## Simulated Outbreak Exercise

|  |  |
| --- | --- |
| Scenario Date: | Thursday 12 October (11:45-14:15, including lunch break) |
| Inject No: | 15 |
| Inject time: | 1 hour 30 minutes |
| Inject Title: | Apply stratified analysis to identify confounding and effect modification |

1. **Learning outcomes:**At the end of the session, participants will be able to:

- Consider the effect of confounding and effect modification on the association between exposure and disease,

- Perform stratified analysis using the Mantel-Haenszel approach

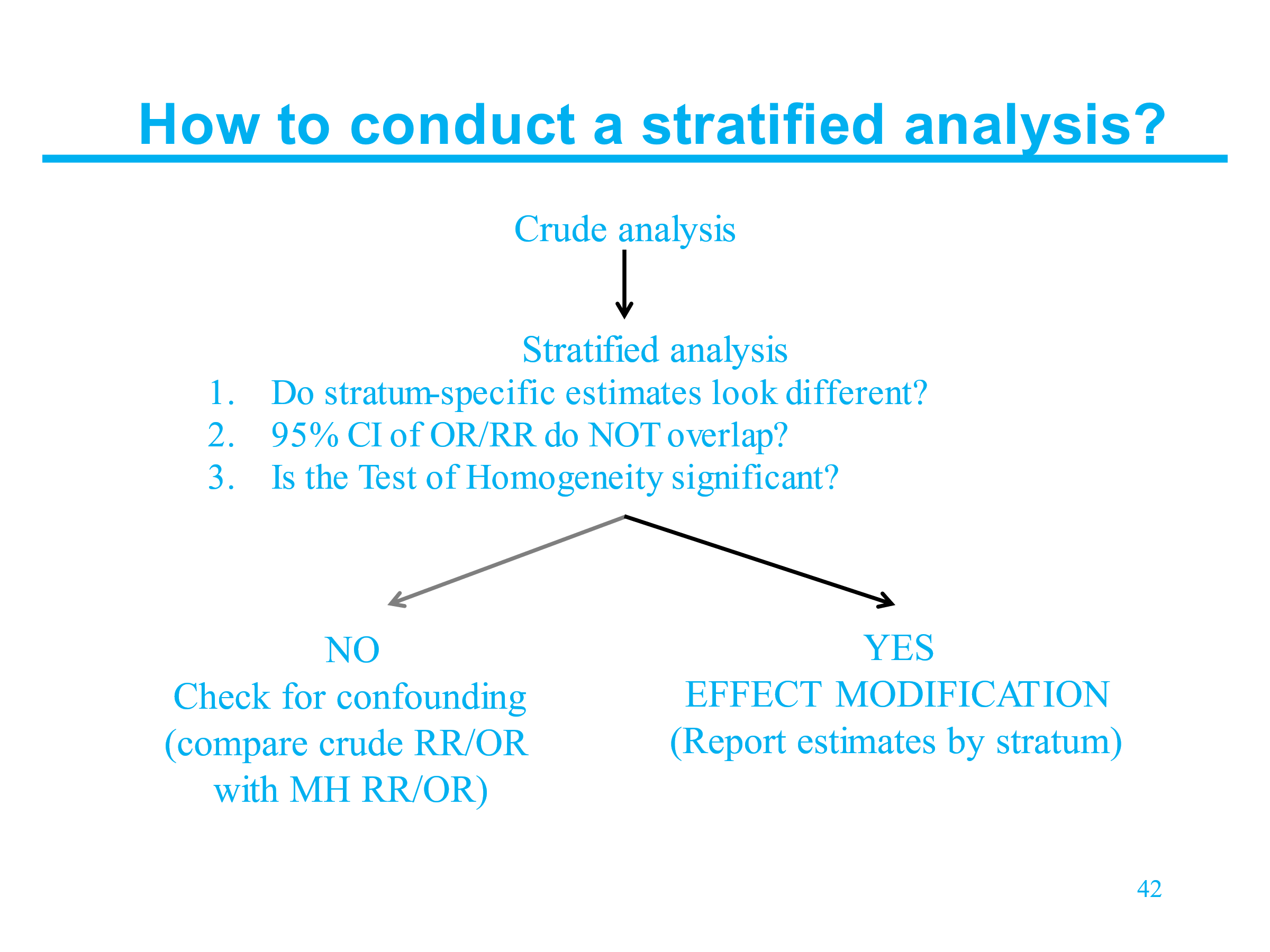
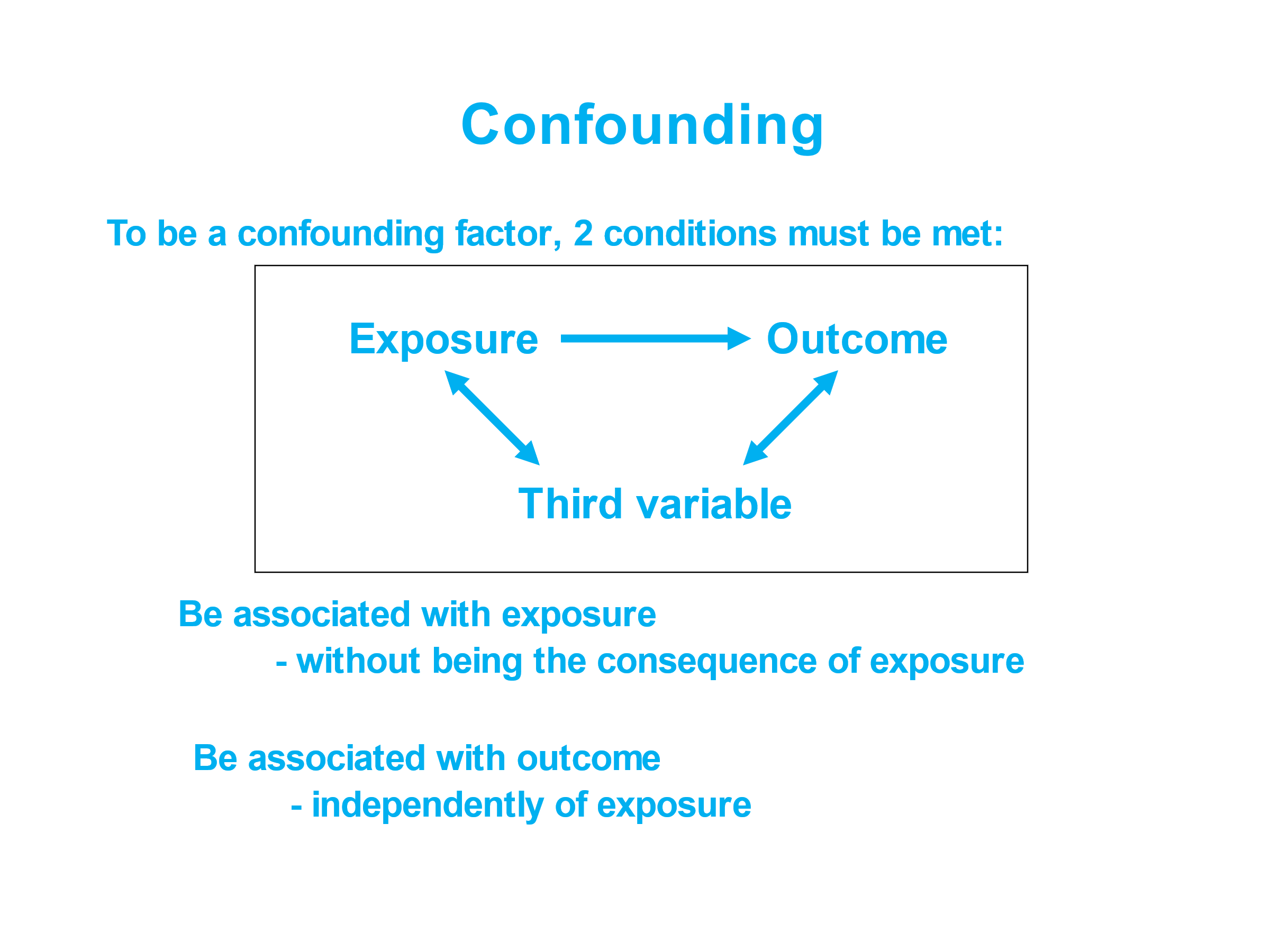
*The use of stratified analysis, is the first step to identify confounding factors and effect modifiers one by one (after, of course, thinking which variables could potentially be confounders or effect modifiers). As the final step, you will be using Regression Models to account for confounding and check for effect modification. We will see these with you in the Multivariable Module (MVA), next year.*

1. **Story/plot description:**From the univariable analysis, it seems that eating pasta and eating veal as well as drinking champagne are associated with the highest risk of becoming ill. There are, however, many other food items that are associated with an increased risk (even if not statistically significant).

You should next think about potential confounders and about effect modification. Think about which variables you might want to check for effect modification or confounding. One common strategy is to base this decision on the results obtained in the univariable analysis and a p-value threshold of 0.20-0.25. Also, food items that are known risk factors for gastroenteritis could also be included regardless of their univariable p-value.

1. **Questions/assignments for the group:**

3.1 Discuss how to identify potential confounders and effect modification. Draw dummy tables before coding to have clear what you want to achieve.



3.2. Install packages (if needed) and load libraries.

3.3. Import your data.

3.4. Consider and assess for confounding and/or effect modification some of the variables. For this, have a look at the relative risk for being a case having eaten a specific food item (for example, veal), when stratified by another variable (for example, pasta). You may consider stratifying by pasta, as it has the highest RR in the univariable analysis.

Facilitator’s note: Have a look at how the team is doing. They should be stratifying veal by pasta. If 10-20 min have passed and they are not at this stage, suggest them to look at that specific stratification.

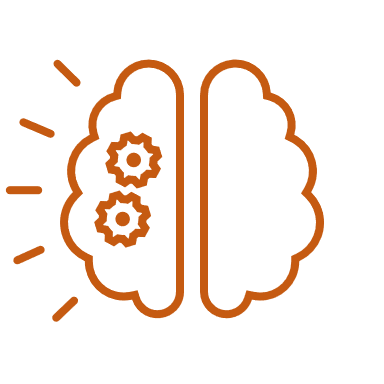
There are many variables in this dataset and it might not make sense to stratify each variable by each other variable on our search for effect modifiers and confounders.

However, we also don't want to be too restrictive as a variable which actually (i.e. causally) is associated with the outcome might not show a significant association at the significant level we decided (say 5%) in the univariable analysis due to confounding. Therefore, we could test all variables statistically significant at the 15%, 20%, or 25% level (specific percentage to be decided by the fellows). In our solutions, we are looking at veal and champagne, stratified by pasta, but fellows may decide to look at others as well.



**Hint**: Use the function csinter() of the EpiStats package.

If we stratify the **effect of veal by pasta** we ask the question: does eating pasta modify or confound the association between eating veal and being a case?



**Stop and reflect**: Are there any indications to make you think there may be effect modification and/or confounding?

Facilitator’s note: We could stratify by pasta (the strongest risk factor in the univariable analysis), to examine if pasta confounds the association between eating veal and being a case. Before stratification, we will need to check if pasta meets the conditions of being a confounder. For a variable to be a confounder it needs to be associated both with the outcome (being a case) and with the exposure (and not be in the causal pathway between exposure and outcome). We know from univariable analysis that pasta is associated with being a case. If we run a Wilcoxon rank sum test, we will see that pasta is also associated with veal (indeed, you can see that most people either had both veal and pasta or neither of these food items, so they are associated with each other).

Use ccinter to stratify and save the object as “pastastrata”.

pastastrata: Within the stratum of the people who ate pasta, veal has no significant effect (RR = 1.20, CI: 0.60 - 2.41). The same holds within the stratum of people who didn’t eat pasta (RR = 1.05, CI = 0.38, 2.92). The adjusted MH-RR also suggests that veal has no effect (RRadj = 1.15, CI: 0.80 - 2.85). To identify confounding, we want to look at the % change between the crude and the adjusted RR. This is given by the csinter output “Adjusted/crude relative change”. The difference between the crude and the MH-RR in this case is >20% suggesting that pasta confounds the association between veal and the disease.

This result suggest that veal is not a risk factor of the disease and that the crude observed effect was due to the confounding effect of pasta.

If you stratify by veal, you see that veal does not confound the association between pasta and the disease. The same applies if you stratify the exposure to pasta by other variables. The above, the higher RR for pasta and the dose response relationship we found earlier for pasta (remember this was optional) provide additional evidence that there was something going on with the pasta with pesto dish!

**Notes for facilitators:**This session will be preceded by a plenary introduction to the proposed approach to stratified analysis (Mantel-Haenszel approach), so fellows should hopefully be able to understand their assignments and apply the suggested functions.

Make fellows think about the extremes: universal exposure (dose response may be useful in these cases) and small cells.

1. **Scenario for the role play:**N/A
2. **Timing:**1,5 hour
3. **Need for materials (logistics)**Laptop, R, Rstudio and R project (with scripts and data)
4. **Deliverables**A script with reproducible code and a paragraph in the draft outbreak report summarizing the stratified analysis, in both methods and results sections.