A Picture is Worth a 1000 Words - Sometimes

Stephen Senn

CRP SANTÉ

CENTRE DE RECHERCHE PUBLIC

MÉTHODOLOGY AND STATISTICS

"In My Statistics Course"

"I was trying to illustrate a three-dimensional analysis of variance design with a free-hand isometric drawing in graph-like form. Just as I was about to finish the drawing I heard several students laughing loudly. One of the students confessed, 'I think I'd prefer a thousand words'."

Jim Vanderplas, *Journal of Irreproducible Results*

Outline

- Some examples of difficult statistical topics that may be explained graphically
  - Regression to the mean
  - Identifiability and individual response
    - Parallel groups trials
    - Cross-over trials
  - Shrinkage
- Some examples where anecdote and analogy may work better
  - Stage migration
  - Invalid inversion
  - Significance incoherence
- A warning about not expecting too much
- Some tentative advice

What not to do

The first time the European Course in Pharmaceutical Statistics was given, an eminent professor of statistics gave the **_introductory_** lecture of one hour for the stats section

By minute 50 or so he had got to this:

The Gauss-Markov Theorem

$$\text{If } \underset{n \times 1}{Y} = \underset{n \times k}{X} \underset{k \times 1}{\beta} + \underset{n \times 1}{\varepsilon}$$

$$\text{with } E[\varepsilon] = \underset{n \times 1}{0}, E[\varepsilon\varepsilon'] = \sigma^2 I_n$$

$$\text{then } \hat{\beta} = (X'X)^{-1}X'Y \text{ is the BLUE estimator of } \beta$$

Regression to the Mean

- If patients are selected for treatment because of some extreme measured value

  - e.g blood pressure, serum cholesterol, Hamilton score

- Even in the absence of causal effects spontaneous movement towards the mean may be expected

- This is an important and regularly overlooked source of bias in medicine and elsewhere

- It is a clear candidate for something we should communicate

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

Diastolic blood pressure

Diastolic blood pressure
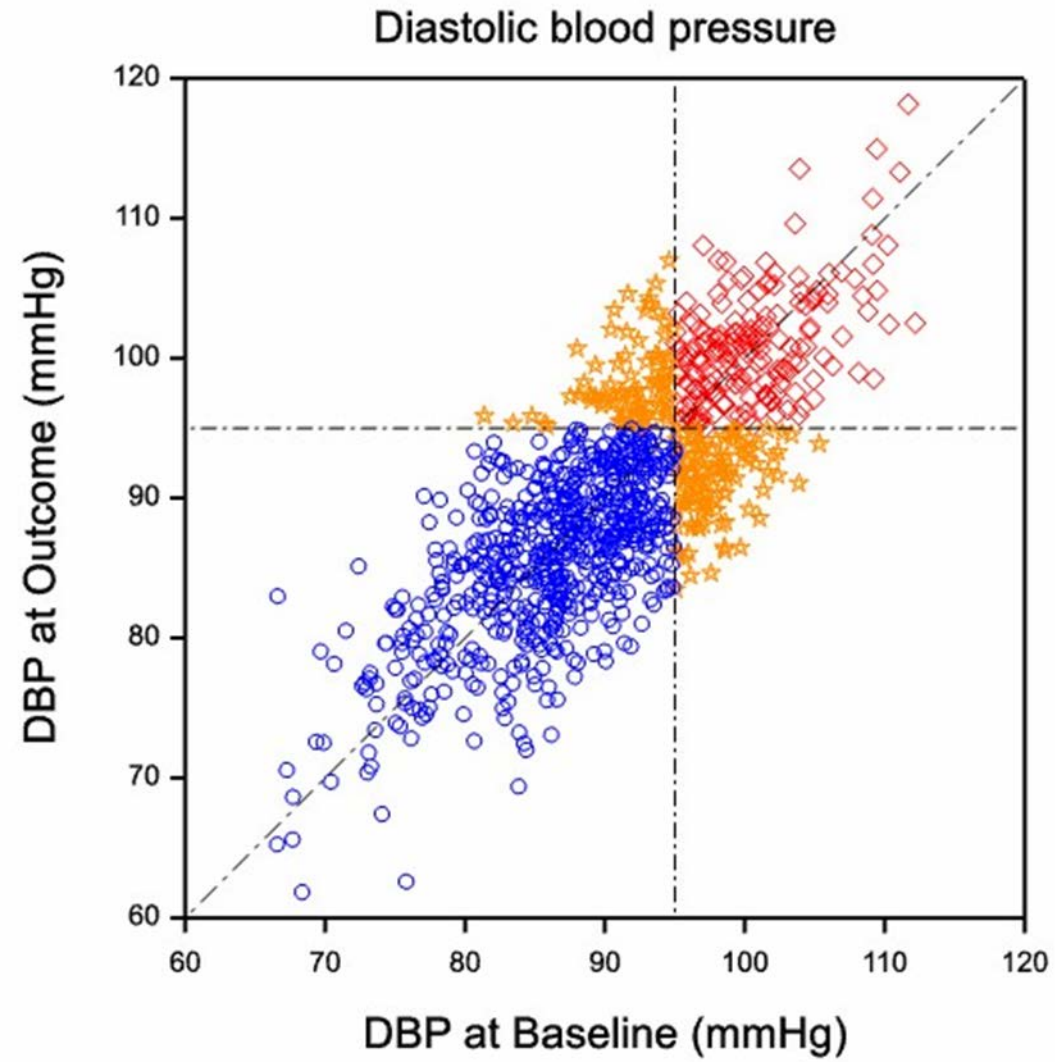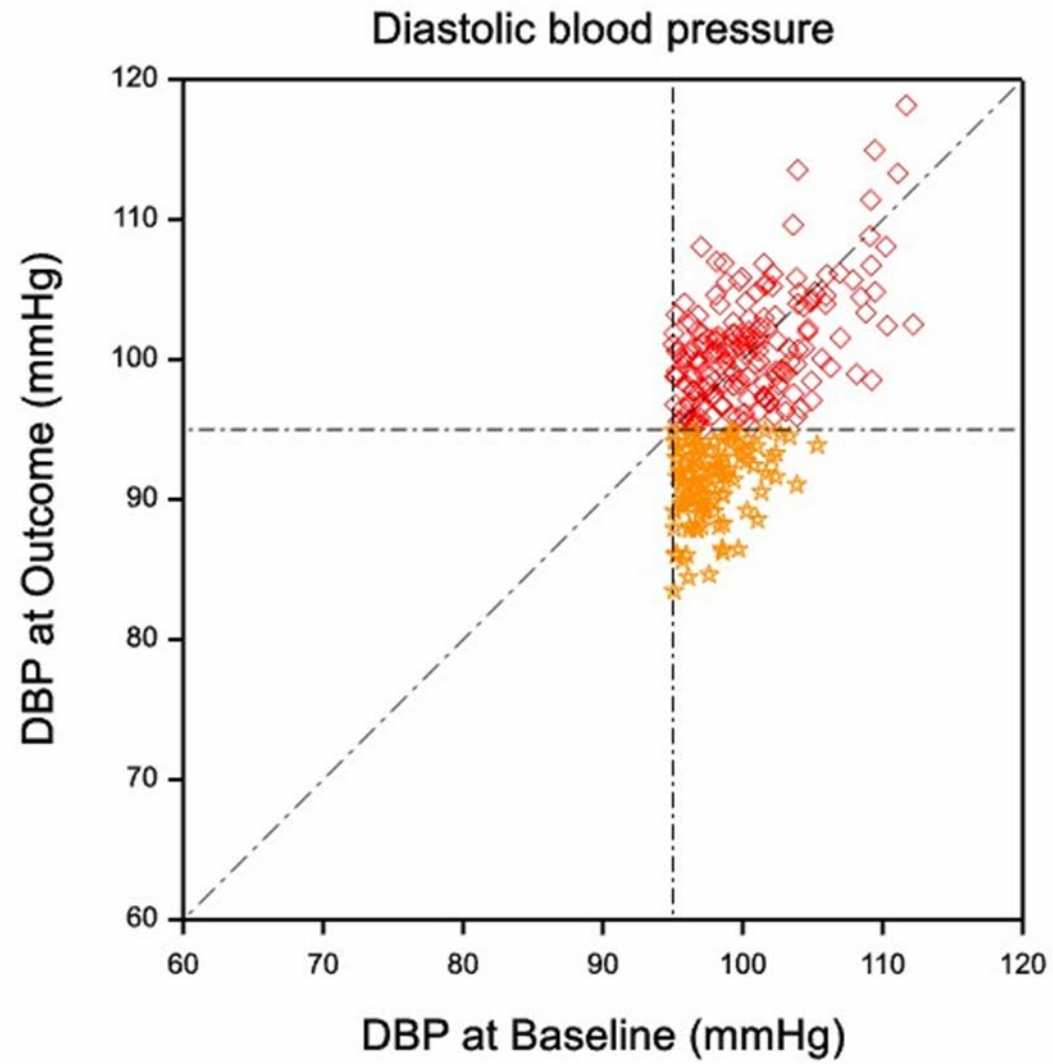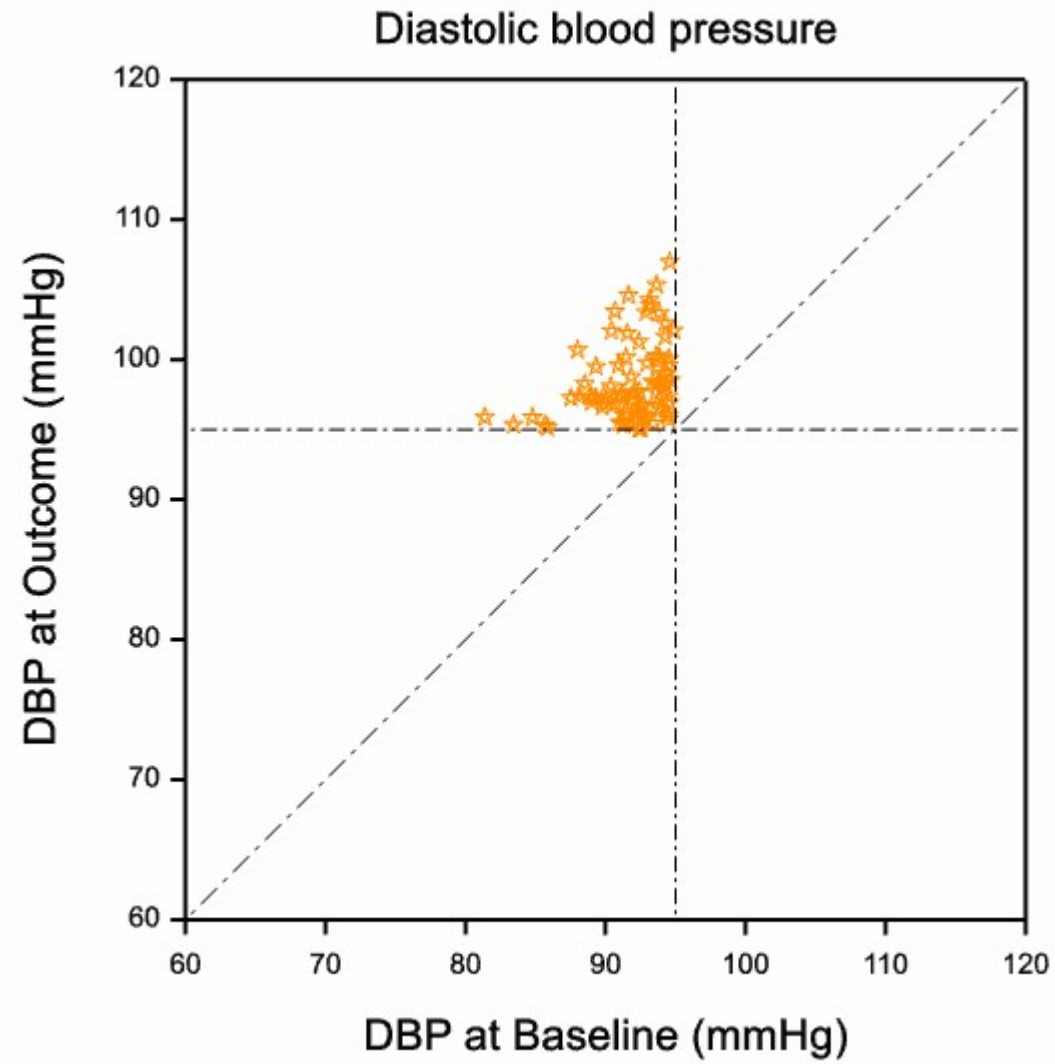
Identifiability and individual response

- Most clinical trials do not permit one to detect individual response
- Failure to understand this is responsible for millions wasted in the pharmaceutical industry
- Wild goose chase for genetic variation and so forth

Sources of Variation in Clinical Trials

Label	Source	Description
A	Between treatments	The difference between treatments averaged over all patients
B	Between patients	The difference between patients given the same treatment
C	Patient-by-Treatment Interaction	The extent to which the effect of treatment varies from patient to patient
D	Within patients	The extent to which the results vary from occasion to occasion for patients given the same treatment

Senn SJ. Individual Therapy: New Dawn or False Dawn. *Drug Information Journal* 2001;35(4):1479-1494.

Identifiability and Clinical Trials

Type of Trial	Description	Identifiable Effects	Error Term
Parallel	Each patient is randomised to receive one treatment	A	B+C+D
Cross-over	Each patient receives each treatment in one period only	A and B	C+D
Repeated cross-overs	Each patient receives each treatment in at least two periods	A and B and C	D

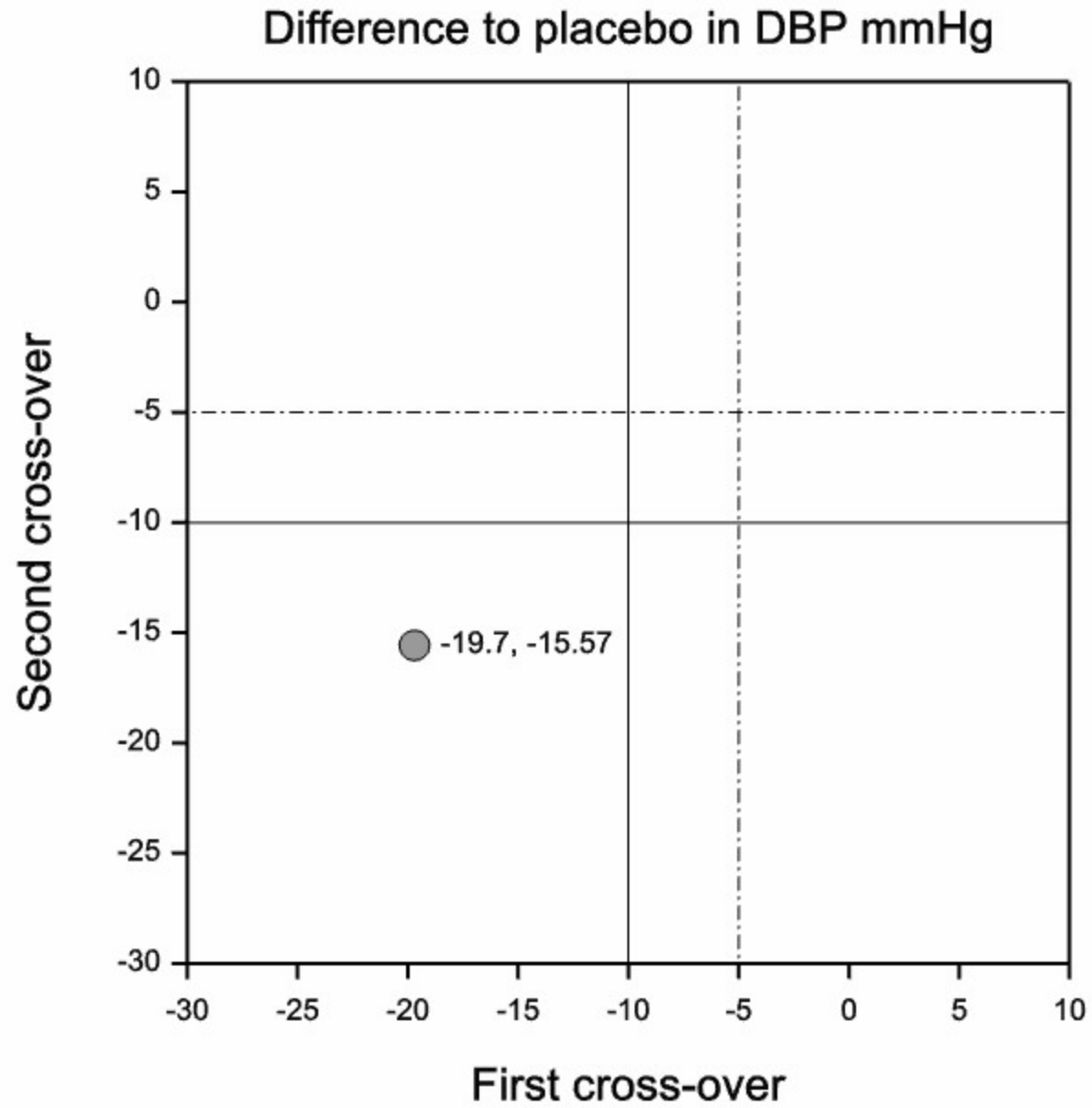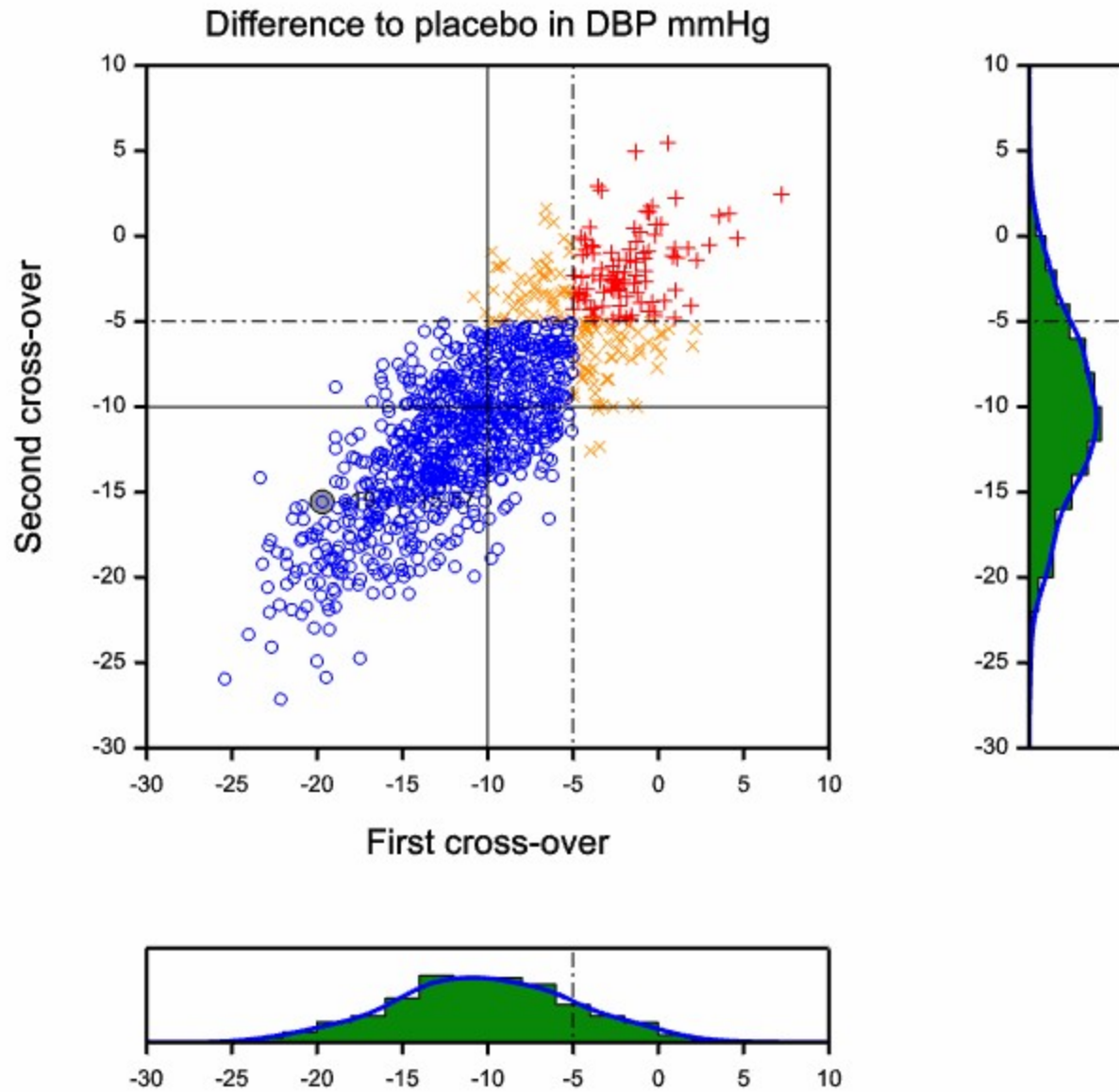Counterfactual experiment
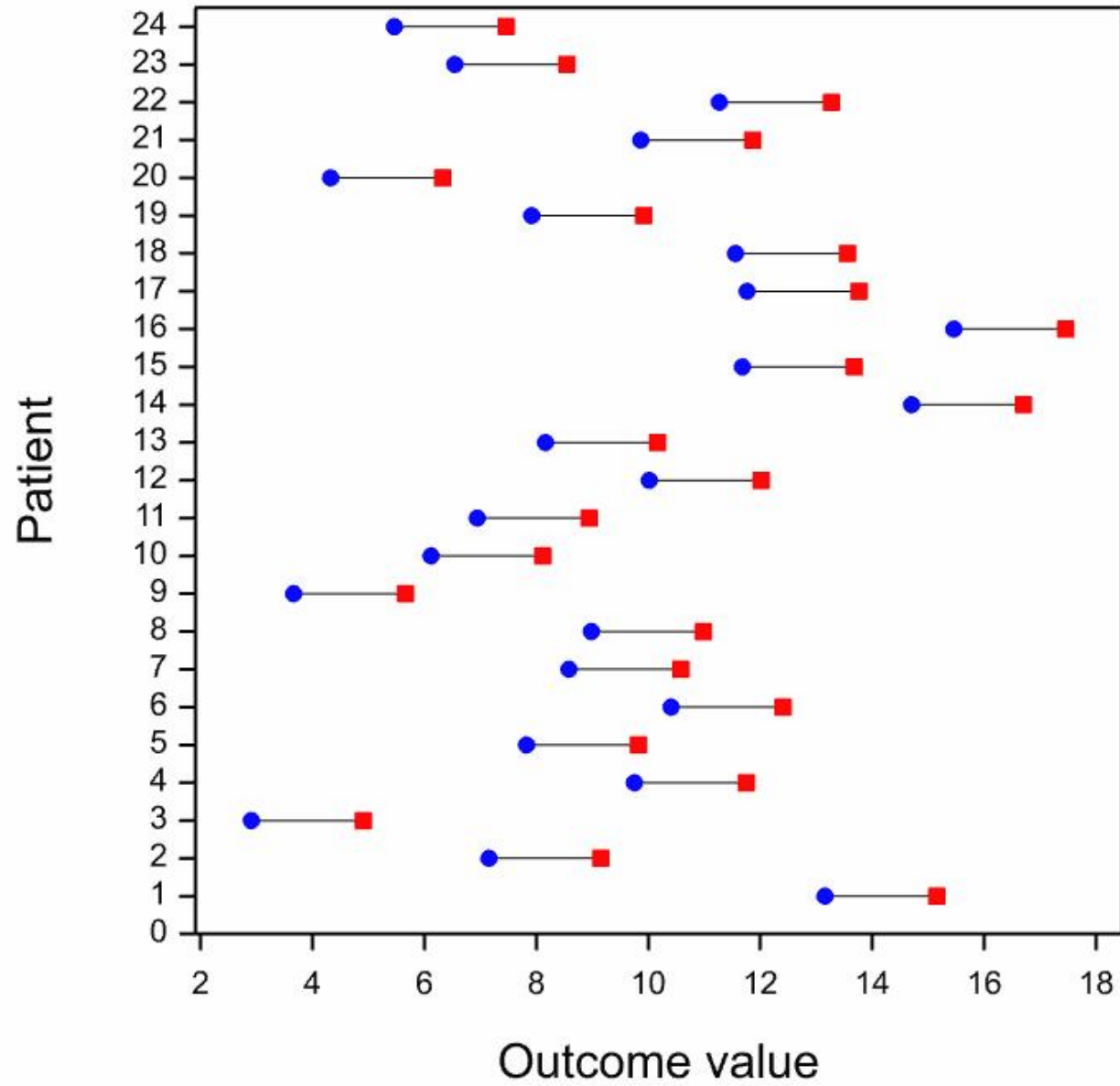
Parallel group trial
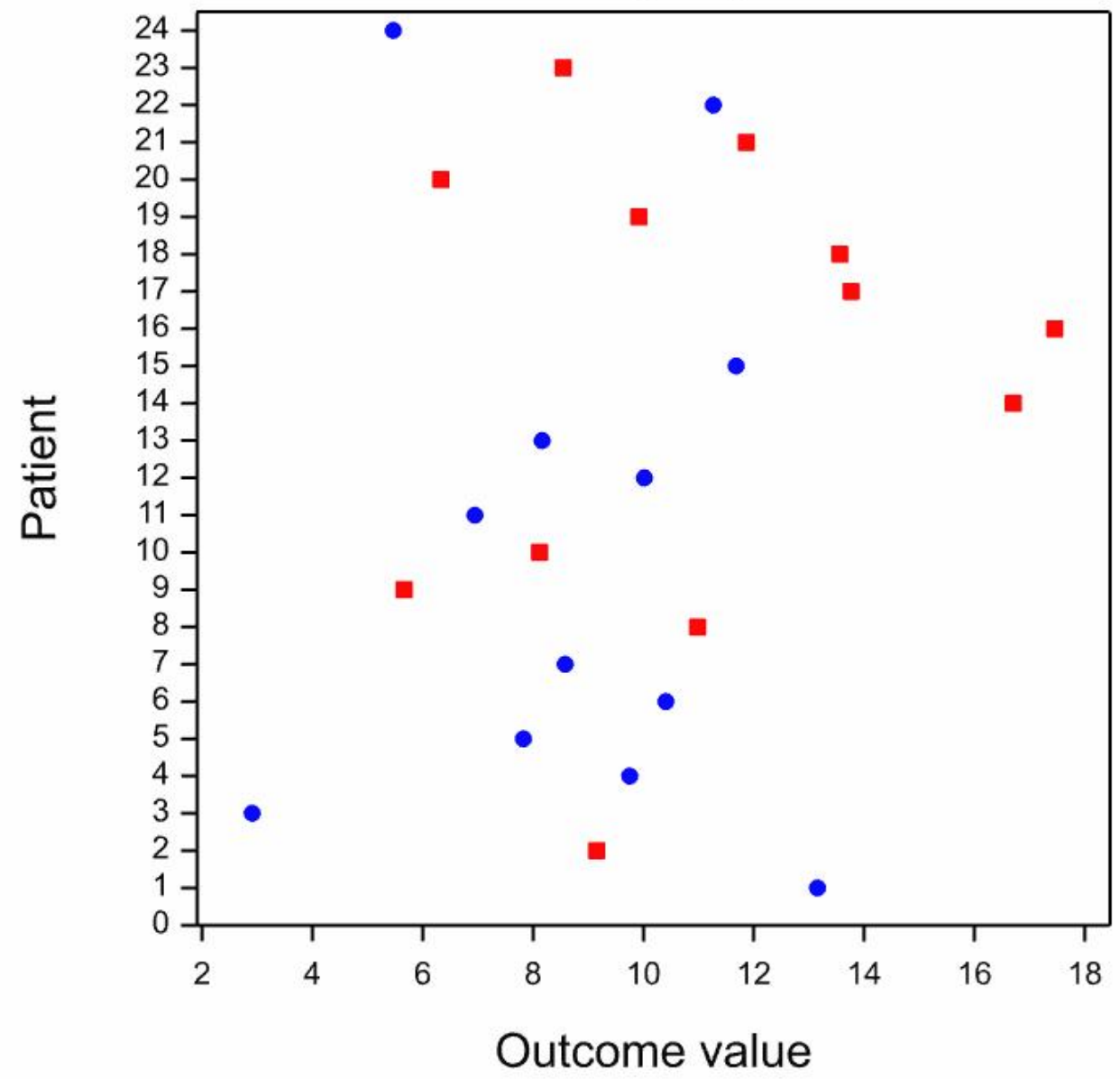
What does it really take to identify individual response?

- Extra-period cross-overs will do the trick
  - Equivalent to sets of n-of-1 trials
- Example of a four period design in hypertension
  - Diastolic blood pressure as main outcome measure
- Patients given placebo and treatment once in each of two pairs of periods
- Definition of 'responder' is 5mm Hg less on treatment than on placebo

Design

Sequence	First Cross-over		Second Cross-over	
	Period			
	1	2	3	4
I	A	B	A	B
II	B	A	B	A
III	A	B	B	A
IV	B	A	A	B

Difference to placebo in DBP mmHg

Difference to placebo in DBP mmHg

Results 1

First Cross-over	Second Crossover		
		Responder	Non-responder	Total
	Responder	781	57	838
	Non-responder	66	96	162
	Total	847	153	1000

Conditional probabilities of observed 'response'

$$\frac{781}{838} = 0.93$$

$$\frac{66}{162} = 0.41$$
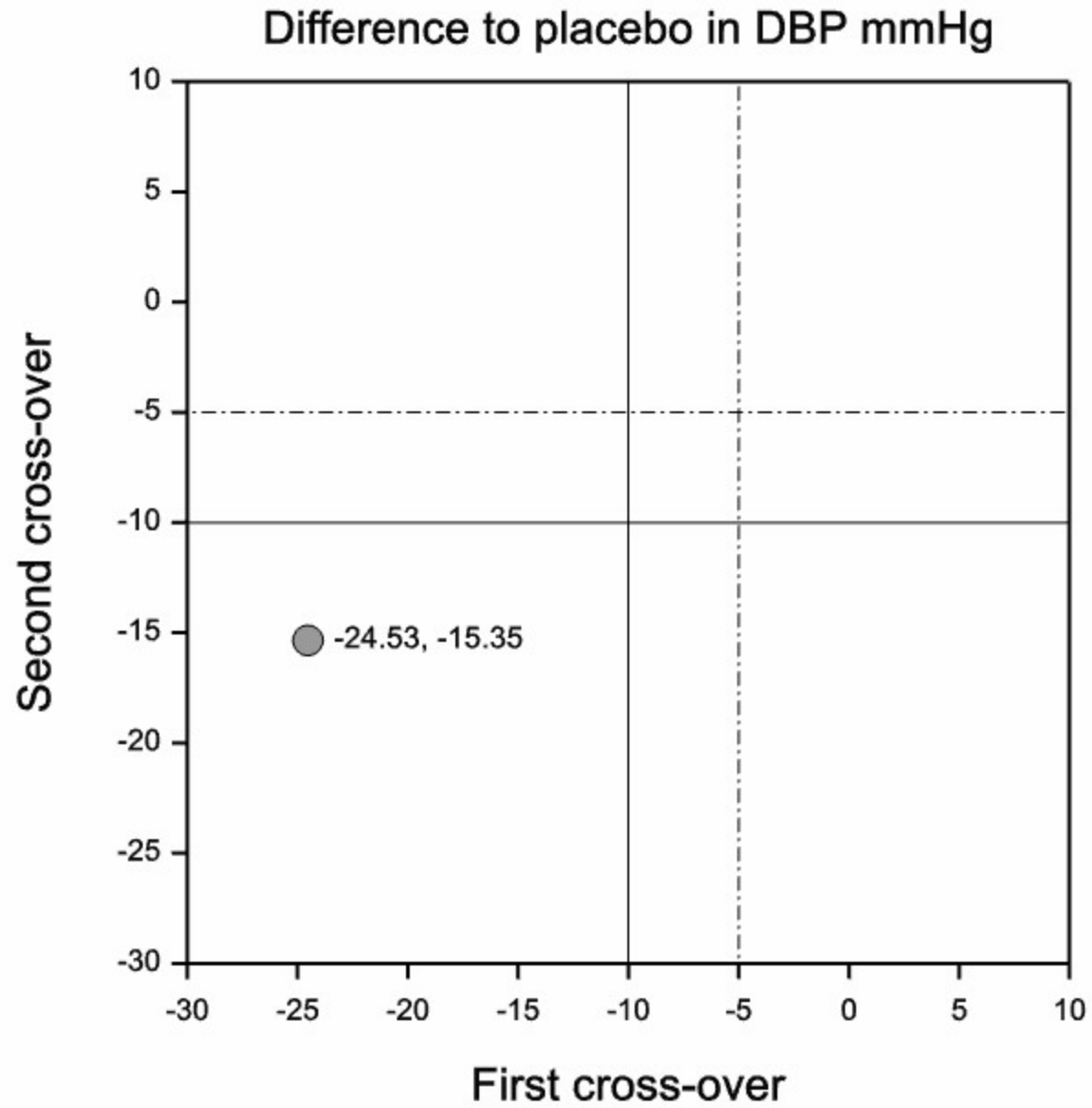
Correlation coefficient is 0.8

Difference to placebo in DBP mmHg

Difference to placebo in DBP mmHg

Results 2

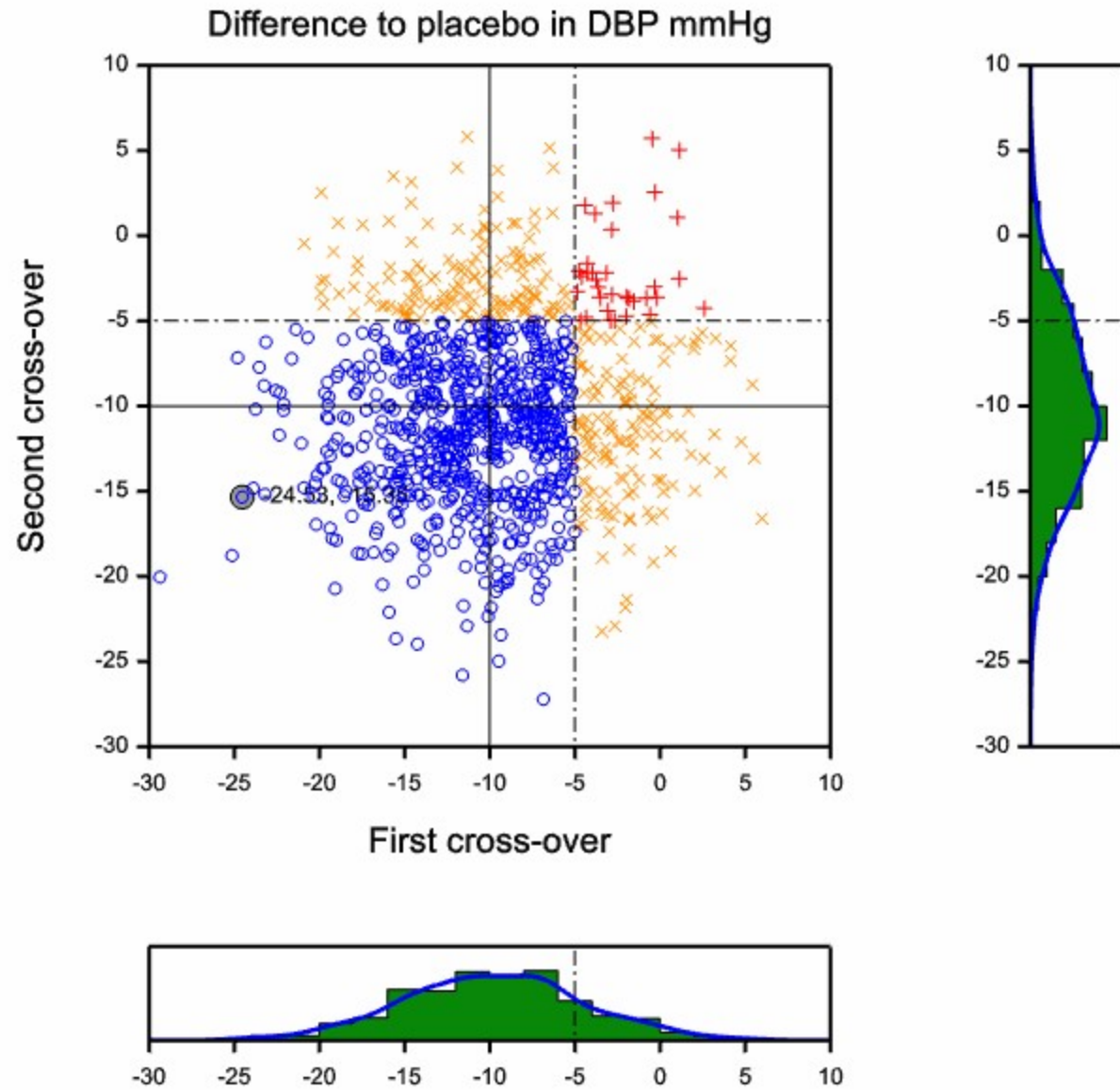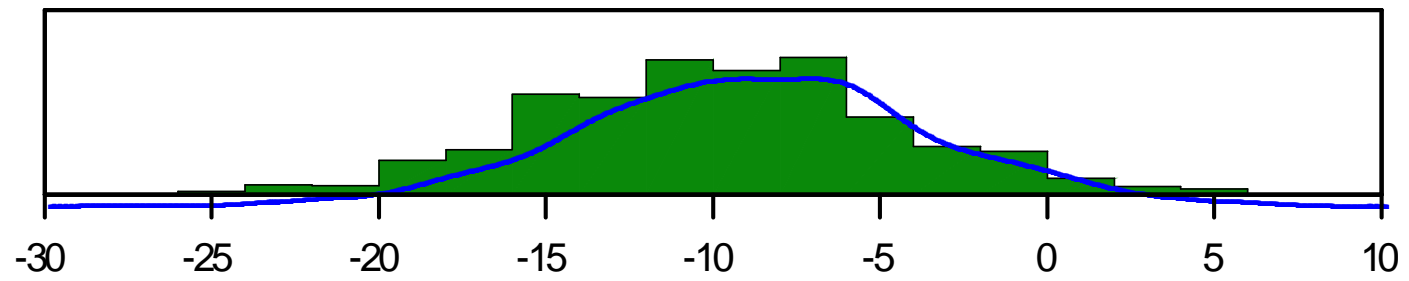| First Cross-over	Second Crossover		
		Responder	Non-responder	Total
	Responder	678	148	826
	Non-responder	140	34	174
	Total	**818**	**182**	**1000**

Correlation coefficient is 0.02

Conditional probabilities of observed 'response'

$$\frac{678}{826} = 0.82$$

$$\frac{140}{174} = 0.80$$

Shrinkage

- Why do so many promising results turn out to be less promising when studied again?

- Particularly noticeable in micro-array work

- Huge effects that largely disappear

- An example of how <u>not</u> to do it follows

How not to do it

τ true ' treatment effect, ϵ experimental error, Y observed effect

$$Y = \tau + \varepsilon$$

$$E[\tau] = \mu, V[\tau] = \gamma^2, E[\epsilon] = 0, V[\epsilon] = \varphi^2 (n)$$

Here $\varphi^2(n)$ is the variance of the experimental result. It is a decreasing function of the number of observations. For example in a parallel group trial with patients allocated with equal probability to one of two groups we have $\varphi^2(n) \cong 4\frac{\sigma^2}{n}$.

The unconditional variance of Y is $V(Y) = \gamma^2 + \varphi^2(n)$ and the ratio of the variance of I to the variance of τ is $\rho(n) = \gamma^2/\{\gamma^2 + \varphi^2(n)\}$.

How not to do it (continued)

Variance-Covariance Matrix

|	τ	Y
τ	γ^2	γ^2
Y	γ^2	$\gamma^2 + \varphi^2(n)$

$$E[Y|\tau] = \mu + \frac{\gamma^2}{\gamma^2}(\tau - \mu) = \tau$$

$$E[\tau|Y] = \mu + \frac{\gamma^2}{\gamma^2 + \varphi^2(n)}(Y - \mu) = \mu + \rho(n)(Y - \mu) \neq Y$$

Don't do this

To put it another way, it does not follow from the fact that Y is an unbiased estimate of τ that on average $\tau = Y$
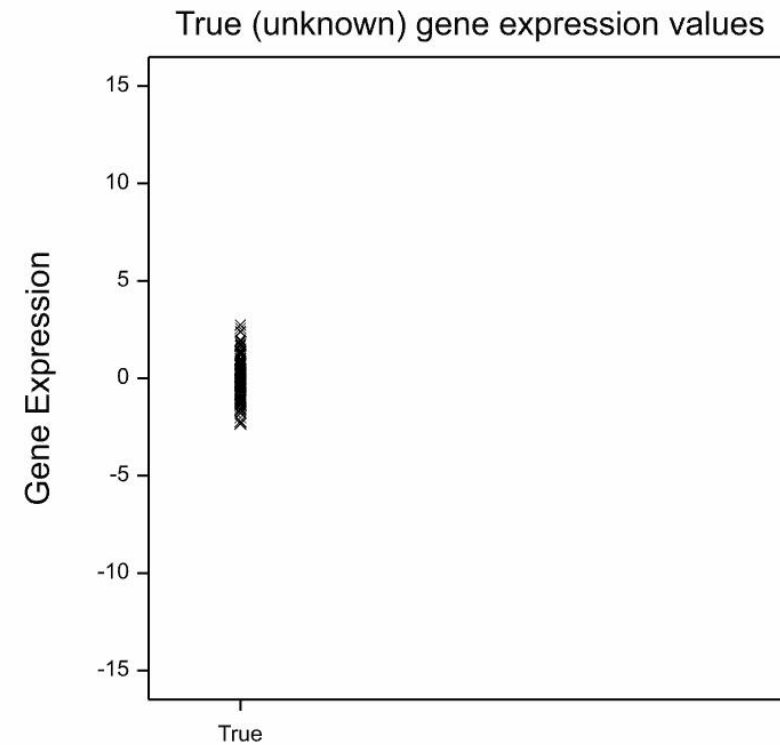
Gene expression 1

- Suppose we have one gene-expression that we could measure many times with high variability

- If the statistic we use to measure this is unbiased this is the sort of thing we will see

- The average position on the right is equal to the true position on the left


Illustration of unbiasedness
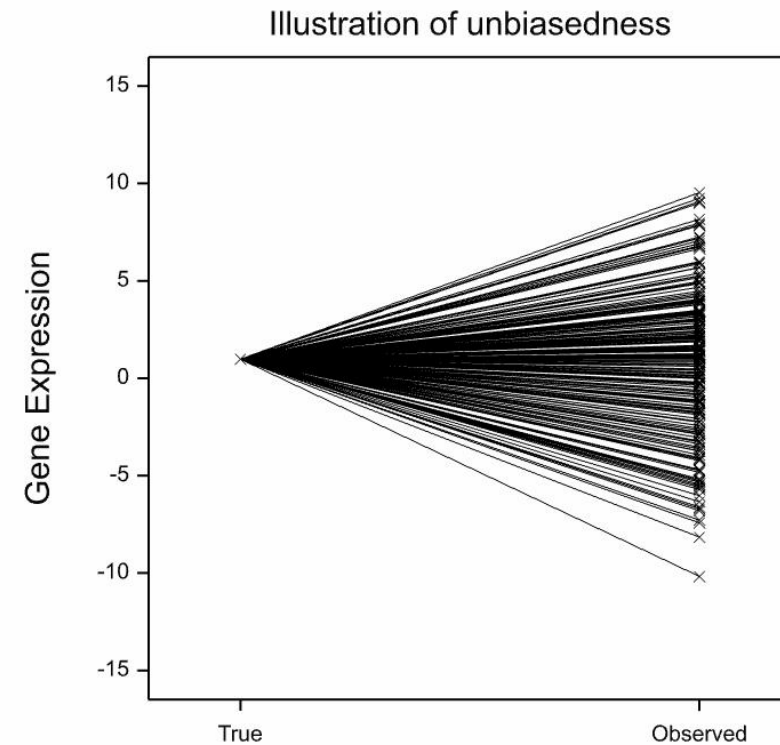
Gene expression 2

- Now consider the distribution of the true (unknown) values of gene expression for many genes

- If we could see these true values we might see something like this

True (unknown) gene expression values

Gene expression 3

- Now let's consider what happens when we measure many genes once

- Now pick an observed value on the right

- If your value was extreme then on average the true value on the left is closer to the mean of all true values than to the value you picked

- Moral: direct unbiasedness does not translate into indirect unbiasedness

- Values shrink

Illustration of shrinkage
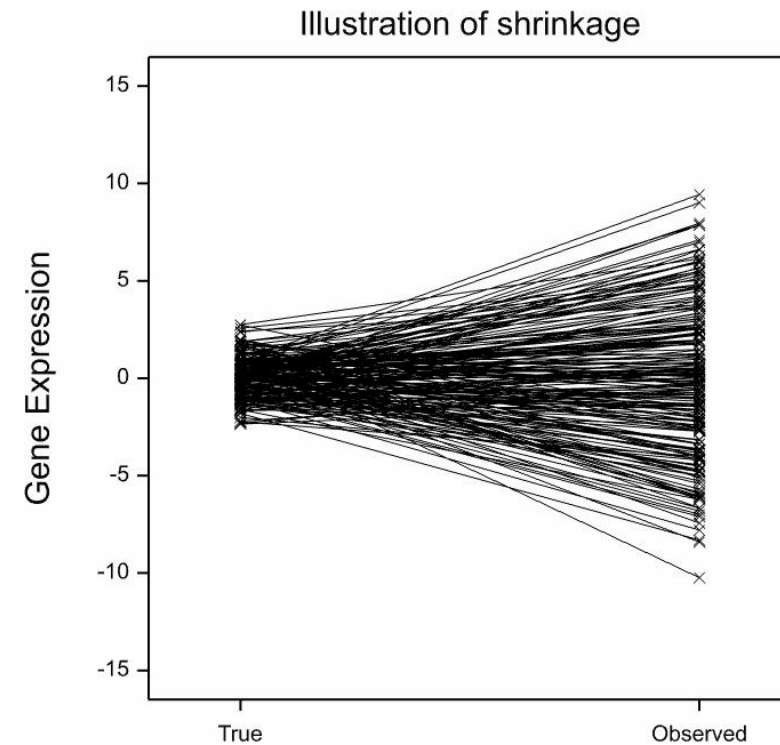
Stage migration

- A phenomenon whereby reclassification of patients can lead to improvement in every subgroup despite no improvement overall
  - Example
    - Change the definition of pre-term for babies from 32 to 34 weeks
- A similar thing can happen with changing intervention strategies
  - Example
    - Increase the proportion of hospital compared to home births

Stage migration phenomenon
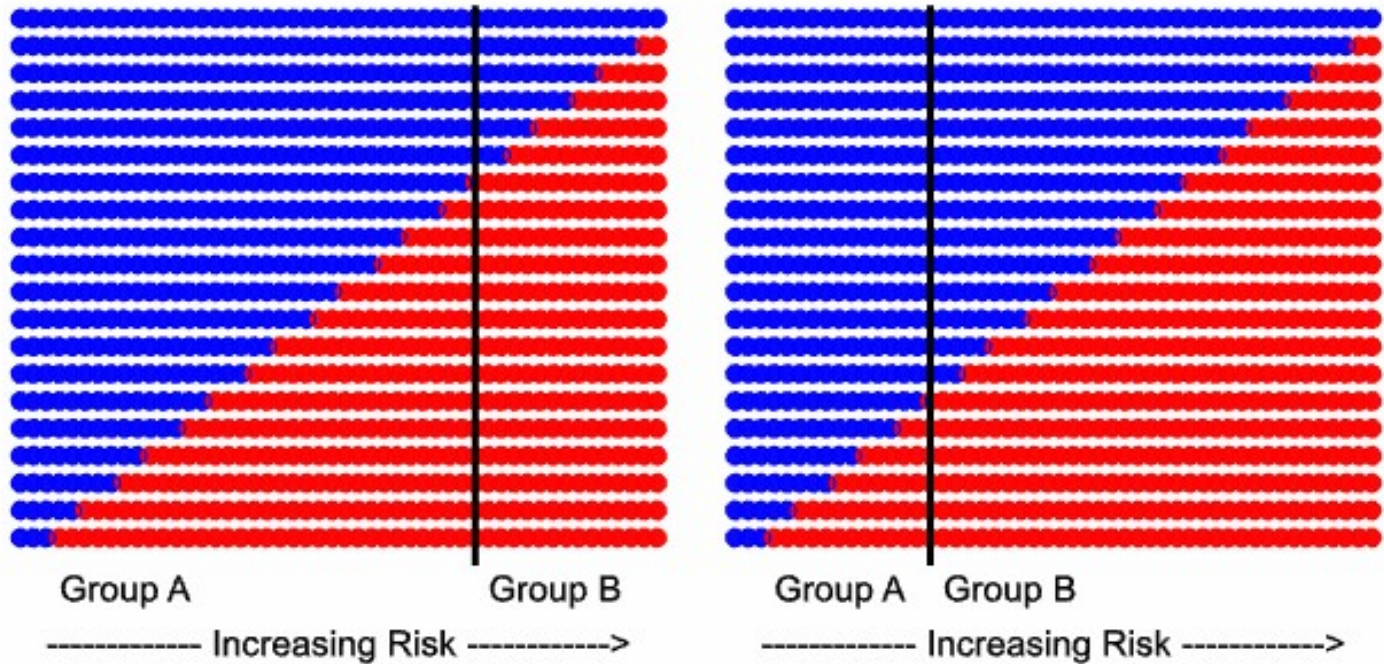
Blue is good outcome
Red is bad outcome
By changing group definition by moving boundary to the left (as in RHS)
we can improve proportion of good outcomes in both groups

Don't do this?

The Will Rogers Phenomenon

When the Okies moved from Oklahoma to California the average intelligence was improved in two states

Invalid inversion

Invalid inversion occurs when you mistake the probability one way for the probability another.

Most common example: mistaking the probability of the data given the hypothesis for the probability of the hypothesis given the data.

$$P(Data|Hypothesis) \neq P(Hypothesis|Data)$$

It is <u>not</u> a good idea to explain this using the probability calculus in an abstract manner. Once people understand the phenomenon, then you can think about such explanations if they are interested.

Instead, make it concrete.

Invalid inversion explanation by concrete example

Most women do not get breast cancer.

It does not follow that most breast cancer victims are not women.

Probability of a randomly chosen woman having breast cancer is not the same as the probability of a randomly chosen breast cancer victim being a woman.

The probability of B given A is not the same as the probability of A given B.

Is the Pope a Catholic?

Yes!

Is a Catholic the Pope?

Almost certainly not

Incoherence in significance tests

- This can happen when a global F-test suggests rejecting a null hypothesis that all treatments are equal; however pair-wise t-tests do not lead to rejection of equality of any pair

- This appears paradoxical

- However, evidential paradoxes occur everywhere in life

- Here's an example

A paradox of parenting

- A mother leaves her two children playing peacefully in the playroom while she goes to prepare lunch
- Soon after she hears that a squabble has broken out and by the time she returns her kids are fighting
  - "He started it"
  - "No she started it"
- She knows that at least one child is guilty of aggressive behavior but she can't know for sure that any given child is
- So in statistics, just as in life, you can sometimes detect that something has happened without being able to pinpoint what

A warning

- All this has its limits
- It is laudable to try and increase mutual understanding by explaining statistical concepts
- Vice versa – we biostatisticians should try understand some basic medical ideas
- But our medical colleagues are not going to turn us into physicians
- We are not going be able to turn most of our medics into statisticians
- Simple explanations are good but this does not mean being simple-minded
- If our medical colleagues think all statistics should be simple where do they stand on immunology?

My advice

- Stick to time
- Be (mentally) in the audience as well as on the stage
- What your audience can understand trumps what you want them to know
- Avoid algebra unless strictly unavoidable
- Think of good illustrations
  - Often graphs
  - Sometimes anecdotes
- Prepare to be disappointed
- Stick to time