Review of: "Measuring the efficacy of a vaccine during an epidemic"

Stephen Senn

Potential competing interests: No potential competing interests to declare.

This is a useful paper that has a closer look at a common measure of the efficacy of a vaccine, to see what needs to be assumed to make fair comparisons. I learned a lot from the paper and there is much with which I am in agreement. There is little point in rehearsing these points. Instead I shall cover some points where I am not in complete agreement.

The first concerns representativeness. The authors state, 'The distribution of the traits (age, census, medical history, etc) of both cohorts must be representative of the whole population' (p2) but experiments don't work like this. Neither random nor any form of representative sampling is involved in clinical trials and indeed it is usually impossible. Instead, the inclusion and exclusion criteria are meant to ensure that patients recruited a) may ethically be offered any of the treatments being compared and b) are presumed capable in principle of showing the effect of treatment. Even observational studies carried out by epidemiologists are rarely intended to be representative (the exception would be health surveys) and that this is so is not usually taken to be a problem. See Rothman et al for a discussion[1]. Instead what is necessary for clinical trials is that the extent to which groups differ from each other, as regards the effect this has on the outcome measure, can be assumed to follow a known random form permitting valid estimates of uncertainty[2]. This is not achieved because the groups are a representative sample of the target population; they are not and could not be. It is achieved because a known random mechanism has allocated (possibly unrepresentative) subjects fairly to treatment.

The authors make the important point that independence can be harder to obtain in studies in infectious epidemiology. However, I am not so optimistic as they are that the difficulties can be avoided. In a double-blind trial it is impossible for results to cluster by treatment except by the degree governed by the allocation algorithm and this is reflected in analysis. For example, in a cluster randomised trial all the patients in one centre will have the same treatment and this centre becomes the primary unit of inference. In a parallel group trial patients may generally be used as the unit of inference. Note, however, that results not only reflect what has actually happened to the patients but also the extent to which the measurement process *reflects* what happened. Independence of this also is guaranteed in a double-blind trial parallel group trial, things could not be organised to have all the treatment group laboratory values processed in the same batch (to given an example). The treatment groups are randomly dispersed among the patients and since the treatment cannot be identified, such grouping is impossible. However, as soon as the trial is open, such guarantees do not apply and it may require careful oversight to make sure that unwanted clustering of procedures does not occur. In a paper of mine on vaccine efficacy[3], I considered six different ways in which clustering could arise in an open study and I suspect that there are many more.

Of five large trials of COVID vaccine efficacy, two, those sponsored by AZ/Oxford University and J&J, and by were double-blind, whereas those of Moderna, Novavax and Pfizer/BioNTech were observer blind. Another important difference was that the J&J vaccine was delivered in a single dose but the others in two. The view might be taken that one should compare the treatment regimens as defined but there are two reasons to be cautious about this. The first is that it is not obvious what time zero should be when counting follow up. Three of the trials started counting about two weeks after the second dose and one (Pfizer/BioNTech) one week after the second dose but J&J started two weeks after the first dose. Personally, although I have some sympathy with a measurement that respects the mechanism of the vaccine, I think that there is much to be said for an intention to treat philosophy of measuring time from the moment of randomisation. The second is that as the pandemic progressed countries moved to using single doses of other vaccine as boosters. Thus, persons who were originally given an AZ/Oxford vaccine in two doses were, in some cases, subsequently given a Moderna 'booster' as a single shot. In other words, judgements were being made about the efficacy of vaccines that when well beyond that to which a per protocol analysis would apply.

Follow up was also different in the five trials and this is where the sort of modelling Cavallo and Scala consider can help in understanding potential problems. Of course, one should not neglect the value of data and there have been a number of studies trying to compare the vaccines fairly. A recent example is given by Hulme et al[4].

As a minor point of style I suggest that the authors replace the words 'kids' with 'children'.

References

1. ^*K. J. Rothman, J. E. Gallacher, E. E. Hatch. (2013).Why representativeness should be avoided. International Journal of Epidemiology, vol. 42 (4), 1012-1014. doi:10.1093/ije/dys223.*
2. ^*Stephen Senn. (2012). Seven myths of randomisation in clinical trials.Statist. Med., vol. 32 (9), 1439-1450. doi:10.1002/sim.5713.*
3. ^*Stephen Senn. (2022). The design and analysis of vaccine trials for<scp>COVID</scp>-19 for the purpose of estimating efficacy. Pharmaceutical Statistics, vol. 21 (4), 790-807. doi:10.1002/pst.2226.*
4. ^*William J Hulme, Elsie M F Horne, Edward P K Parker, Ruth H Keogh, et al. (2023).Comparative effectiveness of BNT162b2 versus mRNA-1273 covid-19 vaccine boosting in England: matched cohort study in OpenSAFELY-TPP. BMJ. doi:10.1136/bmj-2022-072808.*