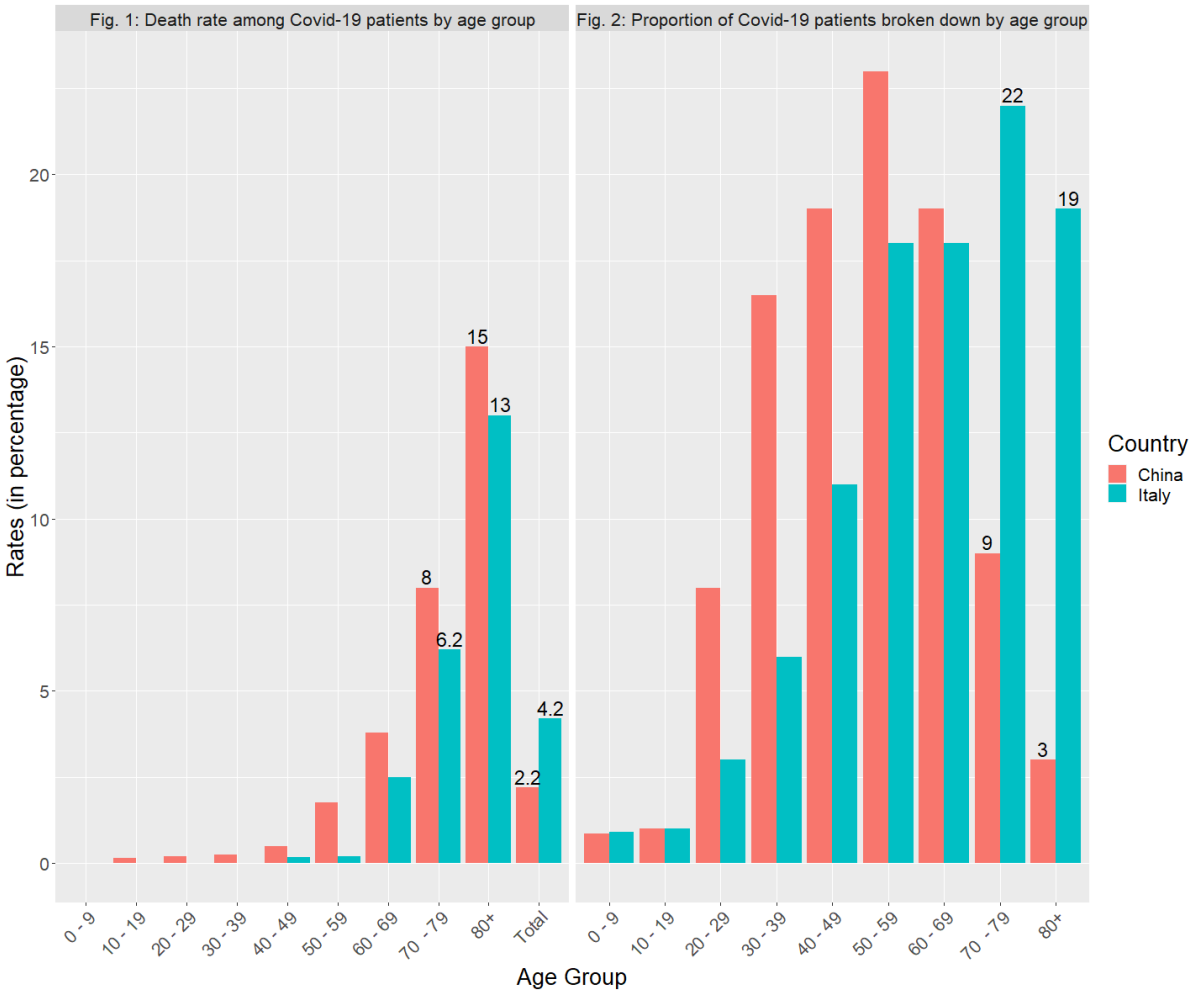
GEA1000 QUANTITATIVE REASONING WITH DATA
TUTORIAL 2

Please work on the problems before coming to class. In class, you will engage in group work.

Case Study 1: COVID in Italy and China

The following figures were obtained from a study[1] investigating the death rate among the COVID-19 patients (Case fatality rate) in Italy and China during the early COVID stages. Observe that for all age groups, death rates among Covid-19 patients in Italy are lower than those in China, but the overall death rate among the Covid-19 patients in Italy is higher than that in China.



1. Let us designate Covid-19 patients aged **70 years and above** as "Old", and all other Covid-19 patients as "Young". Let "D" represent death from Covid-19.

    a. What proportion of patients in Italy are old? How about China? Which country, Italy or China, is positively associated with old patients? 41, 12, Italy

[1] Source: https://arxiv.org/pdf/2005.07180.pdf

b. In Italy, what is the death rate amongst the old patients, rate(D|Old)? Give your answers to 2 decimal places. $(0.062*0.22 + 0.13*0.19) / (0.41) = 9.35\%$

c. From Fig 1, the overall death rate in Italy is 4.2%. Using the basic rule of rates, what must be the possible range of the death rate amongst the young patients in Italy? Is there an association between age and death among Covid-19 patients in Italy?

$$b(0.41) + c(0.59) = 0.042$$
$$c = (0.042 - b(0.41)) / 0.59$$
$$= 0.621\%$$

d. Repeat parts (b) and (c) for China. $(0.08*0.09 + 0.15*0.03) / (0.12) = 9.75\%$

$$b(0.12) + c(0.88) = 0.022$$
$$c = (0.022 - 0.0975(0.12)) / 0.88$$
$$= 1.17\%$$

e. Let's assume the following rough estimates from Fig 1:
In Italy, rate(D|Young) = 0.621%. In China, rate(D|Young) = 1.17%.
Using the information from Q1(a) to (d), explain how it is possible for the overall death rate in Italy to be higher than that in China, despite Italy having a lower death rate in China for every age group, as shown in Fig 1.

Hint: You may use the following table to help you:

	Italy	China
rate(D|Old)	0.0935	0.0975
rate(D|Young)	0.00621	0.0117
rate(D)		
rate(old)	0.41	0.12

Case Study 2: Confounders and Simpsons Paradox

This question is based off a South African longitudinal study of growth of children, referred to as the Birth to Ten study (BTT). Census data of children born during a seven-week period between April and June 1990 were collected in the Johannesburg/Soweto metropolitan area of South Africa. The information collected included the **child's race (White / Black)** and whether they **received medical aid or not (Aid / No Aid)**. Having medical aid for the child is like having health insurance.

Five years later, a follow-up study on the same cohort of children was conducted. However, only 416 out of 1590 of the participants responded, despite medical screening being provided to the children as part of the follow-up study. These 416 children are labelled as "**Traced**", whereas those that did not respond are labelled as "**NotTraced**". Refer to the dataset: *Africa_study.xls*.

2. Use appropriate software to answer the following questions. Give your answers to 2 decimal places.

    a. What can you say about the proportions of whites and blacks in this study?

    b. Determine if there is any association between the variables "Follow Up" and "Medical Aid". Do children with medical aid tend to not follow up?

    c. Determine whether "Race" is a confounder in examining the association between the variables "Follow Up" and "Medical Aid".

    d. In relation to (c), do we observe Simpson's Paradox when investigating the association between the variables "Follow Up" and "Medical Aid"?

The aim of the study was to identify factors related to the emergence of cardiovascular disease (CVD) risk factors in children living in an urban environment in South Africa. Suppose that hypothetically, we have collected the data of these children many years later in a survey, to determine whether they ended up with cardiovascular disease or not, shown under the CVD (Yes/No) column. We are now interested in knowing if medical aid helps to mitigate the risk of CVD.

Thus, the outcome of interest is now "CVD", and the treatment variable is "Medical Aid".

    e. Determine if there is any association between the variables "Medical Aid" and "CVD".

    f. Determine whether "Race" is a confounder in examining the association between "Medical Aid" and "CVD".

    g. In relation to (f), do we observe Simpson's paradox when investigating the association between the variables "Medical Aid" and "CVD"?

Case Study 3: Confounders in an experimental study

Background: Polio, also known as infantile paralysis, is an infectious disease that strikes young children, often causing permanent paralysis. It spreads through person-to-person contact. In the 1950's, American scientist Jonas Salk developed a vaccine that protected monkeys from polio and was safe when injected into human subjects in the laboratory in the 1950's. By 1954, the vaccine was ready to be tested in the real world.

In order to determine if the polio vaccine reduces the risk of polio infection, a cohort of children were invited to take part in a study. However, only some children had parental consent to receive the vaccine, which posed a problem for researchers.

Two different study designs are conducted:

First design: Children with parental consent were divided into two groups – treatment and control. The treatment group was vaccinated, and no placebo was given to the control group. All children without parental consent were placed in the control group as well, since they were not allowed to take the vaccine. This was known as the NFIP study *(you may read up more on this if you wish).*

Second design: Only children with parental consent were considered in this study. Those without consent were excluded. Children with parental consent were assigned into the treatment and control groups based on a 50-50 randomised procedure. The control group received a placebo injection of salt dissolved in water. Doctors involved in the diagnosis were not told which group the children belonged to.

The results of the two studies are tabulated as shown below.

NFIP study	Sample size	Polio +ve	Polio -ve
Vaccinated (with parental consent)	225000	56	224944
Control (parental consent or no parental consent)	725000	391	724609

Randomized controlled trial	Sample size	Polio +ve	Polio -ve
Treatment	200000	56	199944
Control	200000	142	199858

Furthermore, some important characteristics of the different groups were discovered:

- Families who provided consent tended to be of a higher income group and as a result lived in more hygienic conditions.
- Children living in more hygienic conditions were more susceptible to polio as they were not exposed to the virus since young and lacked immunity to the virus.

3a) Calculate the following for both study designs. For rates, leave your answers in percentages corrected to 3 s.f.

(i) Find the conditional rate of getting polio given that they are vaccinated. 56/225000 56/200000

(ii) Find the conditional rate of getting polio given that they are in control group. 391/725000 142/200000

(iii) Comment on the appropriateness of using rates rather than absolute numbers for comparison in the NFIP study. More accurate story

3b) Compare the two study designs by answering the questions below.

(i) Were the children randomly assigned into treatment and control groups?

NFIP study	Randomized controlled trial (RCT)
No. Based on preference	Random Assignment - E.g. a random number generator was done to assigned children with parental consent into the treatment or control group.

(ii) Discuss how the assignment in 3b(i) might affect the results of the studies

NFIP study	Randomized controlled trial (RCT)
They already know	- With large enough samples, all the important characteristics of both treatment and control group subjects will resemble each other very closely - Allows for only the differences to be due to the response to the vaccine.

(iii) Were the subjects/assessors blinded?

NFIP study	Randomized controlled trial (RCT)
It was not stated if the doctors are blinded but the children knew if they were vaccinated or not as no placebo was given.	Double blinding

(iv) Discuss how the blinding in b(iii) might affect the results of the studies

NFIP study	Randomized controlled trial (RCT)
Double-blinding was not done for the NFIP study, which may lead to bias. For example, children in the control group who know that they are not getting the vaccine might result in them taking their own measures to reduce their risk of getting the disease. On the other hand, vaccinated individuals might think they now have full immunity and would be less stringent in health safety.	No bias towards either side, removing most of the confounders associated with the experiment
Remark: In general, the importance of blinding depends on the context of the study. In some cases, whether subjects are blinded or not may not affect the results of the study. Nonetheless, blinding and the use of placebos act as important precautionary measures, especially in situations where it is difficult to tell if blinding can have an impact on the results or not.	

(v) According to the table of results, by how much does the vaccine reduce the polio rate? If we consider the discussions in b(i) to (iv), do you think the vaccine more or less effective than what is stated in the table of results? Consider each study separately.

NFIP study	Randomized controlled trial (RCT)
Factor of 2, but likely better	- The vaccine reduces the polio rate by 0.071% - 0.028% = 0.043% - The actual effect is likely to be similar to 0.043% as the study was conducted with randomised assignment, double blinding and large sample size.