Unmasking the mask studies: why the effectiveness of surgical masks in preventing respiratory infections has been underestimated

Pratyush K. Kollepara,[1,2,*] Alexander F. Siegenfeld,[1,3,*] Nassim Nicholas Taleb,[4] and Yaneer Bar-Yam[1]

[1] *New England Complex Systems Institute, Cambridge, MA*
[2] *Department of Physics, BITS Pilani K K Birla Goa Campus, Goa, India*
[3] *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA*
[4] *Tandon School of Engineering, New York University, Brooklyn, NY*

SUPPLEMENTARY DATA AND METHODS

1. Accounting for non-linearities in the effectiveness of masks

The framework developed in the Methods section of the main text is further developed to determine how the probability of infection is affected when face masks are used.

a. *One mask*

Let γ be the typical amount by which a mask reduces the effective exposure from a single exposure event—i.e. $\tilde{v} \to (1-\gamma)\tilde{v}$.

Since $f(0) = 0$ and $f(v)$ is convex, the simplest possible expression for $f(v)$ is the scale-free form $f(v) = (v/v_0)^\beta$, for some $v_0 > 0$ and $\beta > 1$. In this case, if a mask reduces the viral dose v to v/b, γ can be calculated exactly as $\gamma = 1 - b^{-\beta}$, regardless of v. For small exposures, the infection probability is roughly equal to the effective exposure, which is reduced by a factor greater than b (i.e. $\frac{1}{1-\gamma} > b$) due to convexity ($\beta > 1$), consistent with the analysis in the main text. This effect could potentially be quite large: e.g. for $\beta = 4$, a mask filtering half of the viral particles ($b = 2$) corresponds to a sixteen-fold reduction in effective exposure ($\gamma \approx 0.94$). If we relax this assumption on the function f, then γ becomes an effective parameter that may depend on the distribution of viral doses to which an individual is exposed. Regardless of the precise form of $f(v)$, however, $\frac{1}{1-\gamma} > b$ will always hold due to the convexity of $f(v)$, i.e. masks will always have a disproportionately large effect on the effective exposure (and thus also on the infection probability when the effective exposure is small).

Then, if a mask is worn for a fraction α of all exposures, the total effective exposure will be reduced from $\tilde{v}_T$ to $(1-\alpha\gamma)\tilde{v}_T$. The probability of infection is thus

$$\tilde{p}((1-\alpha\gamma)\tilde{v}_T) = 1 - e^{-(1-\alpha\gamma)\tilde{v}_T} \quad (1)$$

(see Figure 2 of the main text).

Thus, we see that for any fixed γ (mask effectiveness) and $\tilde{v}_T$ (total effective exposure without a mask), the benefit of wearing a mask is a convex function of the fraction α of the exposure events for which it is worn. In other words, wearing a mask x times as often will change the reduction in infection probability by more than a factor of x. Thus, even if masks were 100% effective ($\gamma = 1$), a study in which participants wear masks 10% of the time would need to have sufficient power to detect less than a 10% reduction in the probability of infection. Our statistical power analysis therefore overestimates the true power of the studies.

b. *Two masks*

To the extent that two masks together have an approximately linear effect on the effective exposure (e.g. if one person wearing a mask reduces effective exposure by $1-\gamma_1$ and the second person wearing a mask reduces effective exposure by $1-\gamma_2$, then both wearing masks reduces effective exposure by $1-\gamma_{12} \approx (1-\gamma_1)(1-\gamma_2)$), the effect on the probability of transmission will be super-linear, since the probability of infection $\tilde{p}(\tilde{v})$ is concave in the effective

* These two authors contributed equally.

exposure $\tilde{v}$. In other words, especially for individuals who would have received a large total effective exposure without masks, both the susceptible and infectious individuals wearing masks will have a larger effect than would be calculated if each mask had an independent effect on the probability of transmission.

If the effect of the two masks on the effective exposure is super-linear (i.e. $1-\gamma_{12} < (1-\gamma_1)(1-\gamma_2)$), then the effect on the probability of transmission will be super-linear to an even greater extent. If the effect of the two masks on the effective exposure is sub-linear (i.e. $1-\gamma_{12} > (1-\gamma_1)(1-\gamma_2)$), then whether or not they still have a super-linear effect on the probability of transmission will depend on the total effective exposure.

(Note: Under the simplest possible form for $\tilde{v} = f(v)$, i.e. $f(v) = (v/v_0)^\beta$, if the mask on the infected individual reduces v by a factor of b_1, the mask on the susceptible individual reduces v by a factor of b_2, and together the masks reduce v by a factor of $b_1 b_2$, then the masks will have a linear effect on effective exposure, i.e. $1-\gamma_{12} = (1-\gamma_1)(1-\gamma_2)$. Under other forms for $f(v)$ or assumptions about how the masks affect v, other behavior is possible.)

2. Power analyses

Let p_1 and p_2 be the probabilities of getting infected in the non-mask (size N_1) and mask group (size N_2), respectively. Defining $\epsilon = p_1 - p_2$, the null hypothesis is $H_0 : \epsilon = 0$ and the alternate hypothesis is $H_1 : \epsilon \neq 0$. A test statistic is

$$W = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}_1(1-\hat{p}_1)/N_1 + \hat{p}_2(1-\hat{p}_2)/N_2}} = \frac{\hat{p}_1 - \hat{p}_2}{\hat{s}} \tag{2}$$

where $\hat{p}_1$ and $\hat{p}_2$ refer to the observed fraction of infections, assumed to be normally distributed random variables whose means are p_1 and p_2 (this approximation is asymptotically exact). We use the shorthand $\hat{s}$ for the denominator of W; note that $\hat{s}$ is an estimator for s, where $s^2 = p_1(1-p_1)/N_1 + p_2(1-p_2)/N_2$ is the sum of the asymptotic variances of $\hat{p}_1$ and $\hat{p}_2$. Asymptotically, $W - \epsilon/s$ follows a standard normal distribution. Using the standard notation $\Phi(z_{1-\alpha/2}) = 1 - \alpha/2$ where $\Phi(x)$ is the standard normal cumulative distribution function, the rejection region under H_0 for a significance level of α is given by the union of

$$W < -z_{1-\alpha/2} \text{ and } W > z_{1-\alpha/2} \tag{3}$$

The various studies may use slightly different statistical tests, but the differences between tests should be small and will asymptotically disappear entirely. For any particular values of ϵ and s, the probability $W < -z_{1-\alpha/2}$ is asymptotically given by $\Phi(-z_{1-\alpha/2} - \epsilon/s)$ and the probability $W > z_{1-\alpha/2}$ is asymptotically given by $1 - \Phi(z_{1-\alpha/2} - \epsilon/s) = \Phi(-z_{1-\alpha/2} + \epsilon/s)$. Thus, given ϵ and s, the power, denoted by $1-\beta$ and equal to the probability that the null hypothesis is rejected if it is indeed false, is asymptotically given by

$$1 - \beta = \Phi(-z_{1-\alpha/2} - \epsilon/s) + \Phi(-z_{1-\alpha/2} + \epsilon/s) \tag{4}$$

Under the assumptions that masks are fully effective ($\gamma = 1$) and that the probability of infection p_{inf} decreases linearly with the adherence, the effect of mask usage is

$$p_{\text{inf}} \to p_{\text{inf}}(1-a) \tag{5}$$

where the adherence a is the average fraction of exposure events for which the masks were used (see above; here we use a instead of α for the adherence to avoid confusion with the significance level). Thus, for an infection probability $p_{\text{inf}} = p_1$ in the non-mask group (size N_1), the infection probability in the mask group (size N_2) will be $p_2 = p_1(1-a)$. Thus, by estimating p_1 and a for each study, we can use eq. (4) to find power of each study given the sizes of their non-mask and mask groups, as well as to find the sample size (i.e. total number of participants) that would have been required for 80% power. For the latter estimate, we assume a study design in which the participants are evenly divided between the non-mask and mask groups (i.e. $N_1 + N_2 = 2N_1 = 2N_2$) and rounded up the necessary sample size to the nearest even integer.

For certain studies, some participants in the non-mask group used masks as well. In this case, adherence in both the mask group and non-mask group must be considered. Under the assumption that probability of infection decreases linearly with effective adherence, the probability of infection in the non-mask group p_1 is related to the probability of infection without masks p_0 by $p_1 = p_0(1-\gamma a_1)$ where a_1 is the adherence in the non-mask group and γ is mask effectiveness. Then the probability of infection in the mask group will be $p_2 = p_0(1-\gamma a_2)$ where a_2 is the adherence in the mask group. The net adherence a is defined by $p_2 = (1-\gamma a)p_1$, which yields $a = \frac{a_2 - a_1}{1-\gamma a_1}$. In our analyses we assume $\gamma = 1$, which leads to an overestimate for the net adherence a of

$$a = \frac{a_2 - a_1}{1 - a_1} \tag{6}$$

We estimate p_1 using the observed fraction of infections in the non-mask group $\hat{p}_1$. To check the robustness of our conclusions, we did a sensitivity analysis and found that if $\hat{p}_1$ differs from p_1 by a standard deviation (i.e. if we increase our estimate of p_1 by $\sqrt{\frac{1}{4N_1}}$), all studies that were under-powered ($< 80\%$), except for one [1] remain under-powered. (To ensure robustness we used $\sqrt{\frac{1}{4N_1}}$ as the standard deviation, which is the maximum possible value of the true standard deviation $\sqrt{p_1(1-p_1)/N_1}$.) If $\hat{p}_1$ underestimates p_1 by two standard deviations, another study [2] would have greater than 80% power under our assumptions. It should be noted, however, that these assumptions overestimate the power in multiple ways (fully effective masks, overestimated adherence values, assuming a linear relationship between adherence and effectiveness, and the fact that individuals whose infections were not detected until after the start of the study could have actually been infected before they start of the study, i.e. before the mask intervention was implemented).

A more significant limitation of our analysis is in the difficulty in estimating adherence. Adherence is often reported qualitatively, and even when quantitative, it is reported as the amount of time for which one wears a mask, which may differ from the fraction of exposures for which masks were worn. To account for this difficulty, our strategy has been to consistently overestimate statistical power; to this end, we have erred on the side of overestimating adherence (see table II), and have also used other overestimating assumptions described in the previous paragraph.

TABLE I. Summary of statistical power analysis. Given the adherence levels reported in the studies, the sample size necessary for a statistical power of 80% for a two-tailed test and significance level of 0.05 (assuming participants are equally divided between the non-mask and mask groups) is presented in column XI. The statistical power given the actual sizes of the non-mask and mask groups is presented in column X. These calculations were made for the case in which masks are 100% effective; if masks are effective but not perfectly so, the necessary sample sizes for 80% power (column XI) will be larger, while the statistical powers given the actual sample sizes (column X) will be lower. Studies found to have greater than 80% power are in bold (nos. 15-23), and studies that found a statistically significant reduction in infections in the mask group are italicized (nos. 4, 14-23). Adherence is defined as the fraction of exposure for which masks were used; calculations of adherence for each study are presented in table II. For studies that reported multiple analyses, each analysis is listed as its own entry (e.g. Aiello (2012) [1] performed one analysis in which infection is defined by influenza-like illness (no. 10) and one analysis in which infection is defined by a positive PCR test result (no. 11)). Note that only in the intention-to-treat analyses are participants randomly divided between the non-mask and mask groups; in survey and per-protocol analyses, which group a participant belongs to depends on whether or not that individual reported wearing a mask with a frequency above a threshold decided by the study.

Abbreviations: ITTA: Intention-to-treat analysis; **PPA**: Per-protocol analysis; **ILI**: Influenza-like illness; **ARI**: Acute respiratory infection; **URTI**: Upper respiratory tract infection; **PCR**: Polymerase chain reaction test (nasopharyngeal swab test); **SARS**: Severe Acute Respiratory Syndrome

No. (I)	Name (II)	Year (III)	Mask Use (IV)	Adherence (V)	Size of non-mask group (VI)	Fraction of non-mask group infected (VII)	Size of face mask group (VIII)	Fraction of face mask group infected (IX)	Statistical Power (X)	Required sample size for a power of 0.8 (XI)	Actual sample size (XII)	Primary Outcome (XIII)	Significant reduction in infections in mask group? (XIV)
1	Cowling (ITTA) [3]	2008		0.37	205	0.06	61	0.07	0.118	2910	266	Antibody Test	No
2	MacIntyre (ITTA) [4]	2009	45% people used masks	0.36	100	0.030	94	0.064	0.078	6444	194	ILI	No
3	MacIntyre (PPA) [4]	2009	Less than 2/5 days	0.32	170	0.053	19	0.211	0.066	4636	189	ILI	No
4	*MacIntyre (PPA) [4]*	2009	5/5 days	0.80	170	0.053	18	0.000	0.298	530	188	ILI	Yes
5	Alfelali (ITTA) [5]	2020	Extremely low	0.014	3139	0.019	3199	0.030	0.051	7862388	6338	PCR	No
6	Alfelali (PPA) [5]	2020	~1.6 hours/day	0.068	2200	0.036	1291	0.051	0.067	176108	3491	PCR	No
7	Simmerman (ITTA) [7]	2011		0.23	292	0.226	291	0.227	0.347	1866	583	ILI	No
8	Canini (ITTA) [9]	2010	3.9 hours/day	0.38	158	0.158	148	0.162	0.345	990	306	Antibody Test	No
9	Aiello (ITTA) [2]	2010	~4 hours/day	0.33	177	0.452	99	0.444	0.699	330	276	ILI (Fever)	No
10	Aiello (ITTA) [1]	2012	~5 hours/day	0.42	370	0.138	392	0.117	0.735	884	762	ILI	No
11	Aiello (ITTA) [1]	2012	~5 hours/day	0.42	370	0.043	392	0.031	0.284	3074	762	PCR	No
12	MacIntyre (ITTA) [8]	2016	~4 hours/day	0.33	295	0.010	302	0.003	0.074	22924	597	ILI	No
13	Cowling (ITTA) [10]	2009		0.44	257	0.054	258	0.070	0.269	2234	515	PCR	No
14	*Barasheed (ITTA) [11]*	2014		0.25	53	0.528	36	0.306	0.236	440	89	ILI	Yes
15	***Sung (pre-post) [12]***	2012	Hospital setting	0.80	920	0.103	454	0.033	1.000	260	1374	PCR	Yes
16	***Choudhry (survey) [13]***	2006	Most of the time	0.59	477	0.612	340	0.150	1.000	52	817	ARI	Yes
17	***Al-Jasser (survey) [14]***	2013	Most of the time	0.59	656	0.550	216	0.454	1.000	62	872	URTI	Yes
18	***Suess (ITTA) [15]***	2012		0.80	82	0.232	69	0.087	0.938	102	151	PCR	Yes
19	***Wu (survey) [16]***	2004	Mask usage outdoors	0.80	73	0.630	70	0.386	1.000	22	143	ILI	Yes
20	***Kim (survey) [17]***	2012	Mask usage in schools	0.80	4164	0.057	466	0.030	1.000	486	4630	ILI	Yes
21	***Lau (survey) [18]***	2004	Mask usage in public	0.80	511	0.466	479	0.192	1.000	38	990	Probable SARS	Yes
22	***Lau (survey) [19]***	2004	Mask usage in hospitals	0.80	98	0.173	177	0.079	0.927	144	275	Probable SARS	Yes
23	***Wu (survey) [20]***	2016	Mask usage in hospitals	0.80	10298	0.477	2728	0.423	1.000	36	13026	ILI	Yes

TABLE II: Adherence calculations for each study.

No.	Name	Year	Masks used by	Description and calculation
1	Cowling (ITTA) [3]	2008	Infected patients and their contacts	Household study: 45% of 21 index cases used masks and 21% of 61 contacts wore masks. To overestimate adherence, we assume no transmission occurs while either the index patient or contact is wearing a mask. Neglecting correlations between whether or not the index patient wore a mask and the number of contacts of that index patient, an upper bound for the probability that either a contact or the index patient corresponding to that contact used a mask is 45%+21% = 66% (this is likely an overestimate since households in which index patients wear masks and households in which contacts wear masks are almost certainly not mutually exclusive). In the control group, 30 % of index patients and 1 % of contacts used masks. Those classified as using masks used them often or always; therefore we assume that they used masks for 80% of all exposures, a likely overestimate since the participants were asked to use masks only when they are not sleeping or eating. Therefore, the adherence in the mask group is estimated as $0.66 \times 0.8 = 0.53$, and adherence in the control group is estimated as $0.31 \times 0.8 = 0.25$. This leads to a net adherence of 0.37 according to eq. (6).
2	MacIntyre (ITTA) [4]	2009	Contacts of infected patients	Household study over 5 days: Contacts were told to use masks when in the same room as the index patient. We consider only the surgical mask group (the other group was using P2 masks). On day 3, maximum adherence was reported: 45% of contacts used masks for most of the time. We assume that those who used masks used them for 80% of exposures, a likely overestimate since contacts did not use masks while sleeping, even if the child (index patient) was next to them in bed, and because the contacts could have been infected even if they were not in the same room as the index patient. The adherence is estimated as $0.45 \times 0.8 = 0.36$
3	MacIntyre (PPA) [4]	2009	Contacts of infected patients	Household study over 5 days (see row no. 2): Participants in this arm of the per-protocol analysis used masks for < 2 out of 5 days. Overestimating adherence at 0.8 for 2 days gives adherence $= 2/5 \times 0.8 = 0.32$.
4	MacIntyre (PPA) [4]	2009	Contacts of infected patients	Household study over 5 days (see row no. 2): Participants in this arm of the per-protocol analysis used masks for all 5 days. Overestimating adherence at 0.8 for 5 days gives adherence an adherence of $5/5 \times 0.8 = 0.8$.
5	Alfelali (ITTA) [5]	2020	Susceptible individuals	Hajj study: From figure 2 of the study, we can only obtain approximate numbers since numerical data is not available in the figure. An average across the four days gives us the percentage of people using masks for various amounts of time. Using the upper bounds of the reported time ranges, we compute the average mask usage duration. For the last time range (greater than 3 hours), we assume that masks were used on average for 5 hours. This leads to an average mask use of 0.778 hours and an adherence in the mask group of $0.778/24 = 0.032$. Participants in the control group used masks for 0.438 hours on average, yielding an adherence in the control group of $0.438/24 = 0.018$. The net adherence value is thus 0.014 (eq. (6)). Note that the systematic review [6] uses an older pre-print version of this study.
6	Alfelali (PPA) [5]	2020	Susceptible individuals	Hajj study (see row no. 5): Those who wore masks were compared to those who did not. The average mask use among those who wore masks was 1.637 hours; thus adherence $= 1.637/24 = 0.0682$.
7	Simmerman (ITTA) [7]	2011	Infected patients and their contacts	Household study: We compare the hand-hygiene group with the hand-hygiene + mask group. Only median (and not mean) mask usage was reported for the index and contact individuals; we therefore approximate the mean with the median. The median mask usage for the index patient was 35 minutes. The mean of median mask usage for contacts—parents, siblings and other relations—was 107.9 minutes. We estimate that index patients and contacts were in contact for 10.4 hours per day using data from a similar study [8] (row no. 12). Adherence is therefore estimated as $\frac{107.9+35}{60\times10.4} = 0.23$ (see row no. 1 for why the index and contact mask usages were added together), a likely overestimate, given that the majority of the households resided in small one-bedroom apartments and thus were likely in contact for significantly greater than 10.4 hours per day on average. Furthermore, contacts could have been infected outside of their homes. Also, it was reported that 17.6% of individuals in the control group used masks, meaning it was likely that those in the hand-hygiene-only group did as well (which would further reduce the net adherence).
8	Canini (ITTA) [9]	2010	Infected patients	Household study: Average mask use was 3.9 hours per day (from table 3 of the study). We estimate that index and contact patients were in contact for 10.4 hours per day using data from a similar study [8] (row no. 12). Adherence is therefore estimated as $\frac{3.9}{10.4} = 0.38$, a likely overestimate given that contacts could have been infected outside their homes, or in their homes while not in contact with the index patient.
9	Aiello (ITTA) [2]	2010	Susceptible individuals	University residence hall: Mask usage was recorded inside the residence hall and they were used for 3.92 hours per day. Assuming that residents spent 12 hours outside the halls, we exclude it from the adherence calculation. Adherence $= \frac{3.92}{24-12} = 0.33$, a likely overestimate because participants were only encouraged but not required to use masks outside the residence halls, where they may be infected. In addition, the participants had left the residence halls for spring break, during which they were not required to wear masks.
10	Aiello (ITTA) [1]	2012	Susceptible individuals	University residence hall: Masks were used for 5.08 hours per day. Adherence $= \frac{5.08}{24-12} = 0.42$ (see row no. 9).
11	Aiello (ITTA) [1]	2012	Susceptible individuals	University residence hall: Masks were used for 5.08 hours per day. Adherence $= \frac{5.08}{24-12} = 0.42$ (see row no. 9).
12	MacIntyre (ITTA) [8]	2016	Infected patients	Household study: In the mask group, index patients were in contact with contacts for an average of 10.4 hours, and used masks for an average of 4.4 hours. The adherence in the mask group is thus estimated as $\frac{4.4}{10.4} = 0.42$. In the control group, average mask usage was 1.4 hours; adherence in the control group is thus estimated as $\frac{1.4}{10.4} = 0.13$. Net adherence is thus 0.33 (eq. (6)).
13	Cowling (ITTA) [10]	2009	Infected patients and their contacts	Household study: We compare the hand-hygiene group with the hand-hygiene + mask group. In the hand-hygiene + mask group, 49% of index cases and 26 % of contacts used a mask often or always. We therefore calculate adherence in the hand-hygiene + mask group as $(0.49 + 0.26) \times 0.8 = 0.60$ (see row no. 1). In the hand-hygiene group, 5 % of contacts and 31 % of index cases used masks, which leads to an adherence $= (0.31 + 0.05) \times 0.8 = 0.29$ in the hand-hygiene group. Net adherence is thus 0.44 (eq. (6)).
14	Barasheed (ITTA) [11]	2014	Susceptible individuals	Hajj pilgramage: 36 people were in the face mask group: 8 people never used a mask; 11 people used masks for < 4 hours; 8 people used masks used for 5-8 hours; 9 people used masks for > 8 hours (from table 2 of the study). Using the upper limits of the duration ranges (and 12 hours for the > 8 hour group), adherence $= \frac{1}{36}(8 \times 0/24 + 11 \times 4/24 + 8 \times 8/24 + 9 \times 12/24) = 0.25$.

15	Sung (Pre-post) [12]	2012	Potentially infected individuals	Visitors had to use face masks when they visited patients in their rooms and the incidence of infections was recorded among the patients. Although adherence was not reported, it is reasonable to assume that adherence was high since the study was conducted in a hospital where doctors and health care workers would have ensured that protocols are followed; in addition the visitors were in contact with patients only for a limited duration. We therefore assume an adherence of 0.8
16	Choudhry (Survey) [13]	2006	Susceptible individuals	Survey study for Hajj pilgrims: We consider the group of male pilgrims who reported using masks most of the time, compared to a group who did not use masks. We assume that masks were not used while sleeping or eating, and note that the pilgrims remain susceptible to infection during such activities since they slept in shared tents. Allotting 10 hours per day for sleeping and eating and other activities during which masks were not worn, we estimate the adherence as $14/24 = 0.59$.
17	Al-Jasser (Survey) [14]	2013	Susceptible individuals	Survey study for Hajj pilgrims: We consider the group of male pilgrims who reported using masks most of the time, compared to a group who did not use masks, and therefore estimate adherence as 0.59 (see row no. 16).
18	Suess (ITTA) [15]	2012	Infected patients and their contacts	Household study: From figures 2 and 3 in the study, the average mask usage (across 8 days and across both seasons) among contacts and index patients is 69.4% and 56.4%, respectively. From this data it is not impossible that in every household either the index patient or contacts were wearing masks; using this potential overestimate, we calculate adherence as $1 \times 0.8 = 0.8$ (see row no. 1).
19	Wu (survey) [16]	2004	Susceptible individuals	Survey study: Face mask usage was reported only outside the home. Adherence was reported subjectively – 'Never', 'Sometimes', 'Always' (table 1 of the study). We compare the groups which used masks always and never used masks, and use an adherence value of 0.8 for the 'Always' group, a likely overestimate since participants could have been infected from household contacts.
20	Kim (survey)[17]	2011	Susceptible individuals	Survey study among school children for influenza: Mask usage during school hours was reported as 'continuous', 'irregular', 'not used'. We assume an adherence of 0.8 for the 'continuous' group (and compare the infection rate to the group that did not use masks), a likely overestimate since children could be infected outside of school hours.
21	Lau (survey) [18]	2004	Susceptible individuals	Survey study during SARS epidemic: Mask usage was recorded only for public places. The study considered the frequent use of masks as using a mask, and occasional/seldom/no use was considered as not using a mask. We assume a value of 0.8 for adherence, a likely overestimate since people could have gotten infected at home where mask usage was not recorded and since some mask usage was possible in the non-mask group.
22	Lau (survey) [19]	2004	Susceptible individuals	Survey study during SARS epidemic: Mask usage was recorded only during hospital visits to patients with SARS. We use an adherence value of 0.8 for hospital settings (see row no. 15). For this study 0.8 is likely an overestimate since SARS infection could have occurred outside of the hospital as well.
23	Wu (survey) [20]	2016	Susceptible individuals	Survey study for influenza-like illness. Mask usage was recorded only during hospital visits. We use an adherence value of 0.8 for hospital settings (see row no. 15). For this study 0.8 is likely a substantial overestimate since infection could have occurred outside of the hospital as well.

TABLE III. Studies not included in power analysis.

Name	Year	Reason for exclusion from power analysis
Shin [21]	2018	Study was randomized for testing a common cold drug rather than mask usage, and mask usage was comparable in both of the groups.
Zhang [22]	2013	Unknown adherence and incomplete data.
Jolie [23]	1998	Animal to human transmission: We consider only human to human transmission for our analysis.
Tahir [24]	2019	Animal to human transmission: We consider only human to human transmission for our analysis.
Larson [25]	2010	Mask adherence was reported to be 'poor' but neither the percentage of participants using masks nor the duration of mask usage was reported, so we could not make an estimate for the adherence.
Emamian [26]	2013	Survey study for Hajj pilgrims: Adherence for mask usage was reported only as 'Yes' or 'No'. Even occasional use of mask was considered as 'Yes'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Deris [27]	2010	Survey study for Hajj pilgrims: Adherence for mask usage was reported only as 'Yes' or 'No'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Uchida [28]	2017	Survey study for children. Mask usage was reported as 'using masks at any time or place'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Balaban [29]	2012	Survey study for Hajj pilgrims: Adherence for mask usage was reported only as 'Yes' or 'No'. Since adherence data stratified by frequency and/or duration was not reported, we could not make an estimate for the adherence.
Zein	2002	Study not available.

[1] Aiello AE, Perez V, Coulborn RM, Davis BM, Uddin M, Monto AS. Facemasks, Hand Hygiene, and Influenza among Young Adults: A Randomized Intervention Trial. PLoS ONE. 2012 Jan;7(1):e29744. Available from: https://doi.org/10.1371/journal.pone.0029744.

[2] Aiello AE, Murray GF, Perez V, Coulborn RM, Davis BM, Uddin M, et al. Mask use, hand hygiene, and seasonal

influenza-like illness among young adults: A randomized intervention trial. The Journal of Infectious Diseases. 2010 Feb;201(4):491–498. Available from: https://doi.org/10.1086/650396.

[3] Cowling BJ, Fung ROP, Cheng CKY, Fang VJ, Chan KH, Seto WH, et al. Preliminary Findings of a Randomized Trial of Non-Pharmaceutical Interventions to Prevent Influenza Transmission in Households. PLoS ONE. 2008 May;3(5):e2101. Available from: https://doi.org/10.1371/journal.pone.0002101.

[4] MacIntyre CR, Cauchemez S, Dwyer DE, Seale H, Cheung P, Browne G, et al. Face Mask Use and Control of Respiratory Virus Transmission in Households. Emerging Infectious Diseases. 2009 Feb;15(2):233–241. Available from: https://doi.org/10.3201/eid1502.081166.

[5] Alfelali M, Haworth EA, Barasheed O, Badahdah AM, Bokhary H, Tashani M, et al. Facemask against viral respiratory infections among Hajj pilgrims: A challenging cluster-randomized trial. PLOS ONE. 2020 Oct;15(10):e0240287. Available from: https://doi.org/10.1371/journal.pone.0240287.

[6] Brainard JS, Jones N, Lake I, Hooper L, Hunter P. Facemasks and similar barriers to prevent respiratory illness such as COVID-19: A rapid systematic review. medRxiv. 2020 Apr. Available from: https://doi.org/10.1101/2020.04.01.20049528.

[7] Simmerman JM, Suntarattiwong P, Levy J, Jarman RG, Kaewchana S, Gibbons RV, et al. Findings from a household randomized controlled trial of hand washing and face masks to reduce influenza transmission in Bangkok, Thailand. Influenza and Other Respiratory Viruses. 2011 Feb;5(4):256–267. Available from: https://doi.org/10.1111/j.1750-2659.2011.00205.x.

[8] MacIntyre CR, Zhang Y, Chughtai AA, Seale H, Zhang D, Chu Y, et al. Cluster randomised controlled trial to examine medical mask use as source control for people with respiratory illness. BMJ Open. 2016 Dec;6(12):e012330. Available from: https://doi.org/10.1136/bmjopen-2016-012330.

[9] Canini L, Andréoletti L, Ferrari P, D'Angelo R, Blanchon T, Lemaitre M, et al. Surgical Mask to Prevent Influenza Transmission in Households: A Cluster Randomized Trial. PLoS ONE. 2010 Nov;5(11):e13998. Available from: https://doi.org/10.1371/journal.pone.0013998.

[10] Cowling BJ, Chan KH, Fang VJ, Cheng CKY, Fung ROP, Wai W, et al. Facemasks and Hand Hygiene to Prevent Influenza Transmission in Households. Annals of Internal Medicine. 2009 Oct;151(7):437. Available from: https://doi.org/10.7326/0003-4819-151-7-200910060-00142.

[11] Barasheed O, Almasri N, Badahdah AM, Heron L, Taylor J, McPhee K, et al. Pilot Randomised Controlled Trial to Test Effectiveness of Facemasks in Preventing Influenza-like Illness Transmission among Australian Hajj Pilgrims in 2011. Infectious Disorders - Drug Targets. 2014 Oct;14(2):110–116. Available from: https://doi.org/10.2174/1871526514666141021112855.

[12] Sung AD, Sung JAM, Corbet K, Broadwater G, Hars V, Zanter A, et al. Surgical Mask Usage Reduces the Incidence of Parainfluenza Virus 3 in Recipients of Stem Cell Transplantation. Blood. 2012 Nov;120(21):462–462. Available from: https://doi.org/10.1182/blood.v120.21.462.462.

[13] Choudhry AJ, Al-Mudaimegh KS, Turkistani AM, Al-Hamdan NA. Hajj-associated acute respiratory infection among hajjis from Riyadh. East Mediterr Health J. 2006;12(3-4):300–309. Available from: https://apps.who.int/iris/handle/10665/117087.

[14] Al-Jasser FS, Kabbash I, Almazroa MA, Memish Z. Patterns of diseases and preventive measures among domestic hajjis from Central, Saudi Arabia. Saudi medical journal. 2012;33 8:879–86. Available from: http://www.emro.who.int/emhj-vol-19-2013/volume-19-supplement-1-mass-gathering/patterns-of-diseases-and-preventive-measures-among-domestic-hajjis-from-central-saudi-arabia-complete-republication.html.

[15] Suess T, Remschmidt C, Schink SB, Schweiger B, Nitsche A, Schroeder K, et al. The role of facemasks and hand hygiene in the prevention of influenza transmission in households: results from a cluster randomised trial; Berlin, Germany, 2009-2011. BMC Infectious Diseases. 2012 Jan;12(1). Available from: https://doi.org/10.1186/1471-2334-12-26.

[16] Wu J, Xu F, Zhou W, Feikin DR, Lin CY, He X, et al. Risk Factors for SARS among Persons without Known Contact with SARS Patients, Beijing, China. Emerging Infectious Diseases. 2004 Feb;10(2):210–216. Available from: https://doi.org/10.3201/eid1002.030730.

[17] Kim CO, Nam CM, Lee DC, Chang J, Lee JW. Is abdominal obesity associated with the 2009 influenza A (H1N1) pandemic in Korean school-aged children? Influenza and Other Respiratory Viruses. 2011 Dec;6(5):313–317. Available from: https://doi.org/10.1111/j.1750-2659.2011.00318.x.

[18] Lau J, Tsui H, Lau M, Yang X. SARS Transmission, Risk Factors, and Prevention in Hong Kong. Emerging Infectious Diseases. 2004;10:587 – 592. Available from: https://doi.org/10.3201/eid1004.030628.

[19] Lau JTF, Lau M, Kim JH, Wong E, Tsui HY, Tsang T, et al. Probable Secondary Infections in Households of SARS Patients in Hong Kong. Emerging Infectious Diseases. 2004 Feb;10(2):236–243. Available from: https://doi.org/10.3201/eid1002.030626.

[20] Wu S, Ma C, Yang Z, Yang P, Chu Y, Zhang H, et al. Hygiene Behaviors Associated with Influenza-Like Illness among Adults in Beijing, China: A Large, Population-Based Survey. PLOS ONE. 2016 Feb;11(2):e0148448. Available from: https://doi.org/10.1371/journal.pone.0148448.

[21] Shin K, Wakabayashi H, Sugita C, Yoshida H, Sato K, Sonoda T, et al. Effects of orally administered lactoferrin and lactoperoxidase on symptoms of the common cold. International Journal of Health Sciences. 2018;12:44 – 50.

[22] Zhang L, Peng Z, Ou J, Zeng G, Fontaine RE, Liu M, et al. Protection by Face Masks against Influenza A(H1N1)pdm09 Virus on Trans-Pacific Passenger Aircraft, 2009. Emerging Infectious Diseases. 2013 Sep;19(9). Available from: https://doi.org/10.3201/eid1909.121765.

[23] Jolie R, Bäckström L, Thomas C. Health Problems in Veterinary Students after Visiting a Commercial Swine Farm.

Canadian journal of veterinary research = Revue canadienne de recherche vétérinaire. 1998 02;62:44–8.

[24] Tahir MF, Abbas MA, Ghafoor T, Dil S, Shahid MA, Bullo MMH, et al. Seroprevalence and risk factors of avian influenza H9 virus among poultry professionals in Rawalpindi, Pakistan. Journal of Infection and Public Health. 2019 Jul;12(4):482–485. Available from: https://doi.org/10.1016/j.jiph.2018.11.009.

[25] Larson EL, Ferng YH, Wong-McLoughlin J, Wang S, Haber M, Morse SS. Impact of Non-Pharmaceutical Interventions on URIs and Influenza in Crowded, Urban Households. Public Health Reports. 2010 Mar;125(2):178–191. Available from: https://doi.org/10.1177/003335491012500206.

[26] Emamian M, Hassani A, Fateh M. Respiratory Tract Infections and its Preventive Measures among Hajj Pilgrims, 2010: A Nested Case Control Study. International Journal of Preventive Medicine. 2013;4:1030 – 1035.

[27] Deris ZZ, Hasan H, Sulaiman SA, Wahab MSA, Naing NN, Othman NH. The Prevalence of Acute Respiratory Symptoms and Role of Protective Measures Among Malaysian Hajj Pilgrims. Journal of Travel Medicine. 2010 Mar;17(2):82–88. Available from: https://doi.org/10.1111/j.1708-8305.2009.00384.x.

[28] Uchida M, Kaneko M, Hidaka Y, Yamamoto H, Honda T, Takeuchi S, et al. Effectiveness of vaccination and wearing masks on seasonal influenza in Matsumoto City, Japan, in the 2014/2015 season: An observational study among all elementary schoolchildren. Preventive Medicine Reports. 2017 Mar;5:86–91. Available from: https://doi.org/10.1016/j.pmedr.2016.12.002.

[29] Balaban V, Stauffer WM, Hammad A, Afgarshe M, Abd-Alla M, Ahmed Q, et al. Protective Practices and Respiratory Illness Among US Travelers to the 2009 Hajj. Journal of Travel Medicine. 2012 May;19(3):163–168. Available from: https://doi.org/10.1111/j.1708-8305.2012.00602.x.