The Bangladesh Mask study: a Bayesian perspective

Preprint · May 2022

DOI: 10.13140/RG.2.2.26189.92649

1 author:

Norman Elliott Fenton
Queen Mary, University of London
412 PUBLICATIONS **16,977** CITATIONS

SEE PROFILE

The Bangladesh Mask study: a Bayesian perspective

Norman Fenton[1]

2 May 2022

Abstract

A very large trial, whose results were published in Science, carried out in Bangladesh between 2020 and 2021 has been widely acclaimed as providing the most convincing evidence yet that masks work in reducing Covid-19 transmission and infections. However, the media grossly exaggerated the authors' own conclusions, and sceptical researchers have identified weaknesses in various aspects of the trial and statistical analysis which cast doubts on the significance of the results. The sole focus of this report is to determine what can really be learned about the impact of mask wearing on covid infections from the data in the trial. Using a novel Bayesian causal modelling approach, we find that the claimed benefits do not hold up when subject to this rigorous analysis. At best, one can conclude that there is only a 52% probability that the seropositivity rate among people subject to a mask intervention campaign is lower than those who are not, while there is a 95% chance that a mask intervention campaign would result in anything between 19,240 fewer positives and 18,500 more positives in every 100,000. This means there was no discernible effect of the mask intervention on covid infection. Given that the results of the study have been used explicitly to justify continuing or reintroducing aspects of mask mandates in the USA, UK and elsewhere, the study paper in Science needs to be corrected or withdrawn.

1. Introduction

What has been claimed to be the largest randomized controlled trial to determine the effectiveness of masks in preventing spread of Covid-19 was carried out in rural Bangladesh between November 2020 and April 2021. The trial and its results were first reported in a preprint [1] and subsequently published in Science [2].

In contrast to the only previous randomized controlled trial (in Denmark in 2020) [3] which found no statistically significant benefits of mask wearing in reducing covid transmission or infection, the Bangladesh trial has been widely acclaimed as providing evidence that masks work [4][5][6]. The reporters who trumpeted the 'success of the study' are unlikely to have understood, or even read, the overly complex and often opaque statistical results contained in the original 94-page report. Yet, they were more than happy to parrot the paper summary which states:

> *A randomized-trial of community-level mask promotion in rural Bangladesh during COVID-19 shows that the intervention tripled mask usage and reduced symptomatic SARS-CoV-2 infections, demonstrating that promoting community mask-wearing can improve public health.*

However, sceptical researchers have pointed out multiple weaknesses in the study design (including the curious distinction between different mask types and colours), flaws in the statistical analysis, and how claims by the media grossly exaggerate the authors' own conclusions [7] [8] and [9].

First, it is important to note that the trial was not (as implied in the media reporting) a randomized controlled trial of 340,000 people but was rather a 'cluster randomized' trial of 300 'treatment villages' and 300 'control villages'; in the former there was a mask wearing intervention campaign,

[1] n.fenton@qmul.ac.uk, Queen Mary University of London. Declaration of Interest: The author is also a Director of Agena Ltd whose software is used in the analysis contained in this paper.

while in the latter where there was no intervention. It was the total population of these villages that numbered some 340,000.

If the primary objective of the trial was to determine whether a mask intervention policy led to an increase in mask wearing, then the cluster randomized design makes sense, and indeed there is evidence the mask intervention policy achieved significant success with respect to that objective. But such a result is neither interesting nor useful. We could surely also triple the amount of sweets children ate if we gave them out for free. In fact, the primary objective of the trial was to determine whether mask wearing leads to a reduction in covid infections. While the authors of the study claim the design was well suited to test this *at the community level*, the results have been widely interpreted as demonstrating that mask wearing reduces the risk of covid at the individual level. Indeed the grandiose summary statement above that the intervention "… reduced symptomatic SARS-CoV-2 infections, demonstrating that promoting community mask-wearing can improve public health" confirms this impression. The sole focus of this report is to determine what can really be learned about the impact of mask wearing on covid infections from the data in the trial using a novel Bayesian causal modelling approach [10].

In Section 2 we describe the Bayesian approach and in Section 3 we show that the trial data cannot be used to provide any conclusions at all about whether mask wearing reduces covid infections. We show in Section 4 that, at best, the trial data can help evaluate a weak surrogate hypothesis that people subject to a mask intervention campaign are less likely to test seropositive than those who are not. We also explain the impact on the conclusions of properly accounting for the limited testing and the correlation of outcomes from using clustering of villages. In Section 5 we provide what we believe are the most meaningful results that can be concluded from the trial data: taking full account of the uncertainty inherent in the study, at best one can conclude that there is a 52% probability that the seropositivity rate among people subject to a mask intervention campaign is lower than those who are not (with a 95% risk ratio confidence interval of 0.14 to 6.35). The probability of 52% is way below any usually acceptable levels of significance (these are typically set at 95% or 99%) – it means that there is a 48% probability that the seropositivity rate among those subject to a mask intervention campaign is ***higher*** than those who are not. The risk ratio confidence interval should lie entirely below 1 if there is 'significant' evidence the mask intervention campaign worked. The fact that it ranges from 0.14 (which would mean the seropositivity rate is 7.14 times higher among those not subject to the mask intervention) up to 6.35 (which would mean the seropositivity rate is 6.35 times higher among those subject to the mask intervention) means there is almost no support at all for even the weak surrogate hypothesis about mask interventions.

2. The Bayesian method

To test what we will call the ***ultimate hypothesis*** that mask wearing reduces covid infections, we need to answer the question:

> Is the covid infection rate among mask wearers 'significantly' lower than the rate among non mask wearers?

In a properly constructed randomized controlled trial we would have two approximately equal sized groups of people:

1. Group 1 containing only non mask wearers (the 'control' group)
2. Group 2 containing only mask wearers (the 'treatment' group)

After a suitable period, we would observe for each participant whether they had contracted covid.

To test the 'ultimate hypothesis' using the Bayesian approach, we consider the (unknown) covid infection rate for each of the masked and unmasked population people to be a probability distribution which we learn (using Bayesian inference) from the total number of masked and unmasked respectively in the study and the number of these who get covid.

A superficial look at the key Bangladesh study data (shown in Appendix 1) might lead to the assumption that there were 161211 participants in Group 1 (i.e. non mask wearers) of whom 1106 were known to have got covid. Then, based on standard prior assumptions and Bayesian inference[2] (shown in Appendix 2 Figure 1) we would conclude that the revised probability *p* of getting covid among non mask wearers is a distribution whose mean is 0.687% and whose 95% confidence interval ranges from 0.647% to 0.728%. This means there is a 95% probability that *p* lies between 0.647% to 0.728%.

Similarly, we might assume from the data that there are 174171 participants in Group 2 (mask wearers) and that 1086 of these got Covid. Then we would conclude that the revised probability of getting Covid among mask wearers is a distribution whose mean is 0.624% and whose 95% confidence interval ranges from 0.588% to 0.662%. This is shown graphically in Appendix 2 Figure 2. There is little overlap between the distributions, suggesting intuitively that this provides strong support for the ultimate hypothesis above.

In order to interpret exactly what this means, we use a slightly more complex Bayesian network model (Appendix 2 Figure 3) that separates the two distributions and calculates their Risk Ratio (RR) - defined as $p1$ (probability of masked getting Covid) divided by $p2$ (probability of unmasked getting Covid) and the probability that $p1$ is less than $p2$.

This model tells us that:

- There is a 95% probability that the RR lies between 0.84 and 0.99 (this is the 95% RR CI)[3]
- The probability that p1 < p2 is 98.69%
- The probability that the treatment reduces the infection rate by at least 10% (i.e. the probability that p1< 0.9*p2) is 40.9%

We summarise this information in Table 1 (and use the same format for all subsequent analyses).

Table 1 Results using 'hypothetical data'

	Population	**Covid infected**	**Mean rate**	**95% CI**	**RR CI**	**Prob p1 < p2**
Masked	174171	1086	0.624%	0.588 to 0.662 %		
Unmasked	161211	1106	0.687%	0.647 to 0.728 %		
					0.84 to 0.99	98.68% **Prob p1<0.9*p2:** (40.9%)

So, with this assumed data, there seems to be quite strong evidence (98.68% probability) for the ultimate hypothesis that the covid infection rate of the masked is lower than that of the unmasked. However, it is unlikely (40.9% probability) that the reduction will be more than 10% (and the probability of a more than 20% reduction is just 0.18%). With the Bayesian approach we do not use

[2] All of the Bayesian inference is performed using the AgenaRisk 10 (revision 8607) software with simulation convergence setting 0.0001. Links to access all the models and software are provided at the end of this report.
[3] As shown in Appendix 2 Figure 3, the Risk Ratio confidence interval is simply the 5% and 95% percentile values of the probability distribution computed for the risk ratio node in the model.

p-values for significance as we have actual probabilities associated with our hypotheses. If we wanted to be at least 99% certain of the ultimate hypothesis before declaring the result 'significant' then we have just missed the threshold, but we are well clear if we set a 95% threshold. Of course, even with this 'significance' the absolute risk reduction is small: for every 100,000 masked people we might expect about 624 to get covid compared to 687 out of every 100,000 unmasked. That is an absolute risk reduction of 0.00063, i.e. 63 in 100,000.

Moreover, because of possible confounders we still cannot conclude that it was mask wearing that led to this reduction. Nor can we conclude that rates of more serious outcomes (hospitalisation and death) are lower in the masked since we do not have the data for that.

3. The problems with Bangladesh study data

In the study there were NOT two randomly selected groups of mask wearers and non mask wearers. Rather, the study was based on 600 villages divided randomly into 300 villages whose 174171 people who were 'reached for symptom collection' were defined as the treatment group, and 300 villages whose 161211 people who were 'reached for symptom collection' were defined as the control group (see Appendix Table 1A reproduced from the paper). The treatment group villages received free masks, and various types of intervention to encourage mask-wearing, while the control group villages received none of that. There was, of course, no guarantee that the inhabitants of the treatment villages would wear masks nor that those of the control villages would not. This means that:

> *We do not know how many of the 174171 treatment group participants were really mask wearers nor how many of the 161211 control group participants were really non mask wearers.* ***Hence, the numbers 174171 and 161211 represent, respectively, simply crude surrogates for the number of 'masked' and 'unmasked'.***

There are even more complications when it comes to the numbers infected with covid. The 1086 in the treatment group and 1106 in the control group are the number of people in each of the treatment and control villages respectively who satisfied all the following criteria:

a) self-reported covid-like symptoms; (of whom there were 13,273 in the treatment group and 13,893 in the control group)
b) subsequently agreed to have their blood tested (which narrowed the numbers down to 5006 in the treatment group and 4971 in the control group); and
c) their blood subsequently tested seropositive (which narrowed the numbers down to 1086 in the treatment group and 1106 in the control group)

(it is important to note that, while the other numbers were reported in the paper, the numbers 1086 and 1106 were – curiously - not reported in the paper but had to be inferred, as explained in Appendix 1).

What this means is that:

> *We do not know how many of the participants in each group really contracted covid.* ***Hence, the numbers 1086 and 1106 represent respectively simply crude surrogates for the numbers in each group who contracted covid.***

To see how far these surrogates are from the true information we need, note the following for the treatment group (a similar set of problems apply to the control group):

- The number who actually were masked is an ***unknown proportion*** of the total number of 174171 people in the treatment villages.
- The number testing positive is NOT the number of masked with Covid, but rather the sum of the masked ***and unmasked*** in the treatment villages who first had to report feeling covid-like symptoms, then had to agree to their blood being tested, and then had their blood test seropositive. This test is not a perfect test of a person with covid. Hence, the number testing positive will include some masked and unmasked people who did not actually have covid; and it will wrongly exclude some masked who had covid. And, of course, none of those who were masked and who had covid but were not tested are included in the number testing positive.

So, if we want to learn the probability of masked people getting (symptomatic)[4] covid from the available data then we need to run the full causal Bayesian network model shown in Appendix 2 Figure 4. Because there are so many variables for which there are no observations available, when we run such a model with the limited observed data the posterior distribution for the probability of masked people getting (symptomatic) covid has such a wide 95% confidence interval that it essentially tells us nothing; it is very similar to the posterior probability distribution for unmasked people getting (symptomatic) covid obtained from the equivalent model for the control group.

4. So what can we infer from the available data?

To be able to get any kind of meaningful comparison between the control and treatment groups in the absence of data for all but the orange nodes on Appendix 2 Figure 4, we could attempt to answer the question:

> Is the seropositive rate among people subject to mask intervention procedures 'significantly' lower than the rate among those receiving no intervention?

This would enable us to test the (weak) ***surrogate hypothesis*** that the mask intervention procedures reduce the seropositivity rate.

Now, if it were the case that EVERY participant had been tested and that the number recorded in each group testing seropositive were the numbers observed (i.e. 1106 in the control group and 1086 in the treatment group) then we could easily test the surrogate hypothesis. In fact, the relevant results would be exactly that provided in Section 2, Table 1 (with 'control' replacing 'unmasked', 'treatment' replacing 'masked', and 'testing seropositive' replacing 'covid infected'. Hence, we would have the results shown in Table 2.

Table 2 Results using 'surrogate data' if every participant had been tested

	Population	**Seropositive**	**Mean rate**	**95% CI**	**RR CI**	**Prob p1 < p2**
Treatment	174171	1086	0.624%	0.588 to 0.662 %		
Control	161211	1106	0.687%	0.647 to 0.727 %		
					0.84 to 0.99	98.68% **Prob p1<0.9*p2:** (40.9%)

[4] For simplicity we restrict the interest to symptomatic covid henceforth as the model would double in size if we also wanted to learn the probability of getting covid without symptoms

But, of course, it is NOT the case that every participant was tested. The only ones who were tested were those who both self-reported having covid-like symptoms and who also subsequently agreed to have their blood tested. We will address this issue in the next section, but let us continue with the charade that the number of seropositives is really based on the assumption that everybody was tested.

It turns out that even then we cannot use the raw data presented because the study used cluster-randomization (whole villages rather than individuals). The clustering is problematic because Covid is an infectious disease; it means we cannot consider all participants to be independent because if a person is infected with Covid then it is likely many of those in the village in contact with that person will also be infected. This means the outcomes are likely to be correlated inside a village. As reported by Recht [11]:

> *To capture the correlation among intra-cluster participants, statisticians use the notion of the intra-cluster correlation coefficient ρ. ρ is a scalar between 0 and 1 that measures the relative variance within clusters and between clusters. When ρ=1, all of the responses in each cluster are identical. When ρ=0, the clustering has no effect, and we can treat our assignment as purely randomized. Once we know ρ we can compute an effective sample size: if the villages are completely correlated, the number of samples in the study would be 600. If they were independent, the number of samples would be over 340,000.*

Recht explains why a value of ρ=0.007 is reasonable for the Bangladesh study and that this leads to a 'design effect' of about 5 which means that all the observations (i.e. number of participants and number who are seropositive) must be reduced by a factor of 5 in order to remove the bias from the correlations within villages.

This means that, to take account of the correlation among intra-cluster participants, the relevant revised data we should use is the following

- Control: 221 from 32242 (rather than 1106 from 161211)
- Treatment: 217 from 34834 (rather than 1086 from 174171)

When we run the basic Bayesian model with these revised observations, we get the results shown in Table 3.

Table 3 Results using adjustment for intra-cluster correlation assuming all participants had been tested

	Population	**Seropositive**	**Mean rate**	**95% CI**	**RR CI**	**Prob p1 < p2**
Treatment	34834	217	0.626%	0.545 to 0.711%		
Control	32242	221	0.688%	0.601 to 0.782%		
					0.753 to 1.096	84.16% **Prob p1<0.9*p2:** (38.38%)

Because there are 'less data' to learn from, the results of the Bayesian analysis show that there is now much more uncertainty about whether $p1$ (the probability of seropositivity in the treatment group) is less than $p2$ (the probability of seropositivity in the control group). The probability $p1<p2$ is 84.16%.

So, even if the data were based on everybody having been tested (which they were not), even the results of the surrogate hypothesis would not be considered 'statistically significant' under any normal interpretation. However, it is interesting to note that in an interview [12] with James Lyons-

Weiler the first author of the study Dr Jason Abaluck discusses how they achieved what they believed was 'statistical significance' using a method called 'imputation' (starting at 49:45)[5].
To explain imputation imagine you set up a trial to test if an intervention decreases positivity. You get 10,000 people in the control group and 10,000 in the treatment group. But only 1000 in each group agree to the outcome test. If 100 people in the control group test positive and only 85 in the treatment group the result is certainly not significant (as shown in Appendix 2 Figure 5(a)). With the method of imputation described by Abaluck, we assume that for each group those who refused to get tested would have the same positive rate as those who did get tested. Hence, we assume that 1000 out of 10,000 in the control group test positive and 850 out of 10,000 in the treatment group test positive. That would produce a highly significant result as shown in Appendix 2 Figure 5(b). But the result is bogus since it relies on additional data that are purely imaginary.

Now, we know that only 5006 out of the 13,273 (i.e. 37.7%) of the treatment participants who reported Covid-19 symptoms were tested and that only 4971 out of the 13,893 (i.e. 35.8%) of the control participants who reported Covid-19 symptoms were tested. So, applying the imputation method we would assume that, if all those with symptoms had been tested, then 2879 (instead of 1086) in the treatment group and 3091 (instead of 1106) in the control group would have tested positive. Applying the adjustment for intra-cluster correlation means we would assume:

- Control: 618 (instead of 221) from 32242
- Treatment: 576 (instead of 217) from 34834

When we use these data in the basic model, we get the results shown in Table 4.

Table 4 Impact of imputation on the data adjusted for intra-cluster correlation (assuming everybody was tested)

	Population	**Seropositive**	**Mean rate**	**95% CI**	**RR CI**	**Prob p1<p2**
Treatment	34834	576	1.666%	1.525 to 1.792%		
Control	32242	618	1.92%	1.773 to 2.073%		
					0.770 to 0.966	99.47% **Prob p1<0.9*p2:** (76.86%)

So, with imputation, we increase the probability that $p1<p2$ from an ('insignificant') 84.16% to a 'significant' 99.47%. But, as explained, this is a bogus method introduced to artificially increase significance.

5. Addressing the problem that not everybody was tested

Recall that, to evaluate the surrogate hypothesis that the mask intervention procedures reduce the seropositivity rate we need to answer the following question using the available data:

> Is the seropositive rate among people subject to mask intervention procedures 'significantly' lower than the rate among those receiving no intervention?

To be able to use the available data to infer the probabilities of testing seropositive for people subject to mask intervention procedures and those not, respectively, we need to use (for each) the Bayesian network model shown in Appendix 2 Figure 6 (and note that even this model makes the simplifying assumption that no people without symptoms wrongly report symptoms).

[5] In email correspondence, Dr Abaluck has since stated that "there was no imputation for our primary outcome (that was an auxiliary robustness check reported in an appendix)".

It turns out that, to get results that in any way clearly distinguish between the probabilities of seropositivity in those subject to mask interventions and those not, we have to make some strong prior assumptions without any evidence base to do so. Specifically, we have to make strong prior assumptions about the unobserved variables *probability of reporting if with symptoms* and *probability of getting covid symptoms.* For example, assuming that the former is a Truncated normal with mean 0.5 and variance 0.005 and the latter is a triangle(0, 0.05, 1) distribution, and without the adjustment for cluster correlation we get the posterior distributions shown in Appendix 2 Figure 7. Even with these strong prior assumptions and without the cluster correlation adjustment the results still provide only very weak support for the surrogate hypothesis as shown in Table 5 (in this section we show median rather than mean values of the distributions as they are heavily skewed).

Table 5 Results given strong priors and no adjustment for cluster correlation

	Median seropositivity rate	**95% CI**	**RR CI**	**Prob p1<p2**
Treatment	3.37%	2.60% to 4.74%		
Control	3.90%	2.99% to 5.50%		
			0.563 to 1.327	75.53%

Assuming, uniform priors for the unobserved nodes *probability of reporting if with symptoms; probability of getting covid symptoms;* and *probability of being tested if symptoms reported* and the figures in Table 1A for the observed orange nodes, we get the results shown in Table 6.

Table 6 Uniform prior (and data not adjusted for cluster correlation)

	Median seropositivity rate	**95% CI**	**RR CI**	**Prob p1<p2**
Treatment	5.96%	1.77% to 20.40%		
Control	6.53%	2.05% to 20.91%		
			0.130 to 6.484	53.48%

But applying the cluster correlation factor (i.e. dividing the numbers by 5) we get the results shown in Table 7.

Table 7 Uniform prior using adjustment for cluster correlation

	Median seropositivity rate	**95% CI**	**RR CI**	**Prob p1<p2**
Treatment	6.03%	1.86% to 20.54%		
Control	6.71%	2.04% to 21.10%		
			0.140 to 6.35	52.25%

The Table 7 results are based on the most reasonable assumptions that can be made for this trial data. Based on the 95% confidence intervals for the seropositivity distributions, all we can conclude is that there is a 95% chance that mask intervention would result in anything between 19,240 fewer positives and 18,500 MORE positives among every 100,000 people. The results therefore provide essentially no support even for the weak surrogate hypothesis that the mask intervention procedures reduce the seropositivity rate.

6. Discussion and Summary

The Bangladesh study, when viewed from a Bayesian perspective, does not provide the necessary data to enable us to test the hypothesis that mask wearing reduces the probability of covid infection. It does, however, provide some limited data to test the surrogate hypothesis that the mask intervention procedures reduce the seropositivity rate.

Before discussing the testing of this hypothesis, it is worth questioning whether in fact the primary endpoint chosen (seropositivity reduction) is either clinically meaningful or epidemiologically desirable. In the absence of vaccines which meaningfully or at all reduce infection (which appears to be the reality), exposure to the virus is necessary to gain the quality of immunity which prevents transmission and contributes to population-level immunity to the level which converts the pandemic into endemic equilibrium, thereby minimising the danger for the most vulnerable members of society. Hence public health officials are prone to express satisfaction, rather than concern, with rising antibody levels, as they recognise that the higher these are the closer we are to the end of the pandemic as a significant threat to public health.

Regardless, this paper aims to examine whether the hypothesis based on an endpoint of seropositivity reduction has been proven, and therefore we shall proceed on that basis.

When we take account of the limited testing that was performed and the cluster correlation, we conclude that:

> ***The probability the seropositivity rate is lower in people receiving the mask intervention than those who do not is 52% with risk ratio of 0.14 to 6.35.***

In other words, there is no real statistical support at all because the probability distributions for the treatment and control populations have such wide 95% confidence interval bounds that they are almost indistinguishable. There is a 95% chance that a mask intervention campaign would result in anything between 19,240 fewer positives and 18,500 more positives in every 100,000 people.

To give a feel for just how 'insignificant' the 52% figure is - if you wanted to use it to conclude that the seropositivity rate is lower in people receiving the mask intervention than those who do not - then this would be much like flipping 201 coins, observing 101 'heads' and 100 'tails' and concluding that all coins are more likely to land on heads than tails (assuming a uniform prior for the probability p of heads the probability that p is greater 0.5 is 52.6% in this case).

Moreover, there are other factors which, if properly accounted for, could lead to even greater uncertainty – and possibly even to a higher seropositivity rate in the treatment population. For example:

- The fact that the data are based only on testing people who report symptoms introduces a possible bias: Whether or not a person believes they have 'covid symptoms' is extremely subjective when the symptoms are minor. If two people – a mask wearer and a non mask wearer - have very similar minor symptoms, then intuitively it seems it is less likely that the mask wearer will report having covid symptoms (since they presumably believe wearing the mask avoids catching covid). Even a very small increase (e.g. by 3 or 4%) in mask wearers reporting symptoms could reduce the probability the seropositivity rate is lower in the intervention group to below 50%.

- There is possible bias given that unreachable participants were excluded from the study (this is normally considered bad practice in such trials). There were 4117 participants from the treatment villages who were unreachable and 2627 from the control villages. The fact that 2.3% of the treatment village participants were unreachable compared to only 1.6% of the control village participants suggests there may have been some systematic differences explaining why participants were unreachable. Imagine, for example, if a major reason for being unreachable was 'death'. The fact that a far greater proportion of treatment village participants were unreachable would invalidate the entire study.
- The uncertainty would be greater if we consider asymptomatic covid, as well as symptomatic.

We have explained why unreasonable assumptions may have led the authors of the Bangladesh study to make claims about the benefits of mask wearing that simply do not hold up when subject to rigorous Bayesian analysis. Those claims were further exaggerated in multiple media reports and consequently the study's results have been explicitly cited to justify continuing or reintroducing aspects of mask mandates by CDC [15], IDSA [16] and the UK's National Health Service [17]. In the light of this, the study paper published in Science [2] needs to corrected or withdrawn.

Models and data used in the Bayesian analysis

The models with all the data can be downloaded from http://www.eecs.qmul.ac.uk/~norman/Models/mask_study_models.zip and run using the free trial version of AgenaRisk https://www.agenarisk.com/agenarisk-free-trial

Acknowledgements

With thanks to the following for their comments and help: Clare Craig, Mike Deskevich, Jonathan Engler, Joshua Guetzkow, Steve Kirsch, Scott McLachlan, Martin Neil, Stephen Petty, Joel Smalley.

References

[1] https://www.poverty-action.org/sites/default/files/publications/Mask_RCT____Symptomatic_Seropositivity_083121.pdf
[2] Abaluck, J., Kwong, L. H., Styczynski, A., Haque, A., Kabir, M. A., Bates-Jefferys, E., … Mobarak, A. M. (2022). Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh. Science (New York, N.Y.), 375(6577), eabi9069. https://doi.org/10.1126/science.abi9069
[3] Bundgaard, H., Bundgaard, J. S., Raaschou-Pedersen, D. E. T., von Buchwald, C., Todsen, T., Norsk, J. B., … Iversen, K. (2021). Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers : A Randomized Controlled Trial. Annals of Internal Medicine, 174(3), 335–343. https://doi.org/10.7326/M20-6817
[4] https://www.dailymail.co.uk/health/article-9947923/Study-340-000-people-Bangladesh-finds-masks-effective-preventing-spread-Covid.html
[5] https://www.livescience.com/randomized-trial-shows-surgical-masks-work-curbing-covid.html
[6] https://med.stanford.edu/news/all-news/2021/09/surgical-masks-covid-19.html
[7] https://stevekirsch.substack.com/p/masks-dont-work
[8] https://principia-scientific.com/bangladesh-mask-study-extremely-weak-tea/
[9] https://boriquagato.substack.com/p/bangladesh-mask-study-do-not-believe
[10] Fenton, N. E., & Neil, M. (2018). Risk Assessment and Decision Analysis with Bayesian Networks (2nd ed.). CRC Press, Boca Raton.
[11] http://www.argmin.net/2021/11/29/cluster-power/
[12] https://vimeo.com/695666226
[13] https://www.argmin.net/2021/09/13/effect-size/

[14] http://www.argmin.net/2021/11/23/mask-rct-revisited/
[15] https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/masking-science-sars-cov2.html
[16] https://www.idsociety.org/covid-19-real-time-learning-network/infection-prevention/masks-and-face-coverings-for-the-public/#
[17] https://www.ouh.nhs.uk/working-for-us/staff/covid-staff-faqs-masks.aspx

Appendix 1: The key data in the Bangladesh mask study paper

The key data are summarised in their Table A1 that appeared in their original appendix:

Table A1: Endline Blood Sample Collection

	Treatment Villages	Control Villages	Total
Number of People Identified in Baseline Household Visits	178,288	163,838	342,126
Number of People Reached for Symptom Collection on in the Midline and Endline Visits	174,171	161,211	335,382
Number of People with WHO-defined COVID-19 Symptoms	13,273	13,893	27,166
Number of Symptomatic Endline Blood Samples Collected	5,414	5,538	10,952
Number of Symptomatic Endline Blood Samples Tested	5,006	4,971	9,977

However, what is very strange about this table (inexplicably not picked up by any reviewer before its publication) is that their key surrogate outcome measure (the number of people testing seropositive in each group) is NOT provided. Instead, in the main text, the authors say:

> *Omitting symptomatic participants who did not consent to blood collection, symptomatic seroprevalence was 0.76% in control villages and 0.68% in the intervention villages.*

Intuitively, we could calculate the number of seropositives by multiplying the rates (0.76% and 0.68% respectively for control and treatment villages) which would result in:

- 1225 in control villages
- 1184 in treatment villages

However, as noted by Recht [13]:

> Where do these seropositivity percentages come from? The paper does not make clear what is being counted. Do the authors compute the number of cases in treatment divided by the number of individuals treated? Or do they compute the prevalence in each cluster and average these up? These two different estimates of prevalence can give different answers.

In fact, as later reported by Recht [14] when he got access to the raw data, he was able to calculate that the numbers were:

- 1106 in control villages
- 1086 in treatment villages

This is why the raw data we assume are:

- Control: 1106 from 161211
- Treatment: 1086 from 174171

Appendix 2: The Bayesian network models

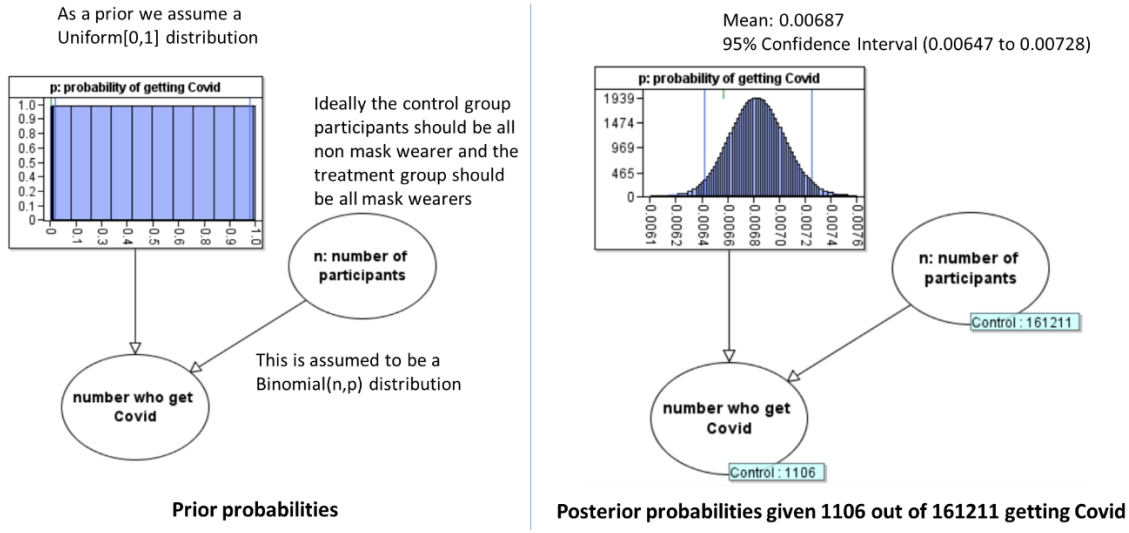

Figure 1 Basic model with a) prior and b) posterior probabilities

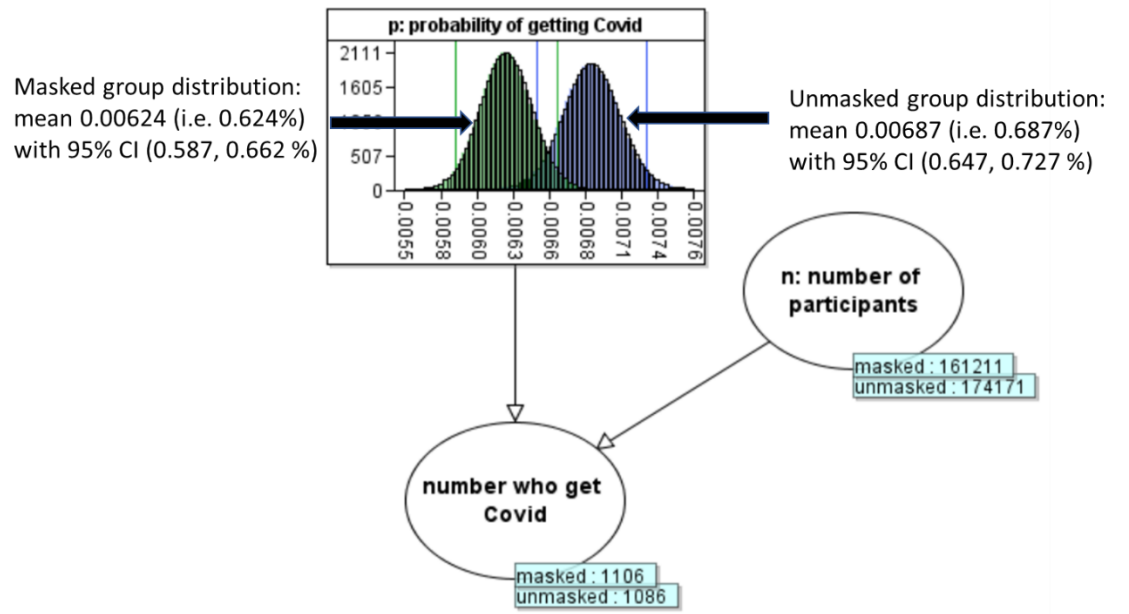

Figure 2 Posterior probabilities for masked v unmasked using 'hypothetical data' (CI stands for Confidence Interval)

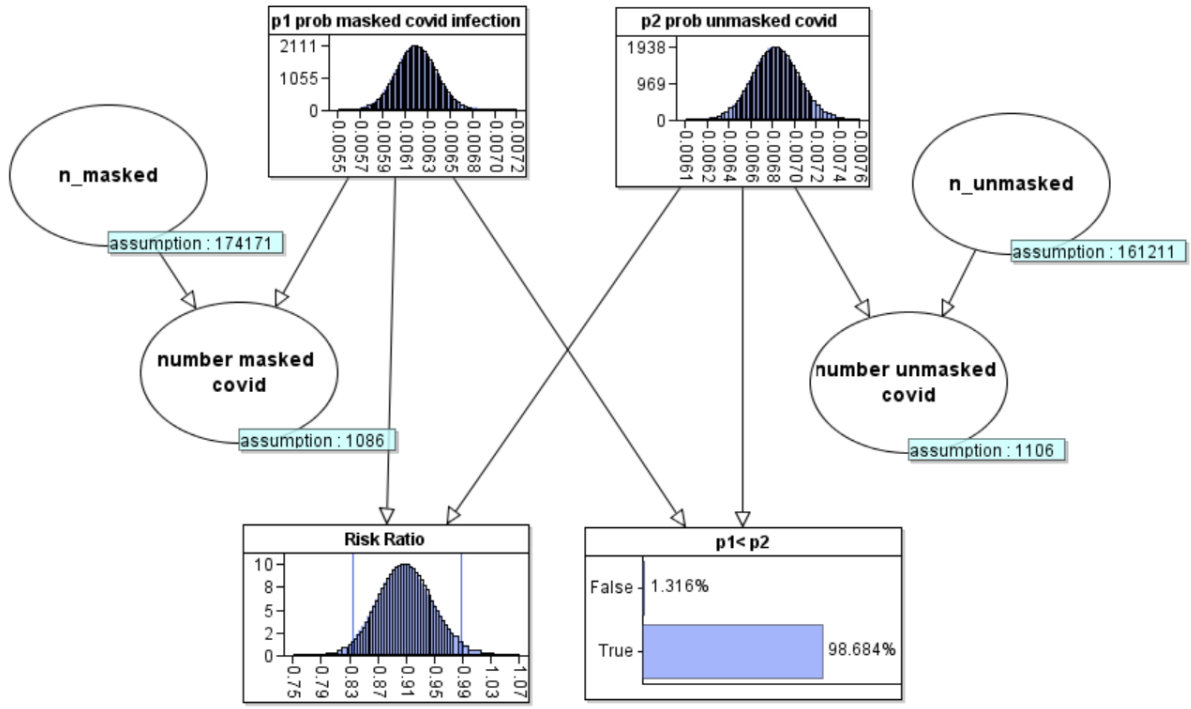

Figure 3 Interpreting the difference between the two distributions

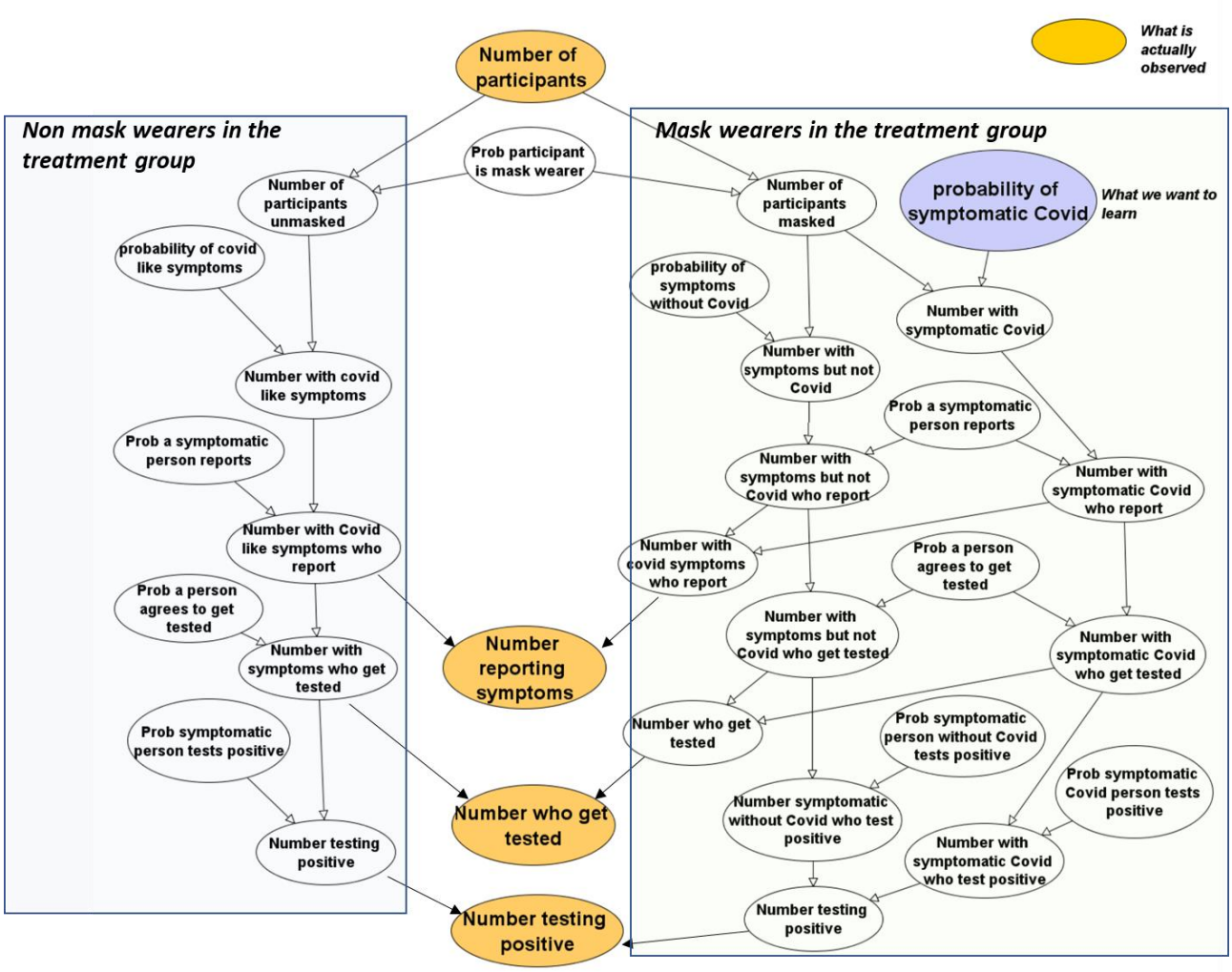

Figure 4 Causal (Bayesian) network of the problem for the treatment villages (a similar model applies to the control villages)

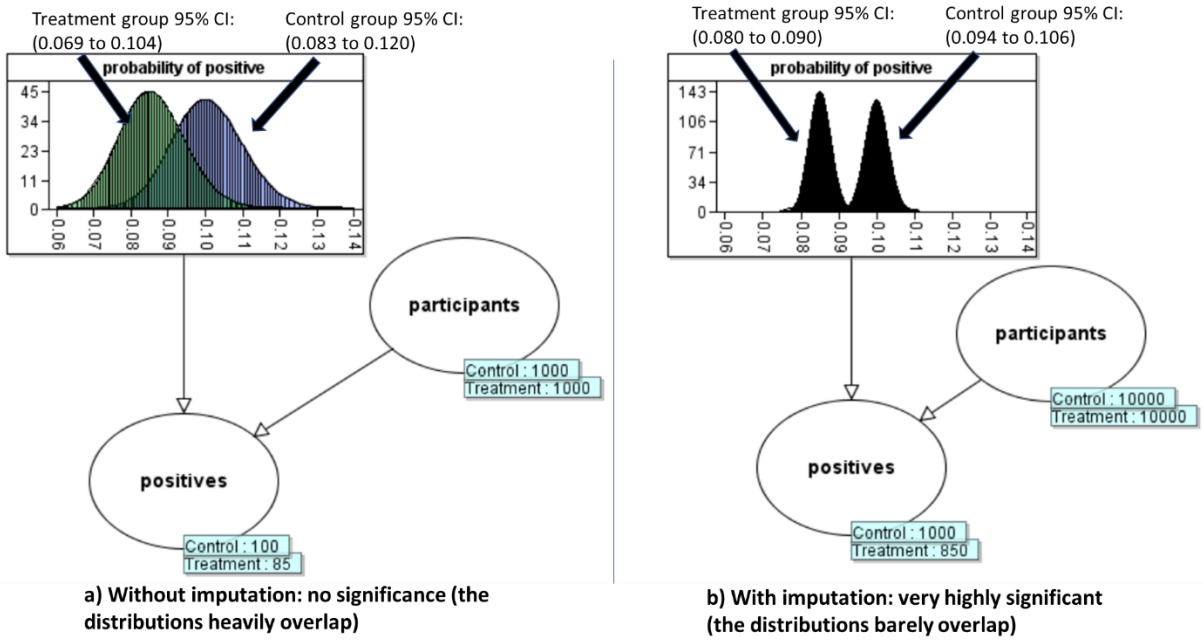

Figure 5 Effect of imputation

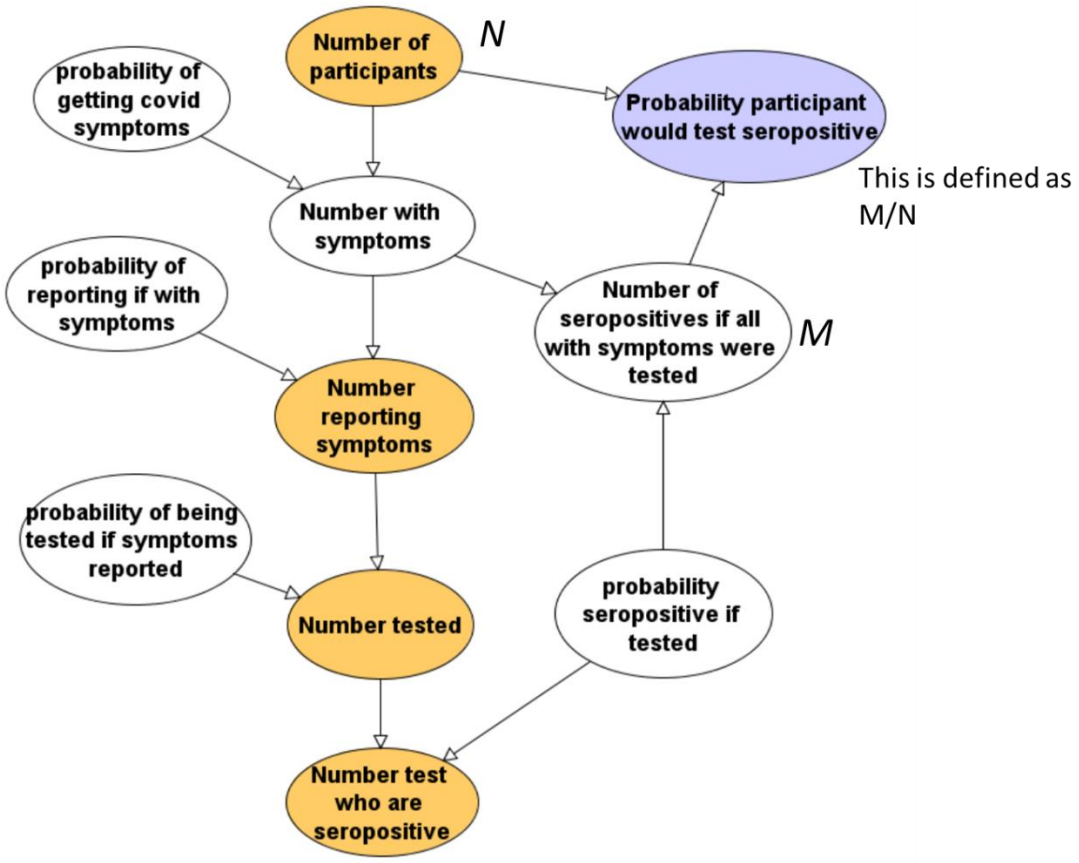

Figure 6 Simplified model needed to infer probability participant would test seropositive

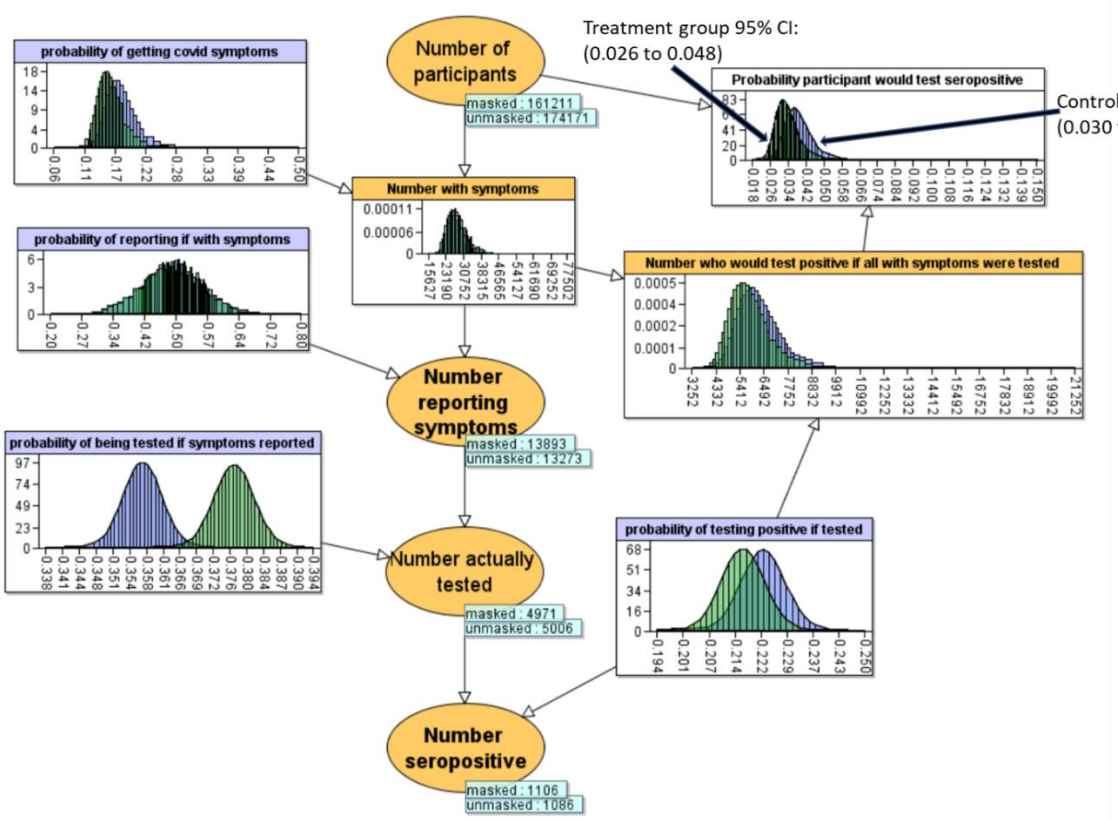

Figure 7 Posterior distributions given strong priors and no adjustment for cluster correlation