Approximate analysis of covariance in trials in rare diseases, in particular rare cancers

Stephen Senn

LUXEMBOURG INSTITUTE OF HEALTH
RESEARCH DEDICATED TO LIFE

Acknowledgements

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"

An apology

This is a work in progress

Outline

- Preliminaries
- Covariate adjustment for the linear model
  - Less relevant for cancer but helps to raise some issues
  - Effects on efficiency
  - Possible approaches
    - External regression
    - Intermediate regression
    - Augmented regression
    - Bayesian approaches?
- Covariate adjustment for non-linear models (e.g. proportional hazards, logistic regression)
  - Greater relevance for cancer
  - What changes?
- Conclusions

Preliminary: three perspectives on estimation

	Perspective		
Issue	1. Experimental	2. Multivariate	3. Regression
Population	All possible randomisations of patients studied	Target population?	Who cares? We're modelling
Mechanism	Randomisation	Sampling (which we shall pretend is random)	Who cares? We're modelling
Stochasticity	Randomisation induced	Multivariate Normal of predictors and outcomes	There's an error term thing there because we know the models don't really fit
Purpose	Causal Did treatment have an effect	Causal/ predictive Did/will treatment have an effect	Causal/predictive Did/will treatment have an effect

The framework I shall choose

I shall follow time-honoured tradition of mixing around these three approaches while pretending to know what I am talking about

Basically, I regard the framework of 1 as most secure but the framework of 3 as being easiest to handle but I am going to have to treat regressors as stochastic and that means I am going to have to shimmy into 2 from time to time.

A fundamental formula and a paradox

$$Var(Y) = E[Var(Y|\mathbf{X})] + Var(E[Y|\mathbf{X}])$$

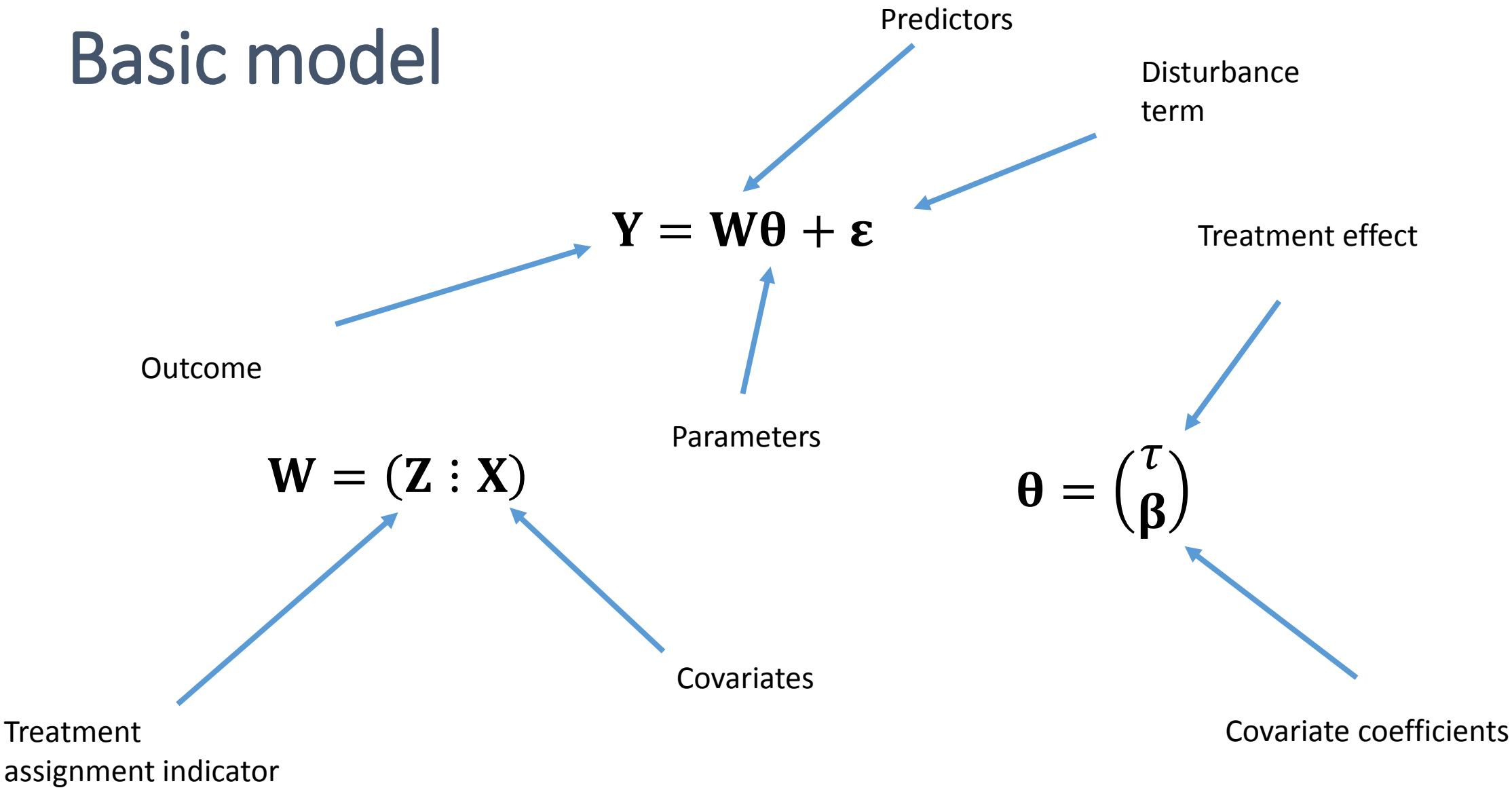
$$E[Var(Y|\mathbf{X})] = Var(Y) - Var(E[Y|\mathbf{X}]) \leq Var(Y)$$

Thus the conditional variance is less than the marginal variance.
Analysis of Covariance is means of conditioning on covariates in the linear model
It would thus seem logical that the variance of the treatment estimate having fitted using ANCOVA should be less than that when you don't condition

Modelling is good!

However, things are not that simple…….

Basic model


Predictors
Disturbance term
$\mathbf{Y} = \mathbf{W}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$
Treatment effect
Outcome
Parameters
$\mathbf{W} = (\mathbf{Z} : \mathbf{X})$
$\boldsymbol{\theta} = \begin{pmatrix} \tau \\ \boldsymbol{\beta} \end{pmatrix}$
Covariates
Treatment assignment indicator
Covariate coefficients

The consequences of adding covariates 1 & 2

First order efficiency

It can be shown that the variance of the estimate of the treatment effect under ANCOVA is

$$\mathrm{var}\left(\hat{\tau}\middle|\mathbf{X}\right)=\left(\frac{1}{1-\frac{n_1 n_2}{n}\mathbf{D}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{D}'}\right)\frac{n}{n_1 n_2}\hat{\sigma}^2_{\mathbf{X}}=\lambda_{\mathbf{X}}\frac{n}{n_1 n_2}\hat{\sigma}^2_{\mathbf{X}}$$

Where **D** is the vector of the mean differences in **X** between treatment groups

Now suppose we have two covariate matrices $\mathbf{X}\subset\mathbf{X}^{+}$ so that the second is just the first with some columns corresponding to some additional covariates added, then the following is the case

$$\lambda_{\mathbf{X}^{+}}\geq\lambda_{\mathbf{X}},\quad E\left[\hat{\sigma}^2_{\mathbf{X}^{+}}\right]\leq E\left[\hat{\sigma}^2_{\mathbf{X}}\right]$$

The consequences of adding covariates 3

Second order efficiency

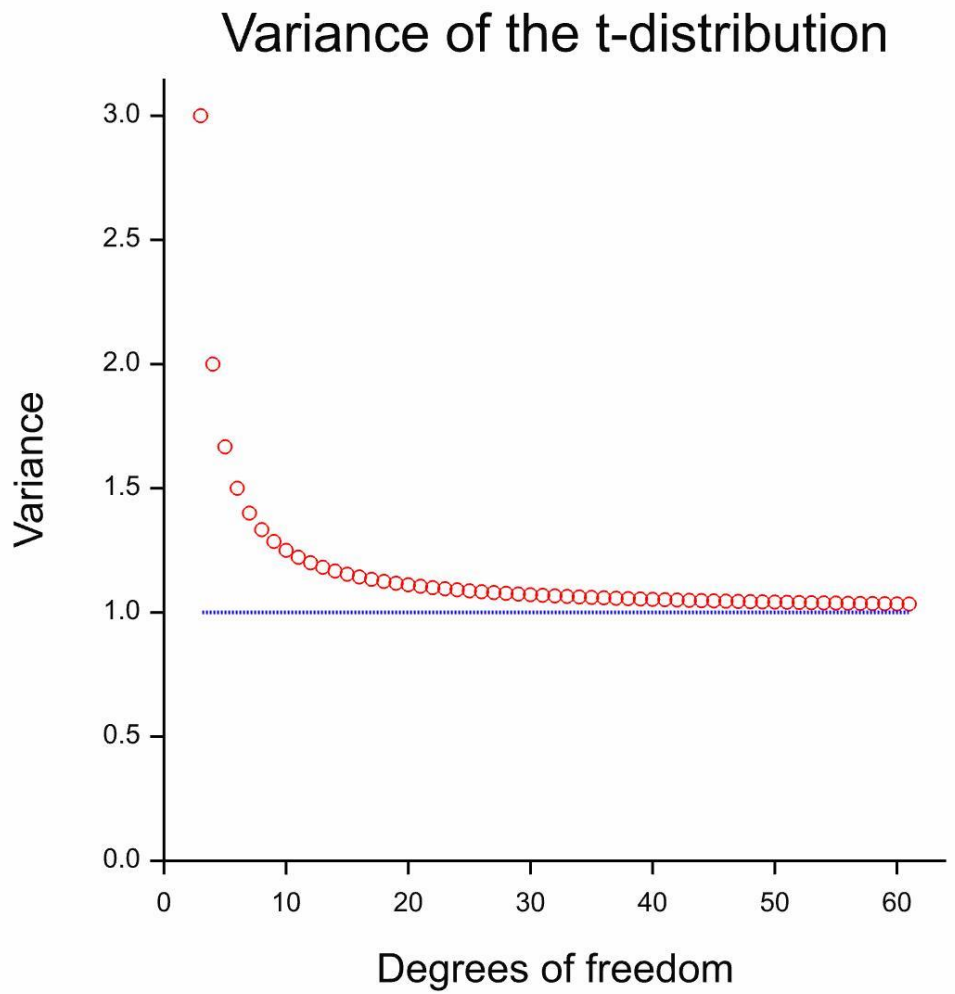
One degree of freedom is lost for every covariate fitted

Perhaps best understood in terms of its effect on the variance of the t distribution

$$\mathrm{var}(t) = \frac{\nu}{\nu - 2} = \frac{N - 2 - k}{N - 4 - k}$$

Where *N* is the number of patients and *k* is the number of covariates

Joint effect of 1st and 2nd order efficiency is controversial (Gilmour & Trinca 2012)



Fisher to Nelder

It must be the peculiarities of your teaching at Cambridge which led you to think that some other method is more authentic……..

Probably, however, you were not taught to regard the fiducial distribution of μ as a frequency distribution at all

3 December 1956, Bennett 1990, p283

To sum up

- Adding predictive covariates to a model makes the residual error smaller
- But it makes the design matrix somewhat less well-conditioned
- Second order efficiency is also affected
    - Fewer degrees of freedom for estimating the error variance
    - Less favourable t-distribution for confidence intervals
- Eventually as we add covariates we lose
- Problem in small trials

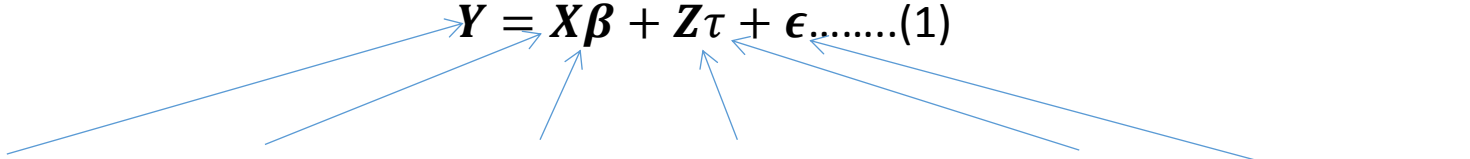
The Gauss-Markov theorem

- This shows that Ordinary Least Squares is optimal if either
  a) you have fixed regressors or
  b) you have stochastic regressors but require conditionally unbiased estimates
- However, if you have stochastics regressors and are happy with marginal unbiasedness you may be able to do better
- Common example is recovering inter-block information
  - The block effects are treated as random
  - We no longer condition on them
- So here are some suggestions

Auxiliary adjustment

Our regression model is as follows

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z}\tau + \boldsymbol{\epsilon} \text{.....(1)}$$



Outcome Covariates Coefficients Indicator Treatment Effect Disturbance

We replace it with

$$\boldsymbol{Y} - \boldsymbol{X\beta}^* = \boldsymbol{Z}\tau + \boldsymbol{\epsilon}^* \text{.....(2)}$$

Where $\boldsymbol{\beta}^*$ is a guesstimate for the vector $\boldsymbol{\beta}$
Since this is a randomised design if we apply ordinary least squares to (2) we have an unconditionally unbiased estimate of τ (it is not conditionally unbiased).
It is true that $\mathrm{E}[Var(\epsilon)] \leq [Var(\epsilon^*)]$ but $\lambda \geq \lambda^*$
Thus, nevertheless, it is possible, due to the reduction in dimensions, that

$$E\left[\operatorname{var}\left(\hat{t}^*\right)\right] < E\left[\operatorname{var}\left(\hat{t}\right)\right]$$

Variations on this

1. Pure adjustment. We use a predicted value using local covariates but slope estimates based on a previous data set.
   Example using table of predicted forced expiratory volume in one second (FEV_1) in asthma based on population surveys and using age, height and sex of the subject

2. Partial adjustment. We use two or more covariates to form a predicted value but the use this as a single covariate in ANCOVA
   Example using predicted FEV_1 based on age, height and sex as a covariate rather than using age, height and sex

3. Augmented regression

The expected value of

$$\lambda = \frac{1}{1-\frac{n_1 n_2}{n}\mathbf{D}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{D}'}$$

Increases as we increase the number of columns but reduces if we increase the number of rows. We create an augmented data matrix consisting of 'other' treatments but the same covariates

Toy example

Cross-over trial in asthma comparing 7 treatments in 5 periods & 158 patients. (Senn et al, 1997)

I have constructed a parallel group trial by dropping all periods except the first.

I have chosen two of the treatments (placebo plus another*) to form a new trial but retained the other 5 to allow me to estimate an external outcome on baseline slope

I have randomly chosen six (three for each treatment) patients and compared ANCOVA to external adjustment

I have repeated this 100 times

*Formoterol ISF 6µg

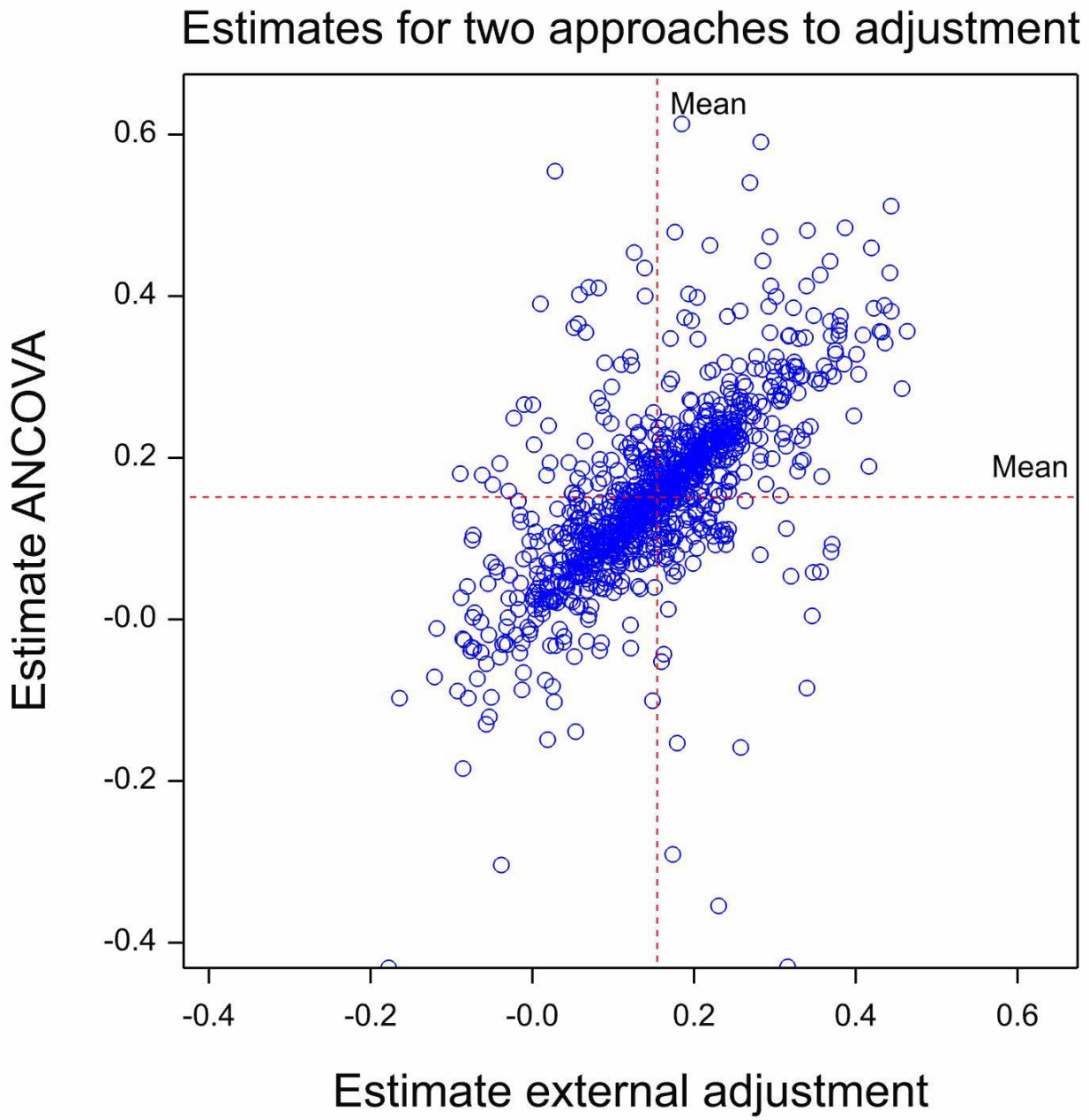
Simulation 1000 samples of size 6

Summary statistics for Estimate external adjustment

Mean =	0.151
Median =	0.150
Variance =	0.0104

Summary statistics for Estimate ANCOVA

Mean =	0.155
Median =	0.150
Variance =	0.0125



Simulation 1000 samples of size 6

Summary statistics for Variance external adjustment

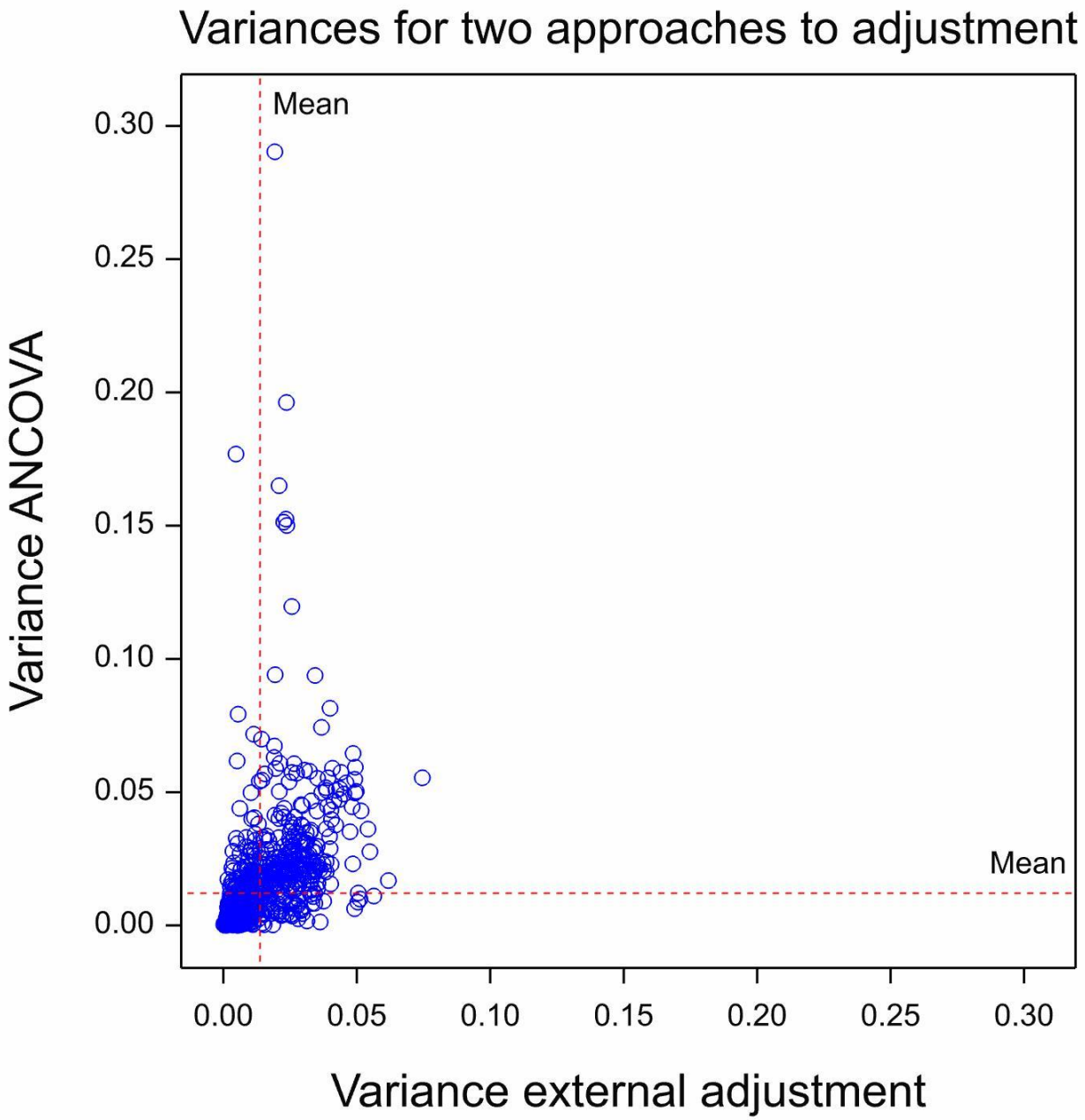
Mean =	0.0121
Median =	0.00716
Minimum =	0.000181
Maximum =	0.0746

NB Variances based on 4 DF

Summary statistics for Variance ANCOVA

Mean =	0.0137
Median =	0.00749
Minimum =	0.0000353
Maximum =	0.290

NB Variances based on 3 DF



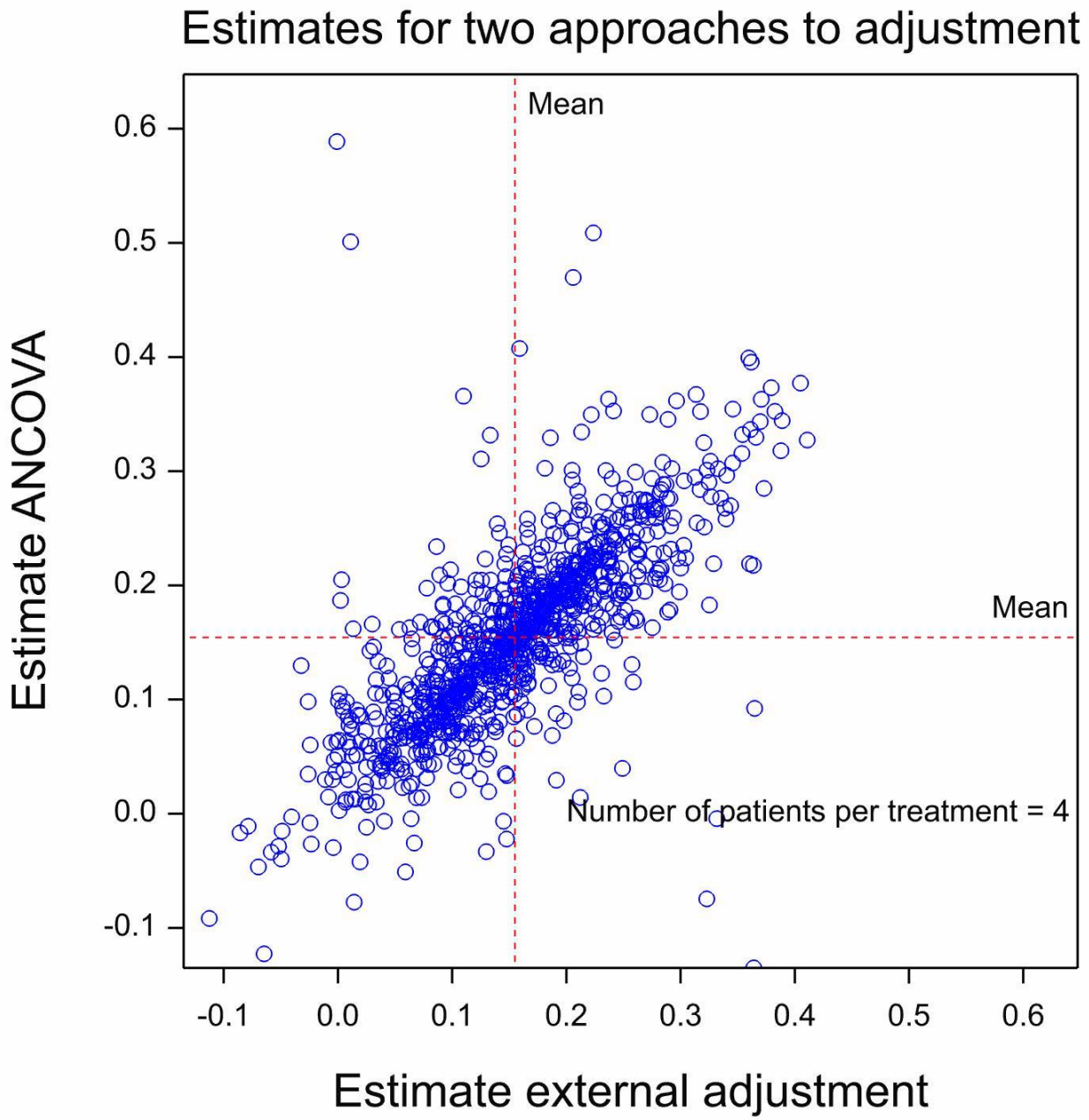
Simulation 1000 samples of size 8

Summary statistics for Estimate external adjustment

Mean =	0.154
Median =	0.154
Variance =	0.00696

Summary statistics for Estimate ANCOVA

Mean =	0.155
Median =	0.155
Variance =	0.00673



Simulation 1000 samples of size 8

Summary statistics for Variance external adjustment

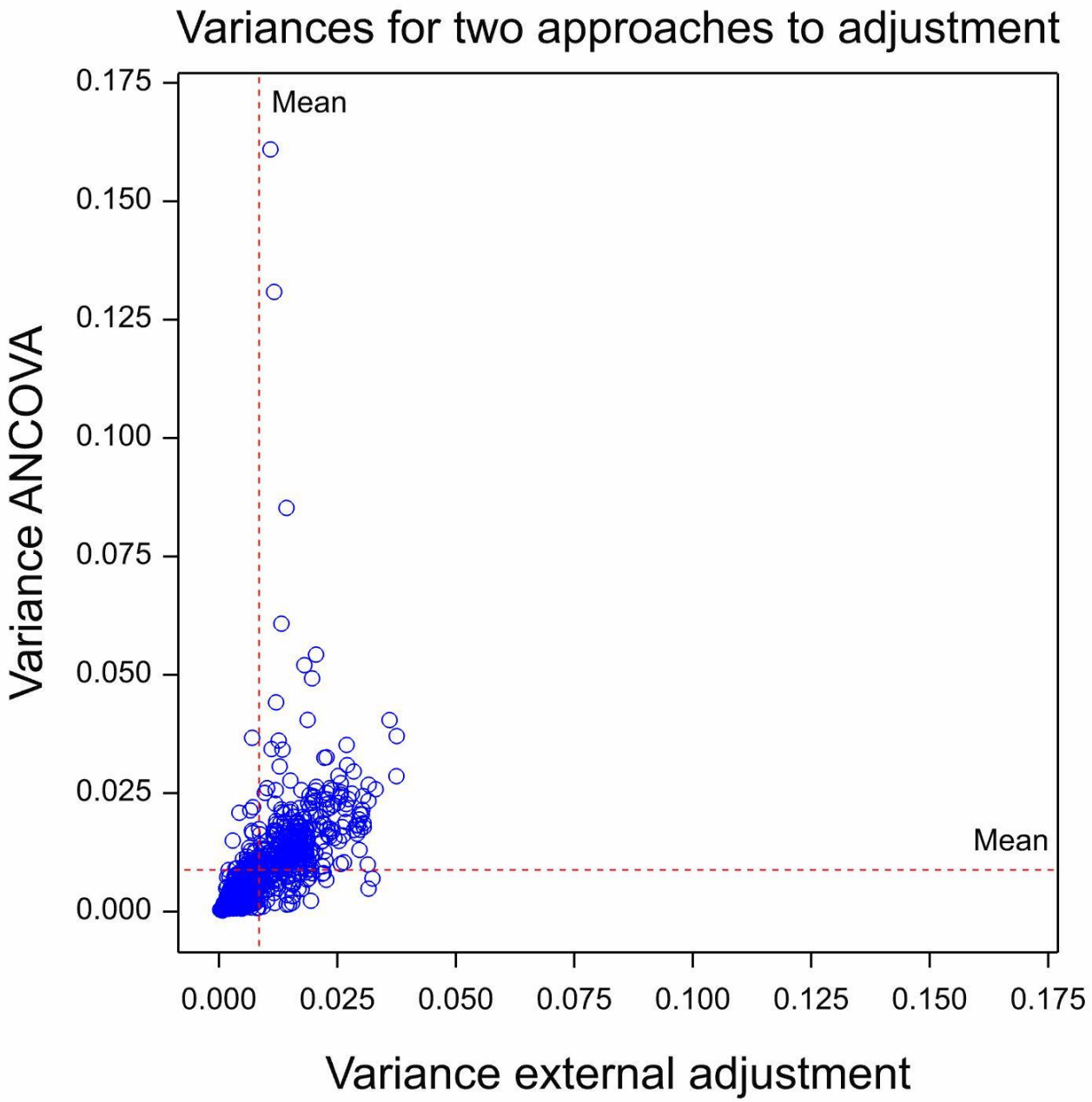
Mean =	0.00878
Median =	0.00596
Minimum =	0.000283
Maximum =	0.0375

NB Variances based on 6 DF

Summary statistics for Variance ANCOVA

Mean =	0.00848
Median =	0.00567
Minimum =	0.000231
Maximum =	0.161

NB Variances based on 5 DF



A Bayesian Perspective?

"A model should be as big as an elephant." Jimmy Savage

A Bayesian view of frequentist models is the following

Any term in a frequentist model has an uninformative prior as to its effect
Any term not in a frequentist model has an informative prior that its effect is zero

The compromise Bayesian approach would be to have partially informative priors

Perhaps a way can be found to use informative priors (ridge regression analogy) that avoids the paradox of information

Conventional Bayesian approach (certainly not)
Lindley & Smith, 1972 (not quite)
Dawid & Fang, 1992 (perhaps)

Work in progress

The world of non-linear models

- The Normal distribution is a two-parameter distribution and this has implications for the linear model
  - Unexplained variation can be swept up in the variance
- The same is not true of certain other distributions/models*
  - Poisson
  - Binomial
  - Proportional hazards
- Variances for certain models always increase if covariates are fitted

Gail et al, *Biometrika* 1984, Robinson & Jewel *Biometrics* 1991, Ford et al *Stats in Med* 1995

However

- Fact that variances increase does not mean power reduces
  - Estimates are biased towards null if predictive covariates omitted
- Furthermore, in predictions space, marginal and conditional models can be more compatible (Lee and Nelder, 2004)
- In other words, the values of adjusting and the potential problems of doing so may be not so dissimilar after all
- This is just speculation on my part
  - Not so much ‘work in progress’ as ‘work that has yet to progress’

Conclusions

- Where diseases are rare, patients are few and large simple trials are not possible
- Further information cannot be obtained by studying more patients but may be obtained by measuring more things
- To the extent that they are predictive covariates can be useful
- But naïve use faces the paradox of conditioning
  - More information appears to be worse than less
- We have to find ways of getting round this to progress
- Nevertheless, statistical modelling is not magic and there are limits