



## MSc/PGDip/PGCert Health Data Science Assignment cover sheet

<b>Module:</b> <b>(Please tick the appropriate box).</b>	<input type="checkbox"/> PMIM102/J Scientific Computing in Healthcare <input type="checkbox"/> PMIM202/J Health Data Modelling <input checked="" type="checkbox"/> PMIM302 Introductory Analysis of Linked Health Data <input type="checkbox"/> PMIM402/J Machine Learning in Healthcare <input type="checkbox"/> PMIM502/J Health Data Visualisation <input type="checkbox"/> PMIM602 Advanced Analysis of Linked Health Data
<b>Module part (if applicable):</b>	Reflective Journal
<b>Title of assignment:</b>	<i>Reflective Journal</i>
<b>Student ID number:</b>	2311233
<b>Word count (if applicable):</b>	n/a
<b>Declaration:</b>	I understand the following conditions which apply throughout this course:  <ol style="list-style-type: none"><li>1. I confirm that I am the sole author of this work.</li><li>2. I understand that proof reading by a third party is not permitted.</li><li>3. I understand the need for academic integrity and that all my submitted work will adhere to its principles.</li><li>4. I understand that the teaching team will take measures to deter, detect and report any academic misconduct.</li><li>5. I agree to my work being submitted to the TurnItIn academic database.</li><li>6. I understand the importance of assignment deadlines and the need to seek help in good time where personal circumstances interrupt my work.</li></ol>
<b>Please copy and paste this declaration onto the front of the submission.</b>	

**1. What was/were the problem(s) experienced and in what specific exercise(s)?**

**Problem Location:** In day 5, from steps 5-7 faced difficulty in understanding examination of confounding variable and risk adjustment of variables.

**Context:** I understand confounding as a form of bias, where a variable has an unexpected effect on the outcome of interest. Confounding by indication is a type that can be controlled by risk adjustment. When it comes to data, questions of the form “how is it affecting the outcome of interest” and how risk adjustment “controls” it, in a practical scenario is difficult for me to visualize and comprehend. So, the question to reflect and investigate is “how does risk adjustment change the outcome in cox regression?”

**2. What was/were the reason(s) or contributing factor(s) for the problem(s) that arose?**

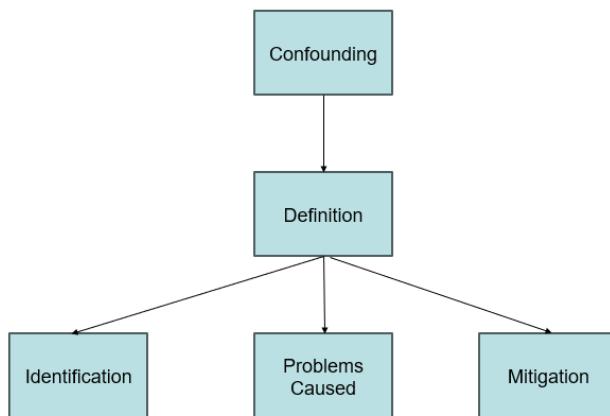
The problem stems from two factors. First, limited knowledge and experience working with risk adjustment in epidemiology. Second, a possibility of misunderstanding the concept of risk adjustment that is leading to unexpected results.

**3. How did you investigate the problem(s) encountered and how did you solve the issue(s)?**

In-order to solve the problem, the investigation was split into two verticals. First, understanding the concept and second, syntax implementation.

**Concept:**

To better understand the concept of confounding variables, drill down technique is used. Drill down technique starts from the main concept and drills down or breaks down into nested concepts that are associated with it. Consider the following visual for reference of the process that I followed.



The process helped to compartmentalize the sections into more easy ingestible parts. I revisited and understood the definition of confounding variables. Understood the difference between effect modifiers. Recognized the shift that occurs in the estimate calculation as a result of the confounding variables. Revisited the examples provided in the workbook and found new ones on the internet about confounding variables in epidemiology setting. I understood that to fix the shift, risk adjustment is necessary. Explored the methods that are used in risk adjustment mentioned in the workbook, especially “charlson index” that is used in the syntax of solution file.

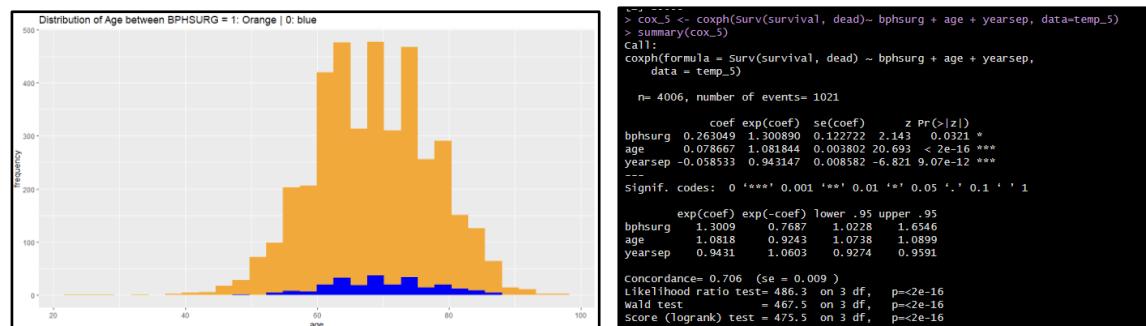
## Syntax:

Age variable is identified as one of the confounding variables. BPHSURG is the outcome of interest. The mean age and year of separation is checked for BPHSURG = 0 and BPHSURG = 1 to understand the potential effects of confounding. The mean value for BPHSURG = 1 seems to be lesser than mean value for BPHSURG = 0. But, how exactly it is confounding from a practical perspective? just because mean is different does it imply that it is confounding?

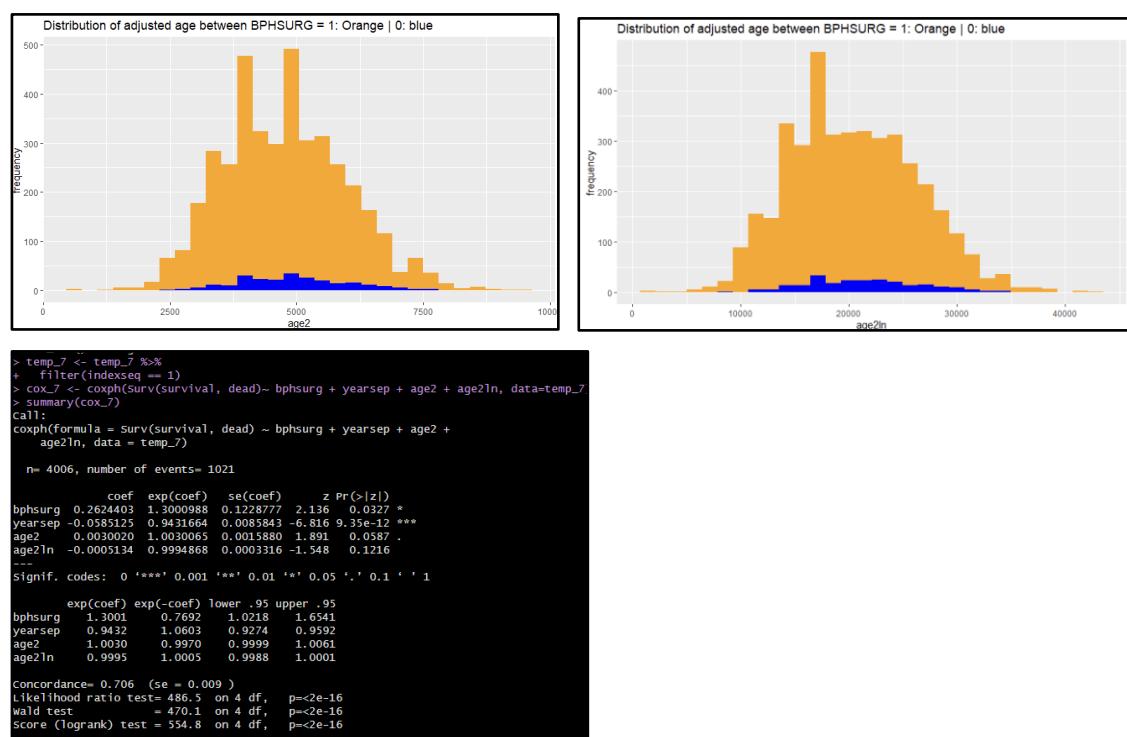
So, what exactly is missing? What would aid to improve understanding?

There was no comparison of cox regression output before and after non-linearity continuous covariate adjustments. This was missing in the solution file. Step 5 does cox regression without including the age variable and in step 7 adjusted age variable is used in cox regression. In the step 6, examination of confounding variable is only limited to comparing the mean values without directly including the age variable in the cox regression equation. It would fill in the gaps if the output of cox regression with and without adjustments of age variable were provided and comparisons were made.

Following is the syntax and output for age variable before adjustments:



Following is the syntax and output for age variable after adjustments:



It is observed that in cox regression age is considered as a significant variable but after adjustment, only the squared age variable is observed as a significant variable. The distribution for age and squared are almost similar. However, the distribution of natural log of age is not similar. This proves that age is confounding variable.

**4. What conceptual and analytical understanding was gained from experiencing the problem(s)?**

**Conceptual Understanding:**

I understood the effects of a confounding variable in practice. The intuition and experience of dealing with confounding variable and its corresponding risk adjustments solidified with the investigation and implementation. It is observed in the before-after analysis, how age comes up as a significant variable in the cox regression at first and later how after adjustments to the variable, it is not significant anymore.

**Analytical Understanding:**

Comparative analysis of syntax output helped to derive better understanding of the implementation of approaches that could be used for long term purposes. The process of before-after analysis developed my thought process to be more objective and critical. The effects of confounding were practically understood by adding on to the solution file syntax. This enabled me to clearly distinguish between the two scenarios and observe the importance of risk adjustments to the confounding variable.

**5. How could your learning(s) from this experience be applied to prevent the problem(s) occurring in future research work?**

First, emphasis of concept reviewing, learning from this experience not only boosted my confidence in the analytical skills and syntax knowledge, but also help develop better understanding of confounding variables, its effects and remediation. It emphasized on checking the syntax to ensure that the implementation does what it is intended to do.

Second, emphasis on readability of the code. The need for better commenting of the code is essential for transfer of knowledge and understanding of the implementation of syntax. The learning from this experience would help save my time and the future syntax reader's time. Better commented syntax would provide clarity of the syntax and easy transfer of knowledge.

Third, emphasis on gap analysis and process. If there was an explanation as above in the solution file, it would have been faster learning and would be more relatable to a tangible output on the concept of confounding.