Clinical trials: three statistical traps for the unwary

Stephen Senn

LUXEMBOURG INSTITUTE OF HEALTH
RESEARCH DEDICATED TO LIFE

Acknowledgements

Acknowledgements

Thanks for inviting me

This work is partly supported by the European Union's 7th Framework Programme for research, technological development and demonstration under grant agreement no. 602552. "IDEAL"

The three

- Regression to the mean
- Invalid inversion
- Misinterpreting ‘response’

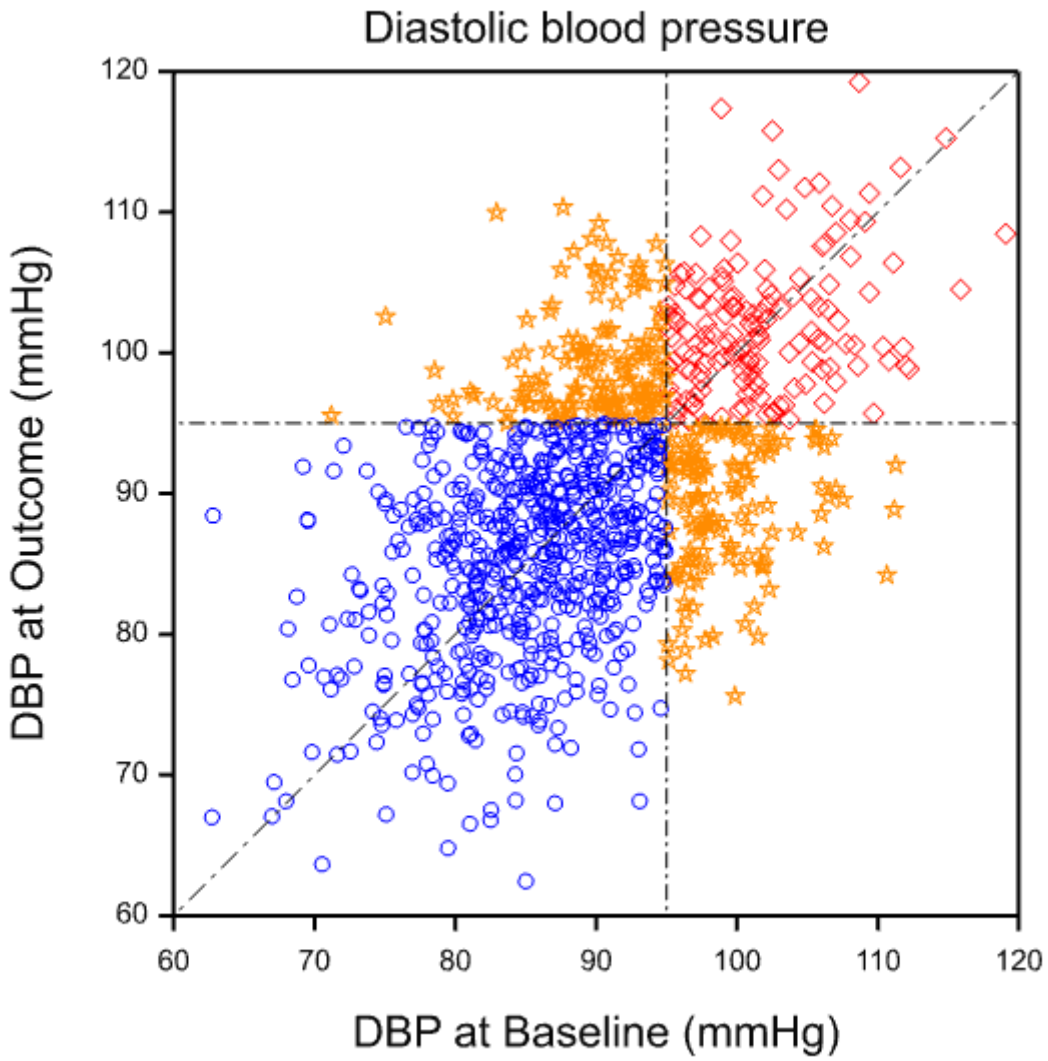
1. Regression to the Mean

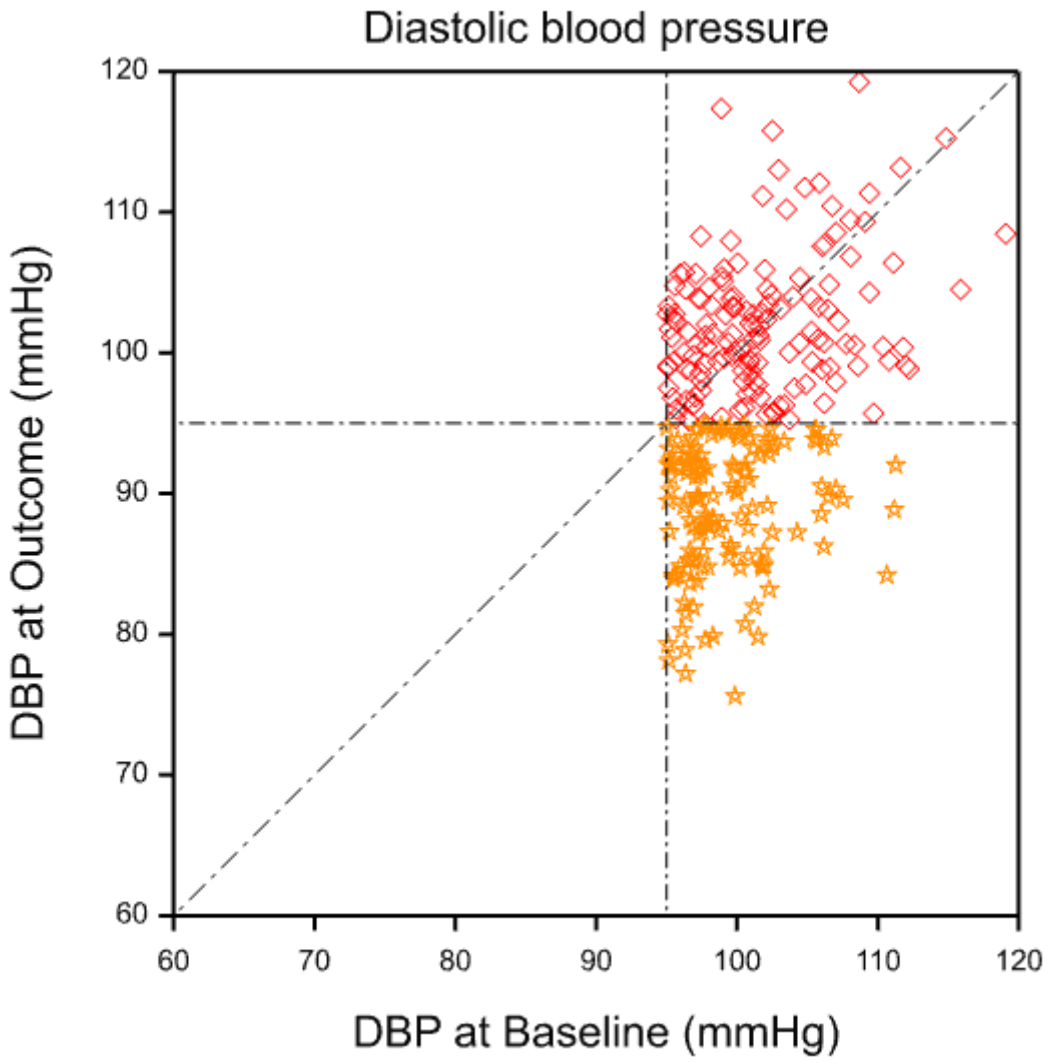
The tendency for extreme things to appear more average when studied again

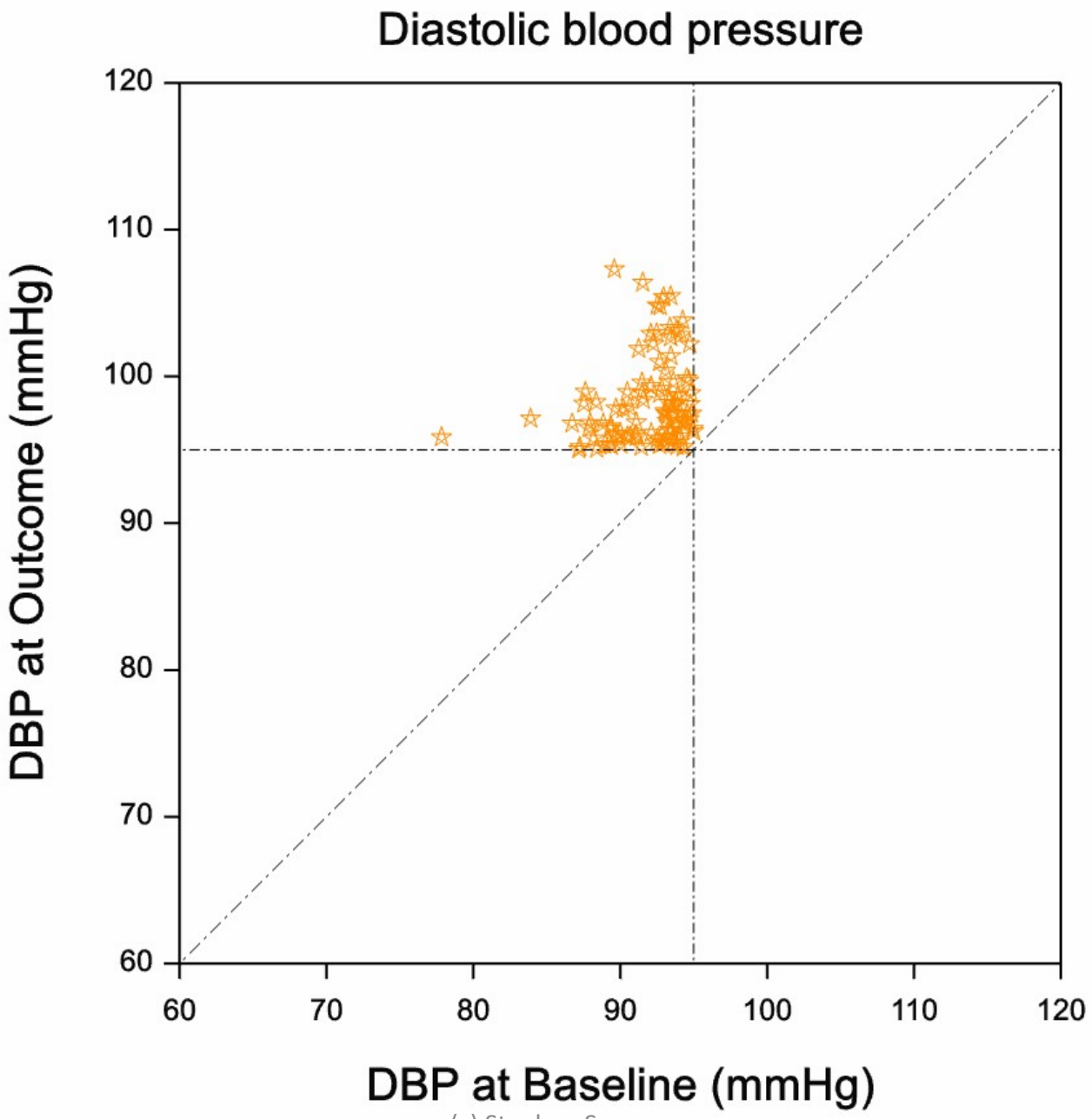
A powerful source of bias in uncontrolled studies

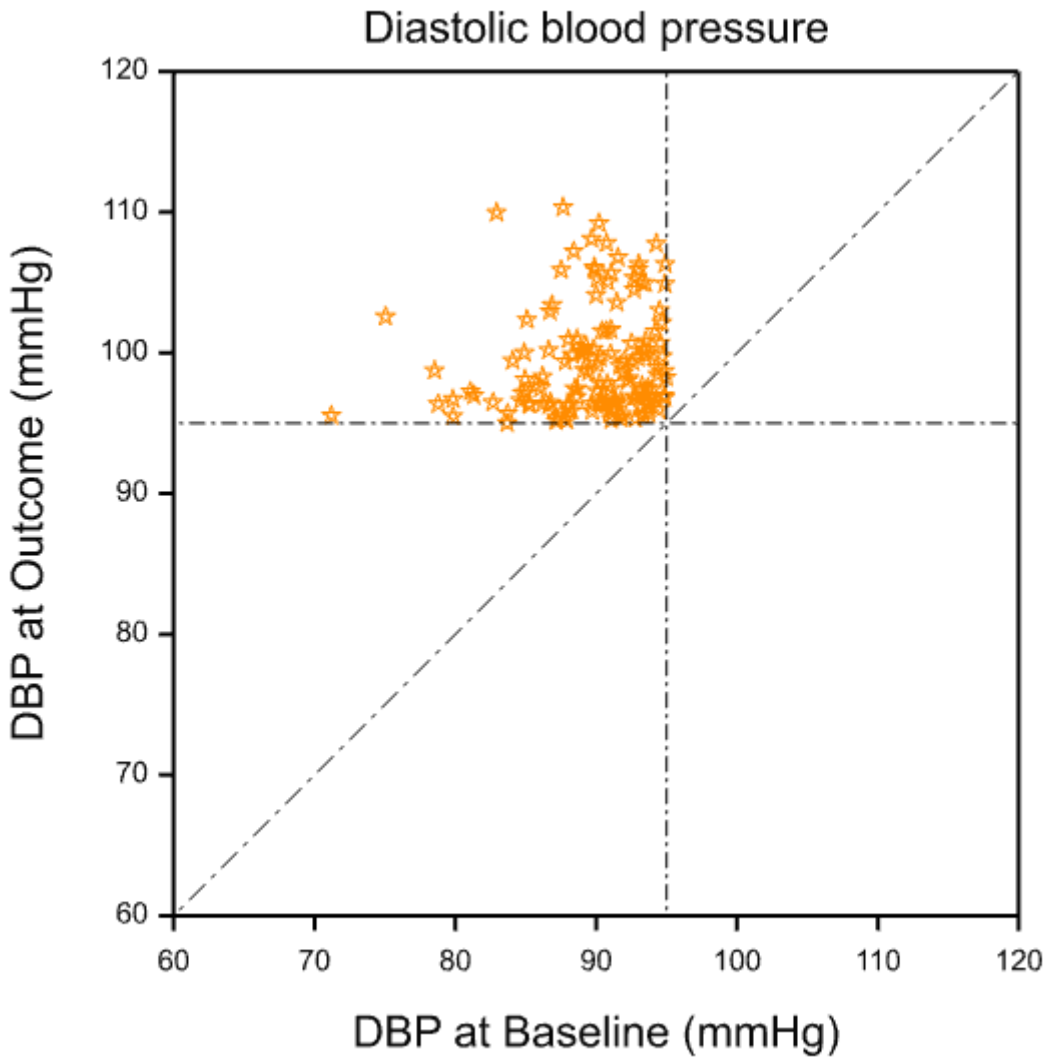
Regression to the Mean
A Simulated Example

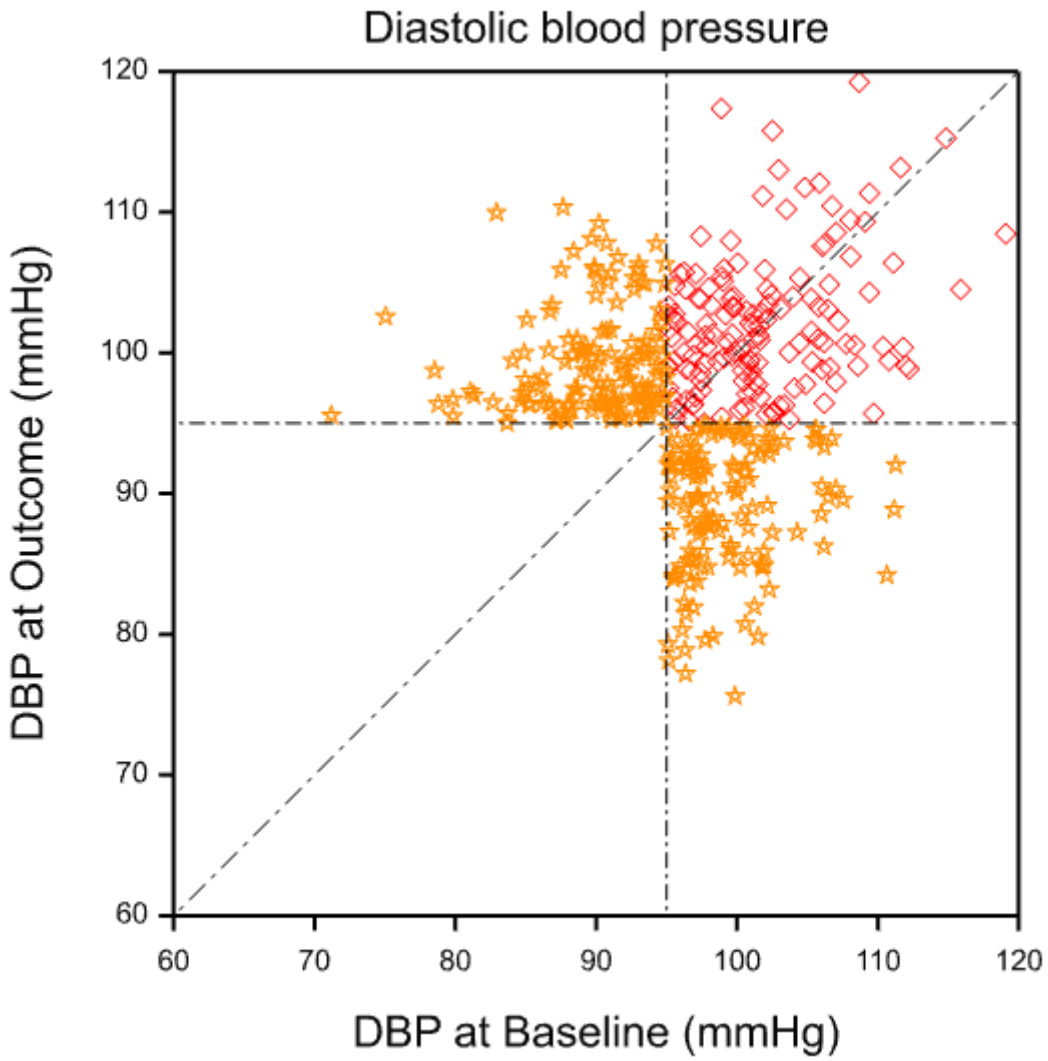
- Diastolic blood pressure (DBP)
  - Mean 90mmHg
  - Between patient variance 50mmHg^2
  - Within patient variance 15 mmHg^2
  - Boundary for hypertensive 95 mmHg
- Simulation of 1000 patients whose DBP at baseline and outcome are shown
  - Blue consistent normotensive
  - Red Consistent hypertensive
  - Orange hypertensive/normotensive or *vice versa*

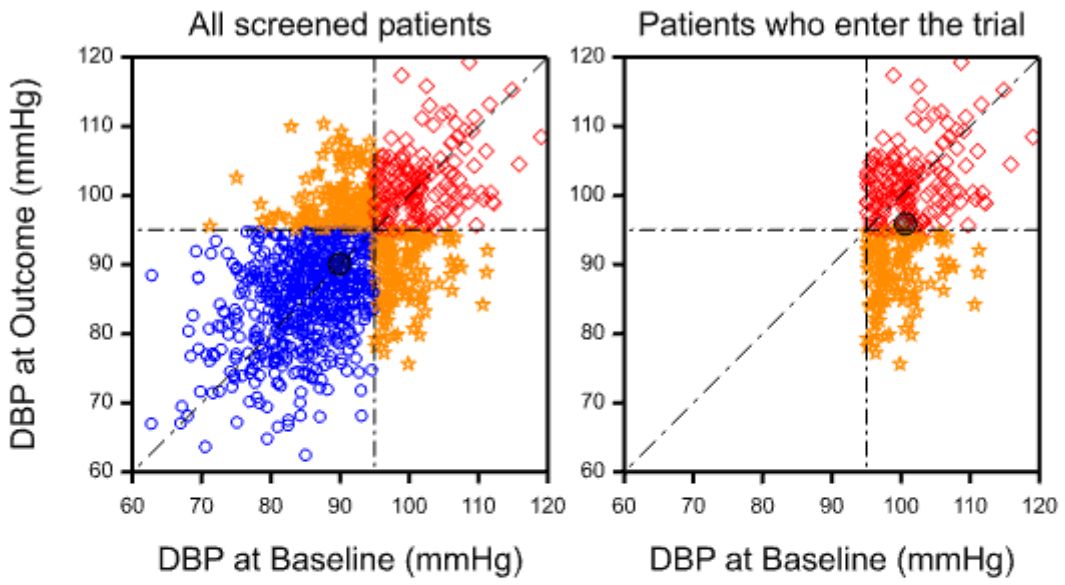

Diastolic blood pressure
DBP at Outcome (mmHg)
DBP at Baseline (mmHg)
120
110
100
90
80
70
60
60
70
80
90
100
110
120


Diastolic blood pressure
DBP at Outcome (mmHg)
DBP at Baseline (mmHg)
120
110
100
90
80
70
60
60
70
80
90
100
110
120


Diastolic blood pressure
DBP at Outcome (mmHg)
DBP at Baseline (mmHg)
120
110
100
90
80
70
60
60
70
80
90
100
110
120


Diastolic blood pressure
DBP at Outcome (mmHg)
DBP at Baseline (mmHg)
120
110
100
90
80
70
60
60
70
80
90
100
110
120


Diastolic blood pressure
DBP at Outcome (mmHg)
DBP at Baseline (mmHg)
120
110
100
90
80
70
60
60
70
80
90
100
110
120



Blue=normotensive at baseline and outcome
Red= hypertensive at baseline and outcome
Orange = inconsistently hyper and normo tensive
Mean at outcome and baseline
dashed diagonal is line of equality at baseline and outcome
LH panel is what we would see if we followed up all patients...
...however we only follow up those hypertensive at baseline
so we see the RH panel
Note that mean at outcome = mean at baseline in LH panel
However mean at outcome < mean at baseline in RH panel

Consequences

- Much of the so-called *placebo effect* may be regression to the mean
- Research findings are often misreported
- Since we usually *define* response in terms of difference from baseline we are in danger of misunderstanding it
  - Such a definition is *not* causal
- Use control!
- Judge by differences to control not to baseline

2. Invalid Inversion

or the Error of the Transposed Conditional

- Invalid inversion occurs when you assume that the probability of A given B is the same as the probability of B given A
- As in ‘The probability that the Pope is a Catholic is one, therefore the probability that a Catholic is the Pope is one’
- This is a common error

The most common example of invalid inversion

- A P-value is the probability of the result given the hypothesis
  - Strictly speaking the probability of a result as extreme or more extreme
- It is not the probability of the hypothesis given the result

A Simple Example

- Most women do not suffer from breast cancer
- It would be a mistake to conclude, however, that most breast cancer victims are not women
- To do so would be to transpose the conditionals
- This is an example of invalid inversion

Some Plausible Figures for the UK

UK Numbers in 1000s		Sex		
		Female	Male	Total
Health Status	Suffering from breast cancer	550	3	553
	Not suffering from breast cancer	30,868	30,371	61,239
		31,418	30,374	61,792

Probability breast cancer given female = 550/31,418=0.018
Probability female given breast cancer =550/553=0.995

A Little Maths

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\text{Unless } P(B) = P(A), P(A|B) \neq P(B|A)$$

So invalid inversion is equivalent to a confusion of the marginal probabilities. The same joint probability is involved in the two conditional probabilities but different marginal probabilities are involved

The Regression Analogue

Predicting Y from X is not the same as predicting X from Y.

$$\beta_{Y|X} = \frac{\sigma_{XY}}{\sigma_X^2}$$

$$\beta_{X|Y} = \frac{\sigma_{XY}}{\sigma_Y^2}$$

Note the similarity with the probability case.
The numerator (the covariance) is a statistic of *joint* variation.
The denominators (the variances) are statistics of *marginal* variation. These marginal statistics are not the same.

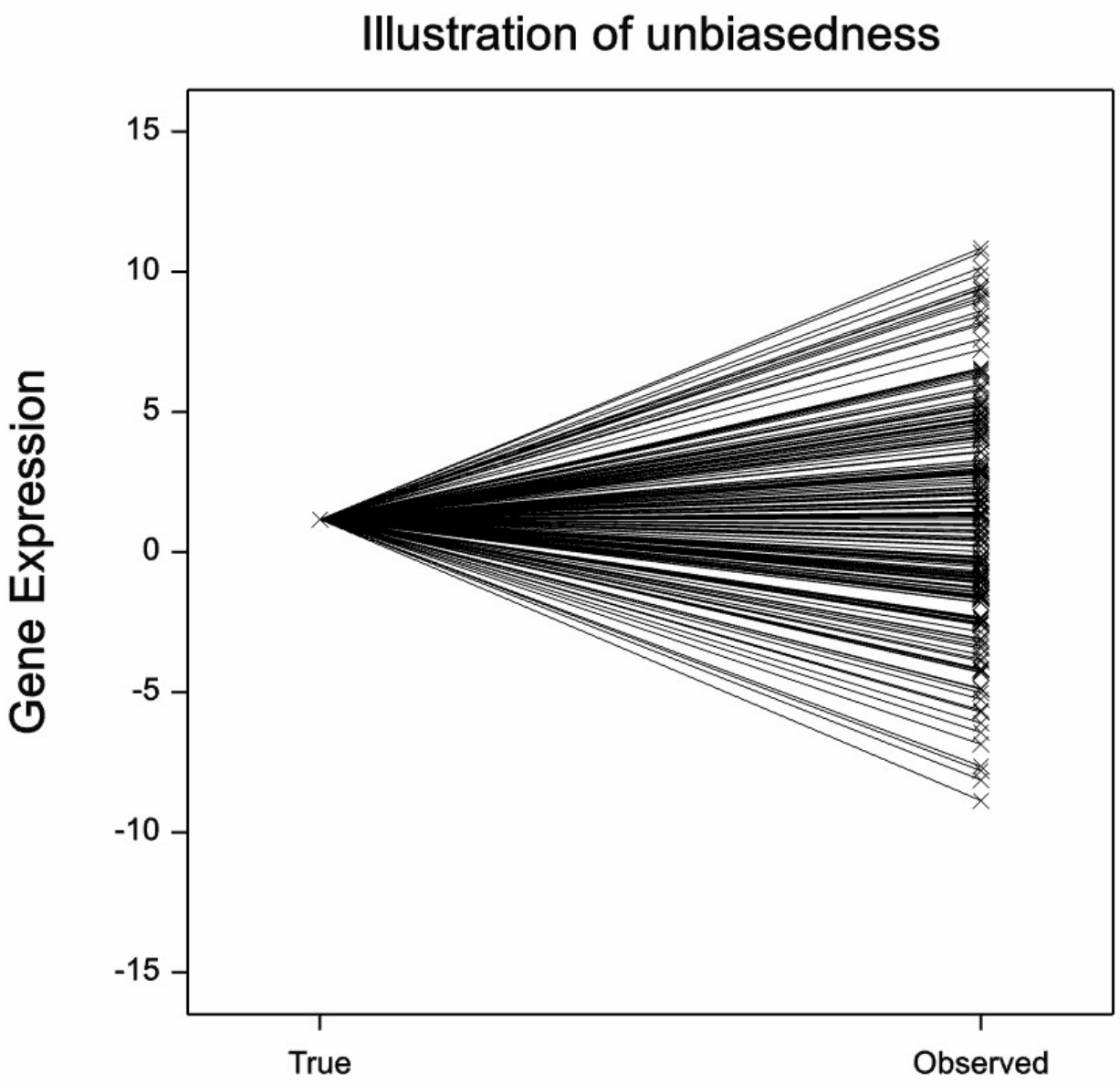


Illustration of unbiasedness
Gene Expression
15
10
5
0
-5
-10
-15
True
Observed

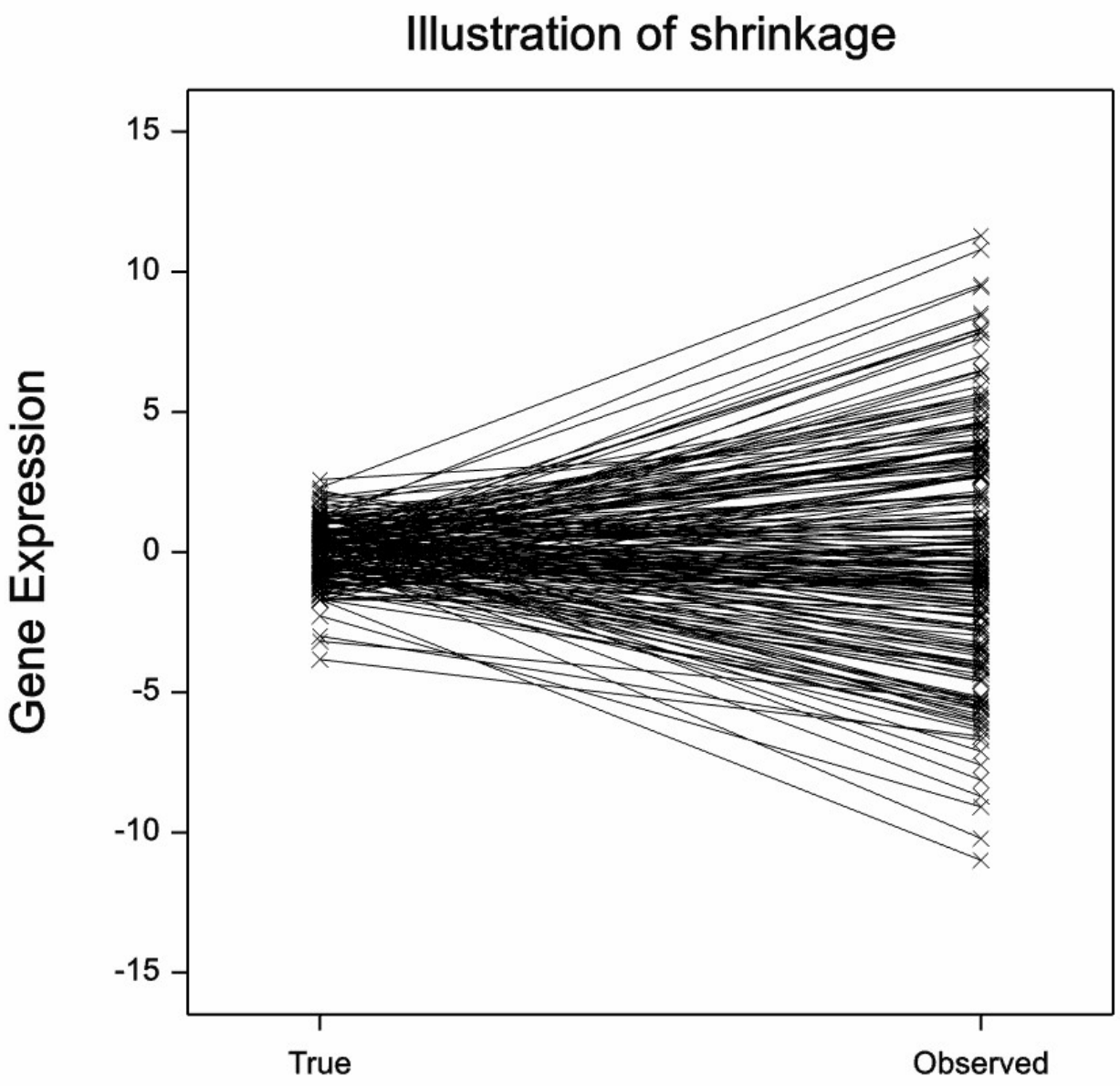


Illustration of shrinkage
Gene Expression
15
10
5
0
-5
-10
-15
True
Observed

Senn's Law

When trying to repeat previous interesting results you can expect to be disappointed – even if you take account of Senn's Law

3. Misinterpreting response

Researchers regularly underestimate that random element of individual response

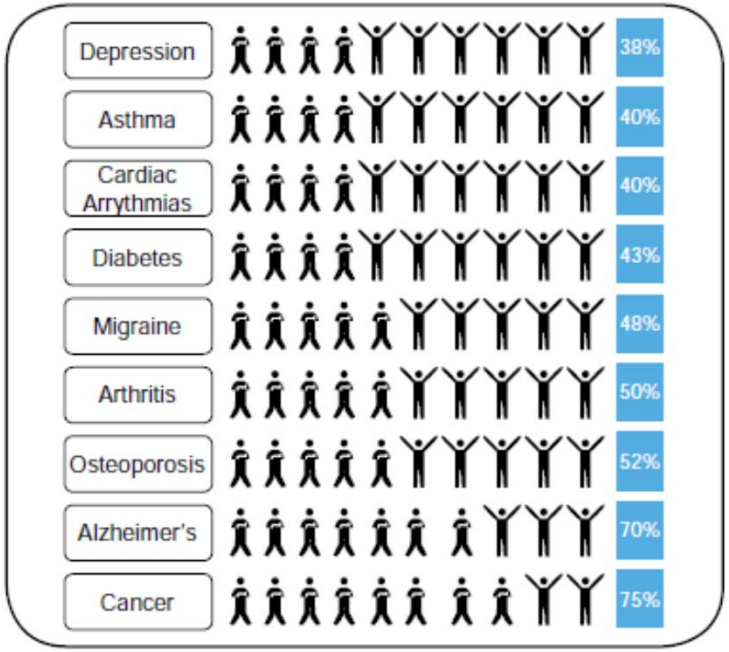
This leads them to over-interpret all differences seen between patients given the same treatments as individual response

A tendency to overhype the potential for personalised medicine Is the consequence

Zombie statistics 1

Percentage of non-responders

What the FDA says



Paving the way for personalized medicine, FDA Oct2013

Where they got it

Table 1. Response rates of patients to a major drug for a selected group of therapeutic areas[1]

Therapeutic area	Efficacy rate (%)
Alzheimer's	30
Analgesics (Cox-2)	80
Asthma	60
Cardiac Arrythmias	60
Depression (SSRI)	62
Diabetes	57
HCV	47
Incontinence	40
Migraine (acute)	52
Migraine (prophylaxis)	50
Oncology	25
Osteoporosis	48
Rheumatoid arthritis	50
Schizophrenia	60

Spear, Heath-Chiozzi & Huff, *Trends in Molecular Medicine*, May 2001

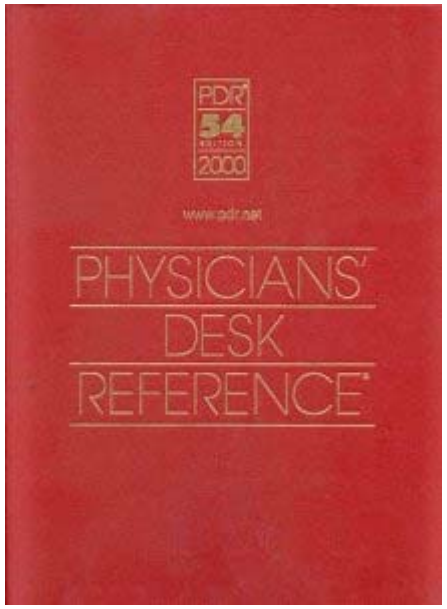
Zombie statistics 2

Where they got it

Where those who got it got it

Table 1. Response rates of patients to a major drug for a selected group of therapeutic areas[1]

Therapeutic area	Efficacy rate (%)
Alzheimer's	30
Analgesics (Cox-2)	80
Asthma	60
Cardiac Arrythmias	60
Depression (SSRI)	62
Diabetes	57
HCV	47
Incontinence	40
Migraine (acute)	52
Migraine (prophylaxis)	50
Oncology	25
Osteoporosis	48
Rheumatoid arthritis	50
Schizophrenia	60



Spear, Heath-Chiozzi & Huff, *Trends in Molecular Medicine*, May 2001

1 Physicians' Desk Reference, 54th Edn., 2000

The Real Truth

- These are zombie statistics
- They refuse to die
- Not only is the FDA's claim not right, it's not even wrong
- It's impossible to establish what it might mean even if it were true

88.2% of all statistics are made up on the spot

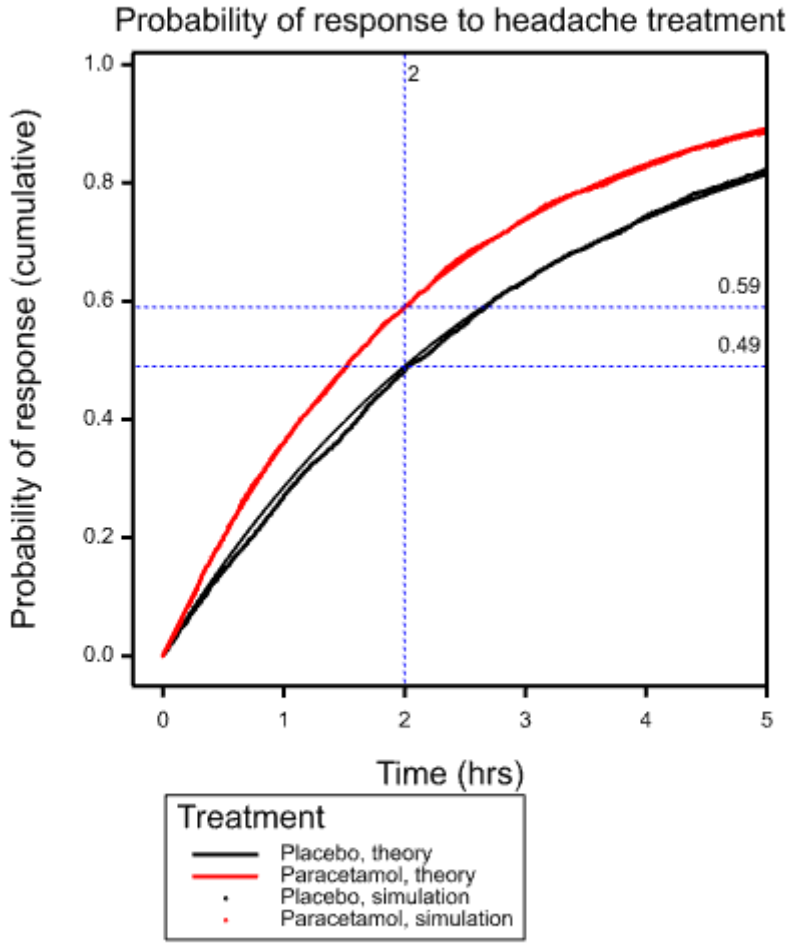
Vic Reeves

An Example

- Cochrane collaboration review of trials of paracetamol in headache
  - 6000 patients in total
- Using a definition of complete response at 2 hours found
  - 59 in 100 taking paracetamol had relief
  - 49 in 100 taking placebo had relief
- Concluded it only worked for 1 in 10
- This is quite wrong

A Simulation to Show Why

- I simulated 6000 patients from an exponential distribution with a mean of about 3
  - Duration of a headache under placebo
- I multiplied each value I generated by ¾
  - Duration of each corresponding headache under paracetamol
- Each patient has a placebo/paracetamol pair
  - The second headache is ¼ less than the first
- But in practice you can only see one of the two
- So I randomly split the patients into two groups
- For one I kept the placebo value and for the other the paracetamol


Probability of response to headache treatment
Probability of response (cumulative)
1.0
0.8
0.6
0.4
0.2
0.0
2
0.59
0.49
0
1
2
3
4
5
Time (hrs)
Treatment
Placebo, theory
Paracetamol, theory
Placebo, simulation
Paracetamol, simulation

Concluding advice

- Be sceptical
- Don’t over-interpret
- Use control
- Think!