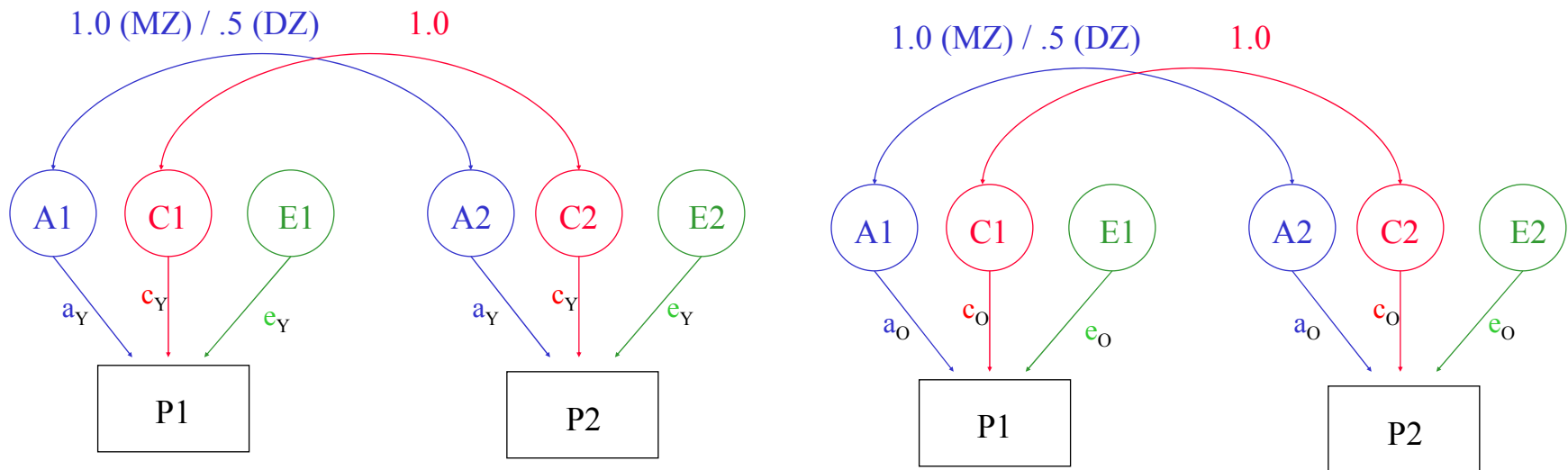
$$c_F = c_M ?$$

$$e_F = e_M ?$$

Age Effects

Young

Old



$$a_Y = a_O ?$$

$$c_Y = c_O ?$$

$$e_Y = e_O ?$$

Exercise I: modifying the script to test for age heterogeneity

- Open **bmi_young.mx** (in \\workshop\dd\)
- This script: young males, 4 groups:
 - 1 = calculation group – matrix declarations
 - 2 = MZ data
 - 3 = DZ data
 - 4 = calculation group – standardized solution
- ADE model, 1 grand mean, so 4 estimated parameters

Exercise I: modifying the script to test for age heterogeneity

- Change this script so it will allow you to estimate ADE in the young and the older cohort by adding four groups for the older cohort
- Then run it
- If done correctly you should get **-2ll = 3756.552** and **df = 1759**

Required modifications for Exercise I

- Copy and paste all 4 groups
- Change **Select if agecat=2** in the two new data groups
- Change **matrices = group 5** in the two new data groups
- Change **#ngroups = 8**

Exercise II: Testing AE model – Significant Differences b/w Young & Old?

- In `bmi_young2.mx`, D has been fixed (it was not significant), so an AE model is estimated
- Check the estimates of A and E in the young and old cohort under the AE model

	A	E
Young - unstandardized		
Young standardized		
Old – unstandardized		
Old - standardized		

	A	E
Young - unstandardized	0.5413	0.1414
Young standardized	0.7924	0.2076
Old – unstandardized	0.4330	0.1815
Old - standardized	0.7046	0.2954

Unstandardized versus standardized effects

	GROUP 1		GROUP 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	

Exercise II: Equality of variance components across age cohorts

Add the option multiples with EQUATE command, use a **get** in between and test whether

- $a_{\text{young}} = a_{\text{old}} ?$
- $e_{\text{young}} = e_{\text{old}} ?$

Exercise II: Fit results

- $a_{\text{young}} = a_{\text{old}} ?$

EQ X 1 1 1 X 5 1 1

end

Chi-squared 4.093

d.f. 1

Probability 0.043

get AE_cohort.mxs

- $e_{\text{young}} = e_{\text{old}} ?$

EQ Z 1 1 1 Z 5 1 1

End

Chi-squared 3.954

d.f. 1

Probability 0.047

	A	E
Young - unstandardized	0.5413	0.1414
Young standardized	0.7924	0.2076
Old – unstandardized	0.4330	0.1815
Old - standardized	0.7046	0.2954

Testing Standardized estimates

- Note: As the standardized parameters are calculated we cannot change them in an option multiple, and cannot use the EQ statement. Instead we need to use a constraint group:

Title G9: Constraint

Constraint

Begin Matrices ;

S comp 1 4 =S4 ;

T comp 1 4 =S8 ;

End matrices ;

Constraint S=T;

Option sat=3760.030, 1761

Option df = -3

End

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 may lose 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)

Definition variables used as covariates

General model with age and sex as covariates:

$$y_i = \alpha + \beta_1(\text{age}_i) + \beta_2 (\text{sex}_i) + \varepsilon$$

Where y_i is the observed score of individual i , α is the intercept or grand mean, β_1 is the regression weight of age, age_i is the age of individual i , β_2 is the deviation of males (if sex is coded 0=female; 1=male), sex_i is the sex of individual i , and ε is the residual that is not explained by the covariates (and can be decomposed further into ACE etc).

Standard model

- Means vector

$$(m \quad m)$$

- 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

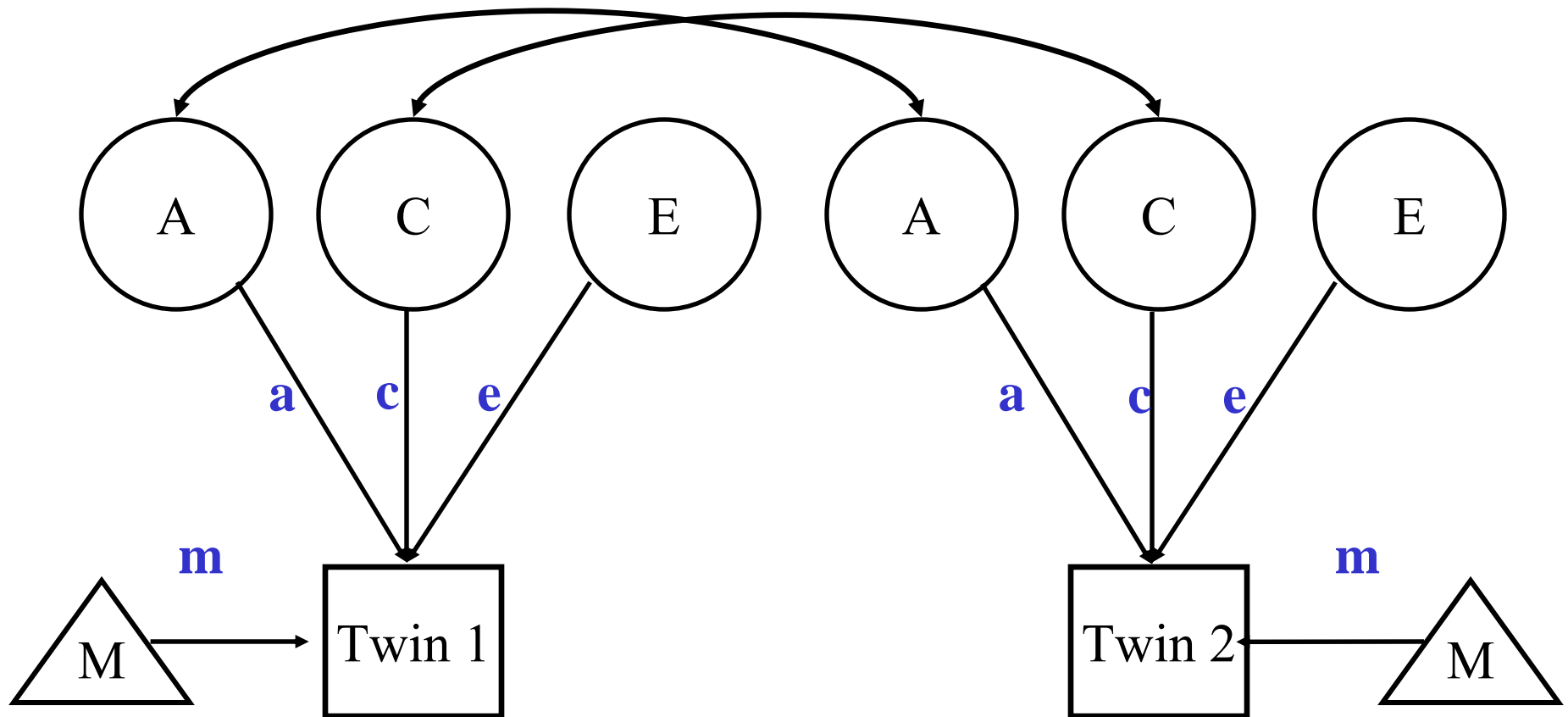
- Means vector

$$\left(m + \beta X_{1i} \quad m + \beta X_{2i} \right)$$

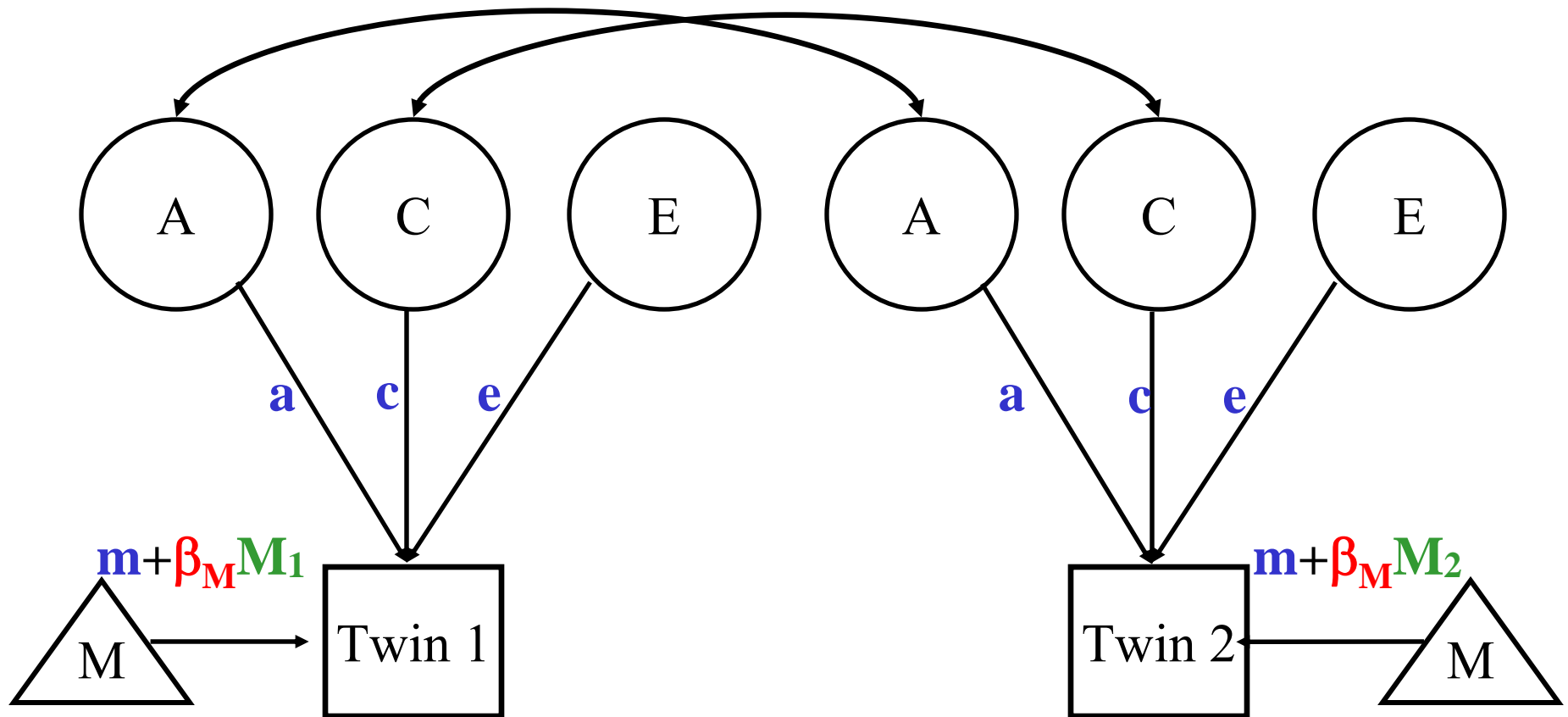
- Covariance matrix

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

Model-fitting approach to GxE



Adding Covariates to Means Model



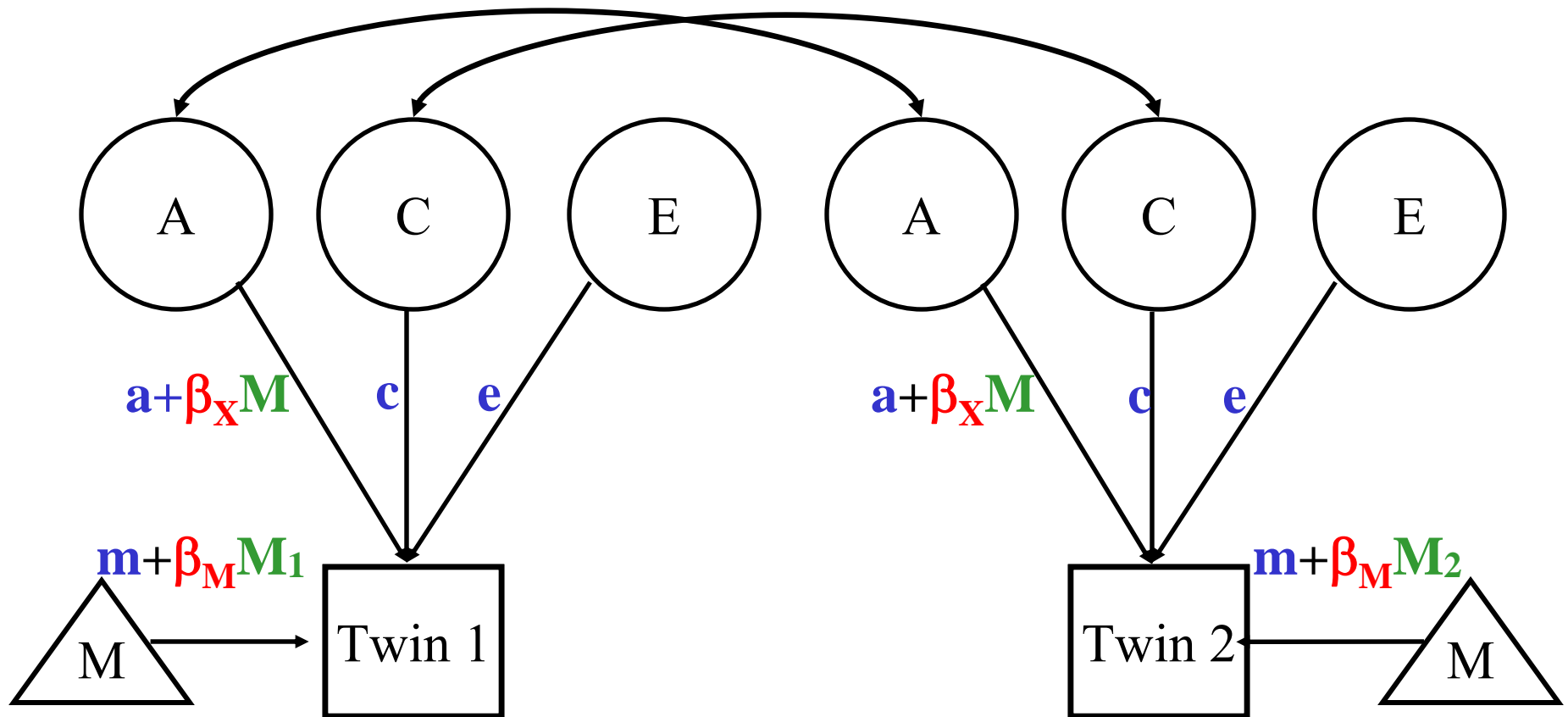
‘Definition variables’ in Mx

create dynamic var/cov structure

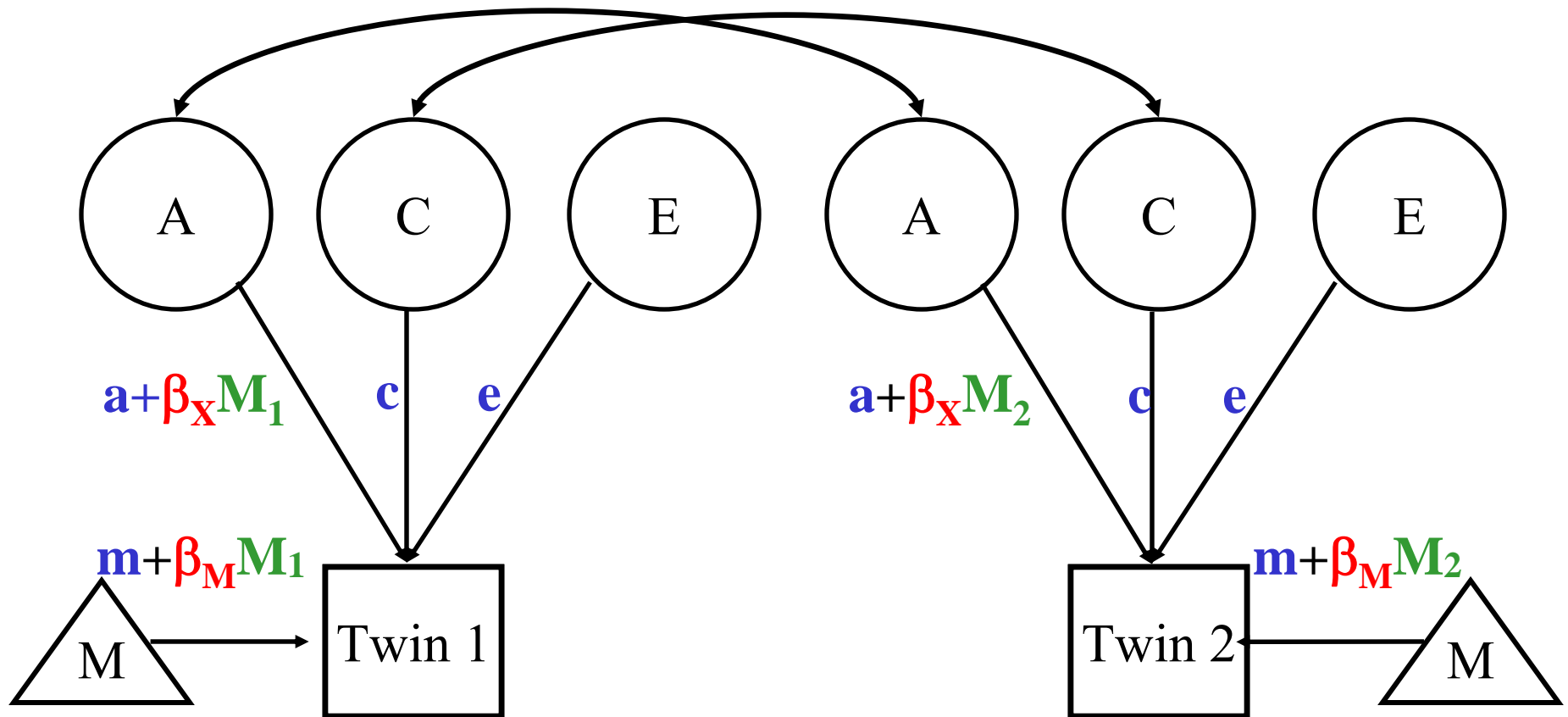
- Common uses:

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)

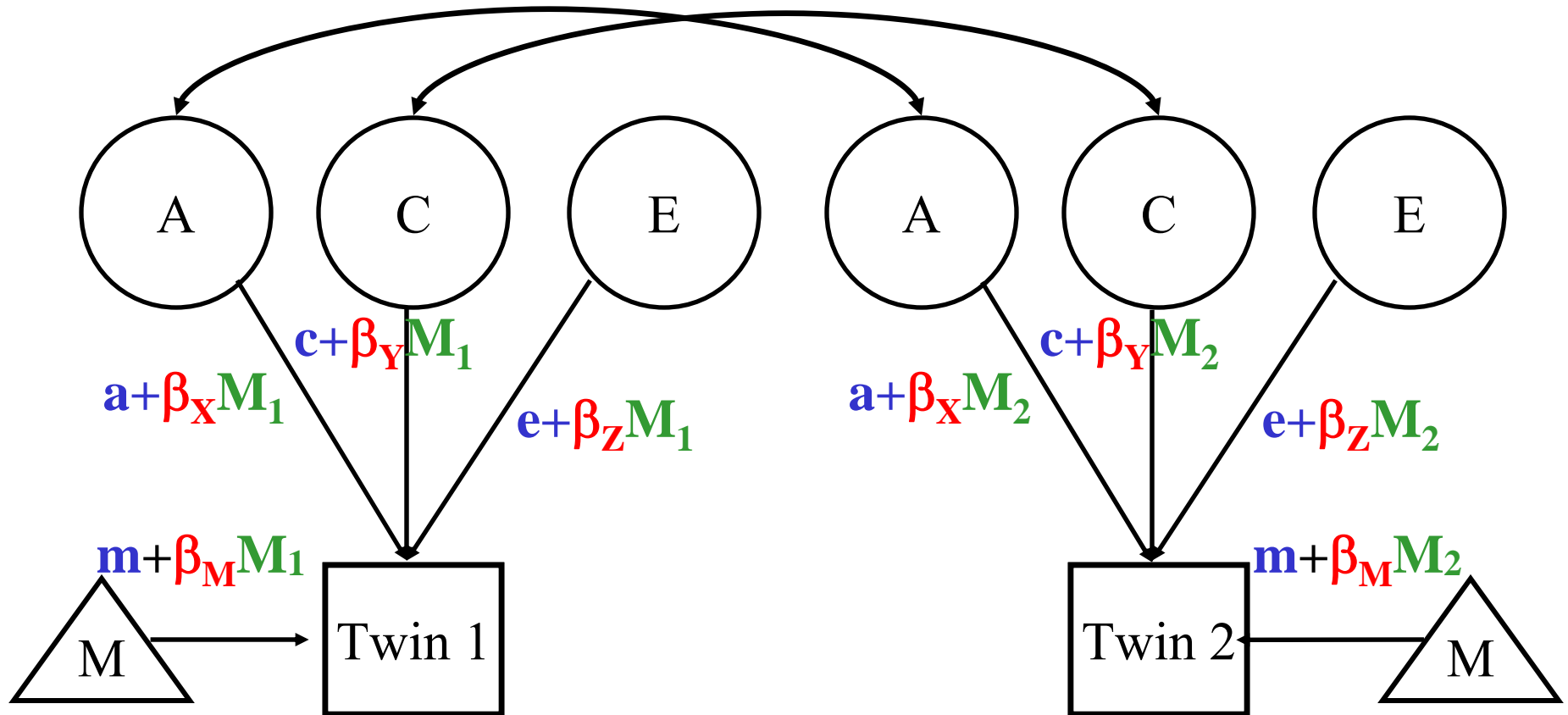
Model-fitting approach to GxE



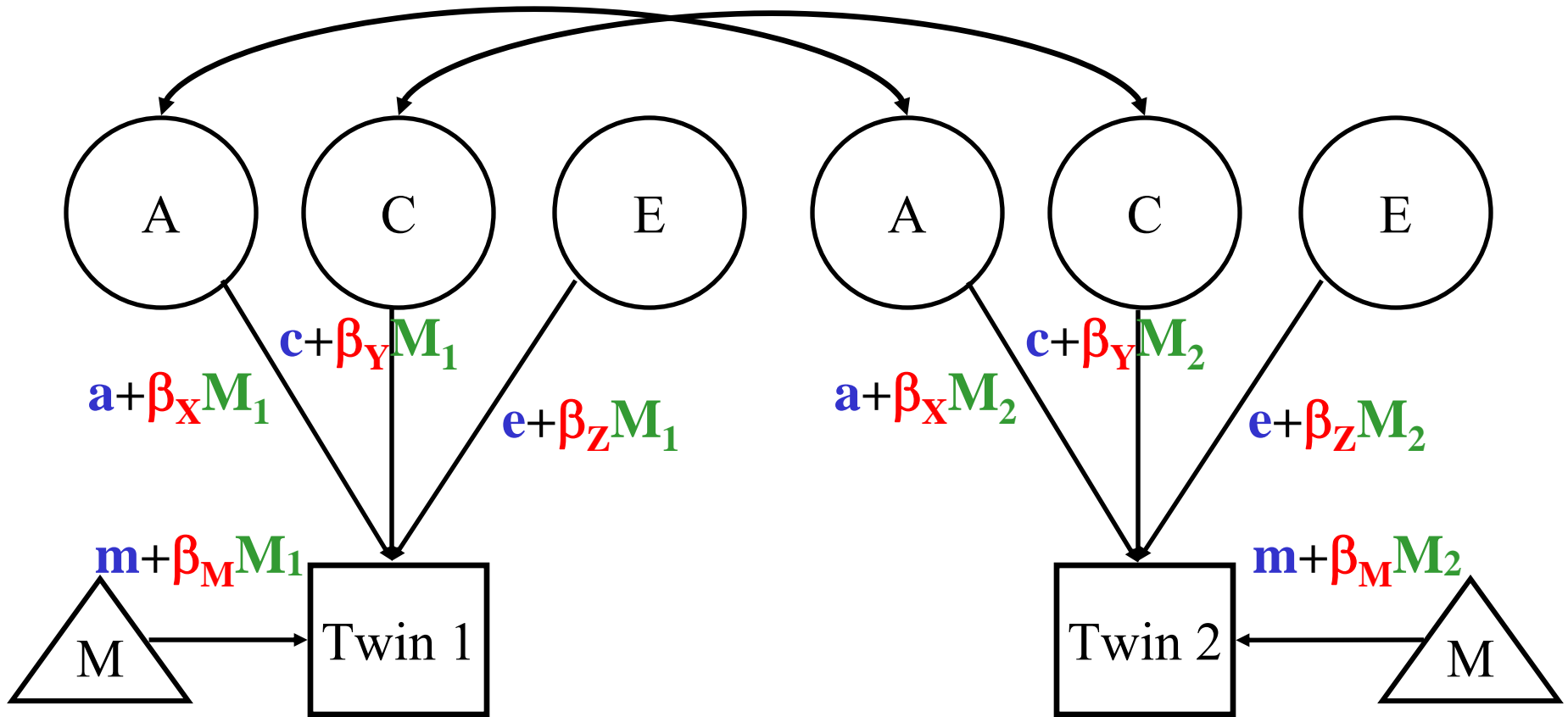
Individual specific moderators



E x E interactions



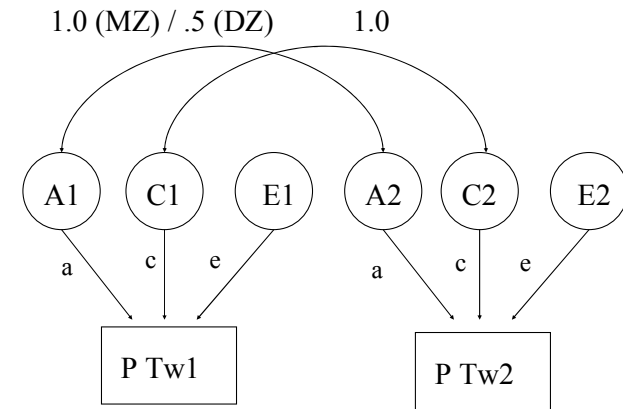
ACE - XYZ - M



Main effects and moderating effects

- Classic Twin Model:

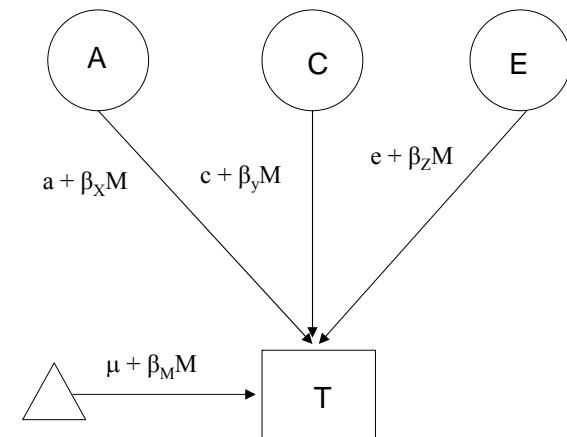
$$\text{Var (P)} = a^2 + c^2 + e^2$$



- Moderation Model:

$$\text{Var (P)} =$$

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



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

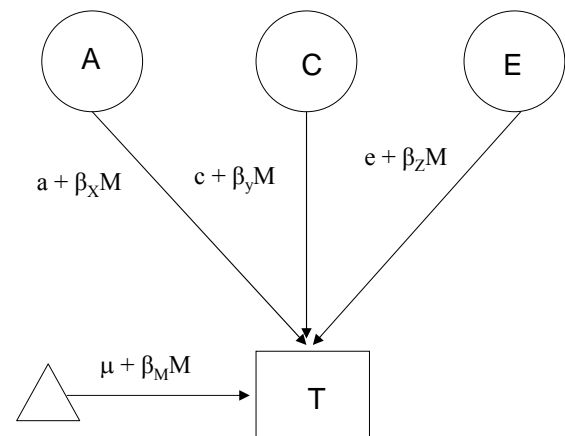
Where M is the value of the moderator and

Significance of β_X indicates genetic moderation

Significance of β_Y indicates common environmental moderation

Significance of β_Z indicates unique environmental moderation

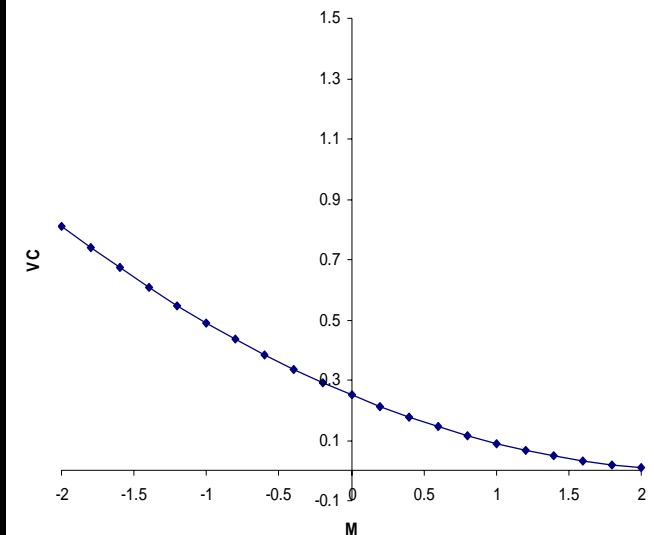
B_M indicates a main effect of the moderator on the mean



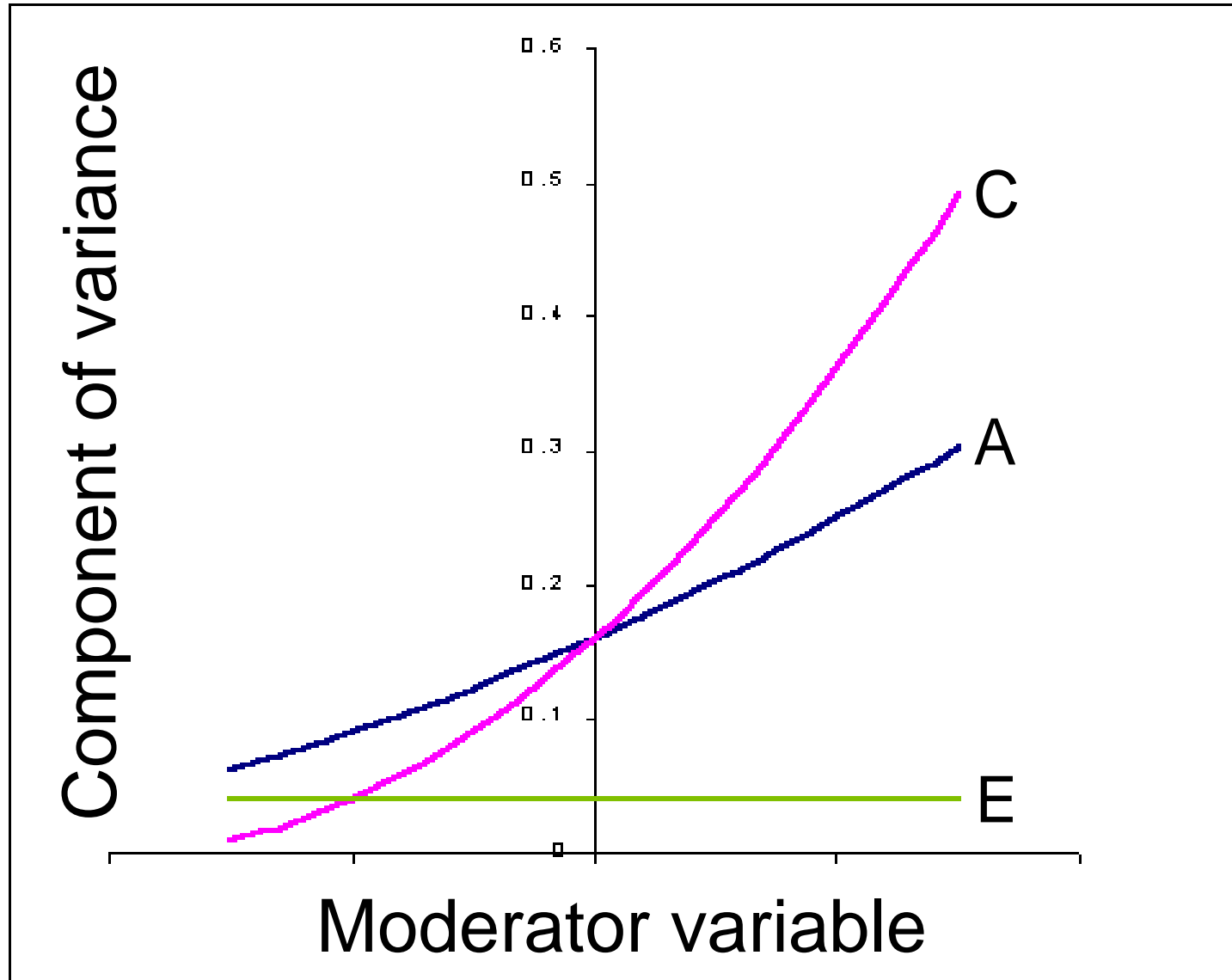
Plotting VCs as Function of Moderator

- 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

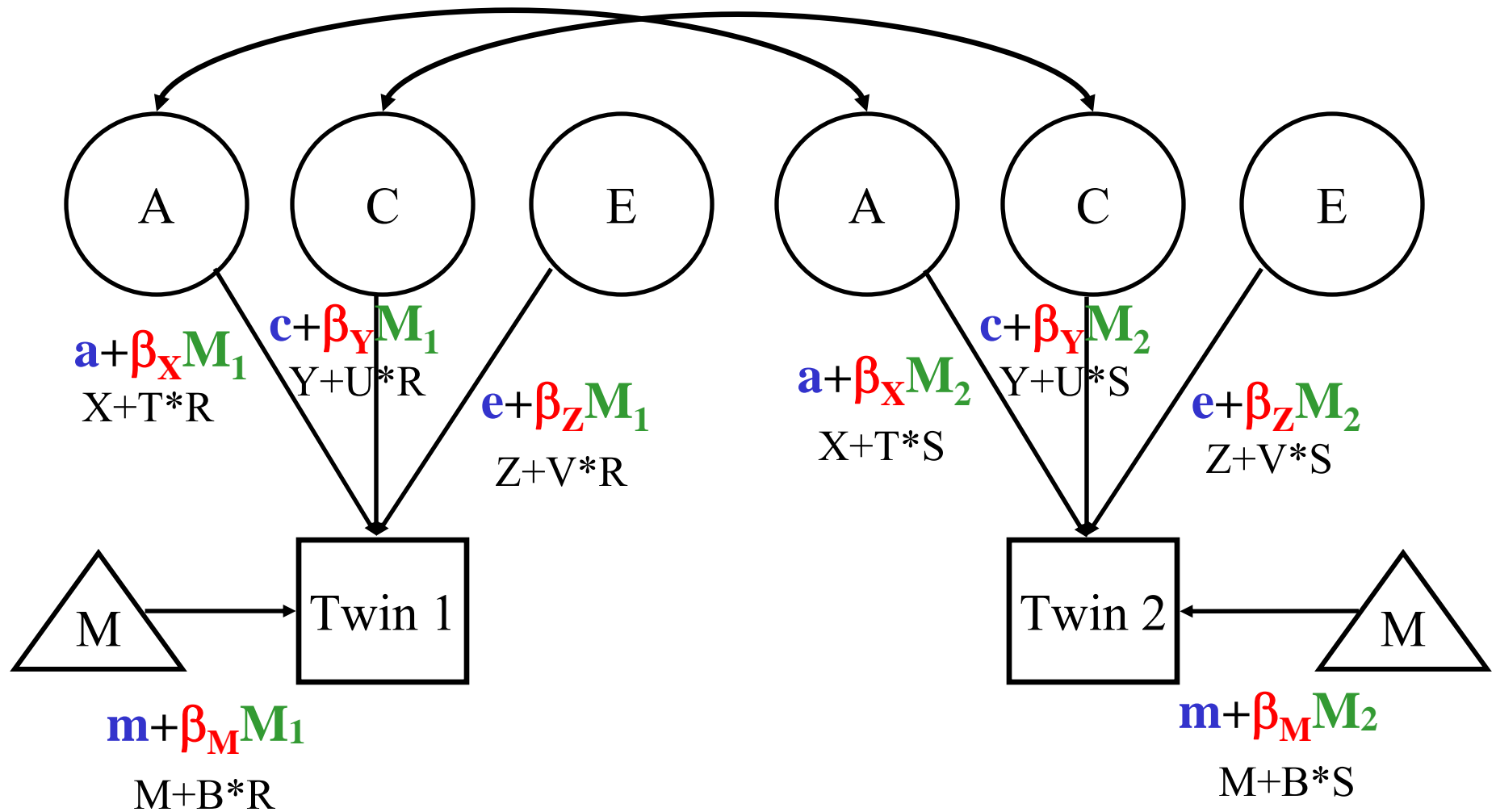
M	$(a+\beta M)^2$	$(a+\beta M)^2$
-2	$(0.5+(-0.2 \times -2))^2$	0.81
-1.5	$(0.5+(-0.2 \times -1.5))^2$	0.73
...		
+2	$(0.5+(-0.2 \times 2))^2$	0.01



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) /

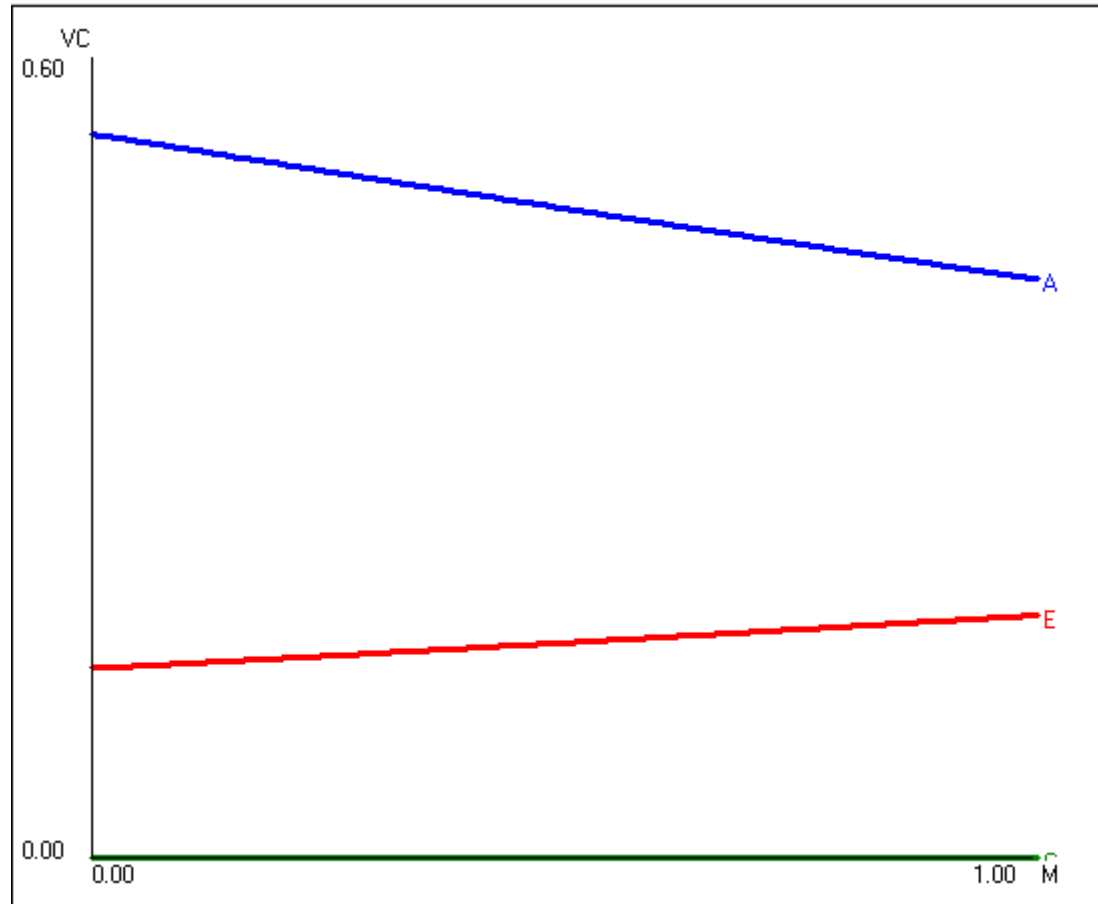
!twin 1 moderator variable
Specify R -1
!twin 2 moderator variable
Specify S -2
Options NO_Output
End

```

Practical

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

Cohort Moderation



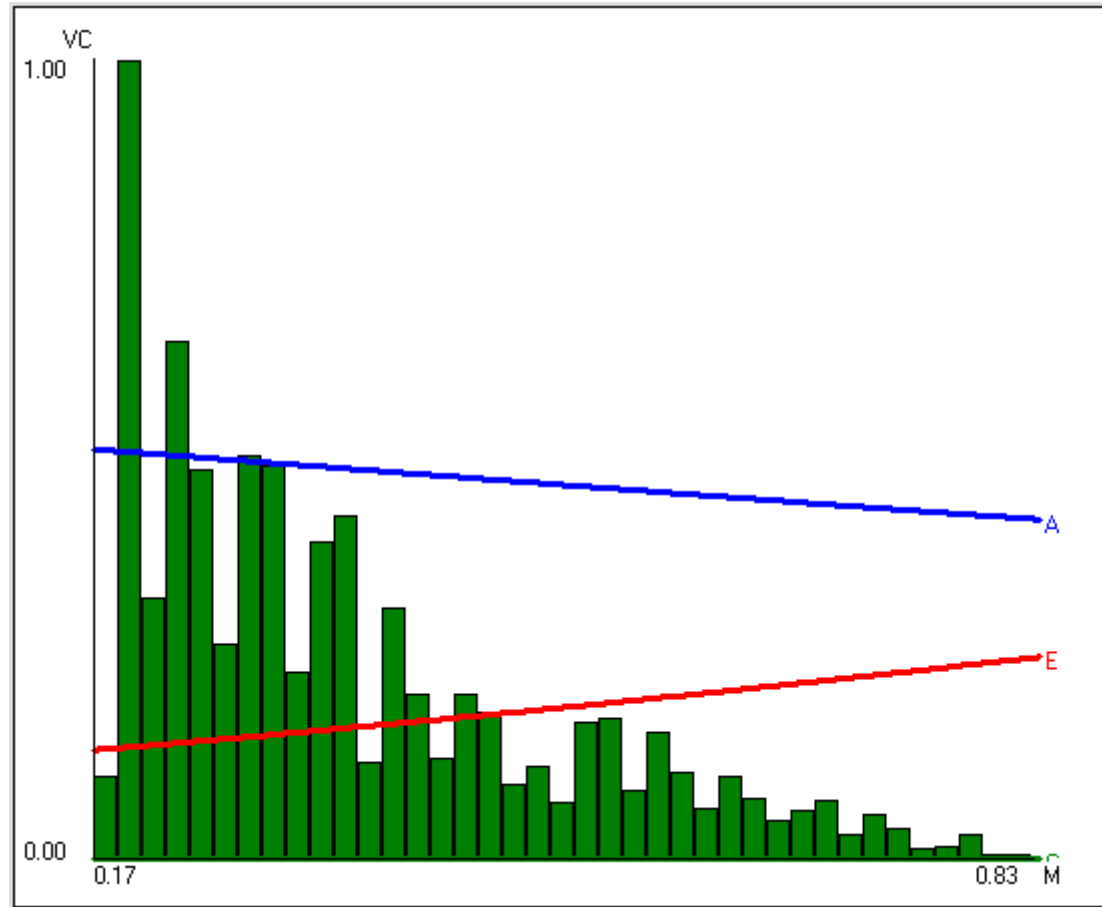
Younger Cohort

Older Cohort

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

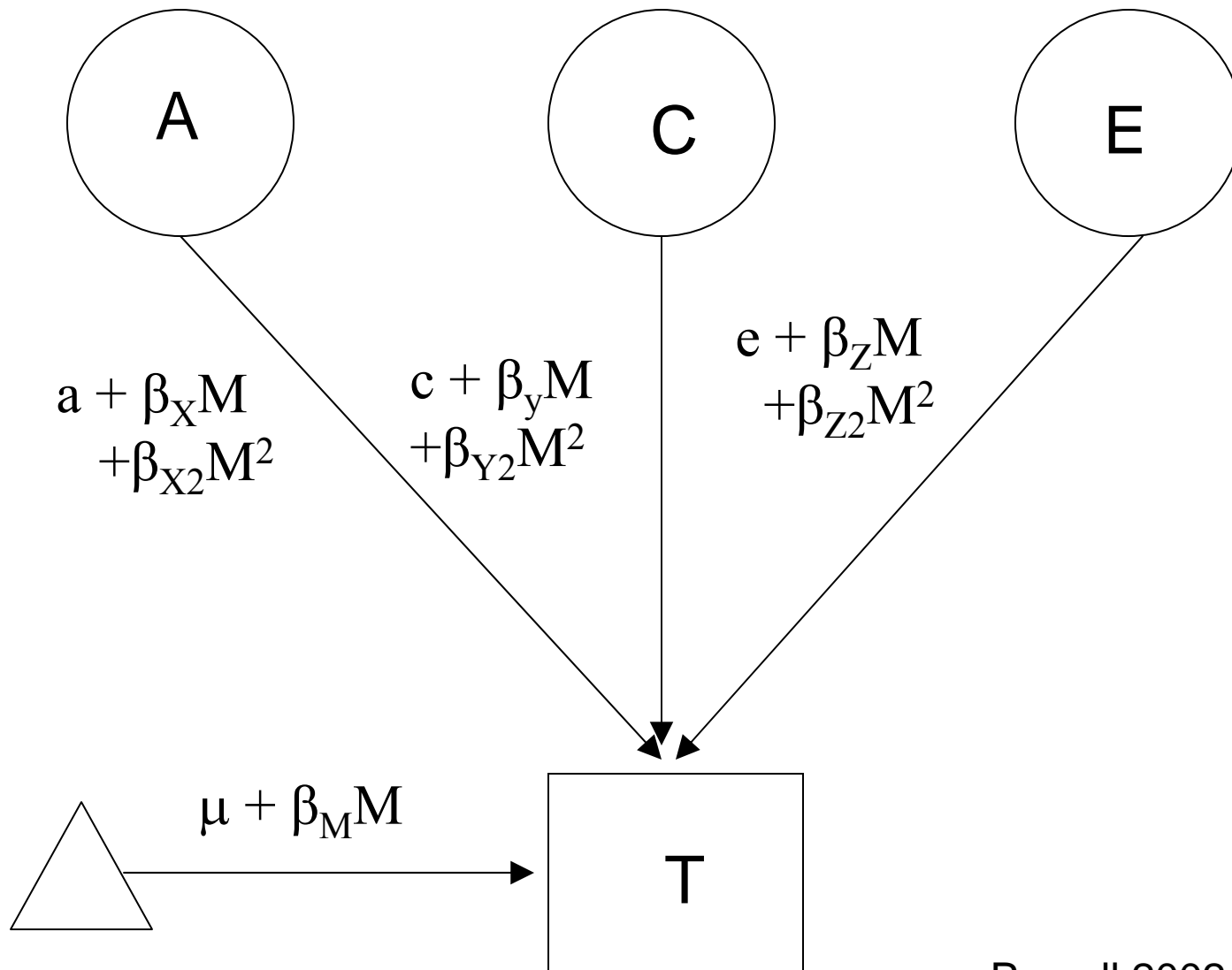
Comparing Results

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

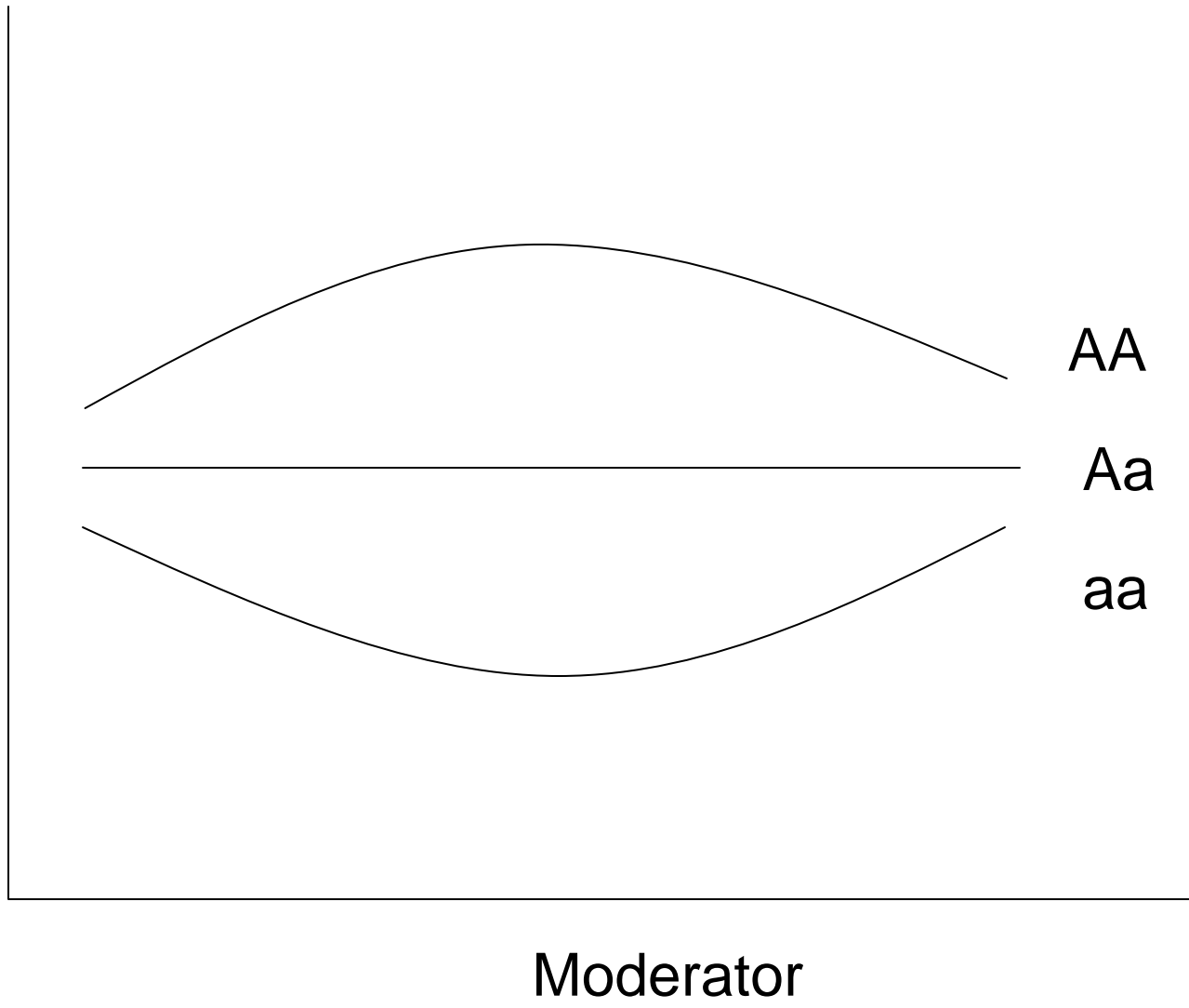
Why is the A moderation NS using the continuous moderator?

- 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



Nonlinear Moderation



This moderation model can be used
to test for gene-environment
interaction

Gene-Environment Interaction

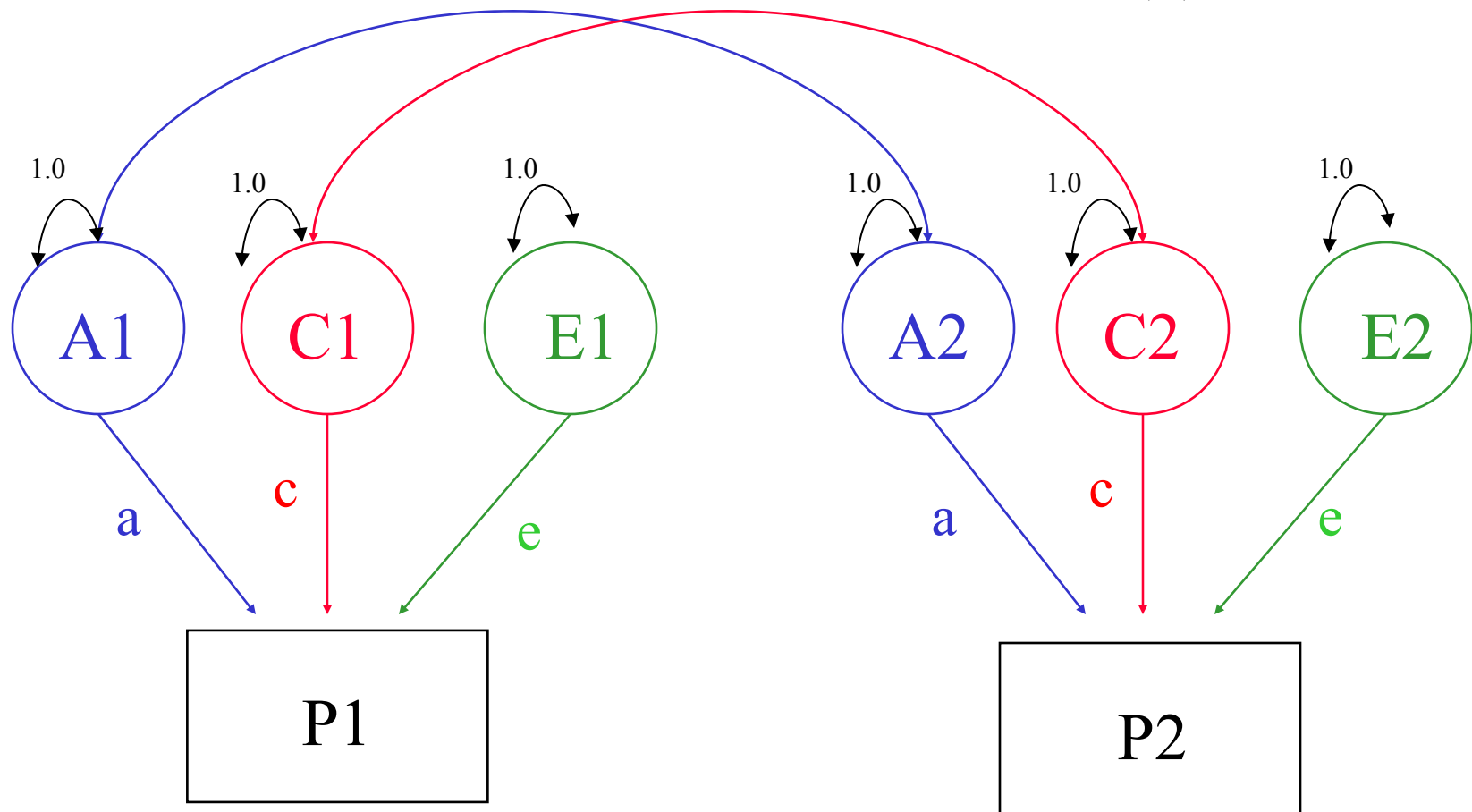
- *Genetic control of sensitivity to the environment*
- *Environmental control of gene expression*
- Bottom line: nature of genetic effects differs among environments

Standard Univariate Model

1.0 (MZ) / .5 (DZ)

1.0

$$P = A + C + E$$
$$\text{Var}(P) = a^2 + c^2 + e^2$$



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	$1 \times 1 = 1$
DZ Pairs/Full Sibs	1	$\frac{1}{2}$	$1 \times \frac{1}{2} = \frac{1}{2}$

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	$1 \times 1 = 1$
DZ Pairs/Full Sibs	1	$\frac{1}{2}$	$1 \times \frac{1}{2} = \frac{1}{2}$

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

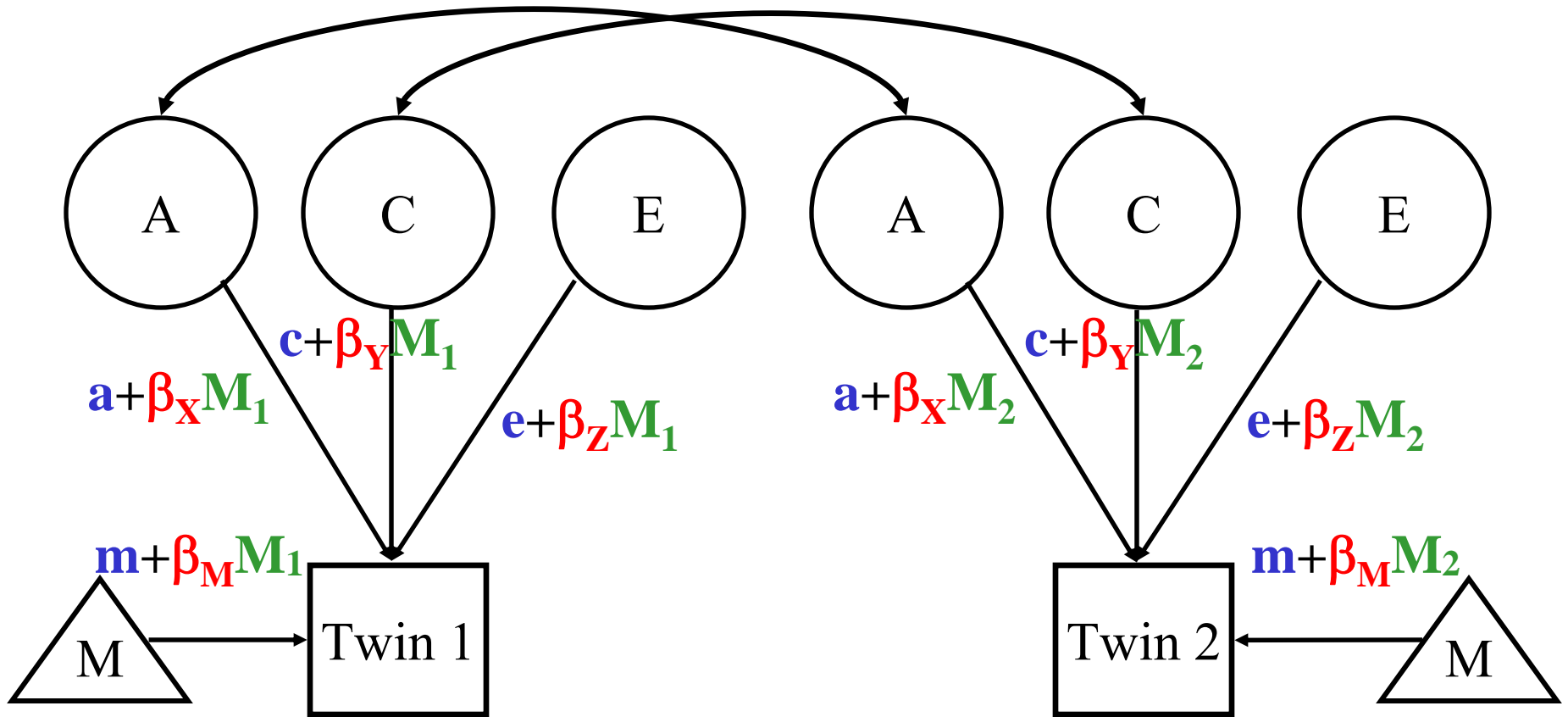
	Unshared (Unique) Environment	Additive Genetic Effects	Genotype x Unshared Environment Interaction
MZ Pairs	0	1	$0 \times 1 = 0$
DZ Pairs/Full Sibs	0	$\frac{1}{2}$	$0 \times \frac{1}{2} = 0$

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	$0 \times 1 = 0$
DZ Pairs/Full Sibs	0	$\frac{1}{2}$	$0 \times \frac{1}{2} = 0$

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

Final Things to Consider

- Don't forget about theory!

Final Things to Consider

- Don't forget about theory!
 - “Moderation in all things....including moderation”
-Mike Neale

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

Heterogeneity Questions II

- Are these differences due to differences in the magnitude of the effects (quantitative)?
 - e.g. Is the contribution of genetic/environmental factors greater/smaller in males than in females?
- Are the differences due to differences in the nature of the effects (qualitative)?
 - e.g. Are there different genetic/environmental factors influencing the trait in males and females?
 - Need OS pairs for this!