When accounting for confounders is just dodgy bookkeeping

Chris Howden
Statistical Consulting Unit (SCU)
@ Sydney Informatics Hub
Core Research Facilities

THE UNIVERSITY OF SYDNEY

Statistical Lead of the Statistical Consulting Unit @ Sydney Uni

Management talk mission statement

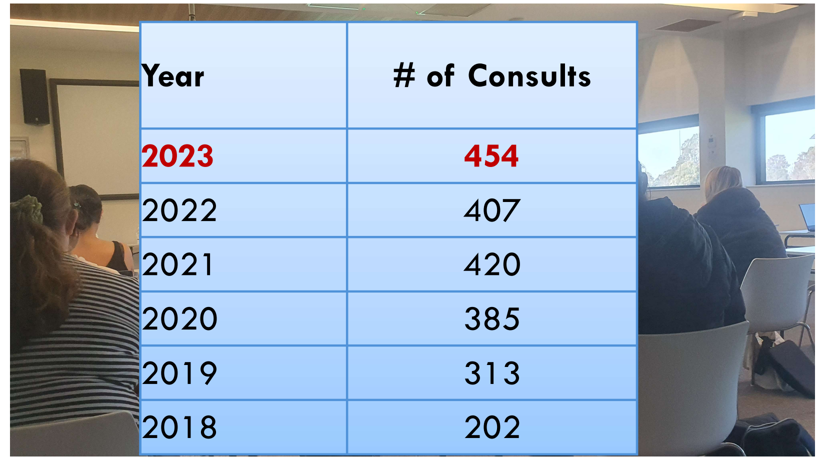
Enable Research Excellence through facilitating higher quality, impactful research and output *by providing statistical expertise to researchers*

In the real world we do that by:

- Teaching researchers to catch their own statistical fish, rather than catching fish for them
- Improving the research culture and practice of the University as a whole
- Researchers' statistical mentors

Consults

Workshops



Year	# of Consults
2023	**454**
2022	407
2021	420
2020	385
2019	313
2018	202

Projects

We often hear researchers say things like

I want to account for confounders/covariates by adding them to the model

I accounted for biological sex by adding it to the model

I removed the effect of biological sex by adding it to the model

I corrected for biological sex by adding it to the model

I controlled for biological sex by adding it to the model

I adjusted for biological sex by adding it to the model

Unfortunately, all they often actually do is just add it as a ***main effect.***

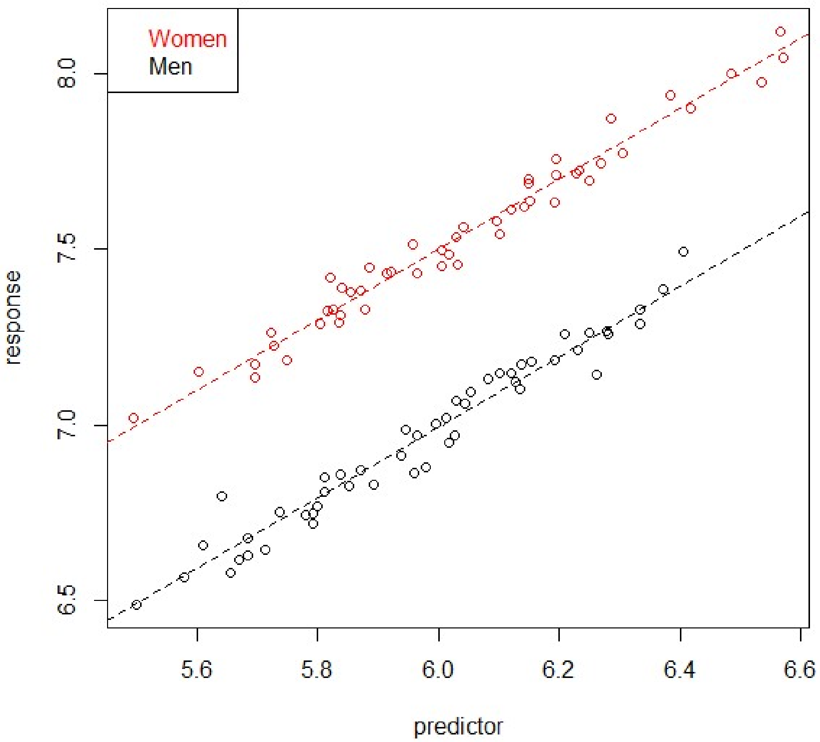
Or in other words adjust some baseline predictor's effect for biological sex e.g. women's effect is 10 points above men's.

It doesn't allow men and women to have different relationships with predictors of interest.

Which is what researchers usually need, and often think they are doing!!

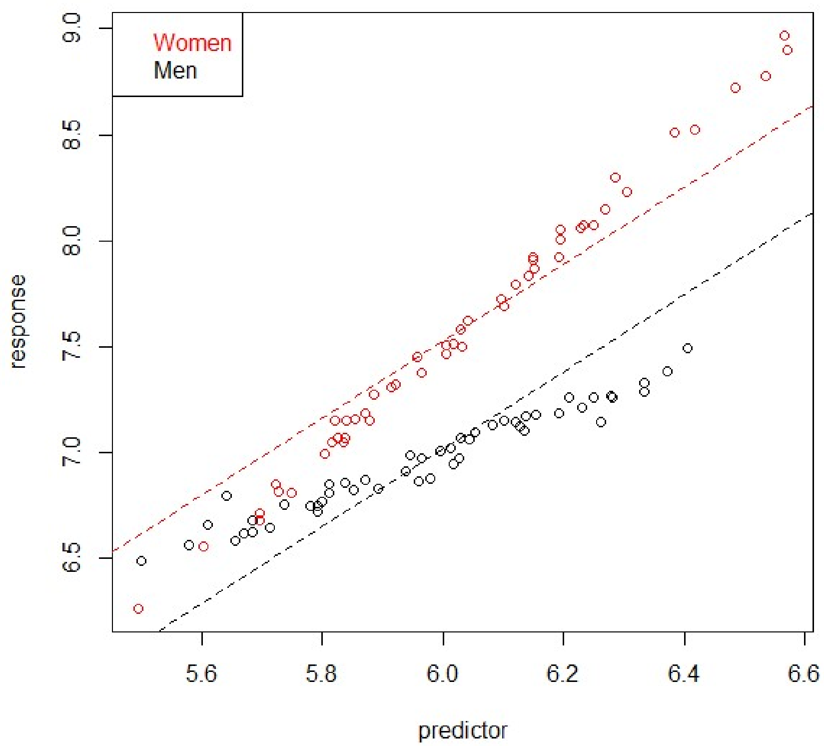
$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 \mathrm{Sex}$$

If there is a *consistent difference* between biological sex then simply fitting a *main effect* works just fine.



$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 \text{Sex}$$

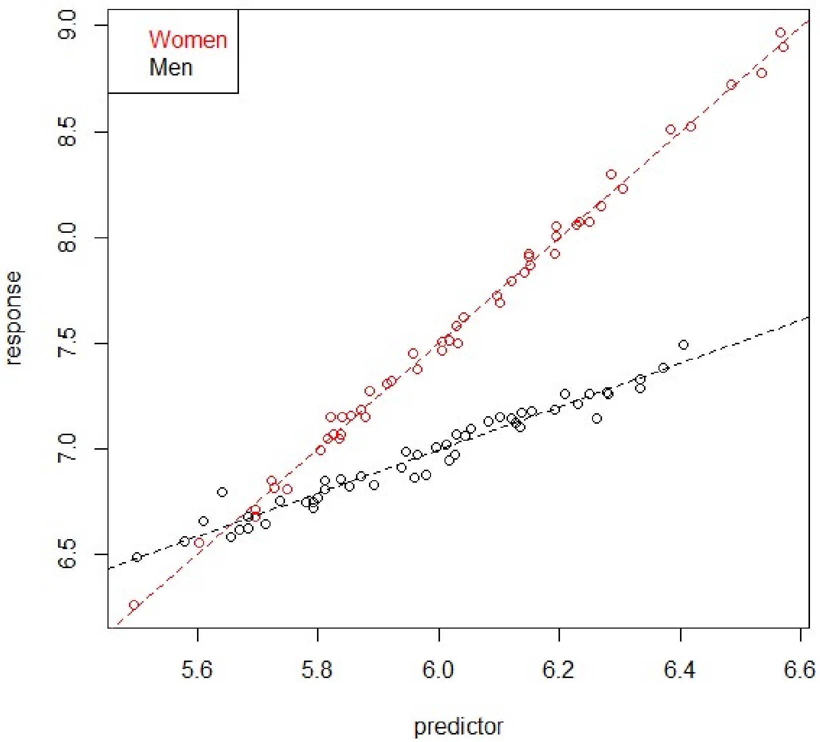
If there is a *consistent difference* between biological sex then simply fitting a *main effect* works just fine.



But not if men and women have a different relationship altogether.

$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 \text{Sex} + \beta_3 X * \text{Sex}$$

If there is a *consistent difference* between biological sex then simply fitting a *main effect* works just fine.



But not if men and women have a different relationship altogether.

That requires an interaction as well.

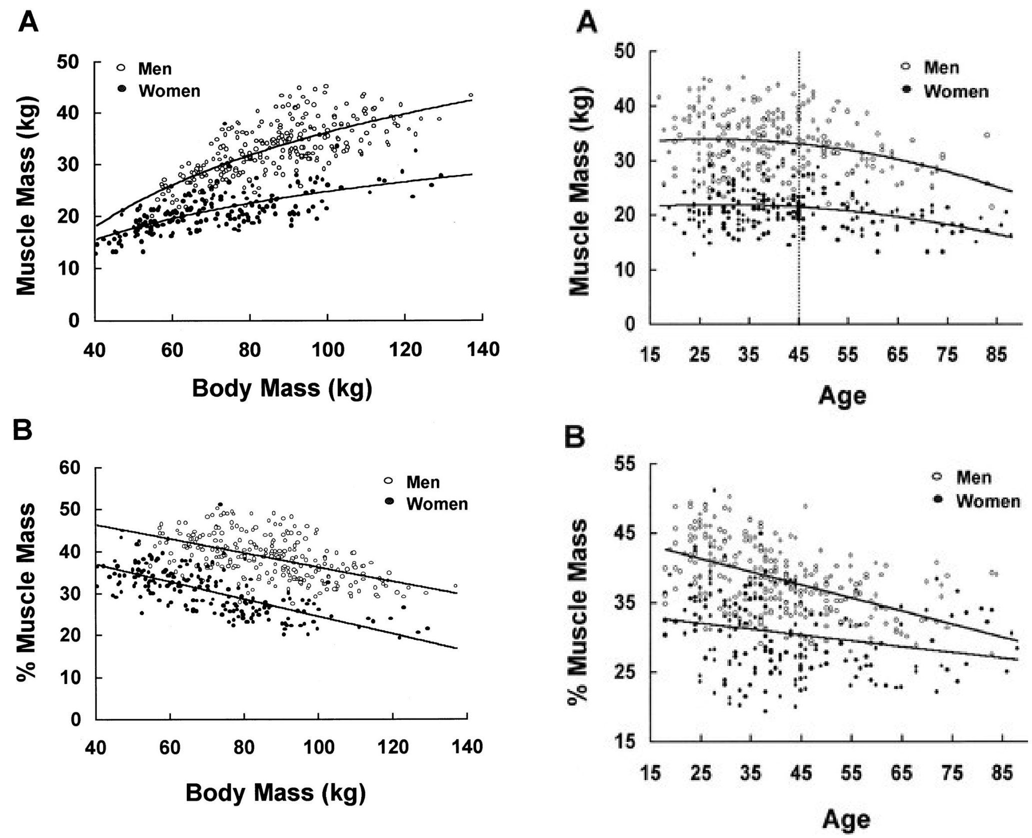
(Or fitting each covariate with a different model, so 2 linear models for this example. But that has drawbacks such as not being able to directly test if their relationship to the predictor is different)

Janssen I., Heymsfield S.B, Wang Z, Ross R. (2000) Skeletal Muscle Mass & Distribution in 468 men and women aged 18-88 years. J. Appl. Physiol.

Skeletal Muscle (SM) is an important factor in many physiological and disease processes *such as the influence of* ***aging on muscle wasting*** *and anabolic effects of training on muscle size.* ***Benchmarks were required for future studies assessing SM status in aging and disease, and to facilitate health policy.***

The 2 plots with roughly parallel lines (lower left and upper right) could possibly be fit with a main effects model.

While the other 2 require an interaction as men and women have strongly different relationships between predictor and response.



Vasan R.S., eta al (2001) Impact of High-Normal Blood Pressure on the Risk of Cardiovascular Disease. NELM.

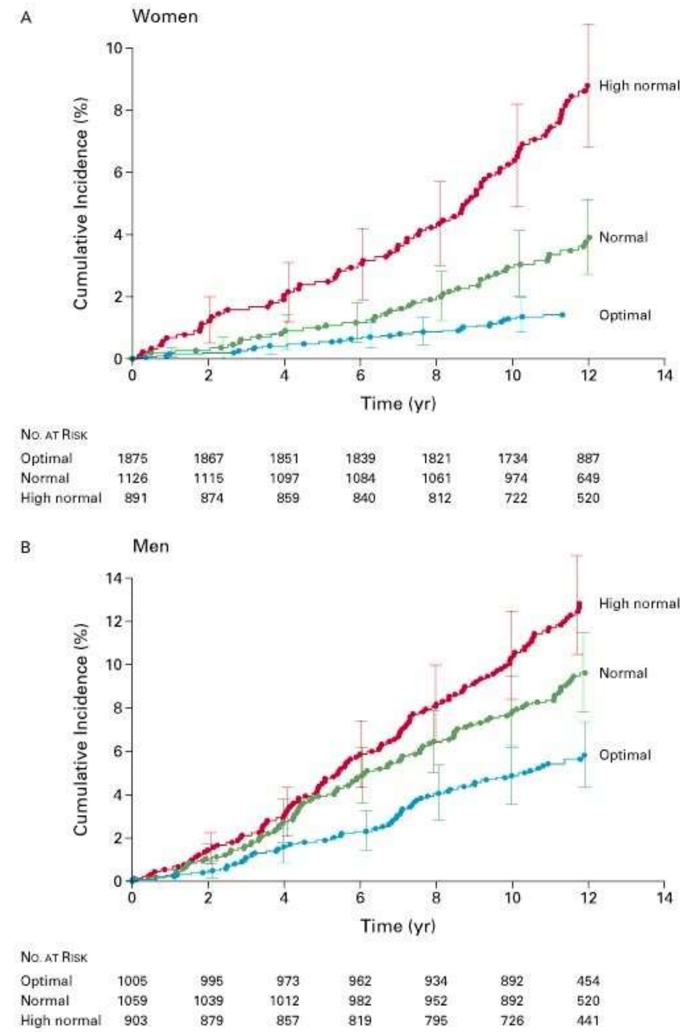
Men could possibly get away with high-normal blood pressure for up to 5 years without their risk of disease being that different to normal (partly due to higher risk of disease in general).

Women with high blood pressure substantially increase their risk immediately.

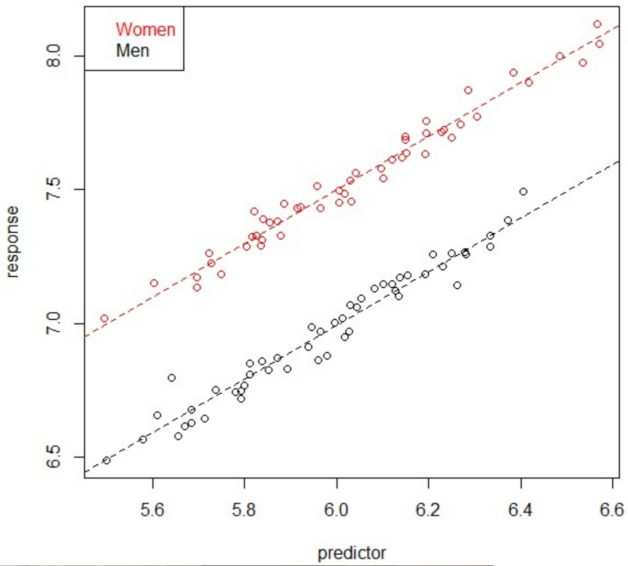
So one might have different polices based on sex:

– women should be supported in checking their blood pressure more often. As there is a greater urgency in lowering their blood pressure they get immediate access to drugs to give them time to make lifestyle changes.

– while men's access could be slightly delayed and they are told to try changing their lifestyle first.



Where's this dodgy bookkeeping come from?



I suspect it comes from the old ANCOVA way of thinking (Analysis of Covariance).

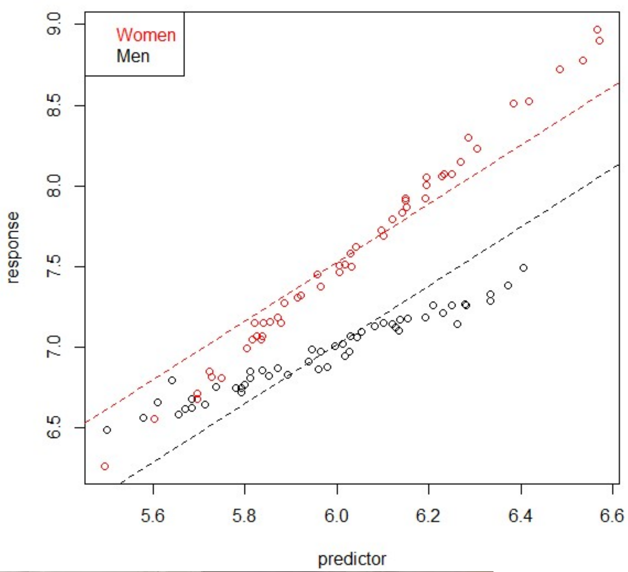
The whole point of ANCOVA was to ***control*** for continuous covariates by showing the consistent relationship between covariate and response can simply be ***adjusted*** up or down based on the predictor of interest (by adding them as main effects).

Which is very useful when we get data like this.

Which meets the ***parallel lines assumption.***

Which ANCOVA workflows ***always*** check for.

Where's this dodgy bookkeeping come from?



BUT somewhere along the way researchers stopped thinking of it as ANCOVA.

They ***forget they needed to check the parallel lines assumption.*** So they started fitting the wrong model.

They also forgot they were controlling for the continuous variable, not the discrete one i.e. biological sex!

ANCOVA morphed from a method that **corrected for a *continuous co-variate if the predictors overall relationship is consistent*** into researchers thinking ***a discrete covariate could always be corrected just be adding it as a main effect!!!***

It's a very real problem!!

1 in 3 consults I'm telling researchers they can't account for covariates by just adding them as main effects. They need to use Exploratory Data Analysis (EDA) and likely also formally test if main effects or interactions are appropriate.

They often look a bit surprised!!

Seems particularly bad in medicine and health.