Slide 1:

In this presentation I'm going to teach you a little bit about statistics and the methodological questions that you might have when you're trying to combine all these data sets and trying to do these big international collaborative research projects.

Slide 2:

We're going to answer 3 big questions. Number 1, what is different about individual data meta analysis and how it differs from traditional aggregate meta analysis. We're then going to unpack that a little bit. We're going to look at IPD analyses and look at the two main types, one stage or a two stage, and I'm going to say why we would want to do a one stage, what are the advantages and disadvantages in comparison to the two stage.

Slide 3:

So just to quickly explain what's different about an IPD analysis. In a traditional meta analysis what we usually do is we search a database and we find all the papers that are relevant for our question that we're interested in. We take all of their results and we combine them and summarise them to make one answer. In comparison for an IPD meta analysis, what we're doing is we're instead getting the individual level data so it's nothing to do with a published paper but what we're instead doing is getting cohorts to send us their data and we're going to combine that in specific way.

Slide 4:

We might wonder what is the point of going to all this extra effort with an IPD analysis. What is the main advantages that we get out of it in comparison to traditional aggregate meta analysis? one thing that might be useful to think about is imagine we have a hypothesis that being born preterm has an effect on your IQ. And what if we think that being male means you are especially susceptible. We might then test this in a multiple regression with an interaction effect. If we were looking at aggregate studies, what we do is something called a meta-regression. We find papers and find what percentage of the participants from each cohort were male and we might then expect that cohorts which have more male participants are going to show lower IQ performance on average. However, this is susceptible to aggregate bias and is a statistical weakness of meta regression. Instead what we might want to do is we might want to take that individual level data from each cohort. We can then perform something like multiple regression and we can test within each cohort does being born male have a significant effect on IQ. Does it interact with being born very preterm to cause interactive moderating effects? This figure that's shown on that right hand side is essentially just showing that a meta regression is really insensitive to finding these interaction effects in comparison to an IPD analysis. There are other advantages of IPD meta analysis in comparison to aggregate ones. One thing you might want to think about is when you're searching papers, you usually have a criteria where you say I only want very preterms or very low birth weight and then you might get to a sticky situation where you define very preterm as less than 32 weeks but a cohort has less than 33 weeks as their criteria so you'd have to usually cut them out because they don't quite fulfil your criteria. But if you were able to get the individual level data from them you could just remove those born at 32 weeks and make sure it's in line with your criteria. You've therefore managed to keep more data that you would have otherwise lost, increasing statistical power and making the analysis more universal. Another advantage is that you can account for and look at missing data. So for example you could do something like multiple imputation so if you had missing data from a cohort you could impute that and that would give your meta analysis greater strength and be able to look at things that a traditional meta analysis wouldn't look at. You'll also be able to harmonise your covariates.

One really nice example of this is when we think about maternal education. In the UK we would have levels based on exams at 16, 18 and then university. Other countries do not have this structure so we cannot directly compare. We can instead harmonise the individual level data to an international standard such as ISCED so all participants now have low, medium or high scores. This make sure you compare like with like. Finally when you actually then go to do your analysis you can do it all in the same way. You're not going to have to take an odds ratio that's published in one paper and compare it to a risk ratio in another cohort/another published paper. So it's much more flexible. You're going to make sure that your comparison is as equivalent as it can be between each cohort.

Slide 5:

So we usually do either a one stage or two stage IPD meta analysis. I want you to look at the figure and you can see BLS, AYLS, VICS, POPS and these are some of the cohorts that we've used here in RECAP. On the left in red is the one stage approach. So what we're going to do is we're going to harmonise the data like Sylvia's shown in the presentation prior and then we're going to bring them together into one big data set, rather than keeping them separate. Then we're going to do a statistical analysis using mixed effects regression. This is sometimes called a hierarchical mixed model or a linear mixed model and what we do is we treat our cohort variable as if it's a random variable. So we're going to be interested with whether very preterms are different to term born controls but we also have to understand our data has a hierarchy. We have to understand some of the participants are coming from a specific cohort and this might have an influence on some of our outcomes. so we do our mixed effects regression and that is essentially our analysis done and we have our meta analysis results. In comparison on the right in green is the two stage approach. This is a little bit different and what we're doing is we're keeping the data a little bit more separate. So what we do in each cohort is we perform a regression. This could be a multiple regression, this might be a logistic regression and we might test whether being born very preterm are different to being term born. We might test other things such as whether being male affects the outcome in comparison to being female. We perform all these analyses separately in each cohort and what we do is we take our summary output, we take our results from the multiple regression (the beta values) and we combine them across cohorts. Then we take that and we get one final estimate which we can combine using either a traditional fixed effects or random effects meta analysis and once we've done that we have our meta analysis results. So again they are slightly different but essentially it's the way you keep the data together or separated.

Slide 6

The differences between one stage and two stage are quite minimal. However there are some potential differences both from a theoretical output kind of approach or from actually a practical doing the analysis approach that I think you might want to be aware of. So a one stage is really flexible and it's so easy to do model testing because if you want to add in a new factor, you just add that into that mixed effects model and it's simple. It simply does the analysis in one go. Where if you want to do it in a two stage you're going to have to write essentially new syntax for every cohort every time you want to change a model or test something new. Then you get into difficult situations where a factor might be significant in one model for one cohort but it's not significant for another cohort and based off model practising techniques should you keep that variable? However the cons of the one stage is that it is a more complex analysis and not everyone is familiar with a linear effects/ hierarchical mixed model and so people might not understand it as much when you're presenting the work. Also it doesn't really produce a forest plot in the same way that you might be used to from when you're reading an aggregate meta analysis. So the advantages of the two stage is that is more simple to understand and usually the results will be similar as a one stage. However,

when you're trying to do something more advanced, we found model testing and model building for the one stage is better.

Slide 7:

To conclude IPD meta analyses are absolutely fantastic and clearly superior to aggregate meta analyses for a whole number of reasons. These can include greater statistical power harmonising data and having consistent inclusion criteria and when we do IPD analyses we do them using one stage or a two stage approach. One stage is more flexible but deviates from traditional meta-analysis whereas a two stage approach can definitely be more time consuming, with model testing especially, but it is easier to understand and easier for a reader especially.