Interactions!

GxG, GxE, MxE, PoOxE, GxMe, PoOxMe

GxG = Gene x Gene interaction

GxE = Gene x Environment interaction

MxE = Maternal genes x Environment interaction

PoOxE = Parent-of-origin x Environment interaction

GxMe = Gene x Methylation interaction

PoOxMe = Parent-of-origin x Methylation interaction

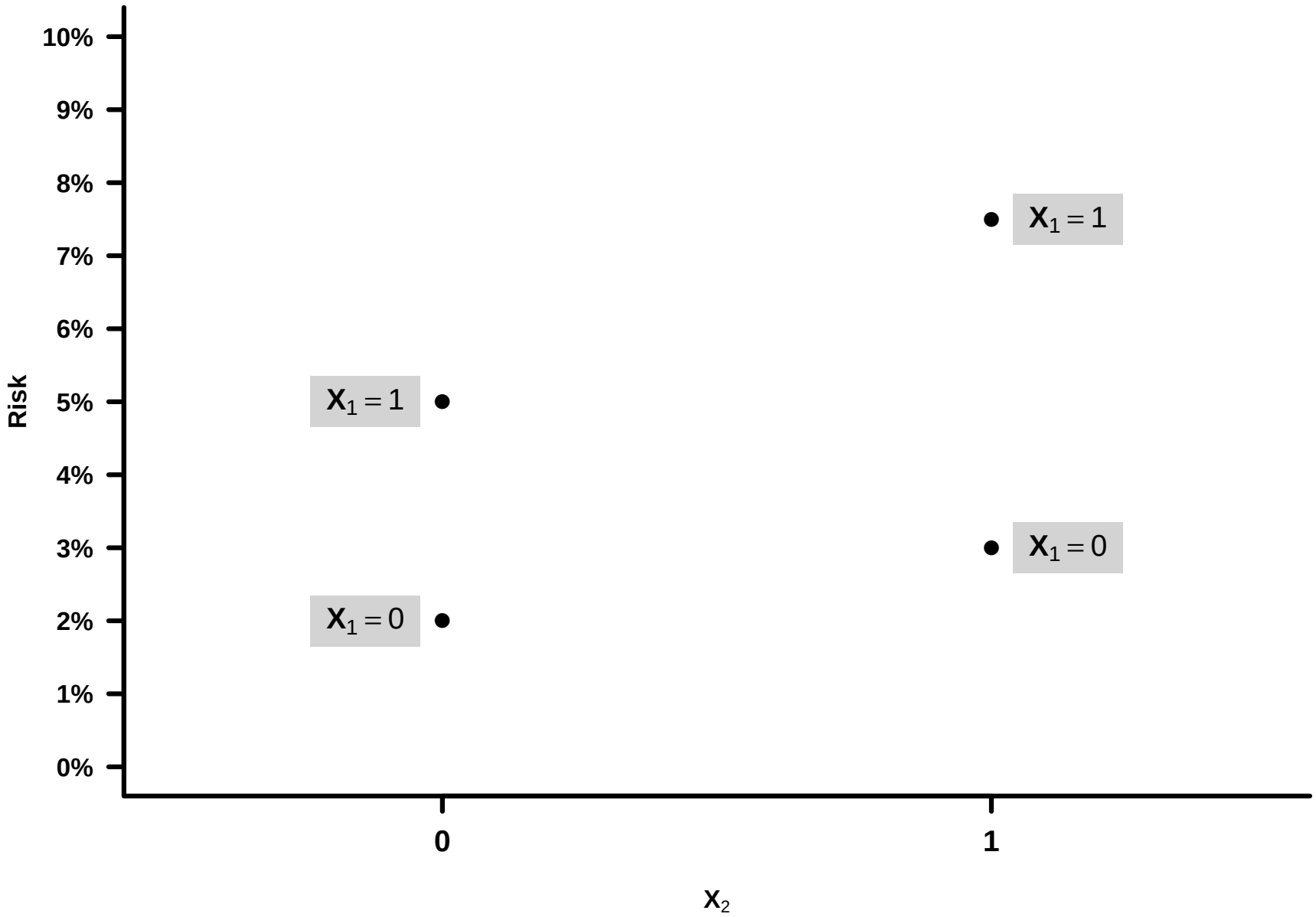
INTERACTIONS

- Also known as “Effect modifications”... but is that the same?
- In genetics, often referred to as epistasis
- Consider the effect of X_1 on Y
- Does the effect change over levels of another variable X_2?

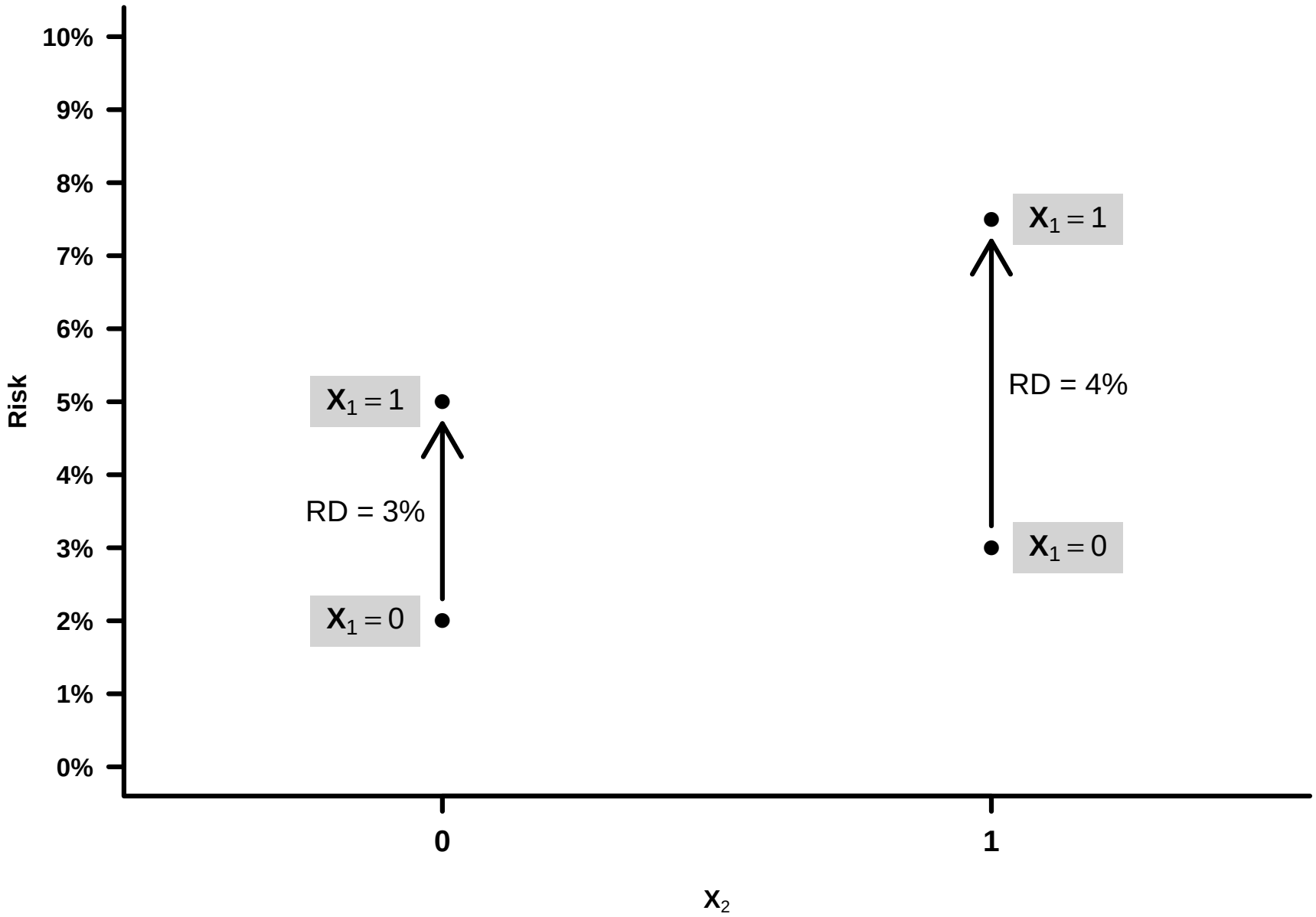
NOTE: X_1 and X_2 can be, for instance:

- X_1: SNP and X_2: SNP
- X_1: SNP and X_2: environmental exposure
- X_1: environmental exposure and X_2: environmental exposure

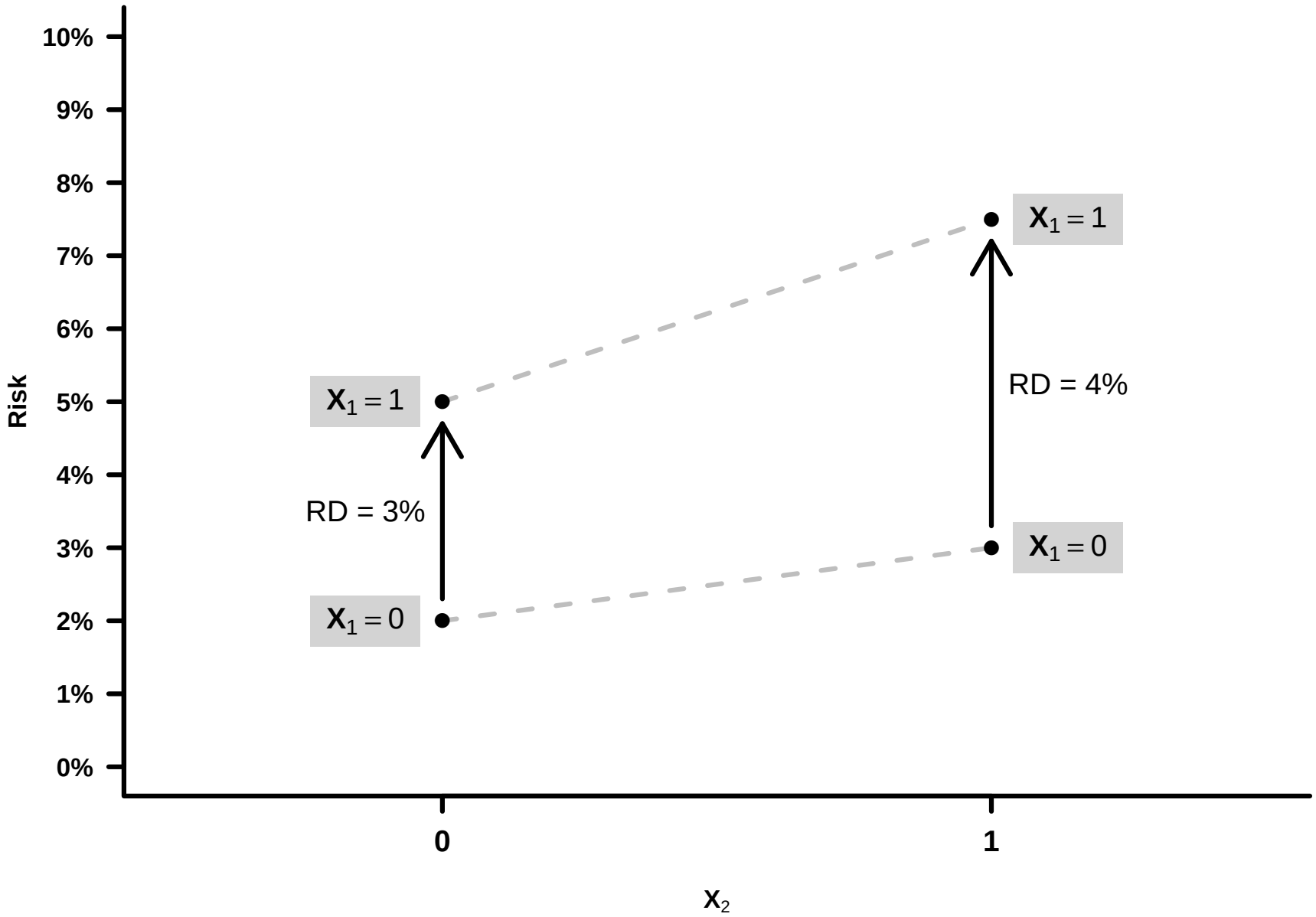
INTERACTIONS, RISK DIFFERENCE (RD)



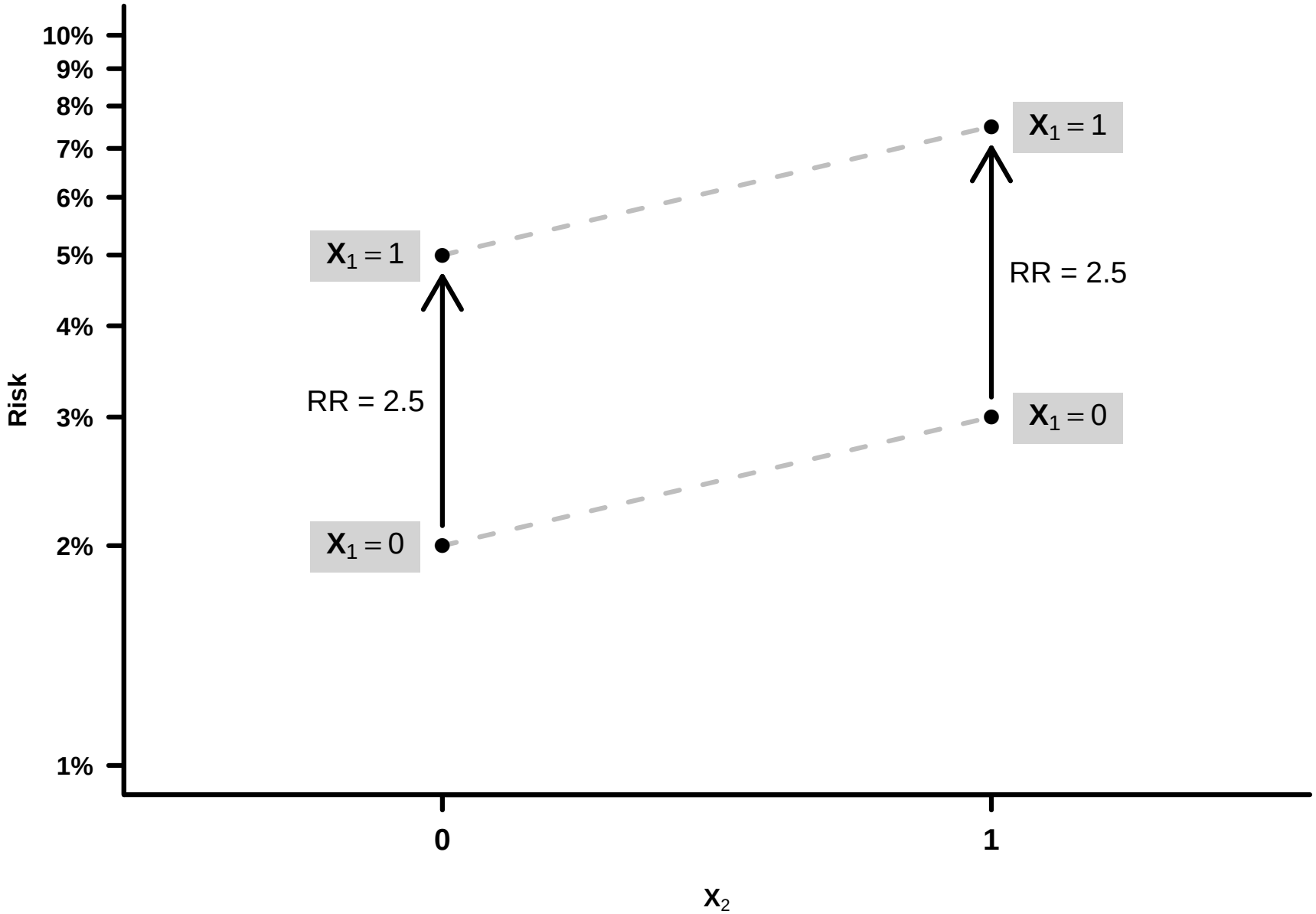
INTERACTIONS, RISK DIFFERENCE (RD)



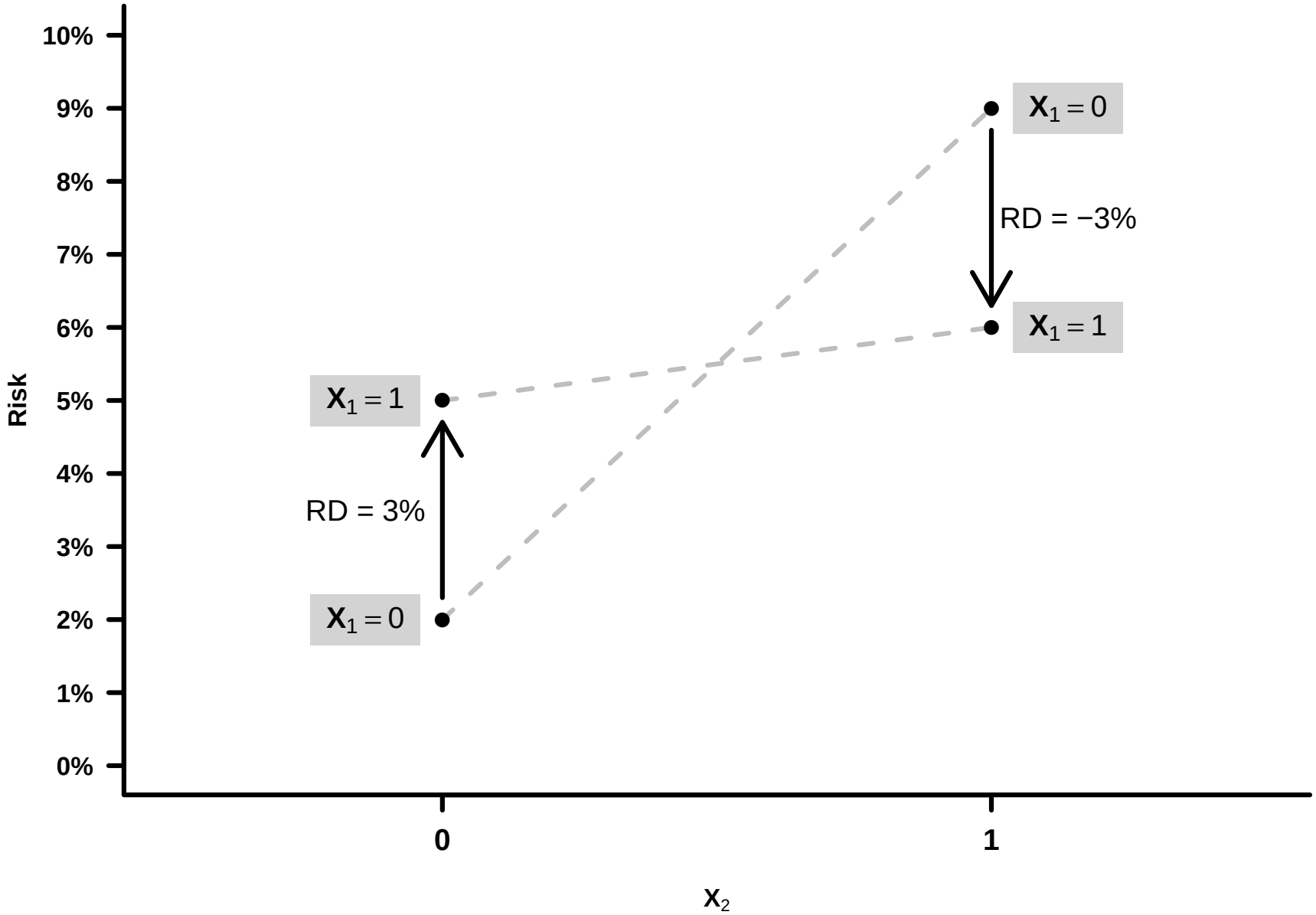
INTERACTIONS, RISK DIFFERENCE (RD)



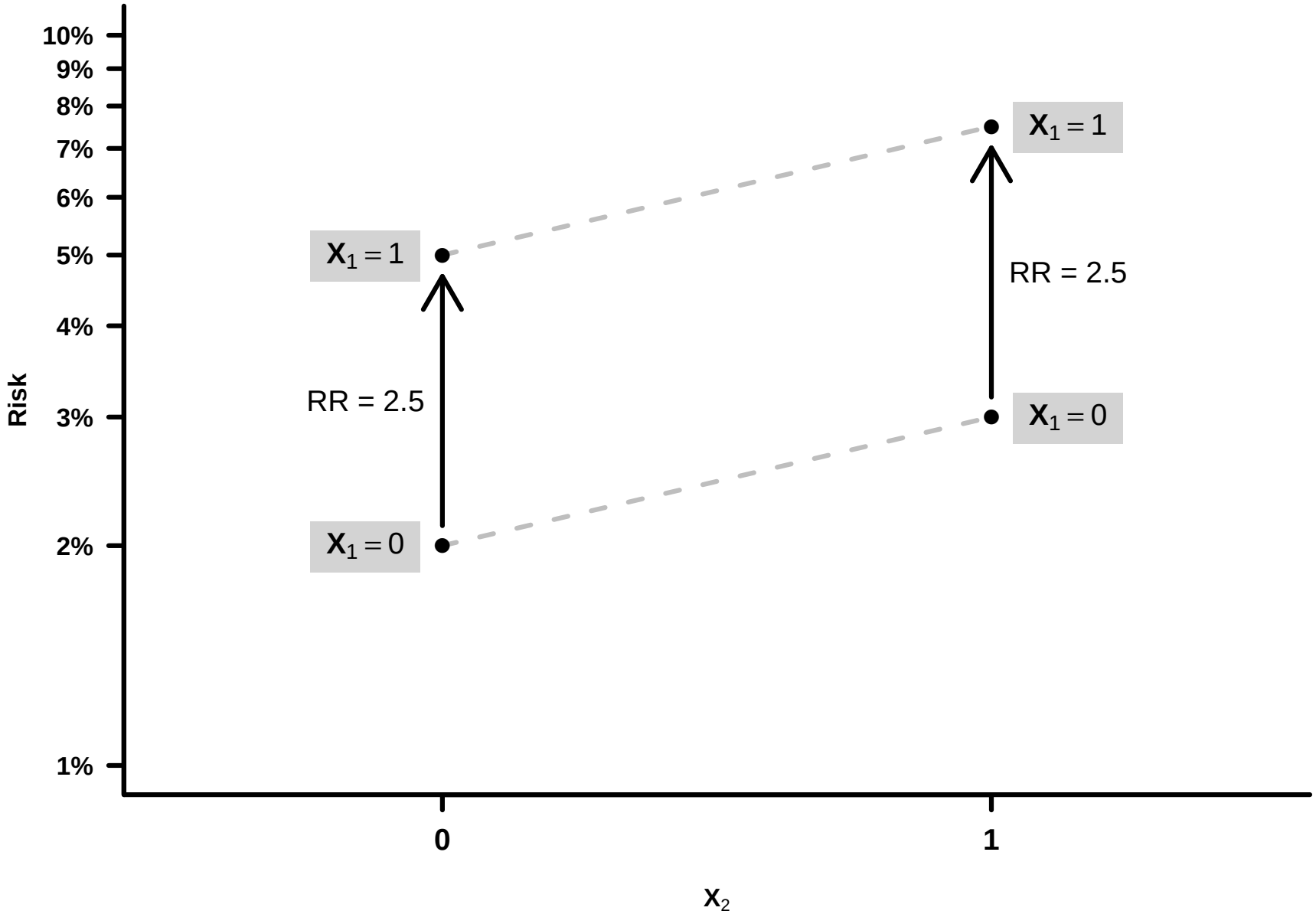
INTERACTIONS, RELATIVE RISK (RR), SCALE DEPENDENCE



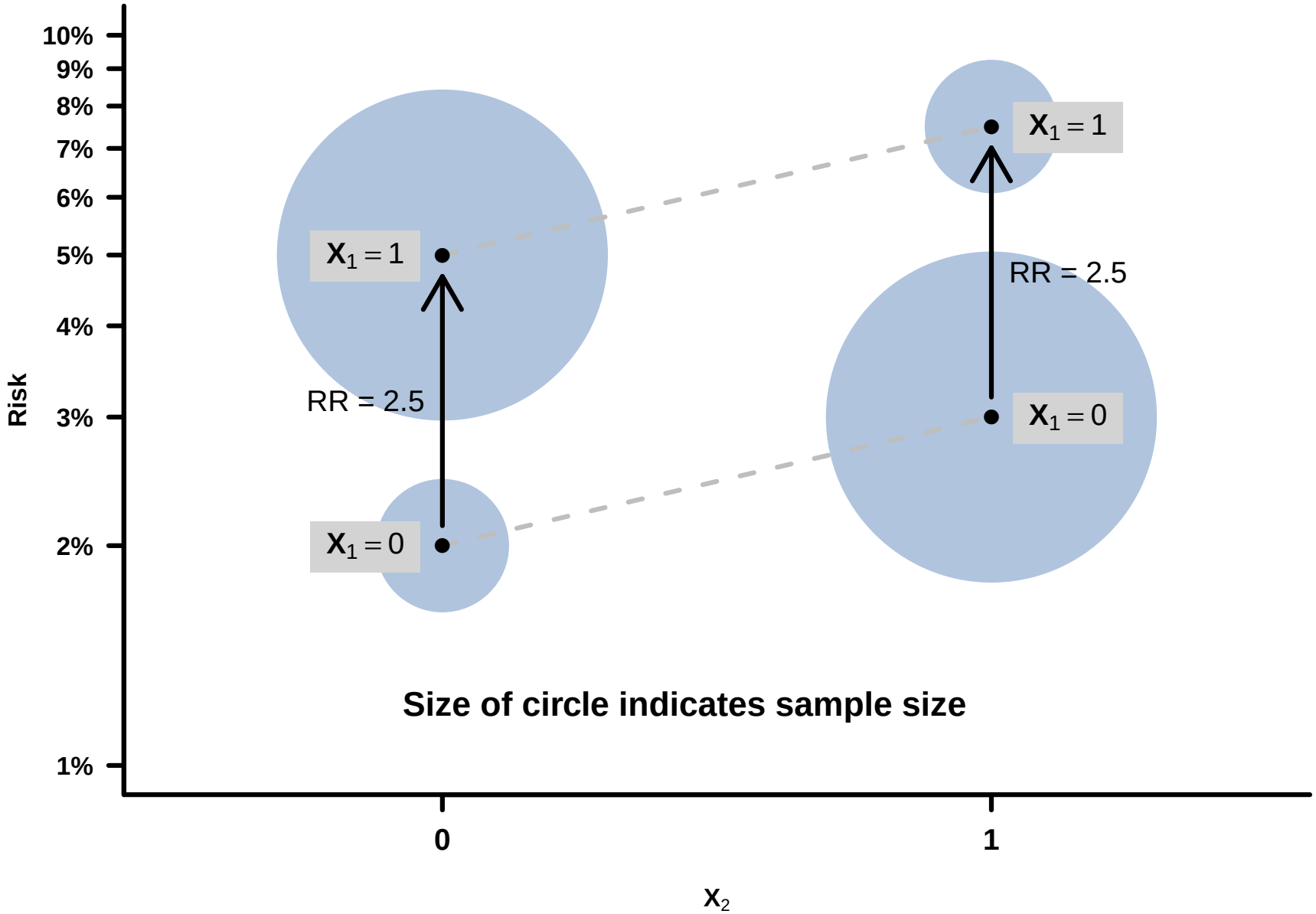
INTERACTIONS, QUALITATIVE



INTERACTIONS VERSUS CONFOUNDING



INTERACTIONS VERSUS CONFOUNDING



INTERACTIONS, BINOMIAL REGRESSION

- Logistic model (logit link):

$$\text{logit}(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underbrace{\beta_3\ x_1 \cdot x_2}_{\text{interaction}}$$

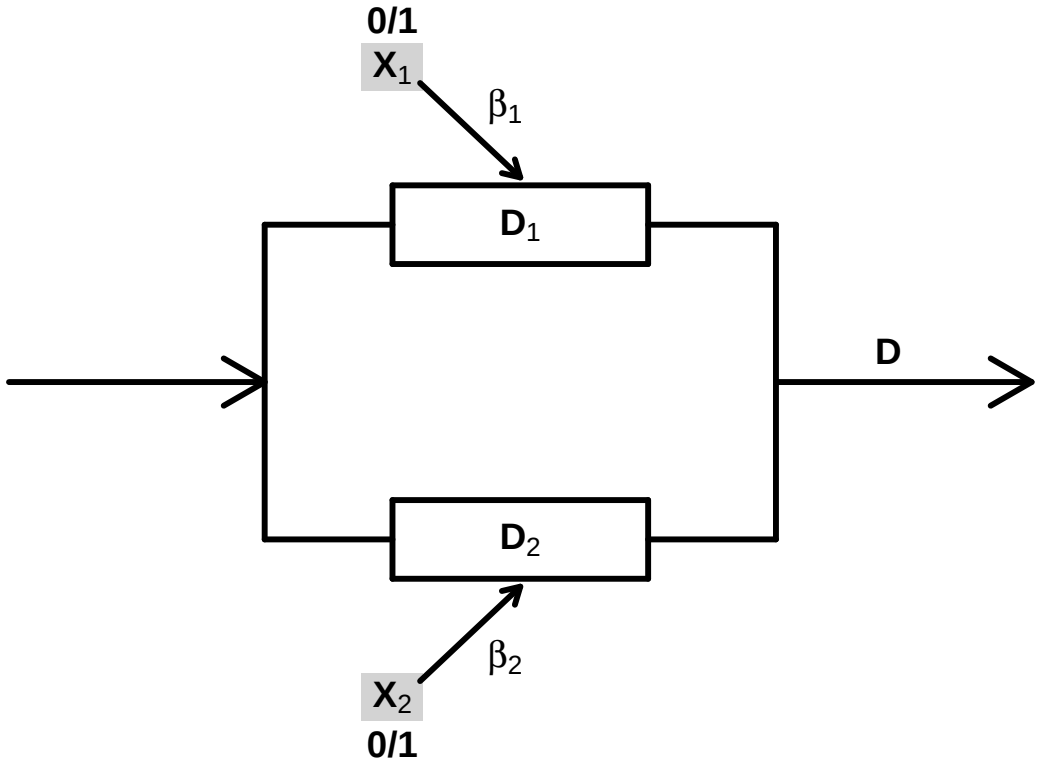
- Relative risk model (log link):

$$\log(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underbrace{\beta_3\ x_1 \cdot x_2}_{\text{interaction}}$$

- Additive model (identity link):

$$p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underbrace{\beta_3\ x_1 \cdot x_2}_{\text{interaction}}$$

INTERACTION: HETEROGENEITY MODEL (INDEPENDENT ACTION)



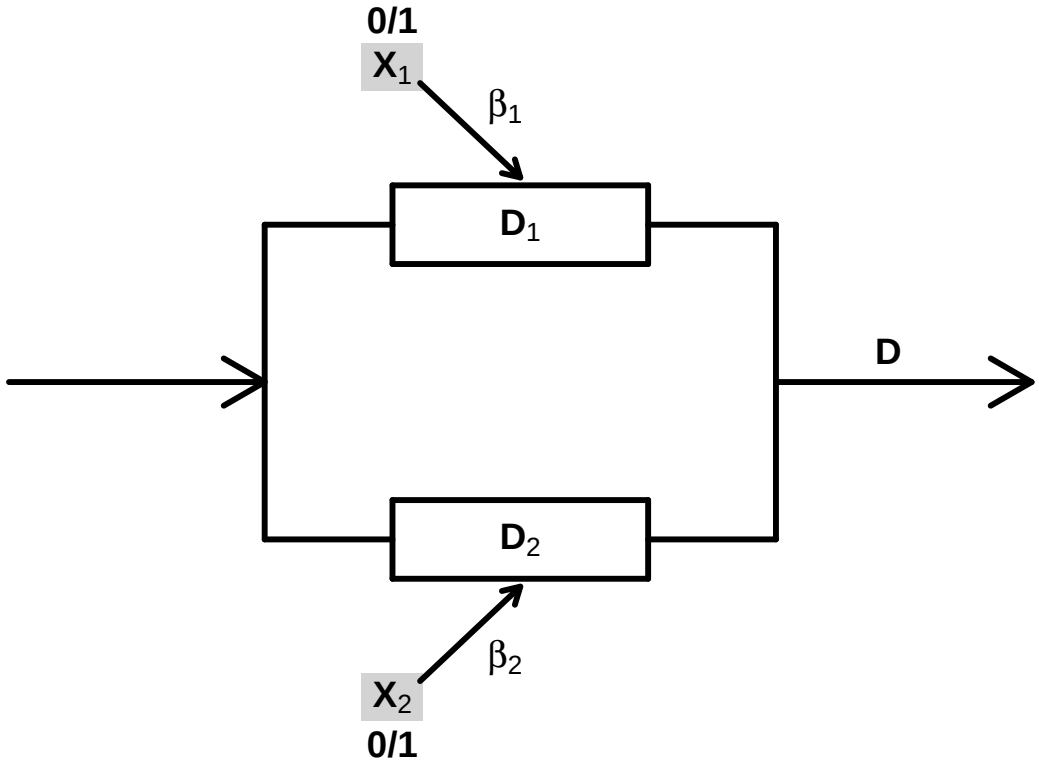
$$P(D) = P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$$

$$P(D_1) = \beta_1 X_1 \qquad P(D_2) = \beta_2 X_2$$

$$P(D) = \beta_1 X_1 + \beta_2 X_2 - \underbrace{\beta_1 \beta_2 \, X_1 \cdot X_2}_{\beta_3 = -\beta_1 \beta_2}$$

Interaction on an additive scale but components act independently!

INTERACTION: HETEROGENEITY MODEL (INDEPENDENT ACTION)

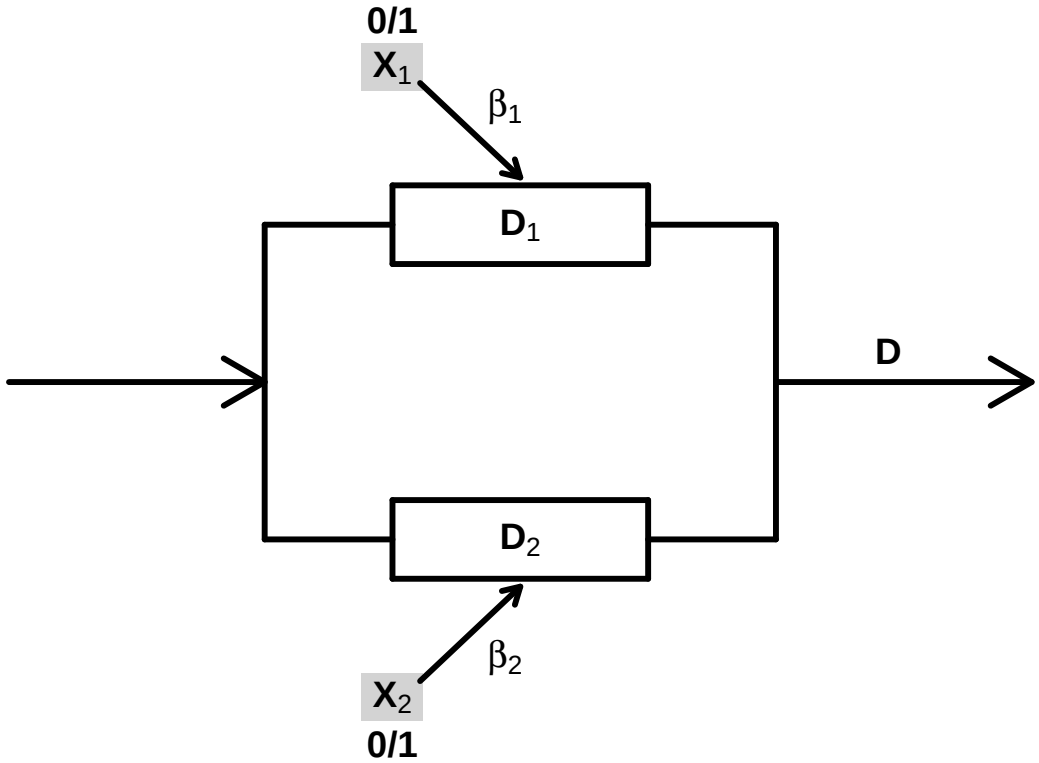


$$P(D) = P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$$

Additiv for $\log(1-p)$

$$1 - P(D) = (1 - \beta_1 X_1)(1 - \beta_2 X_2)$$

INTERACTION: HETEROGENEITY MODEL (INDEPENDENT ACTION)

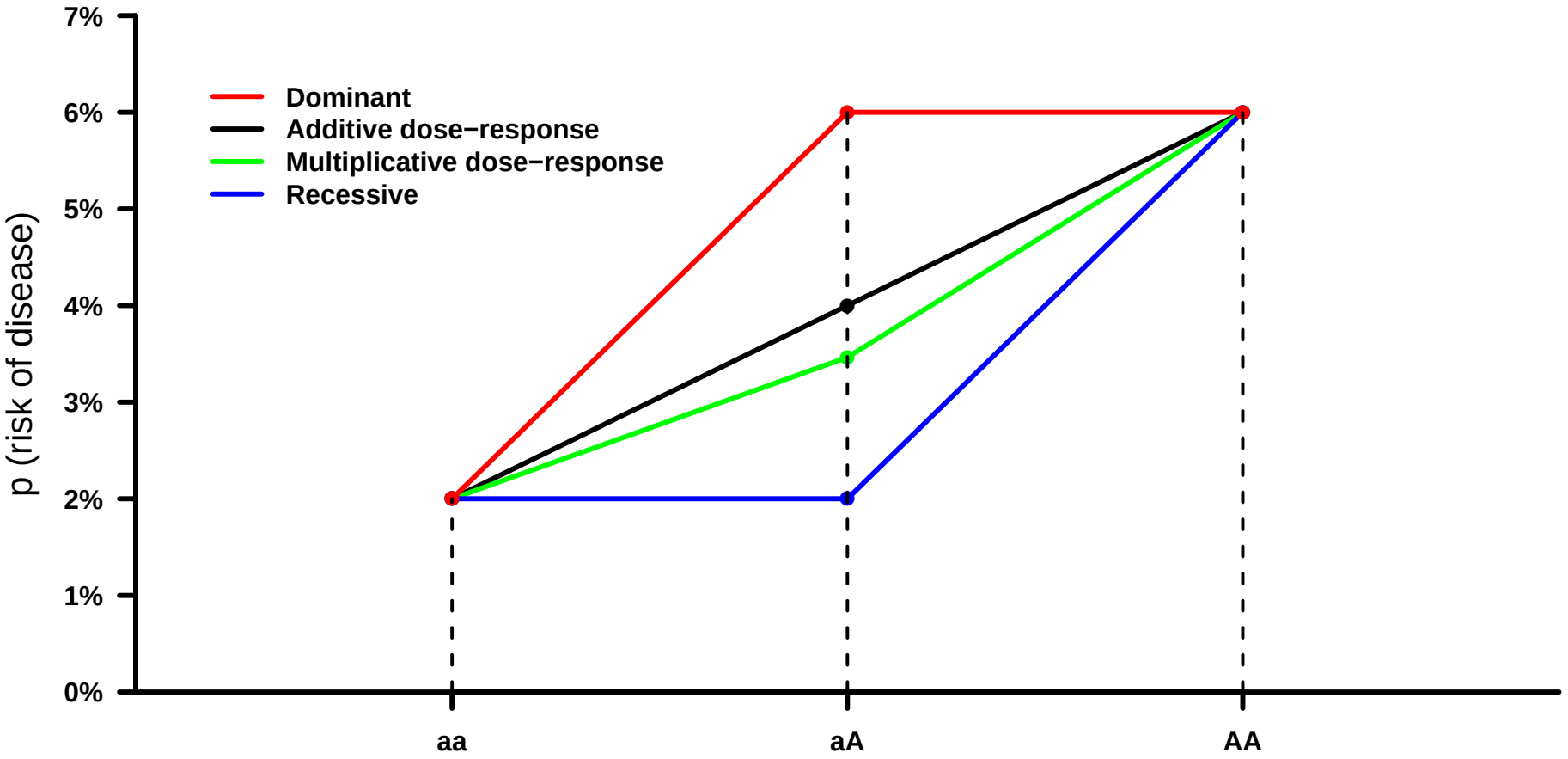


$$P(D) = P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$$

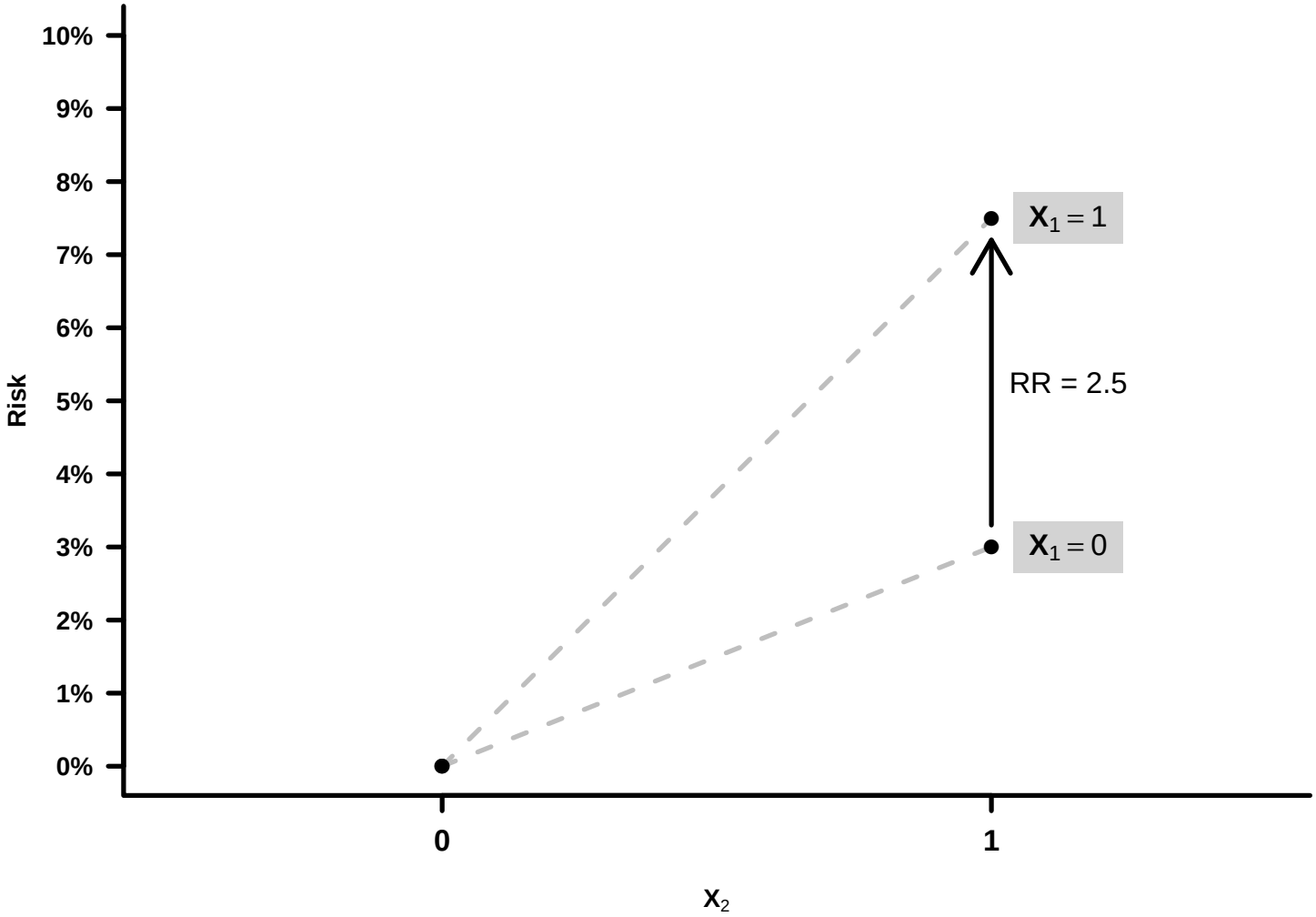
Multiple loci:

$$P(D) = P(D_1 \cup \cdots \cup D_K) = 1 - \prod_i (1 - P(D_i)) = 1 - \prod_i (1 - \beta_i X_i)$$

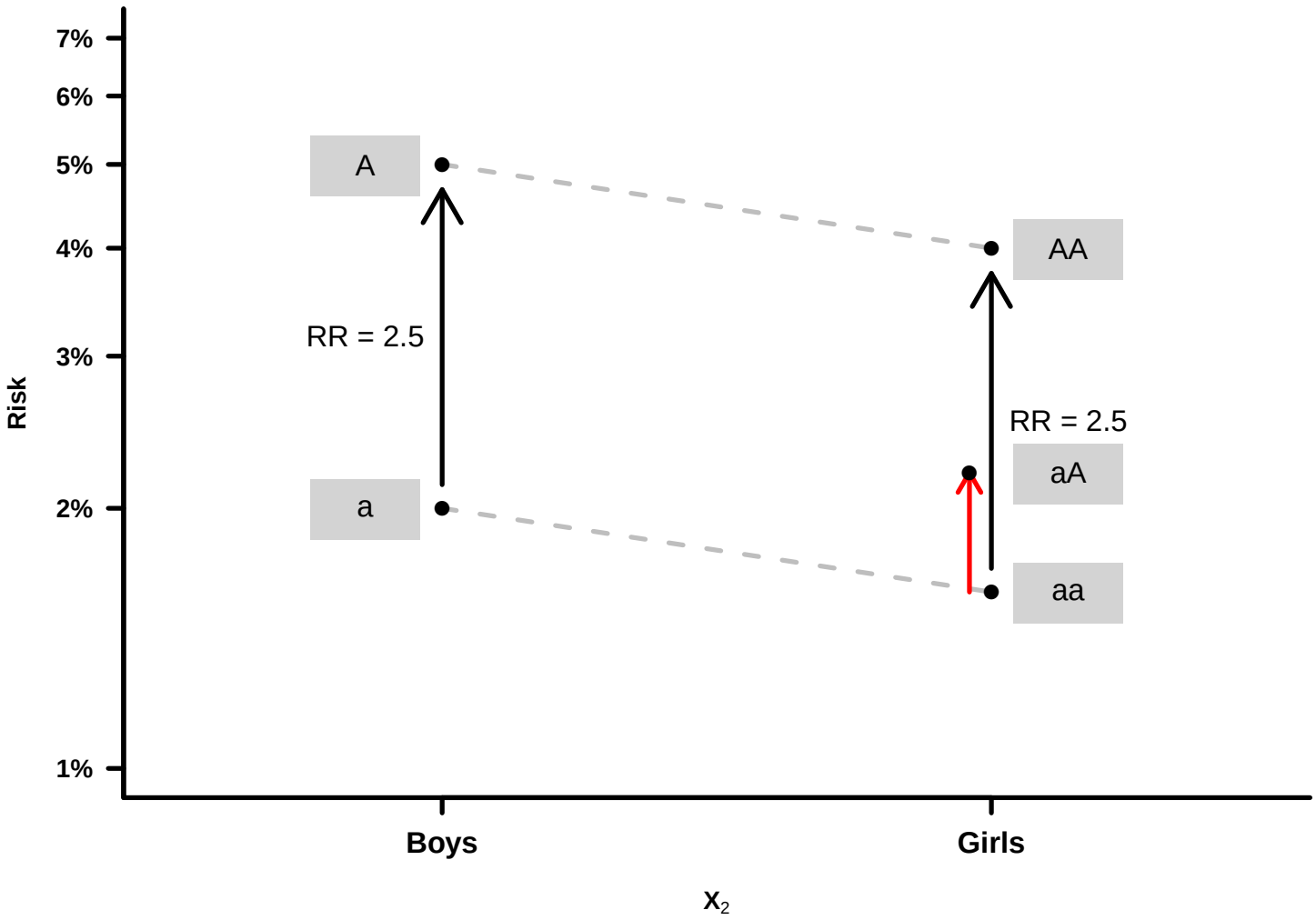
GENETIC INTERACTION MODELS: ALLELE INTERACTIONS



GENETIC INTERACTION MODELS: ONE NECESSARY LOCUS



GENETIC INTERACTION MODELS: X-INACTIVATION



MULTIPLICATIVE VERSUS ADDITIVE

Assume:

- Baseline risk 4%
- $\mathrm{RD} = 1\%$ risk difference for both X_1 and X_2
- $\mathrm{RR} = 1.25$ relative risk for both X_1 and X_2

Additive risk

		X_2 = 0	X_2 = 1
X_1	0	4%	4% + 1% = 5%
	1	4% + 1% = 5%	4% + 1% + 1% = 6%

Multiplicative risk

		X_2 = 0	X_2 = 1
X_1	0	4%	$4\% \cdot 1.25 = 5\%$
	1	$4\% \cdot 1.25 = 5\%$	$4\% \cdot 1.25 \cdot 1.25 = 6.25\%$

MULTIPLICATIVE VERSUS ADDITIVE

Assume:

- Baseline risk 4%
- $\mathrm{RR} = 1.25$ relative risk for both X_1 and X_2
- $\mathrm{RD} = 1\%$ risk difference for both X_1 and X_2

When $X_1 = 1$ and $X_2 = 1$:

ADDITIVE RISK

$$4\% + 1\% + 1\% = 4\% \cdot (1 + 0.25 + 0.25) = 6\%$$

MULTIPLICATIVE RISK

$$4\% \cdot (1 + 0.25) \cdot (1 + 0.25) = 4\% \cdot (1 + 0.25 + 0.25 + 0.0625) = 6.25\%$$

Difference is small when effect is “small” compared to baseline risk

INTERACTIONS: WHO SAID IT WAS EASY?

Original Paper

Human Heredity

Hum Hered 2000;50:334–349

Received: April 7, 1999
Revision received: July 23, 1999
Accepted: August 5, 1999

A Complete Enumeration and Classification of Two-Locus Disease Models

Wentian Li[a] Jens Reich[b]

[a]Laboratory of Statistical Genetics, Rockefeller University, New York, N.Y., USA and
[b]Department of Biomathematics, Max-Delbrück-Centrum, Berlin-Buch, Germany

INTERACTIONS: WHO SAID IT WAS EASY?

Abstract

There are 512 two-locus, two-allele, two-phenotype, fully penetrant disease models. Using the permutation between two alleles, between two loci, and between being affected and unaffected, one model can be considered to be equivalent to another model under the corresponding permutation. These permutations greatly reduce the number of two-locus models in the analysis of complex diseases. This paper determines the number of nonredundant two-locus models (which can be 102, 100, 96, 51, 50, or 58, depending on which permutations are used, and depending on whether zero-locus and single-locus models are excluded). Whenever possible, these nonredundant two-locus models are classified by their

Interactions: Who said it was easy?

> property. Besides the familiar features of multiplicative models (logical AND), heterogeneity models (logical OR), and threshold models, new classifications are added or expanded: modifying-effect models, logical XOR models, interference and negative interference models (neither dominant nor recessive), conditionally dominant/recessive models, missing lethal genotype models, and highly symmetric models. The following aspects of two-locus models are studied: the marginal penetrance

INTERACTIONS: WHO SAID IT WAS EASY?

Example 1:

		Locus 2		
		bb	bB	BB
Locus 1	aa	0	0	0
	aA	0	0	0
	AA	0	0	1

0 = no disease

1 = disease

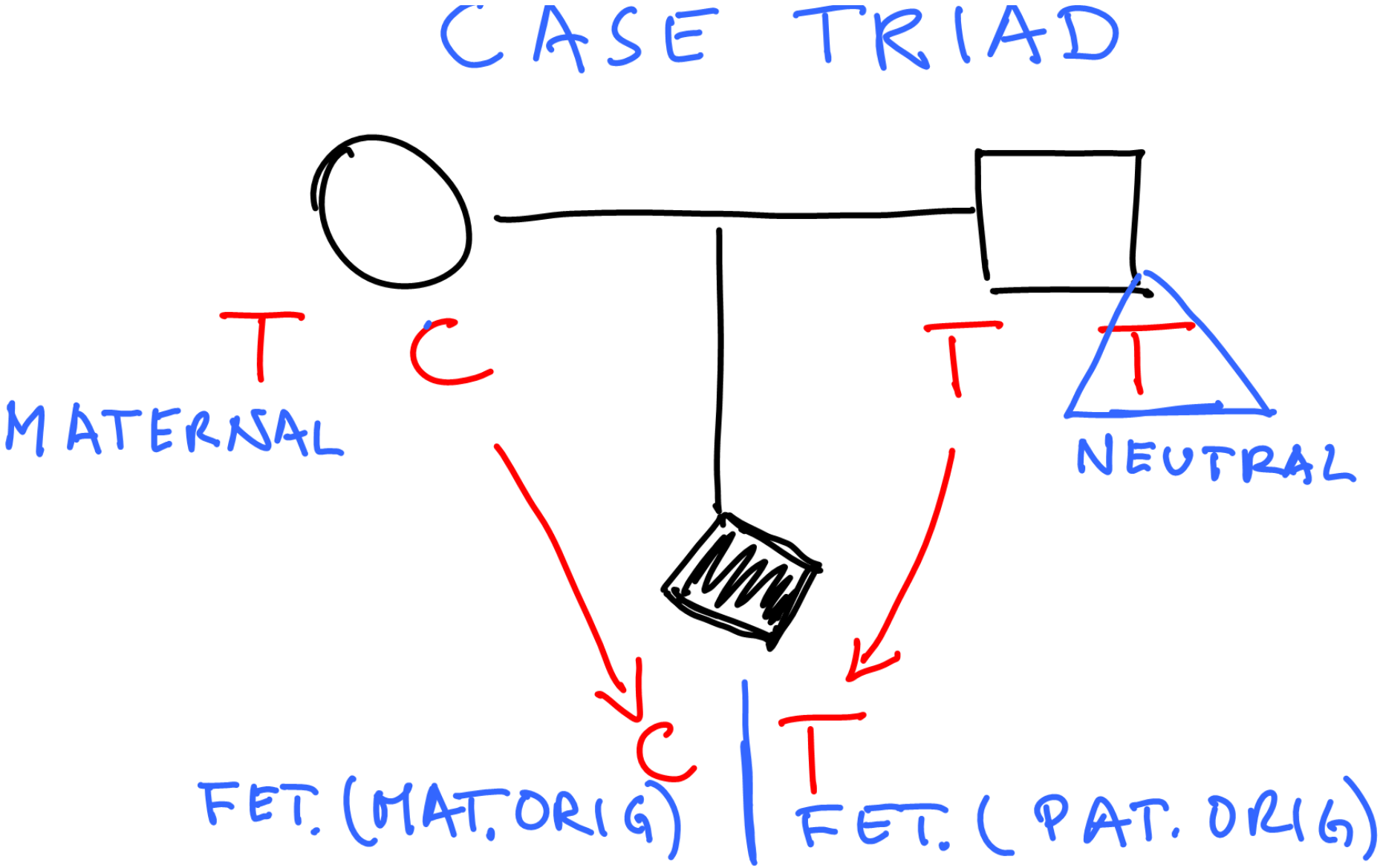
Example 2:

		Locus 2		
		bb	bB	BB
Locus 1	aa	0	1	1
	aA	1	1	1
	AA	1	1	1

0 = no disease

1 = disease

PARENT-OF-ORIGIN EFFECTS



PARENT-OF-ORIGIN EFFECTS

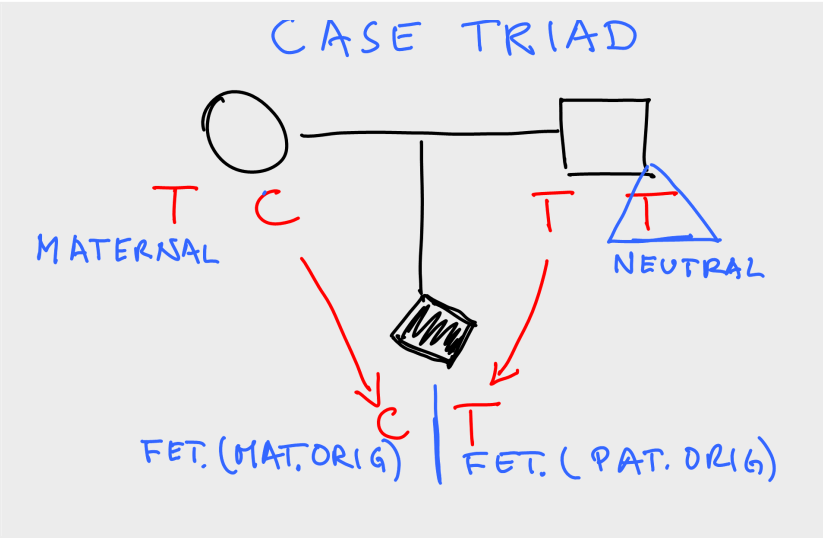
- **Parent-of-origin effect:**

  Interaction between allele effect and parent-of-origin

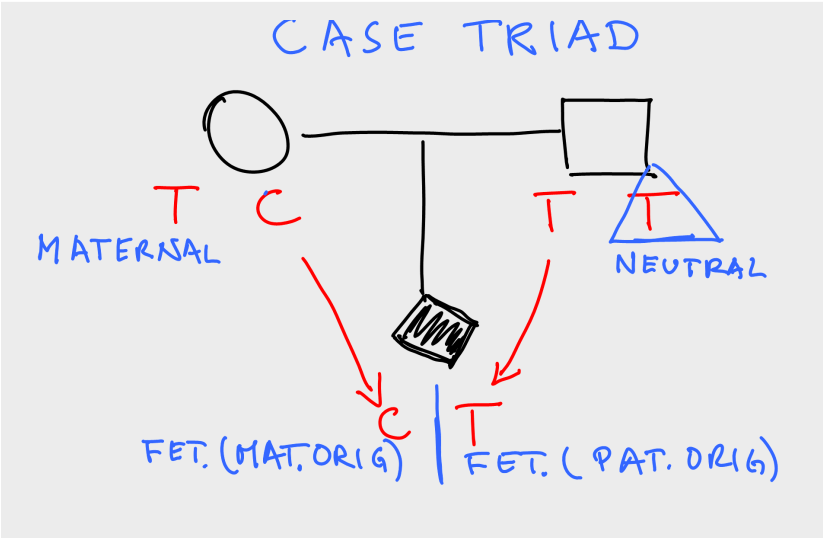
- **GxE:**

  Further interaction with environment

Non-smokers



Smokers



PARENT-OF-ORIGIN EFFECTS, CLEFT LIP/PALATE

RS2964137, KIAA0947

Test effect	Stratum	RR_{cm}	RR_{cf}	RR_{cm}/RR_{cf}
POO effects	S1	0.707	0.936	0.755 (0.505, 1.111)
	S2	0.707	0.936	0.755 (0.505, 1.111)
	S1/S2	1 (-)	1 (-)	1 (-)
GxE effects	S1	0.802	0.802	1 (-)
	S2	0.865	0.865	1 (-)
	S1/S2	0.928 (0.538, 1.603)	0.928 (0.538, 1.603)	1 (-)
POOxE effects	S1	0.533	1.170	0.456 (0.290, 0.705)
	S2	1.915	0.385	4.979 (2.137, 11.253)
	S1/S2	0.278 (0.135, 0.576)	3.038 (1.446, 6.345)	0.092 (0.036, 0.236)

Received: 15 March 2017 | Accepted: 5 September 2017

DOI: 10.1111/ahg.12224

ORIGINAL ARTICLE

WILEY Annals of human genetics

Parent-of-origin-environment interactions in case-parent triads with or without independent controls

Miriam Gjerdevik[1,2] | Øystein A. Haaland[1] | Julia Romanowska[1,3] | Rolv T. Lie[1,4] | Astanand Jugessur[1,2,5] | Håkon K. Gjessing[1,5]

GENE-ENVIRONMENT INTERACTIONS

Exposure variable:

- Environmental exposure variable
- For instance, maternal smoking during pregnancy
- A moderate number of levels, for instance:
  1 = non smokers, 2 = light smokers, 3 = heavy smokers

Objective:

- Will the effect of fetal genes change depending on exposure?
- For instance, genes that modify alcohol metabolism may influence the harmful effect of alcohol consumption

Data:

- We need to include environment/exposure variable in the dataset

Environment (exposure) variable

- Generate a random environment variable with three levels!
  (Only as an illustration)
- Make sure everybody gets the same:

```
set.seed(24)
```

- Draw at random from levels 1, 2, 3:

```
env <- sample(1:3, size = 1659, replace = T)
head(env)
```

```
[1] 1 1 3 2 2 3
```

ENVIRONMENT (EXPOSURE) VARIABLE

Create and dump data frame with variable

```
env <- data.frame(env = env)
write.table(env, "data/env.dat", row.names = F)
```

- **NOTE:** This if of course ad-hoc, for the illustration. The variable should come from original data and be matched to individual id's
- Typically, the environment variable could be the same for all mother, father, and child. Or it could be NA for child and father, etc.

Read data, including environment variable

```
tmp <- genDataRead(file.in = "data/pres.ped",
 file.out = "pres", dir.out = "data",
 format = "ped", cov.file.in = "data/env.dat")
```

Before pre-processing

```
head(tmp$cov.data)
```

```
  id.fam id.c id.f id.m sex cc env
1      1    1    3    2   2  1   1
2      1    2    0    0   2  0   1
3      1    3    0    0   1  1   3
4      2    1    3    2   1  0   2
5      2    2    0    0   2  0   2
6      2    3    0    0   1  1   3
```

- Data frame with character vectors
- Everything read “as is”

Pre-process data, using 3 cores

```
pres.data <- genDataPreprocess(tmp, map.file = "data/pres.map",
 dir.out = "data", ncpu = 3)
```

QUICK PEEK AT THE INTERNAL DATA STRUCTURE

```
head(pres.data$cov.data)
```

```
  id.fam id.m id.f id.c sex.m sex.f sex.c cc.m cc.f
1      1    1    2    1     1     1     2    1    2
2    112    1    2    1     1     1     1    1    2
3    223    1    2    1     1     1     1    2    1
4    334    1    2    1     1     1     1    1    1
5    445    1    2    1     1     1     2    1    2
6    507    1    2    1     1     1     1    2    1
```

```
...  cc.c env.m env.f env.c
...     2     1     3     1
...     1     2     3     2
...     1     3     3     1
...     2     2     2     1
...     2     3     1     3
...     2     1     1     3
```

QUICK PEEK AT THE INTERNAL DATA STRUCTURE

Useful to know (but not really necessary):

- During pre-processing, all data are transformed to Haplin format
- m-f-c format for all covariates and all genetic data
- `cov.data` first contains all covariate data, read "raw"
- During pre-processing, all `cov.data` is recoded and replaced by its numerical codes.
- Codes and frequency counts are stored in pres.dataauxvariables

```
> names(pres.data$aux$variables)
[1] "id.fam" "id.m"   "id.f"   "id.c"   "sex.m"  "sex.f"
 "sex.c"  "cc.m"   "cc.f"  "cc.c"   "env.m"  "env.f"  "env.c"
```

```
> pres.data$aux$variables[["env.c"]]
  1   2   3
192 190 177
```

RUN `haplin` ON STRATIFIED DATA

- `haplinStrat` runs `haplin` on each stratum of environmental variable.
- You need to choose what stratification variable to use.
- If you specify `strata = "env"` you will get `env.c`, i.e. the exposure of the child.
- Or you can specify manually by number. `strata = 11` uses `env.m`, i.e. the exposure of the mother.

Make sure your environment variable is specified correctly in the original file.

```
result <- haplinStrat(data = pres.data, markers = 1,
response = "mult", reference = "ref.cat",
use.missing = T, strata = 11)
```

RUN haplin ON STRATIFIED DATA

```
## Running haplinStrat ##

Reading data from file...
Frequency distribution of selected stratification variable:
  1   2   3
206 172 181

Running Haplin for full data file...Done
Running Haplin on stratum "1"...Done
Running Haplin on stratum "2"...Done
Running Haplin on stratum "3"...Done
```

RUN haplin ON STRATIFIED DATA

- The first element of `result` is for all data
- The remaining elements are for each stratum

```
> names(result)
 [1] "all" "1"   "2"   "3"
```

- Plot all strata and dump to pdf

```
pdf(file = "strataplot.pdf")
lapply(result, plot)
dev.off()
```

- Join to single haptable

```
result1 <- haptable(result)
----- OR:
result1 <- lapply(result, haptable)
result1 <- toDataFrame(result1, reduce = T)
```

TABLE OF RESULTS

	element	marker	alleles	counts	HWE.pv	Original	After.rem.NA
1	all	rs1	c/G	146/3040	0.7075356	559	559
3	1	rs1	c/G	49/1127	0.3118923	206	206
5	2	rs1	c/G	44/934	0.9914135	172	172
7	3	rs1	c/G	53/979	0.7443079	181	181

	After.rem.Mend.inc.	After.rem.unused.haplos	pv.overall	haplos	haplofre
1	559	559	0.4733520	c	0.0418148
3	206	206	0.4907621	c	0.0341654
5	172	172	0.9058180	c	0.0444670
7	181	181	0.6683192	c	0.0479503

	haplofreq.lower	haplofreq.upper	reference	RR.est.	RR.lower	RR.upper
1	0.03107798	0.05593044	-	1.159908	0.7744188	1.756616
3	0.01977345	0.05793304	-	1.286377	0.6281270	2.685070
5	0.02640951	0.07371234	-	1.042735	0.5062032	2.190587
7	0.02941477	0.07677688	-	1.156227	0.5945482	2.288260

	RR.p.value	RRdd.est.	RRdd.lower	RRdd.upper	RRdd.p.value
1	0.4684035	1.345387	0.5997245	3.085701	0.4684035
3	0.4873102	1.654765	0.3945435	7.209603	0.4873102
5	0.9010253	1.087296	0.2562417	4.798672	0.9010253
7	0.6630191	1.336860	0.3534876	5.236135	0.6630191

TEST FOR GENE-ENVIRONMENT INTERACTION

- We can test if there is change from stratum to stratum
- Parameters tested for here, including trend tests:
  1. Haplotype frequencies
  2. Relative risk estimates
- Testing:

```
gxe(result)
```

- P-values from test:

```
          gxe.test      chisq df      pval
1       haplo.freq 0.92303684  2 0.6303258
2            child 0.15885560  2 0.9236447
3 haplo.freq.trend 0.86293402  1 0.3529189
4      child.trend 0.04483292  1 0.8323116
```

- No significant change in haplotype frequencies
- No significant change in child effects

GJERDEVIK ET AL. POOXE 2017

Received: 15 March 2017 | Accepted: 5 September 2017

DOI: 10.1111/ahg.12224

ORIGINAL ARTICLE

WILEY Annals of human genetics

Parent-of-origin-environment interactions in case-parent triads with or without independent controls

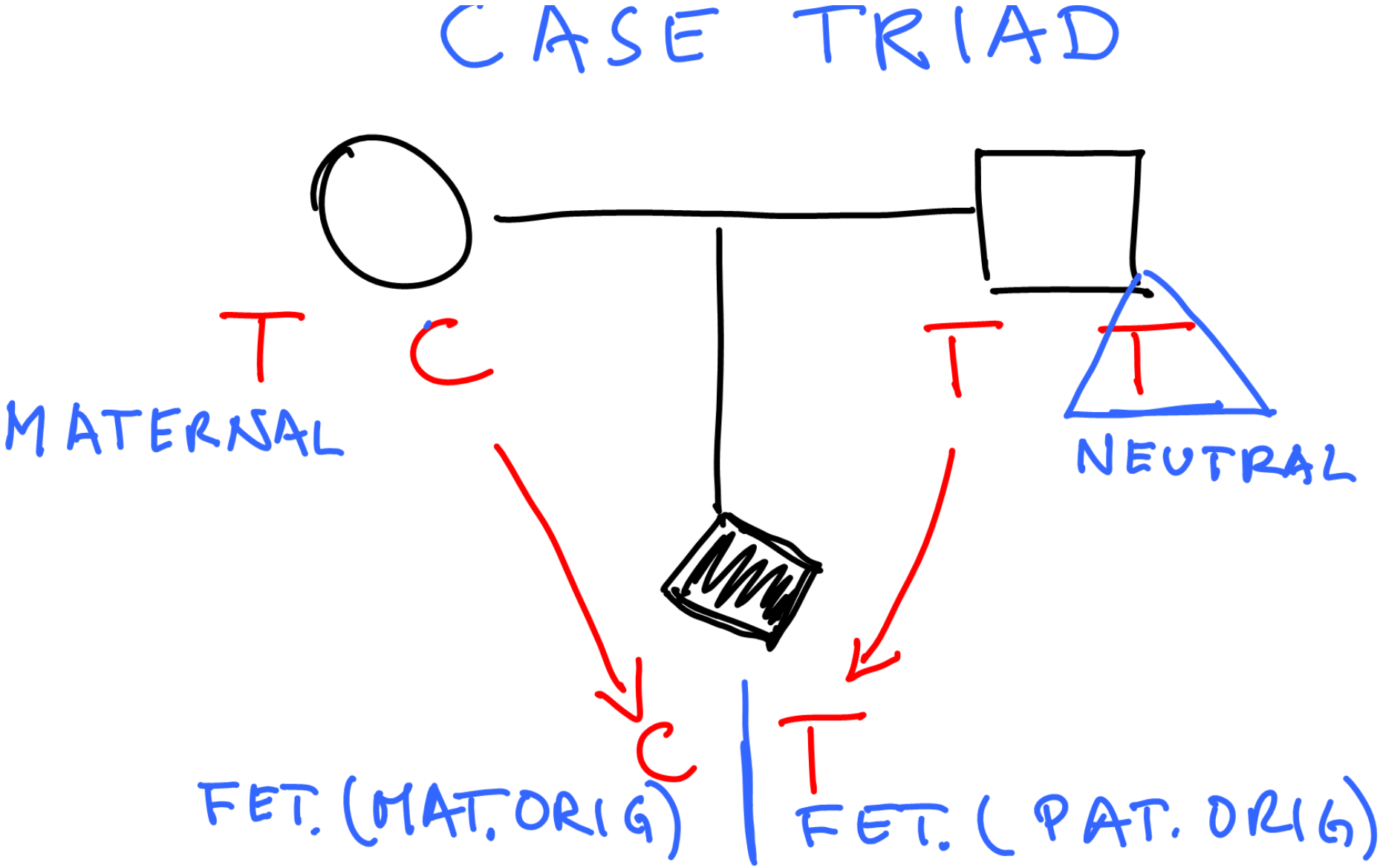
Miriam Gjerdevik[1,2] | Øystein A. Haaland[1] | Julia Romanowska[1,3] | Rolv T. Lie[1,4] | Astanand Jugessur[1,2,5] | Håkon K. Gjessing[1,5]

[1]Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

[2]Department of Genetic Research and Bioin-

Abstract

With case–parent triad data, one can frequently deduce parent of origin of the child's alleles. This allows a parent-of-origin (PoO) effect to be estimated as the ratio of


CASE TRIAD
T
C
T
T
MATERNAL
NEUTRAL
C
T
FET. (MAT. ORIG)
FET. (PAT. ORIG)

PARENT-OF-ORIGIN EFFECTS

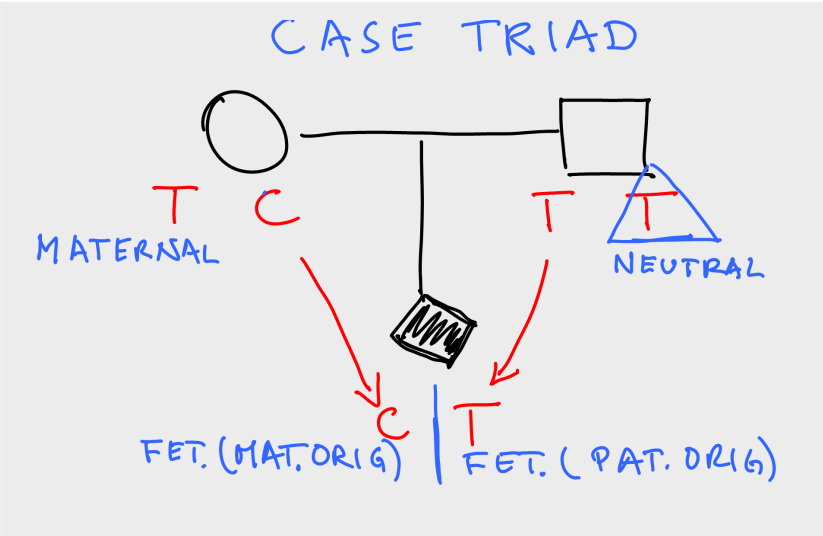
- **Parent-of-origin effect:**

  Interaction between allele effect and parent-of-origin

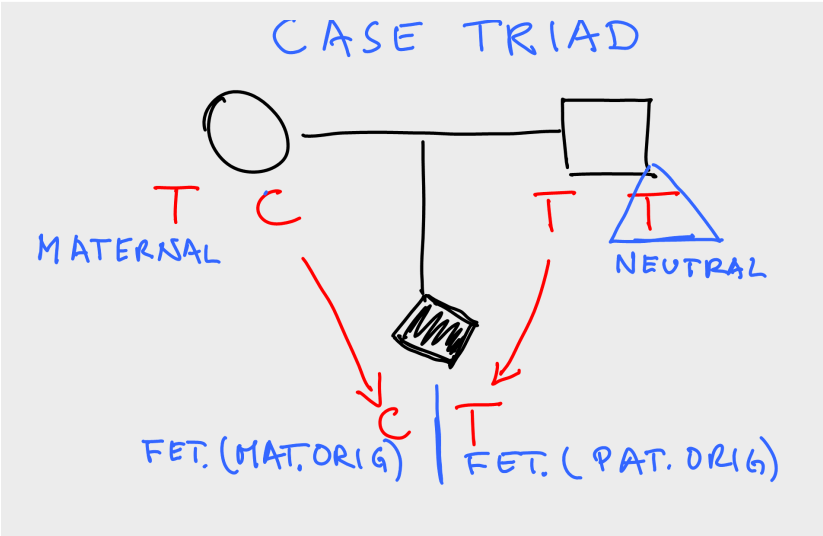
- **GxE:**

  Further interaction with environment

Non-smokers



Smokers



PARENT-OF-ORIGIN EFFECTS

RS2964137, KIAA0947

Test effect	Stratum	RR_{cm}	RR_{cf}	RR_{cm}/RR_{cf}
POO effects	S1	0.707	0.936	0.755 (0.505, 1.111)
	S2	0.707	0.936	0.755 (0.505, 1.111)
	S1/S2	1 (-)	1 (-)	1 (-)
GxE effects	S1	0.802	0.802	1 (-)
	S2	0.865	0.865	1 (-)
	S1/S2	0.928 (0.538, 1.603)	0.928 (0.538, 1.603)	1 (-)
POOxE effects	S1	0.533	1.170	0.456 (0.290, 0.705)
	S2	1.915	0.385	4.979 (2.137, 11.253)
	S1/S2	0.278 (0.135, 0.576)	3.038 (1.446, 6.345)	0.092 (0.036, 0.236)