Estimating the Importance of Maternal Genetic Effects on Offspring Phenotypes with "M-GCTA"

David Evans

Institute for Molecular Bioscience

University of Queensland

(With thanks to Beate St Pourcain, George Davey Smith, Tim York)

In Memoriam

Behav Genet (2014) 44:445–455
DOI 10.1007/s10519-014-9666-6

ORIGINAL RESEARCH

Resolving the Effects of Maternal and Offspring Genotype on Dyadic Outcomes in Genome Wide Complex Trait Analysis ("M-GCTA")

Lindon J. Eaves · Beate St. Pourcain · George Davey Smith · Timothy P. York · David M. Evans

Received: 15 October 2013 / Accepted: 10 June 2014 / Published online: 25 July 2014

Abstract Genome wide complex trait analysis (GCTA) is extended to include environmental effects of the maternal genotype on offspring phenotype ("maternal effects", M-GCTA). The model includes parameters for the direct effects of the offspring genotype, maternal effects and the covariance between direct and maternal effects. Analysis of simulated data, conducted in OpenMx, confirmed that model parameters could be recovered by full information maximum likelihood (FIML) and evaluated the biases that arise in conventional GCTA when indirect genetic effects are ignored. Estimates derived from FIML in OpenMx showed very close agreement to those obtained by restricted maximum likelihood using the published algorithm for GCTA. The method was also applied to illustrative perinatal phenotypes from ~4,000 mother-offspring pairs from the Avon Longitudinal Study of Parents and Children. The relative merits of extended GCTA in contrast to quantitative genetic approaches based on analyzing the phenotypic covariance structure of kinships are considered.

Keywords Maternal effects · Genome wide complex trait analysis · GCTA · Twins · Heritability · Bias · Genetic relatedness · Covariance · Environment · SNPs

Edited by Sarah Medland.

L. J. Eaves (✉) · T. P. York
Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA, USA
e-mail: eaves.lindon@gmail.com

B. St. Pourcain · G. D. Smith · D. M. Evans
MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

B. St. Pourcain
School of Oral and Dental Sciences, University of Bristol, Bristol, UK

B. St. Pourcain
School of Experimental Psychology, University of Bristol, Bristol, UK

G. D. Smith · D. M. Evans
School of Social and Community Medicine, University of Bristol, Bristol, UK

D. M. Evans
Translational Research Institute, University of Queensland Diamantina Institute, Brisbane, QLD, Australia

Background

The recent history of human quantitative genetics has witnessed a remarkable convergence between views of complex trait genetics emerging from approaches that rely on comparing phenotypic resemblance between relatives sharing different degrees of genetic relatedness (see e.g. Fisher 1918; Mather and Jinks 1982; Falconer and McKay 1996) and those made possible by direct characterization of genetic variation at the genomic level, notably genome wide association analysis (GWAS) and genome wide complex trait analysis (GCTA, Yang et al. 2010, 2011a, 2011b). Although there are qualifications and nuances that reflect the relative strength and weaknesses of these two paradigms, their common heuristic recognizes that the heritable contribution to individual differences in quantitative traits reflects the cumulative action of variation at large numbers of genetic variants of small individual effect, widely dispersed across the genome. Such astonishing convergence of quite different approaches may

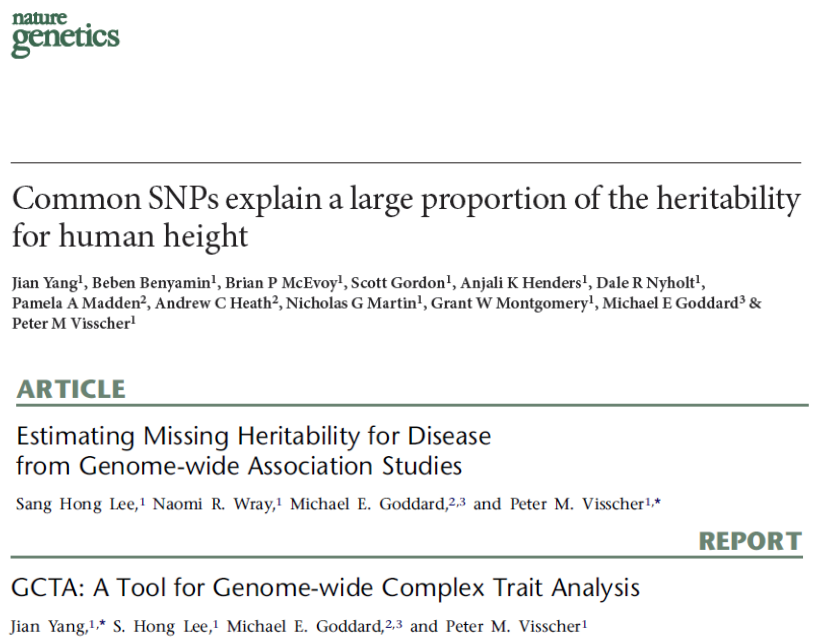
Lindon Eaves

This Session

- GCTA / G-REML
  - What is GCTA / G-REML?
  - Describing Genetic Relationship Matrices (GRMs)
  - Describing the model
- M-GCTA
  - Describing the model
  - Deriving the expected variance
  - Deriving the expected covariance

What is “GCTA” / G-REML?

- Provides an estimate of the amount of phenotypic variance explained by genetic markers on a SNP microarray (“SNP heritability”)
- Uses unrelated individuals
- Originally devised to investigate the missing heritability conundrum in human genetics



nature genetics

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

ARTICLE

Estimating Missing Heritability for Disease from Genome-wide Association Studies

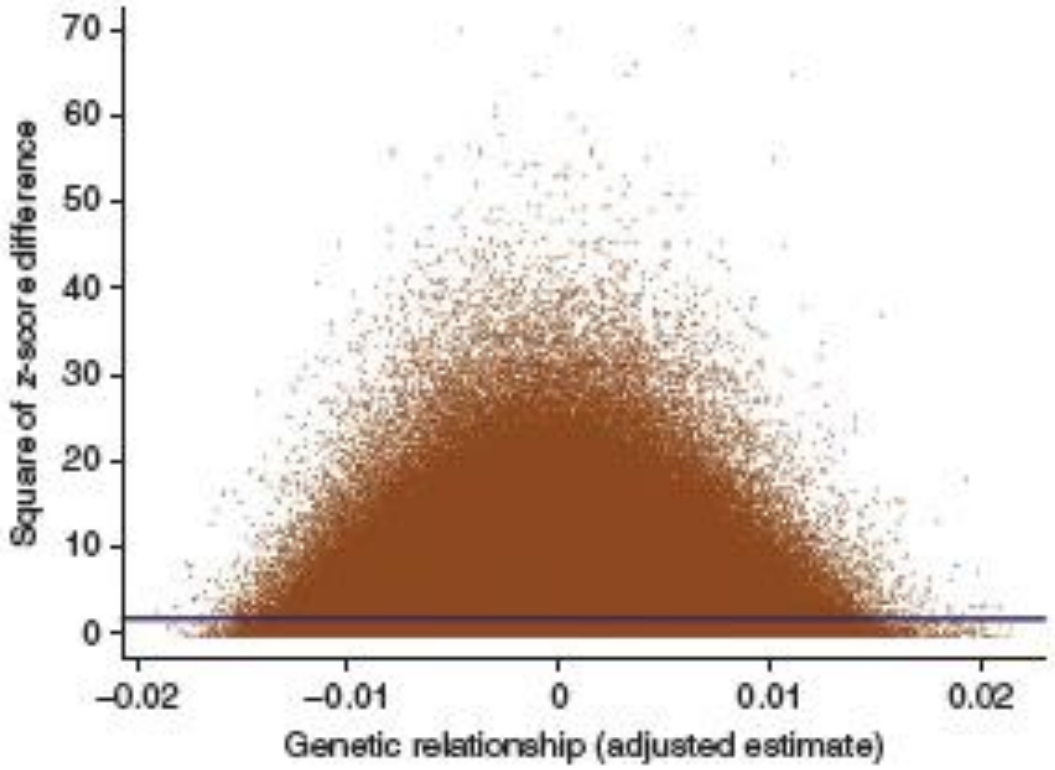
Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

REPORT

GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

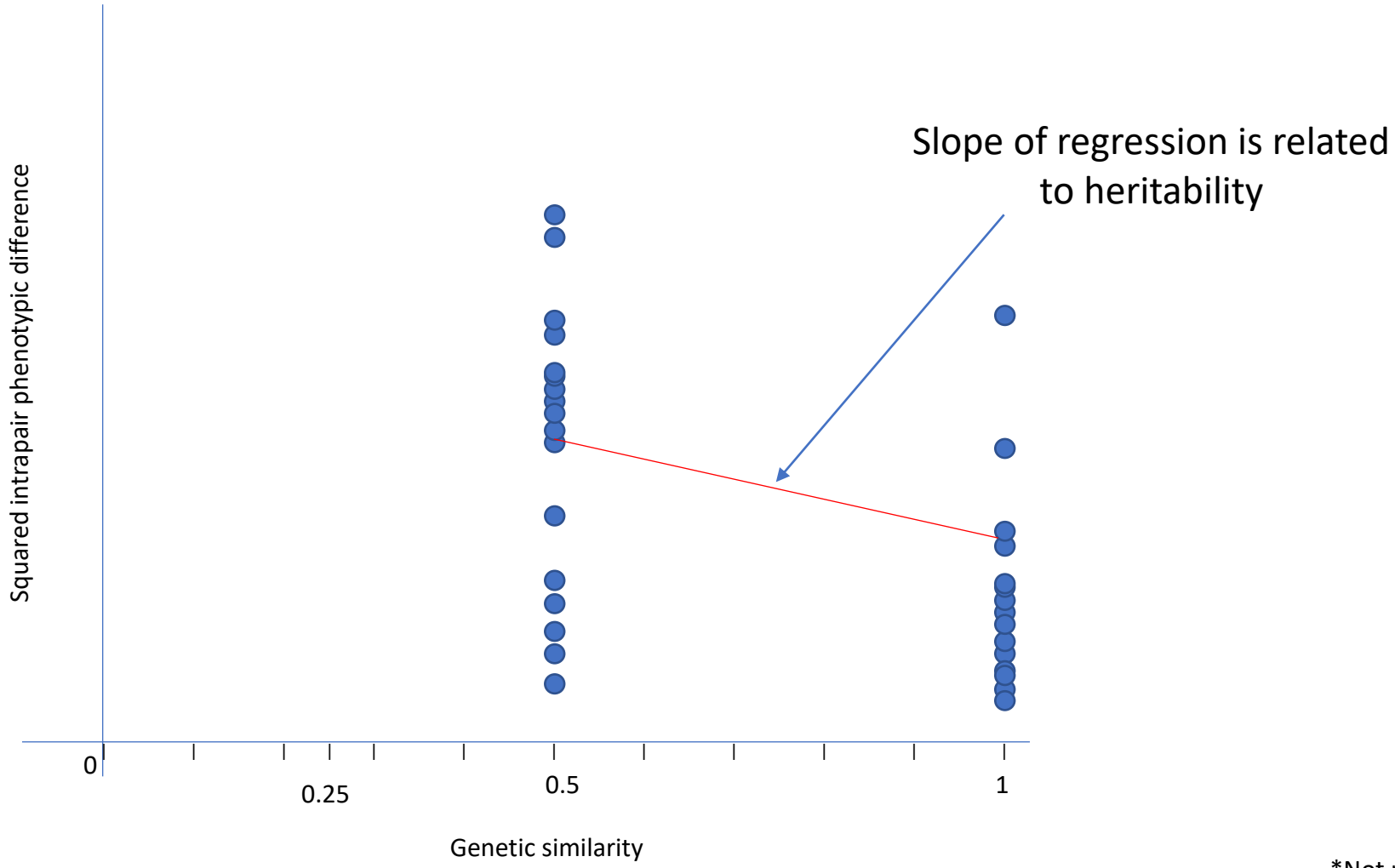
GCTA Intuition



- Unrelated individuals
- If a trait is genetically influenced, then individuals who are more genetically similar should be more phenotypically similar
- Can be thought of like a regression*
- Slope of regression line is proportional to heritability

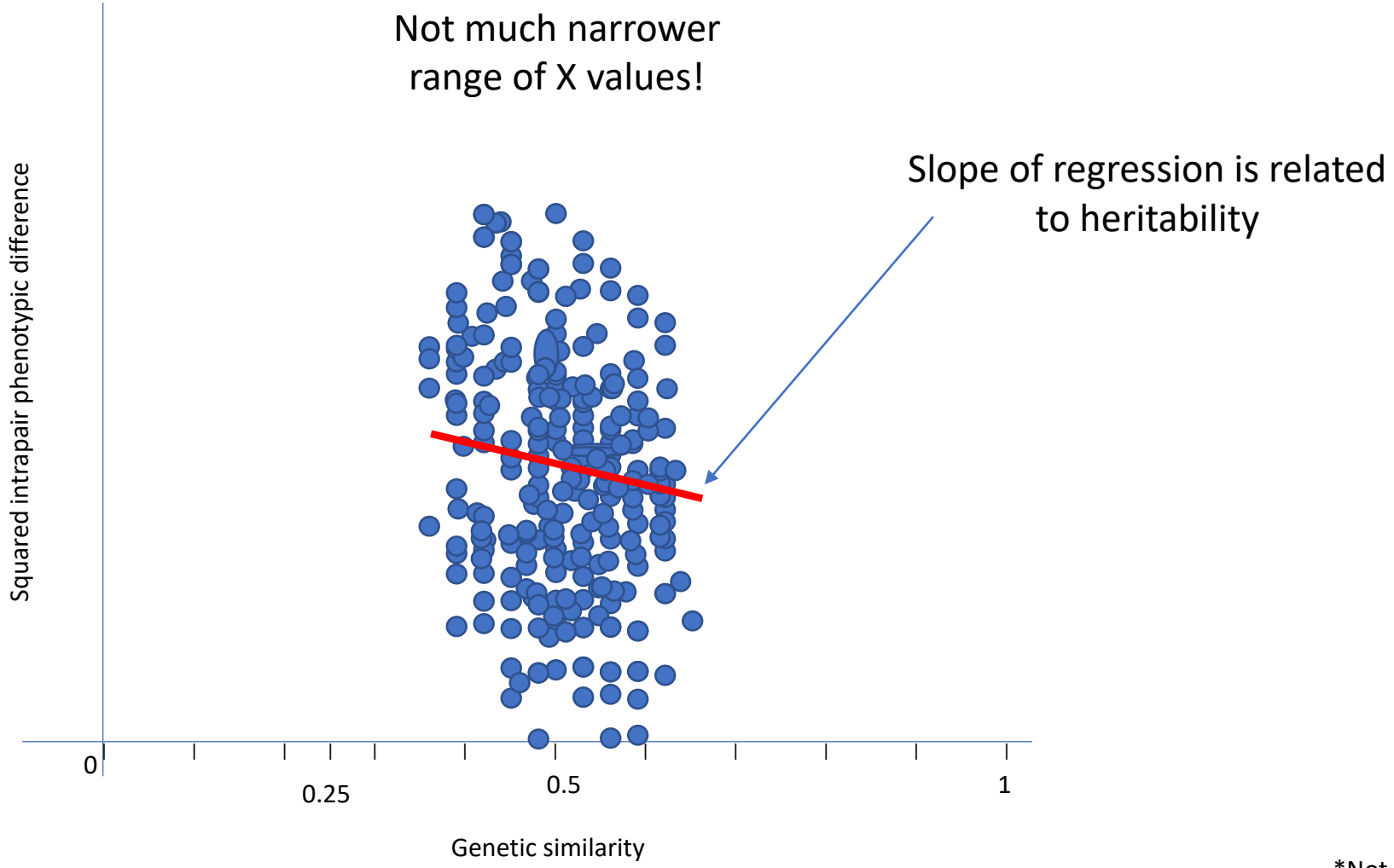
*This is not how the model is fit though!

Classical Twin Design



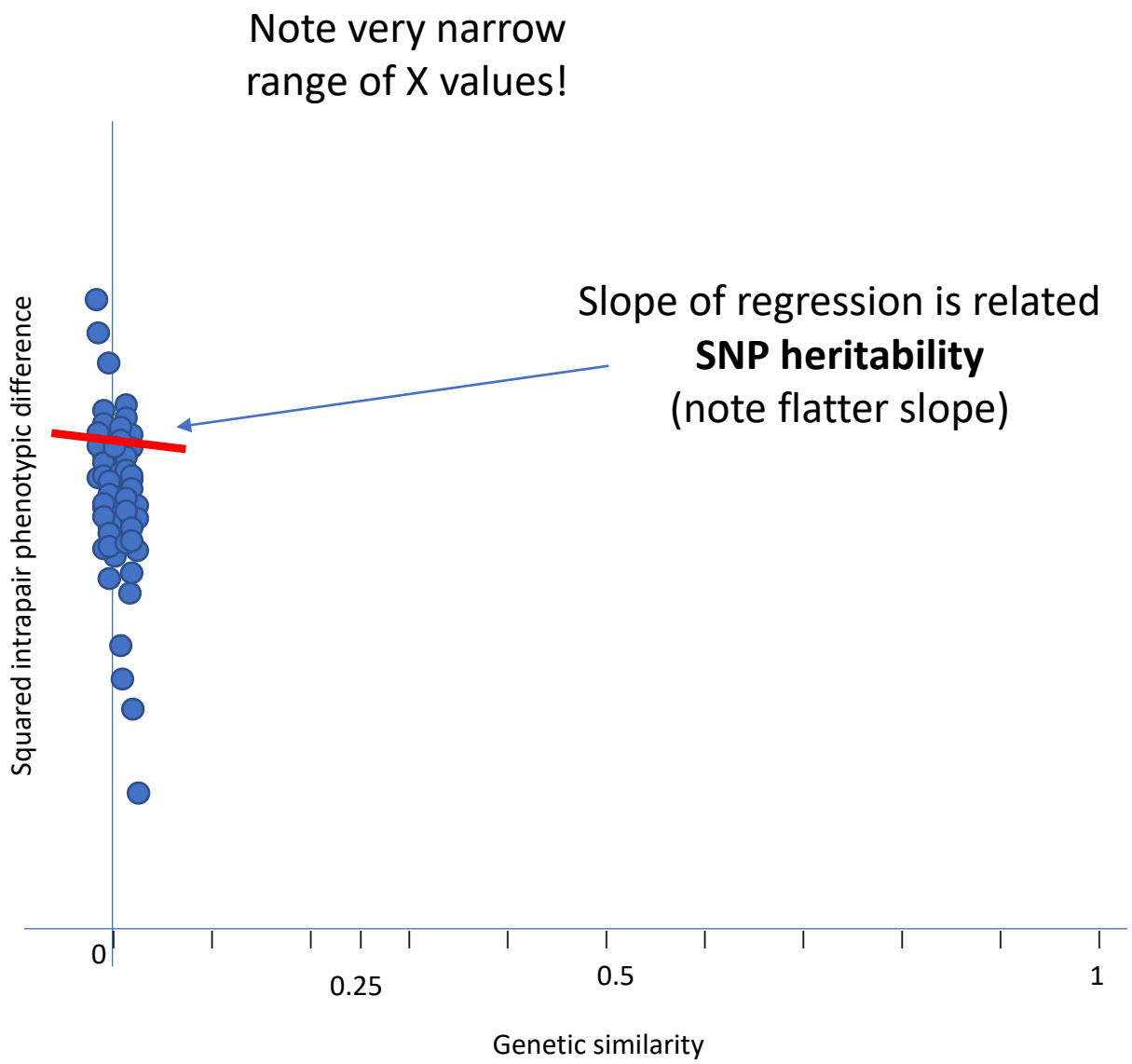
*Not real data

Sibling Pairs and IBD sharing



*Not real data

Unrelated Pairs and IBS sharing



*Not real data

GCTA Process

- Two step process
- Estimate Genetic relatedness matrix (GRM)
  - Exclude one from each pair of individuals who are >2.5% IBS (genetically related individuals more likely to share common environments; individuals who are more genetically similar dominate analysis)
- Estimate variance components (via “REML” or via “FIML”)

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

Think of A_{jk} as like an average genetic correlation across the genome for individuals j and k

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

Recall the formula for a Pearson correlation coefficient:

$$r = \frac{1}{N}\frac{\sum(X-\bar{X})(Y-\bar{Y})}{SD_X SD_Y}$$

The formula for the elements of the GRM is analogous

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

Each dosage x_{ij} (and x_{ik}) is distributed as a binomial variable (like tossing a coin and counting the number of heads)

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

Each dosage x_{ij} (and x_{ik}) is distributed as a binomial variable (like tossing a coin and counting the number of heads)

So each $x_{ij} \sim \text{Bin}(2, p_i)$ with mean $= 2p_i$.

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

Each dosage x_{ij} (and x_{ik}) is distributed as a binomial variable (like tossing a coin and counting the number of heads)

So each $x_{ij} \sim \text{Bin}(2, p_i)$ with mean = $2p_i$ and variance = $2p_i(1\text{-}p_i)$

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

Each dosage x_{ij} (and x_{ik}) is distributed as a binomial variable (like tossing a coin and counting the number of heads)

So each $x_{ij} \sim \text{Bin}(2, p_i)$ with mean = $2p_i$ and variance = $2p_i(1\text{-}p_i)$

(1) GCTA- Genetic Relationship Matrix

$$A_{jk} = \frac{1}{M}\sum_{i=1}^{M}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}.$$

$$r = \frac{1}{N}\frac{\sum(X-\bar{X})(Y-\bar{Y})}{SD_X SD_Y}$$

A_{jk} is the genomic relatedness between individuals j and k

x_{ij} is the dosage $\{x_{ij} = 0,1,2\}$ of the reference allele for SNP i for individual j

x_{ik} is the dosage $\{x_{ik} = 0,1,2\}$ of the reference allele for SNP i for individual k

p_i is the frequency of the reference allele for SNP i

M is the number of markers across the genome

Each dosage x_{ij} (and x_{ik}) is distributed as a binomial variable (like tossing a coin and counting the number of heads)

So each $x_{ij} \sim \text{Bin}(2, p_i)$ with mean = $2p_i$ and variance = $2p_i(1\text{-}p_i)$

(2) Estimate Variance Components

$$\boldsymbol{\Sigma} = \mathbf{A}\sigma^2_a + \mathbf{I}\sigma^2_e$$

$$\begin{bmatrix} \sigma^2_{11} & \sigma_{12} & \dots & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma^2_{22} & \dots & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \dots & \sigma^2_{nn} \end{bmatrix}_{(N \times N)} = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & \dots & a_{nn} \end{bmatrix}_{(N \times N)} \cdot \sigma^2_a + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{(N \times N)} \cdot \sigma^2_e$$

$\sum$ is the expected phenotypic covariance matrix

$\boldsymbol{\sigma^2_a}$ is the additive genetic variance

$\boldsymbol{\sigma^2_e}$ is the unique environmental variance

A is a **GRM** containing pairwise genome-wide genetic "correlation" between individuals

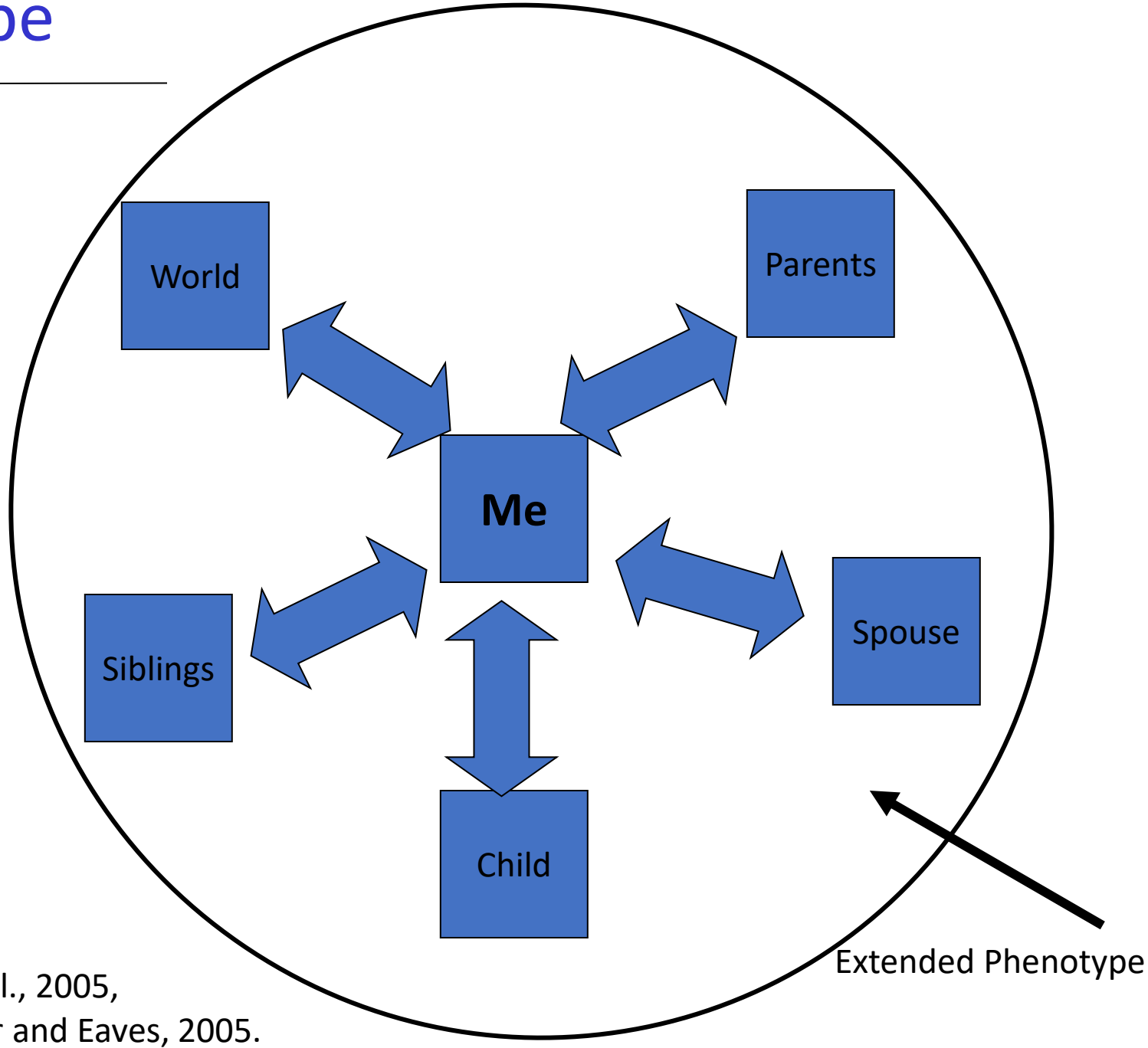
I is an identity matrix

Interpretation of Parameter Estimates

- $\sigma_A^2/(\sigma_A^2 + \sigma_E^2)$ is the estimated "SNP heritability" of the trait
- The proportion of phenotypic variance explained by SNPs on the microarray
- Since most markers on arrays are common, this will primarily reflect common genetic variation, but will also include rare variation that happens to be tagged by the array.
- SNP heritability is different from "heritability" which is typically estimated from twin studies

M-GCTA

Extending the Phenotype

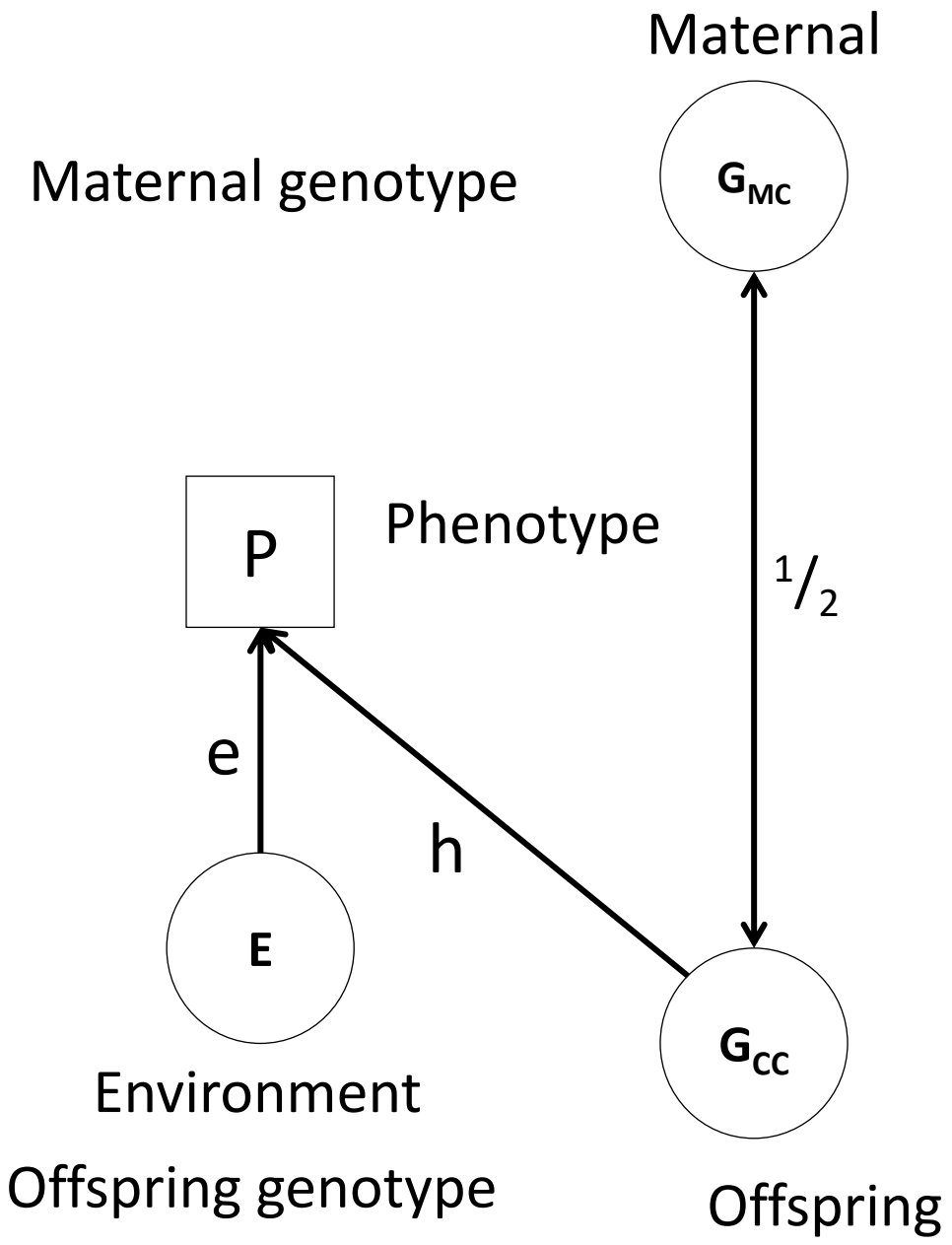


Eaves et al., 2005,
In Kendler and Eaves, 2005.

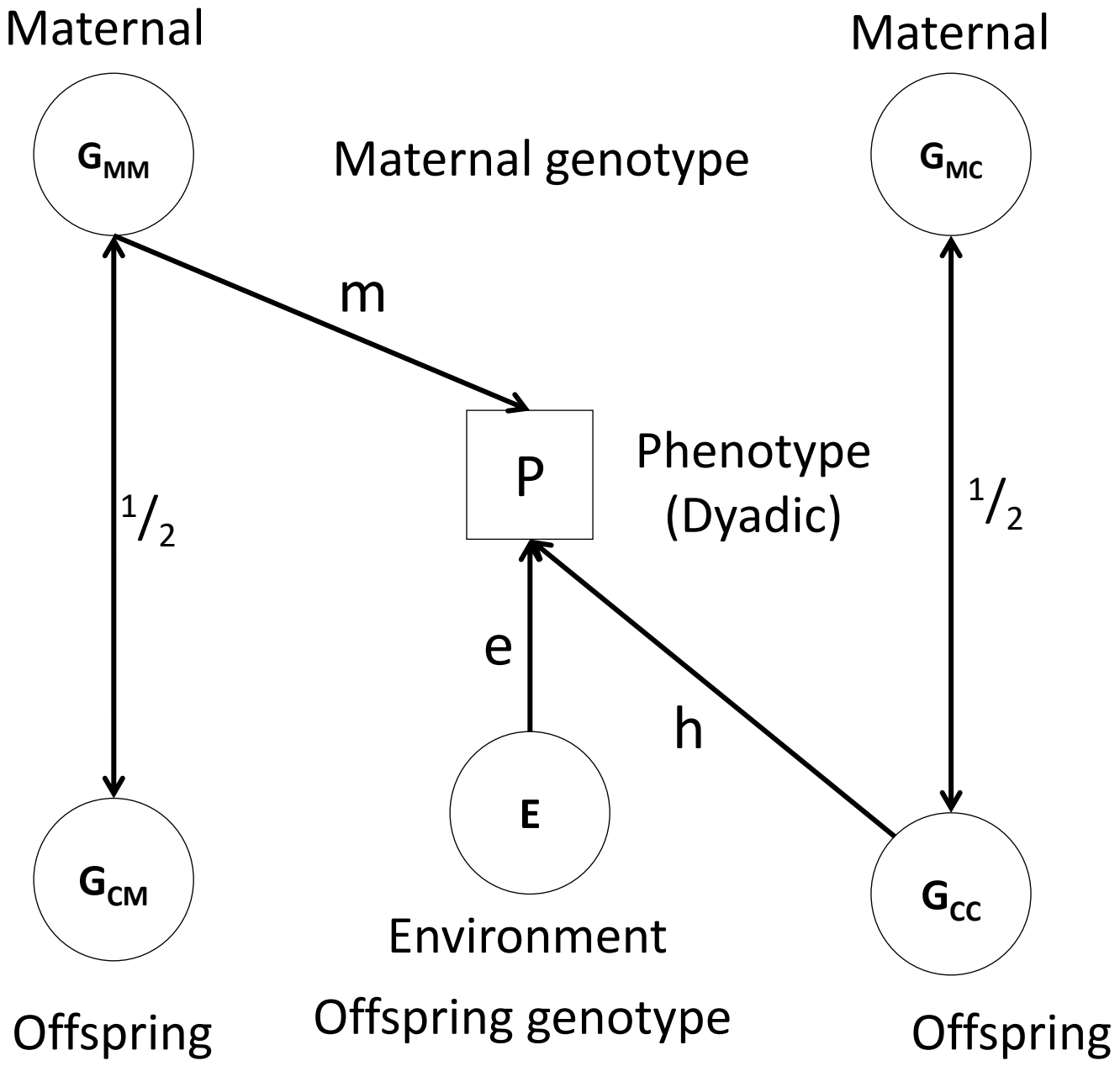
M-GCTA

- An extension of GCTA to estimate the proportion of offspring phenotypic variance explained by the maternal (and child's) genomes
- Requires genome-wide genotyped mother-offspring pairs ("dyads") where the offspring has been phenotyped
- Dyads should be "unrelated" to one another

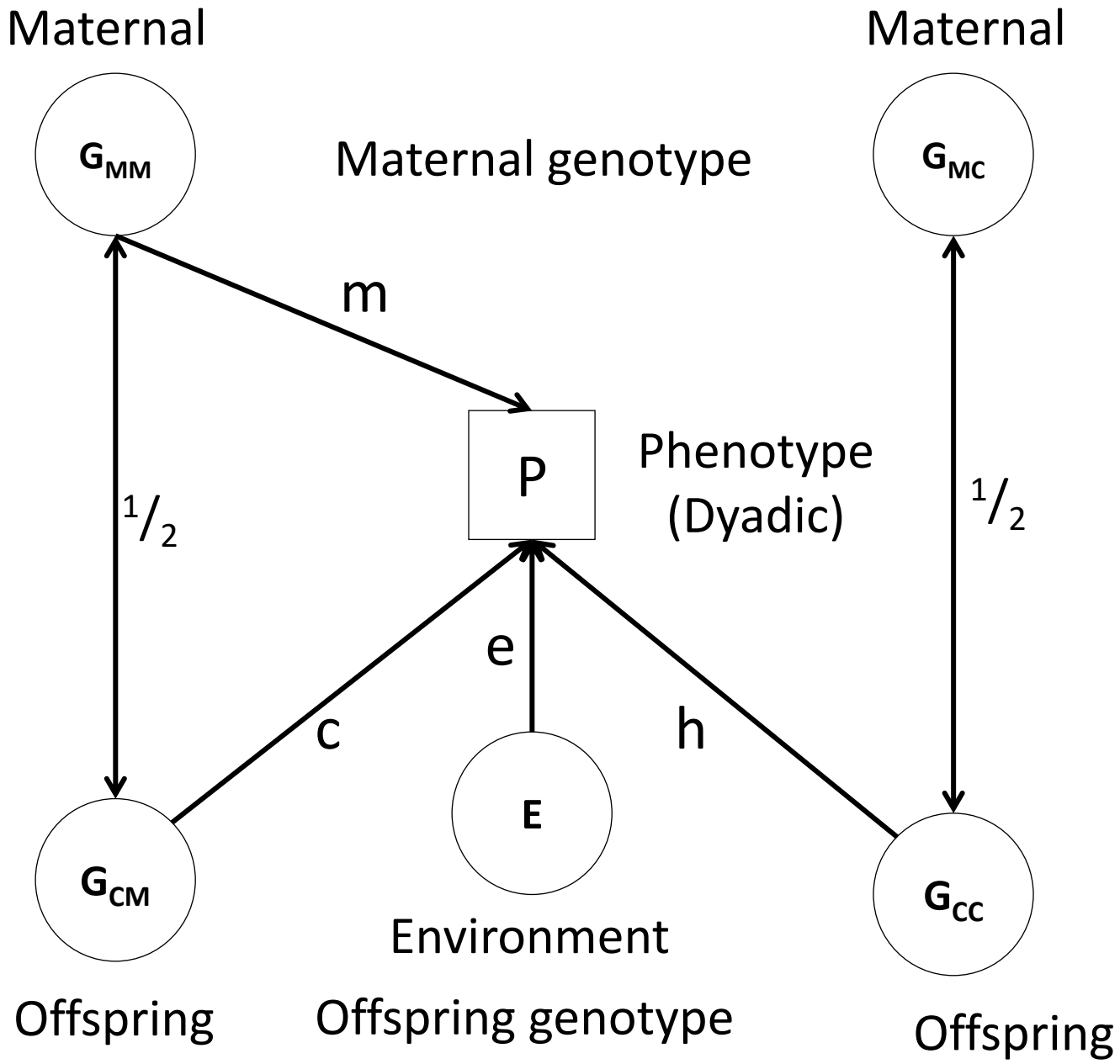
Path model for offspring and maternal effects in mother-child dyads.



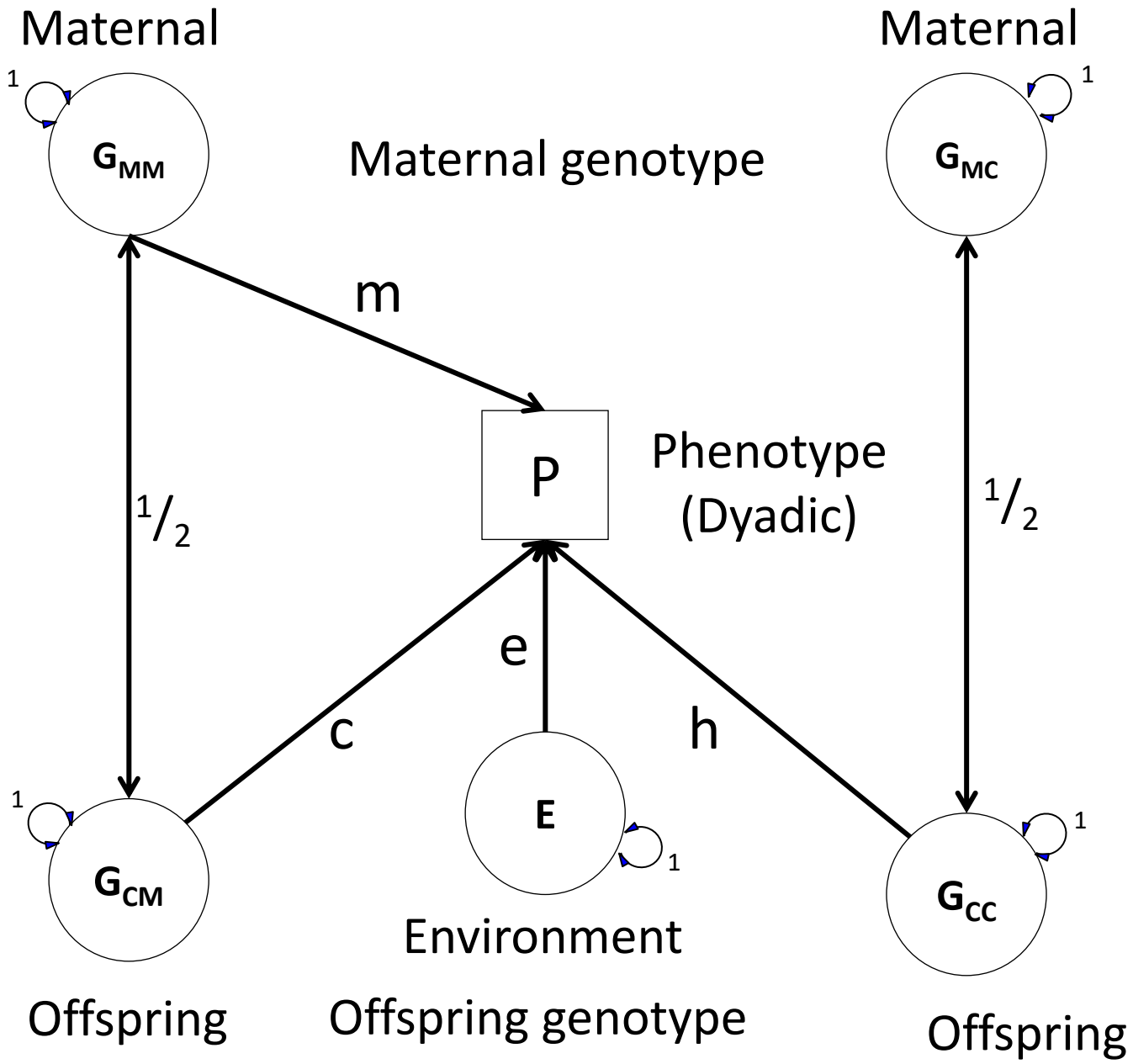
Path model for offspring and maternal effects in mother-child dyads.



Path model for offspring and maternal effects in mother-child dyads.

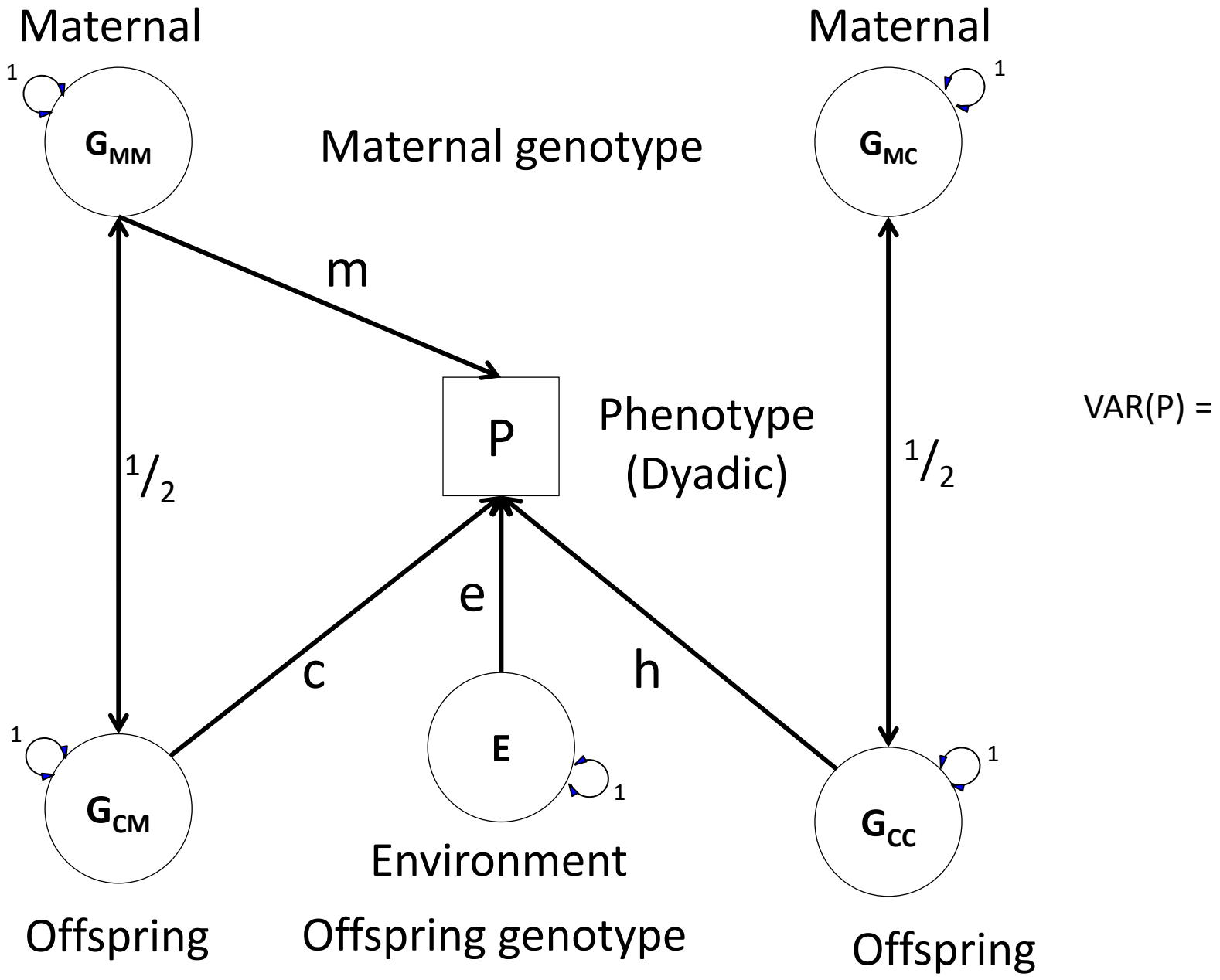


Path model for offspring and maternal effects in mother-child dyads.

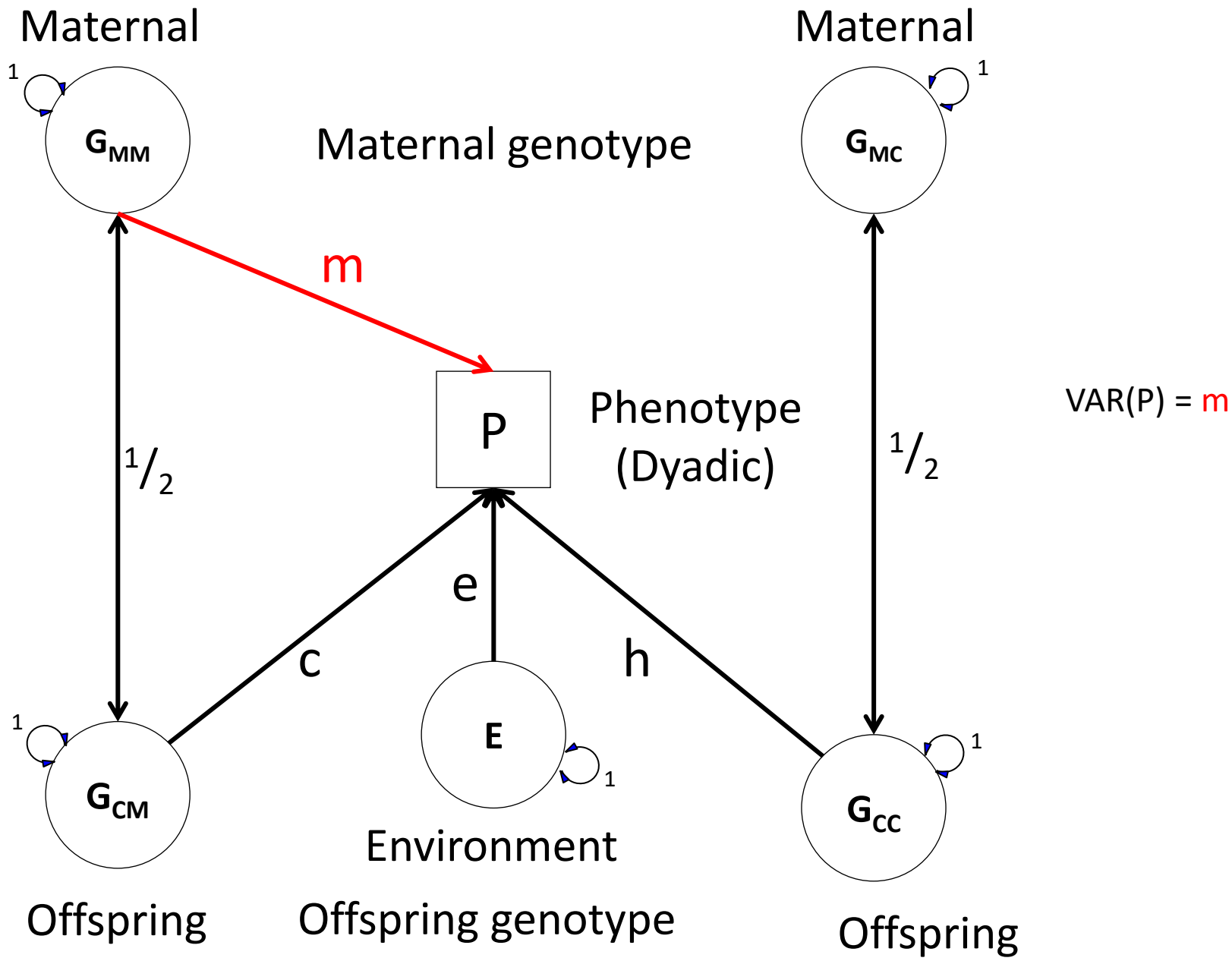


Deriving the Phenotypic Variance

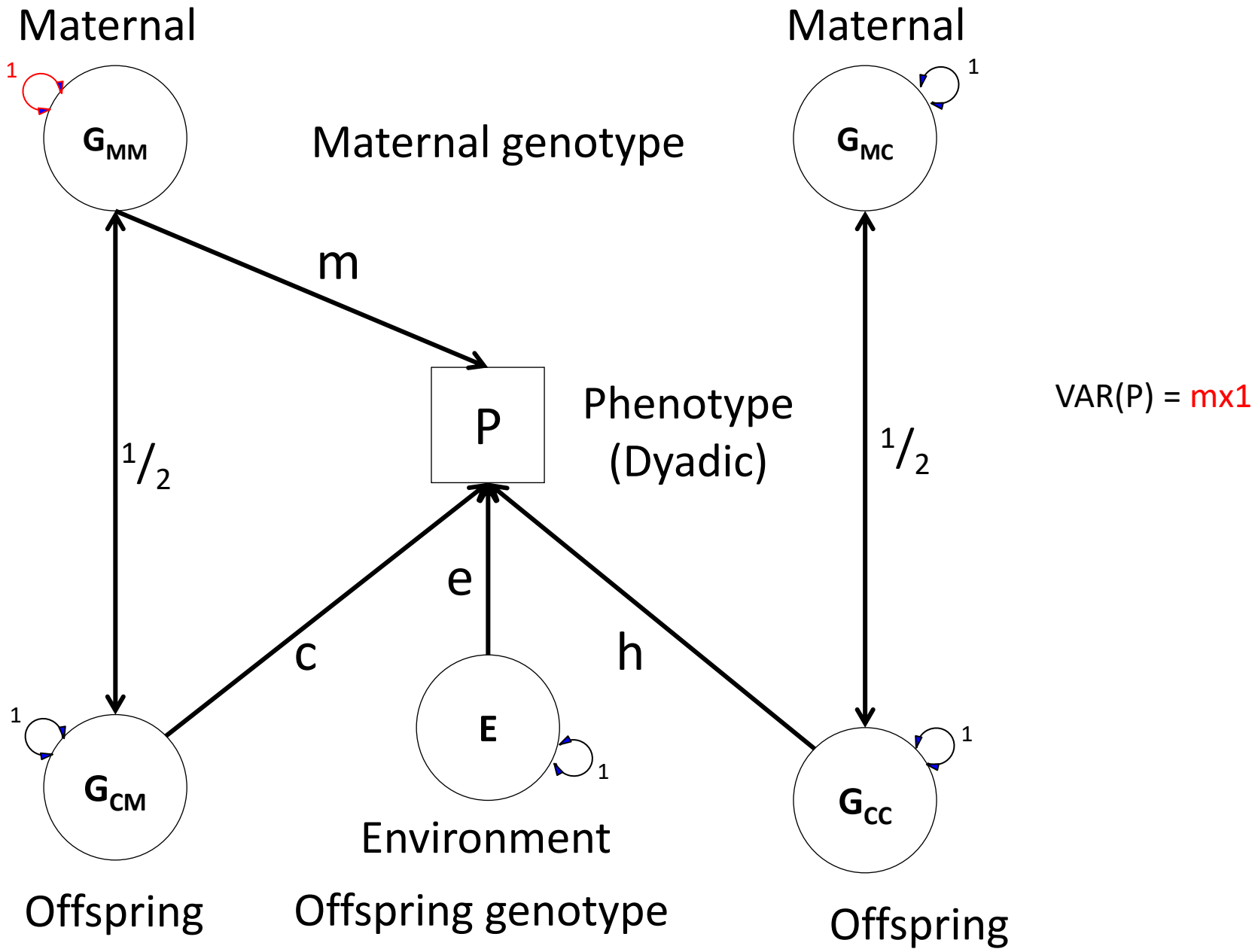
Path model for offspring and maternal effects in mother-child dyads.



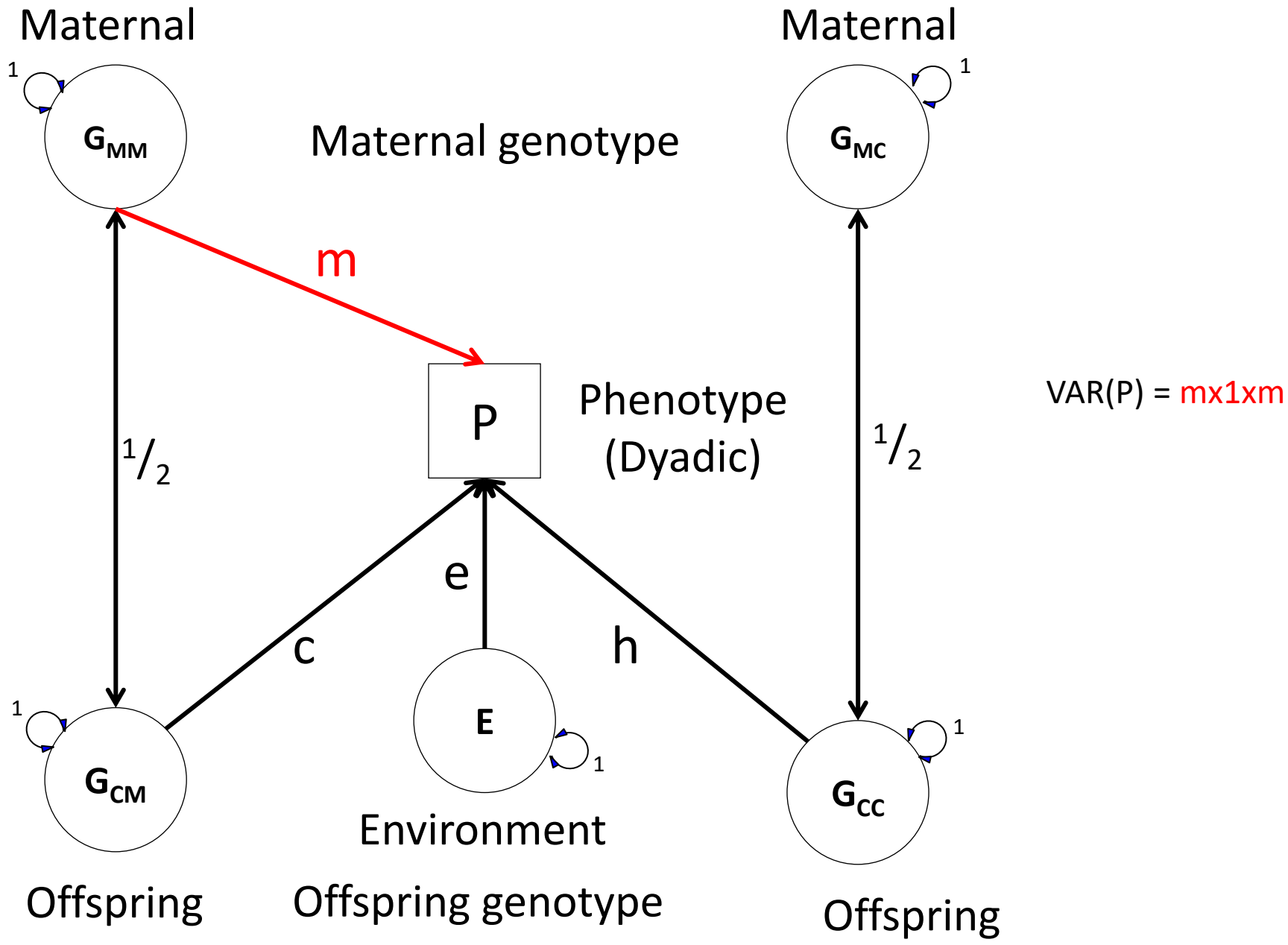
Path model for offspring and maternal effects in mother-child dyads.



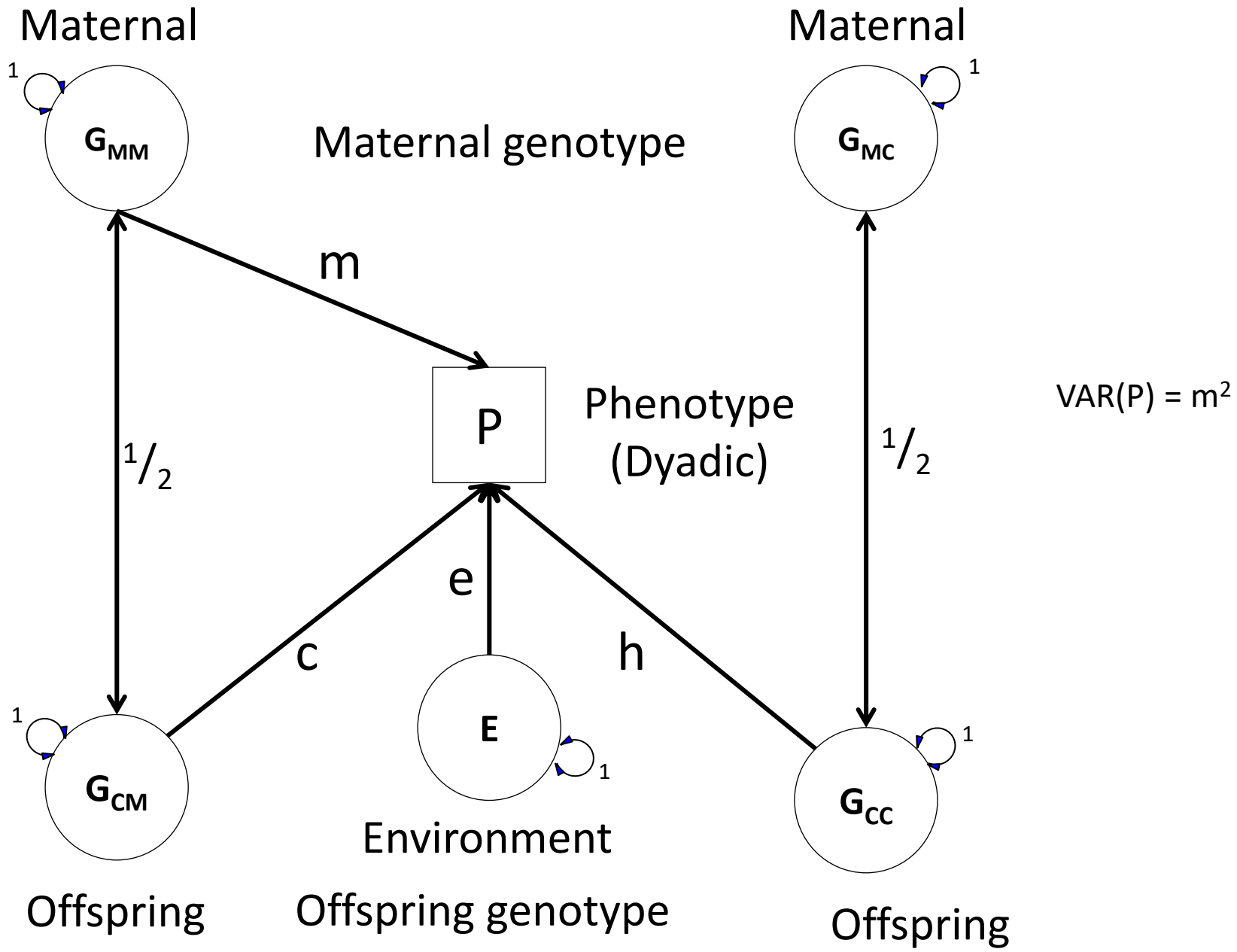
Path model for offspring and maternal effects in mother-child dyads.



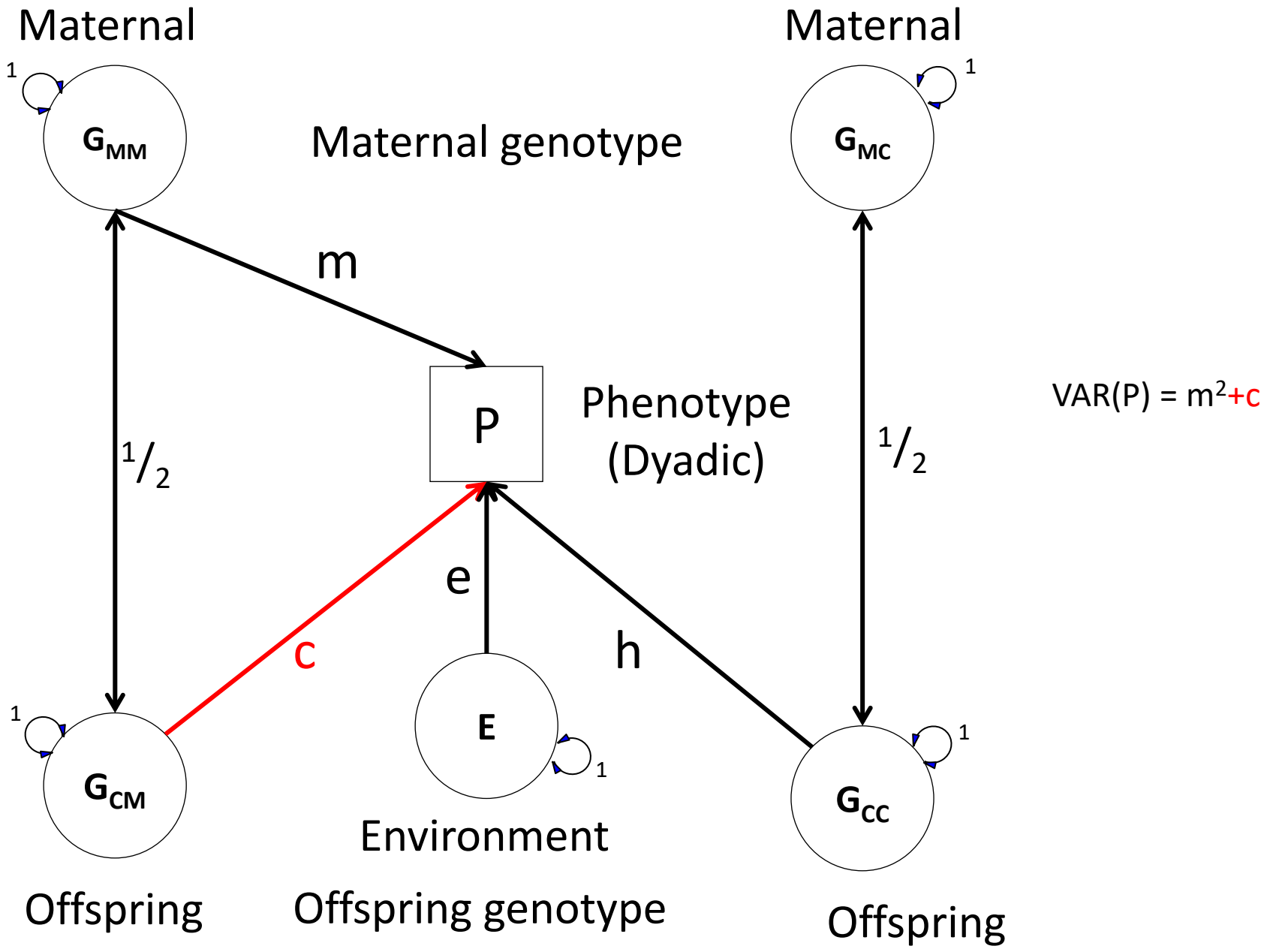
Path model for offspring and maternal effects in mother-child dyads.



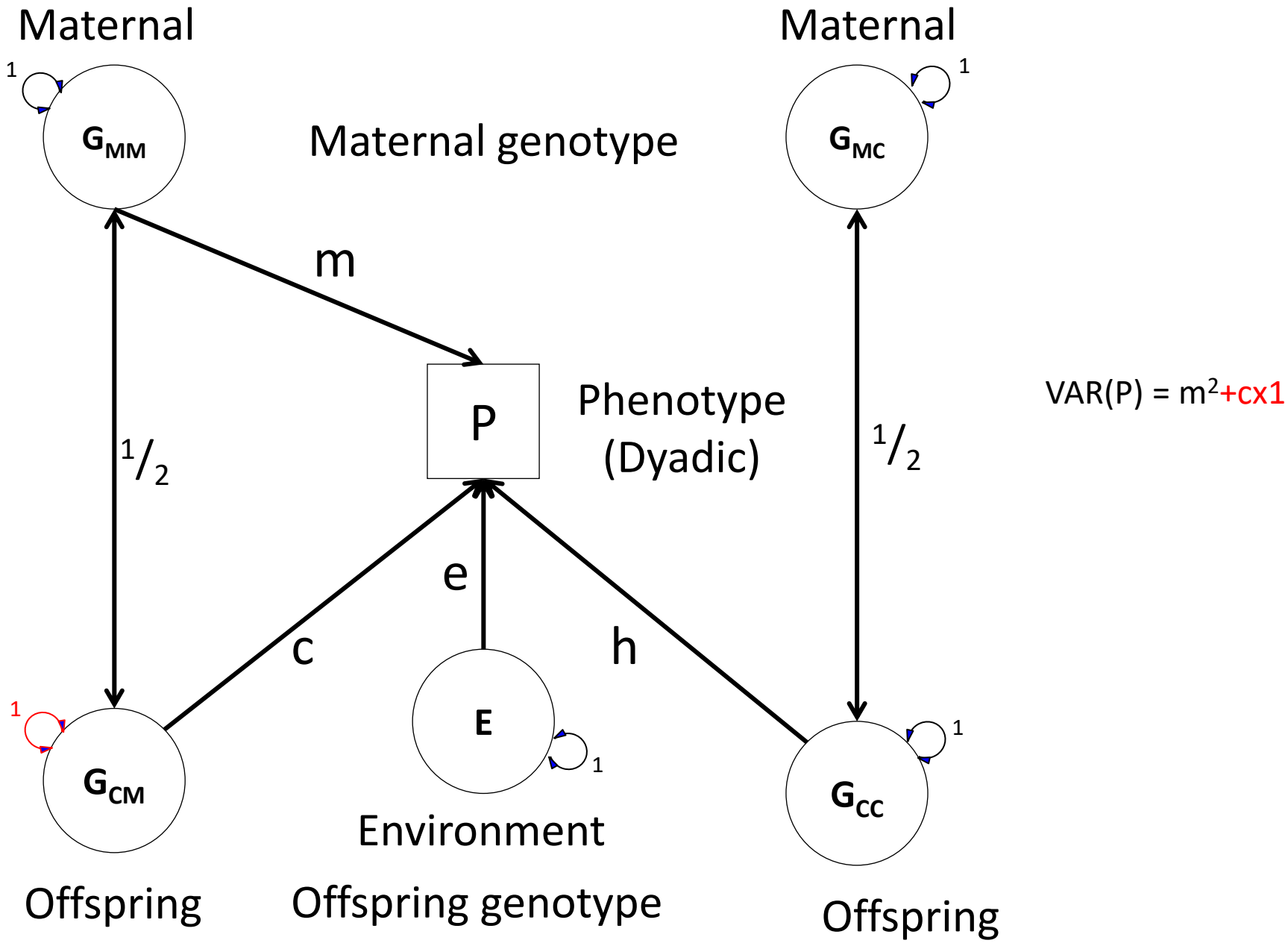
Path model for offspring and maternal effects in mother-child dyads.



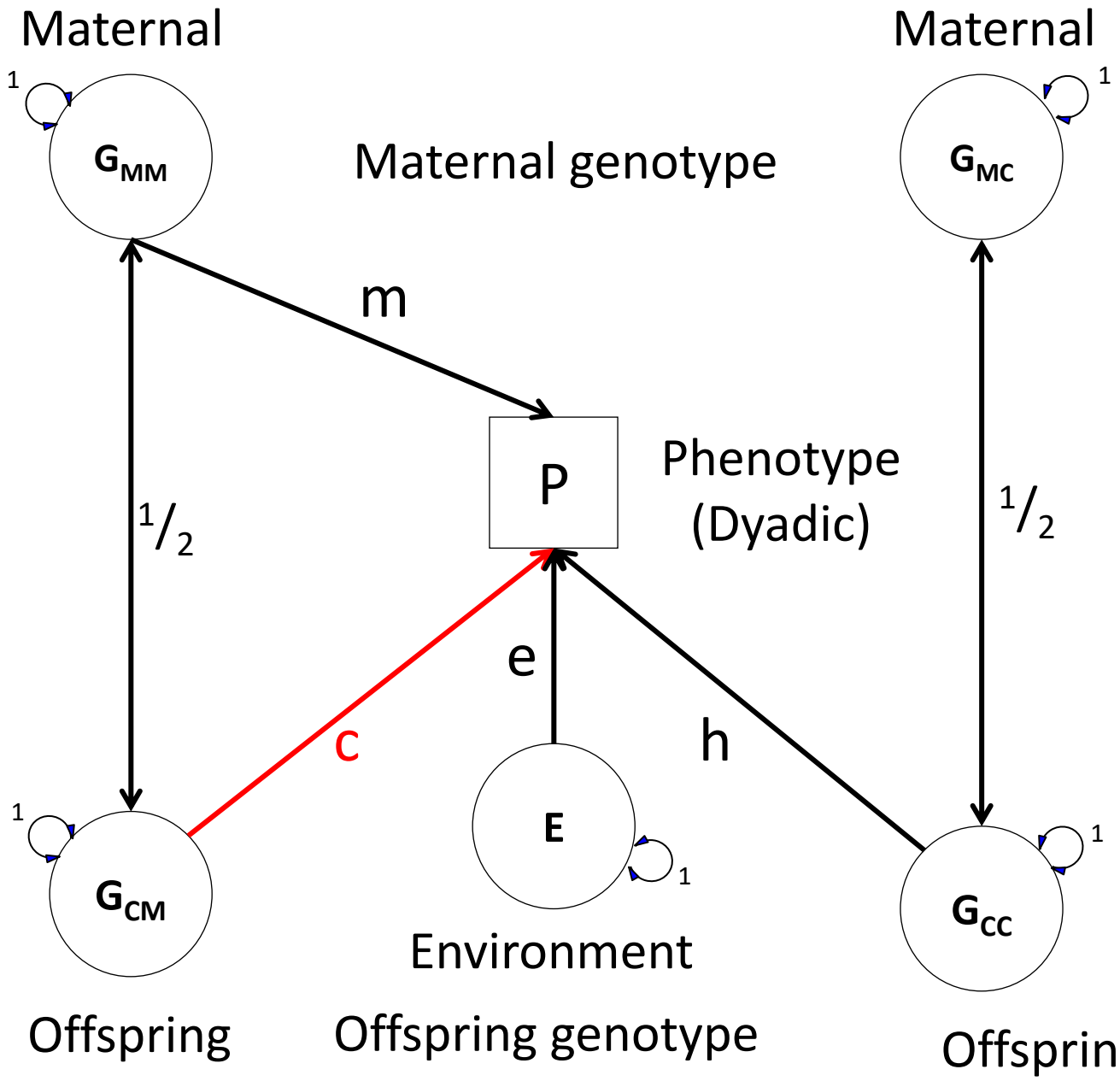
Path model for offspring and maternal effects in mother-child dyads.



Path model for offspring and maternal effects in mother-child dyads.

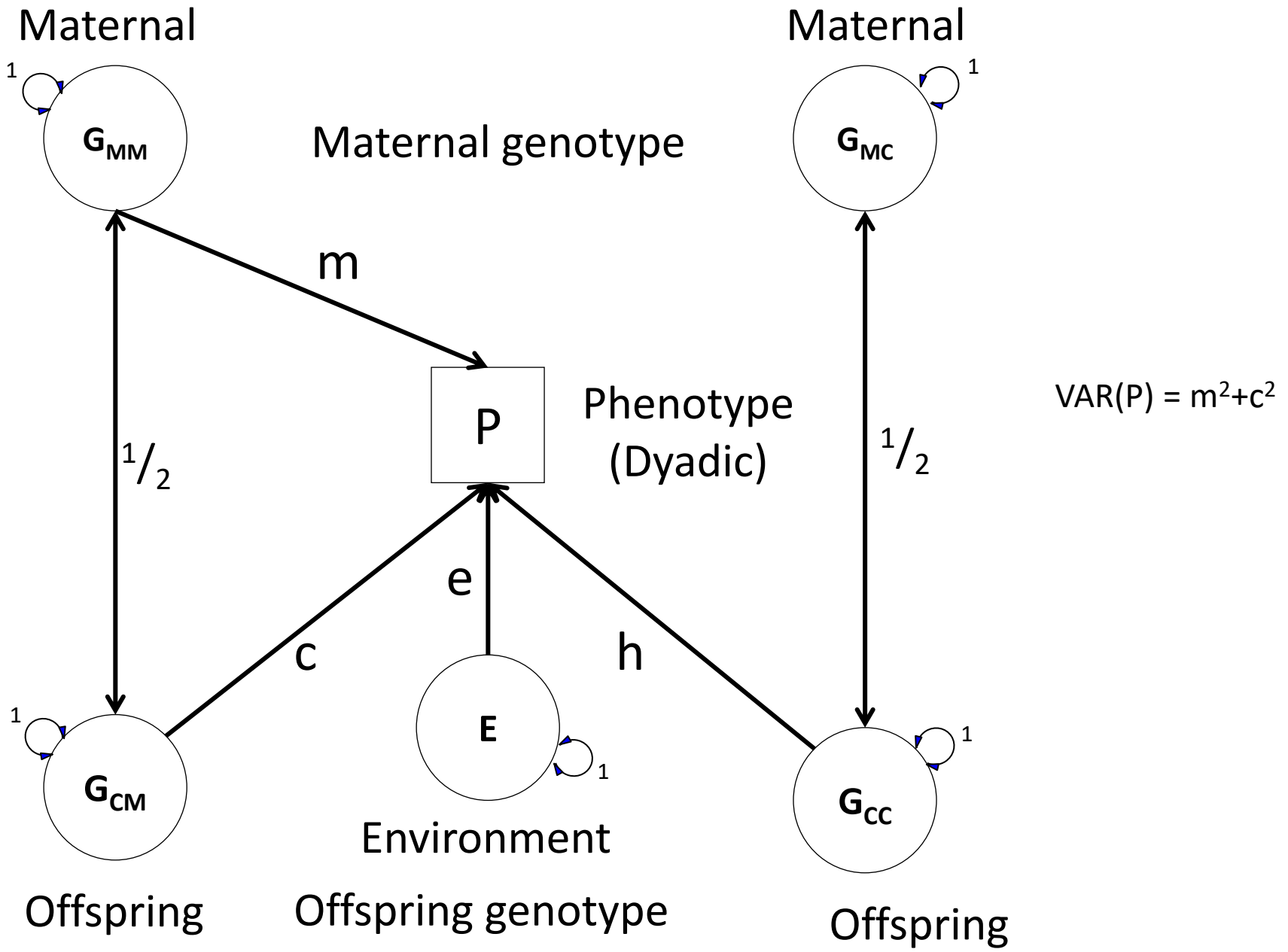


Path model for offspring and maternal effects in mother-child dyads.

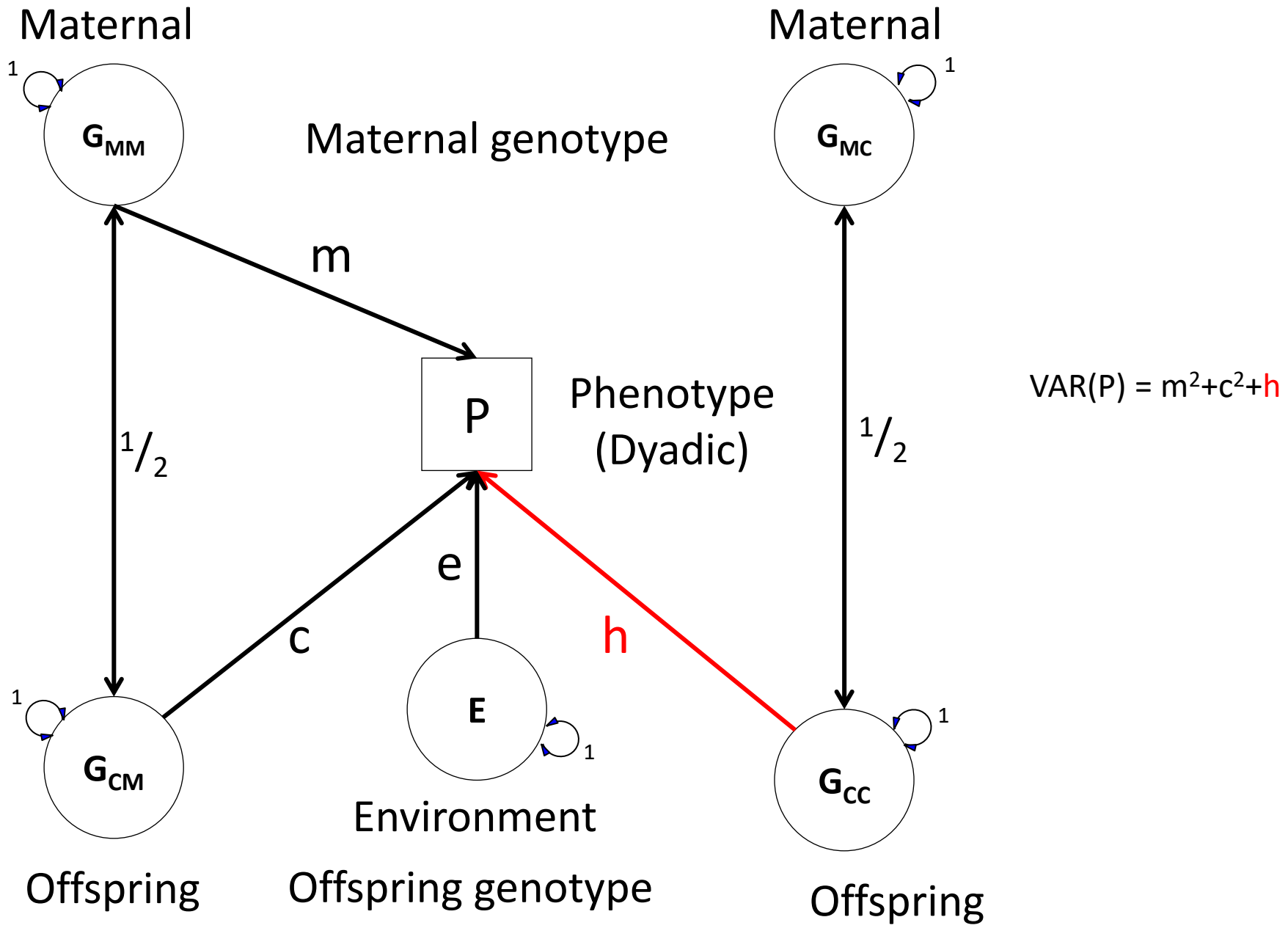


$VAR(P) = m^2 + cx1xc$

Path model for offspring and maternal effects in mother-child dyads.

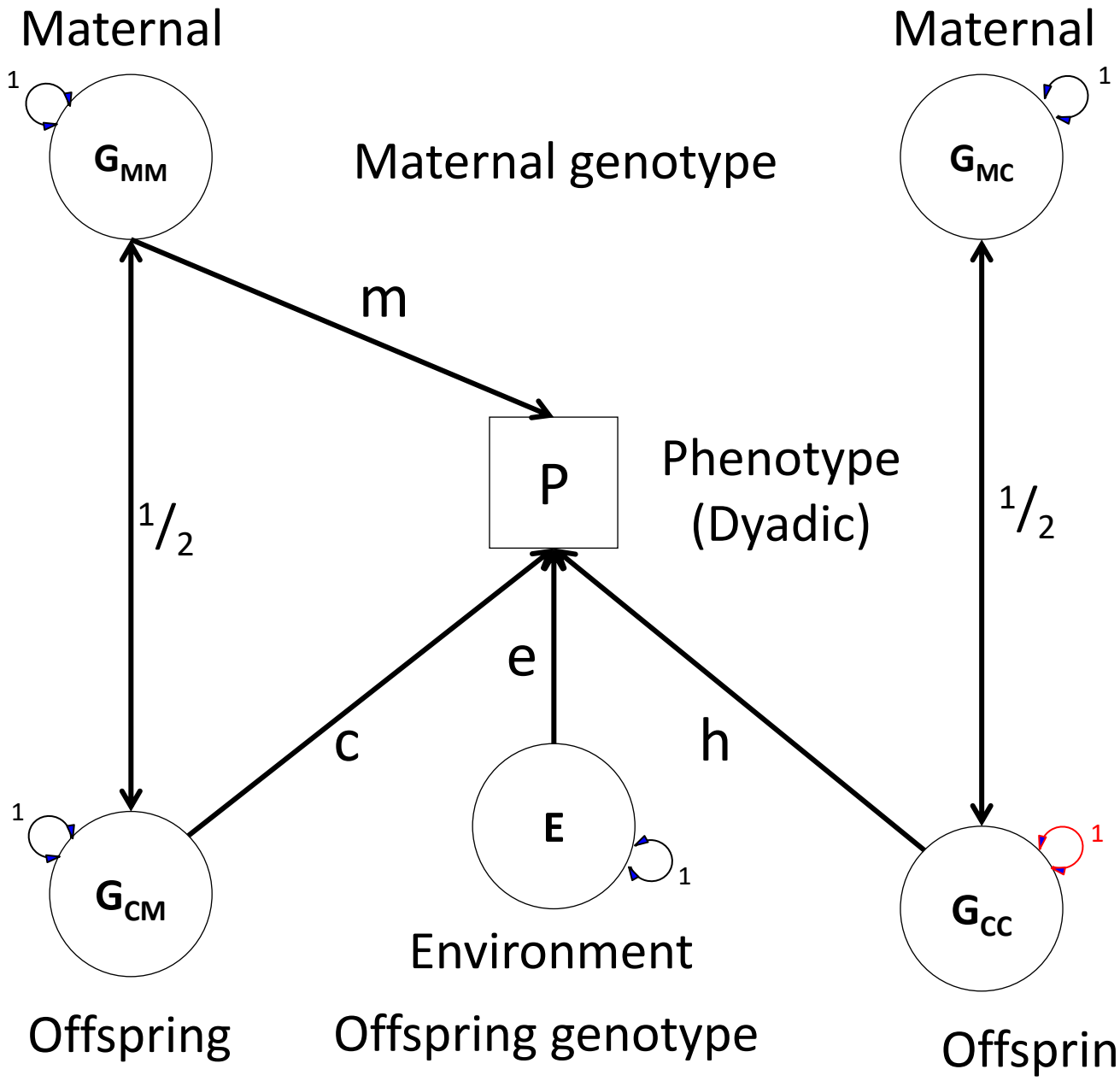


Path model for offspring and maternal effects in mother-child dyads.



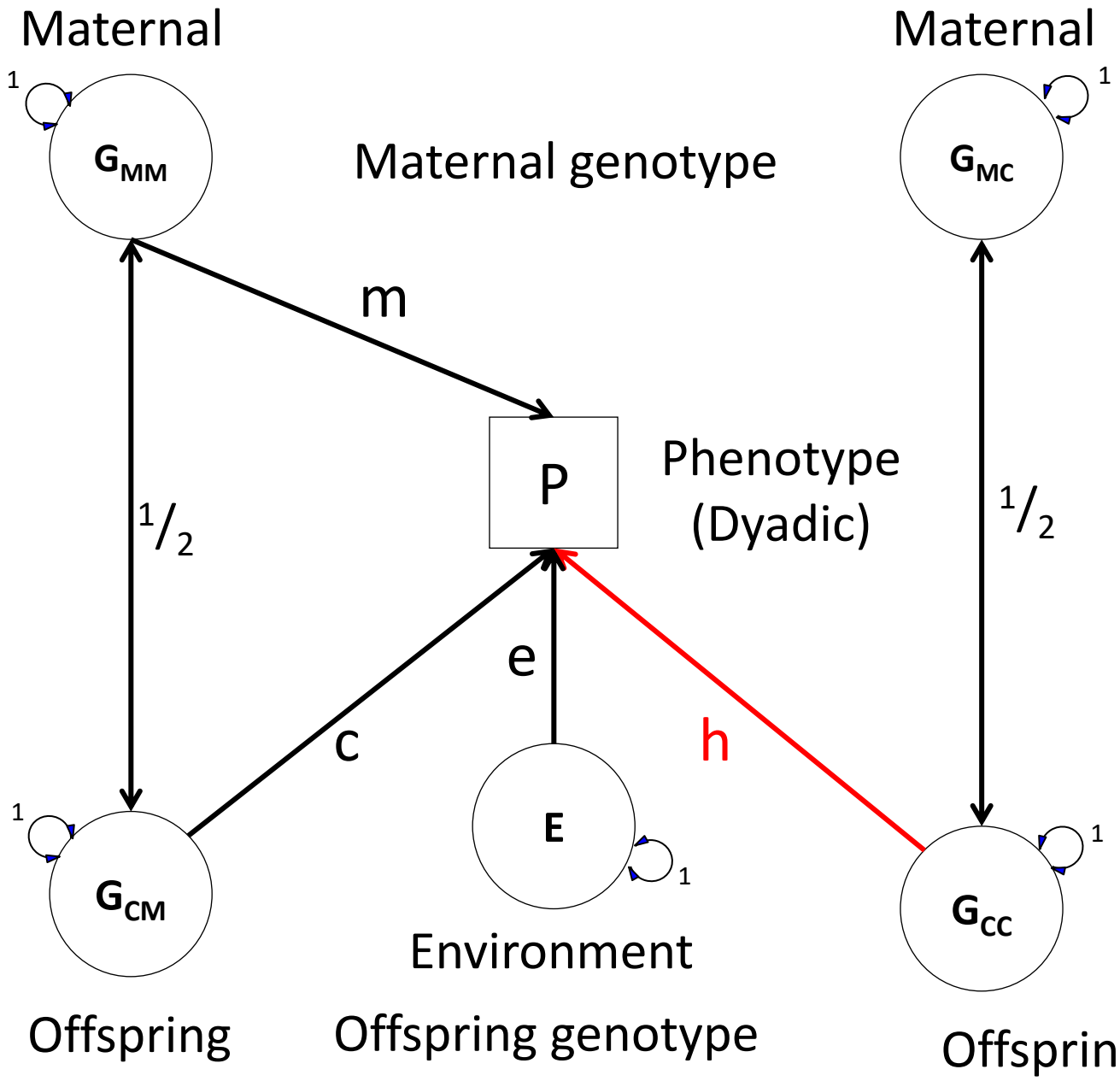
$VAR(P) = m^2+c^2+h$

Path model for offspring and maternal effects in mother-child dyads.



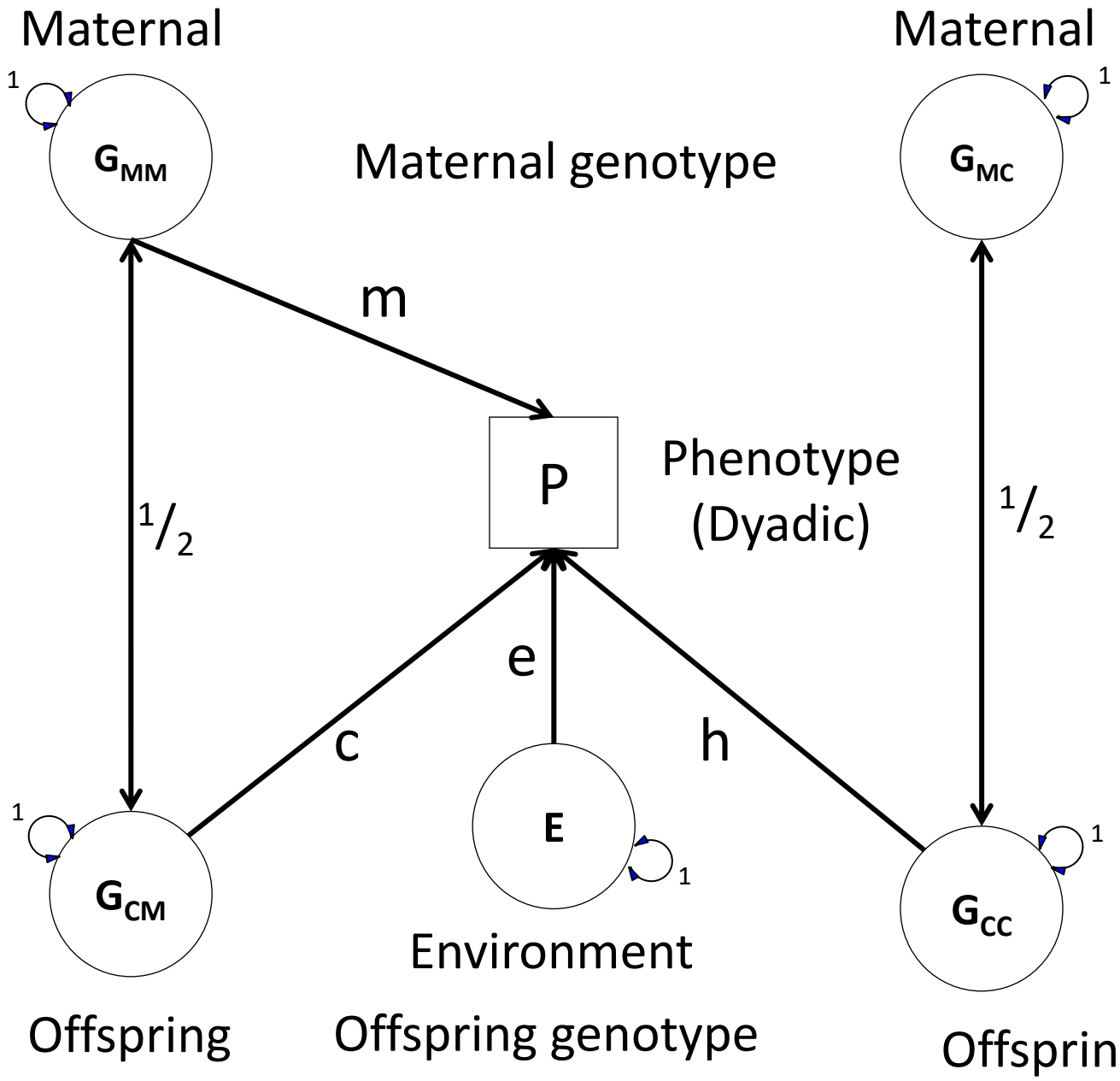
$VAR(P) = m^2+c^2+hx1$

Path model for offspring and maternal effects in mother-child dyads.



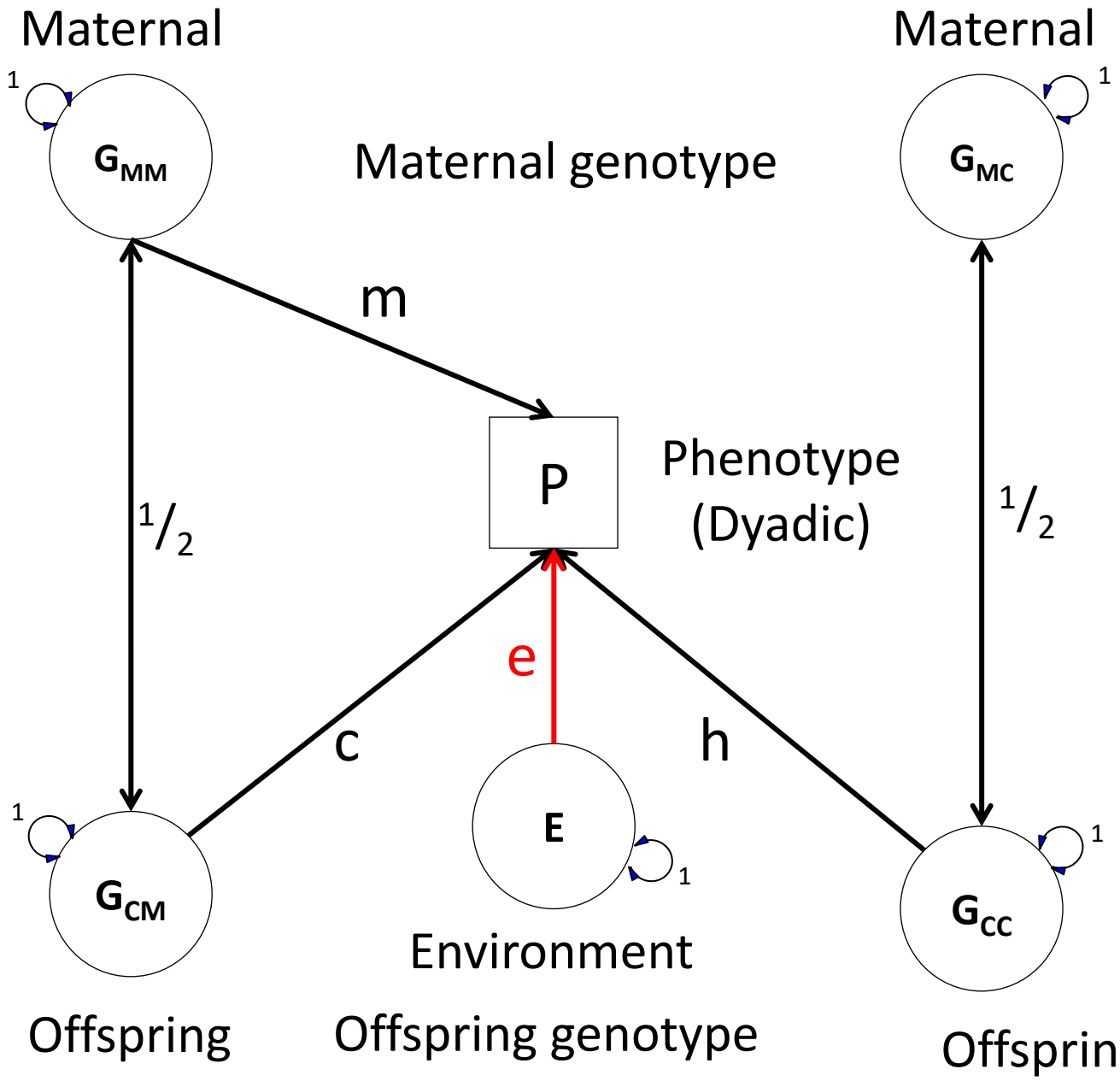
$VAR(P) = m^2+c^2+hx1xh$

Path model for offspring and maternal effects in mother-child dyads.



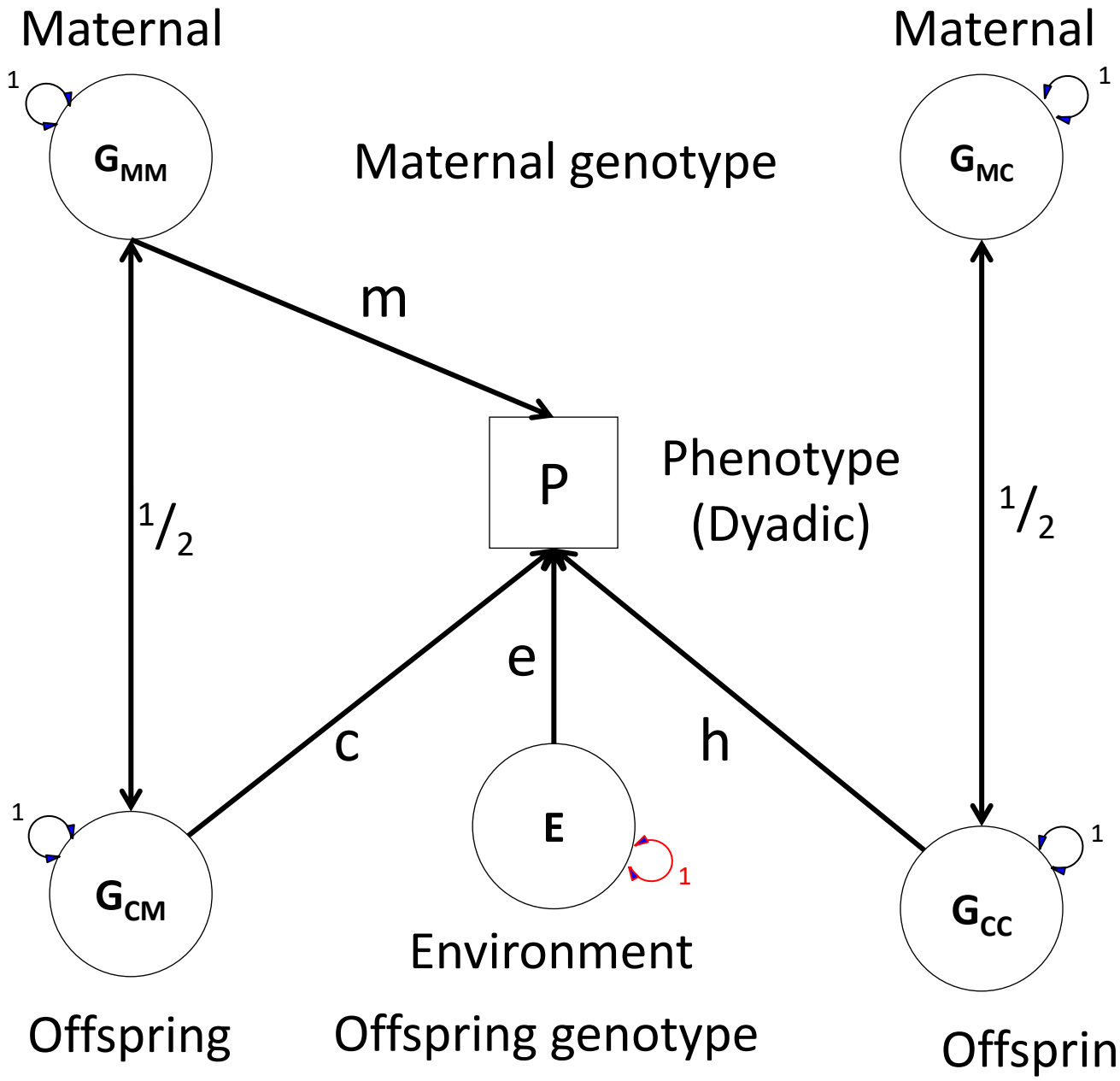
$VAR(P) = m^2+c^2+h^2$

Path model for offspring and maternal effects in mother-child dyads.



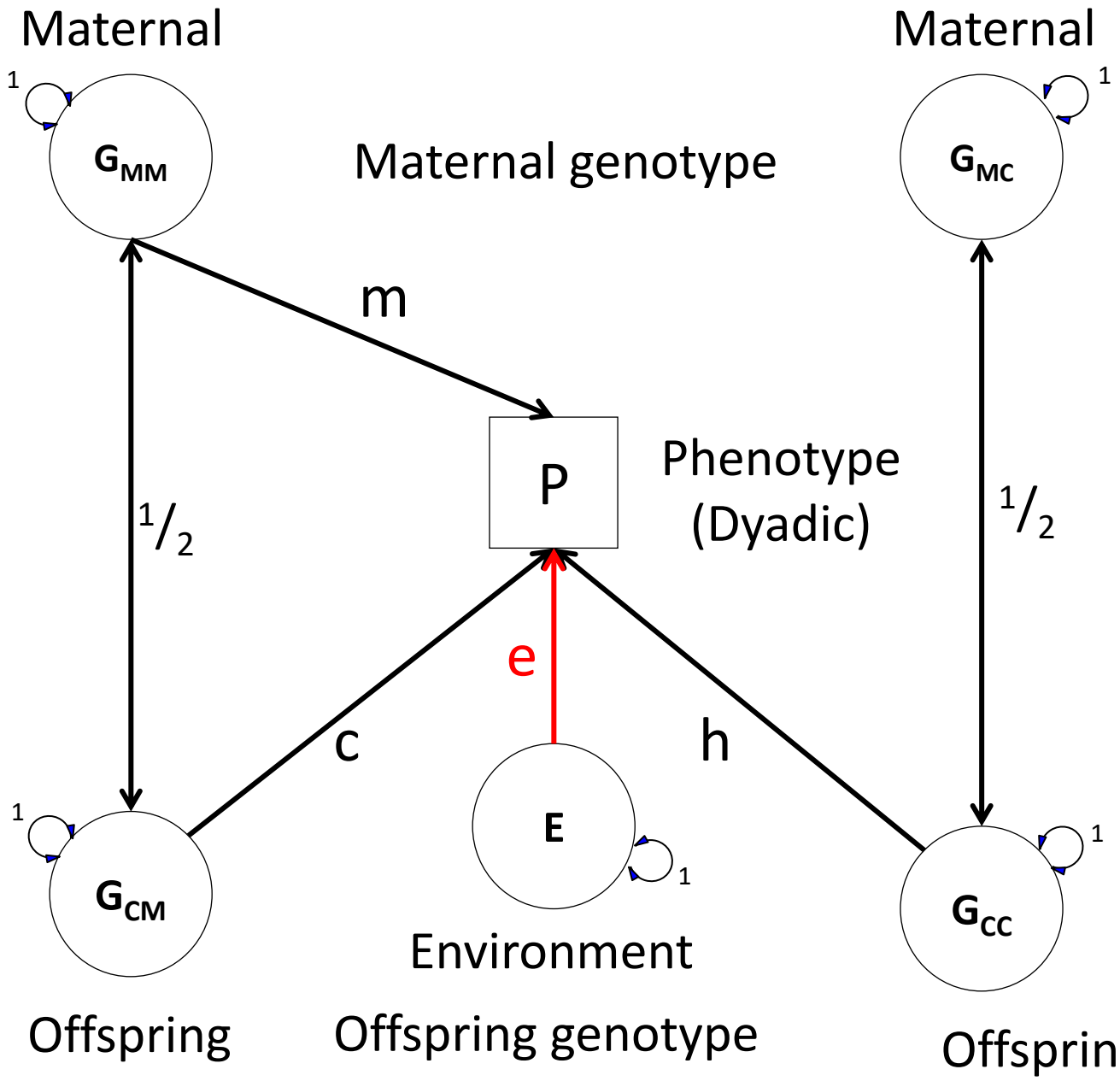
$VAR(P) = m^2+c^2+h^2+e$

Path model for offspring and maternal effects in mother-child dyads.



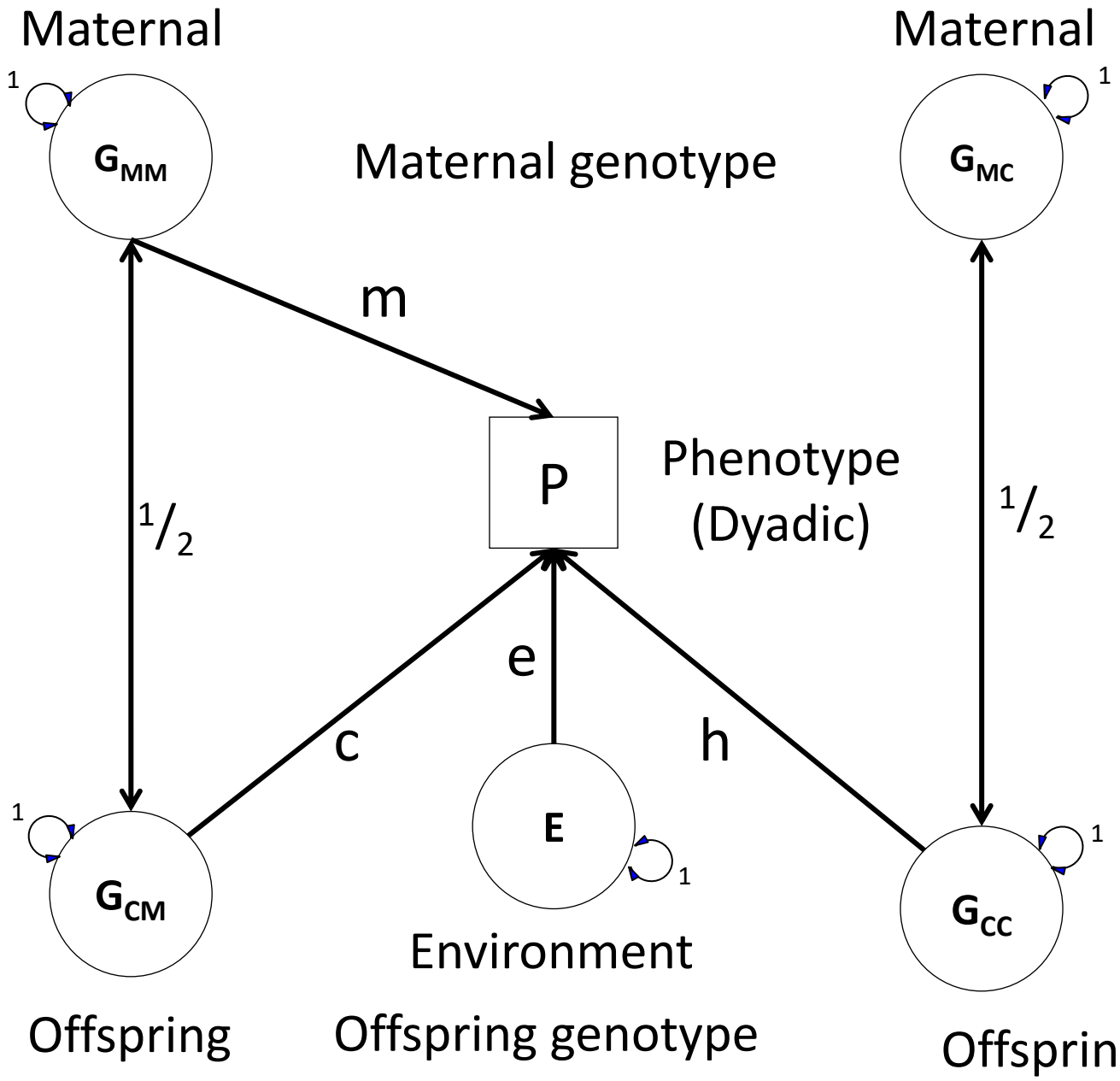
$VAR(P) = m^2+c^2+h^2+ex1$

Path model for offspring and maternal effects in mother-child dyads.



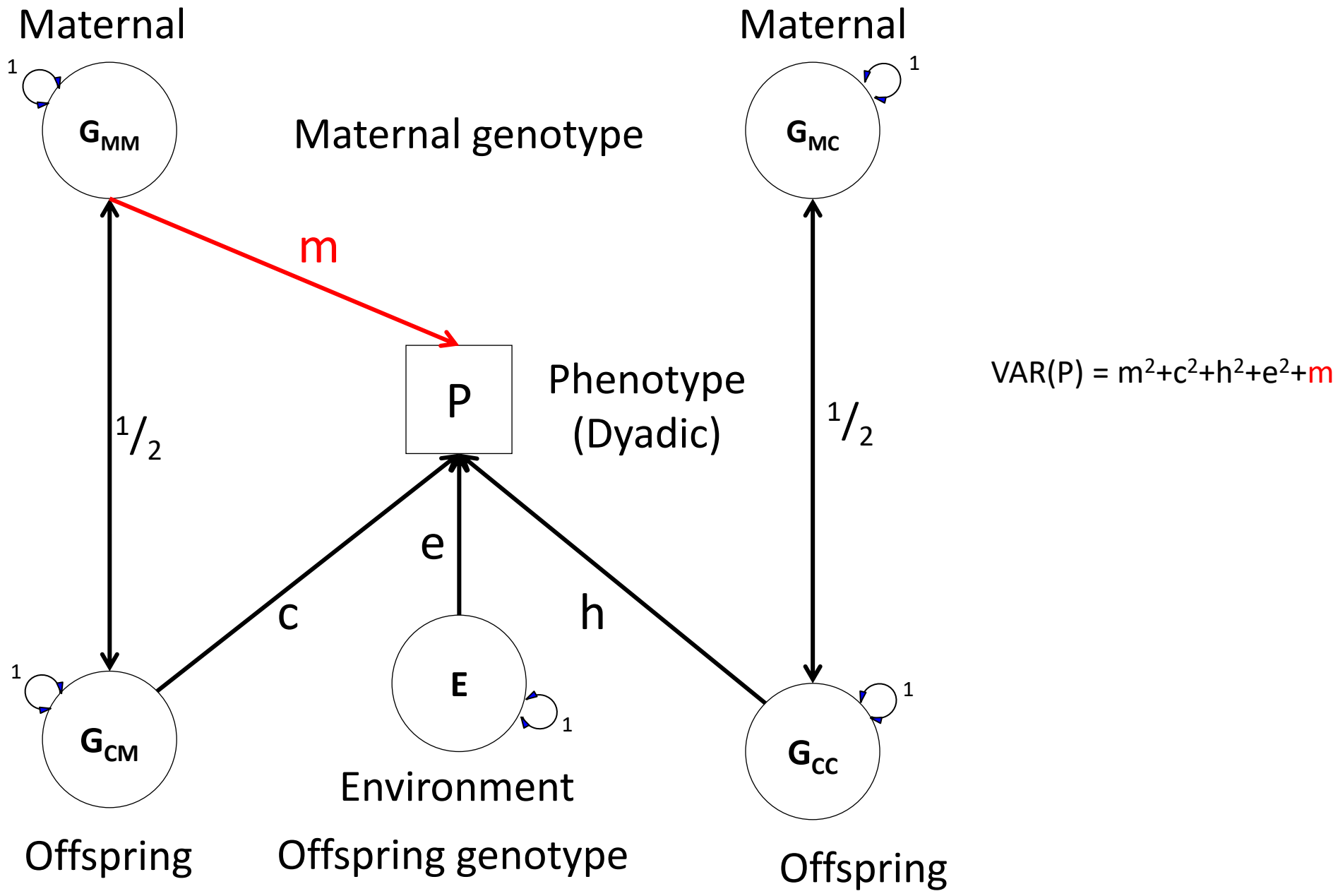
$VAR(P) = m^2+c^2+h^2+ex1xe$

Path model for offspring and maternal effects in mother-child dyads.



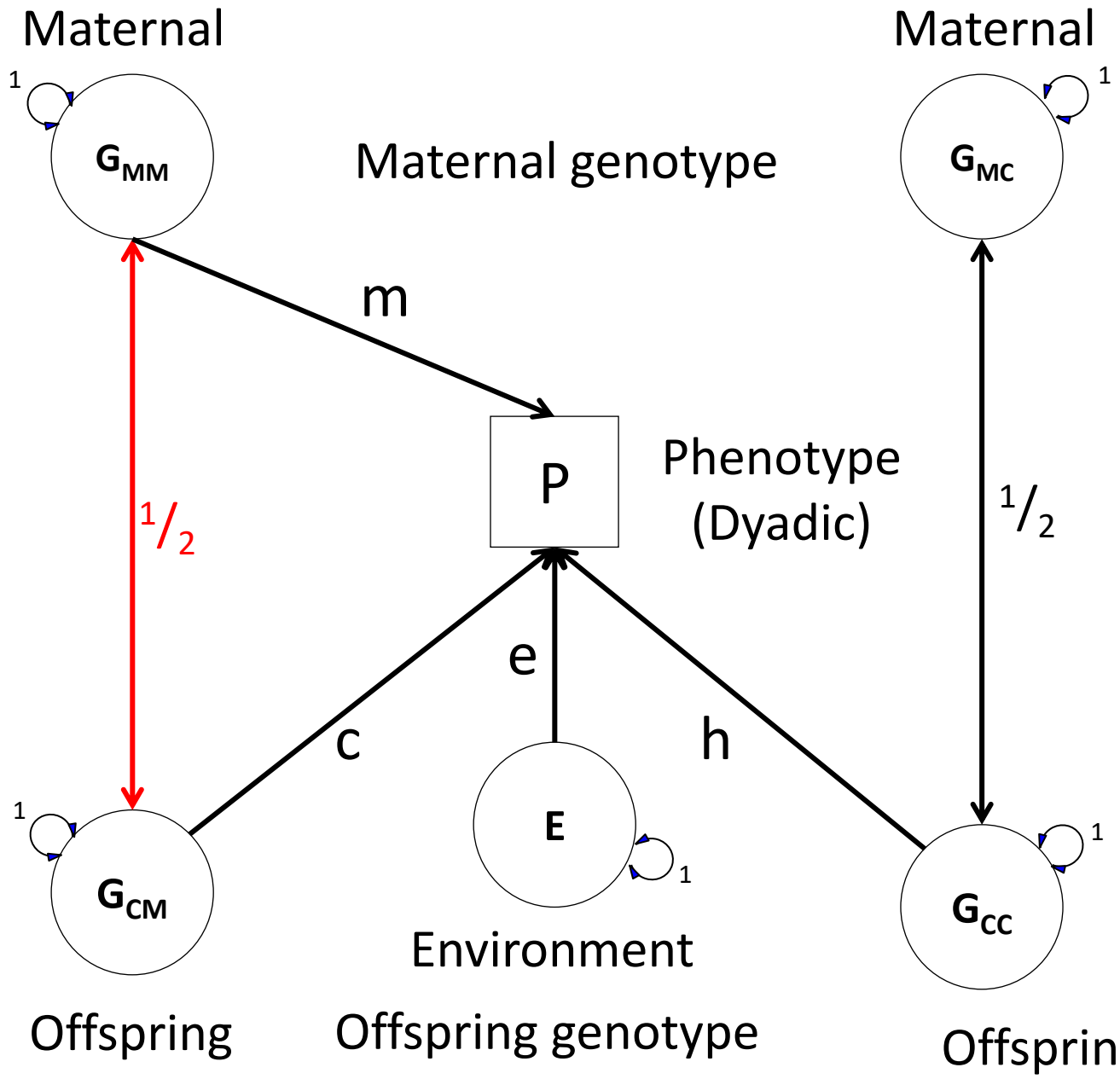
$VAR(P) = m^2+c^2+h^2+e^2$

Path model for offspring and maternal effects in mother-child dyads.



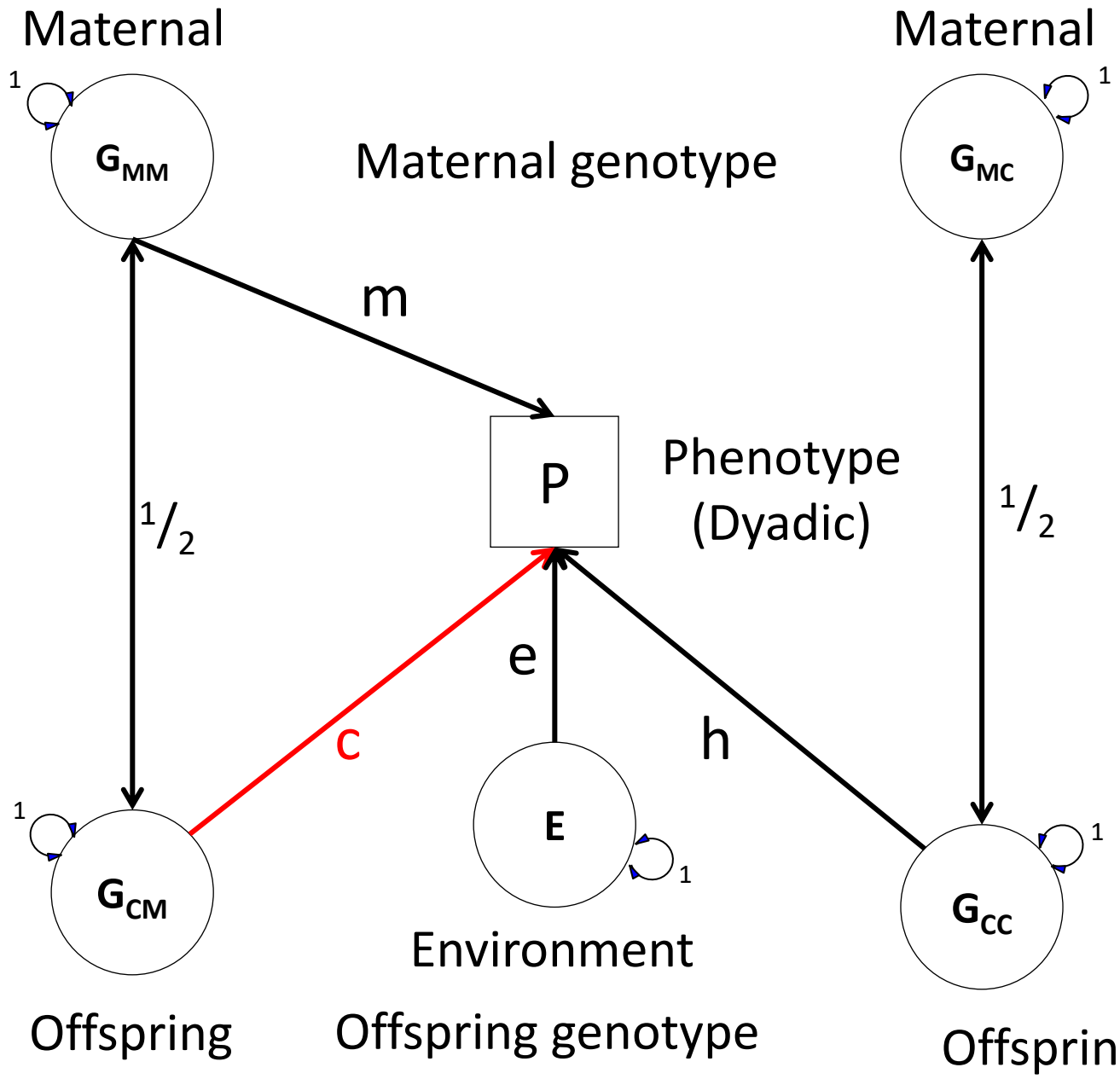
$VAR(P) = m^2+c^2+h^2+e^2+m$

Path model for offspring and maternal effects in mother-child dyads.



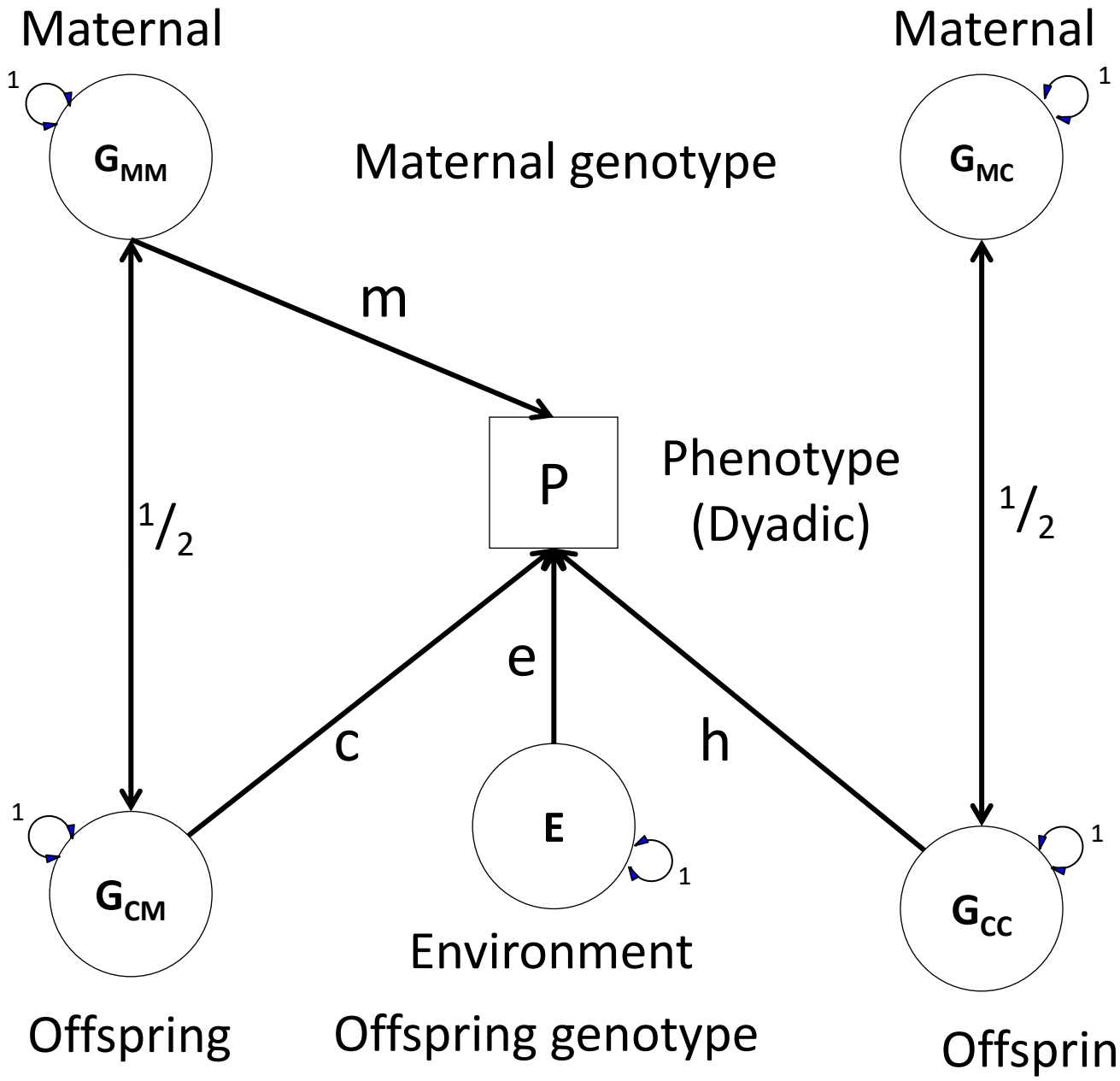
$VAR(P) = m^2+c^2+h^2+e^2+mx½$

Path model for offspring and maternal effects in mother-child dyads.



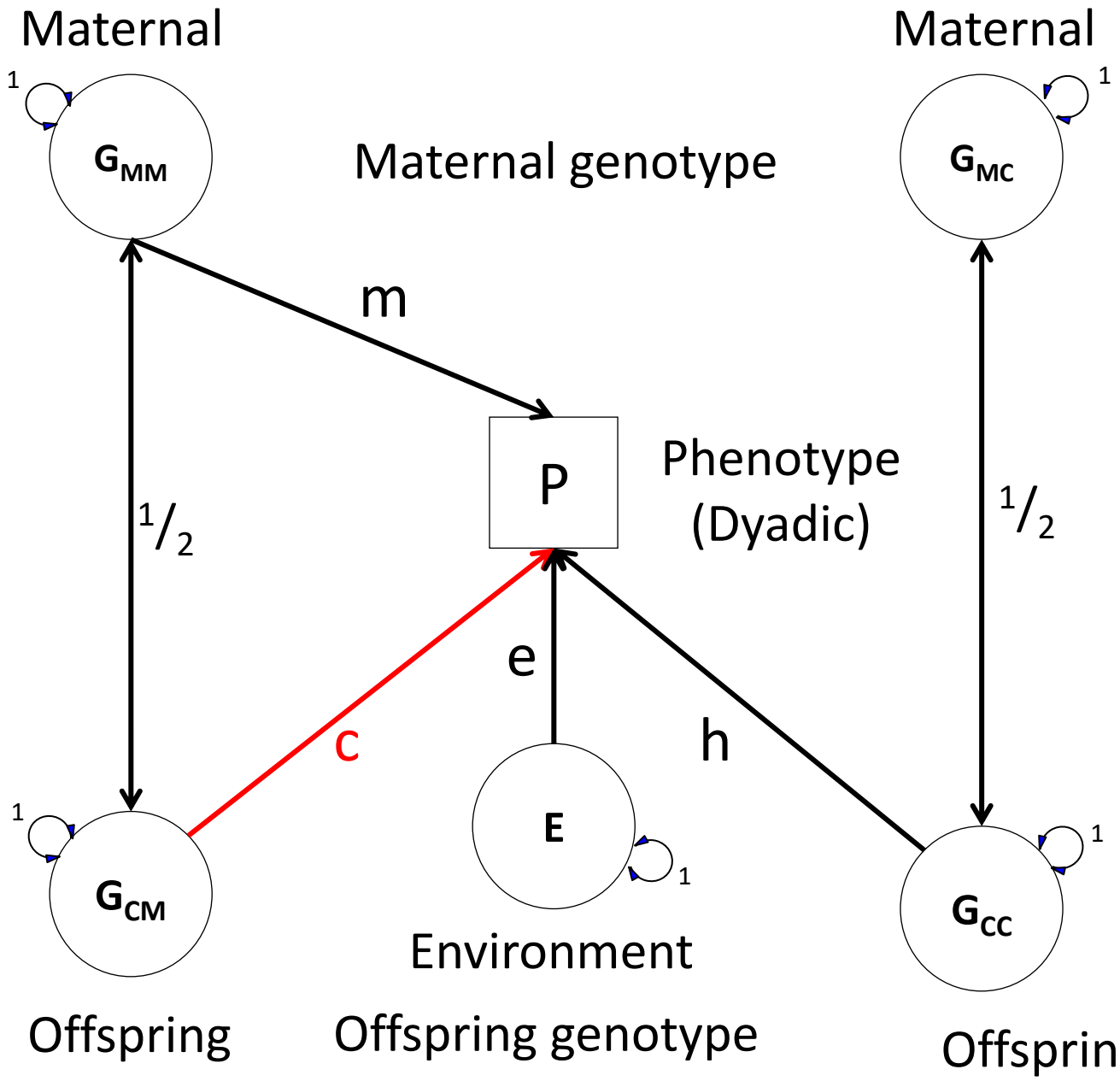
$VAR(P) = m^2+c^2+h^2+e^2+mx½xc$

Path model for offspring and maternal effects in mother-child dyads.



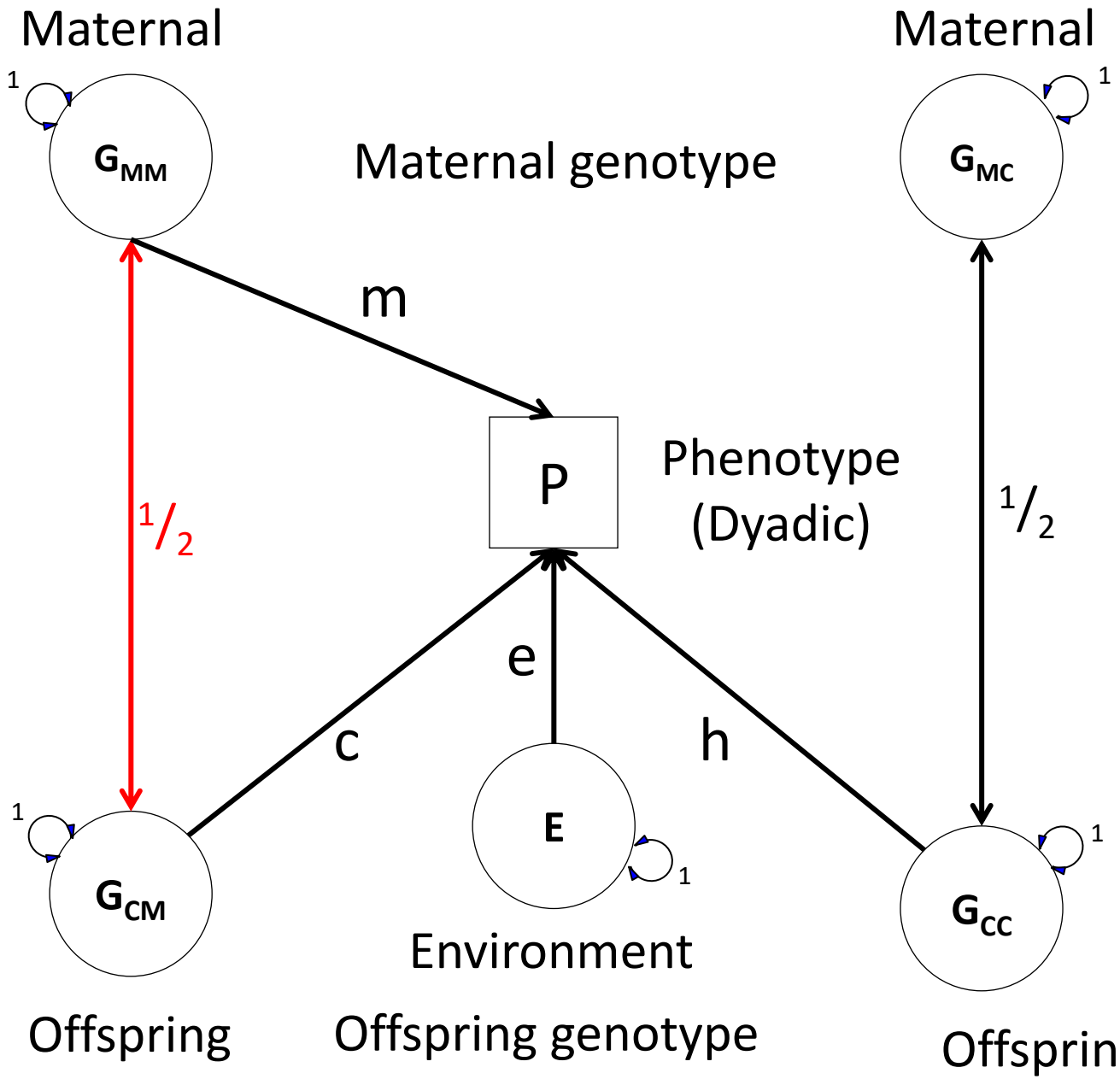
$VAR(P) = m^2+c^2+h^2+e^2+ ½mc$

Path model for offspring and maternal effects in mother-child dyads.

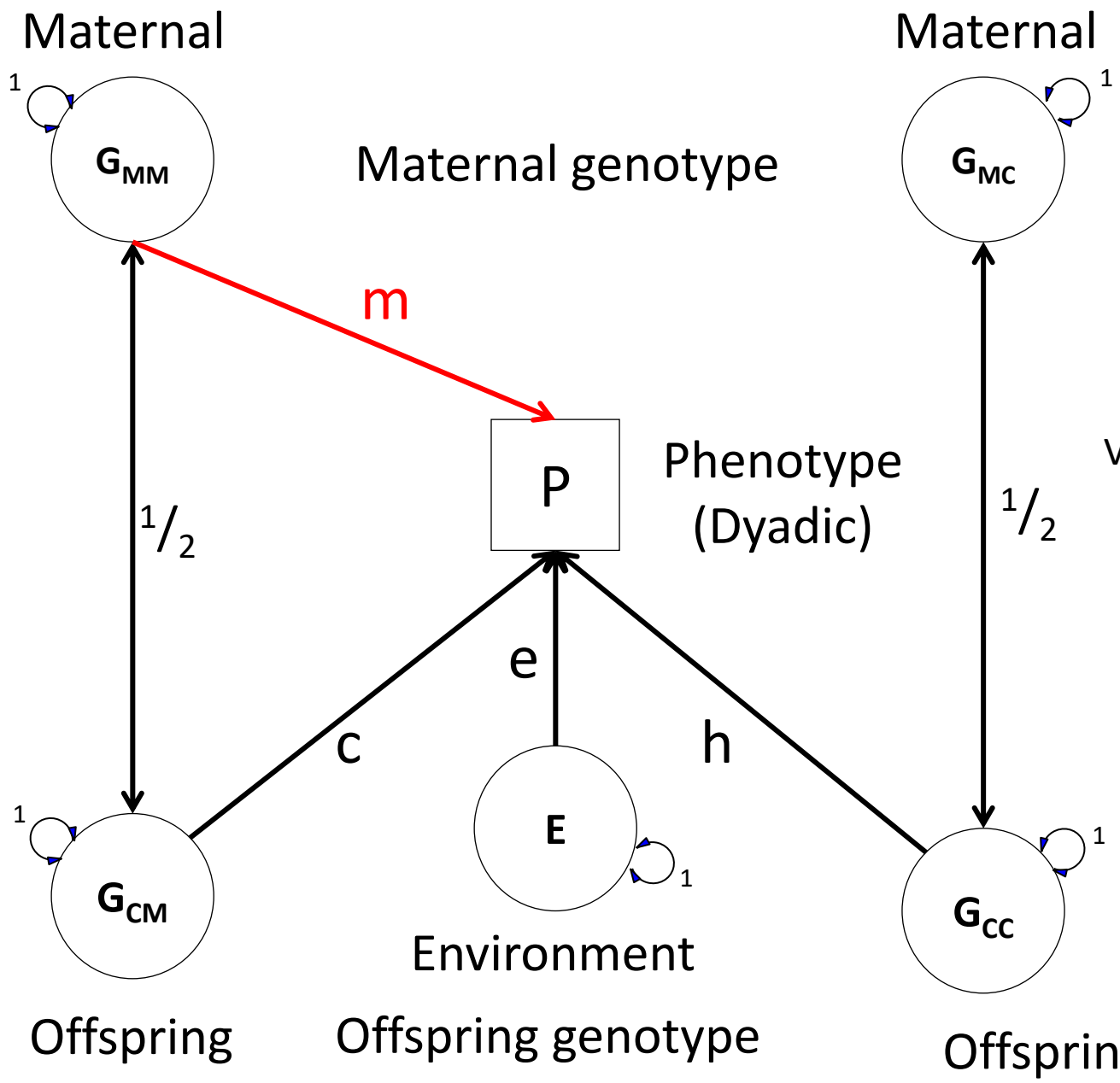


$VAR(P) = m^2+c^2+h^2+e^2+ ½mc+c$

Path model for offspring and maternal effects in mother-child dyads.

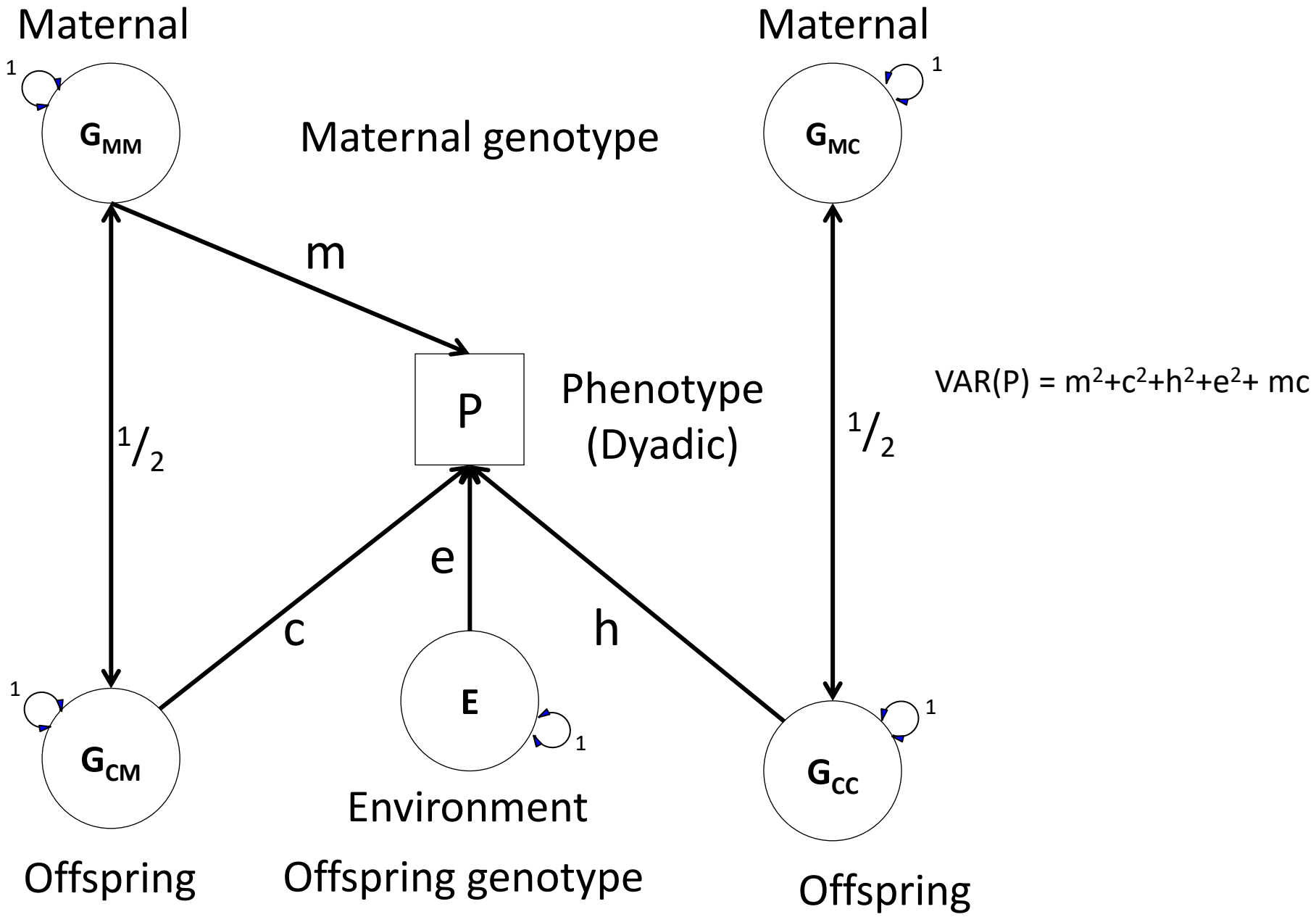


Path model for offspring and maternal effects in mother-child dyads.

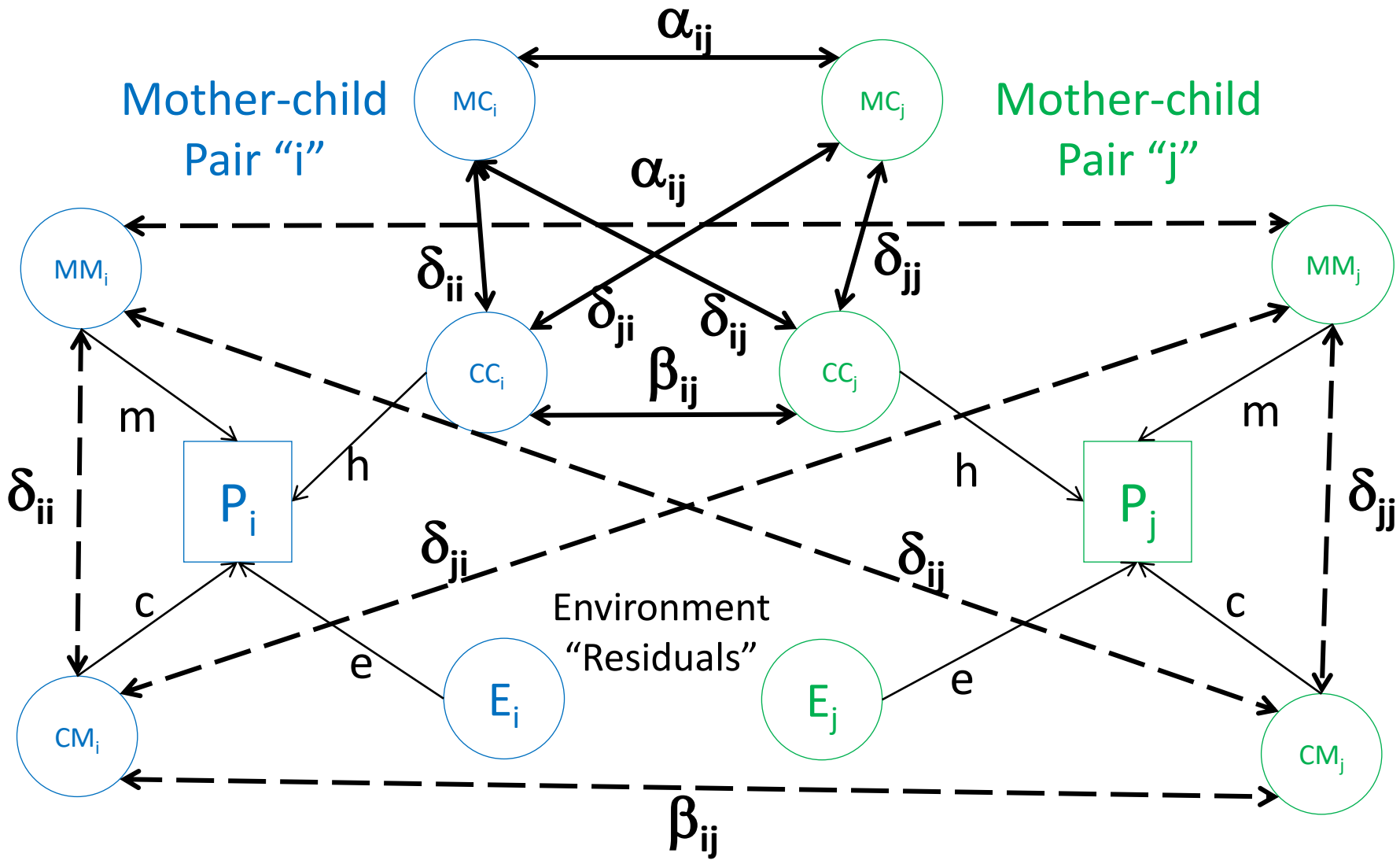


$VAR(P) = m^2+c^2+h^2+e^2+ ½mc+cx½xm$

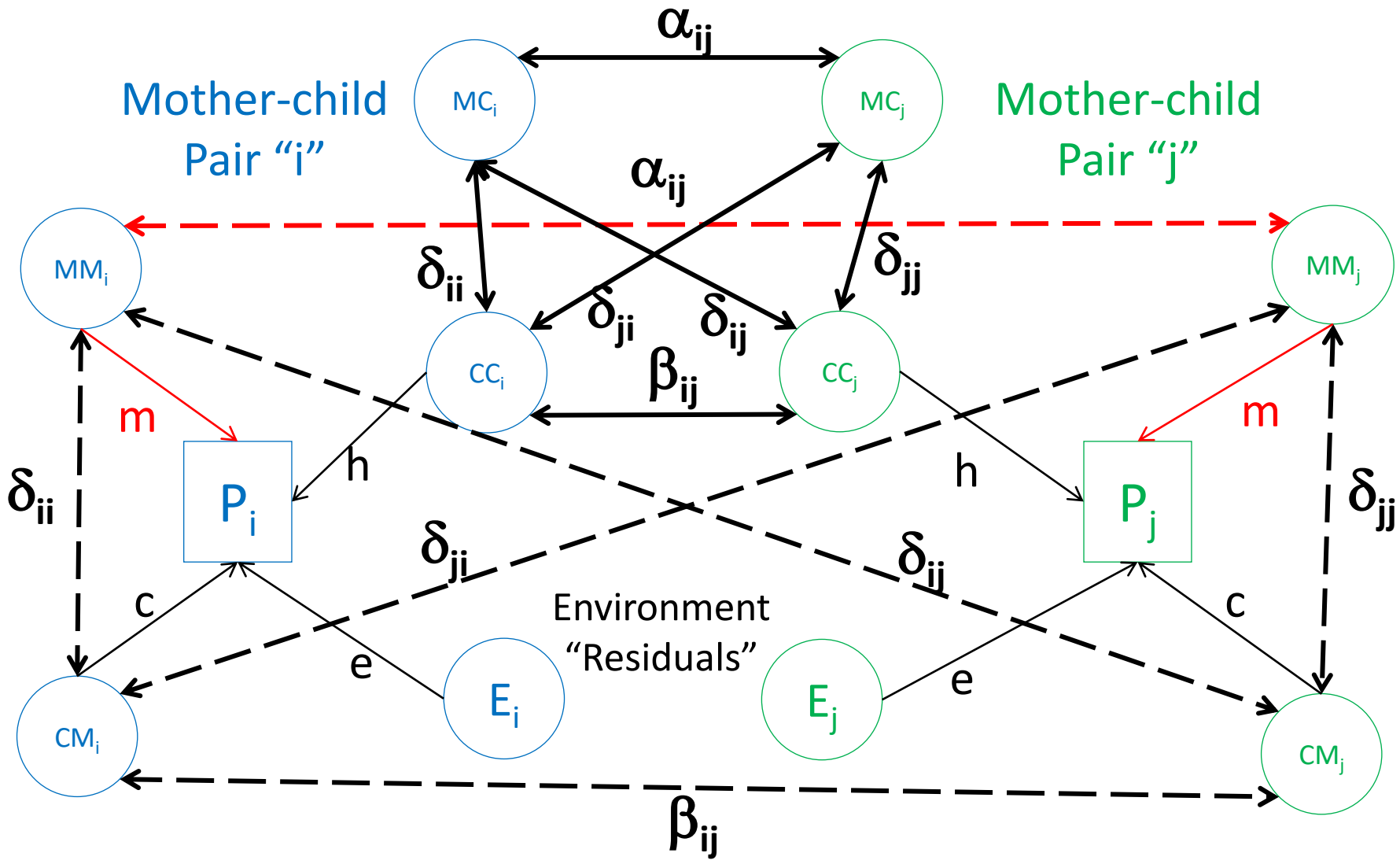
Path model for offspring and maternal effects in mother-child dyads.



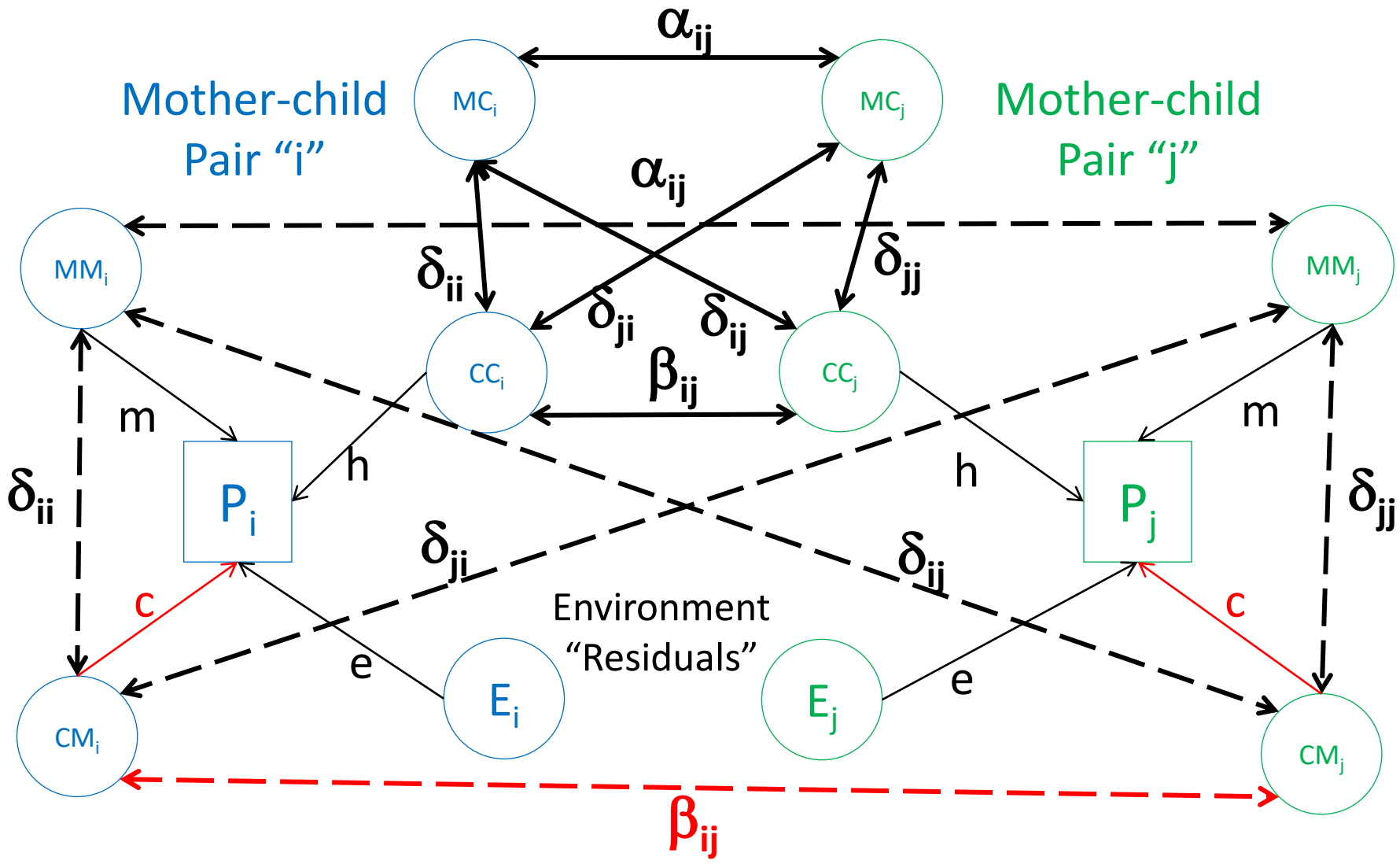
Deriving the Phenotypic Covariance



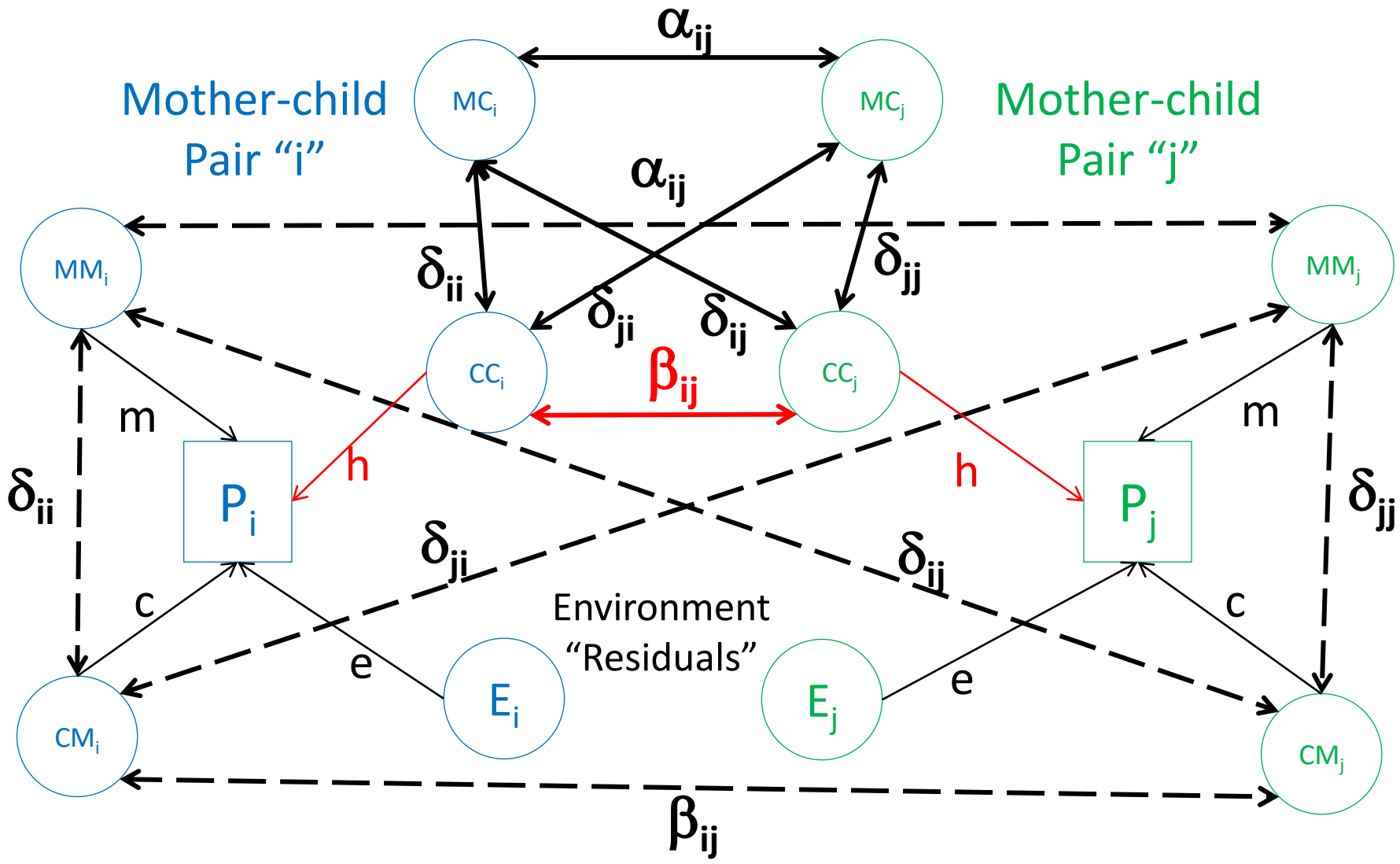
$\text{cov}(P_i, P_j) =$



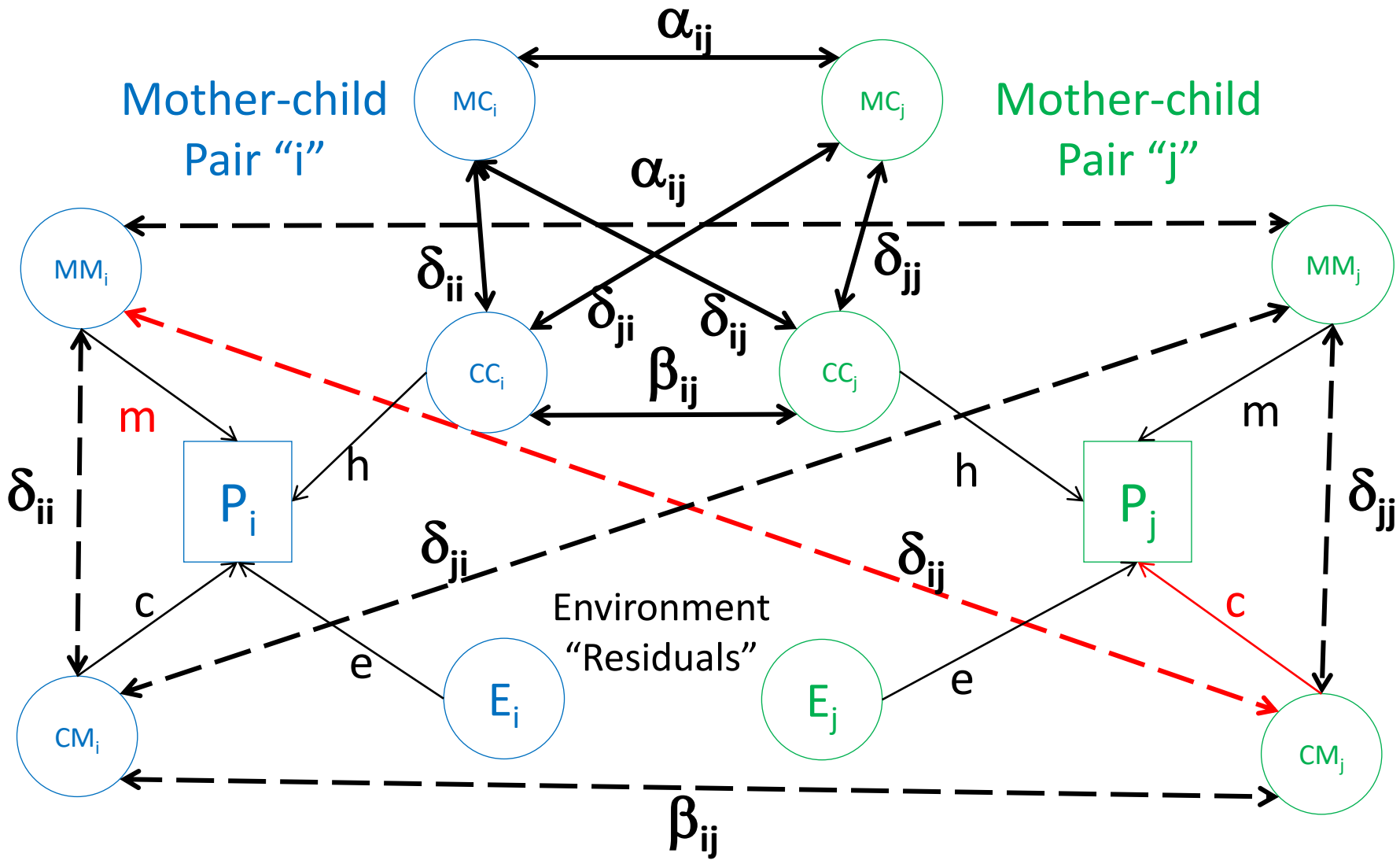
$\text{cov}(P_i, P_j) = \alpha_{ij} m^2$



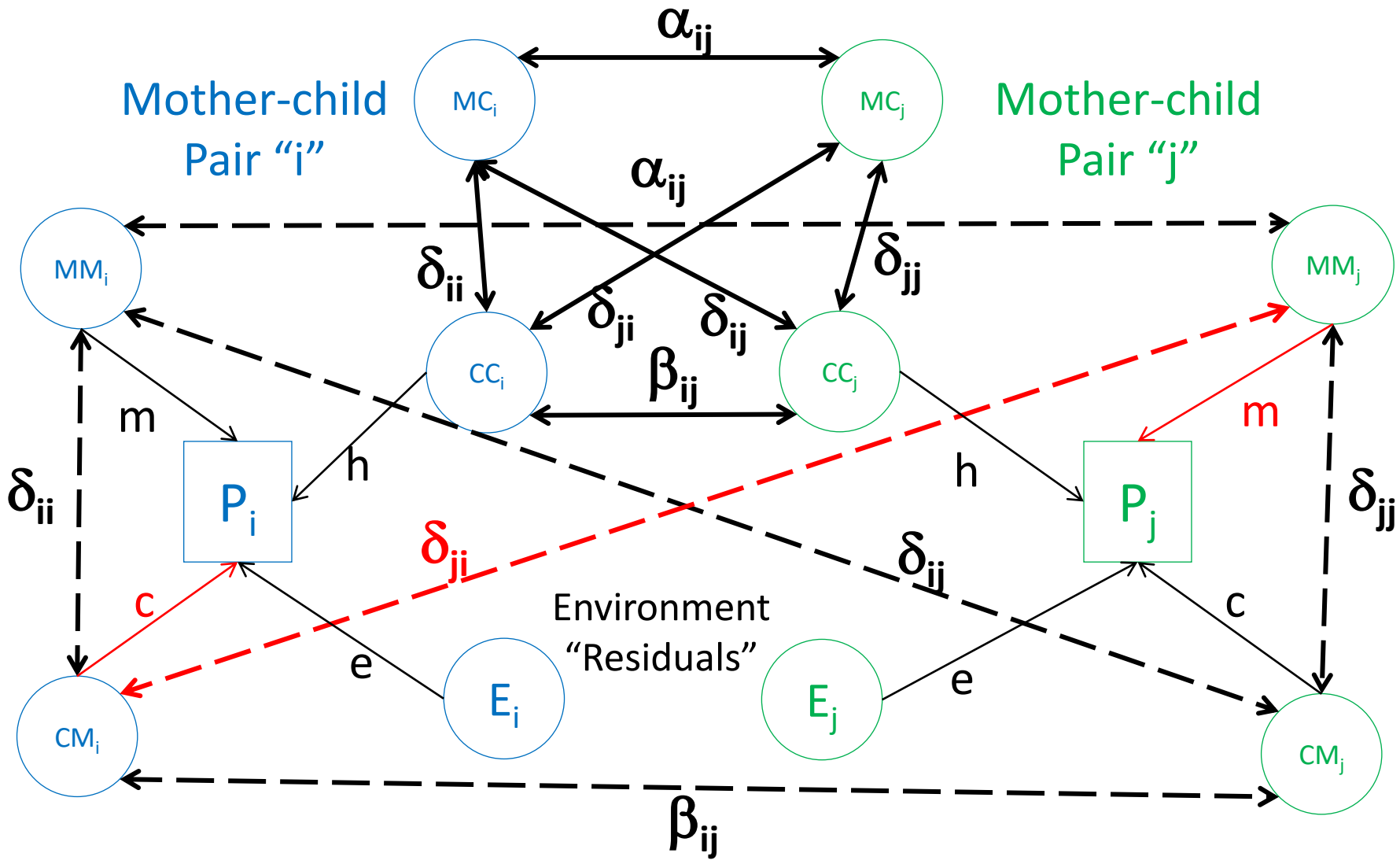
$$\text{cov}(P_i, P_j) = \alpha_{ij}m^2 + \beta_{ij}c^2$$



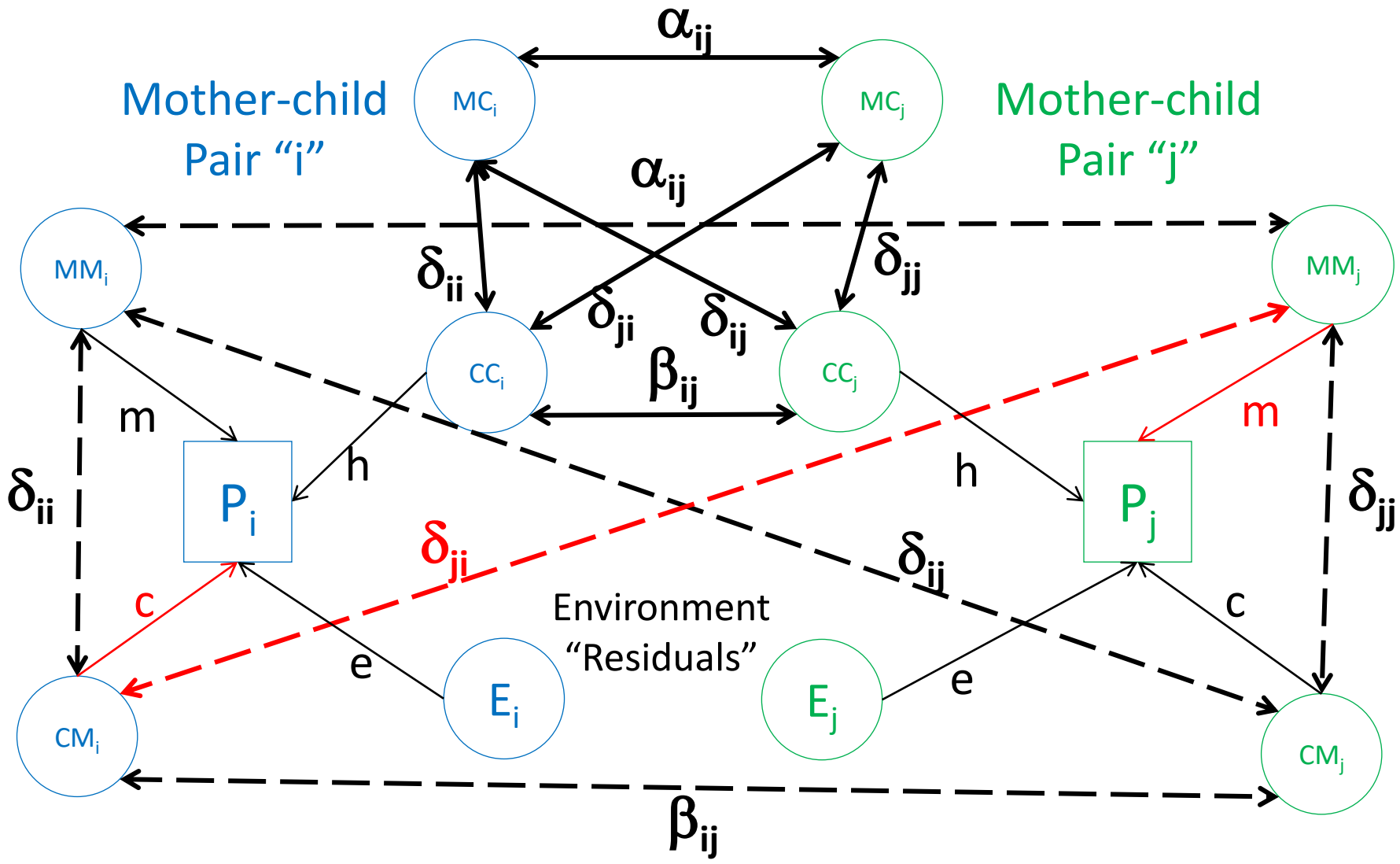
$$\text{cov}(P_i, P_j) = \alpha_{ij}m^2 + \beta_{ij}c^2 + \beta_{ij}h^2$$



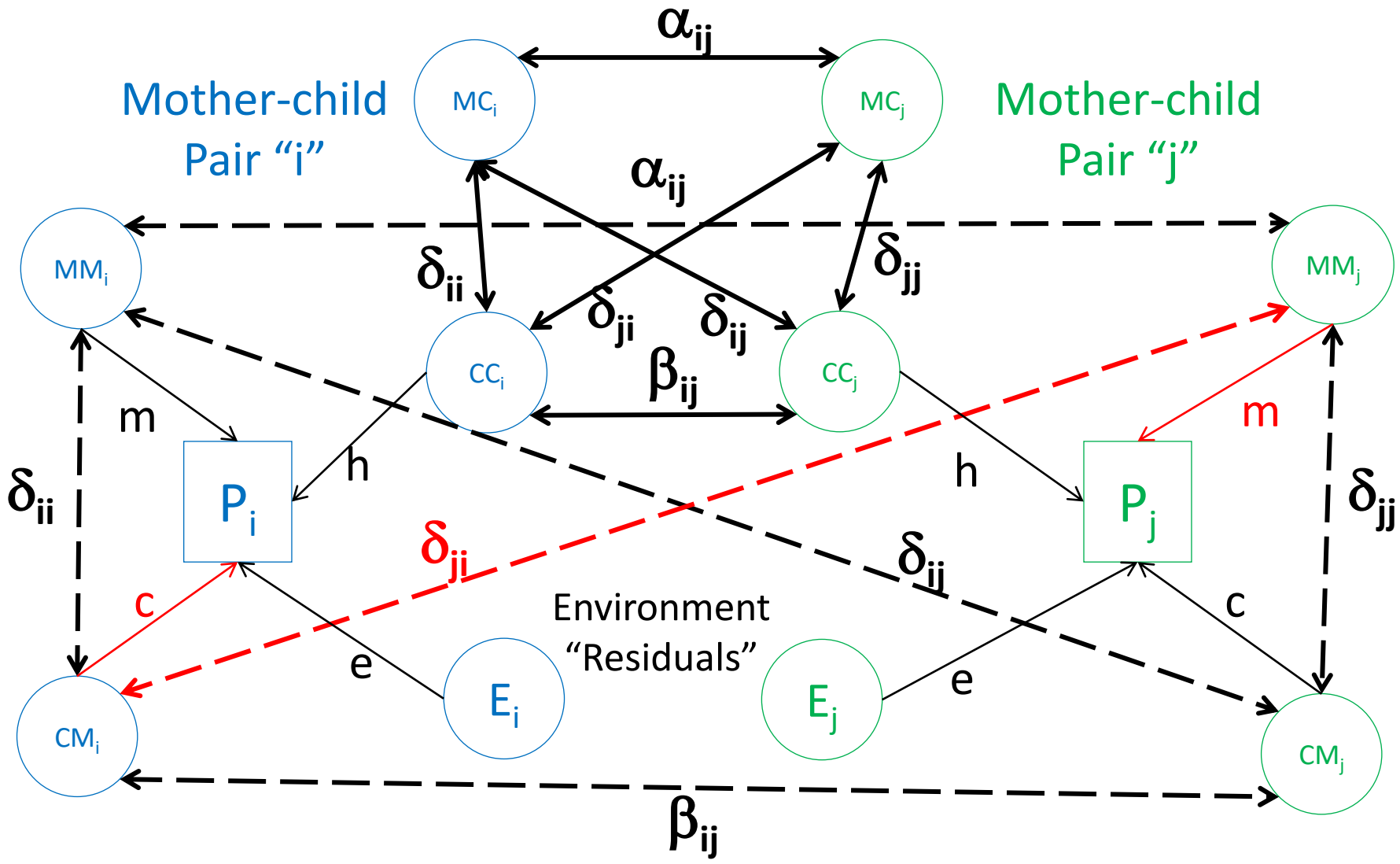
$$\text{cov}(P_i, P_j) = \alpha_{ij}m^2 + \beta_{ij}(c^2 + h^2) + \delta_{ij}mc$$



$$\text{cov}(P_i, P_j) = \alpha_{ij}m^2 + \beta_{ij}(c^2 + h^2) + \delta_{ij}mc + \delta_{ji}mc$$



$$\text{cov}(P_i, P_j) = \alpha_{ij}m^2 + \beta_{ij}(c^2 + h^2) + (\delta_{ij} + \delta_{ji})mc$$



$$\text{cov}(P_i, P_j) = \alpha_{ij}m^2 + \beta_{ij}(c^2 + h^2) + (\delta_{ij} + \delta_{ji})mc$$

$$\text{VAR}(P) = m^2 + c^2 + h^2 + mc + e^2$$

Estimate Path Coefficients*

$$\boldsymbol{\Sigma} = \mathbf{A}m^2 + \mathbf{B}(c^2+h^2) + \boldsymbol{\Delta}mc + \mathbf{I}e^2$$

$\sum$ is the expected phenotypic covariance matrix for the offspring

$\mathbf{m^2}$ is the additive genetic variance from the maternal genome

$\mathbf{c^2+h^2}$ is the additive genetic variance from the child's genome

mc is the covariance arising because of the same genes with (indirect) maternal and (direct) fetal effects

$\mathbf{e^2}$ is the residual variance (includes environmental variance)

A is a **GRM** containing pairwise genome-wide genetic correlation between mothers

B is a **GRM** containing pairwise genome-wide genetic correlation between children

Δ is a **GRM** containing pairwise (twice) the genome-wide genetic correlation between mothers and children

I is an identity matrix

***Note the implicit constraint in that the covariance mc can only be != 0 if both m and c are non-zero**

Formulating the Delta Matrix

$$\Delta = \begin{array}{c} \\ \text{Mother\_1} \\ \text{Mother\_2} \\ \text{Mother\_3} \\ \dots \\ \text{Mother\_n} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Child\_1} & \text{Child\_2} & \text{Child\_3} & \dots & \text{Child\_n} \end{array} \\ \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \dots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \dots & \dots & \delta_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \delta_{n1} & \delta_{n2} & \dots & \dots & \delta_{nn} \end{bmatrix} \\ (n \times n) \end{array} + \begin{array}{c} \\ \text{Child\_1} \\ \text{Child\_2} \\ \text{Child\_3} \\ \dots \\ \text{Child\_n} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Mother\_1} & \text{Mother\_2} & \text{Mother\_3} & \dots & \text{Mother\_n} \end{array} \\ \begin{bmatrix} \delta_{11} & \delta_{21} & \dots & \dots & \delta_{n1} \\ \delta_{12} & \delta_{22} & \dots & \dots & \delta_{n2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \delta_{1n} & \delta_{2n} & \dots & \dots & \delta_{nn} \end{bmatrix} \\ (n \times n) \end{array}$$

Δ is a symmetric matrix derived by adding the mother-child GRM to its transpose

Estimate Variance Components (Alternative Formulation)*

$$\boldsymbol{\Sigma} = \mathbf{A}\sigma^2_m + \mathbf{B}\sigma^2_g + \boldsymbol{\Delta}\sigma^2_q + \mathbf{I}\sigma^2_e$$

$\sum$ is the expected phenotypic covariance matrix

$\boldsymbol{\sigma^2_m}$ is the additive genetic variance from the maternal genome

$\boldsymbol{\sigma^2_g}$ is the additive genetic variance from the child's genome

$\boldsymbol{\sigma^2_q}$ is the covariance arising because of the same genes with (indirect) maternal and (direct) fetal effects

$\boldsymbol{\sigma^2_e}$ is the residual variance (includes environmental variance)

A is a **GRM** containing pairwise genome-wide genetic "correlation" between mothers

B is a **GRM** containing pairwise genome-wide genetic "correlation" between children

Δ is a **GRM** containing pairwise (twice) the genome-wide genetic "correlation" between mothers and children

I is an identity matrix

***This model could be fitted in e.g. the GCTA software package, but it contains no implicit constraint**

Some Empirical Results

Table 4 Results of fitting "GCTA" model for direct and maternal effects to maternal stature and birth length data in the ALSPAC cohort. Results are presented as standardized variance components

	Model	G	M	Q	E	−2lnl	χ^2	d.f.	P %
Maternal stature	Full	0.15 (0.11)	0.72 (0.11)	−0.08 (0.09)	0.21 (0.12)	19825.526	–	–	–
	Q = 0	0.09 (0.08)	0.66 (0.09)	–	0.25 (0.10)	19826.412	0.886	1	35
	M = Q = 0	0.24 (0.08)	–	–	0.76 (0.09)	19881.294	55.768	2	8×10^{-11}
	G = Q = 0	–	0.68 (0.08)	–	0.32 (0.08)	19827.480	1.954	2	38
	M = G = Q = 0	–	–	–	1	19889.516	63.99	3	8×10^{-12}
Birth length	Full	0.13 (0.13)	0.11 (0.13)	0.06 (0.10)	0.70 (0.14)	3553.870	–	–	–
	Q = 0	0.18 (0.10)	0.16 (0.10)	–	0.66 (0.13)	3554.242	0.37	1	54
	M = Q = 0	0.22 (0.10)	–	–	0.78 (0.10)	3556.624	2.75	2	25
	G = Q = 0	–	0.20 (0.10)	–	0.80 (0.10)	3557.646	3.78	2	15
	M = G = Q = 0	–	–	–	1	3561.774	7.90	3	5

Model parameters are: *G* variance due to direct genetic ("offspring/fetal") effects, *M* variance due to indirect genetic effects on offspring phenotype ("maternal effects"), *Q* phenotypic variance due to covariance of direct and indirect genetic effects, *E* residual ("unique environmental effects")

Some Empirical Results

Table 4 Results of fitting "GCTA" model for direct and maternal effects to maternal stature and birth length data in the ALSPAC cohort. Results are presented as standardized variance components

	Model	G	M	Q	E	−2lnl	χ^2	d.f.	P %
Maternal stature	Full	0.15 (0.11)	0.72 (0.11)	−0.08 (0.09)	0.21 (0.12)	19825.526	–	–	–
	Q = 0	0.09 (0.08)	0.66 (0.09)	–	0.25 (0.10)	19826.412	0.886	1	35
	M = Q = 0	0.24 (0.08)	–	–	0.76 (0.09)	19881.294	55.768	2	8×10^{-11}
	G = Q = 0	–	0.68 (0.08)	–	0.32 (0.08)	19827.480	1.954	2	38
	M = G = Q = 0	–	–	–	1	19889.516	63.99	3	8×10^{-12}
Birth length	Full	0.13 (0.13)	0.11 (0.13)	0.06 (0.10)	0.70 (0.14)	3553.870	–	–	–
	Q = 0	0.18 (0.10)	0.16 (0.10)	–	0.66 (0.13)	3554.242	0.37	1	54
	M = Q = 0	0.22 (0.10)	–	–	0.78 (0.10)	3556.624	2.75	2	25
	G = Q = 0	–	0.20 (0.10)	–	0.80 (0.10)	3557.646	3.78	2	15
	M = G = Q = 0	–	–	–	1	3561.774	7.90	3	5

Model parameters are: *G* variance due to direct genetic ("offspring/fetal") effects, *M* variance due to indirect genetic effects on offspring phenotype ("maternal effects"), *Q* phenotypic variance due to covariance of direct and indirect genetic effects, *E* residual ("unique environmental effects")

Some Empirical Results

Table 4 Results of fitting "GCTA" model for direct and maternal effects to maternal stature and birth length data in the ALSPAC cohort. Results are presented as standardized variance components

	Model	G	M	Q	E	−2lnl	χ^2	d.f.	P %
Maternal stature	Full	0.15 (0.11)	0.72 (0.11)	−0.08 (0.09)	0.21 (0.12)	19825.526	–	–	–
	Q = 0	0.09 (0.08)	0.66 (0.09)	–	0.25 (0.10)	19826.412	0.886	1	35
	M = Q = 0	0.24 (0.08)	–	–	0.76 (0.09)	19881.294	55.768	2	8×10^{-11}
	G = Q = 0	–	0.68 (0.08)	–	0.32 (0.08)	19827.480	1.954	2	38
	M = G = Q = 0	–	–	–	1	19889.516	63.99	3	8×10^{-12}
Birth length	Full	0.13 (0.13)	0.11 (0.13)	0.06 (0.10)	0.70 (0.14)	3553.870	–	–	–
	Q = 0	0.18 (0.10)	0.16 (0.10)	–	0.66 (0.13)	3554.242	0.37	1	54
	M = Q = 0	0.22 (0.10)	–	–	0.78 (0.10)	3556.624	2.75	2	25
	G = Q = 0	–	0.20 (0.10)	–	0.80 (0.10)	3557.646	3.78	2	15
	M = G = Q = 0	–	–	–	1	3561.774	7.90	3	5

Model parameters are: *G* variance due to direct genetic ("offspring/fetal") effects, *M* variance due to indirect genetic effects on offspring phenotype ("maternal effects"), *Q* phenotypic variance due to covariance of direct and indirect genetic effects, *E* residual ("unique environmental effects")

Some Empirical Results

Table 4 Results of fitting "GCTA" model for direct and maternal effects to maternal stature and birth length data in the ALSPAC cohort. Results are presented as standardized variance components

	Model	G	M	Q	E	−2lnl	χ^2	d.f.	P %
Maternal stature	Full	0.15 (0.11)	0.72 (0.11)	−0.08 (0.09)	0.21 (0.12)	19825.526	–	–	–
	Q = 0	0.09 (0.08)	0.66 (0.09)	–	0.25 (0.10)	19826.412	0.886	1	35
	M = Q = 0	0.24 (0.08)	–	–	0.76 (0.09)	19881.294	55.768	2	8×10^{-11}
	G = Q = 0	–	0.68 (0.08)	–	0.32 (0.08)	19827.480	1.954	2	38
	M = G = Q = 0	–	–	–	1	19889.516	63.99	3	8×10^{-12}
Birth length	Full	0.13 (0.13)	0.11 (0.13)	0.06 (0.10)	0.70 (0.14)	3553.870	–	–	–
	Q = 0	0.18 (0.10)	0.16 (0.10)	–	0.66 (0.13)	3554.242	0.37	1	54
	M = Q = 0	0.22 (0.10)	–	–	0.78 (0.10)	3556.624	2.75	2	25
	G = Q = 0	–	0.20 (0.10)	–	0.80 (0.10)	3557.646	3.78	2	15
	M = G = Q = 0	–	–	–	1	3561.774	7.90	3	5

Model parameters are: *G* variance due to direct genetic ("offspring/fetal") effects, *M* variance due to indirect genetic effects on offspring phenotype ("maternal effects"), *Q* phenotypic variance due to covariance of direct and indirect genetic effects, *E* residual ("unique environmental effects")

Trio GCTA and the Future

Behavior Genetics (2021) 51:154–161
https://doi.org/10.1007/s10519-020-10036-6

ORIGINAL RESEARCH

Direct and Indirect Effects of Maternal, Paternal, and Offspring Genotypes: Trio-GCTA

Espen Moen Eilertsen[1,6,9] · Eshim Shahid Jami[2] · Tom A. McAdams[3,6] · Laurie J. Hannigan[4] · Alexandra S. Havdahl[1,4,5,6] · Per Magnus[9] · David M. Evans[5,7] · Eivind Ystrom[1,6,8]

Received: 29 June 2020 / Accepted: 7 December 2020 / Published online: 2 January 2021

Abstract
Indirect genetic effects from relatives may result in misleading quantifications of heritability, but can also be of interest in their own right. In this paper we propose Trio-GCTA, a model for separating direct and indirect genetic effects when genome-wide single nucleotide polymorphism data have been collected from parent-offspring trios. The model is applicable to phenotypes obtained from any of the family members. We discuss appropriate parameter interpretations and apply the method to three exemplar phenotypes: offspring birth weight, maternal relationship satisfaction, and paternal body-mass index, using real data from the Norwegian Mother, Father and Child Cohort Study (MoBa).

Keywords Indirect genetic effects · Within-family · Trio-GCTA · MoBa · Gene–environment correlation

Introduction

Most human traits exhibit some degree of heritability (Polderman et al. 2015). Some phenotypes are characteristics not only of individuals, but also depend on the influence of other individuals. While direct genetic effects refer to how the phenotype of an individual depends on their own genotype, indirect genetic effects refer to how it depends on the genotypes of others (McAdam et al. 2014). In this paper we describe a model for separating direct genetic effects from the indirect genetic effects of family members when genome-wide single nucleotide polymorphism (SNP) data have been collected from parent-offspring trios.

As parents transmit half their complement chromosomes to their children, the genomes of parents and offspring are correlated. Because the same genetic variants can have both direct and indirect effects, failing to account for the indirect genetic effects of relatives when attempting to measure heritability can result in misleading quantifications of the importance of direct genetic effects (Eaves et al. 2014; Young et al. 2019).

Indirect genetic effects can also be of interest in their own right. With respect to the focal individual (i.e., the individual whose phenotype is the focus of study), indirect genetic effects are part of the environment and may be of

Edited by Stacey Cherny.

Electronic supplementary material The online version contains supplementary material available at https://doi.org/10.1007/s10519-020-10036-6.

✉ Espen Moen Eilertsen
e.m.eilertsen@psykologi.uio.no

[1] Department of Psychology, University of Oslo, Forskningsveien 3A, 0373 Oslo, Norway

[2] Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[3] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK

[4] Nic Waals Institute, Lovisenberg Diakonale Hospital, Oslo, Norway

[5] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[6] Department of Psychology, PROMENTA Research Center, University of Oslo, Oslo, Norway

[7] University of Queensland Diamantina Institute, University of Queensland, Brisbane, QLD, Australia

[8] PharmacoEpidemiology and Drug Safety Research Group, School of Pharmacy, University of Oslo, Oslo, Norway

[9] Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway