INTERACTIONS

Examples and interpretations

Håkon K. Gjessing

Professor/Principal Investigator

Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo

Department of Global Public Health and Primary Care, University of Bergen

Makerere

Friday, 9 June 2023

AM SOC HUM GENETICS MEETING 2018, SAN DIEGO

CENTRE FOR FERTILITY AND HEALTH

MEANWHILE, IN BERGEN....

INTERACTIONS

The concept of “interaction” appears in numerous settings where a statistical analysis is used to throw light on complex biological systems

Loosely defined:

- One variable modifies the effect of another variable
- Thus, sometimes referred to as an “Effect modifications”
  ... but is that actually the same?

For instance:

- Smoking may be more dangerous to men than to women
- Vaccination may be more “useful” for older people than for younger
- A fetal gene might act differently depending on whether it is inherited from the mother or the father

INTERACTIONS

- Consider the effect of X_1 on Y
- Does the effect change over levels of another variable X_2?

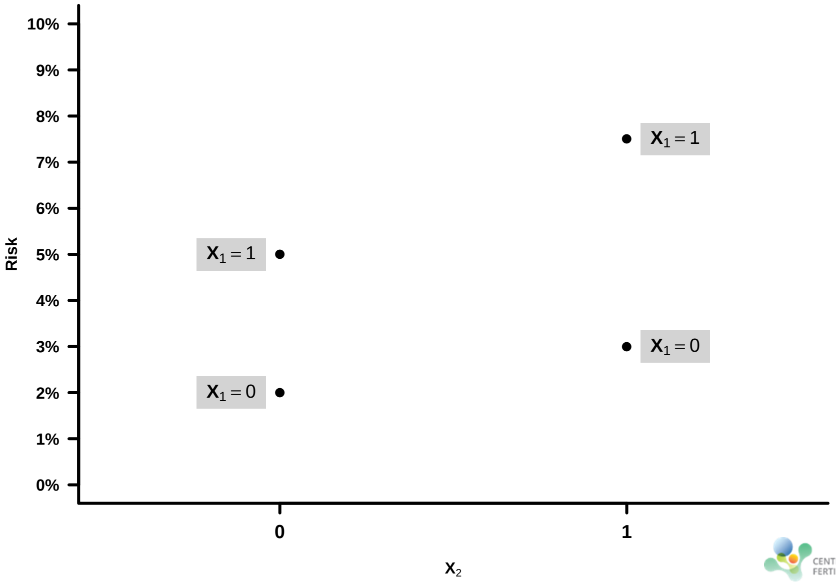
Some questions:

- What is the “correct” way of defining and measuring interactions?
- How does a statistical interaction reflect the underlying biological processes, for instance interactions at a molecular level?
- What are gene-gene (epistatsis) and gene-environment interactions?
- How do interaction analyses increase the required sample size of a study?
- Are statisticians good at social interactions?

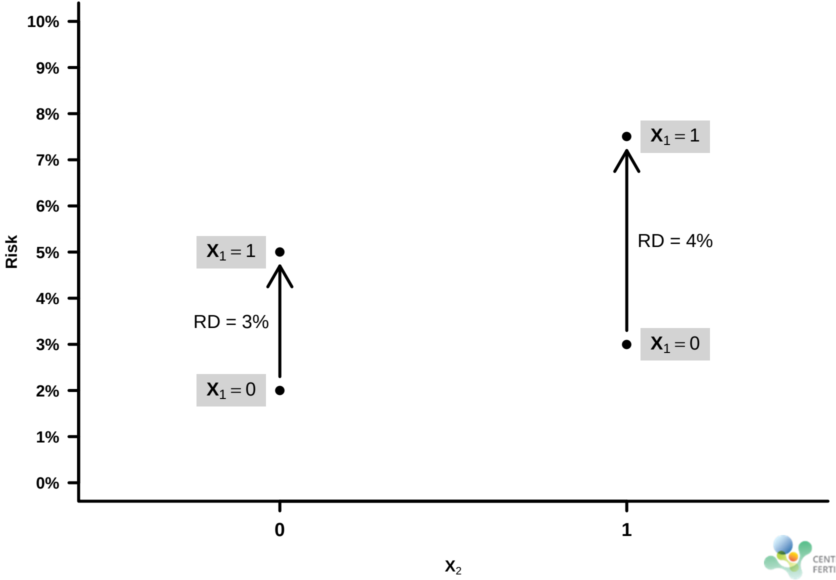
Avoiding the last question, we will see that the answers to the others are sometimes simple, sometimes complex.

CENTRE FOR
FERTILITY AND HEALTH

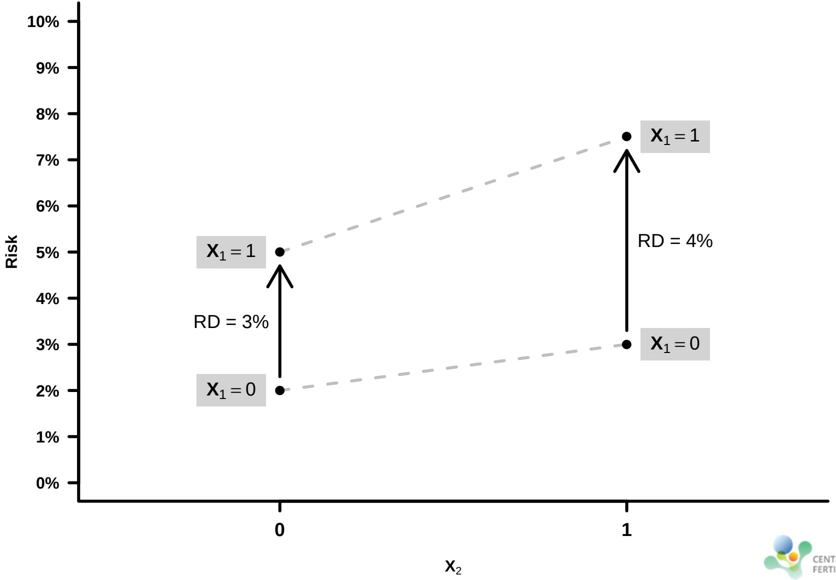
INTERACTIONS, RISK DIFFERENCE (RD)



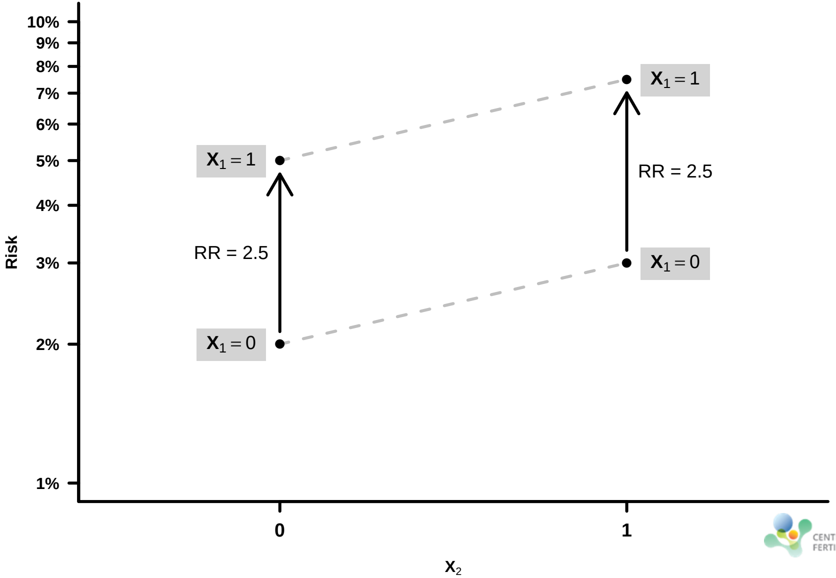
INTERACTIONS, RISK DIFFERENCE (RD)



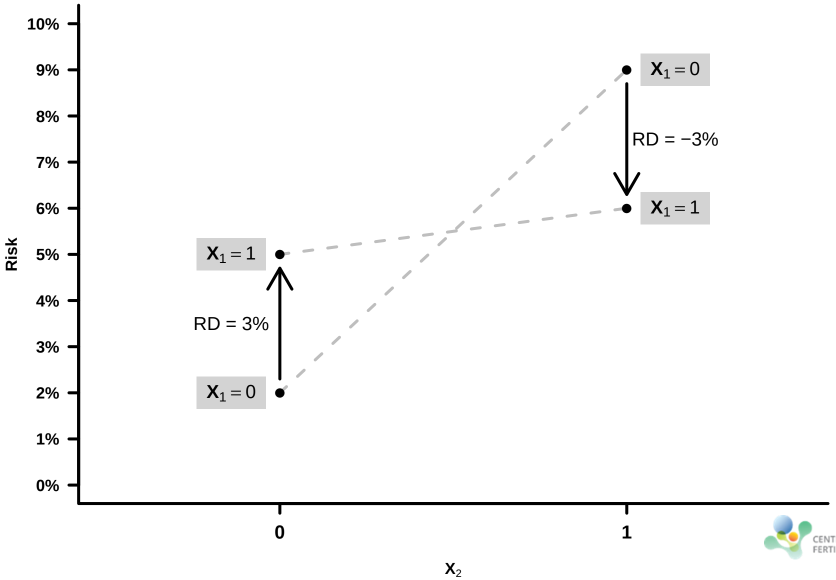
INTERACTIONS, RISK DIFFERENCE (RD)



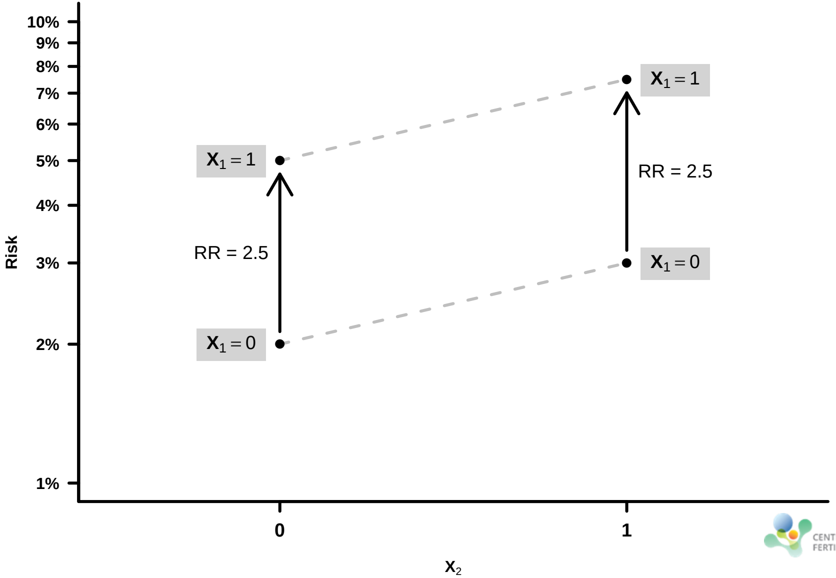
INTERACTIONS, RELATIVE RISK (RR), SCALE DEPENDENCE



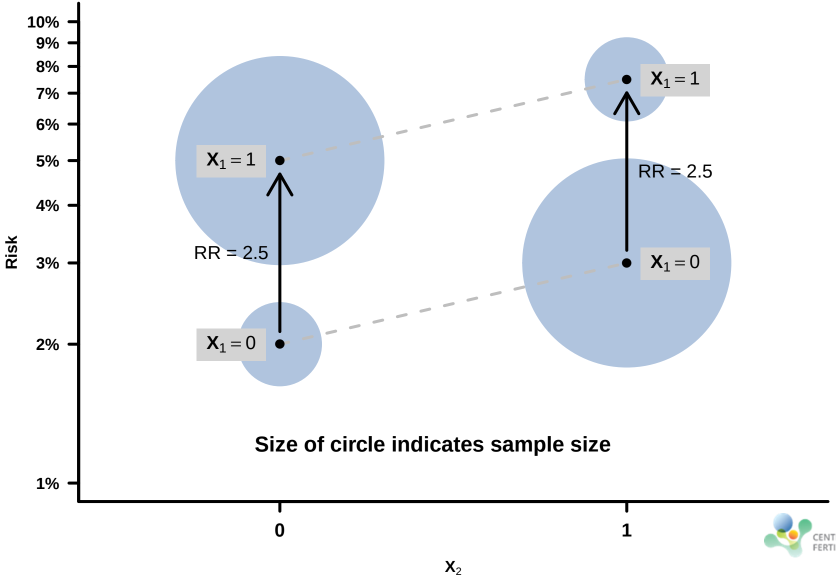
INTERACTIONS, QUALITATIVE



INTERACTIONS VERSUS CONFOUNDING



INTERACTIONS VERSUS CONFOUNDING



NOTE

If

- X_1 changes the effect of X_2 on Y

then **always**

- X_2 changes the effect of X_1 on Y

NOTE

If

- X_1 changes the effect of X_2 on Y

then **always**

- X_2 changes the effect of X_1 on Y

BUT if

- X_1 reverses the effect of X_2 on Y (i.e. “qualitative interaction”)

then it is **not necessarily** the case that

- X_2 reverses the effect of X_1 on Y

(Sigh...)

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):

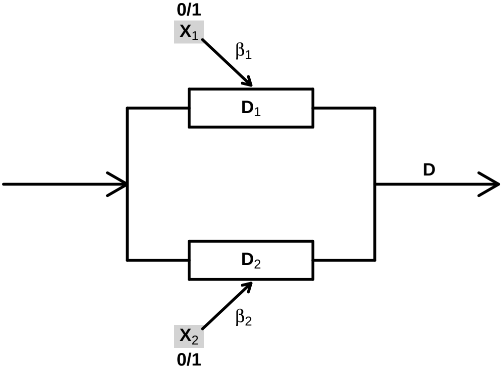
$$\log(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \underbrace{\beta_3\, x_1 \cdot x_2}_{\text{interaction}}$$

- Additive model (identity link):

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

CENTRE FOR
FERTILITY AND HEALTH

INTERACTION: HETEROGENEITY MODEL (INDEPENDENT ACTION)



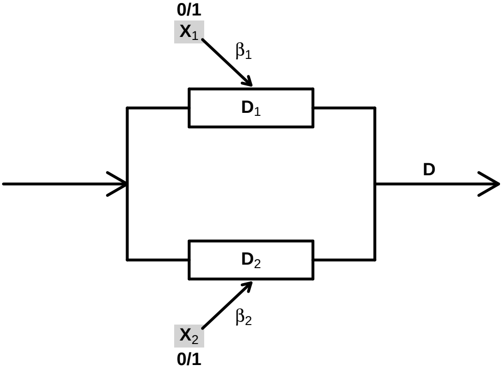
$$P(D) = P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$$

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

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

Interaction on an additive scale but components act independently!

INTERACTION: HETEROGENEITY MODEL (INDEPENDENT ACTION)

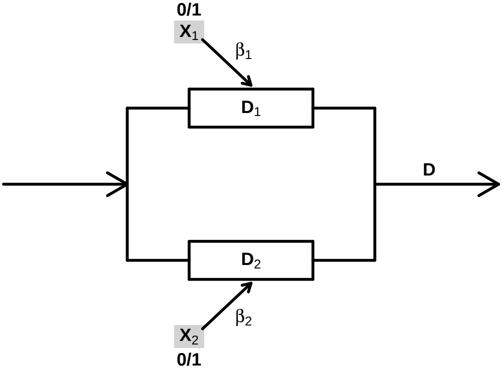


$$P(D) = P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$$

Additiv for $\log(1 - p)$

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

INTERACTION: HETEROGENEITY MODEL (INDEPENDENT ACTION)

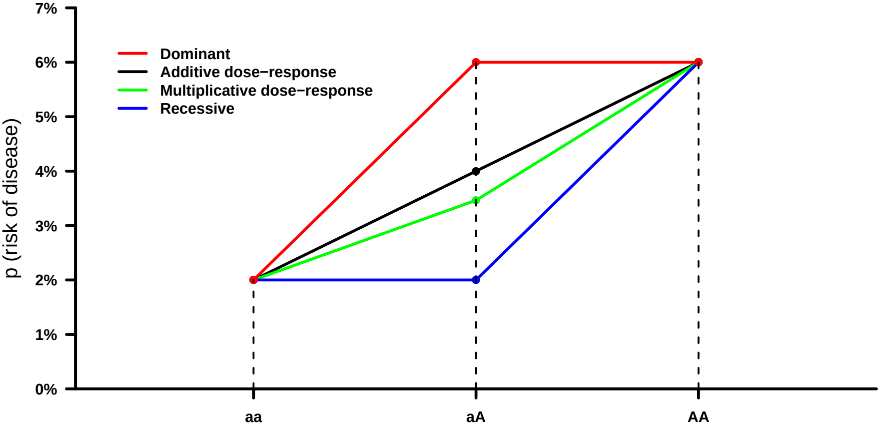


$$P(D) = P(D_1 \cup D_2) = P(D_1) + P(D_2) - P(D_1 \cap D_2)$$

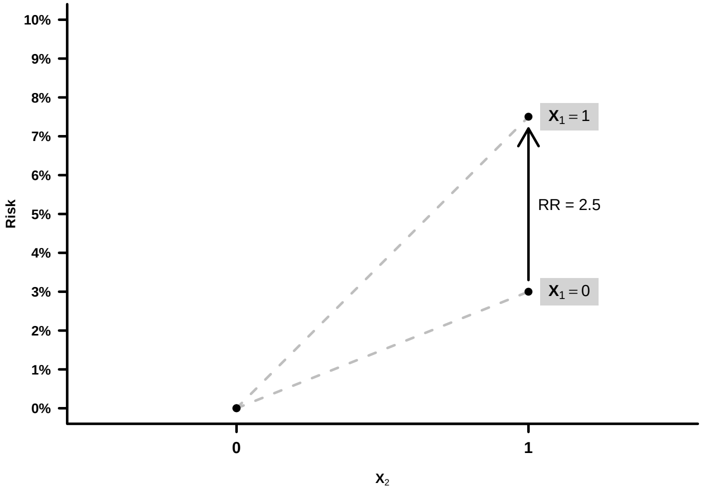
Multiple loci:

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

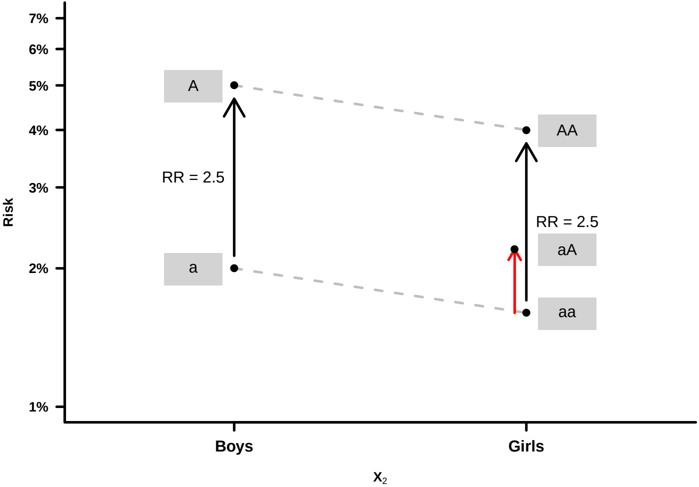
GENETIC INTERACTION MODELS: ALLELE INTERACTIONS



GENETIC INTERACTION MODELS: ONE NECESSARY LOCUS



GENETIC INTERACTION MODELS: X-INACTIVATION



CENTRE FOR FERTILITY AND HEALTH

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

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

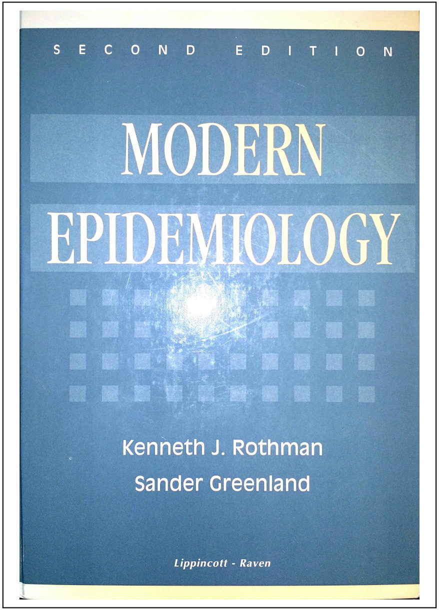
MULTIPLICATIVE RISK

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

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

CENTRE FOR FERTILITY AND HEALTH

DO “CANONICAL” INTERACTIONS EXIST?



SECOND EDITION

MODERN EPIDEMIOLOGY

Kenneth J. Rothman
Sander Greenland

Lippincott - Raven

INTERACTIONS: ALL COMBINATIONS

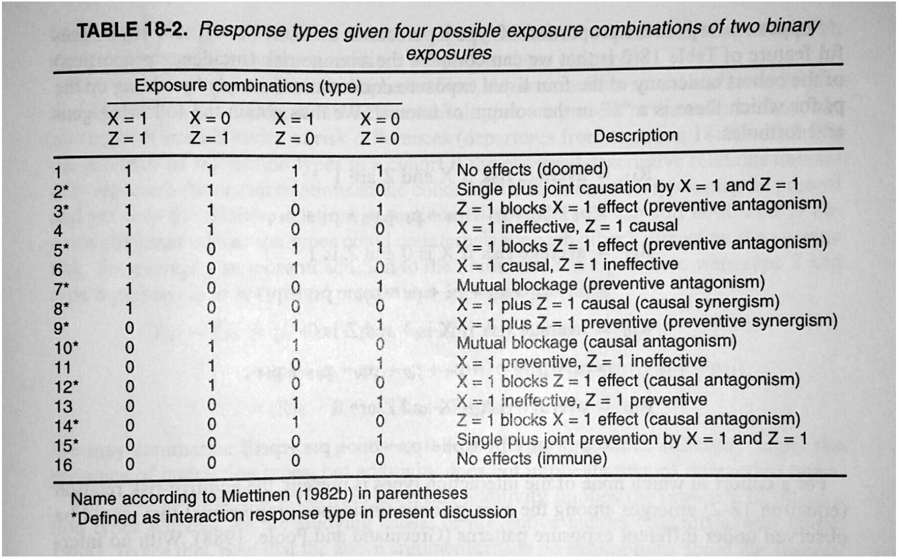


TABLE 18-2. *Response types given four possible exposure combinations of two binary exposures*

	Exposure combinations (type)				
	X = 1 Z = 1	X = 0 Z = 1	X = 1 Z = 0	X = 0 Z = 0	Description
1	1	1	1	1	No effects (doomed)
2*	1	1	1	0	Single plus joint causation by X = 1 and Z = 1
3*	1	1	0	1	Z = 1 blocks X = 1 effect (preventive antagonism)
4	1	1	0	0	X = 1 ineffective, Z = 1 causal
5*	1	0	1	1	X = 1 blocks Z = 1 effect (preventive antagonism)
6	1	0	1	0	X = 1 causal, Z = 1 ineffective
7*	1	0	0	1	Mutual blockage (preventive antagonism)
8*	1	0	0	0	X = 1 plus Z = 1 causal (causal synergism)
9*	0	1	1	1	X = 1 plus Z = 1 preventive (preventive synergism)
10*	0	1	1	0	Mutual blockage (causal antagonism)
11	0	1	0	1	X = 1 preventive, Z = 1 ineffective
12*	0	1	0	0	X = 1 blocks Z = 1 effect (causal antagonism)
13	0	0	1	1	X = 1 ineffective, Z = 1 preventive
14*	0	0	1	0	Z = 1 blocks X = 1 effect (causal antagonism)
15*	0	0	0	1	Single plus joint prevention by X = 1 and Z = 1
16	0	0	0	0	No effects (immune)

Name according to Miettinen (1982b) in parentheses
*Defined as interaction response type in present discussion

Exposures: X(**0**/**1**) and Z(**0**/**1**) **Response: 0/1**

INTERACTIONS: STATISTICAL VERSUS "BIOLOGIC"??

According to Rothman & Greenland:

Statistical interaction

- Departure from additivity on any given scale

"Biologic" interaction

- Additivity on risk scale

	$X = 1$ $Z = 1$	$X = 0$ $Z = 1$	$X = 1$ $Z = 0$	$X = 0$ $Z = 0$	
2^*	1	1	1	0	Single plus joint causation by $X = 1$ and $Z = 1$
3^*	1	1	0	1	$Z = 1$ blocks $X = 1$ effect (preventive antagonism)

Example, 2^* :

- Z is a fatal traffic accident
- X is heavy smoking
- Outcome is death

Dying in a traffic accident prevents death from smoking...

Is that "biologic interaction"??

INTERACTIONS: STATISTICAL VERSUS "BIOLOGIC"??

According to Rothman & Greenland:

Statistical interaction

- Departure from additivity on any given scale

"Biologic" interaction

- Additivity on risk scale

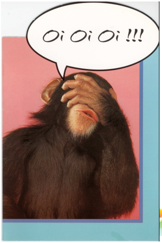
	$X=1$ $Z=1$	$X=0$ $Z=1$	$X=1$ $Z=0$	$X=0$ $Z=0$	
2^*	1	1	1	0	Single plus joint causation by $X=1$ and $Z=1$
3^*	1	1	0	1	$Z=1$ blocks $X=1$ effect (preventive antagonism)

Example, 2^* :

- Z is a fatal traffic accident
- X is heavy smoking
- Outcome is death

Dying in a traffic accident prevents death from smoking...

Is that "biologic interaction"??



CENTRE FOR FERTILITY AND HEALTH

ACTUALLY ONLY 4 BASIC STRUCTURES POSSIBLE

0 = no outcome

1 = outcome

		Z	
		0	1
X	0	1	1
	1	1	1

		Z	
		0	1
X	0	0	0
	1	1	1

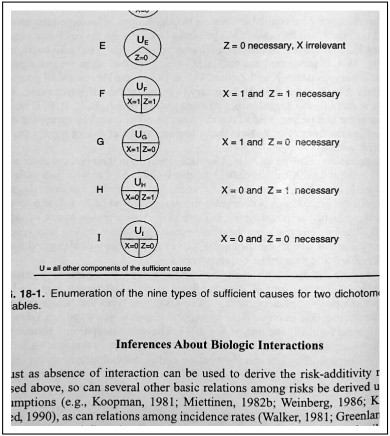
Non-interactions

		Z	
		0	1
X	0	0	1
	1	1	1

		Z	
		0	1
X	0	0	1
	1	1	0

Interactions

CAUSAL PIE (IN THE SKY)



E Z = 0 necessary, X irrelevant

F X = 1 and Z = 1 necessary

G X = 1 and Z = 0 necessary

H X = 0 and Z = 1 necessary

I X = 0 and Z = 0 necessary

U = all other components of the sufficient cause

. **18-1.** Enumeration of the nine types of sufficient causes for two dichotom
ables.

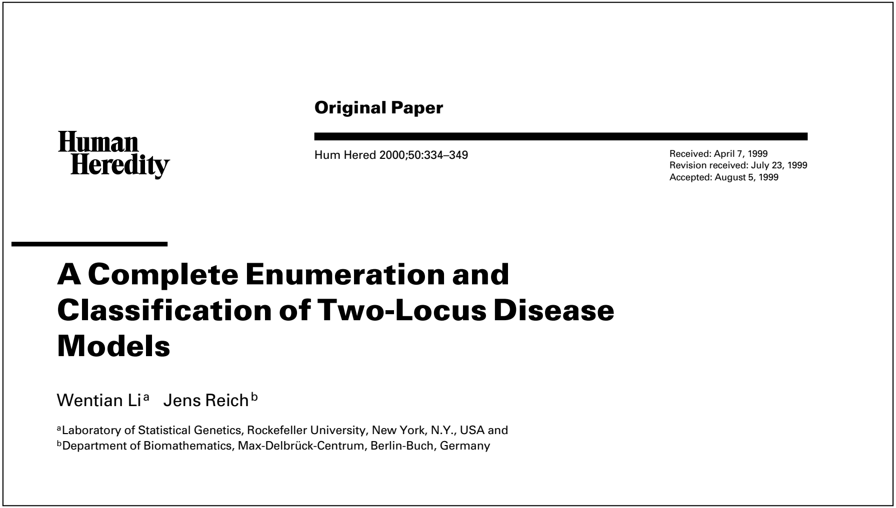
Inferences About Biologic Interactions

ust as absence of interaction can be used to derive the risk-additivity r
sed above, so can several other basic relations among risks be derived u
mptions (e.g., Koopman, 1981; Miettinen, 1982b; Weinberg, 1986; K
ed, 1990), as can relations among incidence rates (Walker, 1981; Greenlan

You will eat, bye and bye,
In that glorious land above the sky;
Work and pray, live on hay,
You'll get pie in the sky when you die

"The Preacher and the Slave" by Joe Hill

INTERACTIONS: WHO SAID IT WAS EASY?



Human Heredity

Original Paper

Hum Hered 2000;50:334–349

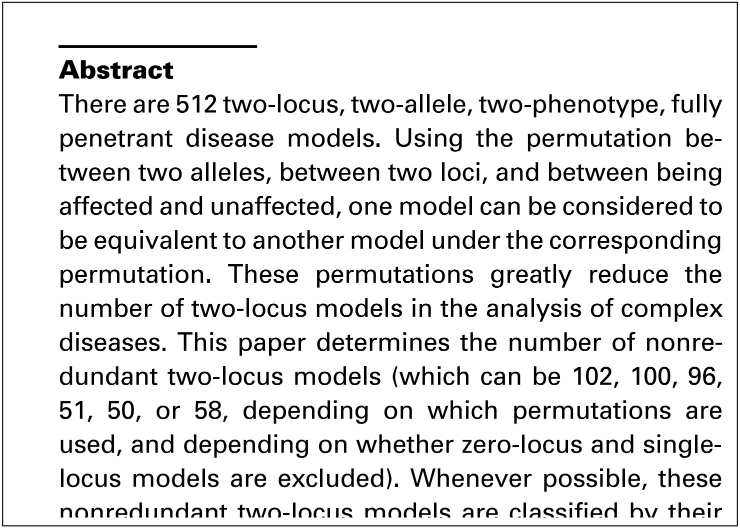
Received: April 7, 1999
Revision received: July 23, 1999
Accepted: August 5, 1999

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

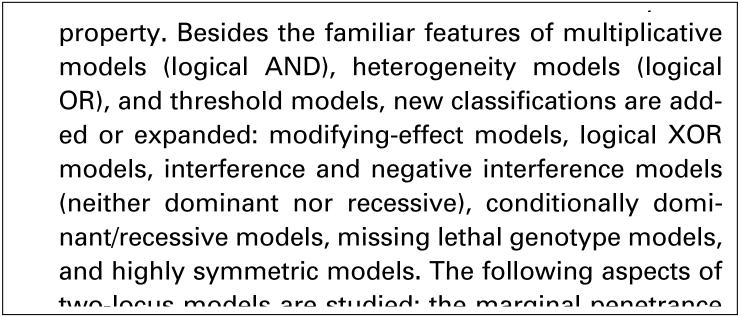
INTERACTIONS: WHO SAID IT WAS EASY?



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
	aa	0	0	0
Locus 1	aA	0	0	0
	AA	0	0	1

0 = no disease

1 = disease

Example 2:

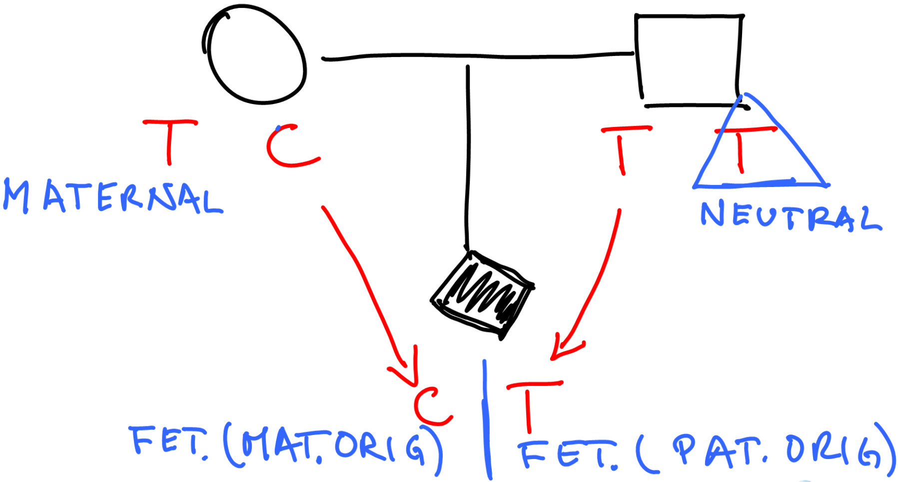
		Locus 2		
		bb	bB	BB
	aa	0	1	1
Locus 1	aA	1	1	1
	AA	1	1	1

0 = no disease

1 = disease

PARENT-OF-ORIGIN EFFECTS

CASE TRIAD



CENTRE FOR FERTILITY AND HEALTH

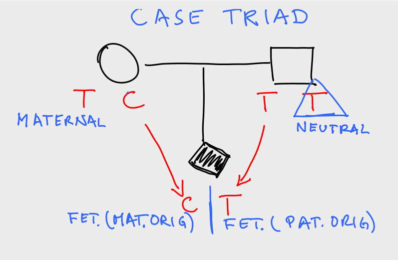
PARENT-OF-ORIGIN EFFECTS

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

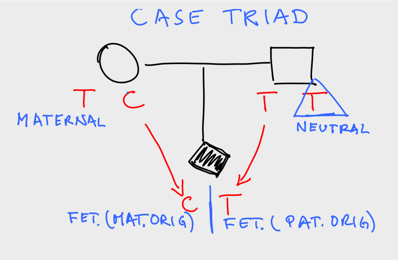
  Interaction between allele effect and parent-of-origin
- **GxE:**

  Further interaction with environment

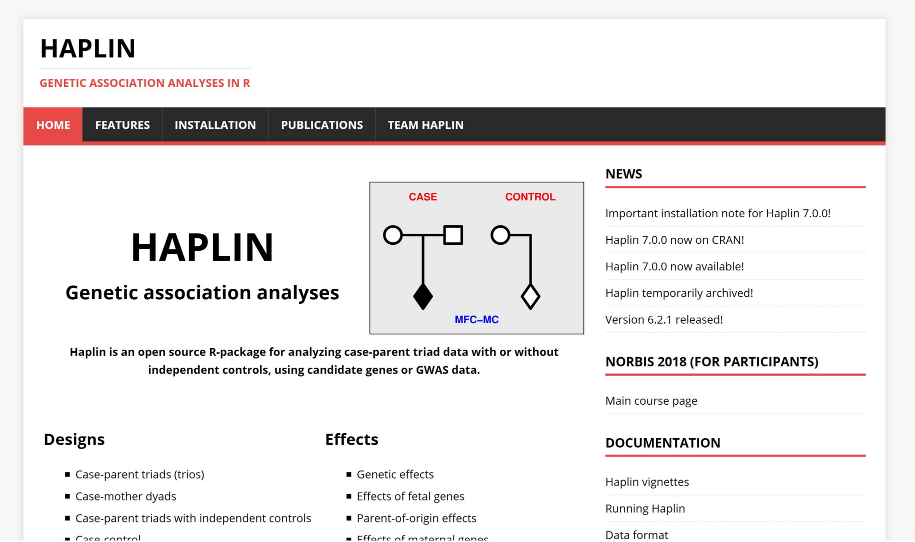
Non-smokers



Smokers



PARENT-OF-ORIGIN EFFECTS



HAPLIN

GENETIC ASSOCIATION ANALYSES IN R

HOME | FEATURES | INSTALLATION | PUBLICATIONS | TEAM HAPLIN

HAPLIN

Genetic association analyses

Haplin is an open source R-package for analyzing case-parent triad data with or without independent controls, using candidate genes or GWAS data.

Designs

- Case-parent triads (trios)
- Case-mother dyads
- Case-parent triads with independent controls
- Case-control

Effects

- Genetic effects
- Effects of fetal genes
- Parent-of-origin effects
- Effects of maternal genes

NEWS

Important installation note for Haplin 7.0.0!

Haplin 7.0.0 now on CRAN!

Haplin 7.0.0 now available!

Haplin temporarily archived!

Version 6.2.1 released!

NORBIS 2018 (FOR PARTICIPANTS)

Main course page

DOCUMENTATION

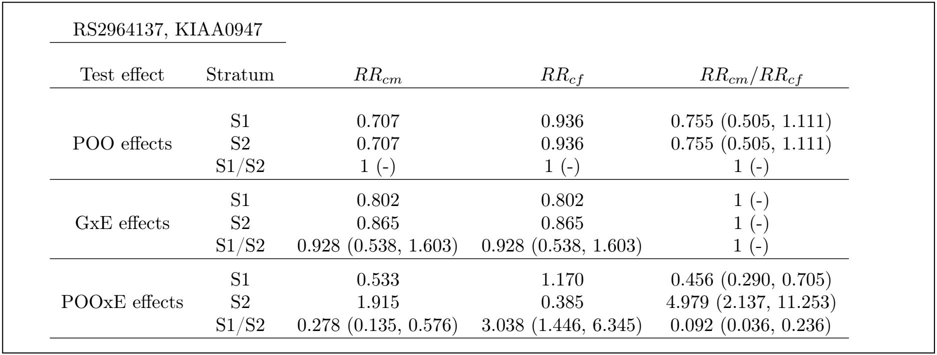
Haplin vignettes

Running Haplin

Data format

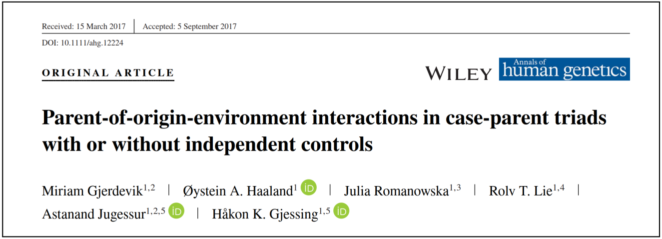
CENTRE FOR FERTILITY AND HEALTH

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

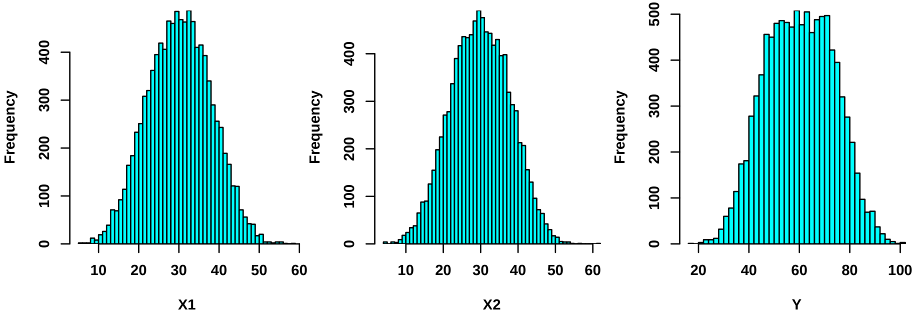
WILEY Annals of human genetics

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]

SURPRISE, SURPRISE(?)

Simulated data, $n = 10000$:



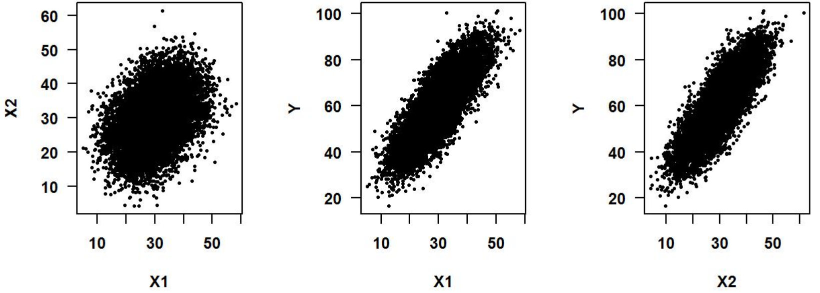
$$Y = X1 + X2 + \varepsilon$$

Data simulated without any interaction

(Cf. Prof. Magne Thoresen, UiO)

CENTRE FOR FERTILITY AND HEALTH

SURPRISE, SURPRISE(?)



Linear regression, including interaction:

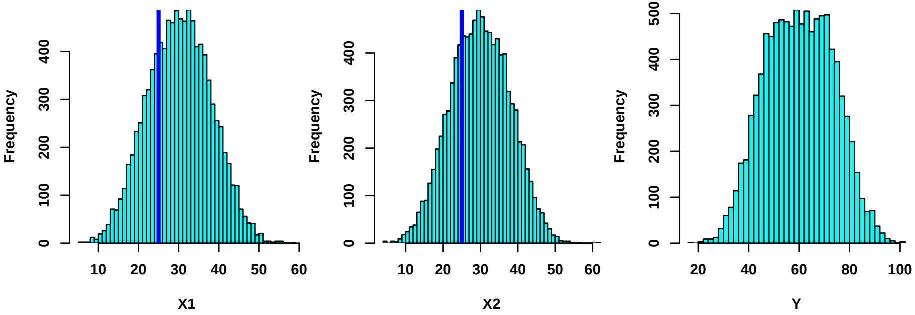
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1128420  0.3990666  -0.283    0.777
X1           1.0038665  0.0135478  74.098   <2e-16 ***
X2           1.0037252  0.0135910  73.852   <2e-16 ***
I(X1 * X2)  -0.0001540  0.0004365  -0.353    0.724
```

As expected, no sign of interaction

SURPRISE, SURPRISE(?)

Cut $X1$ and $X2$ at 25:

$$\text{X1.cat} = \begin{cases} 0 \text{ when } X1 < 25 \\ 1 \text{ when } X1 \geqslant 25 \end{cases} \qquad \text{X2.cat} = \begin{cases} 0 \text{ when } X2 < 25 \\ 1 \text{ when } X2 \geqslant 25 \end{cases}$$



SURPRISE, SURPRISE(?)

New linear regression, using the two categorical variables:

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          40.0916     0.2248  178.36   <2e-16 ***
X1.cat               11.1076     0.3067   36.22   <2e-16 ***
X2.cat               11.1162     0.3118   35.66   <2e-16 ***
I(X1.cat * X2.cat)    6.5513     0.3897   16.81   <2e-16 ***
```

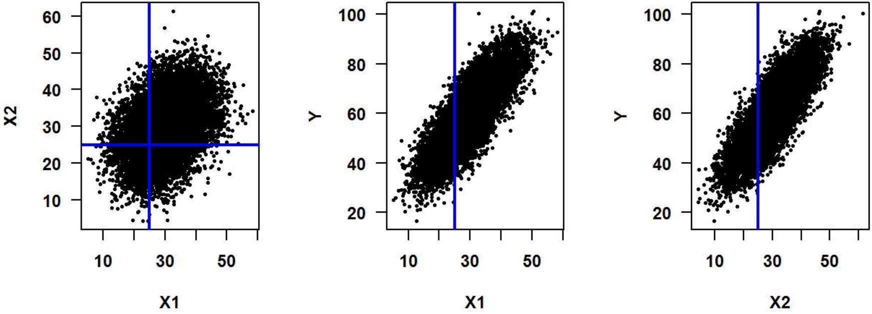
Rounded mean values of Y in categories of $X1.cat$ and $X2.cat$:

		$X2.cat$	
		0	1
$X1.cat$	0	40	51
	1	51	69

		X2.cat	
		0	1
X1.cat	0	40	40 + 11
	1	40 + 11	40 + 11 + 11 + 7

SURPRISE, SURPRISE(?)

		X2.cat	
		0	1
X1.cat	0	40	40 + 11
	1	40 + 11	40 + 11 + 11 + 7



CENTRE FOR FERTILITY AND HEALTH

CAUTION: INTERPRETATION OF MAIN EFFECTS!

$$Y = \text{X1.cat} + \text{X2.cat} + I(\text{X1.cat} * \text{X2.cat})$$

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         40.0916     0.2248  178.36   <2e-16 ***
X1.cat              11.1076     0.3067   36.22   <2e-16 ***
X2.cat              11.1162     0.3118   35.66   <2e-16 ***
I(X1.cat * X2.cat)   6.5513     0.3897   16.81   <2e-16 ***
```

$$Y = \text{X1.cat} + \text{X2.cat}$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.9119     0.1862  203.62   <2e-16 ***
X1.cat       15.1658     0.1919   79.05   <2e-16 ***
X2.cat       15.3095     0.1897   80.72   <2e-16 ***
```

CENTRE FOR FERTILITY AND HEALTH

INTERACTIONS: ALTERNATIVE CODING

Two new variables:

$$\begin{aligned} \mathrm{X2.cat0} &= (1 - \mathrm{X1.cat}) * \mathrm{X2.cat} \\ \mathrm{X2.cat1} &= \mathrm{X1.cat} * \mathrm{X2.cat} \end{aligned}$$

REGRESSION

$$\mathrm{Y} = \mathrm{X1.cat} + \mathrm{X2.cat0} + \mathrm{X2.cat1}$$

INTERACTIONS: ALTERNATIVE CODING

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  40.0916     0.2248  178.36   <2e-16 ***
X1.cat       11.1076     0.3067   36.22   <2e-16 ***
X2.cat0      11.1162     0.3118   35.66   <2e-16 ***
X2.cat1      17.6675     0.2338   75.57   <2e-16 ***
```

		X2.cat	
		0	1
X1.cat	0	40	40 + 11
	1	40 + 11	40 + 11 + 11 + 7

Interpretation:

- $\mathrm{X2.cat0}$ is the effect of X_2 when $X_1 = 0$
- $\mathrm{X2.cat1}$ is the effect of X_2 when $X_1 = 1$