

# Statistical methods in medical research

Marit B. Veierød

Department of Biostatistics

30.10.09

# Contents

- Statistics in medicine/health science
- Effect estimate, confidence interval and p-value
- Randomized clinical trials, epidemiological studies
- Multiple testing issues
- Multivariable methods



- **Application of Statistics to Medicine**
  - One of the 11 most important medical developments of the past thousand years

NEJM 2000;342:42-49



- Use of t-tests decreased from 44% in 1978-79 to 26% in 2004-05.
- Only 21% of the articles are accessible to a reader with only an introductory course in statistics (t-tests, contingency tables, nonparametric tests, epidemiologic statistics, Pearson's correlation coefficient, simple linear regression, analysis of variance, transformations and nonparametric correlations).
- >50% of the papers use relatively sophisticated statistical methods (e.g. survival analysis, multiple regression)

# The American Statistician

2007, vol. 61, no. 1, pp. 47 - 55

## The Use of Statistics in Medical Research: A Comparison of *The New England Journal of Medicine* and *Nature Medicine*

Alexander M. Strasak; Qamruz Zaman; Gerhard Marinell; Karl P. Pfeiffer; Hanno Ulmer

The results of the present study give up-to-date evidence for the expanding use of inferential statistics in modern medical research. Compared with the early results of Emerson and Colditz (1983) for *The New England Journal of Medicine*, a vast increase in the application of inferential methods could be outlined, as the proportion of papers having analytical character more than doubled from 42.0 percent in 1983 to 94.5 percent in 2004. In particular, methods of Survival Analysis, which were virtually absent in 1983, are now in common use in this journal as also recently documented by Horton and Switzer (2005).

Although complexity of statistical analyses in *The New England Journal of Medicine* continued to increase since 1983 and was rather sophisticated in 2004, this does not necessarily hold for papers from *Nature Medicine*, which gives reason to hypothesize that basic science sticks with basic analyses. However, the majority of research reports published in *Nature Medicine* are based on animal studies, whereas *The New England Journal of Medicine* predominately publishes studies on human subjects.

# Why is statistics so important in medical research?

- Statistics provides the optimal methods for designing studies to gain knowledge, and for making inferences from limited empirical data.
- Important: Statistics contribute to the whole research design not merely the statistical analysis

Medical ethics and statistics. Encyclopedia of biostatistics, vol. 4.



## C-vitaminer beskytter ikke mot forkjølelse

C-vitaminer beskytter ikke mot forkjølelse, slår tyske forskere fast.

IQWiG har analysert en rekke forsøk som er utført av Cochrane Collaboration i Storbritannia, der over 11.000 forsøkspersoner har spist C-vitaminpiller.

Forsøkspersonene som begynte å ta høye doser med C-vitaminer straks en forkjølelse meldte seg, ble verken mindre syke eller raskere friske enn dem som spiste såkalte placebopiller som ikke inneholdt C-vitaminer.

De som spiste høye doser med C-vitaminer alt før de ble forkjølet, reduserte sykdomstiden med under en dag på årsbasis. IQWiG advarer derfor folk mot å knaske C-vitaminer og gjør oppmerksom på at den anbefalte dosen per dag bare er på 0,1 gram.



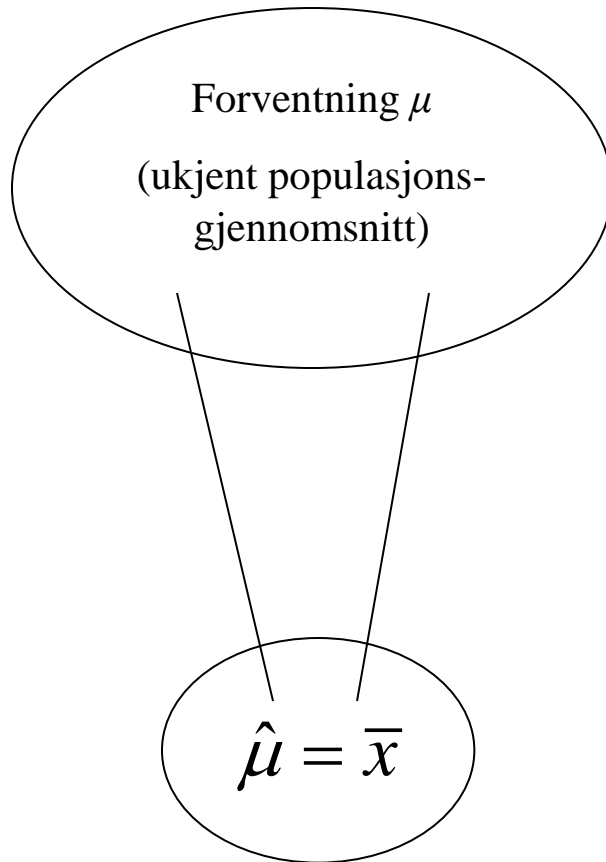
**Hvordan estimere antall  
forkjølelser i løpet av 12 mnd  
blant norske menn og kvinner i  
alder 20-70 år?**

**Spørre alle?**



# Populasjon og utvalg

Jakten på 'lille-my'



## Populasjon:

Alle norske menn og kvinner  
i alderen 20-70 år

## Tilfeldig utvalg:

$n$  tilfeldig valgte menn og kvinner  
i alderen 20-70 år

$X_1, X_2, \dots, X_n$  er ant. forkjølelser i løpet av 12 mnd. for de  $n$  personene i utvalget

- Punktestimering
- Konfidenzintervall estimering
- Hypotesetesting

# Randomized controlled clinical trials (RCTs)

Studies of a medical intervention in which the allocation of patients to the various experimental groups occurs by a chance mechanism.

# Again: Statistical aspects – not only the data analysis!

1. Aim of the study
2. Hypotheses to be tested
3. Design of the study
4. Patient selection criteria
5. Methods for patient evaluation
6. Randomization
7. Required size of study
8. Protocol deviations
9. Plan for statistical analysis
10. Presentation of results/interpretation of trial findings

# RCT

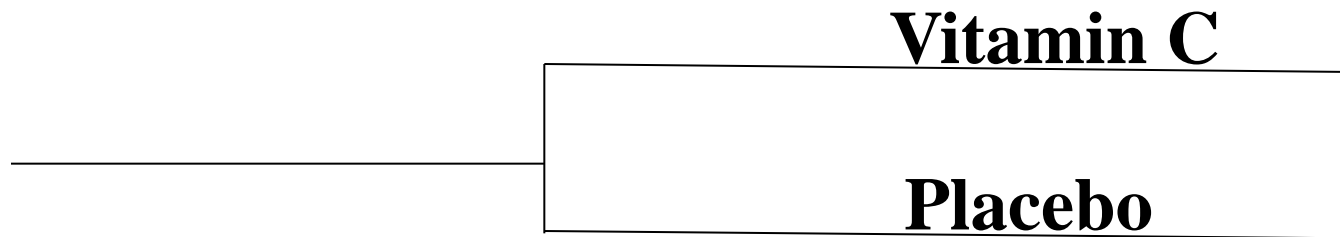
## Forebygger vitamin C forkjølelse?

- Formål
  - Undersøke om behandling med vitamin C har forskjellig effekt fra behandling med placebo på en bestemt sykdom.

Formål  $\longleftrightarrow$  Hypotese

# Parallelgruppe design

Pasientene randomiseres til to grupper



For hver pasient registreres antall forkjølelser i løpet av 12 mnd

# Nullhypotesen

Klinisk forsøk for sammenligning av effekt av to behandlinger i to grupper (her vitamin C og placebo).

$H_0$ : Ingen forskjell i effekt mellom gruppene

$H_A$ : Forskjellig effekt i gruppene

(Tosidig test)

Vi skal utføre en **statistisk test** for å se om våre data tilsier at vi med høy grad av sikkerhet kan forkaste  $H_0$ . Hvis det vi observerer er en tilstrekkelig lite rimelig observasjon dersom  $H_0$  var sann, forkastes  $H_0$  og vi godtar  $H_1$ .



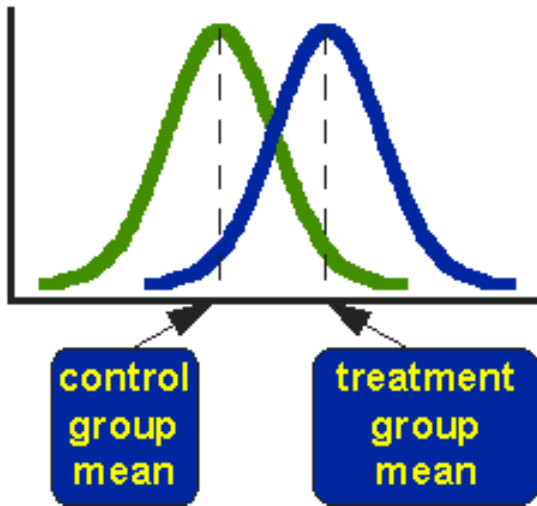
- $X_{11}, X_{12}, \dots, X_{1n_1}$  is a random sample from population 1 with mean  $\mu_1$  and standard deviation  $\sigma_1$ .
- $X_{21}, X_{22}, \dots, X_{2n_2}$  is a random sample from population 2 with mean  $\mu_2$  and standard deviation  $\sigma_2$ .
- The two samples are independent and have equal variance ( $\sigma_1 = \sigma_2$ )

# Null hypothesis

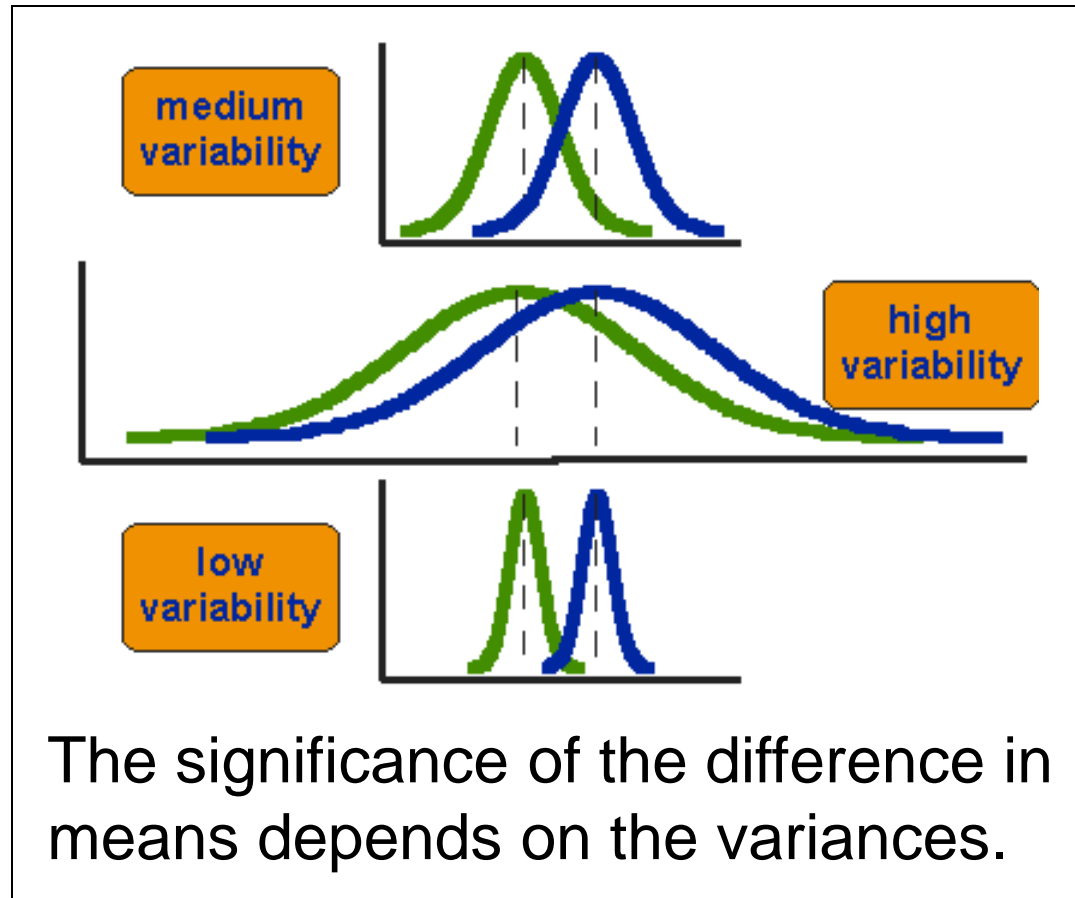
$\mu_1$  and  $\mu_2$  are the unknown means (population means) for the active (vitamin C) and the placebo group

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$



Are the means different?



The significance of the difference in means depends on the variances.

# Test statistic

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s_f \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_f = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Use the Student distribution to find the p-value

# p-value

- The probability to obtain a result at least as extreme as the actually observed one, if the null hypothesis were true
- The null hypothesis is rejected if the p-value is very small.
- The *significance level* is the threshold under which the p-values is considered as small. Often 0.05, or 0.01. If a null hypothesis is rejected, we say the result is significant.

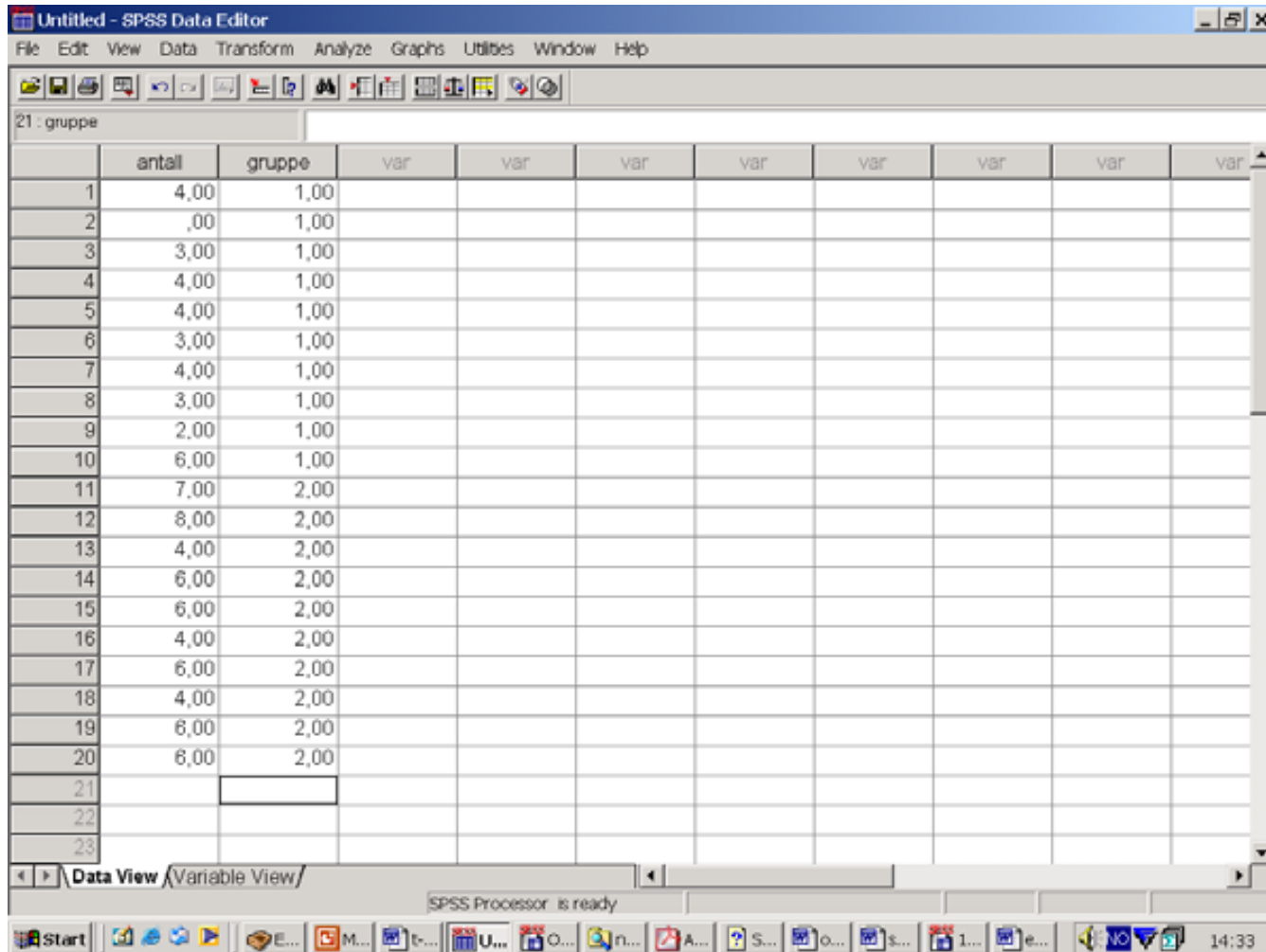
# Signifikansnivå 0.05

- $p \leq 0.05$ : Forkaster  $H_0$  og konkluderer at det er grunnlag for å påstå at de to behandlingene har forskjellig effekt
- $p > 0.05$ : Forkaster ikke  $H_0$ . Det er ikke grunnlag for å påstå at de to behandlingene har forskjellig effekt.

By randomization, 10 subjects are given vitamin C capsules and 10 subjects are given placebo capsules. The number of colds over a 12-month period was recorded.

<b>Vitamin C</b>		<b>Placebo</b>	
$i$	$x_{i1}$	$i$	$x_{i2}$
1	4	1	7
2	0	2	8
3	3	3	4
4	4	4	6
5	4	5	6
6	3	6	4
7	4	7	6
8	3	8	4
9	2	9	6
10	6	10	6

# Statistisk analyse gjøres feks i SPSS



The screenshot shows the SPSS Data Editor interface. The main window displays a data table with the following data:

	antall	gruppe	var	var	var	var	var	var	var	var
1	4,00	1,00								
2	,00	1,00								
3	3,00	1,00								
4	4,00	1,00								
5	4,00	1,00								
6	3,00	1,00								
7	4,00	1,00								
8	3,00	1,00								
9	2,00	1,00								
10	6,00	1,00								
11	7,00	2,00								
12	8,00	2,00								
13	4,00	2,00								
14	6,00	2,00								
15	6,00	2,00								
16	4,00	2,00								
17	6,00	2,00								
18	4,00	2,00								
19	6,00	2,00								
20	6,00	2,00								
21										
22										
23										

The interface includes a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Window, Help), a toolbar, and a status bar at the bottom showing 'SPSS Processor is ready' and the system clock '14:33'.



# T-test

Vitamin C:  $\bar{X}_1 = 3.3$

Placebo:  $\bar{X}_2 = 5.7$

$$\bar{X}_1 - \bar{X}_2 = -2.4$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s_f \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$T = -3.68$ ,  $p = 0.002$   
i.e. reject  $H_0$ .

# Confidence interval

Confidence interval for  $\mu_1 - \mu_2$ :

$$\bar{X}_1 - \bar{X}_2 \pm cS_f \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Vitamin C study:

95% confidence interval is (-3.8, -1.0)

# Test – konfidensintervall

$$H_0 : \mu = a \text{ mot } H_A : \mu \neq a$$

Hvis 95% konfidensintervallet for  $\mu$  ikke dekker  $a$  kan vi forkaste  $H_0$  på 5% nivå.

# Presentasjon av resultater

1. Effektmål
2. 95% konfidensinterval (95% KI)
3. P-verdi

Vårt eksempel: Vi fant en signifikant effekt av vitamin C på antall forkjølelser ( $p=0.002$ ). De som tok vitamin C hadde gjennomsnittlig 2.4 færre forkjølelser enn de tok placebo (95% KI(1.0, 3.8)).

## *Can I use the t-distribution? Assumptions.*

---

Observations (individual expressions) should be

- independent
- normal distributed (at least approximately)

The more subjects, the less important is individual normality, by the central limit theorem.

If few subjects, and we cannot assume normality then

- transform data (log scale etc.)
- Use other tests, that do not require normality; for example non-parametric tests (Wilcoxon-Mann Whitney)

# Oppgave 1

$$n_1 = 13, \bar{x}_1 = 8.066, s_1 = 1.238$$

$$n_2 = 9, \bar{x}_2 = 10.298, s_2 = 1.398$$

$$\bar{x}_1 - \bar{x}_2 = -2.23$$

$$\begin{aligned} & \bar{x}_1 - \bar{x}_2 \pm c s_f \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\ &= 8.066 - 10.298 \pm 2.086 \cdot 1.3044 \cdot \sqrt{\frac{1}{13} + \frac{1}{9}} \\ &= (-3.41, -1.05) \end{aligned}$$

$$\begin{aligned} T &= \frac{\bar{X}_1 - \bar{X}_2}{s_f \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \\ &= \frac{8.066 - 10.298}{1.3044 \cdot \sqrt{\frac{1}{13} + \frac{1}{9}}} = -3.95 \end{aligned}$$

$$p = 0.001$$

**Table 9.4** 24 hour total energy expenditure (MJ/day) in groups of lean and obese women (Prentice *et al.*, 1986)

Lean ( $n = 13$ )	Obese ( $n = 9$ )
6.13	8.79
7.05	9.19
7.48	9.21
7.48	9.68
7.53	9.69
7.58	9.97
7.90	11.51
8.08	11.85
8.09	12.79
8.11	
8.40	
10.15	
10.88	

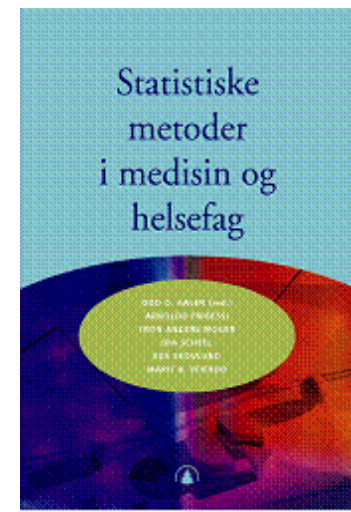
Hvilken hypotese testes her? Hvilke resultater vil du rapportere? Konklusjon?

# Lese mer om sammenligning av gjennomsnitt og t-tester?

s. 362



s. 177



# Signifikansnivå

- Sannsynligheten for å forkaste nullhypotesen selv om den er sann
- Signifikansnivå 5%: Vi forventer at 1 av 20 tester er signifikant selv om det ikke er forskjell på gruppene.

**Falskt positivt funn – forkaste nullhypotesen selv om den er sann**



# Et annet eksempel

En klinisk studie av 4 medisiner blant pasienter med høyt blodtrykk. Systolisk og diastolisk blodtrykk måles før, under og etter en standard fysisk test. Målingene ble gjort ukentlig i en fire måneders periode. Pasientene kan klassifiseres i undergrupper etter alder, kjønn og blodtrykk ved studiestart. Interim analyser kan gjøres etter at hver 20 pasient er vurdert. .... Uendelig mange hypoteser kan testes! Falske positive funn?

# Multiplisitet

- **Flere behandlinger**
- **Multiple endepunkter**
- **Repeterte målinger**
- **Subgruppeanalyser**
- **Interimanalyser**
- **Genetiske studier**

Dess flere signifikanstester, dess større sannsynlighet for falskt positivt funn

# Flere behandlinger

# Sammenligning av >2 grupper

- Eksempel: Vekten til et organ for 30 rotter etter behandling med 4 dietter: 6 rotter ble gitt Diet1, 8 rotter Diet2, 9 rotter Diet3 og 7 rotter Diet4.

Datafilen:

Gr	Vekt
1	4.34
1	4.73
1	4.84
1	4.57
1	4.72
1	4.55
2	4.47
2	4.65
2	4.62
2	4.41
2	4.43
2	4.23
2	4.54
2	4.45
3	4.72
3	4.99
3	5.24
3	5.00
3	4.82
3	4.95
3	5.28
3	4.90
3	4.98
4	4.48
4	5.02
4	4.58
4	4.89
4	4.90
4	4.81
4	5.26

# Variansanalyse

$H_0$ : Det er ingen forskjell mellom gruppene

$H_A$ : Det er forskjell mellom mins to av gruppene

Antar

- Observasjonene innen hver gruppe er normalfordelte
- Observasjonene er uavhengige innen og mellom de  $k$  gruppene
- Lik varians i alle gruppene

# Variansanalyse forts.

group	Mean	N	Std. Deviation	Minimum	Maximum
1,00	4.6250	6	.17672	4.34	4.84
2,00	4.4750	8	.13245	4.23	4.65
3,00	4.9867	9	.17965	4.72	5.28
4,00	4.8486	7	.26220	4.48	5.26

Det er signifikant forskjell på gruppene,  $p < 0.001$ .

## Multiple sammenligninger:

Gruppe 1 og gruppe 2

Gruppe 1 og gruppe 3

Gruppe 1 og gruppe 4

Gruppe 2 og gruppe 3

Gruppe 2 og gruppe 4

Gruppe 3 og gruppe 4

Bonferroni korreksjon: signifikansnivå  $0.05/6$  (SPSS gir  $p \cdot 6$ )

# Multiple endepunkter

Diskuterer vi i artikkelen fra NEJM senere i dag.

s. 264

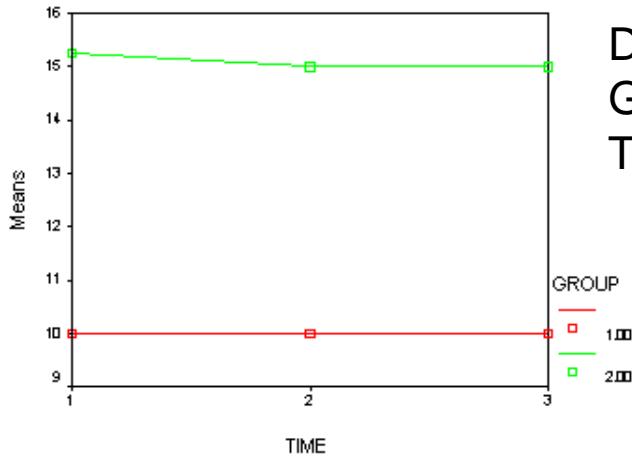


# Repeterte målinger



# Repeated measurements

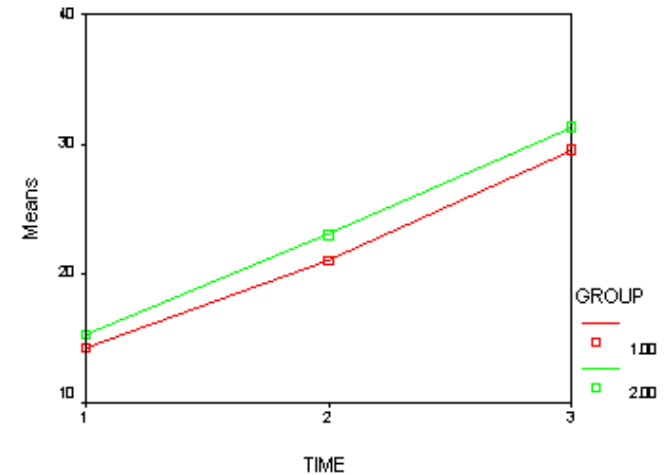
Demo1



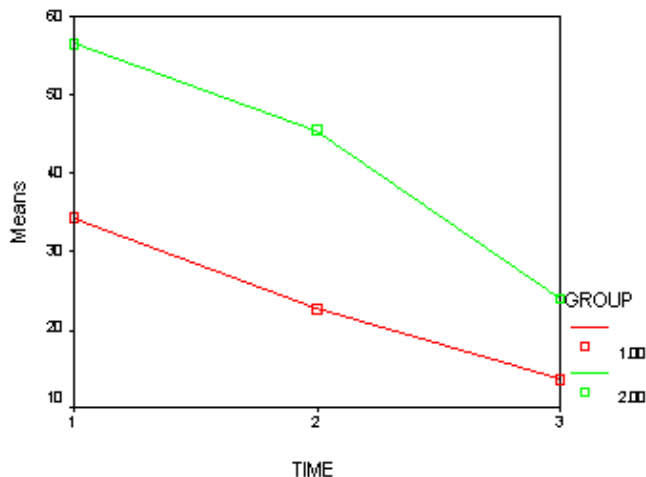
Demo 1:  
Group significant  
Time not significant

Demo 2:  
Group not significant  
Time significant

Demo2

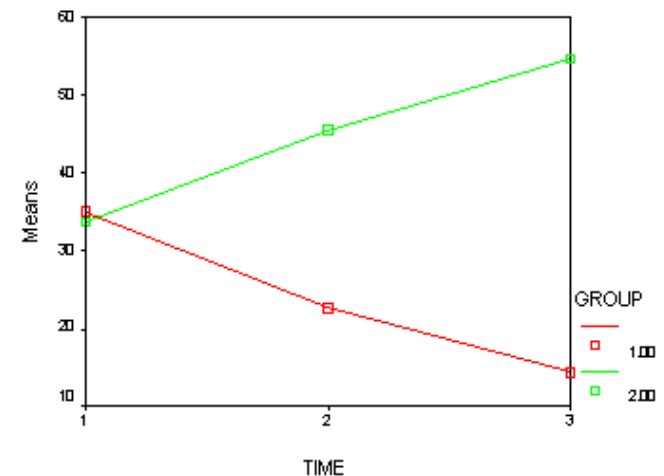


Demo3



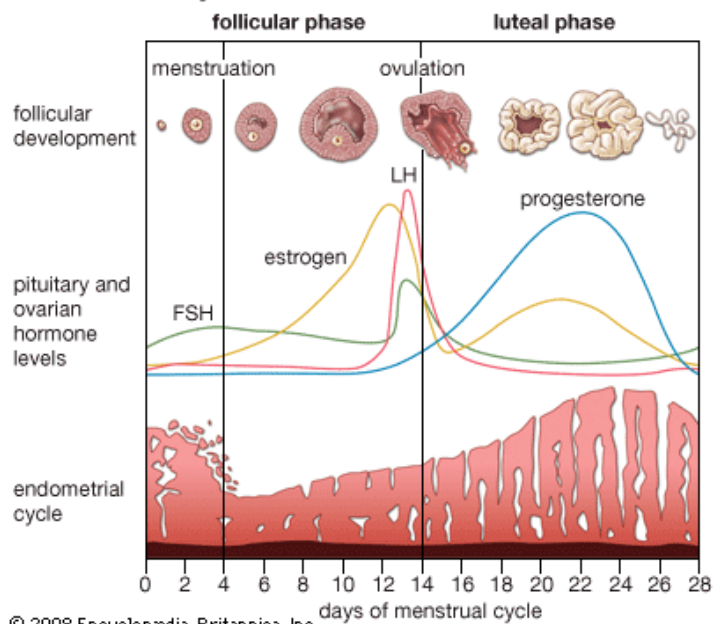
Demos 3 and 4:  
Time\*group significant.  
The groups are changing  
over time but are changing  
in different ways

Demo4



# Repeated measurements

## The menstrual cycle



- Energibalanse fra fødsel og gjennom livet kan være forbundet med økt østrogen og økt brystkreftrisiko.
- Ung alder ved menarche gir økt kumulativ hormon eksponering som kan ha betydning for senere utvikling av brystkreft
- Alder ved menarche og overvekt er risikofaktorer for brystkreft

17  $\beta$ -estradiol målt i daglige spyttprøver gjennom en hel menstruasjonsyklus. n=204 friske norske unge kvinner.

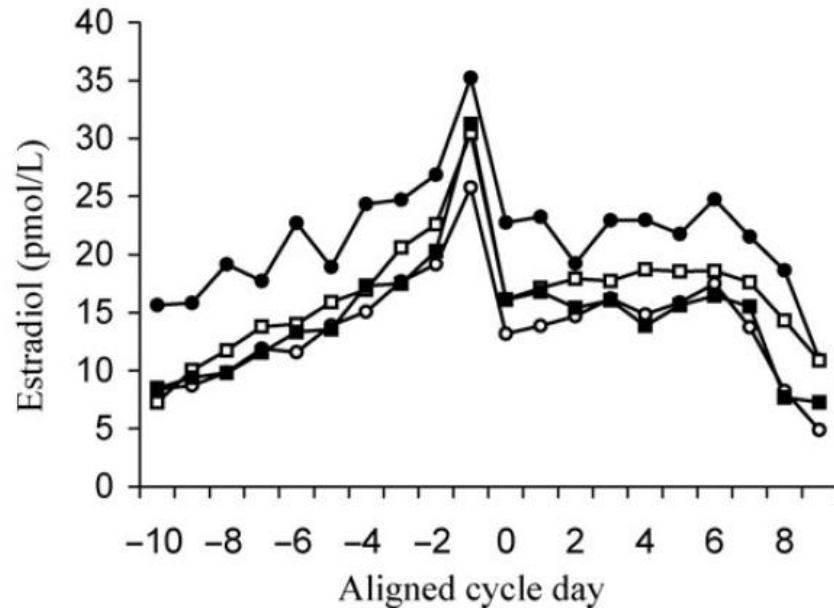
## 17- $\beta$ -Estradiol in relation to age at menarche and adult obesity in premenopausal women

A. Emaus<sup>1,10</sup>, S. Espetvedt<sup>1</sup>, M.B. Veierød<sup>2</sup>, R. Ballard-Barbash<sup>3</sup>, A.-S. Furberg<sup>4,5</sup>, P.T. Ellison<sup>6</sup>, G. Jasienska<sup>7</sup>, A. Hjartaker<sup>8</sup> and I. Thune<sup>1,9</sup>

**Human Reproduction Vol.23, No.4 pp. 919–927, 2008**

# Linear mixed models for repeated measurements.

**D** Groups of WHR and age at menarche



WHR=Waist hip ratio

- 1: Menarche  $\leq 12$  years and WHR  $< 0.8$
- 2: Menarche  $\leq 12$  years and WHR  $\geq 0.8$
- 3: Menarche  $> 12$  years and WHR  $< 0.8$
- 4: Menarche  $> 12$  years and WHR  $\geq 0.8$

Tidlig alder ved første menstruasjon sammen med overvekt gir høye daglige nivåer av estradiol hos unge kvinner.

# Subgruppeanalyser

Separate analyser i undergrupper -  
mange p-verdier - økt risiko for falske  
positive funn

s. 265



# Interimanalyser

s. 266



# Interimanalyse

- Dersom det planlagte antall pasienter er høyt kan det være aktuelt å analysere dataene én el. flere ganger underveis i forsøket. NB interimanalyser må planlegges på forhånd.
- Dersom hver test utføres på 5% nivå, vil den totale sannsynligheten for falske positive funn (å forkaste  $H_0$  selv om den er sann) bli vesentlig høyere enn de ønskede 5%.
- Reduserer nivået for hver analyse.

# Genetiske studier

published online 6 September 2009; doi:10.1038/ng.440

---

# Genome-wide association study identifies variants at *CLU* and *PICALM* associated with Alzheimer's disease

Denise Harold<sup>1,45\*</sup>, Richard Abraham<sup>1,45</sup>, Paul Hollingworth<sup>1,45</sup>, Rebecca Sims<sup>1</sup>, Amy Gerrish<sup>1</sup>, et al.

3941 cases, 7848 controls



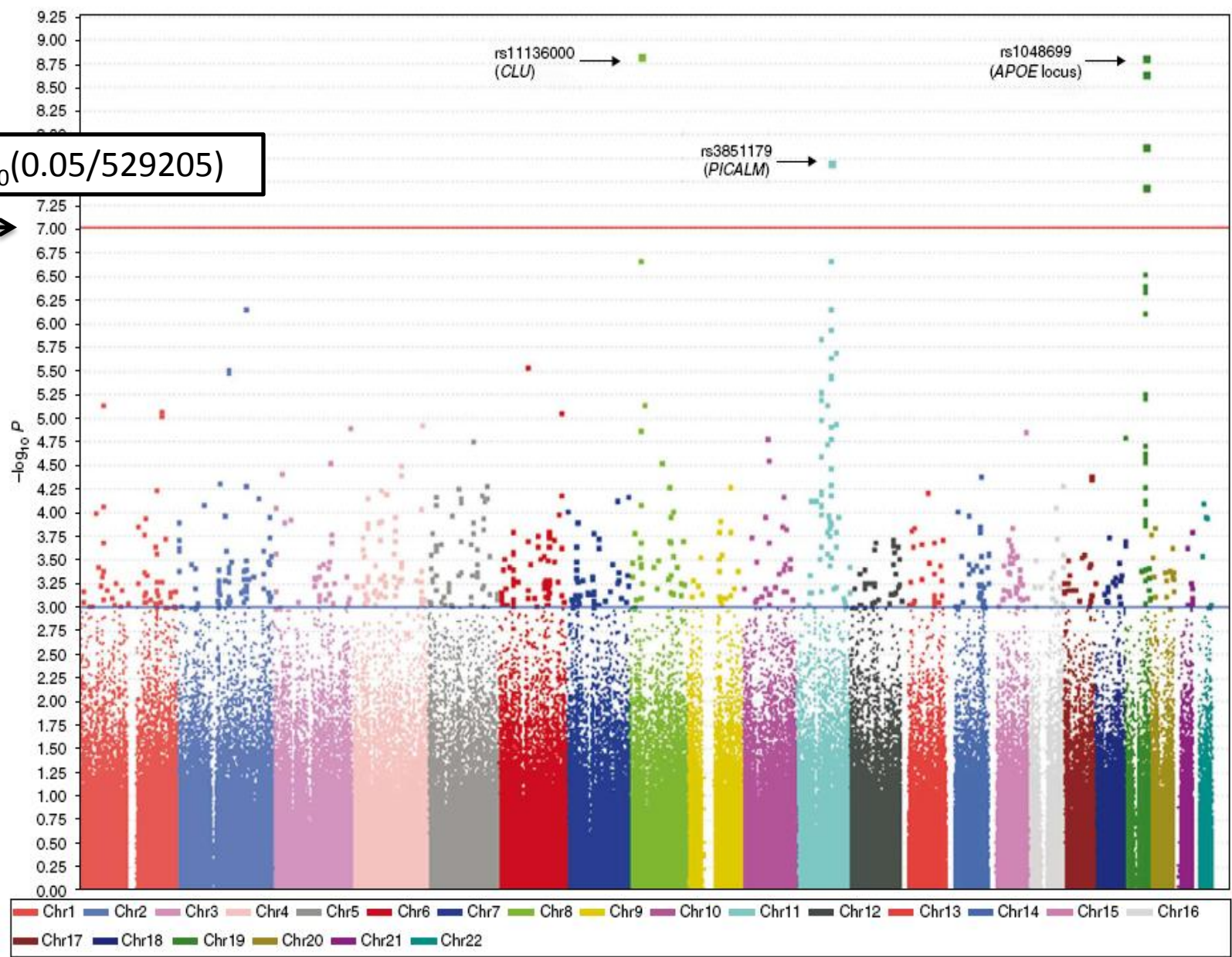
# SNP analysis

- 529,205 SNPs from the human genome.
- Looking for differences between Alzheimer patients and controls.
- SNP: DNA sequence variation when a single nucleotide (G, T, A or C) differs between case and control
- Comparing patients with Alzheimer's disease (cases) and control patients without the disease

# SNP analysis cont.

- Many tests – many false positive results.
- The analysis was based on 529,205 SNPs and a conservative genome-wide significance threshold of  $0.05/529205 = 9.4 \times 10^{-8}$  was used.

$-\log_{10}(0.05/529205)$



**Figure 1** Scatterplot of chromosomal position (*x* axis) against  $-\log_{10}$  GWAS *P* value (*y* axis). The *y*-axis scale has been limited to 9.25 ( $P = 5.6 \times 10^{-10}$ ), although highly significant association was observed with SNPs in the vicinity of the *APOE* locus (for example, rs2075650 with  $P = 1.8 \times 10^{-157}$ ). The threshold for genome-wide significance ( $P \leq 9.4 \times 10^{-8}$ ) is indicated by the red horizontal line. 761 SNPs with  $P \leq 1 \times 10^{-3}$  lie above the blue horizontal line and are listed in **Supplementary Table 2**. The plot was produced using Haploview version 4.0 (ref 49.).

# Regresjonsmodeller

**Tabell 3.1** Regresjonsmodeller mht. målenivå på responsvariabel, forklaringsvariabler og effektestimater

Responsvariabel	Forklaringsvariabler	Regresjonsmetode	Effektestimater
<b>Kontinuerlige data</b>	Kontinuerlige eller kategoriske	Lineær regresjon	Regresjonskoeffisienter, $\hat{\beta}$
<b>Nominale data</b>			
To kategorier	Kontinuerlige eller kategoriske	Logistisk regresjon	Odds ratio, $OR = e^{\hat{\beta}}$
Flere enn to kategorier	Kontinuerlige eller kategoriske	Multinomisk logistisk regresjon	Odds ratio, $OR = e^{\hat{\beta}}$
<b>Ordinale data</b>	Kontinuerlige eller kategoriske	Ordinal logistisk regresjon	Odds ratio, $OR = e^{\hat{\beta}}$
<b>Telledata</b>	Kategoriske eller kontinuerlige	Poisson-regresjon	Insidensrate ratio, $IRR = e^{\hat{\beta}}$
<b>Varighetsdata</b>	Kontinuerlige eller kategoriske	Cox-regresjon	Hasard ratio, $HR = e^{\hat{\beta}}$

# Lineær regresjon

SBP                      Systolisk blodtrykk, mmHg  
ALDER                    År  
KMI                        Kroppsmasseindeks , kg/m<sup>2</sup>  
RØYKING                0=ikke røyker, 1= røyker

32 individer

ID	SBP	ALDER	KMI	RØYKING
1	135	45	18.18	0
2	122	43	20.48	0
3	130	49	19.53	0
4	148	52	23.74	0
5	146	54	18.77	1

Eksempel 3.1 I Veierød og Laake. Regresjonsmodeller og analyse av sammenheng mellom eksponering og sykdom. I Laake, Hjartåker, Thelle, Veierød (ed). Epidemiologiske og kliniske forskningsmetoder.

Dataene kan lastes ned fra [www.forskningsmetode.no](http://www.forskningsmetode.no)

Regresjonsmodell for sammenhengen mellom alder og blodtrykk:

$$SBP = \beta_0 + \beta_1 \cdot ALDER$$

Tabell 3.2 Regresjonsanalyse av systolisk blodtrykk mht. alder

	Regression Coefficients		t	p-value	95% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
Constant	73,615	14,967	4,919	,000	43,048	104,182
alder	1,342	,281	4,779	,000	,768	1,916

Estimerte regresjonskoeffisient for alder er 1.342 ( $\hat{\beta}_1$ ). Systolisk blodtrykk øker i gjennomsnitt 1.34 mmHG for hver 1 års økning i alder (95% KI (0.77, 1.92)). P-verdien er under 0.05 og vi konkluderer at det er signifikant effekt av alder på systolisk blodtrykk.

Regresjonsmodell for sammenhengen mellom alder, kroppsmasseindeks, røyking og blodtrykk:

$$SBP = \beta_0 + \beta_1 \cdot ALDER + \beta_2 \cdot KMI + \beta_3 \cdot RØYKING$$

Tabell 3.6 Regresjonsanalyse av systolisk blodtrykk mht. alder, kroppsmasseindeks og røyking

	Regression Coefficients		t	p-value	95% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
Constant	48,387	11,351	4,263	,000	25,135	71,639
alder	,684	,313	2,186	,037	,043	1,324
kmi	2,476	,664	3,729	,001	1,116	3,836
røyking	10,625	2,859	3,716	,001	4,768	16,482

Regresjonskoeffisienten for alder er nå 0.684 dvs. når vi justerer for kroppsmasseindeks og røyking, reduseres effekten av alder på systolisk blodtrykk noe. Effekten er fortsatt signifikant. Det er også signifikant effekt av kroppsmasseindeks og røyking.

**Tabell 3.6** Regresjonsanalyse av systolisk blodtrykk mht. alder, kroppsmasseindeks og røyking

	Regression Coefficients		t	p-value	95% Confidence Interval for B	
	B	Std. Error			Lower Bound	Upper Bound
Constant	48,387	11,351	4,263	,000	25,135	71,639
alder	,684	,313	2,186	,037	,043	1,324
kmi	2,476	,664	3,729	,001	1,116	3,836
røyking	10,625	2,859	3,716	,001	4,768	16,482

Effekten av røyking:

Systolisk blodtrykk øker i gjennomsnitt 10.625 mmHG (95% KI (4.768, 16.482)) når vi går fra ikke-røykere til røykere. Det er signifikant sammenheng mellom røyking og blodtrykk ( $p=0.001$ ). Det er justert for alder og kroppsmasseindeks.



# Oppgave 2

Vickers and Altman

*BMJ* 2001;323:1123-4

As an illustration, Kleinhenz et al randomised 52 patients with shoulder pain to either true or sham acupuncture.<sup>4</sup> Patients were assessed before and after treatment using a 100 point rating scale of pain and function, with lower scores indicating poorer outcome. There was an imbalance between groups at baseline, with better scores in the acupuncture group (see table).

Pain scores (mean and SD)

	Placebo group (n=27)	Acupuncture group (n=25)
Baseline	53.9 (14)	60.4 (12.3)
Analysis		
Follow up	62.3 (17.9)	79.6 (17.1)
Change score*	8.4 (14.6)	19.2 (16.1)

follow up score =

$$24 + 0.71 \times \text{baseline score} + 12.7 \times \text{group}$$

Dette er en lineær regresjonsmodell. Vi er interessert i regresjonskoeffisienten for gruppe (placebo=0, akupunktur=1) som er 12.7. 95% KI er (4.1, 21.3) og  $p=0.005$ .

Hvordan tolke dette resultatet?



Onsdag, 18.11.09 kl. 10:31



## **B-vitaminet folsyre kan gi økt kreftrisiko**

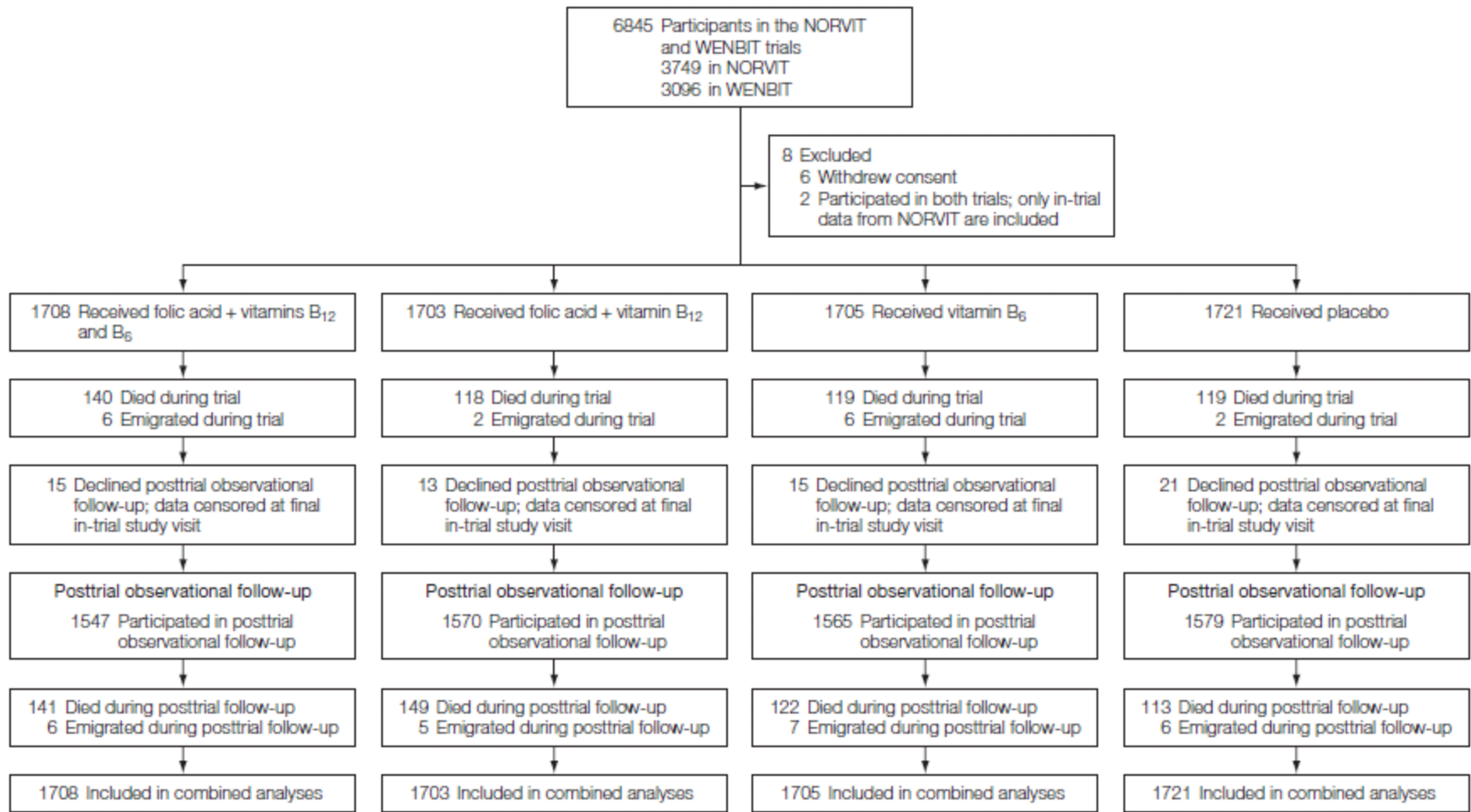
Norsk forskning viser at behandling med folsyre er forbundet med økt risiko for kreft og død blant hjertepasienter.

Data fra to dobbel blinde, placebo kontrollerte, parallell gruppe studier av folsyre, vitamin B og forekomst og dødelighet av koronar hjertesykdom:

Norwegian Vitamin [NORVIT] Trial

Western Norway B Vitamin Intervention Trial [WENBIT]

**Figure.** Flow of Participants Through the NORVIT and WENBIT Trials and Posttrial Observational Follow-up



NORVIT Indicates Norwegian Vitamin Trial; WENBIT, Western Norway B Vitamin Intervention Trial.

# Dataene ble analysert bl.a. i SPSS

Eksempel på kreftinsidens data med oppfølging ut 2002  
(jeg har ikke dataene fra JAMA artikkelen).

ID	INKLUSJONSDATO	SLUTT DATO	PERSON ÅR	KREFT	LUNGE KREFT	PROSTATA KREFT	.	.
1	01011990	31122002	12.91	0	0	0	.	.
2	12061992	18071998	6.07	1	0	1	.	.
3	03031993	25102002	9.59	1	1	0	.	.
4	28101994	13042001	6.43	0	0	0	.	.
5	01021995	31122002	7.83	0	0	0	.	.
.	.	.	.	.	.	.	.	.

# Insidensrate

## Mål for sykdomsforekomst

$$\textit{insidensrate} = \frac{\textit{antall nye tilfeller}}{\textit{totalt antall personår under risiko}}$$

# Effektmål: relativ insidensrate

$$\frac{\textit{insidensraten blant de eksponerte}}{\textit{insidensraten blant de ikke – eksponerte}}$$

Fortolkes som en relativ risiko (*RR*)

# Cox regresjon

- Analyse av levetidsdata
- Antar at effekten av risikofaktorene på levetiden er konstant over tid
- Regresjonskoeffisienten angir hazard rate ratio (HR). Kan fortolkes som relative risiko (RR)

# Results. Ebbing et al. JAMA 2009

**Table 2.** Cancer and Mortality Outcomes and Hazard Ratios<sup>a</sup>

Outcome	Total No.	No. of Cases (Rate per 1000 Observation-Years)				HR (95% CI)	
		Folic Acid + Vitamins B <sub>12</sub> and B <sub>6</sub> (n = 1708)	Folic Acid + Vitamin B <sub>12</sub> (n = 1703)	Vitamin B <sub>6</sub> (n = 1705)	Placebo (n = 1721)	Folic Acid vs Non-Folic Acid Groups	Vitamin B <sub>6</sub> vs Non-Vitamin B <sub>6</sub> Groups
Cancer incidence <sup>b</sup>	629	172 (17.3)	169 (16.7)	151 (14.9)	137 (13.3)	1.21 (1.03-1.41)	1.07 (0.92-1.26)
Colorectal cancer	95	25 (2.5)	22 (2.2)	26 (2.6)	22 (2.1)	1.00 (0.59-1.69) <sup>c</sup>	1.18 (0.69-2.00) <sup>c</sup>
Lung cancer	92	31 (3.1)	25 (2.5)	16 (1.6)	20 (1.9)	1.59 (0.92-2.75) <sup>c</sup>	1.06 (0.62-1.82) <sup>c</sup>
Prostate cancer	165	45 (5.9)	45 (5.8)	36 (4.7)	39 (5.0)	1.21 (0.81-1.81) <sup>c</sup>	0.98 (0.66-1.46) <sup>c</sup>
Hematologic cancer	53	15 (1.5)	16 (1.6)	11 (1.1)	11 (1.1)	1.43 (0.70-2.93) <sup>c</sup>	0.98 (0.48-1.98) <sup>c</sup>
Other cancer	224	56 (5.6)	61 (6.0)	62 (6.1)	45 (4.4)	1.11 (0.79-1.57) <sup>c</sup>	1.13 (0.80-1.60) <sup>c</sup>



Abbreviations: CI, confidence interval; HR, hazard ratio.

<sup>a</sup>Primary end points were cancer incidence, cancer mortality, and all-cause mortality throughout the year 2007. HRs were estimated by Cox proportional hazard regression, unadjusted stratified by trial. Non-folic acid groups or non-vitamin B<sub>6</sub> groups were the reference category. Folic acid groups were assigned to treatment with folic acid plus vitamin B<sub>12</sub> or folic acid plus vitamin B<sub>12</sub> in combination with vitamin B<sub>6</sub>; non-folic acid groups were assigned to treatment with vitamin B<sub>6</sub> alone or placebo. Vitamin B<sub>6</sub> groups were assigned to treatment with vitamin B<sub>6</sub> alone or in combination with folic acid plus vitamin B<sub>12</sub>; non-vitamin B<sub>6</sub> groups were assigned to treatment with folic acid plus vitamin B<sub>12</sub> or placebo. For incidence and mortality of prostate cancer, rates and HRs are calculated among men.

<sup>b</sup>Cancer incidence included any new cancer, except nonmelanoma skin cancer. Only the first of new cancers after randomization in each individual are included. Colorectal cancer denotes cancer classified with *International Statistical Classification of Diseases, 10th Revision (ICD-10)* codes C18 to C21. Lung cancer denotes cancer classified with *ICD-10* code C34. Prostate cancer denotes cancers classified with *ICD-10* code C61. Hematologic cancer denotes cancers classified with *ICD-10* codes C80 to C96.

<sup>c</sup>99% CIs.



# Resultatpresentasjon

**Results** During study treatment, median serum folate concentration increased more than 6-fold among participants given folic acid. After a median 39 months of treatment and an additional 38 months of posttrial observational follow-up, 341 participants (10.0%) who received folic acid plus vitamin B<sub>12</sub> vs 288 participants (8.4%) who did not receive such treatment were diagnosed with cancer (hazard ratio [HR], 1.21; 95% confidence interval [CI], 1.03-1.41; *P* = .02).

I abstractet rapporterte forfatterne effektestimat, 95% KI og p-verdi.

# Modellen: Cox regresjon

$$\ln h(t) = \ln h_0(t) + \beta_1 \cdot \text{Intervensjonsgruppe}$$

$$HR = e^{\hat{\beta}_1}$$

Stratifisert på studie

# Epidemiologi

- Hovedsaklig *observasjonelle* studier. *Eksperimentelle* studier kan sjelden benyttes.
- Eksempel: Studere sammenhengen mellom sigarettøyking og sykdom eller død. Uetisk å randomisere til f.eks. 0, 10, 20 eller 30 sigaretter/dag. I Norge kan dette f.eks. studeres i kohort studier basert på Folkehelseinstituttet sine helseundersøkelser .

# Vi eksponeres aldri for bare en faktor!

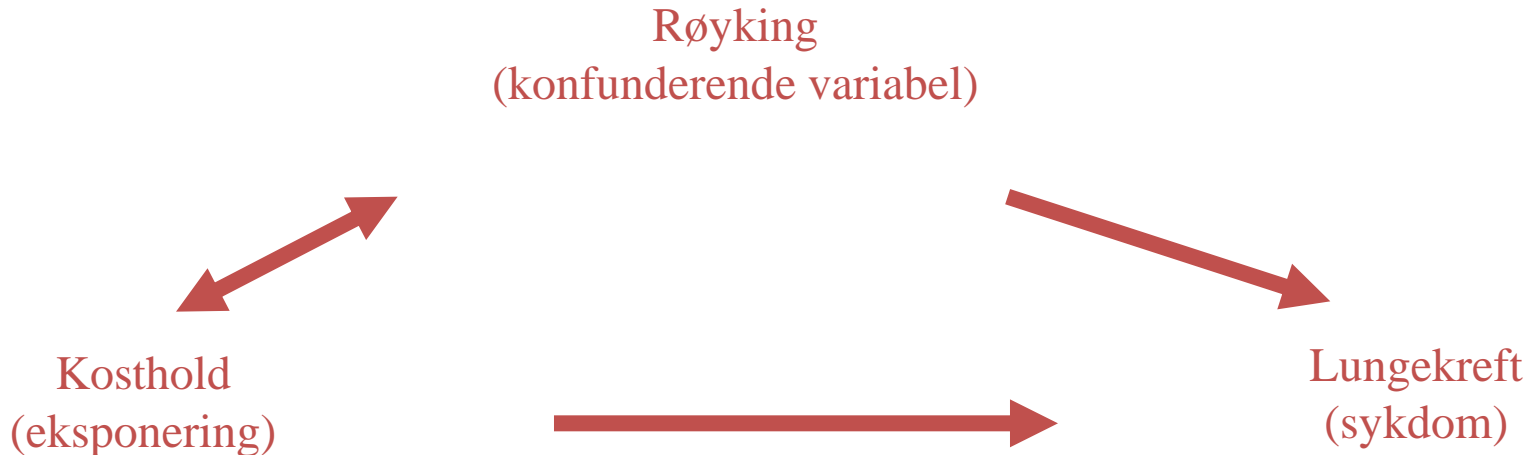
Eksempel:

Vil studere kostholdets betydning for risiko for lungekreft. Må ta hensyn til røyking.

# Konfundering

*Eksempel.* Studie av sammenhengen mellom lungekreft og inntak av frukt og grønnsaker.

- Røyking gir økt risiko for lungekreft
- Det kan være slik at de som spiser minst frukt og grønt er de som røyker.



# Konfundering kan medføre at vi:

- Rapporterer for svak eller for sterk sammenheng
- Finner sammenhenger som ikke er reelle
- Klarer ikke påvise sammenheng

**Multivariable statistiske metoder spiller en sentral rolle!**

# Eksempel på konfundering

## Folate Supplementation and Twin Pregnancies

*Stein Emil Vollset,<sup>\*\*†</sup> Håkon K. Gjessing,<sup>‡</sup> Anne Tandberg,<sup>§</sup> Thorbjørn Rønning,<sup>†</sup> Lorentz M. Irgens,<sup>\*\*†</sup>  
Valborg Baste,<sup>†</sup> Roy M. Nilsen,<sup>†</sup> and Anne Kjersti Daltveit<sup>\*\*†</sup>*

*Epidemiology 2005;16: 201–205*

- Bakgrunn
  - Kvinner som planlegger å bli gravide eller regner med å bli det, bør ta folat for å redusere risikoen for å få barn med ryggmargsbrokk.
  - Studier har vist at kvinner som tar folattilskudd er mer utsatt for tvillingfødsel
- Datamateriale
  - Medisinsk fødselsregister har siden 1998 registrert kosttilskudd og IVF (invitrofertilisering/ prøverørsbefruktning). Per desember 2005 var det registrert 176 042 single og tvillingsvangerskap.

# Forts. folat og tvillingfødsler

	Folattilskudd	Ikke folattilskudd
Tvilling svangerskap	329	2 825
Singel svangerskap	10 748	162 140

$$OR = \frac{329 \cdot 162\,140}{10\,748 \cdot 2\,825} = 1.76$$

95% konfidensintervall (1.57, 1.97)

(Mer om OR i morgen)



# Forts. folat og tvillingfødsler

- 1.5% av svangerskapene var unnfanget ved IVF  
Høyere andel tvillinger ved IVF svangerskap enn ved svangerskap unnfanget på naturlig måte (hhv 27% og 1.4%).  
Folattilskudd 4 ganger vanligere blant kvinner med IVF svangerskap.
- Logistisk regresjon med justering for mors alder og paritet (antall fødsler kvinnen har hatt) ga noe svekket estimat  
 $OR=1.59$  (95% KI (1.41, 1.78)).  
Med justering i tillegg for IVF forsvinner sammenhengen  
 $OR=1.04$  (95% KI (0.91, 1.18)).
- **Sammenhengen mellom folattilskudd og tvillingfødsler var sterkt konfundert av IVF. Konklusjon i publikasjonen: ingen sammenheng mellom folattilskudd forut for befruktning og tvillingfødsler.**

# Modellen: logistisk regresjon

$$p = P(\text{twin})$$

$$\text{Modell 1: } \log \left[ \frac{p}{1-p} \right] = \beta_0 + \beta_1 \cdot \text{folat} \quad \text{gir } OR = e^{\hat{\beta}_1} = e^{0.565} = 1.76$$

$$\text{Modell 2: } \log \left[ \frac{p}{1-p} \right] = \beta_0 + \beta_1 \cdot \text{folat} + \beta_2 \cdot \text{alder} + \beta_3 \cdot \text{paritet}$$

$$\text{gir } OR = e^{\hat{\beta}_1} = e^{0.464} = 1.59$$

**Modell 3:**

$$\log \left[ \frac{p}{1-p} \right] = \beta_0 + \beta_1 \cdot \text{folat} + \beta_2 \cdot \text{alder} + \beta_3 \cdot \text{paritet} + \beta_4 \cdot \text{IVF}$$

$$\text{gir } OR = e^{\hat{\beta}_1} = e^{0.039} = 1.04$$

Tabell 10.1 Observasjonelle epidemiologiske studier. Tidsaspektet, de vanligste effektmålene og analysemetoder

Studiedesign	Tidsaspektet		Effektmål	Analysemetode	
	Fortid	Starttidspunkt			Fremtid
Økologiske studier		Samle inn all informasjon	$r$ $\hat{b}_1, \hat{b}_2, \dots$ $OR$	Korrelasjonskoeffisient <sup>1</sup> Multippel lineær regresjon <sup>2</sup> Logistisk regresjon <sup>3</sup>	
Tverrsnittstudier		Samle inn all informasjon	$r$ $\hat{b}_1, \hat{b}_2, \dots$ $OR$ $RR$	Korrelasjonskoeffisient Multippel lineær regresjon Logistisk regresjon Tabellanalyse <sup>4</sup>	
Kasus-kontrollstudier	Registrer eksponering	Retrospektiv kartlegging	Definere kasus og kontroller	$OR$ Logistisk regresjon	
Kohortstudier		Definer en kohort og registrer eksponering	Prospektiv oppfølging	Observer endepunkt	$HR$ $IRR$ ( $OR$ ) Cox-regresjon <sup>5</sup> Poisson-regresjon <sup>6</sup> (Logistisk regresjon)

<sup>1</sup>Avsnitt 11.2. <sup>2</sup>Avsnitt 11.4. <sup>3</sup>Avsnitt 6.3 og kapittel 12. <sup>4</sup>Avsnitt 6.3 og 6.5. <sup>5</sup>Avsnitt 13.5 ( $HR = \exp(B)$ ). <sup>6</sup>Avsnitt 10.5.

Tabell 10.1 i Aalen, Frigessi, Moger, Skovlund, Veierød. Statistiske metoder i medisin og helsefag

# Mer om feilkilder i epidemiologiske studier i morgen

- **Seleksjons skjevhet**
- **Informasjonsskjevhet**
- **Konfundering**

# Publisering av forskningsresultater

## Tre statistiske mål

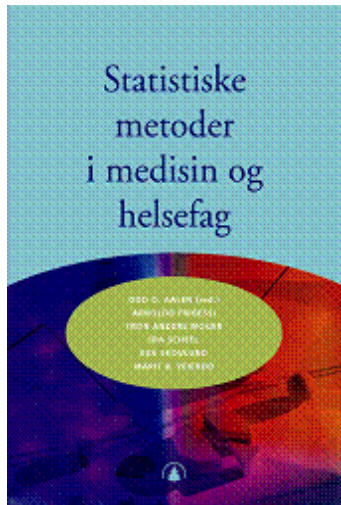
- Effektmålet vi studerer (f.eks. gjennomsnitt, andel, regresjonskoeffisient, relativ risiko, odds ratio)
- Konfidensintervall for effektmålet
- p-verdi for testen på effektmålet

# EQUATOR Network website - the resource centre for good reporting of health research studies

<http://www.equator-network.org/>

- Reporting guidelines
- CONSORT Statement for reporting of RCTs, STROBE for reporting of observational studies in epidemiology etc.

## Mer om statistisk analyse :



Aalen O.O. (red), Frigessi A., Moger T.A, Scheel I., Skovlund E., Veierød M.B.

*Statistiske metoder i medisin og helsefag.*

Gyldendal Akademisk, 2006.

<http://www.med.uio.no/imb/stat/statbok/>

## Mer om forskningsmetoder og statistisk analyse :



Laake P, Olsen B.R., Benestad H.B. (red),

*Forskning i medisin og biofag.*

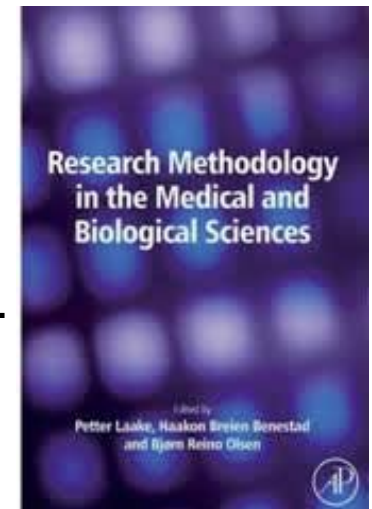
Gyldendal Akademisk, 2008.

[www.forskningsmetode.no](http://www.forskningsmetode.no)

Laake P., Benestad H.B., Olsen B.R. (red.).  
*Research Methodology in the Medical and Biological Sciences.*

Academic Press, London, 2007.

[www.elsevier.com](http://www.elsevier.com)



Mer om forskningsmetode, studiedesign mm:  
Multivariable regresjonsmodeller, se kap 3.



Laake P., Hjartåker H., Thelle D.S., Veierød M.B. (red).  
*Epidemiologiske og kliniske forskningsmetoder*.  
Gyldendal Akademisk, 2007.  
[www.forskningsmetode.no](http://www.forskningsmetode.no)



# Mange statistikk kurs på doktorgradsprogrammet

- MF9130 Innføring i statistikk
- MF9530 Analyse av repeterte/korrelerte kategoriske data
- MF9510 Logistic regression, survival analysis and Cox-regression
- MF9540 Analysis of contingency tables and logistic regression
- MF9120 - Introduction to infectious disease modelling
- .....

# Oppgave

ORIGINAL ARTICLE

# Intensive Lipid Lowering with Simvastatin and Ezetimibe in Aortic Stenosis

Anne B. Rossebø, M.D., Terje R. Pedersen, M.D., Ph.D.,  
Kurt Boman, M.D., Ph.D., Philippe Brudi, M.D., John B. Chambers, M.D.,  
Kenneth Egstrup, M.D., Ph.D., Eva Gerds, M.D., Ph.D.,  
Christa Gohlke-Bärwolf, M.D., Ingar Holme, Ph.D.,  
Y. Antero Kesäniemi, M.D., Ph.D., William Malbecq, Ph.D.,  
Christoph A. Nienaber, M.D., Ph.D., Simon Ray, M.D.,  
Terje Skjærpe, M.D., Ph.D., Kristian Wachtell, M.D., Ph.D.,  
and Ronnie Willenheimer, M.D., Ph.D., for the SEAS Investigators\*

# Design aspects

- A detailed description is given of the design, with an excellent overview in Figure 1.
- Discuss:
  - Exclusions
  - Drop-out
  - “Discontinued placebo, followed per protocol”.  
What does this mean?

# Discuss

- This study is based on a seemingly well founded theory about the anticipated effect of treatment. However, the data do not seem to support this theory. What is the basis for the theory, and how strong is the evidence against it?
- The study ends up showing an increased risk of cancer in the treatment group. How should we relate to this finding?

# Effect measures

- There are at least two levels of effect: *cholesterol modification* and *events* (primary and secondary outcomes and death)
- What type of effect measures do you find at the two levels?

# Figures

- How do you read figures 2 and 3?