Hi everyone and welcome to the second lecture in this lecture series. This is a direct follow-on from the topics I introduced in the first think before you do video. And today I'll be working through a, an example of a pre-registration and taking you step-by-step through each each element of the research checklist that we, that we discussed and unexplored in the first lecture. Again, if you have any questions about the topics covered in this lecture series, my email is here. And we're reproducibility can also be found on, on Twitter. Okay, so the aim of today's lecture is to apply our research checklist to a work example of a pre-registration. But a lot of the content here is just a helpful way to start off with the project and really can help clarify your thinking on the different elements are ingredients of any research study just to provide some context. So today's example is going to be based on one of my PhD projects. So this was a project looking at pubertal development and depression in adolescence. So it's just helpful to have a real-world example of how you might apply this research checklist to a real-life project. But again, the, while I'll discuss in today's lecture is just one example of a project, but there's so many resources online that can help you apply these reproducible tools to your own research. And I just included some links here that are available on the OSF and these will be linked at the bottom of the video as well. But yeah, as I said, there's so many choices and there's a wealth of information and resources available online that I really encourage you to explore. Okay, So again, this is the research checklist that I introduced in the first video of this series. And we'll just kind of break it down step-by-step and give you an example of the information that you might want to include it under each heading to try and make your research as reproducible as possible. Okay, So starting with the first two, I suppose in many ways the most important aspects of this project is your general research question. And then from this specific, testable and concise hypotheses that you want to test in your study. So this information will follow on in your pre-registration. I'm often in the especially in the templates that are available online, you'll include a project title and then also provide a brief description or background to the study. So in many ways this is, can be like a shortened introduction that you'd find in a paper. And what I find quite helpful actually is when you're starting to rise preregistration, you can write this introduction and background and then later use that in your paper or manuscript or whatever projects you are working towards. So even though, as I said in the first video, it might seem that you're investing a lot of time and work, but this work needs to be done anyway. And in many ways you're just changing the distribution of the work. Okay, So e.g. your research question, this can be quite broad about the topic that you're interested in, the relationship or association that you might want to explore. So in the case of the example I'll use today, the overarching research question was, is pubertal timing associated with depressive symptoms during adolescence? So again, this is quite broad and we don't yet know how the researchers are going to define pubertal timing or what they mean by depressant symptoms. But hopefully as we move through the research checklist, it will become clear. You really need to give your reader and anyone looking at your study specific information on how you're defining each of these variables. But we'll come to that in a few slides time. So again, the hypotheses are directly informed by your research question. And I hope you can see here the difference between the research question and your hypotheses. These hypotheses are specific. So the first hypothesis is that there's no association between pubertal timing and depression in adolescents. Then the second hypothesis, or it, sorry, the first one was the null hypothesis. And then hypothesis one is that early pubertal timing is associated with increased depressive symptoms in adolescence. So note here that there's only two concepts here per research question or hypothesis. So this makes it easier to describe any statistical tests later. And your analysis, and again, the hypotheses of your study will change depending on your research

question. But again, just try and imagine in your head, like your, your hypothesis needs to be clear and concise and specific. And if you're hypothesising that there's a direction. So e.g. here, that earlier pubertal timing is associated with increased depressive symptoms. Make sure that that's very clear in your in your pre-registration. Okay, so once you've decided upon your general research question and you've derived your specific hypotheses. The next steps then will pertain to the study design and then in turn the sample size and your variables of interest. But again, the study design will vary depending on the nature of your project. But the key here is to be as detailed as possible. Again, in the case of my example here I was working with pre-existing data, but if you're collecting your own data, then this might be different. There will likely be different. So again, just to give you an example of the detail that you might want to include. And I'm telling the reader where I'm getting my study data from. So it's coming from the adolescent brain cognitive development study, which is a US based longitudinal cohort. I'm telling the reader how large the sample is, what the age range of the sample is, and that's collected across the US. I'm also giving the reader information on what data points are going to be using. A broad introduction to the variables of interest. And you know, this, this, this I suppose will change depending on, as I said, whether or not it's a data collection based project or secondary data analysis, like was the case in my study. But, you know, some general, the general concepts are relevant to every study. So again, the type of Dacia, the content of the dataset, and also giving some more information about the nature of the study. Especially if you're collecting your own data and deriving your own study. Is it a two-group design? Is it a repeated measures design? Are you going to be looking within subjects or between groups? Or if it's a qualitative study, is this a case study or the like? So again, I'm including the picture of a cake here just to remind you that you really want to try and give as much detail as possible so that someone could come along and reproduce your recipe. Okay, so moving on to our variables of interest because our sample size first. So again, this will vary depending on the nature of your study. But again, you want to just kind of, I suppose, take what the information that you've introduced in the study design and then just kinda go into a bit more detail. So again, are you going to use the entire sample? Are you only using a subset and what specific time points you're using? So e.g. if you are recruiting within the community, are you going locally if you're designing your own study? Are you going to recruit to schools? You're going to recruit through clinics if your project has a more clinical focus. So these are all the details that you might want to think about, including in this section. Another thing here is the sample size and a power analysis for your study. So this will vary depending on the nature of your project. But if you're collecting your own data, are you going to have a stopping rule of what's going to determine when the sample size your study, when you're going to stop collecting and also inclusion and exclusion criteria. Are there certain characteristics that you want your study sample to have and how are you going to sample that? So if you're going to recruit via e-mail or via text message, just, again, try and give as much detail as possible here. Okay, So the next three points are going to be grouped together because this relates to the variables of interests and how you're going to define the relationships that you want to study in this, in your project or your paper. And then of course your covariates and how are you going to treat missing data are skewed data or outliers, and these are very important information to include so that another Easter coming along would be able to reproduce the data that you collected and analysed. So I suppose I want to pay particular attention to this aspect of a preregistration because often if you go and read a paper in the literature, there really is so much information lacking in how variables are defined. And sometimes this is due to word limits in a paper or like specific journals and where it lives that they might have. But often the supplementary materials are

an ideal place to include this extra information, especially if you process your data in a specific way. And you really want to try and give as much detail here so that someone else could come along and they understand how you went there. They would have enough information to understand how you went from the raw data that you collected, or you've got access to the kind of finalised variable that you input it into your models. Okay, So e.g. in the case of my study, my independent variable was pubertal timing. So here I'm giving the reader information about this scale or questionnaire. I used the pubertal development scale. What this examines, how many items are in this questionnaire. How my scoring us. Then I wanted to derive a pubertal timing scores. So there are various different ways of doing this, but I'm giving the reader clear information about how I defined us. And this is just an example here, but if this was in your paper or your pre-registration, you would include references here. I'm kinda refer to the literature if there are certain ways of processing data. And importantly, the code for the data cleaning and preparation would ideally be made available here so that not only do you describe how you went from the raw data to the finished project. But you also provide the reader with the tools to go and look at that code and reproduce those. Then my dependent variable is depressive symptoms. So again, similar to how I define my pubertal timing measure, I'm giving the reader information about the measure that I used. How many items were in this, and in this case, we actually created a new kind of scores, the depressive symptoms in a way. And I've given some information here about that actually in the file and pre-registration for the study, I included a lot more information in the supplementary materials about how I scored this variable. But again, some points snows. Detail really is clear here, is key here. And again, as I said, you want to clear journey from the unprocessed to the process data and including a link to a GitHub repository or whatever platform you use to share your code will be super here. Okay, So variables continue, continued. So often in papers, it's most of the time the independent variables and dependent variables are explained in quite a lot of detail. And then there's a little line kind of thrown in our tagged onto the, to the end of a paragraph saying our covariates where age and sex and ethnicity. But often, there seems to be these ways of approaching research where covariates or just use, because they've always been used. But I would really encourage you to pay as much attention to your covariates as independent and dependent variables because in many ways, in, in your models, your covariates are as important as your, as your independent variable. And it's just how you, how you interpret the findings that differentiates your independent variable and your covariates. So again, do you expect your covariates to have an impact on the relationship you're studying and providing rationale for each covariate is really important here, and this can be informed by existing literature. It's often the case that we have missing data in our data sets. And in my case, I use complete case analysis in this project. And then I actually ended up doing a sensitivity analysis to compare kinda complete case analysis where you just exclude anyone with missing data to a method where I imputed any missing data because people with missing data could be different from the participants with complete data. So again, these are all considerations that will be important to have or at least consider in your pre-registration. Same goes for skewed data. If your data violates assumptions and normality distributions, are you going to transform your data in any way then for outliers? So in my study, I excluded data points that were plus or minus three standard deviations from the mean to determine whether or not these extreme values were affecting our results. So I included this as a sensitivity analysis. But again, it's just important to consider all of these different attributes of your dataset and how they might impact your final results and the interpretation of your results. And the pre-registration is kinda provide an ideal platform for you to consider these different aspects of your study. Okay, Moving towards

the end of our research checklist now is what statistical tests are you going to use to test your hypotheses? And then importantly, what criteria will you use to make inferences? So, how are you going to determine whether or not you have supported? Are you how your results of your study provides support for your hypothesis and what would be your criteria of interests. So this could be p-values, effect sizes, confidence intervals. And this is perhaps the most important, if not the most complicated question within a pre-registration. And I would really encourage you to take time to outline this analysis plan and really think quite deeply about how you're going to answer these research questions. And this may change over the course of the project. And especially if you haven't looked at your data, It's difficult to foresee your model parameter issues. But the pre-registration, it's not this kind of binding document. You there is, of course, scope to diverged from the pre-registered analyses. The most important thing, as I said, is that you're just transparent about this. And the OSF now has kind of updates that you can actually use. A tract changed. Kinda software to show how you've diverged from the pre-registered plan. Okay, so in the case of my study, I was looking at the association between pubertal timing and depressive symptoms. And to test this association, I use generalised linear models. And this is because my outcome was kinda had a, had a Poisson distribution and it was counted data. So there is a kind of zero inflation distribution. So this is why I used generalising your models. I've include, included information here about the function I'm using an AR. Obviously there's lots of different ways of testing these associations. But again, the most important thing here is that you're just providing as much information as possible. I've also given some information on the p-values and whether or not I'm going to correct for multiple comparisons. If I am going to do this, what method that I'm using. And then importantly the inference criteria. So what, what kind of evidence will I use to support or to examine whether or not my hypotheses have been supported. In my case, I was looking at p-values and effect sizes. And so the regression, the Beta coefficient. And this will end. You can also use things like odds ratios. This will just depend on the nature of your projects, but these are things that you can find out as you go through your project and as you think about your data analysis plan. But the main point here is just to include as much information as possible. And also this makes it much easier than, because when you run all your models, you get all of this data and it can be a bit overwhelming sometimes to try and figure out how you're going to assess whether or not your hypotheses are supported. So again, just some points to note here. And again, while you will pre-register your confirmatory or your primary analysis, that doesn't mean that you aren't just limited to these analyses. You can go and do many other exploratory analyses or post-hoc analyses depending on the initial results from your confirmatory analyses. But the most important thing here is that this is just report it as an exploratory test on your transparent about it being kind of a follow-on analysis rather than the starting point which often can happen in research. And as Laura mentioned in the introduction, this is known as harking. Hypothesising after the results are known, which is something that we want to avoid in research. Okay, So just coming to the end of our video now that the main takeaway point that I'd like you to take from this video is the question, does my recipe or does my preregistration have enough information for someone else to make the same study as me. So these are our contact details. Please feel free to reach out if you have any questions about the material covered in this video and any of the videos in this lecture series. But also, I would encourage you to look at the resources available online on the OSF. And in the next video, I'm going to take be taking a deeper dive into reproduce book publishing method known as registered reports. And this content will be quite similar to what's covered in this topic. But I suppose answering more specific are covering more specific information relevant for a registered report. So I hope

that this video was helpful and best of luck in your own reproducible research journey. Thanks very much.