Studi clinici longitudinali: metodi di analisi statistica

Massimo Borelli

21 aprile 2016

UNIVERSITÀ
DEGLI STUDI DI TRIESTE

SOCIETA' DEI MATEMATICI
E NATURALISTI DI MODENA
www.socnatmatmo.unimore.it

2005, John Ioannidis

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims [6–8]. However, this should not be surprising. It can be proven that most claimed research findings are false. Here I will examine the key factors that influence this problem and some corollaries thereof.

The Essay section contains opinion pieces on topics of broad interest to a general medical audience.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a p-value less than 0.05. Research is not most appropriately represented and summarized by p-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on p-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2×2 table in which research findings are compared against the gold standard of true relationships in a scientific field. In a research field both true and false hypotheses can be made about the presence of relationships. Let R be the ratio of the number of "true relationships" to "no relationships" among those tested in the field. R is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R + 1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2×2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

Citation: Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8): e124.

Abbreviation: PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, and Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America. E-mail: jioannid@cc.uoi.gr

Competing Interests: The author has declared that no competing interests exist.

DOI: 10.1371/journal.pmed.0020124

2009, Marcia Angell

It is simply no longer possible to believe much of the clinical research that is published, or to rely on the judgment of trusted physicians or authoritative medical guidelines. I take no pleasure in this conclusion, which I reached slowly and reluctantly over my two decades as an editor of The New England Journal of Medicine.

2015, Richard Horton

The case against science is straightforward: much of the scientific literature, perhaps half, may simply be untrue. Afflicted by studies with small sample sizes, tiny effects, invalid exploratory analyses, and flagrant conflicts of interest, together with an obsession for pursuing fashionable trends of dubious importance, science has taken a turn towards darkness.

2016, Ron Wasserstein

ASA News
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • www.amstat.org • www.twitter.com/AmstatNews

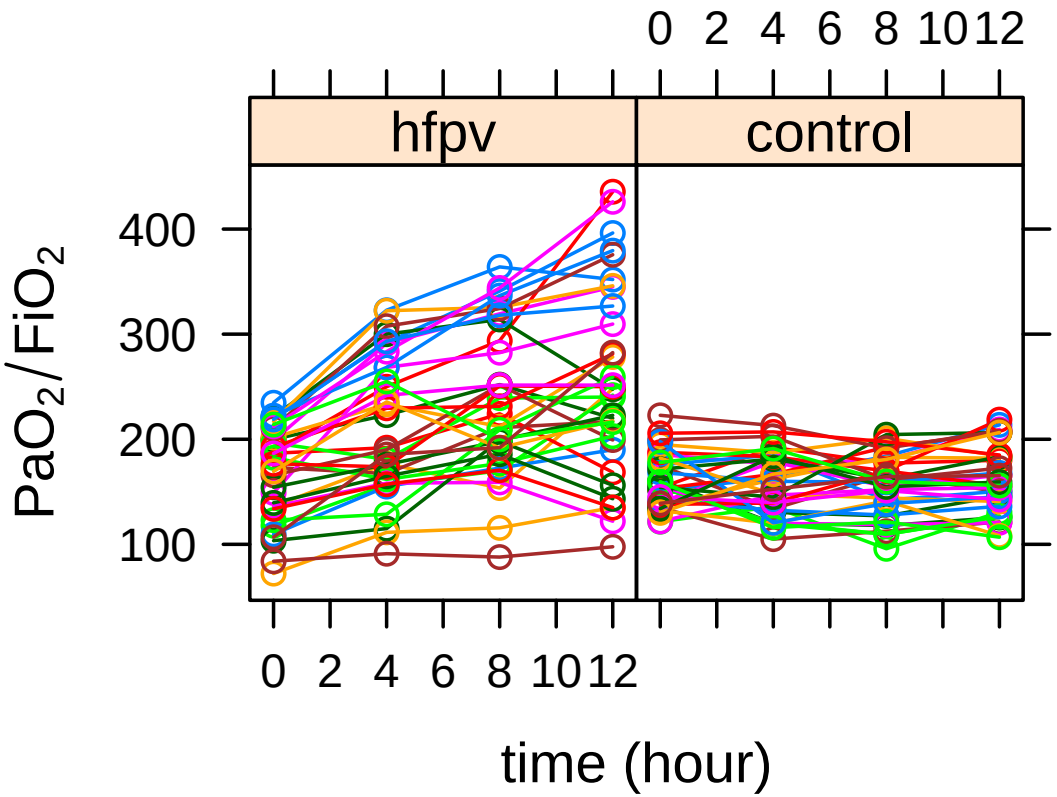
AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND *P*-VALUES

Provides Principles to Improve the Conduct and Interpretation of Quantitative Science

March 7, 2016

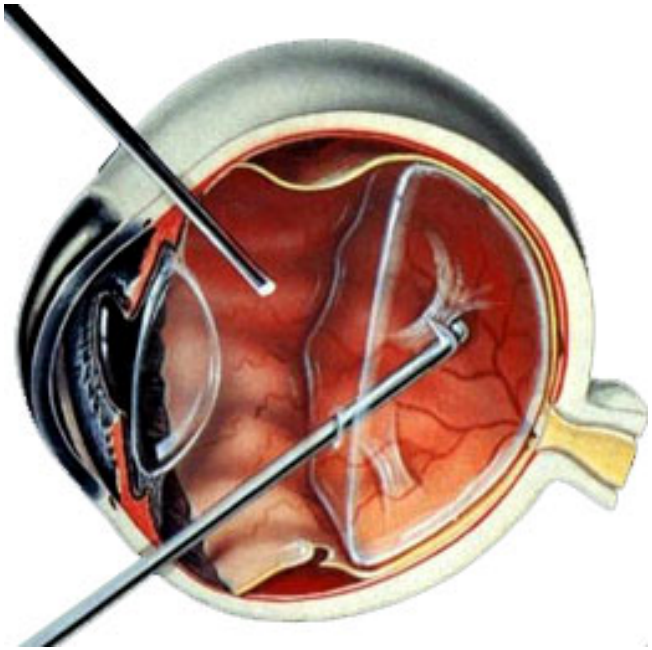
Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The ASA statement is intended to steer research into a '**post p<0.05 era**'.

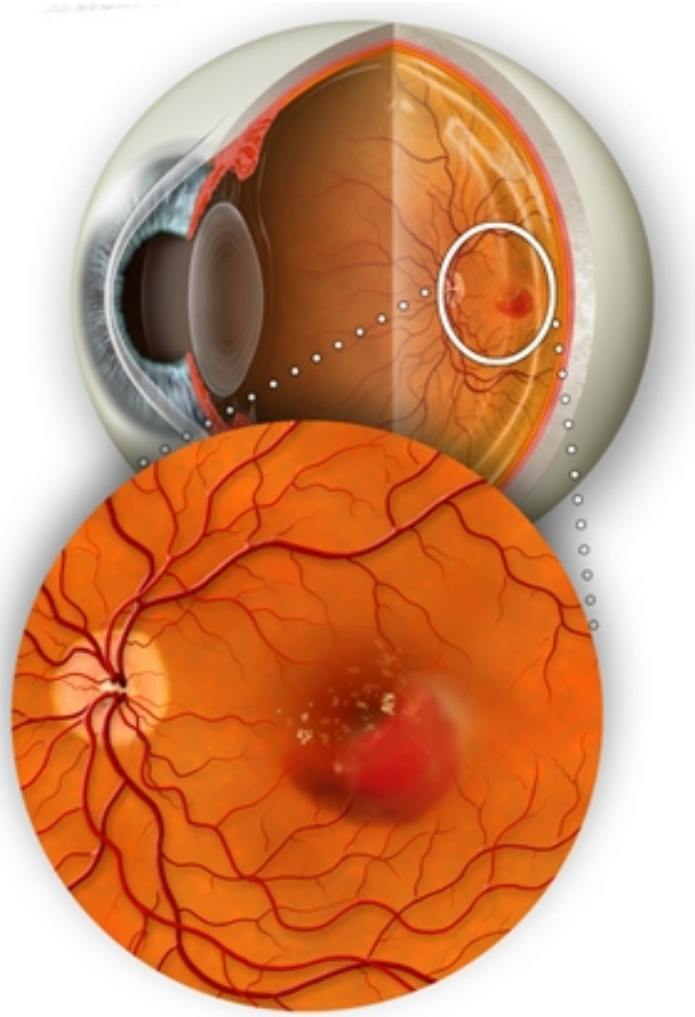
what does it mean 'longitudinal'?



U. Lucangelo et al., 2011. *Early Short-Term Application of High-Frequency Percussive Ventilation*.

clinical topics





the **vitrectomia** dataset

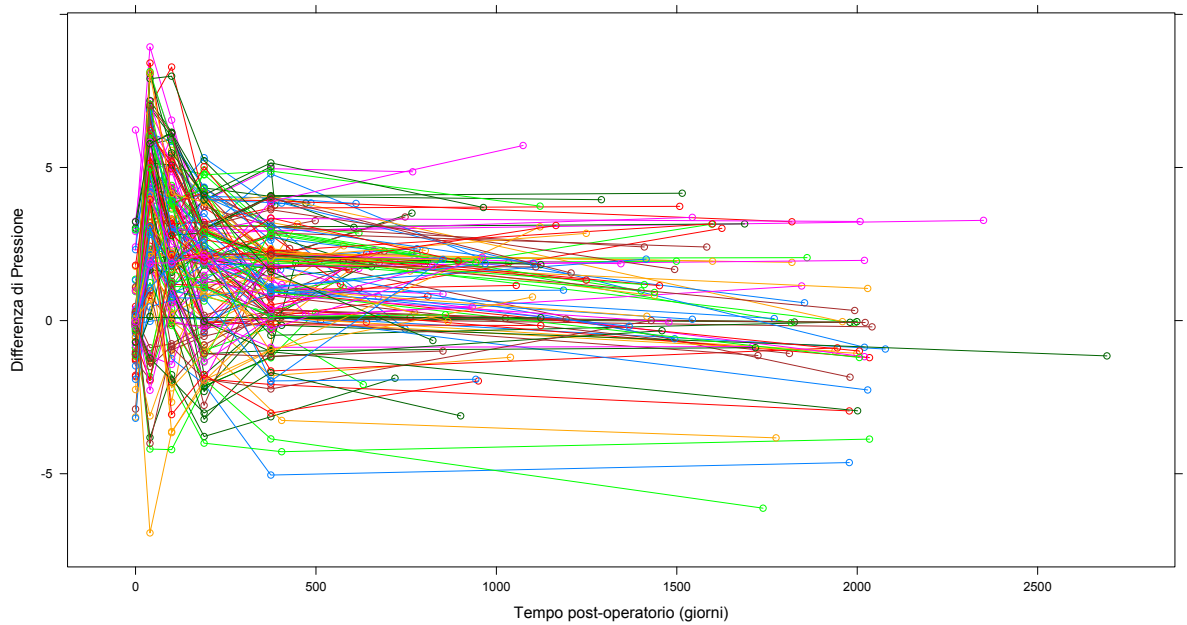
	Soggetto	Nascita	Datavisita	Tipovisita	Occhio
979	164	11760	41908	Apreop	OS
980	164	11760	41948	B30	OS
981	164	11760	42008	C90	OS
982	164	11760	42098	D180	OS
983	164	11760	42283	E365	OS
984	164	11760	42352	Finale	OS

	PTrattato	PControllo	Sesso	Intervento	Gauge
979	12	12	M	Mer	27
980	8	12	M	Mer	27
981	12	12	M	Mer	27
982	15	12	M	Mer	27
983	13	12	M	Mer	27
984	11	13	M	Mer	27

the **maculopatia** dataset

Subject	AcVis	Time	Gender	Age	Eye	Drug	Injection
1	0.63	0.00	F	68	sx	Beva	15
2	0.40	0.00	F	82	dx	poli	14
3	0.20	0.00	F	71	dx	Ranib	18
4	0.20	0.00	M	64	dx	poli	25
5	0.25	0.00	F	83	dx	poli	7
6	0.63	0.00	M	79	dx	poli	8
..	..	..	..	..	..	..	4
280	NA	72.00	F	80	sx	Ranib	3
281	NA	72.00	F	85	sx	Beva	3
282	NA	72.00	F	88	sx	poli	9

time profiles



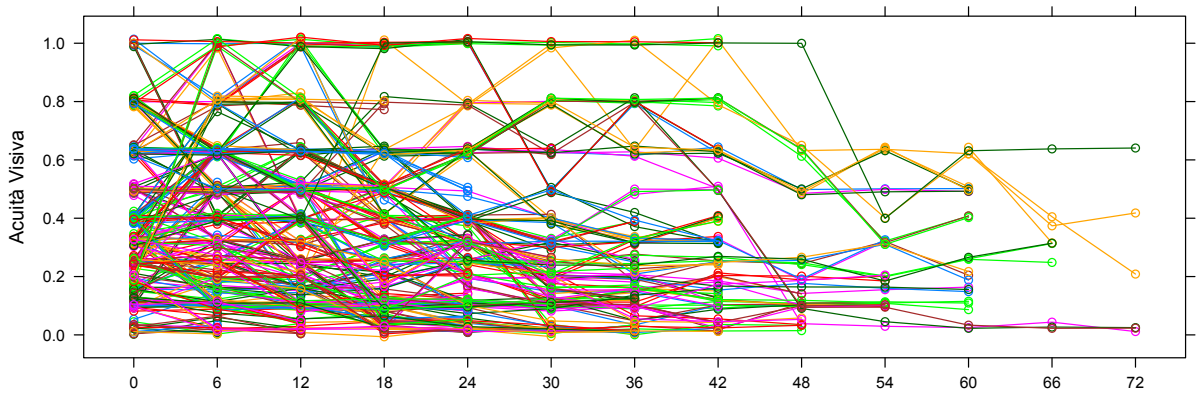
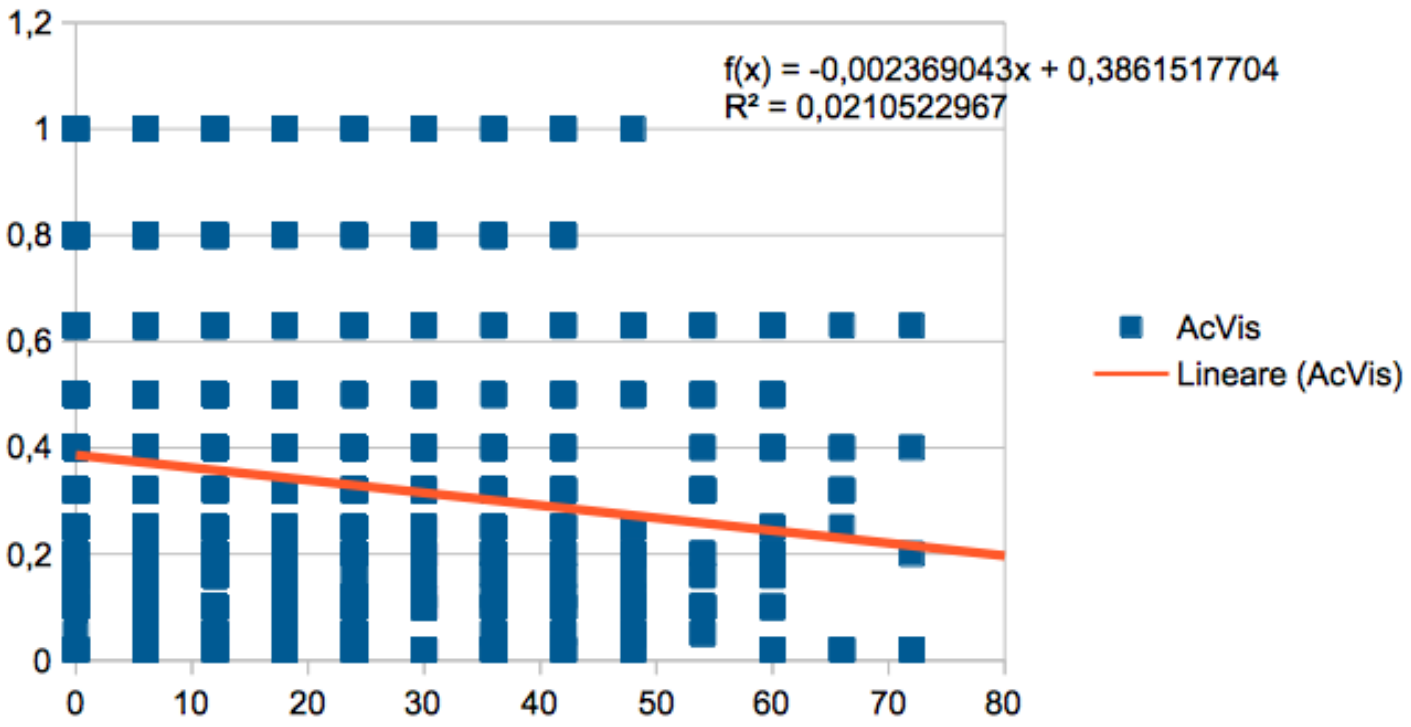


Table of contents

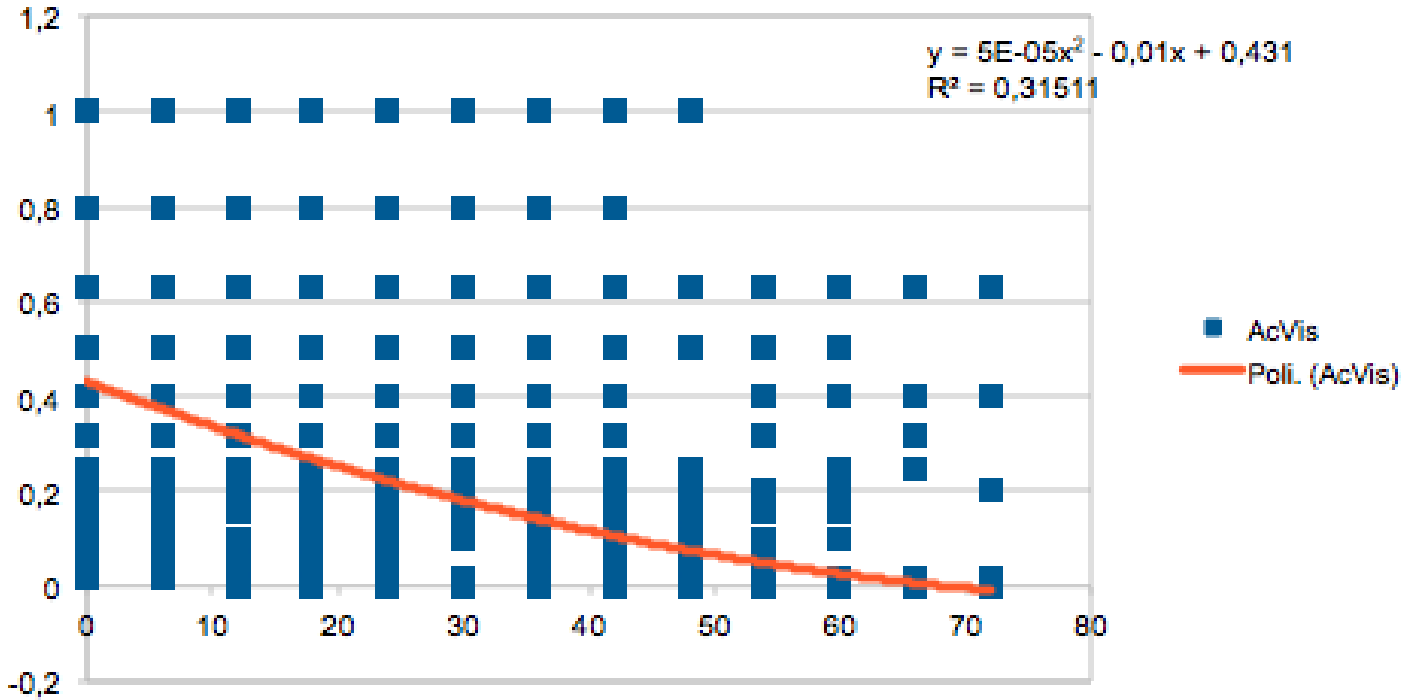
a wrong analysis

	A	B	C	D	E	F	G	H	I	J
1	Subject	Time	AcVis	Gender	Age	Eye	Drug	Injection		
2	1	0	0,63	Femmina	68	sinistro	Bevacizumab	15		
3	2	0	0,4	Femmina	82	destro	polifarmacologica	14		

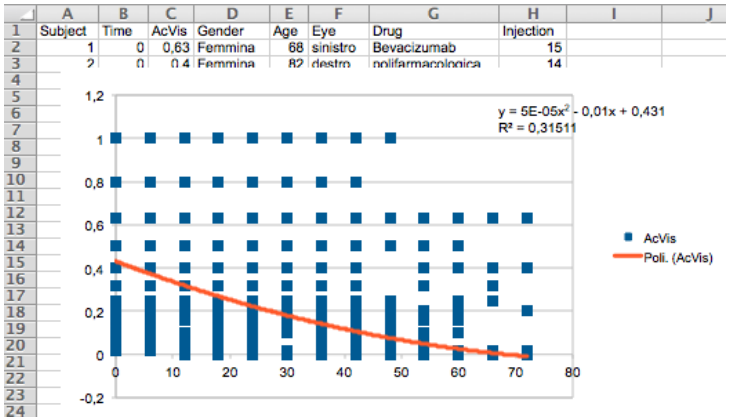


a wrong analysis

	A	B	C	D	E	F	G	H
1	Subject	Time	AcVis	Gender	Age	Eye	Drug	Injection
2	1	0	0,63	Femmina	68	sinistro	Bevacizumab	15
3	2	0	0,4	Femmina	82	destro	polifarmacologica	14



a wrong conclusion



Subject	Time	AcVis	Gender	Age	Eye	Drug	Injection
1	0	0,63	Femmina	68	sinistro	Bevacizumab	15
2	0	0,4	Femmina	82	destro	polifarmacologica	14

wrong!

The second order polynomial regression exhibit a stronger R^2 determination coefficient (0.30 vs. 0.02 in the linear case), therefore we deduce that the terapy **has an effect in slowing down** the disease

Why that analysis is wrong?

we have to face different difficulties

- to 'reduce repeated information' into *one* number
- assure reliable inference
  - managing the *twins effect* :-)
  - managing the *latent variables* / hierarchical structure

the *twins effect* :-)

$\text{alice} = 73.6$

$\text{ellen} = 73.8$

	gemella	peso
1	alice	73.60
2	alice	73.40
3	alice	74.10
4	ellen	73.80
5	ellen	73.50
6	ellen	74.60

	gemella	peso
1	alice	73.60
2	alice	73.40
3	alice	74.10
4	alice	73.50
5	alice	73.20
..	..	..
21	alice	74.20
22	ellen	73.80
23	ellen	73.50
..	..	..
42	ellen	74.50

$alice = 73.6$

$ellen = 73.8$

	gemella	peso
1	alice	73.60
2	alice	73.40
3	alice	74.10
4	ellen	73.80
5	ellen	73.50
6	ellen	74.60

	gemella	peso
1	alice	73.60
2	alice	73.40
3	alice	74.10
4	alice	73.50
5	alice	73.20
..	..	..
21	alice	74.20
22	ellen	73.80
23	ellen	73.50
..	..	..
42	ellen	74.50

$\nexists p - value$ $p - value = 0.54$ $p - value = 0.02$

explanation of the phenomenon

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

the most dangerous equation

the most dangerous equation

BMC Health Services Research

BioMed Central

Correspondence

Open Access

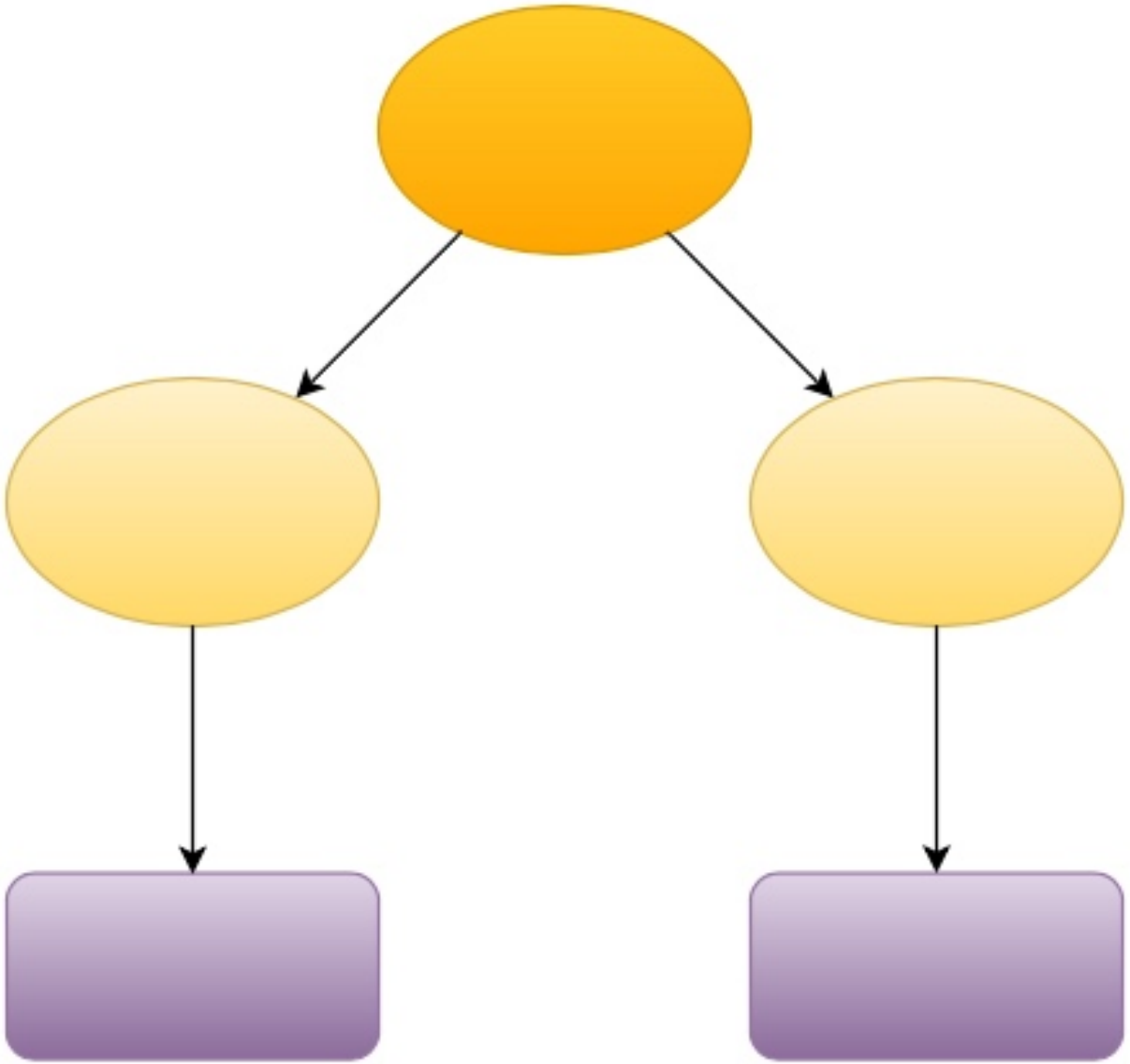
The most dangerous hospital or the most dangerous equation?

Yu-Kang Tu*[1,2] and Mark S Gilthorpe[1]

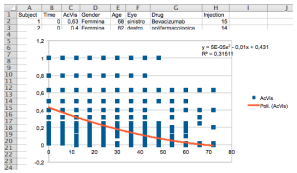
Results: A close examination of the information reveals a pattern which is consistent with a statistical phenomenon, discovered by the French mathematician de Moivre nearly 300 years ago, described in every introductory statistics textbook: namely that variation in performance indicators is expected to be greater in small Trusts and smaller in large Trusts. From a statistical viewpoint, the number of deaths in a hospital is not in proportion to the size of the hospital, but is proportional to the square root of its size. Therefore, it is not surprising to note that small hospitals are more likely to occur at the top and the bottom of league tables, whilst mortality rates are independent of hospital sizes.


A
B


4



a wrong conclusion

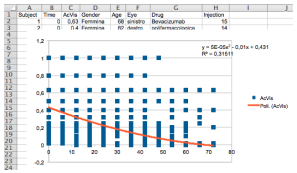


so, what was wrong?

- to have considered each point as an 'independent' observation, forgetting that there is information carried in (i.e. previous patient conditions)

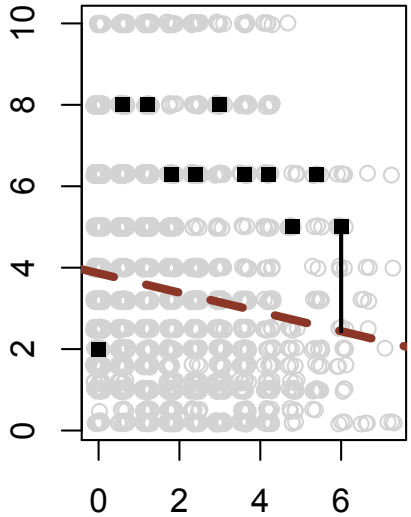
The second order polynomial regression exhibit a stronger R^2 determination coefficient (0.30 vs. 0.02 in the linear case), therefore we deduce that the terapy **has an effect in slowing down** the disease

a wrong conclusion



so, what was wrong?

- to have considered each point as an 'independent' observation, forgetting that there is information carried in (i.e. previous patient conditions)

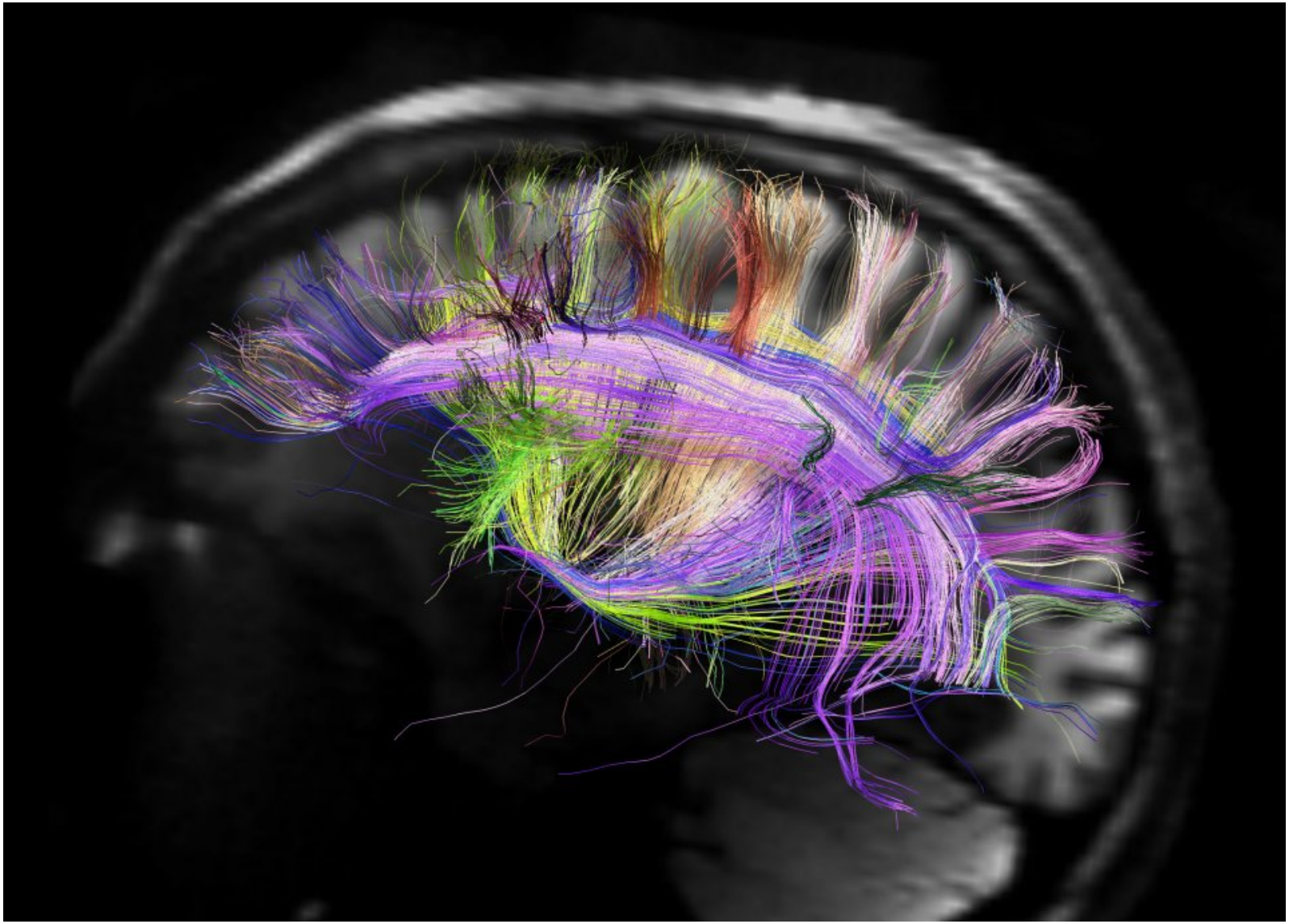


the mixed-effects models

good news

a **mixed-effects model** allows to

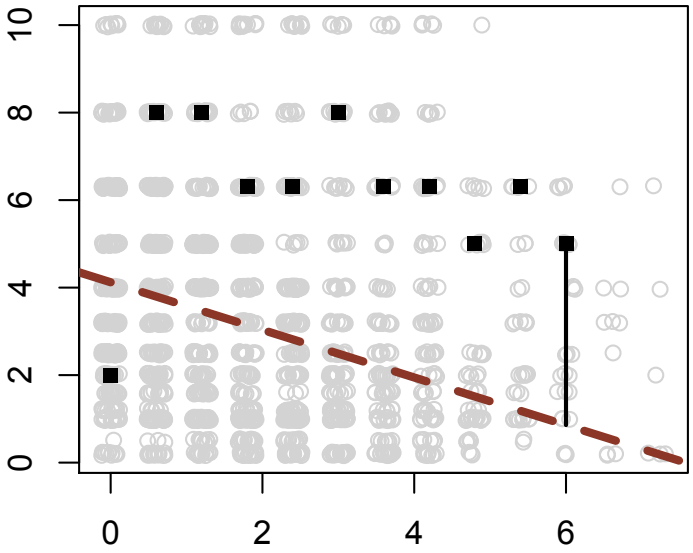
- obtain (population) time-evolution estimates from the (random sample) observations
  - **fixed effects**
- to take in account the patient-level time-evolution within the (random) sample observed
  - **random effects**

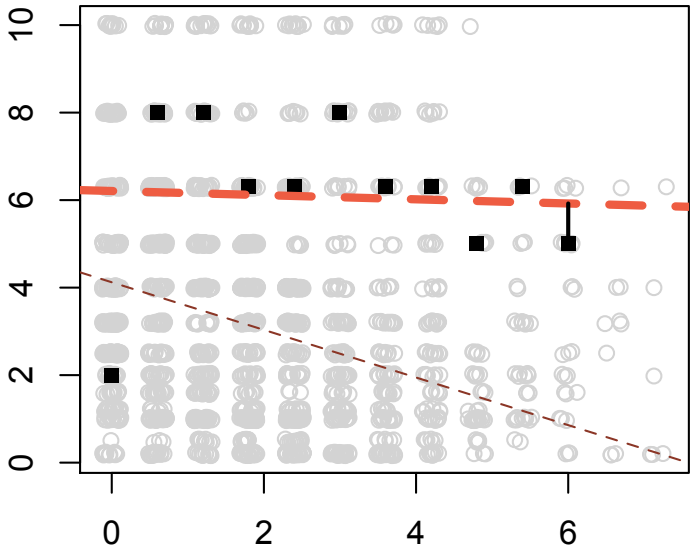


what are we going to talk now?

1. the **idea behind** a mixed-effects model
2. to explain the difference between **fixed** and **random** effects
3. *hard* – to pursuit a proper **model selection**

the idea behind





$$y = mx + q + \varepsilon$$

$$y = (m + \beta)x + (q + \alpha) + \varepsilon$$

fixed effects vs. random effects

$$y = mx + q + \varepsilon$$

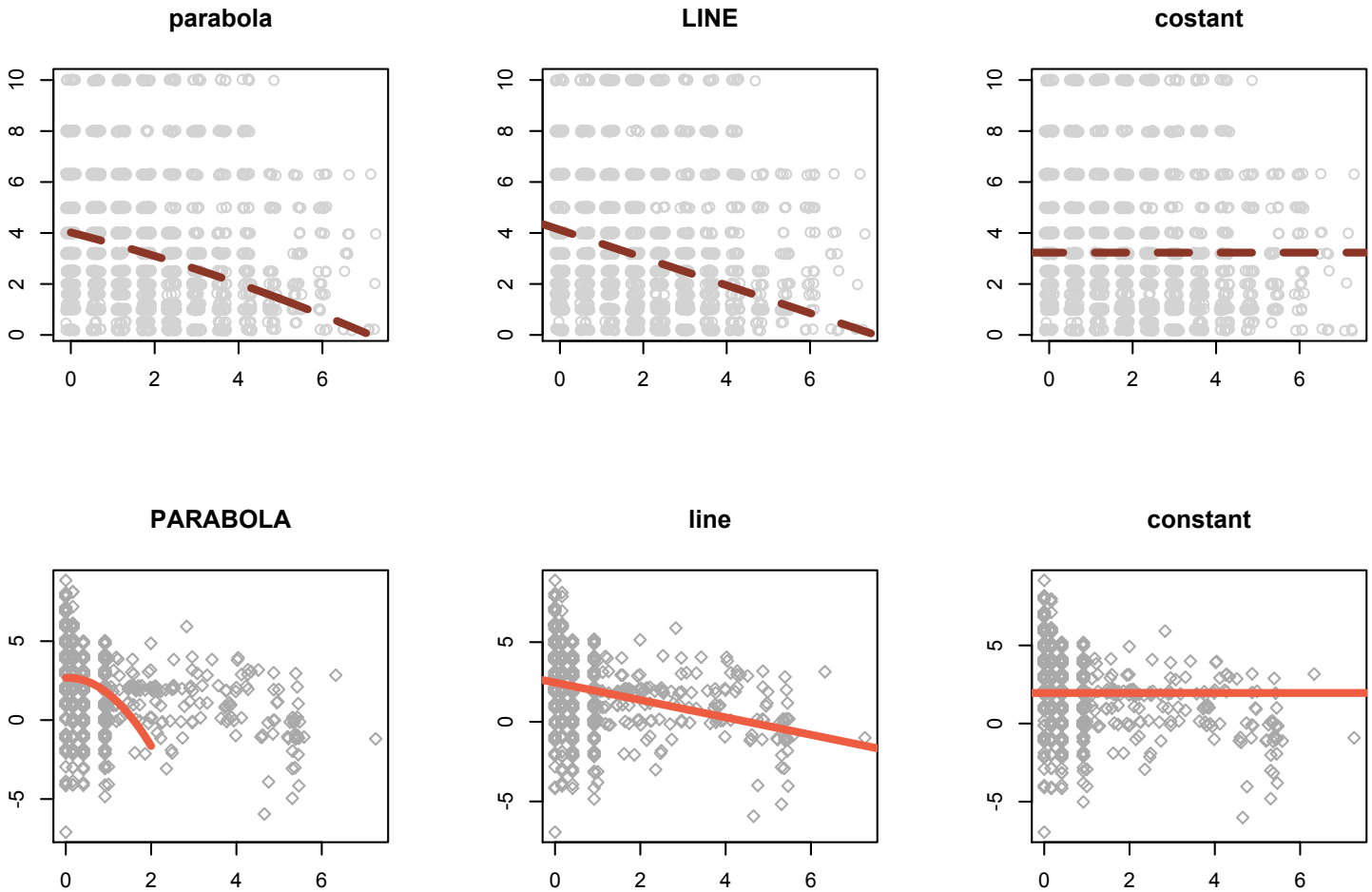
$$y = (m + \beta)x + (q + \alpha) + \varepsilon$$

- m, q are (population) **fixed effects**
- α, β are (patient) **random effects**
- ε is the residual **random effects**

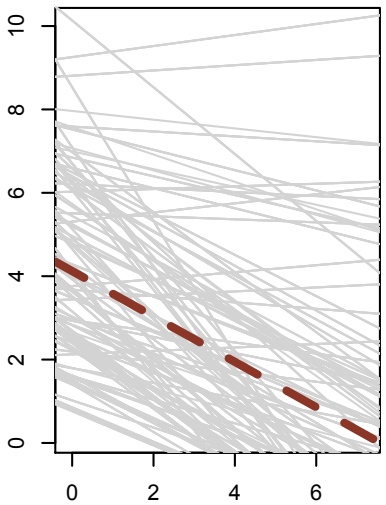
$$\alpha, \beta, \varepsilon \sim N(0, ...)$$

$$cor(\alpha, \beta) = ...$$

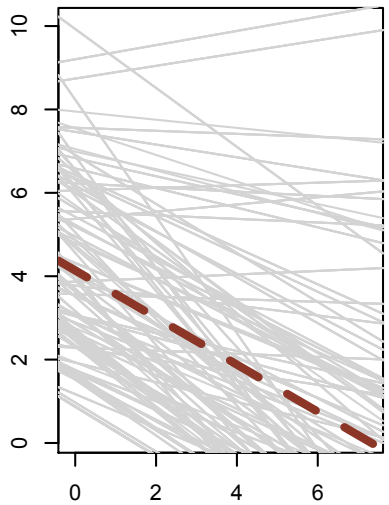
hard - model selection / fixed effects



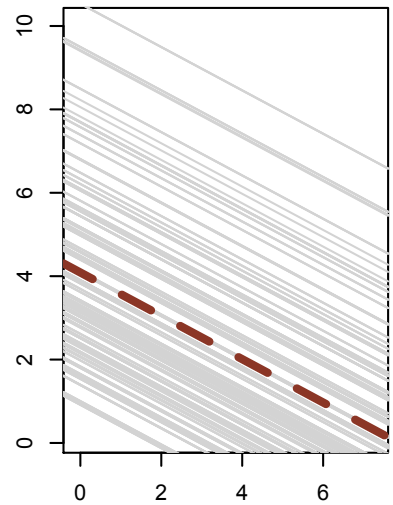
hard - model selection / random effects



$$(m+\beta)x+(q+\alpha)$$
$$cor(\alpha,\beta)=\rho$$
$$\varepsilon$$



$$(m+\beta)x+(q+\alpha)$$
$$cor(\alpha,\beta)=0$$
$$\varepsilon$$



$$mx+(q+\alpha)$$
$$\varepsilon$$

hard - model selection

Three way are commonly exploited to pursuit a model selection

- \- deviance analysis (under Maximum Likelihood estimates)
- \- information criteria (e.g. AIC)
- \+ parametric bootstrap

Note: first two methods properly work only on fixed effects
J. Faraway, 2016, ISBN 9781498720960.

hard - parametric bootstrap

```
Ricco = lmer(AcVis ~ 1 + Time + I(Time^2) + (1|Subject),
                  REML = FALSE)
Povero = lmer(AcVis ~ 1 + Time + (1|Subject),
                  REML = FALSE)

Lrts = 2*(logLik(Ricco)-logLik(Povero))

###

howmany = 1000; Lrdistrib = numeric(howmany)

for(i in 1:howmany)
{
simul = simulate(modelloPovero)
bootRicco = lmer(simul ~ Time + I(Time^2) + (1|Subject),
                  REML = FALSE)
bootPovero = lmer(simul ~ Time + (1|Subject),
                  REML = FALSE)
Lrdistrib[i] = 2*(logLik(bootRicco)-logLik(bootPovero))
}

pvalue = sum(Lrdistrib>Lrts)/howmany
```

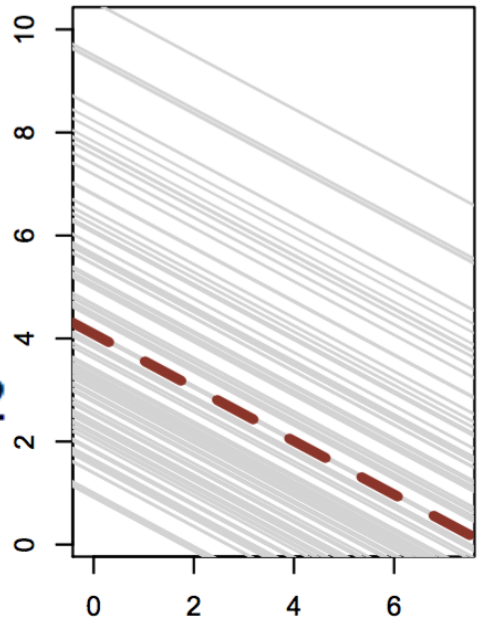
hard - model selection

with R: model summary

```
library(lme4)

mixed0 = lmer(AcVis ~ 1 + (1 | Subject))
summary(mixed0)

mixed1 = lmer(AcVis ~ 1 + Time + ( 1 | Subject))
summary(mixed1)

mixed2 = lmer(AcVis ~ 1 + Time + I(Time^2) + ( 1 | Subject))
summary(mixed2)
```

with R: to understand summary

```
Random effects:
 Groups   Name        Variance Std.Dev.
 nsubject (Intercept) 4.891    2.211
 Residual             1.862    1.364
Number of obs: 1817, groups:  nsubject, 2

Fixed effects:
            Estimate Std. Error t value
(Intercept)  4.08801    0.14161   28.87
Time        -0.51963    0.02327  -22.33
```



$$y = mx + (q + \alpha) + \varepsilon$$

insight: the bayesian approach

please, note the differences (1/3)

```
wrong1 = lm(AcVis ~ 1 + Time)
summary(wrong1)
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.86152    0.09482  40.725  < 2e-16 ***
Time        -0.23690    0.03792  -6.248 5.18e-10 ***
```

please, note the differences (2/3)

```
mixed1 = lmer(AcVis ~ 1 + Time + (1|Subject))
summary(mixed1)
```

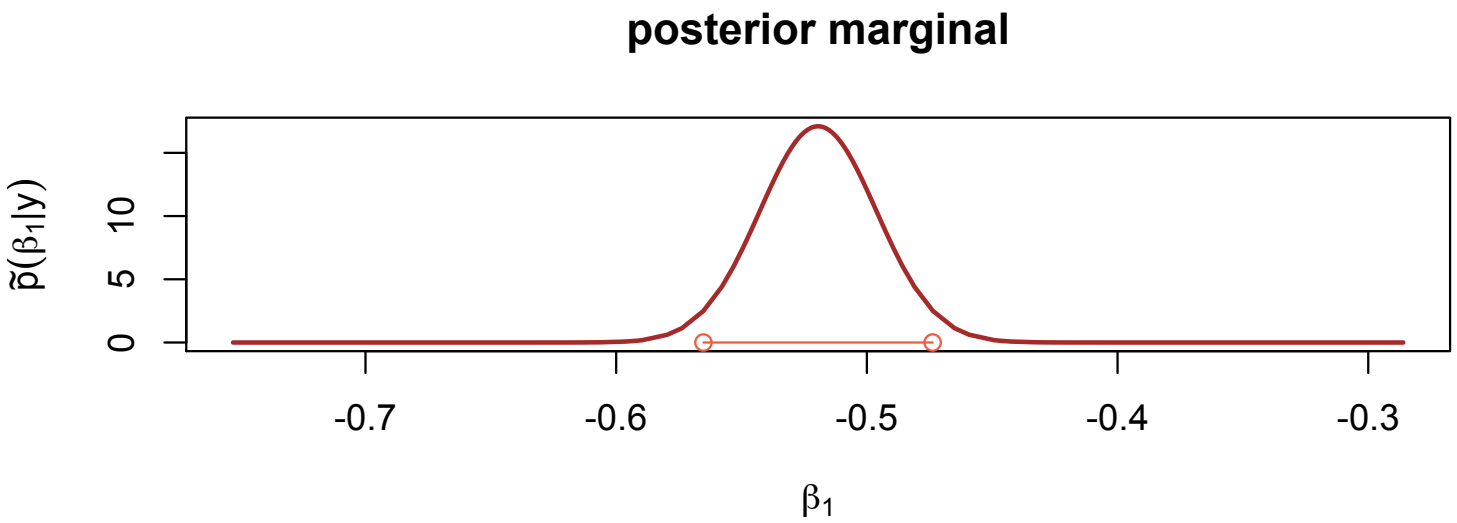
```
Random effects:
 Groups   Name        Variance Std.Dev.
 nsubject (Intercept) 4.891    2.211
 Residual             1.862    1.364
Number of obs: 1817, groups:  nsubject, 282

Fixed effects:
            Estimate Std. Error t value
(Intercept)  4.08801    0.14161   28.87
Time        -0.51963    0.02327  -22.33
```

please, note the differences (3/3)

```
formula = AcVis ~ 1 + Time + f(Subject, model = iid)
output = inla(formula, family = gaussian)
summary(output)
```

```
Fixed effects:
             mean     sd 0.025quant 0.5quant 0.975quant    mode kld
(Intercept)  4.0879 0.1414     3.8102   4.0879     4.3656  4.0878   0
Time        -0.5194 0.0233    -0.5652  -0.5194    -0.4736 -0.5195   0

Random effects:
Name    Model
 Subject   IID model

Model hyperparameters:
                                                mean     sd 0.025quant 0.5qu
Precision for the Gaussian observations 0.5377 0.0194     0.5004   0.5
Precision for Subject                   0.2063 0.0187     0.1717   0.2
```

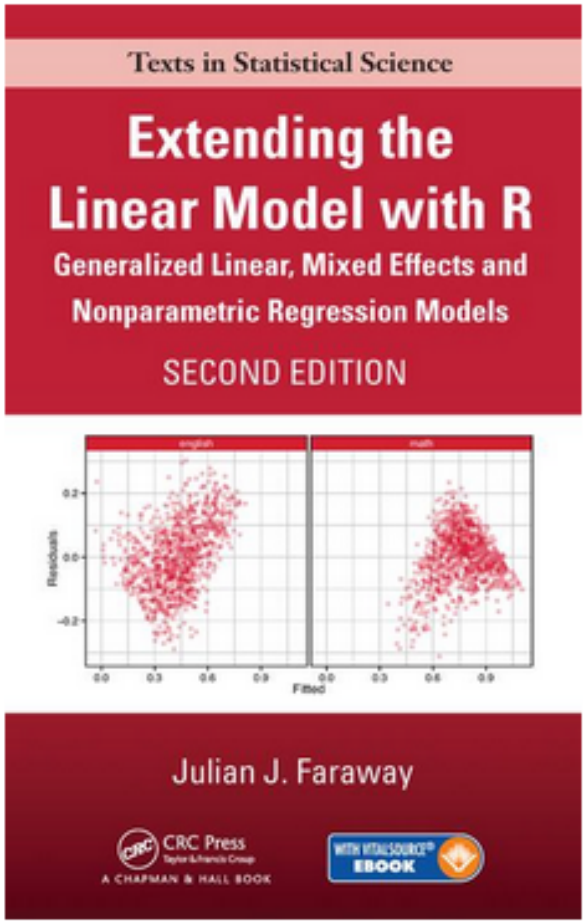
great advantage

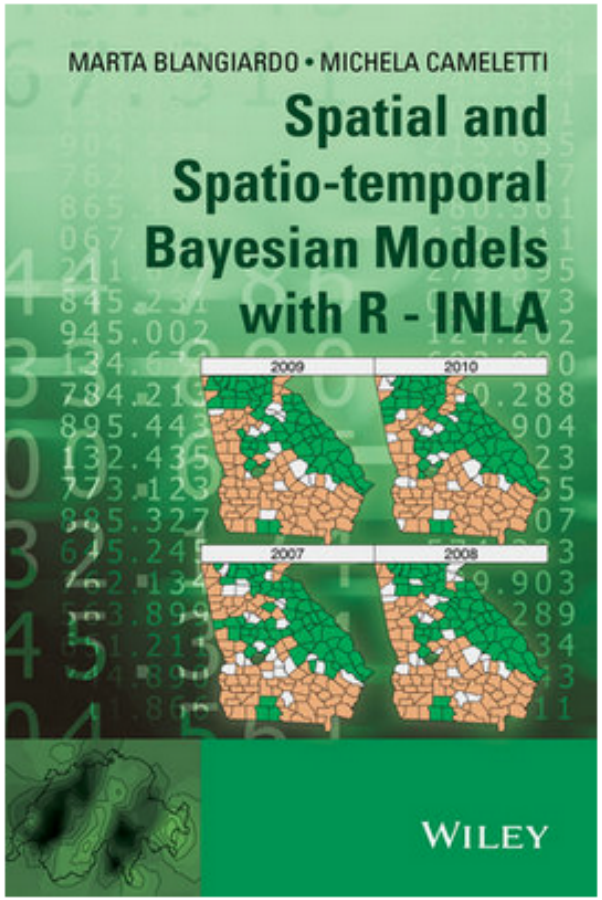


Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	
(Intercept)	4.0879	0.1414	3.8102	4.0879	4.3656	4
Time	-0.5194	0.0233	-0.5652	-0.5194	-0.4736	-0

some textbooks





ringraziamenti

Arjuna, Federico, Roberto: youtube.com/medicinatrieste

Massimo Borelli
borelli@units.it

www.dmi.units.it/borelli/