

[A blog about statistical musings](#)

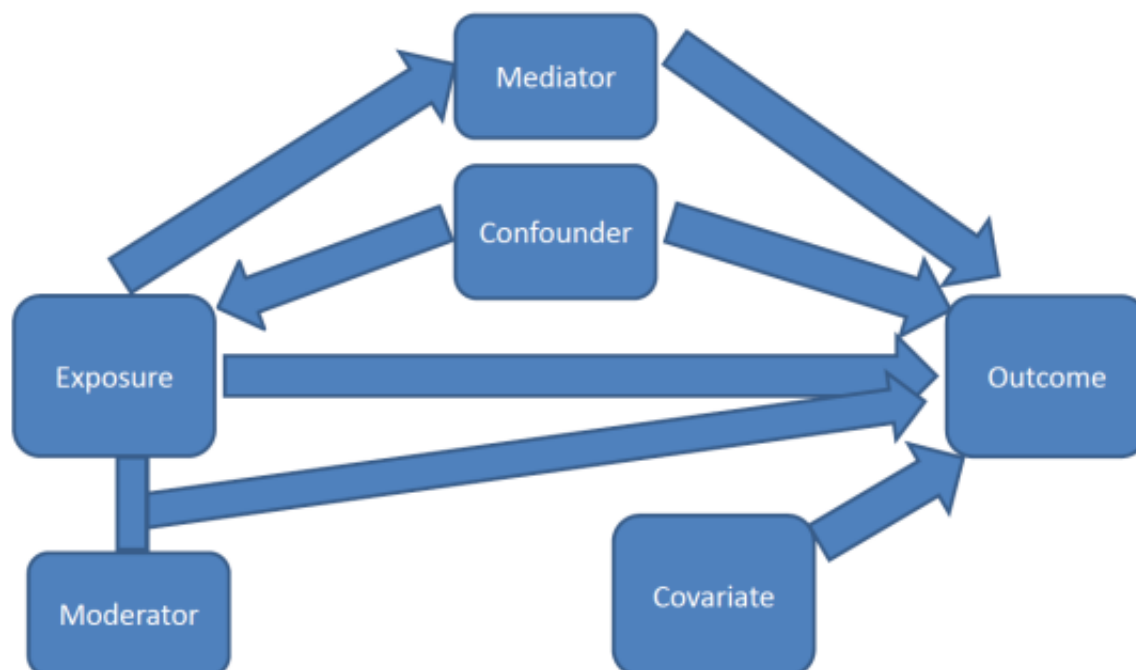
# Confounders, mediators, moderators and covariates

I recently put together some slides to explain mediators and mediation analysis to some people who knew slightly less than I did on the topic. Explaining things you barely understand is the bread and butter of a statistician, so I took to it like a duck to plum sauce.

I started looking for some nice examples that would describe what a mediator was. I found plenty. Of course, it was also important to preempt confusion between similar and related terms, and since mediators and confounders are regularly mixed up I also looked for nice examples of confounders. Again, i found lots of good examples. For completeness I looked for examples of moderators and covariates. What struck me was that there were nice examples for each term separately but i couldn't find a really comprehensive example that included them all and clearly delineated the difference between them. So I put my mind to it. And failed.

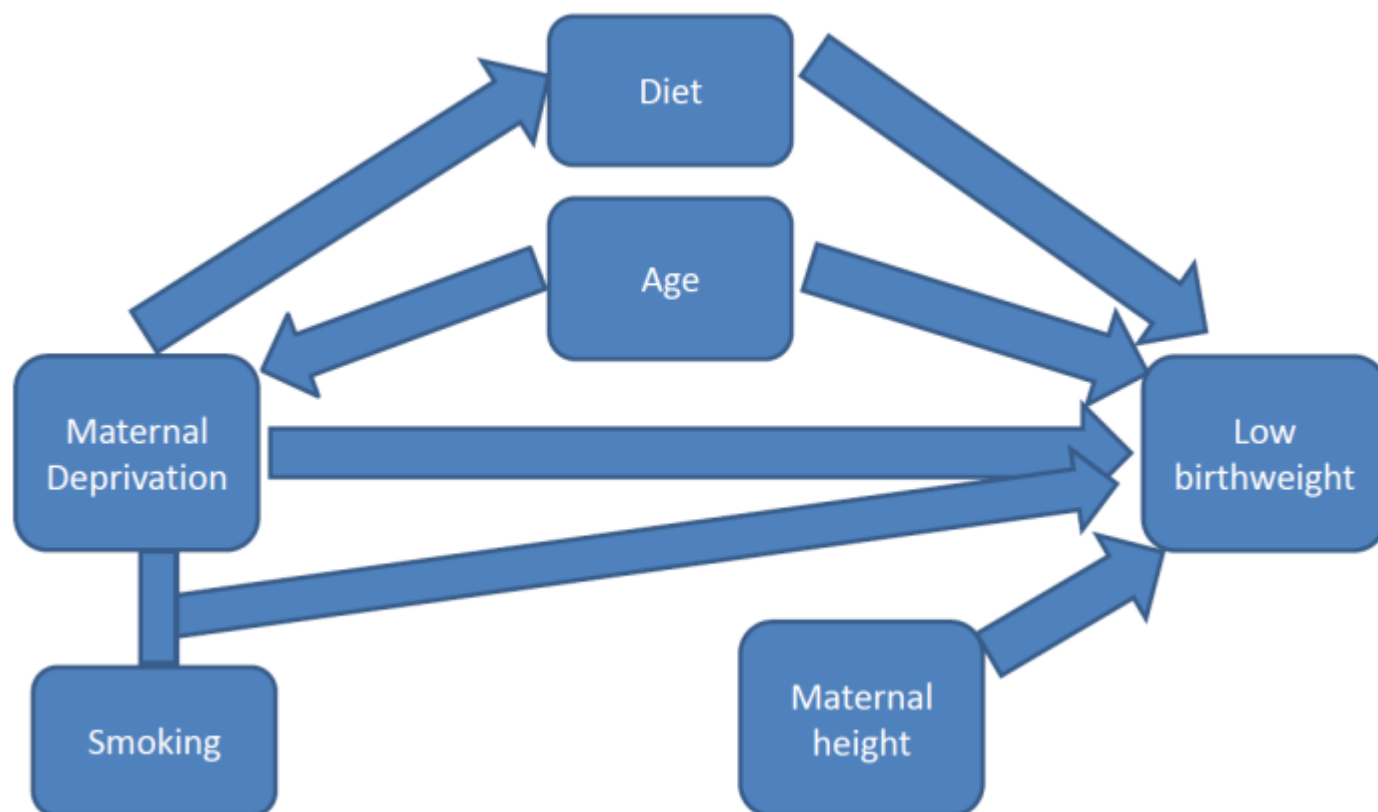
But then I asked someone cleverer than me (one of my former PhD supervisors) and he provided the bones of a nice example to which I have added. I can't claim that the result is as clean and clear as I would like, but it is the best I have and I would welcome corrections and clarifications.

So here is the basic formulation.



<https://significantlystatistical.files.wordpress.com/2014/12/slide-31.png>

We have an exposure (or treatment if you are trials minded) that we think is associated with an outcome of some type. Mediators and confounders are similar except for the direction of effect between them and the exposure/treatment. Mediators are additionally characterised by lying on the causal pathway between exposure and outcome. Moderators are simply interaction terms that change the size or direction (or both) of the effect of the exposure on outcome. I have represented them here using a vertical line between exposure and moderator that feeds into an arrow leading to outcome (which isn't conventional but represents the relationship better). Meanwhile, covariates are variables that might affect outcomes but are not associated with anything else. So far, so confusingly boring. But hopefully grounding this stuff in a concrete example will help. Here's the one I have.



Okay, so maternal deprivation is associated with mothers giving birth to babies with lower birthweight. This is well established. So well established I can't find the reference. Let's move on.

Next the mediator. The causal pathway through which deprivation might act could be through diet, i.e. being poor might mean mothers can't afford good food and their diet consequently suffers. This sub-optimal diet is the same diet nourishing the unborn child who is smaller as a result. Mediation analysis can formally test whether this hypothesis is true.

The confusion between mediators and confounders arises from the fact that both have associations between the exposure and outcome. Now the confounder I have chosen is age. According to **figure 2 here** (<http://bmjopen.bmj.com/content/4/4/e004514.full.pdf+html>), there is an association between maternal age and deprivation. Logically, we do not allow deprivation to influence someone's age, so the arrow only goes from age to deprivation. I like that about this example. We are also assuming that maternal age is associated with low birthweight **which it seems to be** (<http://www.ncbi.nlm.nih.gov/pubmed/2892726>).

Smoking is what I have chosen for my moderator. Now **this study** (<http://www.ncbi.nlm.nih.gov/pubmed/22404858>) investigated black smoke (so air pollution) and did indeed find a significant interaction (so

moderation) between deprivation and black smoke exposure on birthweight. This means that deprived mothers who smoke produce babies even smaller than the separate expected effects of smoking and deprivation. I'm gonna go ahead and call that a win.

Finally, there may be additional covariates that we want to control for such as maternal height. This could just as easily be called genetic factors or something else, but the point being that maybe there are things we need to account for as we think they are related to outcome (but nothing else).

Of course while ~~cherry picking~~ systematically reviewing the literature I found this



A simplified pathway diagram

(<http://www.ncbi.nlm.nih.gov/pubmed/17108301>) which observes that there is an interaction between maternal age and deprivation. This would indicate that age may be a moderator as well as a confounder. And smoking and deprivation are also linked. And this really is where it becomes clear that in most situations the pathway diagram is complicated and uncertain. Some of the relationships will be stronger than others, there will be associations between practically all of the variables and there will be feedback loops all over the place.

Moreover, I have come to the conclusion that mediators and confounders are kind of intuitive. People understand them when given an example. Clear examples of interactions/moderation are rarer and I suppose that is because interactions are counter-intuitive. They are effects that are different from the sum of their parts and real-life examples that everyone recognises are not commonplace (prove me wrong commenters – I'd love some better examples).

Nevertheless, and I'd like to finish on an ill-deserved upbeat tone, the discipline of clearly describing what we think the relationships and pathways are is incredibly powerful and important to hone. In particular pre-specifying the order we expect things will happen allows us to identify and test causal hypotheses. In most situations, the process of committing this to paper is the hardest part. It is only once that step has been made that statistics can really begin to be applied (of course statistical thinking can help you get there too). To paraphrase Indiana Jones' dad\* once the pathway is clear the appropriate statistical approach often presents itself. (I would like to thank Zoe Kelson and Daniel Farewell for constructive comments on earlier drafts)



<https://significantlystatistical.files.wordpress.com/2014/12/solution-presents-itself.jpg>

\* I couldn't get a picture of Indiana Jones' dad with this text. Here's one of Han Solo instead.

[About these ads \(https://wordpress.com/about-these-ads/\)](https://wordpress.com/about-these-ads/)



**DECEMBER 12, 2014****DECEMBER 10, 2015** **MARKJAMESKELSON**

## 3 thoughts on “Confounders, mediators, moderators and covariates”

**EpiAnn says:**

**JULY 7, 2015 AT 3:52 PM**

1. Found this through google while looking for mediation moderation vs. confounding interaction–suspect I’ll be a regular visitor!

**REPLY**

**markjameskelson says:**

**JULY 7, 2015 AT 4:51 PM**

1. Thanks Ann. Must blog more regularly!

**REPLY**

**hopeLESS says:**

**OCTOBER 5, 2015 AT 9:02 AM**

2. wow, this is exactly what I am looking for. Thank you. I have no suspicion to become a regular...

**REPLY**

**Create a free website or blog at WordPress.com. The Gazette Theme.**

%d bloggers like this: