Statistical report:

Causal Graphs (worth 40/100 marks)

Question 1: To find the values in a scenario where it's mediated by age, we need to use CausalModelMediator.

```
Using Mediator Model:
Exact ACE (P(F='YES'|Country='Italy') - P(F='YES'|Country='China')) =  0.11230984519563034
Estimated ACE (P(F='YES'|Country='Italy') - P(F='YES'|Country='China')) =  0.11800000000000224
```

Question 2: To find the values in a scenario where it's confounded by age, we need to use CausalModelConfounder.

```
Using Confounder Model:
Exact ACE Sum_over_Ages (P(Age)*(P(F='YES'|Country='Italy',Age) - P(F='YES'|Country='China',Age))) =  0.07010325890807337
Estimated ACE Sum_over_Ages (P(Age)*(P(F='YES'|Country='Italy',Age}) - P(F='YES'|Country='China',Age}))) =  0.0813587235872356
```

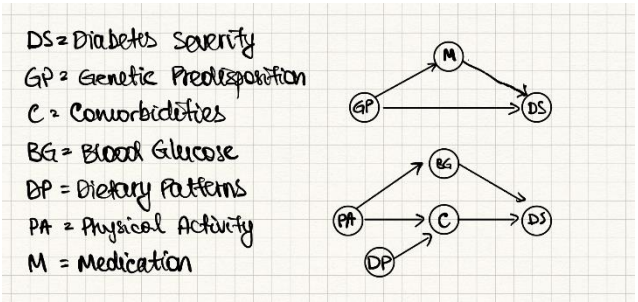
Question 3: From the result presented above, we can see that the model assumption changes our final result. Our question asks us to estimate the *true* effect of a country on fatality. Although Q2 provides us with granularity by adding an age component to our calculation, I do not find it necessary in the context of the given question. We care about the country's impact on fatality, which means that using a causal model with a mediator will suffice to arrive at a correct conclusion. My preference goes to using Exact ACE and here is why. We do not have a huge Bayesian Net for this question and computationally it's possible to calculate all probabilities accurately. As such, there is no hindrance to us in using exact ACE. However, we covered it in class, truly enormous sets of data are hard to calculate because BN's become uncontrollably large. In such a case, we may consider resorting to the Estimated ACE as it will produce a result that is "good enough" to use in our argument.

Question 4:

Air pollution would be a great confounder since it independently affects the likelihood of people having respiratory illnesses which COVID is classified as (and thus contributed to the measurement of fatality), and it is not influenced by any of our other variables while it also does not directly cause COVID.

Level of exercise (measured in days, as an example) would be a great mediator since it will be dependent on the age group in question (hence, we can say it is influenced by another variable) while it also does indicate whether you have better chances of survival.

Creating a Causal Graph (worth 10/100)



In graph the on the top, we observe M as a mediator. There is a logical sequence between these variables. If you get a condition at the time of birth, you might be prescribed medication to avoid a bad case of diabetes. As such, M depends on GP but it can change the outcome of severity depending on whether you take medication as prescribed or not. Contributing to the severity of your case of diabetes.

In the graph on the bottom, we have a confounding variable PA. It does not depend on anything else whether you decide to exercise or not. But it is beneficial as it's proven to decrease chances of getting diabetes if you exercise regularly. On its own, it does not affect the severity of diabetes, but

it can contribute to either improving or bringing down your health conditions which will affect if you get diabetes.

C is an example of a collider variable (in the second graph at the bottom right). If we exercise AND take care of our nutrition, we significantly decrease our chances of catching any type of illness thanks to a very strong immune system. Hence, your predisposition to having any sickness or medical condition is highly correlated to your overall lifestyle. On the other hand, if you do get another illness, it will compromise your immune system making you more prone to complications. Either by attaining diabetes OR causing worsening of your diabetes condition.