Genetic Nurture: statistical designs and practical estimation

Perline Demange & Michel G Nivard

5/28/2020

Contents

Genetic Nurture . . . 2
The simulation . . . 2
Simulating the genetic data set . . . 2
Simulating traits **in absence of** any indirect effect . . . 5
Simulating traits **in the presence of** an indirect effect . . . 5
Estimating (in)direct genetic effects . . . 6
Genotyped parents and offpring . . . 6
Model used in Kong et al. . . . 6
Model inspired by Conley et al. . . . 7
Model inspired by Warrington et al. . . . 7
Genotyped siblings . . . 8
Model used in Selzam et al. . . . 8
Genotyped parents of adoptees, or adoptees . . . 9
Model used in Cheesman et al. . . . 9
Model used in Domingue et al. . . . 10
Compare results . . . 10
Review the results in the **absence** of an indirect effect. . . . 10
Review the results in the **presence** of an indirect effect. . . . 11
estimating direct, and indirect, effects in the presence of population stratification. . . . 12
Simulating traits **in absence of** any indirect effect . . . 12
Simulating traits **in the presence of** an indirect effect . . . 13
Estimating (in)direct genetic effects . . . 13
Genotyped parents and offpring . . . 13
Model used in Kong et al. . . . 13
Model inspired by Conley et al. . . . 14
Model inspired by Warrington et al. . . . 15

Genotyped siblings . . . 16
Model used in Selzam et al. . . . 16
Genotyped parents of adoptees, or adoptees . . . 16
Model used in Cheesman et al. . . . 16
Model used in Domingue et al. . . . 17
Compare results . . . 18
Review the results in the **absence** of an indirect effect. . . . 18
Review the results in the **presence** of an indirect effect. . . . 19
Speculative variations, extensions and comparisons. . . . 19
Literature . . . 20

Genetic Nurture

There has been a recent interest in estimating the influence of direct genetic effects (my genome influences my outcomes in life), and indirect genetic effect (the genome of my parents and/or siblings influence my outcomes in life) on complex traits using polygenic risk scores (PRS) and data of first-degree relatives and/or adopted children. The indirect effect is sometimes described as "Genetic Nurture" a term which implies a process active child rearing or "Dynastic effects" a term which implies an economic process where the inter-generational transmission of wealth and power. Because your genome, and that of your parents are correlated, the presence of an indirect genetic effect the presence of gene-environment correlation. There are several ways to estimate the direct genetic effect accounting for the possible presence of an indirect genetic effect.

Here we document existing designs, and designs inspired by previous work which we adapted slightly. We do so by proving and discussing the regression equations used to fit each design and the estimates of the direct and indirect effect.We directly apply each design to simulated data so the reader can follow along. Each of the designs relies on either biological or adopted relatives, and either directly in a model, or by comparison across models, corrects a PRS analysis for the fact that parents may influence the outcome in their child, and in the case of biological children the parent and offspring PRS are correlated.

The simulation

All these designs rely on slightly different familial relations to estimate the direct and indirect effects, so we must simulate the required genotypes in simulated families. We simulate 100 bi-allelic SNPs in linkage equilibrium (i.e. uncorrelated) for 10000 fathers 10000 mothers, we generate a focal child, a sibling, a non-transmitted genotype (not transmitted to the focal child) and an adopted sibling/child.

We also simulate 100 effect sizes, and compute the PRS based on these effect sizes. We assume that the direct and indirect genetic effects are correlated 1, which need not be the case. We assume no assortative mating. In the future we plan to run extra simulations with assortment.

Simulating the genetic data set

We sample true effects for the 100 SNPs on the phenotype from a standard normal distribution. We sample minor allele frequencies from a uniform distribution between .1 and .5. We set the sample size to 10000. Then, we sample SNP 1 for mother, father, and from the mother and father sample a child. I immediately create the non-transmitted genotype (i.e. the genotype that was NOT transmitted to the child). We then sample a sibling, and an adopted sibling. Having created the first SNP, then a loop creates SNP 2 to 100 and append these to the genotype data set. The data set contains for each person (n = 10000) the number of risk alleles (coded 0/1/2) for 100 SNPs.

```r
# Make 100 effect sizes for the effect of 100 SNPs on 1 trait

true_eff <- rnorm(100)

GWAS_eff <- sqrt(.2)*true_eff + sqrt(.8)*rnorm(100)

m.error <- cor(GWAS_eff,true_eff)

# maker true MAF's for 100 SNPs
maf <- runif(100,.1,.5)

# make bi-alleic SNP calls from 100 SNPs in 10000 mothers, 10000 fathers
# and their kids, a non-transmitted PRS and an adopted kid
n <- 10000

# make the first "SNP" for n people:
mothers <-  rbinom(n,size = 2,prob = maf[1])
fathers <-  rbinom(n,size = 2,prob = maf[1])
# make child & non -transmitted allele:
ft <- rbinom(n,size = 1,prob = fathers/2)
mt <- rbinom(n,size = 1,prob = mothers/2)
children <- ft + mt
ntc <- (fathers - ft) + (mothers- mt)

# make some sibs (for sib design)
ft <- rbinom(n,size = 1,prob = fathers/2)
mt <- rbinom(n,size = 1,prob = mothers/2)
sibs <- ft + mt

# Make some adoptees (i.e drawn fully independently from the rest):
adoptees  <- rbinom(n,size = 2,prob = maf[1])

# make SNP 2 to 100:
for(i in 2:100){

mother <- rbinom(n,size = 2,prob = maf[i])
father <- rbinom(n,size = 2,prob = maf[i])
ft <- rbinom(n,size = 1,prob = father/2) # draw 1 allele fromt he father
mt <- rbinom(n,size = 1,prob = mother/2) # draw 1 allele form the mother
child  <- ft + mt # make the child SNP
nt <- (father - ft) + (mother- mt) # make the non-transmitted genotype

mothers <- cbind(mothers,mother) # add SNP to file
fathers <- cbind(fathers,father)
children <- cbind(children,child)
ntc <- cbind(ntc,nt)

# For the sib design, make a SNP for a second child:
ft <- rbinom(n,size = 1,prob = father/2)
mt <- rbinom(n,size = 1,prob = mother/2)
sib  <- ft + mt

sibs <- cbind(sibs,sib) # add the the data
```

```
# For the adoptee design make a SNP for the adoptee:
adoptee  <- rbinom(n,size = 2,prob = maf[i])
adoptees <- cbind(adoptees,adoptee)
}
```

Lets have a look at the first 5 lines of the data set for a single SNP:

```
kable(head(cbind(mother,father,child,nt,sib,adoptee)))
```

mother	father	child	nt	sib	adoptee
0	1	0	1	0	0
1	1	2	0	1	2
2	2	2	2	2	1
2	0	1	1	1	0
1	2	1	2	1	1
0	0	0	0	0	1

We then multiply the risk allele count by the true effect size and sum it to get the genetic liability (sub g or _g), we do the same for the "GWAS" effect size to gat a PRS (sub prs or _prs). We use a matrix algebra to ompute the scores.

```
# multiply the beta and the SNPs and sum to a perfect PRS:
mother_g <- true_eff %*% t(mothers)
father_g<- true_eff %*% t(fathers)
child_g<- true_eff %*% t(children)
ntc_g <-  true_eff %*% t(ntc)
sib_g <- true_eff %*% t(sibs)
adoptee_g <-  true_eff %*% t(adoptees)


#scale genetic scores:
mother_g <- scale(t(mother_g))
father_g <- scale(t(father_g))
child_g <- scale(t(child_g))
ntc_g <- scale(t(ntc_g))
sib_g <- scale(t(sib_g))
adoptee_g <- scale(t(adoptee_g))


# multiply the beta and the SNPs and sum to a perfect PRS:
mother_prs <- GWAS_eff %*% t(mothers)
father_prs<- GWAS_eff %*% t(fathers)
child_prs<- GWAS_eff %*% t(children)
ntc_prs <-  GWAS_eff %*% t(ntc)
sib_prs <- GWAS_eff %*% t(sibs)
adoptee_prs <-  GWAS_eff %*% t(adoptees)


#scale genetic scores:
mother_prs <- scale(t(mother_prs))
```

```
father_prs <- scale(t(father_prs))
child_prs <- scale(t(child_prs))
ntc_prs <- scale(t(ntc_prs))
sib_prs <- scale(t(sib_prs))
adoptee_prs <- scale(t(adoptee_prs))
```

We can check whether the correlations between the various "true" PRS are as we expect based on theory (and they are):

```
# check whether everyone is related to eachother as expected:

cors <-cor(cbind(mother_g,father_g,child_g,ntc_g,sib_g,adoptee_g))
rownames(cors) <-  c("mother","father","child","non-transmitted",
                     "sibling","adopted sib")
colnames(cors) <-  c("mother","father","child","non-transmitted",
                     "sibling","adopted sib")
kable(round(cors,2))
```

	mother	father	child	non-transmitted	sibling	adopted sib
mother	1.00	-0.02	0.50	0.50	0.49	-0.01
father	-0.02	1.00	0.49	0.50	0.50	-0.02
child	0.50	0.49	1.00	0.01	0.50	-0.01
non-transmitted	0.50	0.50	0.01	1.00	0.50	-0.01
sibling	0.49	0.50	0.50	0.50	1.00	-0.01
adopted sib	-0.01	-0.02	-0.01	-0.01	-0.01	1.00

Simulating traits in absence of any indirect effect

We can now (re) use the PRS, add some environmental effect (50/50) and explore different scenario's the first scenario is one where there is only a direct genetic effect and no indirect genetic effects. All designs should retrieve only a direct genetic effect of the simulated magnitude.

```
### No indirect effects:

mother_t1 <- sqrt(.5)*mother_g + sqrt(.5)*rnorm(n)
father_t1 <- sqrt(.5)*father_g + sqrt(.5)*rnorm(n)
child_t1 <- sqrt(.5)*child_g + sqrt(.5)*rnorm(n)
sib_t1 <- sqrt(.5)*sib_g + sqrt(.5)*rnorm(n)
adoptee_t1 <- sqrt(.5)*adoptee_g + sqrt(.5)*rnorm(n)
```

Simulating traits in the presence of an indirect effect

```
### With indirect effects:

mother_t2  <- sqrt(.5)*mother_g  + sqrt(.5)*rnorm(n)
father_t2  <- sqrt(.5)*father_g  + sqrt(.5)*rnorm(n)
child_t2   <- sqrt(.5)*child_g   + sqrt(.1)*mother_t2 + sqrt(.1)*father_t2 + sqrt(.3)*rnorm(n)
sib_t2     <- sqrt(.5)*sib_g     + sqrt(.1)*mother_t2 + sqrt(.1)*father_t2 + sqrt(.3)*rnorm(n)
adoptee_t2 <- sqrt(.5)*adoptee_g + sqrt(.1)*mother_t2 + sqrt(.1)*father_t2 + sqrt(.3)*rnorm(n)
```

Estimating (in)direct genetic effects

We review and discuss a number of design that aim to estimate the direct and indirect genetic effect. We explain the intuition behind the designs, and explicate how they estimate the direct and indirect genetic effects.

The designs can be grouped into 3 broad catagories, first there are those designs that relie on **genotyped parents and their biological offspring**. Two of these, the model propose by **Kong et al. (2018)** and the model inspired by **Conley et al. (2015)** require both parents and the offspring be genotyped, which the offspring must be phenotyped, the third model inspired by innovative work by **Warrington et al. (2018)** can be estimated using far less data: The model can be estiated if a parental genotype, a parental phenotype and offspring phenotype are available. The model can also be estimated when 1 parental and offspring genotypes and offspring phenotype are available. When relying on a single parental genotype, and this genotype is always the maternal or paternal genotype, the assumption must be made the indirect effect is equal for both parents, or absent for the parent that isn't availble. For maternity related variables such as birthweight one could probably savely assume the maternal indirect effect is far bigger than the paternal indirect effect.

The second broad catagory of designs relies on siblings where both siblings are gneotyped and phenotypes. The design, applied to PRS by **Selzam et al. (2018)** leverages the fact that within sibship analysis nutralizes all betwene sibling effect (among which are the effects on parent on their offspring).

The third broad catagory of models relies on the contrast between an adoption relation and a biological relation between parents and offspring. In the design proposed by **Cheesman et al. (2018)** the predictino of a PRS in adopteed and those that are raised by biological relatives are compared whereas in **Domingue et al. (2020)** the relation betwene a parental PRS and the phenotype in either adopted or biological offspirng are contrasted.

Genotyped parents and offpring

Model used in Kong et al.

The design presented in **Kong et al. (2018)** uses genotyped offspring and parents to compute the offspring PRS, and a PRS based on the parental allele's that weren't transmitted to the child. The intuition behind the design is that in the absence of indirect effect, the non-transmitted allele's should not relate to the offspring outcome.

$$phenotype = \beta_t * PRS_t + \beta_{nt} * PRS_{nt} + e$$

In R we can run this regression for both simualted phenotypes, for the phenotype without indirect effects (t1) and the phenotyped with indirect effecs (t2)

```
# perform transmitted non-transmitted PRS analysis:
kong_t1 <- summary(lm(child_t1 ~ child_prs + ntc_prs))
kong_t2 <- summary(lm(child_t2 ~ child_prs + ntc_prs))
```

In the model as defined by **Kong et al.** the estimated direct effect are:

$$direct = \beta_t - \beta_{nt}$$

```
direct_kong_t1 <- kong_t1$coef[2,1] - kong_t1$coef[3,1]
direct_kong_t2 <- kong_t2$coef[2,1] - kong_t2$coef[3,1]
```

$$indirect = \beta_{nt}$$

```
indirect_kong_t1 <-kong_t1$coef[3,1]
indirect_kong_t2 <- kong_t2$coef[3,1]
```

Model inspired by Conley et al.

A design discussed and applied in **Conley et al. (2015)** is to condition a phenotype on the child's PRS and the parental PRS (one parent or both). We take the liberlty to extend on it a bit, and include both maternal and paternal PRS.

$$phenotype_{child} = \beta_{child} * PRS_{child} + \beta_{mother} * PRS_{mother} + \beta_{father} * PRS_{father} + e$$

We can estimate the model using the following code:

```
# perform transmitted non-transmitted PRS analysis:
conley_t1 <- summary(lm(child_t1 ~ child_prs + father_prs + mother_prs))
conley_t2 <- summary(lm(child_t2 ~ child_prs + father_prs + mother_prs))
```

the direct effect and indirect effect are define as:

$$direct = \beta_{child}$$

```
direct_conley_t1 <- conley_t1$coef[2,1]
direct_conley_t2 <- conley_t2$coef[2,1]
```

$$indirect = .5 * (\beta_{mother} + \beta_{father})$$

```
indirect_conley_t1 <- .5*(conley_t1$coef[3,1] + conley_t1$coef[4,1])
indirect_conley_t2 <- .5*(conley_t2$coef[3,1] + conley_t2$coef[4,1])
```

Model inspired by Warrington et al.

In the context of GWAS a close cousin of this design has been appied to Birthweight revealing specific maternal and offspring effects on birthweight (**Warrington et al., 2018**). Here we restrict ourself to discussing the details of PRS analysis, but all design we discuss could be applied per locus in a GWAS. What we do adapt from **Warrington et al.** is an application where, given certain assumptions, we can estimate the direct and indirect effect of a PRS on an outcome if we abserve a parental PRS, the parental phenotype and the offspring phenotype but no offspring PRS. The design can also be applied when we observe parental and offspring PRS, and only offspring phenotype.

The Warrington model requires 2 seperate regressions be performed:

$$phenotype_{child} = \beta_{offspring} * PRS_{mother} + e$$

and:

$$phenotype_{mother} = \beta_{own} * PRS_{mother} + e$$

OR

$$phenotype_{child} = \beta_{own} * PRS_{child} + e$$

```
warrington_t1_off <- summary(lm(child_t1 ~ mother_prs))
warrington_t1_own <- summary(lm(child_t1 ~ child_prs))

warrington_t2_off <- summary(lm(child_t2 ~ mother_prs))
warrington_t2_own <- summary(lm(child_t2 ~ child_prs))
```

the direct effect and indirect effect are defined as:

$$direct = beta_{own} - (2 * \beta_{offspring} - \beta_{own})$$

```
direct_warrington_t1 <- (warrington_t1_own$coef[2,1] -
                         (2*warrington_t1_off$coef[2,1] - warrington_t1_own$coef[2,1]))
direct_warrington_t2 <- (warrington_t2_own$coef[2,1] -
                         (2*warrington_t2_off$coef[2,1] - warrington_t2_own$coef[2,1]))
```

$$indirect = (2 * \beta_{offspring} - \beta_{own})$$

```
indirect_warrington_t1 <-  (2*warrington_t1_off$coef[2,1] - warrington_t1_own$coef[2,1])
indirect_warrington_t2 <-  (2*warrington_t2_off$coef[2,1] - warrington_t2_own$coef[2,1])
```

Genotyped siblings

Model used in Selzam et al.

The design used in **Selzam et al. (2019)** relies on genotyped sibling pairs and involves regressing the difference in sibling phenotype on the mean sibling PRS (between family effect) and the per sibling deviance of the mean family effect (within family effect). The intuition is that in the absence of an indirect genetic effect, raising a person's genetic liability by 1 risk allele increases the outcome by a fixed amount regardless of whether this person is compared to his or her sibling or to the general population, whereas an indirect (Parental) effect influences both siblings regardless of their genotype at the locus.

$$between = .5 * (PRS_{child} + PRS_{sibling})$$

$$within = PRS_{child} - between$$

$$phenotype_{child} = \beta_{within} * within + \beta_{between} * between + u + e$$

```
# perform sib PRS analysis:
between <- .5*(child_prs + sib_prs)
within <- child_prs - between

selzam_t1 <- summary(lm(child_t1 ~ between + within))
selzam_t2 <- summary(lm(child_t2 ~ between + within))
```

The estimated direct and indirect effect are:

$$direct = \beta_{within}$$

```
direct_selzam_t1 <- selzam_t1$coef[3,1]
direct_selzam_t2 <- selzam_t2$coef[3,1]
```

$$indirect = .75 * (\beta_{between} - \beta_{within})$$

```
indirect_selzam_t1 <- .75*(selzam_t1$coef[2,1]-selzam_t1$coef[3,1])
indirect_selzam_t2 <- .75*(selzam_t2$coef[2,1]-selzam_t2$coef[3,1])
```

Genotyped parents of adoptees, or adoptees

Model used in Cheesman et al.

The design used by **Cheesman et al. (2020)** relies on a comparison between PRS prediction between adoptees and biological offspring. The intuition is the following: In the presence of an indirect genetic effect after birth both adoptees and biological offspring are influenced by the parent that raises them. This induces a correlation between PRS of the rearing parent and either adopted or biological offspring of that parent, however since the PRS of adoptee and rearing parent are uncorrelated, a regression of the adoptee phenotype on the adoptee PRS gives us an unconfounded estimate of the direct effect. A regression of the phenotype of biological offspring on their PRS gives us an estimate that consists of the direct, and a part of the indirect genetic effect.

$$phenotype_{adoptee} = beta_{adoptee} * PRS_{adoptee} + e$$

$$phenotype_{biological-child} = beta_{biological-child} * PRS_{biological-child} + e$$

```
# perform adoption PRS analysis:
cheesman_t1_adopt <- summary(lm(adoptee_t1 ~ adoptee_prs ))
cheesman_t1_bio <- summary(lm(child_t1 ~ child_prs ))

cheesman_t2_adopt <- summary(lm(adoptee_t2 ~ adoptee_prs ))
cheesman_t2_bio <- summary(lm(child_t2 ~ child_prs ))
```

And the direct and indirect effects are defined as:

$$direct = \beta_{adoptee}$$

```
direct_cheesman_t1 <- cheesman_t1_adopt$coef[2,1]
direct_cheesman_t2 <- cheesman_t2_adopt$coef[2,1]
```

$$indirect = (beta_{biological-child} - \beta_{adoptee})$$

```
indirect_cheesman_t1 <- cheesman_t1_bio$coef[2,1] - cheesman_t1_adopt$coef[2,1]
indirect_cheesman_t2 <- cheesman_t2_bio$coef[2,1] - cheesman_t2_adopt$coef[2,1]
```

Model used in Domingue et al.

A design used by **Domingue et al. (2020)** is a slight variation of the adoption design where the phenotype of adopted and biological offspring are regressed on that rearing parents genotype. This design again leverages the fact that parent and biological offspring have correlated PRS whereas parent and adoption offspring have uncorrelated PRS.

$$phenotype_{adoptee} = beta_{adoption} * PRS_{rearing-parent} + e$$

$$phenotype_{biological-child} = beta_{biological} * PRS_{rearing-parent} + e$$

```
# perform adoption PRS analysis:
domingue_t1_adop_offs <- summary(lm(adoptee_t1 ~ father_prs ))
domingue_t1_bio_offs <- summary(lm(child_t1 ~ father_prs ))

domingue_t2_adop_offs <- summary(lm(adoptee_t2 ~ father_prs ))
domingue_t2_bio_offs <- summary(lm(child_t2 ~ father_prs ))
```

And the direct and indirect effects are defined as:

$$direct = 2 * (beta_{biological} - \beta_{adoption})$$

```
direct_domingue_t1 <- 2*(domingue_t1_bio_offs$coef[2,1]-domingue_t1_adop_offs$coef[2,1])
direct_domingue_t2 <- 2*(domingue_t2_bio_offs$coef[2,1]-domingue_t2_adop_offs$coef[2,1])
```

$$indirect = (\beta_{adoption})$$

```
indirect_domingue_t1 <- domingue_t1_adop_offs$coef[2,1]
indirect_domingue_t2 <- domingue_t2_adop_offs$coef[2,1]
```

Compare results

Review the results in the absence of an indirect effect.

We cooelct all direct and indirect effect estimates, and the true effect as specified in the simulation, in a single object so we can esily compare the results.

```
direct_effects <- c(m.error*sqrt(.5),
                    direct_kong_t1,
                    direct_conley_t1,
                    direct_warrington_t1,
                    direct_selzam_t1,
                    direct_cheesman_t1,
                    direct_domingue_t1
                    )
```

```
indirect_effects <- c(0,
                      indirect_kong_t1,
                      indirect_conley_t1,
                      indirect_warrington_t1,
                      indirect_selzam_t1,
                      indirect_cheesman_t1,
                      indirect_domingue_t1
                      )
```

Lets compare the results:

```
tab <-  cbind(round(direct_effects,3),
              round(indirect_effects,3))
rownames(tab) <- c("Truth","Kong et al.","Conley et al.","Warrington et al.",
                   "Selzam et al.","Cheesman et al.","Domingue & Fletcher")
  colnames(tab) <-  c("Direct effect", "Indirect Effect")

kable(tab)
```

	Direct effect	Indirect Effect
Truth	0.353	0.000
Kong et al.	0.345	-0.005
Conley et al.	0.345	-0.005
Warrington et al.	0.354	-0.014
Selzam et al.	0.359	-0.020
Cheesman et al.	0.333	0.007
Domingue & Fletcher	0.354	-0.008

Review the results in the presence of an indirect effect.

We cooelct all direct and indirect effect estimates, and the true effect as specified in the simulation, in a single object so we can esily compare the results.

```
direct_effects <- c(m.error*sqrt(.5),
                    direct_kong_t2,
                    direct_conley_t2,
                    direct_warrington_t2,
                    direct_selzam_t2,
                    direct_cheesman_t2,
                    direct_domingue_t2
                    )
```

Then I compute the indirect effects some authors are explicit about how this is done, others aren't in which case I determined how to compute it myself:

```
indirect_effects <- c(m.error*sqrt(.1)*sqrt(.5),
                      indirect_kong_t2,
                      indirect_conley_t2,
                      indirect_warrington_t2,
                      indirect_selzam_t2,
```

```
                    indirect_cheesman_t2,
                    indirect_domingue_t2
                    )
```

Lets compare the results:

```
tab <-  cbind(round(direct_effects,3),
           round(indirect_effects,3))
rownames(tab) <- c("Truth","Kong et al.","Conley et al.","Warrington et al.",
                   "Selzam et al.","Cheesman et al.","Domingue & Fletcher")
  colnames(tab) <-  c("Direct effect", "Indirect Effect")

kable(tab)
```

	Direct effect	Indirect Effect
Truth	0.353	0.112
Kong et al.	0.341	0.106
Conley et al.	0.341	0.105
Warrington et al.	0.352	0.093
Selzam et al.	0.351	0.095
Cheesman et al.	0.325	0.120
Domingue & Fletcher	0.332	0.113

estimating direct, and indirect, effects in the presence of population stratification.

WORK IN PROGRESS

Simulating traits in absence of any indirect effect

We can now (re) use the PRS, add some environmental effect (50/50) and explore different scenario's the first scenario is one where there is only a direct genetic effect and no indirect genetic effects. All designs should retrieve only a direct genetic effect of the simulated magnitude.

```
### Make the residual vairance correlate to the genetic liability:

mother_e <- sqrt(n)
father_e <- sqrt(n)
child_e <- sqrt(.2) * mother_e + sqrt(.2) * father_e + sqrt(.6) * sqrt(n)
sib_e <-   sqrt(.2) * mother_e + sqrt(.2) * father_e + sqrt(.6) * sqrt(n)
adoptee_e <- sqrt(.2) * mother_e + sqrt(.2) * father_e + sqrt(.6) * sqrt(n)

### No indirect effects:

mother_t1 <- sqrt(.5)*mother_g + sqrt(.5)*mother_e
father_t1 <- sqrt(.5)*father_g + sqrt(.5)*father_e
child_t1 <- sqrt(.5)*child_g + sqrt(.5)*child_e
sib_t1 <- sqrt(.5)*sib_g + sqrt(.5)*sib_e
adoptee_t1 <- sqrt(.5)*adoptee_g + sqrt(.5)*adoptee_e
```

Simulating traits in the presence of an indirect effect

```
### With indirect effects:




mother_t2  <- sqrt(.5)*mother_g  + sqrt(.5)*rnorm(n)
father_t2  <- sqrt(.5)*father_g  + sqrt(.5)*rnorm(n)
child_t2   <- sqrt(.5)*child_g   + sqrt(.1)*mother_t2 + sqrt(.1)*father_t2 + sqrt(.3)*rnorm(n)
sib_t2     <- sqrt(.5)*sib_g     + sqrt(.1)*mother_t2 + sqrt(.1)*father_t2 + sqrt(.3)*rnorm(n)
adoptee_t2 <- sqrt(.5)*adoptee_g + sqrt(.1)*mother_t2 + sqrt(.1)*father_t2 + sqrt(.3)*rnorm(n)
```

Estimating (in)direct genetic effects

We review and discuss a number of design that aim to estimate the direct and indirect genetic effect. We explain the intuition behind the designs, and explicate how they estimate the direct and indirect genetic effects.

The designs can be grouped into 3 broad catagories, first there are those designs that relie on **genotyped parents and their biological offspring**. Two of these, the model propose by **Kong et al. (2018)** and the model inspired by **Conley et al. (2015)** require both parents and the offspring be genotyped, which the offspring must be phenotyped, the third model inspired by innovative work by **Warrington et al. (2018)** can be estimated using far less data: The model can be estiated if a parental genotype, a parental phenotype and offspring phenotype are available. The model can also be estimated when 1 parental and offspring genotypes and offspring phenotype are available. When relying on a single parental genotype, and this genotype is always the maternal or paternal genotype, the assumption must be made the indirect effect is equal for both parents, or absent for the parent that isn't availble. For maternity related variables such as birthweight one could probably savely assume the maternal indirect effect is far bigger than the paternal indirect effect.

The second broad catagory of designs relies on siblings where both siblings are gneotyped and phenotypes. The design, applied to PRS by **Selzam et al. (2018)** leverages the fact that within sibship analysis nutralizes all betwene sibling effect (among which are the effects on parent on their offspring).

The third broad catagory of models relies on the contrast between an adoption relation and a biological relation between parents and offspring. In the design proposed by **Cheesman et al. (2018)** the predictino of a PRS in adopteed and those that are raised by biological relatives are compared whereas in **Domingue et al. (2020)** the relation betwene a parental PRS and the phenotype in either adopted or biological offspirng are contrasted.

Genotyped parents and offpring

Model used in Kong et al.

The design presented in **Kong et al. (2018)** uses genotyped offspring and parents to compute the offspring PRS, and a PRS based on the parental allele's that weren't transmitted to the child. The intuition behind the design is that in the absence of indirect effect, the non-transmitted allele's should not relate to the offspring outcome.

$$phenotype = \beta_t * PRS_t + \beta_{nt} * PRS_{nt} + e$$

In R we can run this regression for both simualted phenotypes, for the phenotype without indirect effects (t1) and the phenotyped with indirect effecs (t2)

```
# perform transmitted non-transmitted PRS analysis:
kong_t1 <- summary(lm(child_t1 ~ child_prs + ntc_prs))
kong_t2 <- summary(lm(child_t2 ~ child_prs + ntc_prs))
```

In the model as defined by **Kong et al.** the estimated direct effect are:

$$direct = \beta_t - \beta_{nt}$$

```
direct_kong_t1 <- kong_t1$coef[2,1] - kong_t1$coef[3,1]
direct_kong_t2 <- kong_t2$coef[2,1] - kong_t2$coef[3,1]
```

$$indirect = \beta_{nt}$$

```
indirect_kong_t1 <-kong_t1$coef[3,1]
indirect_kong_t2 <- kong_t2$coef[3,1]
```

Model inspired by Conley et al.

A design discussed and applied in **Conley et al. (2015)** is to condition a phenotype on the child's PRS and the parental PRS (one parent or both). We take the liberlty to extend on it a bit, and include both maternal and paternal PRS.

$$phenotype_{child} = \beta_{child} * PRS_{child} + \beta_{mother} * PRS_{mother} + \beta_{father} * PRS_{father} + e$$

We can estimate the model using the following code:

```
# perform transmitted non-transmitted PRS analysis:
conley_t1 <- summary(lm(child_t1 ~ child_prs + father_prs + mother_prs))
conley_t2 <- summary(lm(child_t2 ~ child_prs + father_prs + mother_prs))
```

the direct effect and indirect effect are define as:

$$direct = \beta_{child}$$

```
direct_conley_t1 <- conley_t1$coef[2,1]
direct_conley_t2 <- conley_t2$coef[2,1]
```

$$indirect = .5 * (\beta_{mother} + \beta_{father})$$

```
indirect_conley_t1 <- .5*(conley_t1$coef[3,1] + conley_t1$coef[4,1])
indirect_conley_t2 <- .5*(conley_t2$coef[3,1] + conley_t2$coef[4,1])
```

Model inspired by Warrington et al.

In the context of GWAS a close cousin of this design has been appied to Birthweight revealing specific maternal and offspring effects on birthweight (**Warrington et al., 2018**). Here we restrict ourself to discussing the details of PRS analysis, but all design we discuss could be applied per locus in a GWAS. What we do adapt from **Warrington et al.** is an application where, given certain assumptions, we can estimate the direct and indirect effect of a PRS on an outcome if we aboserve a parental PRS, the parental phenotype and the offspring phenotype but no offspring PRS. The design can also be applied when we observe parental and offspring PRS, and only offspring phenotype.

The Warrington model requires 2 seperate regressions be performed:

$$phenotype_{child} = \beta_{offspring} * PRS_{mother} + e$$

and:

$$phenotype_{mother} = \beta_{own} * PRS_{mother} + e$$

OR

$$phenotype_{child} = \beta_{own} * PRS_{child} + e$$

```
warrington_t1_off <- summary(lm(child_t1 ~ mother_prs))
warrington_t1_own <- summary(lm(child_t1 ~ child_prs))

warrington_t2_off <- summary(lm(child_t2 ~ mother_prs))
warrington_t2_own <- summary(lm(child_t2 ~ child_prs))
```

the direct effect and indirect effect are defined as:

$$direct = beta_{own} - (2 * \beta_{offspring} - \beta_{own})$$

```
direct_warrington_t1 <- (warrington_t1_own$coef[2,1] -
                         (2*warrington_t1_off$coef[2,1] - warrington_t1_own$coef[2,1]))
direct_warrington_t2 <- (warrington_t2_own$coef[2,1] -
                         (2*warrington_t2_off$coef[2,1] - warrington_t2_own$coef[2,1]))
```

$$indirect = (2 * \beta_{offspring} - \beta_{own})$$

```
indirect_warrington_t1 <-  (2*warrington_t1_off$coef[2,1] - warrington_t1_own$coef[2,1])
indirect_warrington_t2 <-  (2*warrington_t2_off$coef[2,1] - warrington_t2_own$coef[2,1])
```

Genotyped siblings

Model used in Selzam et al.

The design used in **Selzam et al. (2019)** relies on genotyped sibling pairs and involves regressing the difference in sibling phenotype on the mean sibling PRS (between family effect) and the per sibling deviance of the mean family effect (within family effect). The intuition is that in the absence of an indirect genetic effect, raising a person's genetic liability by 1 risk allele increases the outcome by a fixed amount regardless of whether this person is compared to his or her sibling or to the general population, whereas an indirect (Parental) effect influences both siblings regardless of their genotype at the locus.

$$between = .5 * (PRS_{child} + PRS_{sibling})$$

$$within = PRS_{child} - between$$

$$phenotype_{child} = \beta_{within} * within + \beta_{between} * between + u + e$$

```
# perform sib PRS analysis:
between <- .5*(child_prs + sib_prs)
within <- child_prs - between

selzam_t1 <- summary(lm(child_t1 ~ between + within))
selzam_t2 <- summary(lm(child_t2 ~ between + within))
```

The estimated direct and indirect effect are:

$$direct = \beta_{within}$$

```
direct_selzam_t1 <- selzam_t1$coef[3,1]
direct_selzam_t2 <- selzam_t2$coef[3,1]
```

$$indirect = .75 * (\beta_{between} - \beta_{within})$$

```
indirect_selzam_t1 <- .75*(selzam_t1$coef[2,1]-selzam_t1$coef[3,1])
indirect_selzam_t2 <- .75*(selzam_t2$coef[2,1]-selzam_t2$coef[3,1])
```

Genotyped parents of adoptees, or adoptees

Model used in Cheesman et al.

The design used by **Cheesman et al. (2020)** relies on a comparison between PRS prediction between adoptees and biological offspring. The intuition is the following: In the presence of an indirect genetic effect after birth both adoptees and biological offspring are influenced by the parent that raises them. This induces a correlation between PRS of the rearing parent and either adopted or biological offspring of that parent, however since the PRS of adoptee and rearing parent are uncorrelated, a regression of the adoptee phenotype on the adoptee PRS gives us an unconfounded estimate of the direct effect. A regression of the phenotype of biological offspring on their PRS gives us an estimate that consists of the direct, and a part of the indirect genetic effect.

$$phenotype_{adoptee} = beta_{adoptee} * PRS_{adoptee} + e$$

$$phenotype_{biological-child} = beta_{biological-child} * PRS_{biological-child} + e$$

```
# perform adoption PRS analysis:
cheesman_t1_adopt <- summary(lm(adoptee_t1 ~ adoptee_prs ))
cheesman_t1_bio <- summary(lm(child_t1 ~ child_prs ))

cheesman_t2_adopt <- summary(lm(adoptee_t2 ~ adoptee_prs ))
cheesman_t2_bio <- summary(lm(child_t2 ~ child_prs ))
```

And the direct and indirect effects are defined as:

$$direct = \beta_{adoptee}$$

```
direct_cheesman_t1 <- cheesman_t1_adopt$coef[2,1]
direct_cheesman_t2 <- cheesman_t2_adopt$coef[2,1]
```

$$indirect = (beta_{biological-child} - \beta_{adoptee})$$

```
indirect_cheesman_t1 <- cheesman_t1_bio$coef[2,1] - cheesman_t1_adopt$coef[2,1]
indirect_cheesman_t2 <- cheesman_t2_bio$coef[2,1] - cheesman_t2_adopt$coef[2,1]
```

Model used in Domingue et al.

A design used by **Domingue et al. (2020)** is a slight variation of the adoption design where the phenotype of adopted and biological offspring are regressed on that rearing parents genotype. This design again leverages the fact that parent and biological offspring have correlated PRS whereas parent and adoption offspring have uncorrelated PRS.

$$phenotype_{adoptee} = beta_{adoption} * PRS_{rearing-parent} + e$$

$$phenotype_{biological-child} = beta_{biological} * PRS_{rearing-parent} + e$$

```
# perform adoption PRS analysis:
domingue_t1_adop_offs <- summary(lm(adoptee_t1 ~ father_prs ))
domingue_t1_bio_offs <- summary(lm(child_t1 ~ father_prs ))

domingue_t2_adop_offs <- summary(lm(adoptee_t2 ~ father_prs ))
domingue_t2_bio_offs <- summary(lm(child_t2 ~ father_prs ))
```

And the direct and indirect effects are defined as:

$$direct = 2 * (beta_{biological} - \beta_{adoption})$$

```
direct_domingue_t1 <- 2*(domingue_t1_bio_offs$coef[2,1]-domingue_t1_adop_offs$coef[2,1])
direct_domingue_t2 <- 2*(domingue_t2_bio_offs$coef[2,1]-domingue_t2_adop_offs$coef[2,1])
```

$$indirect = (\beta_{adoption})$$

```
indirect_domingue_t1 <- domingue_t1_adop_offs$coef[2,1]
indirect_domingue_t2 <- domingue_t2_adop_offs$coef[2,1]
```

Compare results

Review the results in the absence of an indirect effect.

We cooelct all direct and indirect effect estimates, and the true effect as specified in the simulation, in a single object so we can esily compare the results.

```
direct_effects <- c(m.error*sqrt(.5),
                    direct_kong_t1,
                    direct_conley_t1,
                    direct_warrington_t1,
                    direct_selzam_t1,
                    direct_cheesman_t1,
                    direct_domingue_t1
                    )
```

```
indirect_effects <- c(0,
                    indirect_kong_t1,
                    indirect_conley_t1,
                    indirect_warrington_t1,
                    indirect_selzam_t1,
                    indirect_cheesman_t1,
                    indirect_domingue_t1
                    )
```

Lets compare the results:

```
tab <-  cbind(round(direct_effects,3),
            round(indirect_effects,3))
rownames(tab) <- c("Truth","Kong et al.","Conley et al.","Warrington et al.",
                   "Selzam et al.","Cheesman et al.","Domingue & Fletcher")
  colnames(tab) <-  c("Direct effect", "Indirect Effect")

kable(tab)
```

	Direct effect	Indirect Effect
Truth	0.353	0.000
Kong et al.	0.339	0.004
Conley et al.	0.339	0.004
Warrington et al.	0.344	-0.001
Selzam et al.	0.345	-0.003
Cheesman et al.	0.335	0.008
Domingue & Fletcher	0.349	-0.001

Review the results in the presence of an indirect effect.

We cooelct all direct and indirect effect estimates, and the true effect as specified in the simulation, in a single object so we can esily compare the results.

```
direct_effects <- c(m.error*sqrt(.5),
                    direct_kong_t2,
                    direct_conley_t2,
                    direct_warrington_t2,
                    direct_selzam_t2,
                    direct_cheesman_t2,
                    direct_domingue_t2
                    )
```

Then I compute the indirect effects some authors are explicit about how this is done, others aren't in which case I determined how to compute it myself:

```
indirect_effects <- c(m.error*sqrt(.1)*sqrt(.5),
                    indirect_kong_t2,
                    indirect_conley_t2,
                    indirect_warrington_t2,
                    indirect_selzam_t2,
                    indirect_cheesman_t2,
                    indirect_domingue_t2
                    )
```

Lets compare the results:

```
tab <-  cbind(round(direct_effects,3),
            round(indirect_effects,3))
rownames(tab) <- c("Truth","Kong et al.","Conley et al.","Warrington et al.",
                   "Selzam et al.","Cheesman et al.","Domingue & Fletcher")
  colnames(tab) <-  c("Direct effect", "Indirect Effect")

kable(tab)
```

	Direct effect	Indirect Effect
Truth	0.353	0.112
Kong et al.	0.341	0.106
Conley et al.	0.341	0.106
Warrington et al.	0.358	0.087
Selzam et al.	0.357	0.089
Cheesman et al.	0.332	0.113
Domingue & Fletcher	0.353	0.106

Speculative variations, extensions and comparisons.

The equivalence between the models discussed above can break down in interesting ways, ways that are informative about the developmental process. For example the adoption design as applied by **Cheeesman et al.** assumes the indirect effect plays out after birth, if there is a maternal influence on a child's outcome

during pregnancy (as is the case for birth-weight for example) this indirect effect is not accounted for by the adoption design. This also means that in the presence of a prenatal effect, the results form the adoption and the other designs, will begin to diverge. This divergence could be used to detect the presence of prenatal indirect effects. It is likely these types of analyses could require rather large samples, and preferably the data for multiple designs be collected in a uniform fashion and in a single population to ensure any subtle differences population structure in measurement of genotype or phenotype doesn't confound the estimate of a prenatal indirect effect.

There are other variations on the adoption design one could consider, such as analysis of PRS effects on people reared by one biological and one adopted parent (in case of sperm or egg donation, or non-paternity), or people reared by one or two non-biological parent from a certain age onward (in case of rearing by stepparent(s)). Again this section is specifically for speculation, and numerous practical and theoretical issues would need to be considered. One of the mayor issue to consider (either when applying the adoption design, or when considering variations of it) is the sample selection that occurs when considering populations that differ in more ways then one, adoption, and the other processes that we discussed aren't random processes, and both adoptive parents, and adoptees may differ from the general population. This selection could influence results in unexpected ways.

Literature

Cheesman, R. et al. (2020). Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. *Psych Science*. 31(5) 582-591

Conley, D. et al. (2015). Is the effect of parental education on offspring biased or moderated by genotype?. *Sociological Science*, 2, 82.

Domingue, B. W., & Fletcher, J. (2020). Separating Measured Genetic and Environmental Effects: Evidence Linking Parental Genotype and Adopted Child Outcomes. *Behaviour Genetics* (AOP).

Kong, A. et al. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424-428.

Selzam, S. et al. (2019). Comparing within-and between-family polygenic score prediction. *The American Journal of Human Genetics*, 105(2), 351-363.

Warrington, N.M. et al. (2018). Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. *International journal of epidemiology*, 47(4), 1229-1241.