



## Risk factors for COVID-19

By Matthew Edwards for

**COVID-19 Actuaries Response Group – Learn. Share. Educate. Influence.**

### Summary

Very little ‘multivariate’ analysis has been done to quantify the effect of COVID-19 risk factors simultaneously. Hence there is potential to misunderstand the operation of risk factors because of confounding – where the effect of one risk factor is distorted by the effect of another.

Age and biological sex are clearly risk factors. Obesity, hypertension and medical conditions such as diabetes appear to be risk factors, but their effect has been exaggerated in much commentary because they are strongly correlated with age and sex. Ethnicity also appears to be a risk factor. The evidence regarding smoking is unclear.

COVID-19 risk factors can operate at two levels: “morbidity”, the risk of developing a severe form of COVID-19, and “mortality”, the risk that someone already unwell due to COVID-19 then dies. Risk factor effects should be compared against normal (all-cause) mortality.

Understanding risk factors better can help with understanding the underlying pathogenesis, and with evaluating strategies to move beyond lockdown.

### Introduction

A primary objective of epidemiology in respect of any disease is to establish the ‘risk factors’ that affect morbidity (succumbing to the disease) and mortality (dying from the disease). For most established diseases, for instance major cancers or heart disease, we now take for granted the ability to analyse large, well-curated datasets using relatively advanced statistical techniques in order to identify and quantify the risk factors.

The short ‘history’ of COVID-19 means that, to date, very little ‘multivariate’ analysis has been done to quantify the effect of risk factors simultaneously. Risk factors have mostly been considered in isolation, through simple ‘one-dimensional’ analyses. This creates potential to misunderstand the operation of risk factors regarding COVID-19 because of confounding – where the effect of one risk factor is distorted by the effect of another.

#### *Example of confounding*

In the example below, we set out how the effect of one risk factor can be distorted by another major risk factor with which it is correlated. The numbers used are ‘made up’ for simplicity, but they are broadly plausible. We have based it on a population split into two age bands, and two risk levels (smoker and non-smoker). We start with two elements: the population split across this 2x2 combination of age bands and smoker status, and some assumed mortality relationship between them (with smoker mortality higher than non-smoker mortality).

So in this example, we have:

Population	Smokers	Non-smokers
Age 50-80	100	400
Age >80	50	450
Total	150	850

Mortality	Smokers	Non-smokers
Age 50-80	1.4%	1.0%
Age >80	10.5%	7.5%

Multiplying the population by the mortality rates, we expect the following deaths:

Deaths	Smokers	Non-Smokers
Age 50-80	1.4	4.0
Age >80	5.3	33.8
Total	6.7	37.8

However, if we now compare overall smoker and non-smoker mortality, we see that both are almost identical ( $\frac{6.7}{150} = 4.47\%$ ;  $\frac{37.8}{850} = 4.45\%$ ).

So, although we know that smokers have higher mortality this is not apparent from the one-way analysis. This is a direct result of a correlation in the data between age and smoking status – the lives in the example were not distributed evenly across the four ‘cells’.

We can see from this how the effect of a factor such as smoking, if damaging in the context of COVID-19, could be masked by the effect of age – and hence it would appear from a simple look at the data that the factor had no effect on COVID-19 mortality. Conversely, it is also possible for insignificant factors to appear influential.

### Terminology

We should note that the meaning of “risk factor” varies across professions. Actuaries generally use it to denote something measurable with a material impact on the risk in question. Hence age is a risk factor for all-cause mortality despite just being a proxy for physiological degeneration.

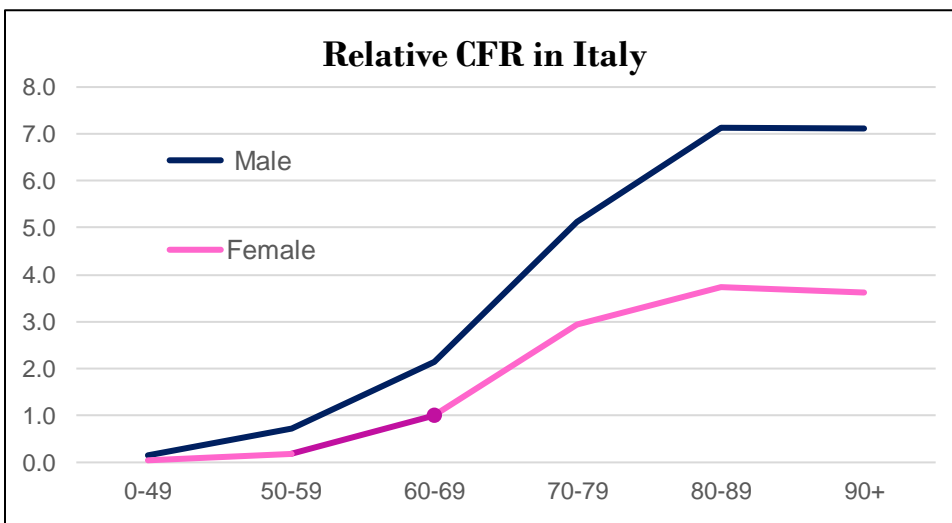
Likewise, in the context of COVID-19 we are interested in the observed relationships, even if features such as age are proxies for poorly understood ‘real’ risk factors such as the strength of the immune system.

### What we know about risk factors so far

#### Age and sex

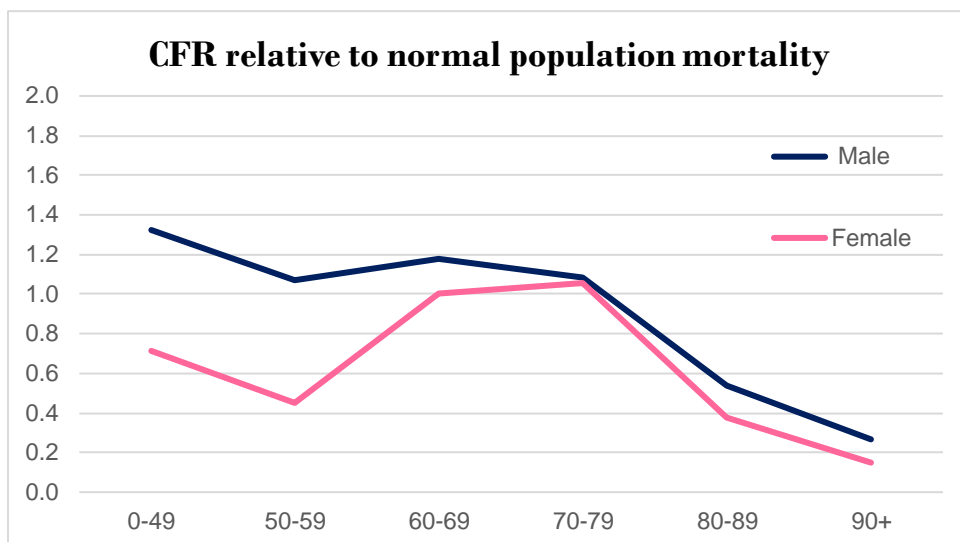
The one thing we can be sure of from data accumulated so far is that age and sex are very strong risk factors. For instance, using the most recent Italian data<sup>i</sup> (because of the age and sex breakdown, and the sadly large volumes) shows us the following relative age and sex case fatality rate (‘CFR’).

We present it as relative to avoid, in this context, the many uncertainties regarding the absolute numbers – for instance, the effect of unknown cases, as discussed in [a previous bulletin](#).



Here, compared with the 'Female 60-69' band taken as a reference point (the choice of reference point is arbitrary), we can see both the clear increase of the CFR with age and the effect of sex.

However, although the above relativities show the age and sex impacts of COVID-19, it is useful to interpret them in the context of normal (all-cause mortality) age and sex effects. If we normalise for Italian population mortality, we find the effects above not only disappear but to some extent reverse.



So, there is a strong age and sex effect with COVID-19, but it is not out of line with normal mortality. If anything, the high-age effect seen with COVID-19 is less marked than for normal mortality.

#### *Other risk factors*

Obesity, hypertension and diabetes appear to be risk factors, with their prevalence in COVID-19 cases and deaths noted by (for instance) the Istituto Superiore di Sanità in Italy and the Centers for Disease Control and Prevention in North America.

In the UK, the Intensive Care National Audit and Research Centre ('ICNARC') publishes weekly reports showing the characteristics of COVID-19 patients admitted to intensive care units (ICU), and also the characteristics of patients for whom outcomes are recorded (those who have been discharged from ICU, or died).

Looking first at obesity, we see that in the most recent report obese patients account for 38% of ICU cases. This compares with UK population obesity prevalence of 30% (above age 65). Obese patients have a 50% chance of surviving ICU, consistent with other patients.

So there seems to be a material morbidity effect, although it is not overwhelming given the population statistics, and even this increased propensity may be partly explained by other factors. There is no discernible mortality effect.

Two factors that have started to receive more attention are ethnicity and socio-economic status.

Using the same data, we find that the two most deprived of the five quintiles used in the report account for 50% of ICU cases. 40% would be expected if there were no socio-economic effect. The most deprived patients have a 47% chance of surviving ICU, compared with 51% for less deprived groups.

Again, a material morbidity effect, though there will be confounding with obesity in particular (prevalence is higher among more deprived groups). There is only a small mortality effect, and it is lower than the socio-economic mortality differentials seen in the UK for all-cause mortality.<sup>ii</sup>

Considering ethnicity, the ICNARC report shows that non-White patients account for 34% of ICU cases. This compares with a 23% non-White proportion of the population in the patients' locations. Non-white patients have a 45% chance of surviving ICU, compared to 52% for whites.

The non-white admissions are higher than expected, and it is not plausible that this is caused by the non-white population being older (in fact the reverse is true<sup>iii</sup>). It seems that ethnicity is a strong risk factor for morbidity and may also be significant for mortality.

Finally, one risk factor that has surprised many has been smoking. So far, there has been no evidence from 'one-dimensional' analysis showing that smokers clearly exhibit higher morbidity or mortality from COVID-19. Indeed, there has been speculation that smoking may confer some degree of protective benefit.

This highlights the importance not only of being aware of the dangers of confounding, but also bearing in mind the need to interpret data with regard to clinical and biological plausibility. It seems unlikely that a smoking history would result in lower mortality risk for instance, as smokers will generally have damaged respiratory systems from prolonged tobacco use.

### What multivariate analyses are there?

One 'by patient' multivariate analysis is a New York hospital study, 'Factors associated with hospitalization and critical illness among 4,103 patients with COVID-19 disease in New York City' by Petrilli et al<sup>iv</sup>.

This brought out the following main results in respect of hospitalisation, where the multivariate analysis allows results to be presented as 'odds ratios' (i.e. showing the increased chance of hospitalisation from the characteristic in question, allowing at the same time for the effect of other factors in the data). The odds ratios are expressed relative to some assumed 'base level', e.g. for age, higher age band effects are expressed relative to those for the 19-44 years banding).

Characteristic	Odds ratio	Compared with base of
Age 45-54	2.6	19-44 years
Age 55-64	4.2	(as above)
Age 65-74	10.9	(as above)
Age 75+	66.8	(as above)
Diabetes	2.8	No diabetes
Male sex	2.8	Female
Hypertension	[1.2]	Absence of hypertension
Obesity	4.3	BMI < 30
Ethnicity	1.4 (Asian) – 2.0 (other)	White
Tobacco use	0.7	None

A few comments:

- The hypertension result is shown in square brackets as it was not statistically significant, but we thought it interesting to observe how small the effect is compared with other factors, given how much the condition has been noted in other analyses;
- The obesity result shown is for BMI 30-39.9, above BMI of 40 the effect is higher (6.2);
- For ethnicity, African-Americans showed no significant difference, but as with hypertension it is interesting to see the result anyway – the risk factor was [0.88];

- The tobacco result is interesting.

The study also considered severe outcomes (defined as needing intensive care, or ventilation, or discharge to hospice, or death). What is interesting at this point is that very little was found to be significant here. Many of the medical conditions that were significant for hospitalisation (such as diabetes) were not significant.

If we look at age, we see:

Age	Odds ratio
To 18	6.3
19-44	Base (reference) band
45-64	[Not significant]
55-64	1.3
65-74	1.9
75+	2.6

Regarding ethnicity, the study found:

Ethnicity	Odds ratio
White	Base (reference) band
African-American	0.6
Asian	1.9
Other	[Not significant]

Another multivariate analysis we are aware of looks at a large number of hospitalised UK patients, with particular regard to prior medical history<sup>v</sup>. The event studied was survival vs death, with the following factors found significant for mortality:

Factor	Odds ratio
Age < 50	Base (reference) band
Age 50-69	4.0
Age 70-79	9.6
Age 80+	13.6
Sex = female	0.8
Chronic cardiac disease	1.3
Chronic pulmonary disease	1.2
Chronic kidney disease	1.2
Cancer	1.2
Obesity	1.4
Dementia	1.4

The study did not appear to consider social class, ethnicity, hypertension or smoking habits.

Overall, the effects of prior medical conditions appear low compared with one-way COVID-19 analyses we have seen, and this would be consistent with confounding between medical history and age. The sex effect (20% lower mortality for women) appears low compared with other results.

## Conclusion

We are only now starting to have enough data to look at risk factors properly, through multivariate analyses. At the moment, before we have good studies in a range of countries to consider, the most important thing is to be aware of the issue of confounding, and hence avoid 'double-counting' some aspects.

The multivariate analyses we have seen show results that are appreciably different from the one-way analyses of COVID-19 morbidity and/or mortality, these differences being generally consistent with the confounding mentioned (for instance, hypertension 'disappears' as a factor because it was confounded with age).

It is also advisable to consider the issue of biological plausibility of any results.

Understanding risk factors better can help with understanding the underlying pathogenesis, and with evaluating strategies to move beyond lockdown.

Matthew Edwards

1 May 2020

---

<sup>i</sup> Author's simple calculations from data in 'Epidemia COVID-19, Aggiornamento nazionale 23 aprile 2020', [www.iss.it](http://www.iss.it)

<sup>ii</sup> eg the work by Cairns and Kleinow

[https://www.actuaries.org.uk/system/files/field/document/E3%20Kleinow\\_LifeConf\\_2018V2.pdf](https://www.actuaries.org.uk/system/files/field/document/E3%20Kleinow_LifeConf_2018V2.pdf)

<sup>iii</sup> <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest#main-facts-and-figures>

<sup>iv</sup> Factors associated with hospitalization and critical illness among 4,103 patients with COVID-19 disease in New York City, Christopher M. Petrilli, Simon A. Jones, Jie Yang, Harish Rajagopalan, Luke F. O'Donnell, Yelena Chernyak, Katie Tobin, Robert J. Cerfolio, Fritz Francois, Leora I. Horwitz, medRxiv 2020.04.08.20057794

<sup>v</sup> Features of 16,749 hospitalised UK patients with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol, medRxiv 2020.04.23.20076042