

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354601308>

# Paradoxes in the reporting of Covid19 vaccine effectiveness: Why current studies (for or against vaccination) cannot be trusted and what we can do about it

Preprint · September 2021

DOI: 10.13140/RG.2.2.32655.30886

CITATION

1

READS

5,789

3 authors:



**Norman Elliott Fenton**

Queen Mary, University of London

386 PUBLICATIONS 15,497 CITATIONS

SEE PROFILE



**Martin Neil**

Queen Mary, University of London

197 PUBLICATIONS 6,102 CITATIONS

SEE PROFILE



**Scott Mclachlan**

Queen Mary, University of London

62 PUBLICATIONS 337 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



KNIFE (Knowledge Discovery from Health Use Data) [View project](#)



Bayes and the Law [View project](#)

# Paradoxes in the reporting of Covid19 vaccine effectiveness

## *Why current studies (for or against vaccination) cannot be trusted and what we can do about it*

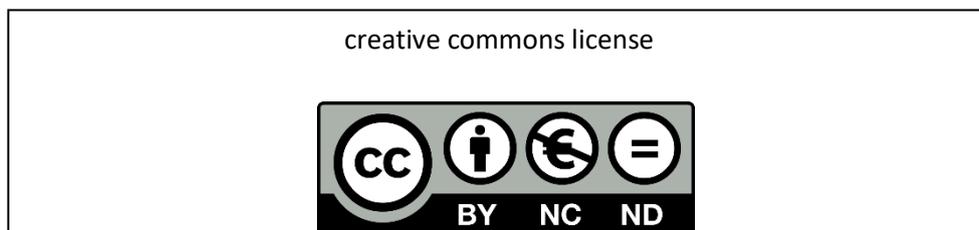
Norman Fenton, Martin Neil and Scott McLachlan

Risk Information and Management Research  
School of Electronic Engineering and Computer Science,  
Queen Mary University of London

15 Sept 2021

### Abstract

Given the limitations of the randomized controlled trials (RCTs) for Covid19 vaccines, we must increasingly rely on data from observational studies to determine vaccine effectiveness. But over-simplistic reporting of such data can lead to obviously flawed conclusions due to statistical paradoxes. For example, if we just compare the total number of Covid19 deaths among the vaccinated and unvaccinated then we are likely to reach a different conclusion about vaccine effectiveness than if we make the same comparison in each age category. But age is just one of many factors that can confound the overall results in observational studies. Differences in the way we classify whether a person *is vaccinated* or *is a Covid19 case* can also result in very different conclusions. There are many critical interacting causal factors that can impact the overall results presented in studies of vaccine effectiveness. Causal models and Bayesian inference can in principle be used to both explain observed data and simulate the effect of controlling for confounding variables. However, this still requires data about relevant factors and much of these data are missing from the observational studies (and the RCTs). Hence their results may be unreliable. In the absence of such data, we believe the simplest and most conclusive evidence of vaccine evidence is to compare all-cause deaths for each age category between those who were unvaccinated and those who had previously had at least one vaccine dose.



The randomized controlled trials (RCTs) to establish the safety and effectiveness of Covid19 vaccines produced impressive results (Polack et al., 2020) but were inevitably limited in the way they assessed safety (Folegatti et al., 2020)<sup>1</sup> and are effectively continuing (Ledford, Cyranoski, & Van Noorden, 2020; Singh et al., 2021) . Ultimately, the safety and effectiveness of these vaccines will be determined by real world observational data over the coming months and years.

However, data from observational studies on vaccine effectiveness can easily be misinterpreted leading to incorrect conclusions. For example, we previously noted<sup>2</sup> the Public Health England data shown in Figure 1 for Covid19 cases and deaths of vaccinated and unvaccinated people up to June 2021. Overall, the death rate was three times higher in the vaccinated group, leading many to conclude that vaccination increases the risk of death from Covid19. But this conclusion was wrong for this data because, in each of the different age categories (under 50 and 50+), the death rate was **lower** in the vaccinated group.

<50 category			
	Cases	Deaths	%
vaccinated	19693	2	0.010
unvaccinated	52846	6	0.011

>50 category			
	Cases	Deaths	%
vaccinated	7499	68	0.907
unvaccinated	976	38	3.893

Aggregated			
	Cases	Deaths	%
vaccinated	27192	70	0.257
unvaccinated	53822	44	0.082

Figure 1 Data from Public Health England, June 2021

This is an example of Simpson’s paradox (Pearl & Mackenzie, 2018). It arises here because most vaccinated people were in the 50+ category where most deaths occur. Specifically: a) a much higher proportion of those aged 50+ were vaccinated compared to those aged <50; and b) those aged 50+ are much more likely to die.

So, as shown in Figure 2(a), ‘age’ is a **confounding** variable. While it is reasonable to assume that death is dependent on age, in a proper RCT to determine the effectiveness of the vaccine we would need to break the dependency of vaccination on age as shown in Figure 2(b), by ensuring the same proportion of people were vaccinated in each age category.

<sup>1</sup> Some participants and sites were unblinded and non-randomised and others were effectively unblinded when they received paracetamol prior to jab

<sup>2</sup> <https://probabilityandlaw.blogspot.com/2021/06/simpsons-paradox-in-interepretation.html>

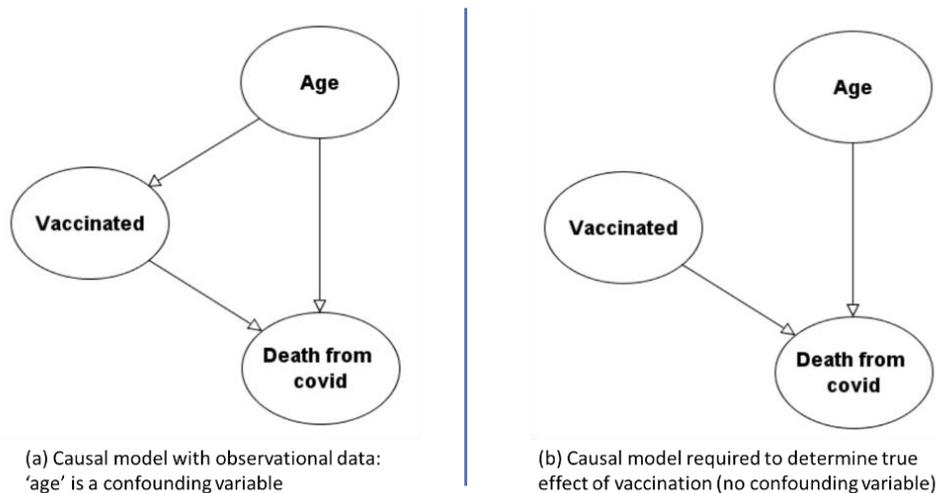


Figure 2 Causal model reflecting the observed data

The Appendix demonstrates how this causal model, and Bayesian inference, can both explain the paradox and avoid it (by simulating an RCT). Using the model in Figure 2 (b), which avoids the confounding effect of age, we conclude (based only on the data in this study) that the (relative) risk of death is four times higher in the unvaccinated (0.417%) than the vaccinated (0.104%), meaning the absolute increase in risk of death is 0.313% greater for the vaccinated.

An excellent article by Jeffrey Morris<sup>3</sup> demonstrates the paradox in more detail using more recent data from Israel.

Clearly confounding factors like age (and also comorbidities) must, therefore, always be considered to avoid underestimating vaccine effectiveness data. However, the conclusions of these studies are also confounded by failing to consider non-Covid deaths, which will **overestimate** the safety of the vaccine if there were serious adverse reactions.

In fact, there are many other confounding factors that can compromise the results of any observational study into vaccine effectiveness (Krause et al., 2021). By 'compromise' we mean not just over- or underestimate effectiveness, but - as in the example above - may completely reverse the results if we fail to adjust even for a single confounder (Fenton, Neil, & Constantinou, 2019).

In particular, the following usually ignored confounding factors will certainly **overestimate** vaccine effectiveness. These include:

- The *classification of Covid19 deaths and hospitalizations*. For those classified as Covid19 cases who die (whether due to Covid19 or some other condition), there is the issue of whether the patient is classified as dying 'with' Covid19 or 'from' Covid19. There may be differences between vaccinated and unvaccinated in the way this classification is made. The same applies to patients classified as Covid19 cases who are hospitalized.
- The *number of doses and amount of time since last dose* used to classify whether a person has been vaccinated. For example, any person testing positive for Covid19 or dying of any cause within 14 days of their second dose is now classified by the CDC as 'unvaccinated' (CDC, 2021). While this definition may make sense for determining effectiveness in preventing Covid19 infections, it may drastically overestimate vaccine safety; this is because most serious adverse reactions from vaccines in general occur in the first 14 days (Scheifele, Bjornson, & Johnston, 1990; Stone, Rukasin,

<sup>3</sup> <https://www.covid-datascience.com/post/israeli-data-how-can-efficacy-vs-severe-disease-be-strong-when-60-of-hospitalized-are-vaccinated>

Beachkofsky, Phillips, & Phillips, 2019) and the same applies to Covid19 vaccines (Farinazzo et al., 2021; Mclachlan et al., 2021). There is also growing evidence that people hospitalized for any reason within 14 days of a vaccination are classified as unvaccinated and, for many, as Covid19 cases<sup>4</sup>.

- The *accuracy of Covid19 testing and Covid19 case classification*. These are critical factors since there may be different testing strategies for the unvaccinated compared to the vaccinated. For example, in the large observation study of the Pfizer vaccine effectiveness in Israel (Haas et al., 2021) unvaccinated asymptomatic people were much more likely to be tested than vaccinated asymptomatic people, resulting in the unvaccinated being more likely to be classified as Covid19 cases than vaccinated<sup>5</sup>.

Even if we wish to simply study the effectiveness of the vaccine with respect to avoiding Covid infection (as opposed to avoiding death or hospitalization) there are many more factors that need to be considered than currently are. To properly account for the interacting effects of all relevant factors that ultimately impact (or explain) observed data we need a causal model such as that in Figure 3.

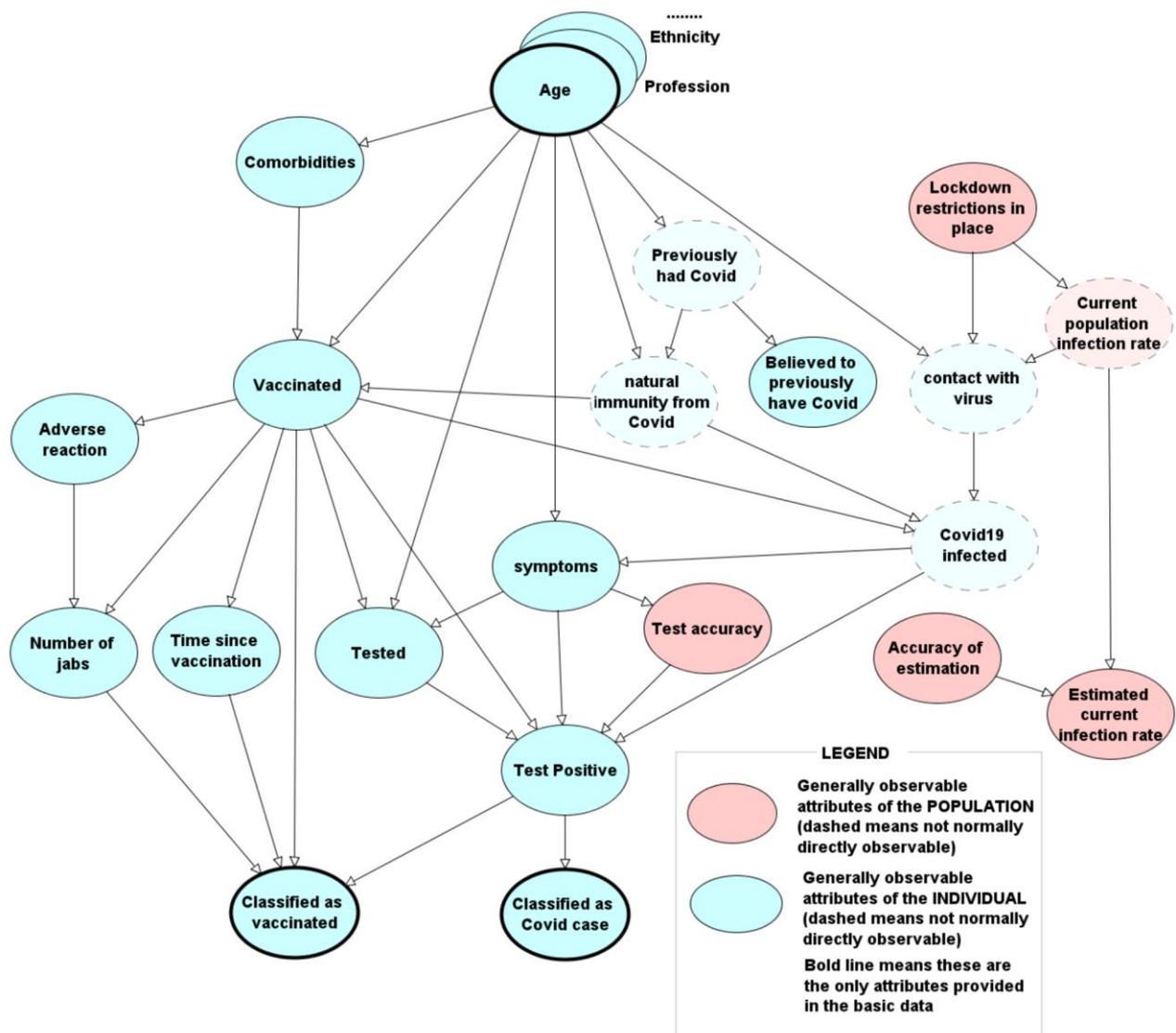


Figure 3 Causal model to determine vaccine effectiveness

<sup>4</sup> <https://www.bitchute.com/video/IXrcpFe4V4U2/>

<sup>5</sup> <https://probabilityandlaw.blogspot.com/2021/05/important-caveats-to-pfizer-vaccine.html>

As in the simple model of Figure 2, the nodes in the model shown in Figure 3 correspond to relevant factors (some of which relate to individuals – like age, and some of which relate to the population – like whether lockdowns are in place) and an arc from one node to another means there is a direct causal/influential dependence in the direction of the arc. For example: younger people – and those who have immunity from previous Covid infection – are less likely to be vaccinated than older people; older people are more likely to have comorbidities and more likely to have symptoms if they are infected. However, while those factors and relationships are widely considered in observational studies, most of the other factors in the model are not.

The first thing to note is that the model makes clear the critical distinction between whether a person is **Covid19 infected** (something which is not easily observable) and whether they are **classified as a Covid19 case** (i.e. the ones who are recorded as cases in any given study). The latter depends not just on whether they are genuinely infected but also on the accuracy of the testing and whether they are vaccinated. If (as in the Israel study described above) the unvaccinated are subject to more extensive (and potentially inaccurate) testing, then they are more likely to be erroneously classified as a case. The model also makes clear the critical distinction between those who have been vaccinated (at least once) and those classified as vaccinated in the study. The latter depends on the number of doses, time since last dose, and whether the person tests positive. Moreover, whether a person gets more than one dose will depend on whether they suffered an adverse reaction first time; those who do and who do not get a second dose are generally classified as unvaccinated - and this will compromise any studies of risk associated with the vaccine. Indeed, even the results of randomized controlled trials were compromised both by ‘removing’ those who died within 14 days of the second vaccination and ‘losing’ many subjects after the first dose<sup>6</sup>.

The causal model makes clear that a person cannot become infected with the virus unless they come into contact with it. The latter depends not just on age, ethnicity and profession (so young people who live, work and travel in crowded environments are more likely to come into contact with the virus as are any people in a hospital environment) but also on changing population factors like lockdown restrictions in place and current population infection rate. Assuming a person comes into contact with the virus, whether they get infected depends on whether they have natural immunity and whether they are vaccinated.

If we had relevant data on all of the factors in the model then, as in the case of the simple model in the Appendix, we can capture the probabilistic dependence between each node and its immediate parents, and then use Bayesian inference to determine the true effect of vaccination. In principle, this enables us to properly explain all observed data, adjust for all confounding factors, and provide truly accurate measures of effectiveness. The problem is that several key variables are generally unobservable directly while many of the easily observable variables are simply not recorded. While we can incorporate expert judgment with observed statistical data to populate the model, this can be extremely complex and subjective.

Moreover, if you think the model is already very complex, then it should be noted that it is far from fully comprehensive. Even before we consider all the additional factors and relationships needed to consider the outcomes of hospitalization and death (and the accuracy of reporting these), the model does not take account of: different treatments given; different morbidities and lifestyle choices; seasons over which data are collected; different strains of the virus; and many other factors. Nor does it account for the fact that all observational data are biased (or ‘censored’) in the sense that it only contains information on people who are available for the study; so, for example, studies in particular countries will largely contain people of a specific ethnicity, while all studies will generally exclude certain classes of people (such as the

---

<sup>6</sup> Some of the covid vax trials were unblinded, others were only single-blinded. Yet more were non-randomised and others were accidentally unblinded when the treatment recipients were given paracetamol prior to their covid jab

homeless). This means that, while such studies could be useful in determining effectiveness at a 'local' level, their conclusions are not generalizable. Indeed, they may be completely unreliable because of another paradox (called collider or Berkson's paradox) unless we have explicitly adjusted for this as described in (Fenton, 2020).

Given the impossibility of controlling for all these factors in randomized trials, and the overwhelming complexity of adjusting for them from observational data there is little we can reliably conclude from the data and studies so far. And we have not even mentioned the general failure of these studies to consider the impact and trade-offs of safety on effectiveness.

So, what can we do about this mess? We believe there is an extremely simple and objective solution: if we ignore the **cost** of vaccination, then ultimately we can all surely agree that the vaccine is effective overall if there are fewer deaths (from any cause) among the vaccinated than the unvaccinated. This combines both effectiveness and safety since it encapsulates the trade-off between them. It is not perfect, because there could be systemic differences in treatments given to vaccinated and unvaccinated<sup>7</sup>, but it completely bypasses the problem of classifying Covid19 'cases' which, as we have noted, compromises all studies so far.

So, provided that we can agree on an objective way to classify a person as vaccinated (and we propose that, for this purpose, the fairest way is to define anybody as vaccinated if they have received at least one dose), then all we need to do is compare all-cause mortality rates in different age categories of the vaccinated v unvaccinated over a period of several months<sup>8</sup>.

A recent analysis does indeed look at all-cause deaths in vaccinated and unvaccinated (Classen, 2021). The study shows that, for all three of the vaccines for which data were available, all-cause deaths is significantly higher in the vaccinated than the unvaccinated. However, this study did not account for age and hence its conclusions are also unreliable.

We could immediately evaluate the effectiveness to date of vaccines in the UK by simply looking at the registered deaths since the start of the vaccination programme in December 2020. All we need to know for each registered death is the person's age and whether they received at least one dose of the vaccine before death. Although a longer period would, of course, be better it is still sufficiently long to show a real effect if the vaccines work as claimed and if Covid19 is as deadly as claimed.

Moving forward we should certainly be collecting this simple data, but our concern is that (in many countries) the 'control group' (i.e. unvaccinated) may soon not be large enough for such a simple evaluation.

---

<sup>7</sup> There are multiple anecdotal reports that Australian hospitals are now giving ivermectin only to vaccinated patients

<sup>8</sup> <https://probabilityandlaw.blogspot.com/2021/06/why-all-studies-so-far-into-risks-andor.html>

## References

- CDC. (2021). COVID-19 Breakthrough Case Investigations and Reporting | CDC. Retrieved September 15, 2021, from <https://www.cdc.gov/vaccines/covid-19/health-departments/breakthrough-cases.html>
- Classen, B. (2021). US COVID-19 Vaccines Proven to Cause More Harm than Good Based on Pivotal Clinical Trial Data Analyzed Using the Proper Scientific Endpoint, "All Cause Severe Morbidity." *Trends in Internal Medicine*, 1(1), 1–6. Retrieved from <https://www.scivisionpub.com/pdfs/us-covid19-vaccines-proven-to-cause-more-harm-than-good-based-on-pivotal-clinical-trial-data-analyzed-using-the-proper-scientific--1811.pdf>
- Farinazzo, E., Ponis, G., Zelin, E., Errichetti, E., Stinco, G., Pinzani, C., ... Zalaudek, I. (2021). Cutaneous adverse reactions after mRNA COVID-19 vaccine: early reports from Northeast Italy. *Journal of the European Academy of Dermatology and Venereology*, 35(9), e548–e551. <https://doi.org/10.1111/jdv.17343>
- Fenton, N. (2020). Why most studies into COVID19 risk factors may be producing flawed conclusions - and how to fix the problem. *ArXiv*. <https://doi.org/http://arxiv.org/abs/2005.08608>
- Fenton, N. E., Neil, M., & Constantinou, A. (2019). *Simpson's Paradox and the implications for medical trials*. Retrieved from <http://arxiv.org/abs/1912.01422>
- Folegatti, P. M., Ewer, K. J., Aley, P. K., Angus, B., Becker, S., Belij-Rammerstorfer, S., ... Oxford COVID Vaccine Trial Group. (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet (London, England)*, 396(10249), 467–478. [https://doi.org/10.1016/S0140-6736\(20\)31604-4](https://doi.org/10.1016/S0140-6736(20)31604-4)
- Haas, E. J., Angulo, F. J., McLaughlin, J. M., Anis, E., Singer, S. R., Khan, F., ... Alroy-Preis, S. (2021). Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet (London, England)*, 397(10287), 1819–1829. [https://doi.org/10.1016/S0140-6736\(21\)00947-8](https://doi.org/10.1016/S0140-6736(21)00947-8)
- Krause, P. R., Fleming, T. R., Peto, R., Longini, I. M., Figueroa, J. P., Sterne, J. A. C., ... Henao-Restrepo, A.-M. (2021). Considerations in boosting COVID-19 vaccine immune responses. *The Lancet*, 0(0). [https://doi.org/10.1016/S0140-6736\(21\)02046-8](https://doi.org/10.1016/S0140-6736(21)02046-8)
- Ledford, H., Cyranoski, D., & Van Noorden, R. (2020). *The UK has approved a COVID vaccine — here's what scientists now want to know*. Retrieved from [https://www.nature.com/articles/d41586-020-03441-8?utm\\_source=Nature+Briefing&utm\\_campaign=597ee8dba8-briefing-dy-20201203&utm\\_medium](https://www.nature.com/articles/d41586-020-03441-8?utm_source=Nature+Briefing&utm_campaign=597ee8dba8-briefing-dy-20201203&utm_medium)
- Mclachlan, S., Osman, M., Dube, K., Chiketero, P., Choi, Y., & Fenton, N. (2021). *Analysis of COVID-19 vaccine death reports from the Vaccine Adverse Events Reporting System (VAERS) Database Interim: Results and Analysis*. Retrieved from <http://dx.doi.org/10.13140/RG.2.2.26987.26402>
- Pearl, J., & Mackenzie, D. (2018). *The book of why : the new science of cause and effect*. New York: Basic Books.
- Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., ... C4591001 Clinical Trial Group. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *The New England Journal of Medicine*, 383(27), 2603–2615. <https://doi.org/10.1056/NEJMoa2034577>
- Scheifele, D. W., Bjornson, G., & Johnston, J. (1990). Evaluation of adverse events after influenza vaccination in hospital personnel. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne*, 142(2), 127–130. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2295029>
- Singh, J. A., Kochhar, S., Wolff, J., Atuire, C., Bhan, A., Emanuel, E., ... Upshur, R. E. G. (2021). Placebo use and unblinding in COVID-19 vaccine trials: recommendations of a WHO Expert Working Group. *Nature Medicine*, 27(4), 569–570. <https://doi.org/10.1038/s41591-021-01299-5>
- Stone, C. A., Rukasin, C. R. F., Beachkofsky, T. M., Phillips, E. J., & Phillips, E. J. (2019). Immune-mediated adverse reactions to vaccines. *British Journal of Clinical Pharmacology*, 85(12), 2694–2706. <https://doi.org/10.1111/bcp.14112>

## Appendix

The prior probabilities based on the study data are shown in Figure 4

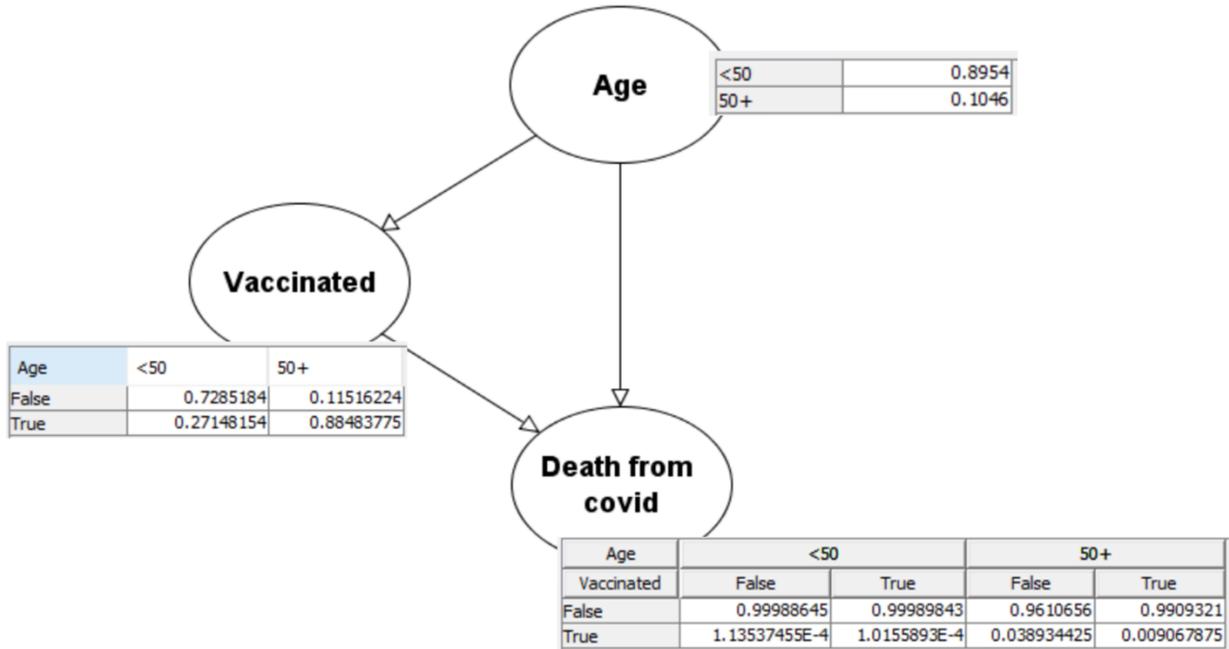


Figure 4 Causal model as a Bayesian network with probability tables taken from the observed data

This results in the so-called marginal probabilities shown in Figure 5.

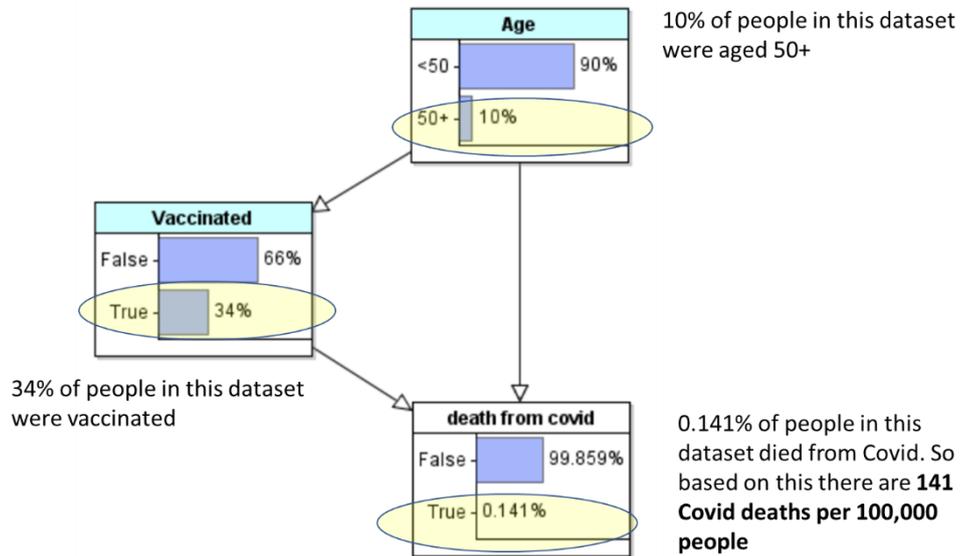
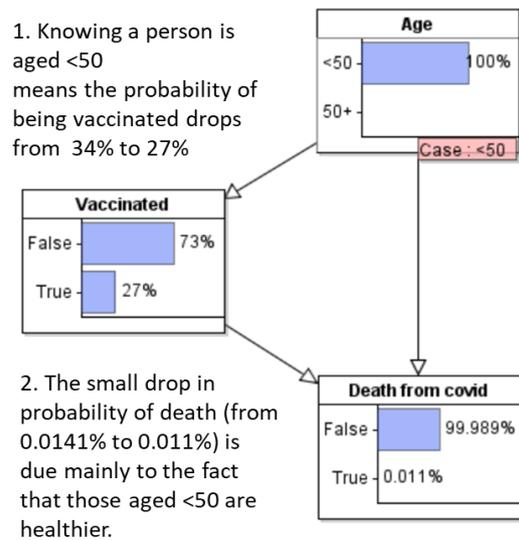
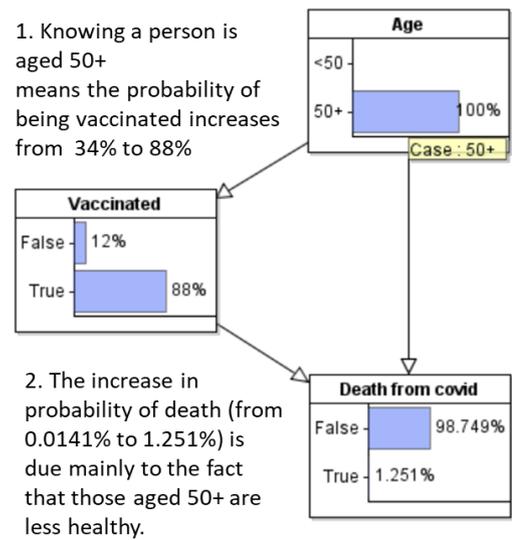


Figure 5 Marginal probabilities (Age and Vaccinated are rounded to 0 decimal places)

By entering observations on Age we can see the overall effect on probability of vaccinated and death as shown in Figure 6.



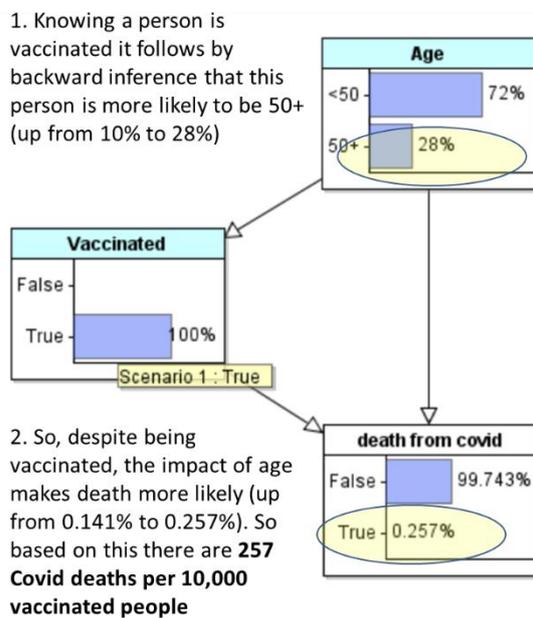
(a) Age < 50



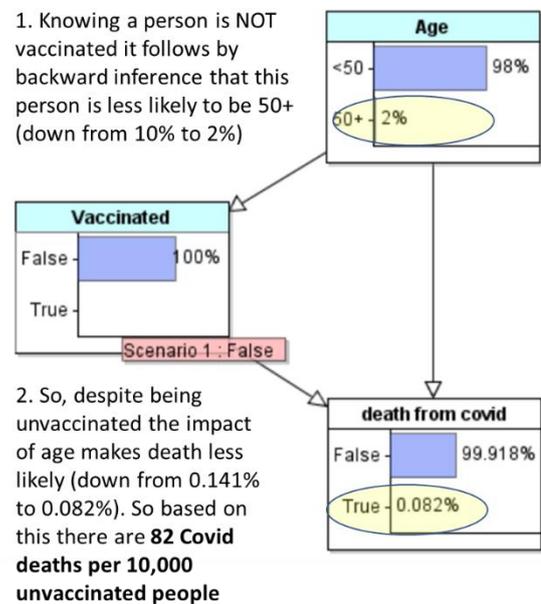
(b) Age 50+

Figure 6 Overall impact of age on probability of vaccinated and death

However, the real power of the Bayesian network comes in the backward inference shown in Figure 7 that enables us to determine the impact of vaccination status on age as well as death.



(a) Vaccinated



(b) Unvaccinated

Figure 7 Impact of vaccination status on age and death

Here we see (as noted in the original data) that the vaccinated are four times more likely to die than the unvaccinated. But this is explained by the vaccinated being much more likely to be 50+

Next we use the model to see the impact of vaccination respectively on those aged <50 and those aged 50+. In Figure 8 we see that, for those aged <50 there is a small decrease on probability of death among the vaccinated.

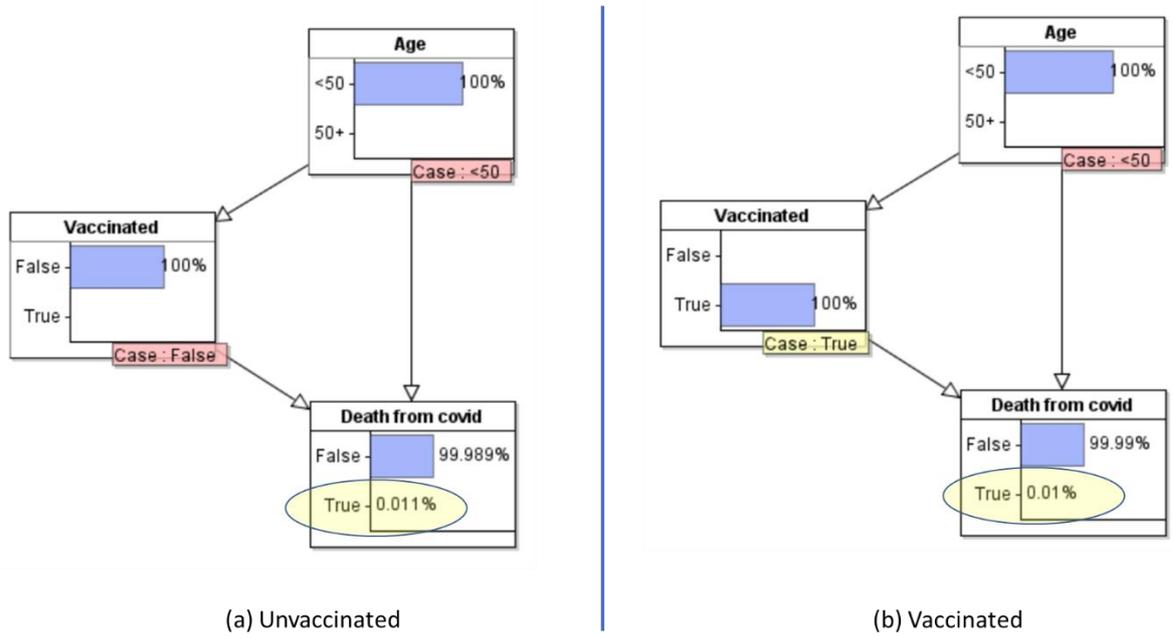


Figure 8 Impact of vaccination on those age <50: small decrease in probability of death for vaccinated

In Figure 9 we see that, for those aged 50+ there is a large decrease on probability of death among the vaccinated.

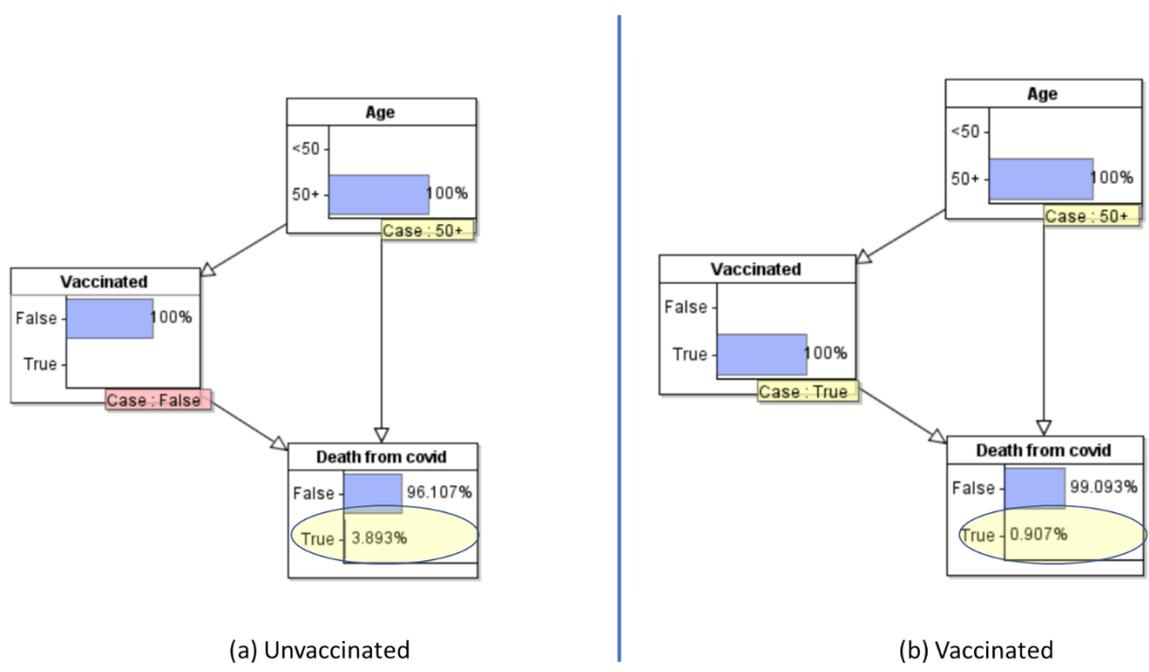


Figure 9 Impact of vaccination on those age 50+: large decrease in probability of death for vaccinated

So, in contrast to Figure 7 which suggested vaccination increased the probability of death overall, we now see that in each age category vaccination decreases the probability of death.

In order to avoid the confounding effect of age we can simulate the effect of the vaccination 'intervention' by simply removing the dependency from 'age' to 'vaccinated' as shown in Figure 10.

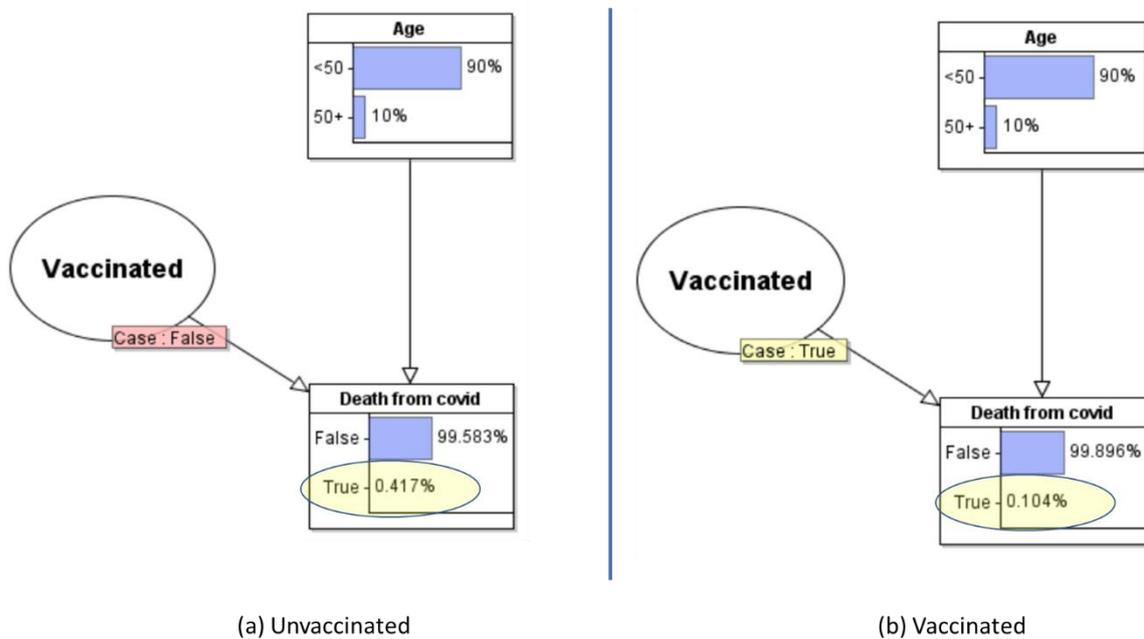


Figure 10 Unconfounded impact of vaccination: probability of death decreases from 0.417% to 0.104%

So there is a decrease in probability of death from 0.417% in the unvaccinated to 0.104% in the vaccinated. This means that (using the notion of 'relative risk') the unvaccinated are four times more likely to die than the vaccinated. However, in absolute terms, the unvaccinated have a 0.313% greater probability of dying.