Review of "Smoking, vaping and hospitalization for COVID-19"

Carl Phillips

These statistics are the epitome of "huge if true". These results are either breathtakingly important (suggesting smoking is practically a vaccine against bad COVID-19 outcomes) or completely worthless (because the validity of the data is so suspect). There is no in-between. The authors try to create an in-between, and this is not valid.

(The mandatory star rating required in order to publish this review creates quite the irony on this point. This paper is either a 5-star discovery or a 1-star misleading report on data that is so flawed that it should never be analyzed. Forced to provide a scalar as I sit here now, I split the difference and gave it 3.)

I have written a full review of this paper and posted it (along with some brief related observations) here: https://antithrlies.com/2020/04/04/can-smoking-protect-you-against-covid-19/

To summarize the main points:

-Because this is so huge if true, the normal poor standards of epidemiology research are not good enough.

-The authors should make clear what their analysis really implies (if the data is taken at face value): Smoking is *hugely* protective against bad COVID-19 outcomes. They either need to say this or declare that their data is so uncertain that they can conclude nothing. It is not possible to draw the conclusion that the data support the claim of a null relationship, as is implied (not quite explicitly stated) in the text. This is a misinterpretation of the results.

-The authors need to explicitly point out that the quality of the data is highly suspect, as a result of unknown but conceivably fatal levels of exposure misclassification and and odd form of selection bias. They should investigate further to the extent possible (searching out additional information about Chinese hospitalization numbers, reviewing what has

been written about the Chinese data by Chinese and knowledgable Westerners, assessing the quality of these medical records, etc.), and attempt to quantify the potential bias in the results. Having done so, they should decide either that they trust the data to be informative, and report what their calculations show (with whatever quantified adjustment they determine is optimal) and what that implies, or decide that they cannot do this and report that they think it is impossible to learn something from this data. It epistemically and logically invalid to say, in effect, "we do not trust our data to support an affirmative conclusion, so it therefore supports the null conclusion."

-Potential confounding is a trivial issue compared to the data quality uncertainty, and should basically be ignored. More important, the claim that hypothesized detrimental effects of smoking on COVID-19 outcomes, via various pathways, attenuate the implications of these results is flatly wrong. These statistics (again, if the data is valid) show the *net* effect of smoking on COVID-19 outcomes. Any detrimental effect is already captured in them, not something that should be subtracted from them. If there are detrimental pathways, via accumulated disease burden or whatever, then the results suggest that current smoking is even *more* protective against COVID-19 outcomes (though it is a bit more complicated than that, with other possible combinations of effects). The authors get this backward.

-The averaging together of the various datasets is not appropriate. There is inadequate justification for the necessary assumption that they are all measures of the same phenomenon (same population, same methodology, etc.).