

A Generalized Multinomial Probabilistic Model for SARS-CoV-2 Infection Prediction and Public Health Intervention Assessment in an Indoor Environment

Victor OK Li^{1,#,*}, Jacqueline CK Lam^{1,#,*}, Yuxuan Sun¹, Yang Han¹, Kelvin Chan¹,
Shan-shan Wang¹, Jon Crowcroft², Jocelyn Downey¹, Qi Zhang¹

¹Department of Electrical and Electronic Engineering

The University of Hong Kong, Hong Kong

²Department of Computer Science and Technology

The University of Cambridge, Cambridge, UK

[#]Equal Contributions

^{*}Corresponding and Senior Authors. Address: Department of Electrical and Electronic Engineering, The University of Hong Kong, Pok Fu Lam, Hong Kong. Email: vli@eee.hku.hk; jcklam@eee.hku.hk

Abstract

SARS-CoV-2 Omicron has become the predominant variant globally. Current infection models are limited by the need for large datasets or calibration to specific contexts, making them difficult to cater for different settings. To ensure public health decision-makers can easily consider different public health interventions (PHIs) over a wide range of scenarios, we propose a generalized multinomial probabilistic model of airborne infection to systematically capture group characteristics, epidemiology, viral loads, social activities, environmental conditions, and PHIs, with assumptions made on social distancing and contact duration, and estimate infectivity over short time-span group gatherings. This study is related to our 2021 work published in Nature Scientific Reports that modelled airborne SARS-CoV-2 infection (Han, Lam, Li, et al., 2021).¹ It is differentiated from former works on probabilistic infection modelling in terms of the following: (1) predicting new cases arising from more than one infectious in a gathering, (2) incorporating additional key infection factors, and (3) evaluating the effectiveness of multiple PHIs on SARS-CoV-2 infection simultaneously. Although our results reveal that limiting group size has an impact on infection, improving ventilation has a much greater positive health impact. Our model is versatile and can flexibly accommodate other scenarios by allowing new factors to be added, to support public health decision-making.

Keywords: SARS-CoV-2 Infection, Ventilation, Vaccination, Outdoor, Indoor, Masking, Public Health Interventions, Multinomial Probabilistic Model, Generalized Model, High Infectivity Variant, Omicron

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

1. Introduction

The SARS-CoV-2 Omicron variant, first detected in November 2021 in Botswana, has now become the globally dominant variant.² Omicron carries more than 30 mutations in the spike glycoprotein and evades over 85% of tested neutralizing antibodies,³ resulting in a higher risk of reinfection due to its greater ability to evade immunity from prior infection and reduced vaccine effectiveness.⁴

SARS-CoV-2 variants are primarily transmitted through close interpersonal contact.⁵ Non-clinical public health interventions (PHI), such as social distancing and lockdowns, have been introduced in many countries to reduce transmission,⁶ but have met with varying degrees of success in implementation and enforcement⁷ and carry economic and social costs. To inform public health decisions on the type and timing of PHIs, we have sought to create a simple, comprehensive and versatile SARS-CoV-2 infection model that can estimate infections over a wide range of scenarios and be easily updated to reflect changing key empirical characteristics of SARS-CoV-2, its modes and rate of transmission, and public health measures.

Research Gaps and Aims of the Study: Current models of SARS-CoV-2 transmission primarily rely on viral reproduction rate and Susceptible-Infected-Recovered (SIR) / Susceptible-Exposed-Infectious-Recovered (SEIR) transmission frameworks/models.⁸ These models compute the instantaneous reproductive numbers, parameterize the contact matrix, and then fit the models to epidemic curves. This approach is able to estimate infections across a large population group but requires infection case data over a longer time span (i.e. several days). In the real-world, data limitations may arise, especially over short time spans – for example, data may be delayed, inaccurate, incomplete or even not obtainable at all – and result in a potentially severe loss of accuracy.⁹

In contrast, other SARS-COV-2 infection models that statistically estimate viral infection risks are often context-based and lack generalizability. For example, Moritz *et al.* (2021) developed a transmission model based on data from an experimental indoor mass gathering and computational fluid dynamics model (CFD) simulations¹⁰ and Singanayagam *et al.* (2022) estimated the risk of community-based transmission in the United Kingdom.¹¹ However, their reference to specific contexts/scenarios restricts their transferability to other settings and CFD models in particular are highly geometrically-specific and determining distribution values across different room settings can be very time-consuming. Additionally,

both studies were relatively limited in providing a comprehensive and generalizable account of the risk factors affecting SARS-CoV-2 transmission/infection in different contexts.

To date, few probabilistic models have been developed to uncover the transmission patterns of SARS-CoV-2, especially for small group activities. One notable example is Tupper *et al.* (2020),¹² which proposed a simple mathematical model of transmission rate based on social contact (contact rate, duration, and mixing level) and calibrated this for different outbreaks. However, the model was limited to new infections arising from a single infectious individual only and did not incorporate a broader range of biological, behavioral, and environmental risk factors affecting infection in different scenarios, such as variant infectivity, vaccination or ventilation conditions.

In this study, we propose a multinomial probabilistic model to predict infections arising from short time span group activities. The model can comprehensively capture key factors affecting SAR-CoV-2 infection; be adapted to model different scenarios; and incorporate additional factors affecting transmissibility as research and circumstances develop.

Factors Affecting Contacts: Close and prolonged indoor contact can increase viral transmission of SARS-CoV-2.¹³ The longer an infectious person remains in close contact with others, the more likely an exposed (or uninfected) person is to become infected. Similarly, more infectious people or more people in a room increases the number of close contacts between infectious persons and exposed persons and the number of likely infections.

Factors Affecting Transmissibility: SARS-CoV-2 infection requires a relatively low initial infectious dose compared to other respiratory viruses¹⁴. Multiple factors affect transmissibility, including: the infectivity of the viral variant, the contagiousness of infected individuals, the susceptibility of exposed individuals, the nature of contact, the contact environment, and any infection reduction measures undertaken.^{11,14,15}

At the biological level, despite the virology and physiology of SARS-CoV-2 are yet to be fully understood, several papers have noted differences in infectivity between SARS-CoV-2 variants.¹⁶⁻¹⁸ Thus, Blanquart *et al.* (2021) observed a difference in viral loads between vaccinated and unvaccinated individuals for Delta, though this was statistically significant only for asymptomatic infections.¹⁹ The effectiveness of vaccinations against emerging variants is the subject of ongoing investigation^{11,18,20-23} but studies, such as Levine-Tiefenbrum *et al.* (2021), suggest lower viral loads during breakthrough infections for at least some types of vaccine.²⁰

At the behavioral level, physical proximity and duration of exposure influence the probability of infection,^{10,24,25,10,24} as does the activity undertaken – for instance, singing produces a significantly higher viral load than normal breathing.^{15,26,27} Wearing face masks or using shielding can reduce the risk of infection,^{10,28-32} with effectiveness depending inter alia on mask material and compliance.^{30,33}

At the environmental level, ventilation plays a critical role in air-borne SARS-CoV-2 transmission^{10,15,34-36} by directing, dispersing and diluting expelled infectious droplets/aerosols.³⁷ Outdoor settings with better natural air flows than indoor spaces can further reduce airborne viral transmission.³⁸

There are likely further risk factors. In particular, demographic influences are believed to be significant in determining infection outcomes,^{39,40} but since the relationship between demographic factors and of SARS-CoV-2 infection risk has not been fully quantified at the time of writing this paper, the proposed model does not take demographic factors into account.

Research Questions: Firstly, given M infectious persons and a total of N persons ($N-M$ exposed persons) in a room, we consider the number of exposed persons who are expected to become infected in a time period T . Secondly, we examine how partitioning a large group into smaller groups affects the average number of exposed persons becoming infected. Thirdly, we investigate how the probability of becoming infected is impacted by changes in each of the following conditions: the activity of the infectious persons and of the exposed persons (including breathing only, speaking, etc.), masking, vaccination, ventilation, viral load, and variant infectivity.

Novelties and Significance of Study: The proposed model has a simple and versatile multinomial probabilistic structure that can systematically capture known infection risk factors, including: group characteristics (group size); biological factors (infectivity of the variant, vaccinated/unvaccinated); human behaviours (duration of exposure, activities undertaken, masked/unmasked); and environmental conditions (ventilation). The model is readily extended to include additional epidemiological risk factors and its bottom-up structure allows a wide range of scenarios to be considered. Compared to earlier probabilistic models, the proposed model is more comprehensive, capturing cases with one or multiple infectious persons and incorporating more key risk factors affecting transmission and infection.

Our proposed model determines expected infection cases under different circumstances, offering, for example, quick and useful public health guidance on how to reduce infections

during large group gatherings. This is particularly important at the present time as cities increasingly seek to end social distancing policies and resume normal daily social and economic activities as much as possible, while minimizing the risks of large-scale infection and the resultant adverse health consequences.

2. Results

2.1 Basic model: M infectious persons in one single group gathering of N persons

Suppose there are N persons in a room with M infectious persons, i.e., there are N-M exposed (healthy) persons, with each infectious person making L contacts at random in a time period T, and infecting each exposed (healthy) person, with probability p for each contact. We are interested in I, the number of exposed persons newly infected. The formulae derived in Eq(1) and Eq(2) in the Methods section give the distribution of I and E(I), the expected average of I, respectively.

Figure 1a shows the average number of infections E(I) for group size N ranging from 1 to 100, for time duration (in hours) T = 1, 2, 3, and 4, and for p = 1. Here we have assumed that M = 1, i.e., there is only one infectious person, and that this infectious person will make one contact every 15 minutes. As expected, the average number of infected persons increases with the number of persons in the room and with T, until E(I) reaches a saturation point. In Figure 1a, for each T, the 10% saturation point is indicated with a diamond and the 5% saturation point is indicated with a circle. A 10% saturation point means the lowest N such that when N increases by 1, E(I) increases by less than 10%. Similarly, a 5% saturation point means the lowest N such that when N increases by 1, E(I) increases by less than 5%. When T = 1 (L = 4), 10% and 5% saturation occur at N = 9 and 11, respectively; when T = 4 (L = 16), 10% and 5% saturation points occur at N = 32 and 44, respectively. Table 1a shows the average number of infected persons when the size of the group increases from 50 to 500 for time durations from one to four hours.

Table 1b and Figure 1b-1d show that given the same duration T, as M/N increases, the number of new cases, E(I), due to incoming participants to the room who are infectious, also increases, until it reaches a peak. For T = 1, this peak occurs at M/N = 30% for N = 50, 100, and 500. For T = 2, 3, and 4, this peak occurs at M/N = 20% for N = 50, 100, and 500. Subsequent to the peak, E(I) decreases, as the number of exposed becomes smaller as M/N increases. When M/N is 100%, there are no exposed persons to be infected. Given the same group size, N, and an infectious percentage, M/N, a smaller T (e.g., T=1) always produces a lower infection number, E(I), as compared to a larger T (e.g., T=2, 3, 4).

Table 1a. E(I) vs N, when N = 50, 100, and 500, for T = 1, 2, 3, 4, M=1, and p=1.

No. of People in a Room (N)	E(I)			
	T = 1	T = 2	T = 3	T = 4
50	3.88	7.45	10.74	13.77
100	3.94	7.72	11.36	14.84
500	3.99	7.94	11.87	15.76

Table 1b. E(I) vs M/N, when N = 50, 100, and 500, for M/N=20% and 30%, T = 1, 2, 3, 4, M=1, and p=1.

M/N vs N	T=1	T=2	T=3	T=4
N=50; M/N=20%	22.47	32.31	36.63	38.52
N=100; M/N=20%	44.49	64.24	73.00	76.89
N=500; M/N=20%	220.70	319.63	363.97	383.85
N=50; M/N=30%	24.84	32.05	34.14	34.75
N=100; M/N=30%	49.30	63.88	68.19	69.46
N=500; M/N=30%	244.96	318.48	340.54	347.16

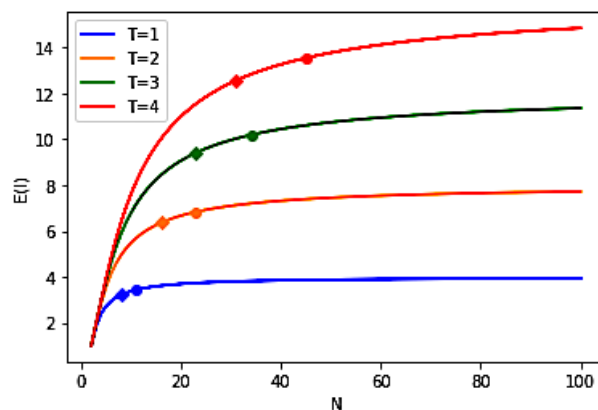


Figure 1a. E(I) vs N, for T = 1, 2, 3, and 4, M=1, and p=1 (Diamond indicates 10% saturation and circle indicates 5% saturation for a given curve).

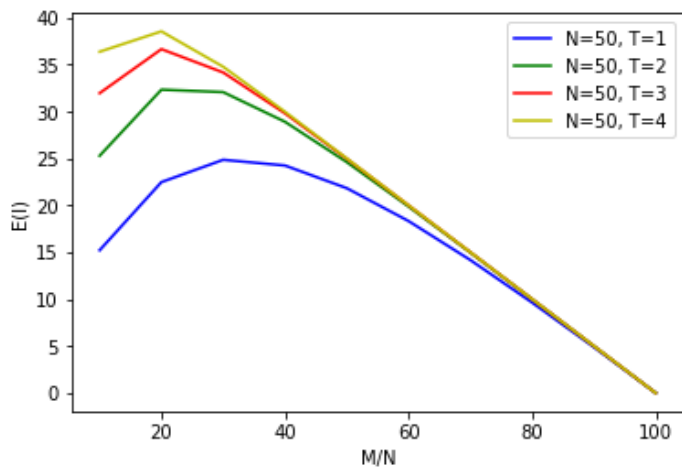


Figure 1b. $E(I)$ vs M/N , for $T = 1, 2, 3,$ and $4, N=50,$ and $p=1.$

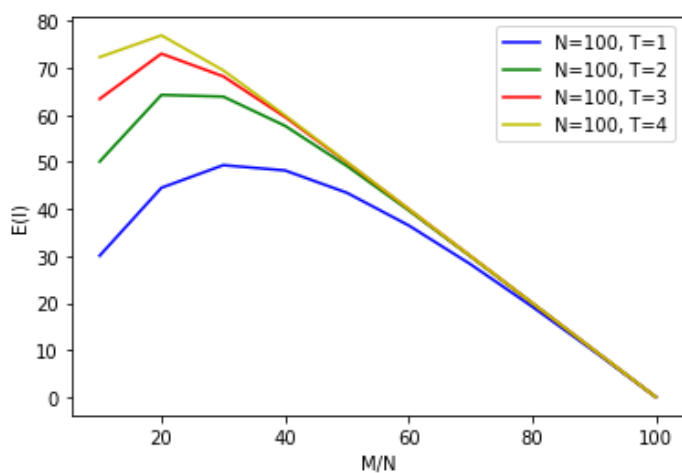


Figure 1c. $E(I)$ vs M/N , for $T = 1, 2, 3,$ and $4, N=100,$ and $p=1.$

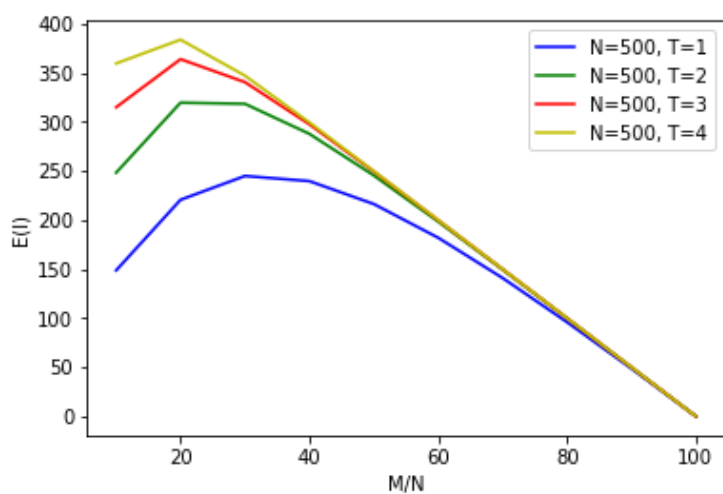


Figure 1d. $E(I)$ vs M/N , for $T = 1, 2, 3,$ and $4, N=500,$ and $p=1.$

2.2 Partition model: M infectious persons in one single indoor meeting with N persons partitioned at random into multiple rooms

We consider three different group gathering sizes, mimicking different real world social settings, such as classrooms, social gatherings, and conferences. The formulae derived in the Methods section allow us to determine the average number of persons who become infected when a single group gathering of size N with M infectious persons is partitioned at random into smaller groups and placed in separate rooms.

Consider a gathering of size $N = 50$ and the cases where infectious persons number $M = 5, 10, \text{ and } 15$ amongst these 50 persons, i.e., $M/N = 10\%, 20\%, \text{ and } 30\%$ respectively; as all individuals are in one room, we consider this as partition size 1, i.e. $PAR = 1$. Next, repeat but randomly partition the group of $N = 50$ into two separate rooms of 25, i.e. $PAR = 2$, and similarly for five separate rooms of 10, i.e. $PAR = 5$. Finally, for each case calculate $E(I)$, the expected number of persons who become infected. The results, together with those of larger groups $N = 100$ and $N = 500$, are shown in Table 2 (for $p = 1$) and demonstrate that partitioning has a very limited effect in reducing $E(I)$. The greatest reduction in $E(I)$ is achieved when the percentage of infectious persons M/N is 30% (Figure 2). $E(I)$ as a proportion of N declines slightly as group size N increases but this is highly constrained. Details of the methodology can be found in the Methods section, for the case $p = 1$.

Table 2. The expected number of people becoming infected $E(I)$ for $p = 1, T=1, M/N = 10\%, 20\% \dots, 100\%, N = 50, 100, 500$ and $PAR = 1, 2, 5$.

	No. of Partitions (PAR)	Partition Size	E(I) for $p = 1$									
			% Infectious (M/N)									
			10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Small Group Gathering $N = 50$	1	50	15.21	22.47	24.84	24.24	21.82	18.32	14.16	9.63	4.88	0
	2	25	14.78	21.92	24.32	23.79	21.46	18.05	13.98	9.55	4.87	0
	5	10	13.45	20.20	22.65	22.37	20.40	17.37	13.66	9.45	4.86	0
Medium Group Gathering $N = 100$	1	100	30.04	44.49	49.30	48.18	43.44	36.50	28.25	19.22	9.74	0
	2	50	29.63	43.96	48.79	47.75	43.09	36.24	28.05	19.10	9.71	0
	5	20	28.38	42.35	47.23	46.40	42.02	35.46	27.61	18.94	9.69	0
Large Group Gathering $N = 500$	1	500	148.72	220.70	244.96	239.72	216.37	181.99	140.96	95.96	48.65	0
	2	250	148.33	220.19	244.47	239.30	216.03	181.72	140.76	95.82	48.56	0
	5	100	147.14	218.66	242.99	238.02	215.00	180.93	140.17	95.42	48.43	0

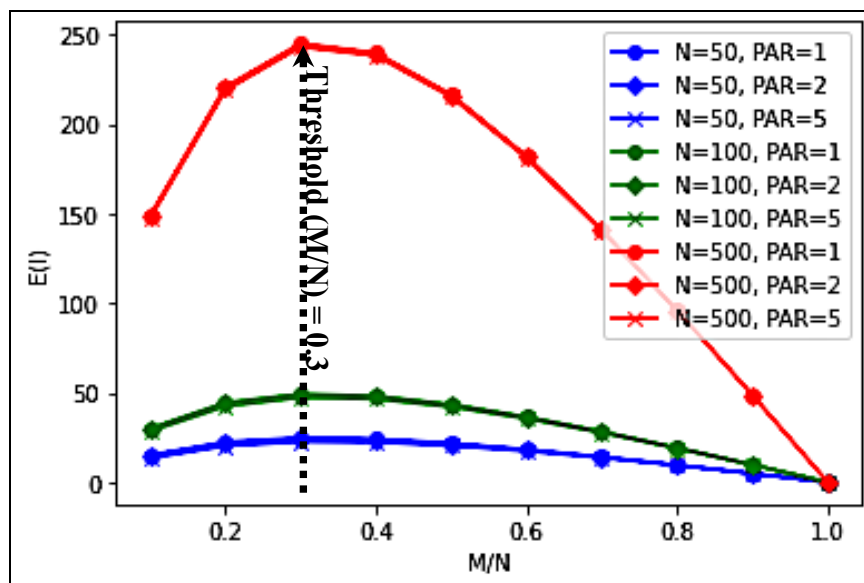


Figure 2. The expected number of people becoming infected for $p = 1$, $T=1$, $M/N = 10\%$, 20% , ..., 100% , $N = 50, 100, 500$ and $PAR = 1, 2, 5$.

2.3 Average number of people getting infected in an indoor setting under different values of infection probability p

Our model allows us to explore how $E(I)$ changes for different values of infection probability p , reflecting different social, environmental, and public health measures. Tables 3a-3b and Figures 3a-3b show the expected number of people infected for different group size N , partition size PAR , infection probability p and ratios of infectious persons to total group size M/N , when the duration of the gathering is 1 hour ($T = 1$). Peak $E(I)/N$ is relatively insensitive to group size and number of partitions. Peak $E(I)$ is [moderately] sensitive to M/N : for $N = 100$, $PAR = 1$, peak $E(I)$ occurs at $M/N = 50\%$ for $p = 0.1$ and $p = 0.3$, $M/N = 30\%$ for $p = 0.5$ and $p = 1$. The rate of change in $E(I)$ peaks when $p = 0.5$, for all group sizes and all number of partitions.

Table 3a. The expected number of people infected $E(I)$ for $p = 0.1, 0.3, 0.5$; $T = 1$; $M/N = 10\%$, 30% , 50% ; $N = 50, 100, 500$ and $PAR = 1, 2, 5$.

	No. of Partitions (PAR)	Partition Size	E(I) for $p = 0.1$			E(I) for $p = 0.3$			E(I) for $p = 0.5$		
			% Infectious (M/N)			% Infectious (M/N)			% Infectious (M/N)		
			10%	30%	50%	10%	30%	50%	10%	30%	50%
Small Group Gathering $N = 50$	1	50	1.80	4.04	4.62	5.20	10.79	11.47	8.35	16.08	16.04
	2	25	1.79	3.99	4.52	5.15	10.63	11.23	8.21	15.80	15.70
	5	10	1.76	3.86	4.26	4.97	10.13	10.57	7.78	14.92	14.79
Medium Group Gathering $N = 100$	1	100	3.57	7.99	9.15	10.29	21.37	22.75	16.50	31.87	31.84
	2	50	3.56	7.95	9.06	10.23	21.21	22.51	16.37	31.60	31.51
	5	20	3.53	7.82	8.77	10.06	20.73	21.78	15.97	30.75	30.53
Large Group Gathering $N = 500$	1	500	17.68	39.66	45.40	51.00	106.02	112.99	81.76	158.20	158.26
	2	250	17.67	39.61	45.31	50.94	105.86	112.75	81.63	157.93	157.94
	5	100	17.64	39.48	45.02	50.78	105.39	112.03	81.24	157.11	156.97

Table 3b. The expected number of people infected $E(I)$ for $p = 0.7, 0.9, 1.0$, $T = 1$, $M/N = 10\%, 30\%, 50\%$, $N = 50, 100, 500$ and $PAR = 1, 2, 5$.

	No. of Partitions (PAR)	Partition Size	$E(I)$ for $p = 0.7$			$E(I)$ for $p = 0.9$			$E(I)$ for $p = 1$		
			% Infectious (M/N)			% Infectious (M/N)			% Infectious (M/N)		
			10%	30%	50%	10%	30%	50%	10%	30%	50%
Small Group Gathering $N = 50$	1	50	11.25	20.24	19.07	13.94	23.49	21.08	15.21	24.84	21.82
	2	25	11.01	19.84	18.70	13.58	23.00	20.72	14.78	24.32	21.46
	5	10	10.26	18.60	17.67	12.45	21.46	19.64	13.45	22.65	20.39
Medium Group Gathering $N = 100$	1	100	22.24	40.13	37.90	27.54	46.60	41.95	30.04	49.30	43.44
	2	50	22.01	39.74	37.54	27.20	46.13	41.59	29.63	48.79	43.09
	5	20	21.31	38.56	36.46	26.14	44.68	40.50	28.38	47.23	42.02
Large Group Gathering $N = 500$	1	500	110.16	199.24	188.58	136.37	231.51	208.89	148.72	244.96	216.37
	2	250	109.93	198.87	188.23	136.04	231.05	208.54	148.33	244.47	216.03
	5	100	109.26	197.73	187.16	135.04	229.67	207.48	147.14	242.99	215.00

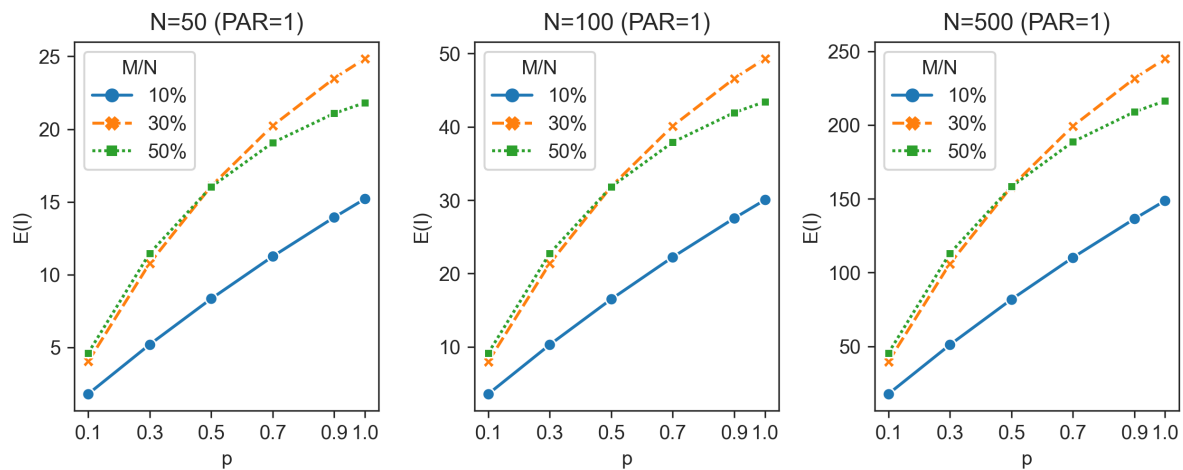


Figure 3a. The expected number of people infected $E(I)$ for $p = 0.1, 0.3, 0.5, 0.7, 0.9, 1.0$ (y-axis), under $T = 1$; group size $N = 50, 100, 500$; $PAR = 1$; $M/N = 10\%, 30\%, 50\%$

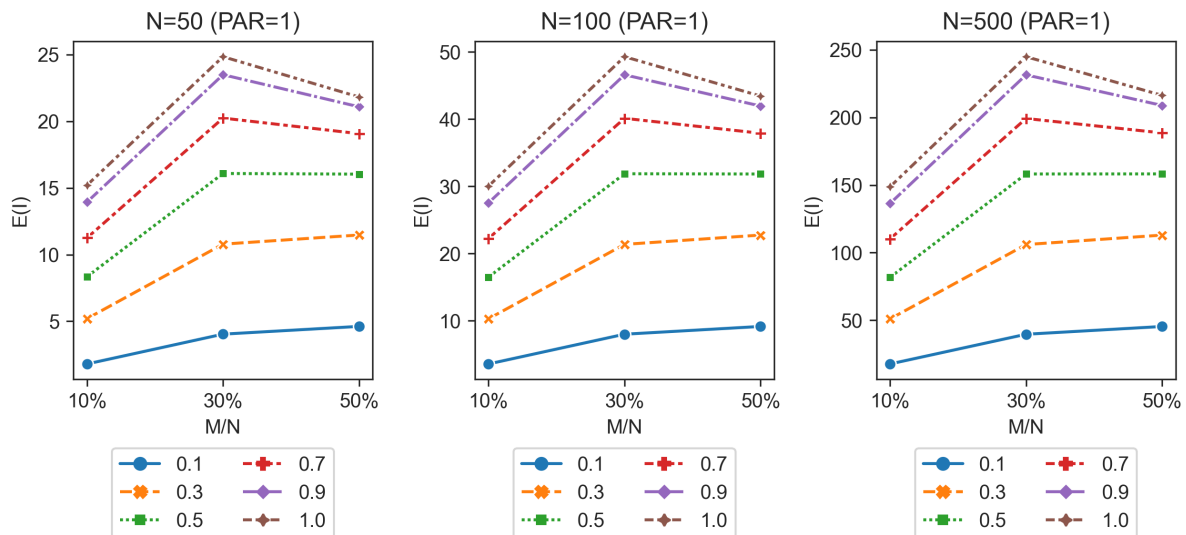


Figure 3b. The expected number of people infected $E(I)$ for $M/N = 10\%, 30\%, 50\%$ (y-axis), under $T = 1$; group size $N = 50, 100, 500$; $PAR = 1$; $p = 0.1, 0.3, 0.5, 0.7, 0.9$

2.4 The probability of infection p under different interaction activities and public health interventions

Table 4a shows the probability of infection p under four different indoor "activities", namely: (1) neither exposed persons nor infectious persons speak (i.e. they only breathe passively), (2) exposed persons do not speak but infectious persons do, (3) exposed persons speak but infectious persons do not speak, and (4) both exposed persons and infectious persons speak. Each scenario is considered under 8 different public health interventions (PHIs), reflecting different parameter values for Masking Factor (MF), Vaccination Factor (VAF), Ventilation Factor (VEF), and Variant Infectivity Factor (CVF). For each of the resultant 32 scenarios, the probability of infection, p , and the relative reduction in p (i.e. the ratio of p under that particular scenario to p under the applicable baseline scenario), p -ratio, are shown. For each activity, the first row shows the baseline scenario without any public health intervention, namely all persons do not wear masks ($MF = 1$), all persons have not been vaccinated ($VAF = 1$), the room has minimum mechanical ventilation ($VEF = 1$), the variant is Omicron ($CVF = 1$), and one "quanta" of the virus is inhaled ($Quantum = 1$). Subsequent rows reflect different PHIs and combinations of PHI, with scenarios h and k reflecting a hypothetical new SARS-CoV-2 variant that is 10x more transmissible than Omicron.

Our results show that for non-speaking activities, the probability of infection p can be substantially reduced from the baseline scenario (Scenario 1a) by 87.75% when three PHIs – masking, three-jab vaccination, and maximum ventilation – are implemented (Scenario 1e). The same measures lead to similar reductions in p of 78.39% from the baseline scenario (Scenarios 3a, 3e) in situations where exposed persons speak but infectious persons do not speak.

However, the effectiveness of these PHIs is severely compromised when infectious persons speak. When exposed persons do not speak but infectious persons do, the reduction in p from the baseline scenario (Scenario 2a) drops to 35.37% (Scenario 2e), while when both exposed persons and infectious persons speak, the reduction from the baseline scenario (Scenario 4a) drops to 5.33% (Scenario 4e), such that neither individual nor a combination of public health interventions lead to any meaningful reduction in the probability of infection.

We also consider the effect of individual PHIs on the probability of infection. Their ordinal ranking is consistent across activities, with maximum mechanical ventilation (MaxMV) having the greatest effect on reducing p , followed by masking (Mask), and vaccination (VAC). As above, these have a meaningful effect only when infectious persons do not speak. When infectious persons speak, then even MaxMV reduces p by only 1.83% (Scenario 2d) from the

baseline scenario (Scenario 2a) and when both infectious persons and exposed persons speak, then no PHI is effective (Scenarios 4a, 4d).

If variant infectivity (CVF) is increased ten-fold from Omicron, then even in the most favorable situation of infectious and exposed persons doing nothing but breathing, the strongest PHI combination (Mask plus Vac plus MaxMV) reduces p by just 12.48% (Scenario 1h), as compared to the Baseline (Scenario 1a) with no PHIs. In all other situations, PHIs cease to have any meaningful impact in reducing infections (Scenarios 2h, 3h, and 4h).

Table 4a. The p-ratio and reduction in p for different situations and public health interventions

1. Scenario 1: Exposed Persons and Infectious Persons – Both No Speaking										
Public Health Interventions (PHIs)	AFE	AFI	MF	VAF	VEF	CVF	Viral Load (Quanta)	p	p-ratio	Reduction in p due to PHIs
1a. Baseline	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.63	1.00	
1b. Mask	1.00	1.00	0.41	1.00	1.00	1.00	0.41	0.34	0.53	47.00%
1c. Vac	1.00	1.00	1.00	0.64	1.00	1.00	0.64	0.47	0.75	25.47%
1d. MaxMV	1.00	1.00	1.00	1.00	0.31	1.00	0.31	0.27	0.42	57.83%
1e. Mask+Vac+MaxMV	1.00	1.00	0.41	0.64	0.31	1.00	0.08	0.08	0.12	87.75%
1f. Mask+Vac+MedMV	1.00	1.00	0.41	0.64	0.51	1.00	0.13	0.12	0.20	80.40%
1g. Mask+Vac+NatV	1.00	1.00	0.41	0.64	0.82	1.00	0.21	0.19	0.30	69.70%
1h. Mask+Vac+MaxMV+NewVariant	1.00	1.00	0.41	0.64	0.31	10.00	0.81	0.55	0.88	12.48%
1i. Outdoor	1.00	1.00	1.00	1.00	0.03	1.00	0.03	0.03	0.05	94.65%
1j. Mask+Vac+Outdoor	1.00	1.00	0.41	0.64	0.03	1.00	0.01	0.01	0.01	98.59%
1k. Mask+Vac+Outdoor+NewVariant	1.00	1.00	0.41	0.64	0.03	10.00	0.09	0.09	0.14	86.47%
2. Scenario 2: Exposed Persons No Speaking and Infectious Persons Speaking										
Public Health Interventions (PHIs)	AFE	AFI	MF	VAF	VEF	CVF	Viral Load (Quanta)	p	p-ratio	Reduction in p due to PHIs
2a. Baseline	1.00	12.90	1.00	1.00	1.00	1.00	12.90	1.00	1.00	
2b. Mask	1.00	12.90	0.41	1.00	1.00	1.00	5.26	0.99	0.99	0.52%
2c. Vac	1.00	12.90	1.00	0.64	1.00	1.00	8.22	1.00	1.00	0.03%
2d. MaxMV	1.00	12.90	1.00	1.00	0.31	1.00	4.00	0.98	0.98	1.83%
2e. Mask+Vac+MaxMV	1.00	12.90	0.41	0.64	0.31	1.00	1.04	0.65	0.65	35.37%
2f. Mask+Vac+MedMV	1.00	12.90	0.41	0.64	0.51	1.00	1.71	0.82	0.82	18.15%
2g. Mask+Vac+NatV	1.00	12.90	0.41	0.64	0.82	1.00	2.74	0.94	0.94	6.44%
2h. Mask+Vac+MaxMV+New Variant	1.00	12.90	0.41	0.64	0.31	10.00	10.39	1.00	1.00	0.00%
2i. Outdoor	1.00	12.90	1.00	1.00	0.03	1.00	0.44	0.36	0.36	64.16%
2j. Mask+Vac+Outdoor	1.00	12.90	0.41	0.64	0.03	1.00	0.12	0.11	0.11	89.11%
2k.Mask+Vac+Outdoor+NewVariant	1.00	12.90	0.41	0.64	0.03	10.00	1.15	0.68	0.68	31.56%

3. Scenario 3: Exposed Persons Speaking and Infectious Persons No Speaking										
Public Health Interventions (PHIs)	AFE	AFI	MF	VAF	VEF	CVF	Viral Load (Quanta)	p	p-ratio	Reduction in p due to PHIs
3a. Baseline	2.82	1.00	1.00	1.00	1.00	1.00	2.82	0.94	1.00	
3b. Mask	2.82	1.00	0.41	1.00	1.00	1.00	1.15	0.68	0.73	27.31%
3c. Vac	2.82	1.00	1.00	0.64	1.00	1.00	1.80	0.83	0.89	11.30%
3d. MaxMV	2.82	1.00	1.00	1.00	0.31	1.00	0.87	0.58	0.62	38.03%
3e. Mask+Vac+MaxMV	2.82	1.00	0.41	0.64	0.31	1.00	0.23	0.20	0.22	78.39%
3f. Mask+Vac+MedMV	2.82	1.00	0.41	0.64	0.51	1.00	0.37	0.31	0.33	66.89%
3g. Mask+Vac+NatV	2.82	1.00	0.41	0.64	0.82	1.00	0.60	0.45	0.48	52.05%
3h.Mask+Vac+MaxMV+NewVariant	2.82	1.00	0.41	0.64	0.31	10.00	2.27	0.90	0.95	4.63%
3i. Outdoor	2.82	1.00	1.00	1.00	0.03	1.00	0.10	0.09	0.10	90.17%
3j. Mask+Vac+Outdoor	2.82	1.00	0.41	0.64	0.03	1.00	0.03	0.02	0.03	97.35%
3k.Mask+Vac+Outdoor+NewVariant	2.82	1.00	0.41	0.64	0.03	10.00	0.25	0.22	0.24	76.30%
4. Scenario 4: Exposed Persons and Infectious Persons – Both Speaking										
Public Health Interventions (PHIs)	AFE	AFI	MF	VAF	VEF	CVF	Viral Load (Quanta)	p	p-ratio	Reduction in p due to PHIs
4a. Baseline	2.82	12.90	1.00	1.00	1.00	1.00	36.38	1.00	1.00	
4b. Mask	2.82	12.90	0.41	1.00	1.00	1.00	14.84	1.00	1.00	0.00%
4c. Vac	2.82	12.90	1.00	0.64	1.00	1.00	23.17	1.00	1.00	0.00%
4d. MaxMV	2.82	12.90	1.00	1.00	0.31	1.00	11.28	1.00	1.00	0.00%
4e. Mask+Vac+MaxMV	2.82	12.90	0.41	0.64	0.31	1.00	2.93	0.95	0.95	5.33%
4f. Mask+Vac+MedMV	2.82	12.90	0.41	0.64	0.51	1.00	4.81	0.99	0.99	0.81%
4g. Mask+Vac+NatV	2.82	12.90	0.41	0.64	0.82	1.00	7.73	1.00	1.00	0.04%
4h. Mask+Vac+MaxMV+NewVariant	2.82	12.90	0.41	0.64	0.31	10.00	29.31	1.00	1.00	0.00%
4i. Outdoor	2.82	12.90	1.00	1.00	0.03	1.00	1.25	0.71	0.71	28.61%
4j. Mask+Vac+Outdoor	2.82	12.90	0.41	0.64	0.03	1.00	0.33	0.28	0.28	72.24%
4k. Mask+Vac+Outdoor+NewVariant	2.82	12.90	0.41	0.64	0.03	10.00	3.25	0.96	0.96	3.87%

Table 4b. E(I) for different activities and public health interventions, for fixed parameter values of group size N, duration T and ratio of infectious persons to total persons M/N

1. Scenarios for Exposed-Breathing and Infectious-Breathing Only										
Public Health Interventions (PHIs)	AFE	AFI	MF	VAF	VEF	CVF	Viral Load (Quanta)	p	E(I) w. N=50 (500), M/N=0.3, T=1	E(I) w. N=50 (500), M/N=0.5, T=1
1a. Baseline	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.632	18.94 (186.38)	18.17 (179.60)
1e. Mask+Vac+MaxMV	1.000	1.000	0.408	0.637	0.310	1.000	0.081	0.077	0.30 (2.93)	0.35 (3.48)
2. Scenarios for Exposed-Breathing and Infectious-Speaking Only										
2a. Baseline	1.000	12.90	1.000	1.000	1.000	1.000	12.90	1.000	24.8 (244.96)	21.8 (216.37)
2e. Mask+Vac+MaxMV	1.000	12.90	0.408	0.637	0.310	1.000	1.039	0.646	19.21 (189.11)	18.37 (181.55)
3. Scenarios for Exposed-Speaking and Infectious-Breathing Only										
3a. Baseline	2.820	1.000	1.000	1.000	1.000	1.000	2.820	0.940	24.05 (237.09)	21.40 (212.06)
3e. Mask+Vac+MaxMV	2.820	1.000	0.408	0.637	0.310	1.000	0.227	0.203	7.72 (75.82)	8.49 (83.57)
4. Scenarios for Exposed-Speaking and Infectious-Speaking Only										
4a. Baseline	2.820	12.90	1.000	1.000	1.000	1.000	36.39	1.000	24.8 (244.96)	21.8 (216.37)
4e. Mask+Vac+MaxMV	2.820	12.90	0.408	0.637	0.310	1.000	2.931	0.947	24.15 (238.04)	21.44 (212.59)

3. Discussion and Conclusions

3.1 Time spent indoors

For a given number of infectious persons M in a room with a total of N persons, the expected number of infected persons $E(I)$ always increases with time T spent in the room. All other factors being equal, shorter gatherings are preferred.

The incremental increase in $E(I)$ for an additional period varies with M/N . From Figures 1b-1d (reflecting $p = 1$), for $T = 1$, $E(I)$ initially increases with M/N until $E(I)$ peaks at $M/N = 30\%$ and then decreases for higher M/N . Dividing the curve into 3 sections – a left-hand section where $E(I)$ is increasing, peak $E(I)$, and a right-hand section where $E(I)$ is decreasing – it can be seen that as T increases, the left-hand section shifts up and flattens, peak $E(I)$ increases and shifts leftwards (i.e. occurs at lower values of M/N), and the right-hand section more rapidly approaches linear decline. As meetings become longer, there is less sensitivity as to whether a single infectious person or multiple infectious persons are initially present because the increase in exposure time tends towards saturation, such that everyone will be infected either way. This shifts policy emphasis; it is always desirable to have fewer infectious persons in a room but, whereas in shorter meetings there is benefit in reducing M/N , in longer meetings most benefit occurs from ensuring $M/N = 0$. As a consequence, longer meetings would benefit from more stringent control measures, such as prior PCR testing of individuals rather than rapid antibody tests (RAT), to ensure that $M = 0$, as well as smaller groups to limit the probability of a single infectious person entering the room – a risk that increases with group size.

Finally, the impact of N on $E(I)$ is highly constrained which can be seen visually from the similarity across Figures 1b, 1c and 1d for $N = 50, 100$ and 500 respectively.

3.2 Group partitioning

The effect of group partitioning or restricting group sizes is very modest. Referring to Table 3a, we illustrate this for $N = 100$ and $p = 0.5$. For $M/N = 30\%$, $E(I)$ falls from 31.87 for 1 partition ($PAR = 1$) to 30.75 when the group is divided into 5 partitions ($PAR = 5$), i.e., an absolute decrease of 1.12 or relative decrease of 3.51%. Intuitively, the effect is limited because partitioning does not directly stop infection. Instead, the effect is achieved via reducing the probability of finding a new person to infect for a given total group size: more partitions mean fewer people within each partition and hence a reduced maximum number of people that a given infectious person can infect, so saturation within a partitioned group occurs more quickly.

For cases where the infected proportion of the group is low, reaching saturation is relatively unlikely, so partitioning has little effect. Higher p causes saturation to be more likely, increasing the effect of partitioning. Thus, for $N = 100$ and $M/N = 10\%$, the relative decrease in $E(I)$ for $PAR = 1$ to $PAR = 5$ is 1.12% for $p = 0.1$ but increases to 5.53% for $p = 1$.

For cases where the infected proportion of the group is already high, reaching saturation is relatively likely, so partitioning again has little effect. Since lower p causes saturation to be less likely, this increases the effect of partitioning. Thus, for $N = 100$ and $M/N = 50\%$, the relative decrease in $E(I)$ for $PAR = 1$ to $PAR = 5$ is 4.15% for $p = 0.1$ but decreases to 3.27% for $p = 1$.

Notwithstanding the limited effectiveness of restricting group size, the use of such measures should consider the phase of infection cycle and probability of infection. For example, for a highly infectious variant ($p \rightarrow 1$), restricting group size is relatively more effective during the early phases of an outbreak, whereas for a less infectious variant ($p \rightarrow 0$), restricting group size is relatively more effective during later phases.

In the real-world, restrictions on group size have been amongst the most widely implemented PHIs to control the pandemic. Several studies have suggested that such measures have been effective in reducing transmission, with restrictions on small groups being more effective than on large groups. For instance, Sharma *et al.* (2021) found that restricting gatherings to 2 people was considerably more effective than restricting gatherings to 10 people, which in turn was slightly more effective than restricting gatherings to 30 people, with stricter mask wearing policies being more effective than restricting gatherings to 10 people.⁴¹ Haug *et al.* (2020) found that the most effective PHI was cancelling small gatherings (including gatherings of up to 50 people, closures of shops and restaurants, and mandatory home working).⁴² There are several reasons for the apparent discrepancy between these findings and our model. Most importantly, the studies do not distinguish between frequency and per event impact: even a very modest effect repeated many times can have a considerable aggregate impact. Although M/N , the proportion of individuals infected early in an outbreak is generally small (<0.01), meetings seldom occur with social distancing in place, so there can be a large impact as infection spreads through sequential meetings at low N . Additionally, the restrictions themselves (especially restricting gatherings to 2 people) are likely to affect people's behavior and reduce the number of gatherings (which is outside the scope of this model's one-shot structure).

Real-world studies also suggest that the distribution of transmission follows an exponential distribution rather than a uniform distribution. For example, Laxminarayan *et al.*

(2020) estimated that 5% of infected individuals accounted for 80% of cases in the Indian states of Tamil Nadu and Andhra Pradesh and Adam *et al.* (2020) found that 19% of cases caused 80% of local transmissions in Hong Kong.^{43,44} “Super-spreaders” likely reflect a combination of biological, behavioral and environmental factors. Interpreted in the framework of this model, super-spreaders may exhibit much greater viral shedding (it has been suggested that individuals infected by Omicron may exhale levels of virus 1000 times higher than the levels of Alpha or Delta strains⁴⁵) and be part of larger and more numerous partitions than typical individuals. Hence, earlier interventions that restrict group size N and movement (for example, geographically isolating areas or quarantining), may act to mitigate such cases and significantly reduce transmission, though relative effectiveness may decline as infection rates increase.

3.3 Speaking interactions

Table 4a shows that speaking has a large effect on the probability of infection p , especially if infectious persons speak. In the baseline scenario in which neither infectious persons nor exposed persons speak (Scenario 1a), viral load quantum = 1 (by definition) and $p = 0.63$. By comparison, viral load = 2.82 and $p = 0.94$ when only exposed persons speak (Scenario 3a) and viral load = 12.90 and $p = 1.00$ when only infectious persons speak (Scenario 2a). Since the effect of exposed persons speaking and the effect of infectious persons speaking interact multiplicatively, when both exposed and infectious persons speak (Scenario 4a), viral load = 36.38 and $p = 1.00$.

Under maximum PHI (masking, vaccination, and maximum ventilation), when both infectious persons and exposed persons speak (Scenario 4e), viral load is reduced to 2.93 and p is reduced to 0.95. The reduction in viral load is much larger than the reduction in infection probability, reflecting the viral load being far above the minimum needed to cause infection. This model treats infection as a binary variable but in practice there may be benefits to lowering the viral load, though the relationship between viral load and infection probability-severity is not yet definitively understood.¹²

For Omicron, infectious persons speaking essentially negates maximum PHI (viral loads and infection probability under Scenario 3e are similar to those of no speaking and no PHI under Scenario 1a) and both infectious and exposed persons speaking overwhelms maximum PHI (viral loads are approximately 3 times higher and probability of infection is approximately 50% higher under Scenario 4e than no speaking and no PHI under Scenario 1a).

This motivates widespread frequent Rapid Antigen Testing (RAT). Even with sensitivity rates of 70%, RAT and self-isolation (with considerably less than 100% compliance)

is likely to mitigate infection more than maximum levels of PHI in conversational settings (p declines from 1.00 with no PHI under Scenario 4a to 0.95 with maximum PHI under Scenario 4e). If public health authorities seek to control the spread of SARS-CoV-2, this may be the only alternative to remote living. The scope and appeal of restricting in-person conversation seems limited; conversation is intrinsic to human society and the nature and purpose of most in-person gatherings. Modern society is ill-adapted for monastic silence. For the relatively few settings in which communication is naturally one-way – for example, meet-and-greet receptions and lecturers/teachers in lecture halls/classrooms – it seems prudent for the speaking individuals to take additional precautions given the strong asymmetrical effect of an infectious person speaking, but as a policy, it seems highly questionable whether PHIs to permit gatherings provided verbal communication is one-way would be broadly practical or even desirable (in many cases, this would defeat the point of gathering).

3.4 Public health interventions and consideration of outdoor gatherings and potential future high-infectivity variants

As PHIs – masking, vaccination, and ventilation – combine multiplicatively, the greatest reduction in probability of infection p comes when all three measures are introduced. The most effective individual PHI is maximum ventilation (MaxMV), followed by wearing masks and then by vaccination. When there is no speaking (Scenario 1), baseline infection probability is 0.632 (Scenario 1a), falling to 0.077 under maximum PHI (scenario 1e), 0.267 under MaxMV (Scenario 1d), 0.335 under masking (Scenario 1b), and 0.471 under vaccination (Scenario 1c). As discussed previously, all interventions are greatly affected by speaking but this ordinal ranking is preserved (Scenarios 2-4). In the Supplementary Materials, we consider the threshold value of p such that exponential growth is avoided early in an infection cycle; for the assumptions here (notably $T = 1$), threshold p is approximately 0.3.

Improving ventilation for indoor environments warrants greater public health attention. This reduces infection probability by more than face masks or vaccinations and its effect is incremental to other PHIs. A similar alternative is the use of air filtration systems, such as portable High Efficiency Particulate Air (HEPA) cleaners,¹² though efficacy will depend on ventilation in the sense of air in the room being well mixed. Addressing ventilation is also likely less contentious, since this avoids many issues related to not wearing face masks or being vaccinated, and becomes even more important in the absence of masking or vaccination.

Switching from indoors venues to outdoors (Outdoor) has an effect similar to the impact of conversing. Baseline infection probability with no speaking and no PHI is 0.632 (Scenario

1a); with speaking and no PHI, infection probability increases to 1.00 (Scenario 4a); but with speaking and meeting outdoors, infection probability falls to 0.714 (Scenario 4i). With face masks and vaccination, meeting outdoors further reduces infection probability to 0.278 (Scenario 4j). Clearly, this is subject to weather and climate but when and where appropriate, for example in sub-tropical/tropical regions or in warmer months in temperate regions, moving outdoors is a quick, effective and low-cost measure.

Face masks are relatively affordable and now widely available but have met with resistance from a significant proportion of Western society and mandatory masking may be sufficiently publicly unpopular as to prevent the re-introduction of such measures.

Vaccines play a crucial role in mitigating the severity of illness but are less effective in preventing infection relative to other measures. Additionally, in the developing world, the cost of vaccines and limited supporting infrastructure mean vaccination may not be an option, whilst in Western countries, there are significant segments of society unwilling to be vaccinated.

Lastly, we observe that a hypothetical new variant with infectivity an order of magnitude greater than Omicron would pose serious challenges to PHIs. In a no speaking setting, Omicron has an infection probability of 0.63 with no PHI (Scenario 1a), whilst this hypothetical new variant would have an infection probability of 0.55 under maximum PHI (masking, vaccination, and maximum ventilation; Scenario 1h), assuming the practical challenges discussed above can be overcome. But if both infectious persons and exposed persons speak, then even moving outdoors will be insufficient ($p = 0.96$ under Scenario 4k).

3.5 Limitations

Practical application of the model would benefit from refining parameterization and the introduction of additional parameters, for example to distinguish between different types of masks or vaccines. The model structure can be extended to more accurately capture real-world dynamics. One important dimension is heterogeneity, for example to reflect different levels of contagiousness and susceptibility as well as non-uniform population distancing and contact times. Additionally, the model could be extended from a “one-shot” structure, which ends when people leave a room, to a “multi-shot” structure, which considers people leaving a room and entering new rooms.

3.6 Conclusions

We have proposed a multinomial probabilistic model to capture infection dynamics on a bottom-up individual basis. This helps to make the model relatively straightforward to

understand and easy to adjust for variations in parameters or additional parameters. As discussed in the Limitations section, this framework can be further developed but we believe it can provide policy-makers with quick and intuitive guidance to assist in scenario planning.

This model underscores the immense practical challenges of eliminating SARS-CoV-2, as high SARS-CoV-2 infectivity greatly raises the cost of allowing even a single infectious person to come into contact with uninfected individuals. In this regard, whilst frequent mass testing and self-isolation can help to control outbreaks, as numbers increase, the probability of avoiding such instances may become untenably small.

Improving ventilation – whether by making physical alterations, moving meetings outdoors or using air filtration systems – should be given much greater attention than at present. Compared to other public health interventions, ventilation can have a larger impact on reducing infection, is incrementally beneficial (not mutually exclusive), and is likely to be less challenged in implementation. This is particularly relevant as societies seek to resume normal human interaction while controlling the spread of SARS-CoV-2, as our model also suggests that in-person conversation has an order of magnitude impact on infection probability (in scenarios where this increase is not constrained by the new infection probability approaching 1) and on viral load.

Finally, our model suggests partitioning groups has a limited impact on reducing transmission within a gathering. To the extent that studies suggest restrictions on group size have had an impact, this suggests a large hidden cost since the channel is not via reduced spread per event (as commonly perceived) but rather by affecting the number of gatherings.

4. Methods

4.1 Basic model: M infectious persons in a room with N total persons, with each infectious person making L contacts at random in a time period T, and infecting each exposed (healthy) person with probability p for each contact

Suppose we have a room of N persons, M infectious persons, and N-M exposed (healthy) persons, labelled 1, 2, ..., N-M. The M infectious persons circulate among the N persons in a room, conducted over a duration T, and each infectious person will make L contacts with the others in the room at random, with the possibility of meeting another infectious person or with an exposed person more than once. Assuming infectious individuals will infect any exposed person with probability p each time a contact is made, what is E(I), the expected number of I, the number of exposed persons infected?

According to the US CDC²⁵, individuals who are within six feet of an infectious person, via a face-to-face contact, for a total of 15 minutes or more, will be at risk of infection. Hence, we selected 15 minutes as a time of interest. Hence, in a 1-hour meeting event every infectious person entering a room will have an opportunity of contacting four persons. Therefore, we assume if one infectious person moves around the room and talks to different people freely, at random, every 15 minutes, he will make four personal contacts. The infectious person may talk to the same person in the room repeatedly. We assume that the infectious person will infect each person with a probability p for every single personal contact, lasting for about 15 minutes.

4.1.1 Expected number of exposed subjects infected

Now we compute the probability of no one being infected after one contact. This single contact is equally likely to be caused by any of the M infectious persons, i.e., it is caused by infectious person 1 with the probability $1/M$. The infectious person 1 will cause zero infection if he/she meets another infectious person, with the probability $(M-1)/(N-1)$, or if he/she meets an exposed (healthy) person, with probability $(N-1-(M-1))/(N-1)$, and does not infect the exposed person, with the probability $(1-p)$. So the infectious person 1 not causing any infection is given by $(M-1)/(N-1) + (N-M)/(N-1) \times (1-p)$. Due to the independence between different infectious persons, the infectious person 2, 3, ..., M will not cause any new infection with the same probability. So in one contact the probability that no new infection is caused is as follows:

$$p_0 = \frac{1}{M} \left(\frac{M-1}{N-1} + \frac{N-M}{N-1} (1-p) \right) + \dots + \frac{1}{M} \left(\frac{M-1}{N-1} + \frac{N-M}{N-1} (1-p) \right)$$

$$= \frac{M-1}{N-1} + \frac{N-M}{N-1} (1-p)$$

The probability of each of the $N-M$ exposed (healthy) persons being infected should be the same as follows:

$$p_1 = \frac{1-p_0}{N-M}$$

Then we can use a multinomial distribution to determine how many of those exposed to SARS-CoV-2 are getting productively infected. In each of the ML contacts, we have $N-M+1$ outcomes: Exposed person 1 gets infected with probability p_1 , exposed person 2 gets infected with probability p_1, \dots , exposed person $(N-M)$ gets infected with probability p_1 , whilst no one

gets infected with probability p_0 . Note that here we consider only the infections obtained via this contact, without considering whether the exposed person has gotten infected before or not.

Next, we have a multinomial $(ML, p_0, p_1, \dots, p_1)$ distribution. Let X_0 be the number of times out of the ML contacts that no one gets infected, X_j be the number of times the exposed person j is infected, for $j = 1, \dots, N-M$. Then we have

$$P(X_0 = L_0, X_1 = L_1, \dots, X_{N-M} = L_{N-M}) = \frac{(ML)!}{L_0!L_1!\dots L_{N-M}!} p_1^{ML-L_0} p_0^{L_0} \quad \text{Eq(1)}$$

Now the number of exposed persons infected is $\sum_{j \neq 0} \text{Ind}(L_j > 0)$ where $\text{Ind}(\cdot)$ is the indicator function. The expected number of people getting infected is:

$$E(I) = \sum_{L_0 + \dots + L_{N-M} = ML} \left[\frac{(ML)!}{L_0! \dots L_{N-M}!} p_1^{ML-L_0} p_0^{L_0} \sum_{j \neq 0} \text{Ind}(L_j > 0) \right] \quad \text{Eq(2)}$$

Now we simplify Eq(2) for computational purpose. We separate the situations according to the different values of I . Let $P^i = \mathbb{P}(I=i)$.

First, consider the situation, when $L_1 = \dots = L_{N-M} = 0$. There is only one permutation, i.e., $L_0 = ML, L_1 = \dots = L_{N-M} = 0$ that satisfies this condition. From Eq(1),

$$\begin{aligned} P^0 &= \mathbb{P}(L_0 = ML, L_1 = \dots = L_{N-M} = 0) \\ &= (p_0)^{ML} \end{aligned}$$

Second, consider the situation, when $L_0 = ML - j, L_1 = j, L_2 = \dots = L_{N-M} = 0$,

$$\begin{aligned} \text{From Eq(1), } \mathbb{P}(L_0 = ML - j, L_1 = j, L_2 = \dots = L_{N-M} = 0) &= \frac{(ML)!}{(ML-j)!j!} p_1^j p_0^{ML-j} = \\ &= \binom{ML}{j} p_1^j p_0^{ML-j} \end{aligned}$$

Hence, summing over the possible values of j ,

$$\begin{aligned} P^1 &= \sum_{j=1}^{ML} \binom{ML}{j} p_1^j p_0^{ML-j} \\ &= \sum_{L_0 + L_1 = ML} \frac{(ML)!}{L_0! L_1!} p_1^{ML-L_0} p_0^{L_0} - P^0 \\ &= (p_1 + p_0)^{ML} - P^0 \end{aligned}$$

Then, given that $\mathbb{P}(L_k \neq 0 \text{ and } L_j = 0 \text{ for } j \neq k) = P^1$ for all ks .

Now, consider the situation, when $L_1 \neq 0, \dots, L_k \neq 0, L_{k+1} = \dots = L_{N-M} = 0$,

$$\begin{aligned} P^k &= \mathbb{P}(L_1 \neq 0, \dots, L_k \neq 0, L_{k+1} = \dots = L_{N-M} = 0) \\ &= \sum_{L_0 + \dots + L_k = ML} \frac{(ML)!}{L_0! L_1! \dots L_k!} p_1^{ML-L_0} p_0^{L_0} - P^0 \\ &= (kp_1 + p_0)^{ML} - C_k^1 P^{k-1} - \dots - C_k^k P^0 \\ &\text{for } 1 < k \leq \min(ML, N - M) \end{aligned}$$

If $ML < N - M$, at most ML people will get infected, hence, $P^j = 0$ for $j > ML$.

The expected infection number is:

$$E(I) = \sum_k P^k \times C_{N-M}^k \times k$$

4.1.2 Partition model: M infectious persons in one single group gathering of N persons, partitioned at random into R multiple rooms

When computing the expected number of exposed subjects infected for the partition model, we assume that each of the M infectious persons are equally likely to be in each of the R rooms and enumerate the different cases of how the M infectious persons are distributed in these R rooms. For example, suppose there are 5 rooms, the probability of each infectious person in each room is $1/5$. If the number of infectious people is 10, then the probability that all the 10 infectious people are in one room is $5 \times (1/5)^{10} = 1/5^9$.

To compute the expected number of exposed subjects infected, we compute the probabilities of all the enumerated cases, and the expected number of exposed subjects infected in each case. Then we can use the law of total expectation to get the result.

4.1.3 Probability of infection p

Our goal is to find the probability p that an exposed person will get infected when he/she is in contact with an infectious person. We want to factor in different scenarios/public health measures, such as (1) the activities conducted by the exposed person, and the infectious person, (2) masking, (3) vaccination, (4) ventilation, and (5) variant infectivity.

We start with Eq(3), developed in Buonanno *et al.*¹⁵, using a modified Wells-Riley equation:

$$p = 1 - e^{-IR \int_0^T n(t) dt} \tag{Eq(3)}$$

where IR = inhalation rate of the exposed person (m^3/h)

$n(t)$ = number of quanta/ m^3 at time t

T = time of exposure (h)

For relatively short exposure times, for instance, a few hours, we assume that $n(t)$ is a constant n , and Eq(3) can be reduced to:

$$p = 1 - e^{-IRnT} = 1 - e^{-no. \text{ of quanta inhaled}} \quad \text{Eq(4)}$$

From Eq(4), we can see that, by setting the number of quanta inhaled as one, one virus quantum is the dose of airborne droplet nuclei required to cause infection in $1 - e^{-1} = 63\%$ of exposed persons.

We can use viral quanta as the variable to study different scenarios of interest by properly modifying IR and n . IR will increase with the activity level of the exposed, while n (viral quanta/m³ in the air) will increase with the activity level of the infectious person and decrease when the infectious person is vaccinated. n will also decrease with ventilation and air filtering. Masking by either the exposed or the infectious person will also decrease the effective n as some of the virus will be blocked by masking.

Hence Equ. (4) can be modified as follow:

$$p = 1 - e^{-IR \times AFE \times MF \times nT \times AFI \times VAF \times VEF \times CVF} \quad \text{Eq(5)}$$

where IR has been modified by AFE , the activity factor of the exposed, while n , the number of viral quanta emitted by the infectious person is modified by AFI , the activity factor of the infectious person, VAF , the vaccination factor of the infectious person, VEF , the ventilation and filtering factor, and CVF , the SARS-CoV-2 variant factor. MF is the masking factor.

Next we study how to determine each of these modifying factors:

AFE: Buonanno *et al.* (2020)¹⁵ found the inhalation rate of five different activities of the exposed in an indoor setting, namely, resting, standing, light exercise, moderate exercise, and heavy exercise, averaged between males and females, to be 0.49, 0.54, 1.38, 2.35, and 3.30 m³/h, respectively. Thus, if we take resting as the baseline (with an *AFE* of 1), then the *AFE* for the five activities will be 1, 1.10, 2.82, 4.80, and 6.73, respectively. In Table 4, we assume the activity of speaking is similar to light exercise.

AFI: Coleman *et al.* (2021)²⁶ compared the different viral loads emitted by the infectious person when they are breathing, talking, and singing, and found viral loads of 1960, 25312, and 32831 virus gene copies, in a period of 30 minutes. Coleman *et al.* (2022) actually reported the viral loads for the latter two activities for a period of 15 minutes each. To match the value reported for the activity of breathing for 30 minutes, we doubled the reported values to get the 30-minute values. If we take breathing as the baseline, with $AFI = 1$, then the AFI for speaking and singing will be 12.9 and 16.8.

MF: Talic *et al.* (2021)⁴⁶ surveyed 72 studies that assessed the effectiveness of public health measures in reducing the incidence of SARS-CoV-2 transmission, and SARS-CoV-2 mortality, and based on data of 35 of such studies, which evaluated the effect of individual health measures, rather than a package of measures, estimated that the reduction of the incidence of SARS-CoV-2 with masking is 0.47. Taking no masking as the baseline, *MF* for masking is 0.408, which is obtained as the ratio of the estimated viral quanta with masking and without masking.

VAF: Accorsi *et al.* (2022)⁴⁷ estimated the C_t values for the Delta and Omicron variants of SARS-CoV-2, and found that for the Delta variant, the C_t values for unvaccinated and vaccinated (three doses) were 18.28 and 19.07, respectively. The corresponding numbers were 18.71 and 19.35 for the Omicron variant. Singanayagam *et al.*¹¹ quoted a conversion equation between C_t and viral load, published by Public Health England, as follow:

$$\text{Viral load (copies /ml)} = 133e^{(37.933 - C_t)/1.418}$$

If the unvaccinated is the baseline, with a *VAF* of 1, then we can easily calculate the *VAF* for vaccinated infectious person as 0.573 for the Delta variant, and 0.637 for the Omicron variant.

VEF: After a detailed study of the literature covering ventilation rates and SARS-CoV-2 infections in the US and EU/UK, we have decided to focus our ventilation setting on the classroom setting within the US (given the variation in ventilation rates over different environmental settings and building codes in different countries). ACH_T is used as the ventilation metric, defined as the rate at which room air is replaced by recirculated and outdoor air as determined by measuring the decay of a tracer species (such as CO_2). From McNeill *et al.* (2022)⁴⁸, we obtained the values for maximum mechanical ventilation rate (MaxMV) at ACH_T 8.7, minimum mechanical ventilation rate (MinMV) 2.7, median ventilation rate (MedMV) 5.3, and natural ventilation rate (NatV) 3.3 for the school classroom setting based in the US. Using the MinMV of 2.7 as the baseline, we can calculate the *VEF* for MaxMV, MedMV, and NatV as 0.310, 0.509, and 0.818, respectively. For outdoor ventilation (Outdoor), based on a thorough review of reported Covid-19 cases by Bulfone *et al.* (2021)³⁸, the odds of transmission indoors are 18.7 times higher than outdoors. Using Eq(5), the ratio of Viral Load or quantum, Q , between Outdoor and Indoor, is 0.0344, which is derived from the following equations:

$$\begin{aligned} p_{\text{indoor}}/p_{\text{outdoor}} &= 18.7 \\ (1 - e^{-Q_{\text{indoor}}}) / (1 - e^{-Q_{\text{outdoor}}}) &= 18.7 \end{aligned}$$

where

$$Q_{\text{indoor}} = \text{viral quanta indoors}$$
$$Q_{\text{outdoor}} = \text{viral quanta outdoors}$$

Using baseline of $Q_{\text{indoor}} = 1$ virus quantum, we have:

$$1 - e^{-1} = 18.7 - 18.7 e^{-Q_{\text{outdoor}}}$$
$$18.7 e^{-Q_{\text{outdoor}}} = 18.7 - 1 + e^{-1} = 18.068$$
$$e^{-Q_{\text{outdoor}}} = 0.9662; Q_{\text{outdoor}} = 0.0344$$
$$Q_{\text{outdoor}}/Q_{\text{indoor}} = 0.0344$$

CVF: It is well-known that different SARS-CoV-2 variants will have different levels of infectivity. A recent study by the University of Hong Kong¹⁷ found that that Omicron infects and multiplies 70 times faster than the Delta variant and the original coronavirus in the human bronchus, which may explain why Omicron may transmit faster between humans than previous variants. Thus, if the baseline of Omicron is assigned a CVF of 1, then the CVF of Delta will be 1/70. In the future, we may have variants that are even more infectious than Omicron, in which case the CVF of such a variant will be bigger than 1.

Based on these factors, we can compute p-ratio, the ratio of p compared to the baseline, for four sets of scenarios, as summarized in Table 4 above.

Acknowledgements

This research is supported in part by the Theme-based Research Scheme of the Research Grants Council of Hong Kong, under Grant No. T41-709/17-N.

Data Availability

All data used in this study are included in the article.

Code Availability

No custom code was used in the data analysis.

Author contributions statement

VL and JL were jointly responsible for full conceptualization and methodological design of a generalized multinomial probabilistic model for SARS-CoV-2 infection prediction, to cater for public health policy intervention evaluation; wrote the first draft; and acquired research funding. YS contributed partially to literature review, refined the model and performed computations. YH contributed partially to the draft in relation to the sensitivity of the results to infection probability, and suggested computational simplifications. KC edited the draft and contributed partially to the discussions and conclusions. SW assisted in reference formatting, contributed partially to literature review, and proposed the Wells-Riley equation for infection risk assessment. JC observed that ventilation is one of the most significant factors driving COVID infection, and this inspiration has guided some of the key arguments in this paper. JD inspired us to frame the risk of infection based on viral loads, and contributed partially to the discussion and the abstract. QZ performed computations based on the model developed by VL.

Additional information

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- 1 Han, Y. *et al.* Outdoor PM_{2.5} concentration and rate of change in COVID-19 infection in provincial capital cities in China. *Scientific Reports* **11**, 23206 (2021). <https://doi.org/10.1038/s41598-021-02523-5>
- 2 Viana, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **603**, 679-686 (2022). <https://doi.org/10.1038/s41586-022-04411-y>
- 3 Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657-663 (2022). <https://doi.org/10.1038/s41586-021-04385-3>
- 4 Keeton, R. *et al.* T cell responses to SARS-CoV-2 spike cross-recognize Omicron. *Nature* **603**, 488-492 (2022). <https://doi.org/10.1038/s41586-022-04460-3>
- 5 Crawford, F. W. *et al.* Impact of close interpersonal contact on COVID-19 incidence: Evidence from 1 year of mobile device data. *Sci Adv* **8**, eabi5499-eabi5499 (2022). <https://doi.org/10.1126/sciadv.abi5499>
- 6 Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257-261 (2020). <https://doi.org/10.1038/s41586-020-2405-7>
- 7 Petherick, A. *et al.* A worldwide assessment of changes in adherence to COVID-19 protective behaviours and hypothesized pandemic fatigue. *Nature Human Behavior* **5**, 1145-1160 (2021). <https://doi.org/10.1038/s41562-021-01181-x>
- 8 Leung, K., Wu, J. T. & Leung, G. M. Real-time tracking and prediction of COVID-19 infection using digital proxies of population mobility and mixing. *Nat Commun* **12**, 1501 (2021). <https://doi.org/10.1038/s41467-021-21776-2>
- 9 Roda, W. C., Varughese, M. B., Han, D. & Li, M. Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model* **5**, 271-281 (2020). <https://doi.org/10.1016/j.idm.2020.03.001>
- 10 Moritz, S. *et al.* The risk of indoor sports and culture events for the transmission of COVID-19. *Nature Communications* **12**, 5096 (2021). <https://doi.org/10.1038/s41467-021-25317-9>
- 11 Singanayagam, A. *et al.* Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The Lancet Infectious Diseases* **22**, 183-195 (2022). [https://doi.org/10.1016/s1473-3099\(21\)00648-4](https://doi.org/10.1016/s1473-3099(21)00648-4)
- 12 Tupper, P., Boury, H., Yerlanov, M. & Colijn, C. Event-specific interventions to minimize COVID-19 transmission. *Proceedings of the National Academy of Sciences* **117**, 32038-32045 (2020).
- 13 Goodwin, L. *et al.* Which factors influence the extent of indoor transmission of SARS-CoV-2? A rapid evidence review. *J Glob Health* **11**, 10002 (2021). <https://doi.org/10.7189/jogh.11.10002>
- 14 Karimzadeh, S., Bhopal, R. & Nguyen Tien, H. Review of infective dose, routes of transmission and outcome of COVID-19 caused by the SARS-CoV-2: comparison

- with other respiratory viruses. *Epidemiology and Infection* **149**, e116 (2021).
<https://doi.org/10.1017/S0950268821001084>
- 15 Buonanno, G., Stabile, L. & Morawska, L. Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment. *Environment International* **141** (2020). <https://doi.org/10.1016/j.envint.2020.105794>
- 16 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827 e819 (2020).
<https://doi.org/10.1016/j.cell.2020.06.043>
- 17 LKS Faculty of Medicine at The University of Hong Kong (HKUMed). *HKUMed finds Omicron SARS-CoV-2 can infect faster and better than Delta in human bronchus but with less severe infection in lung*,
<<https://www.med.hku.hk/en/news/press/20211215-omicron-sars-cov-2-infection>> (2021).
- 18 Kislaya, I. *et al.* Comparative Effectiveness of Coronavirus Vaccine in Preventing Breakthrough Infections among Vaccinated Persons Infected with Delta and Alpha Variants. *Emerg Infect Dis* **28**, 331-337 (2022).
<https://doi.org/10.3201/eid2802.211789>
- 19 Blanquart, F. *et al.* Characterisation of vaccine breakthrough infections of SARS-CoV-2 Delta and Alpha variants and within-host viral load dynamics in the community, France, June to July 2021. *Euro Surveill* **26** (2021).
<https://doi.org/10.2807/1560-7917.ES.2021.26.37.2100824>
- 20 Levine-Tiefenbrun, M. *et al.* Initial report of decreased SARS-CoV-2 viral load after inoculation with the BNT162b2 vaccine. *Nature Medicine* **27**, 790-792 (2021).
<https://doi.org/10.1038/s41591-021-01316-7>
- 21 Dean, N. E. & Halloran, M. E. Protecting the herd with vaccination. *Science* **375**, 1088-1089 (2022). <https://doi.org/doi:10.1126/science.abo2959>
- 22 Prunas, O. *et al.* Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel. *Science* **375**, 1151-1154 (2022).
<https://doi.org/doi:10.1126/science.abl4292>
- 23 Hayek, S. *et al.* Indirect protection of children from SARS-CoV-2 infection through parental vaccination. *Science* **375**, 1155-1159 (2022).
<https://doi.org/doi:10.1126/science.abm3087>
- 24 Jones, N. R. *et al.* Two metres or one: what is the evidence for physical distancing in covid-19? *BMJ* **370**, m3223 (2020). <https://doi.org/10.1136/bmj.m3223>
- 25 Centers for Disease Control and Prevention of the United States. *Interim Guidance on Developing a COVID-19 Case Investigation & Contact Tracing Plan: Overview*,
<<https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/appendix.html>> (2022).
- 26 Coleman, K. K. *et al.* Viral Load of SARS-CoV-2 in Respiratory Aerosols Emitted by COVID-19 Patients while Breathing, Talking, and Singing. *Clinical Infectious Diseases* (2021). <https://doi.org/10.1101/2021.07.15.21260561>
- 27 Sayampanathan, A. A. *et al.* Infectivity of asymptomatic versus symptomatic COVID-19. *The Lancet* **397**, 93-94 (2021). [https://doi.org/10.1016/s0140-6736\(20\)32651-9](https://doi.org/10.1016/s0140-6736(20)32651-9)
- 28 Riley, J., Huntley, J. M., Miller, J. A., Slaichert, A. L. B. & Brown, G. D. Mask Effectiveness for Preventing Secondary Cases of COVID-19, Johnson County, Iowa, USA. *Emerg Infect Dis* **28**, 69-75 (2022). <https://doi.org/10.3201/eid2801.211591>
- 29 Khosronejad, A. *et al.* Fluid dynamics simulations show that facial masks can suppress the spread of COVID-19 in indoor environments. *AIP Advances* **10**, 125109 (2020). <https://doi.org/10.1063/5.0035414>

- 30 Abaluck, J. *et al.* Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh. *Science* **375**, eabi9069 (2022).
<https://doi.org/10.1126/science.abi9069>
- 31 Dubois, B. *et al.* Clinical diagnosis of Alzheimer's disease: recommendations of the International Working Group. *The Lancet Neurology* **20**, 484-496 (2021).
[https://doi.org/10.1016/s1474-4422\(21\)00066-1](https://doi.org/10.1016/s1474-4422(21)00066-1)
- 32 Freeman, A. L. *et al.* Expert elicitation on the relative importance of possible SARS-CoV-2 transmission routes and the effectiveness of mitigations. *BMJ Open* **11**, e050869 (2021). <https://doi.org/10.1136/bmjopen-2021-050869>
- 33 Howard, J. *et al.* An evidence review of face masks against COVID-19. *Proceedings of the National Academy of Sciences* **118**, e2014564118 (2021).
<https://doi.org/doi:10.1073/pnas.2014564118>
- 34 Allen, J. G. & Ibrahim, A. M. Indoor Air Changes and Potential Implications for SARS-CoV-2 Transmission. *JAMA* **325**, 2112-2113 (2021).
<https://doi.org/10.1001/jama.2021.5053>
- 35 Guo, M. *et al.* Review and comparison of HVAC operation guidelines in different countries during the COVID-19 pandemic. *Building and Environment* **187** (2021).
<https://doi.org/10.1016/j.buildenv.2020.107368>
- 36 Blocken, B. *et al.* Ventilation and air cleaning to limit aerosol particle concentrations in a gym during the COVID-19 pandemic. *Building and Environment* **193** (2021).
<https://doi.org/10.1016/j.buildenv.2021.107659>
- 37 Schijven, J. *et al.* Quantitative microbial risk assessment for airborne transmission of SARS-CoV-2 via breathing, speaking, singing, coughing, and sneezing. *Environmental Health Perspectives* **129**, 047002 (2021).
- 38 Bulfone, T. C., Malekinejad, M., Rutherford, G. W. & Razani, N. Outdoor Transmission of SARS-CoV-2 and Other Respiratory Viruses: A Systematic Review. *The Journal of Infectious Diseases* **223**, 550-561 (2020).
<https://doi.org/10.1093/infdis/jiaa742>
- 39 Pijls, B. G. *et al.* Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies. *BMJ Open* **11**, e044640 (2021).
<https://doi.org/10.1136/bmjopen-2020-044640>
- 40 Venkatesan, P. The changing demographics of COVID-19. *The Lancet Respiratory Medicine* **8** (2020). [https://doi.org/10.1016/s2213-2600\(20\)30461-6](https://doi.org/10.1016/s2213-2600(20)30461-6)
- 41 Sharma, M. *et al.* Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. *Nature Communications* **12** (2021).
<https://doi.org/10.1038/s41467-021-26013-4>
- 42 Haug, N. *et al.* Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behavior* **4**, 1303-1312 (2020).
<https://doi.org/10.1038/s41562-020-01009-0>
- 43 Laxminarayan, R. *et al.* Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691-697 (2020).
- 44 Adam, D. C. *et al.* Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine* **26**, 1714-1719 (2020).
- 45 Lai, J. *et al.* Evolution of SARS-CoV-2 Shedding in Exhaled Breath Aerosols. *medRxiv* (2022).
- 46 Talic, S. *et al.* Effectiveness of public health measures in reducing the incidence of Covid-19, SARS-CoV-2 transmission, and Covid-19 mortality: systematic review and meta-analysis. *BMJ* **375**, e068302 (2021). <https://doi.org/10.1136/bmj-2021-068302>

- 47 Accorsi, E. K. *et al.* Association between 3 doses of mRNA COVID-19 vaccine and symptomatic infection caused by the SARS-CoV-2 Omicron and Delta variants. *JAMA* **327**, 639-651 (2022).
- 48 McNeill, V. F. *et al.* Room-level ventilation in schools and universities. *Atmospheric Environment: X* **13**, 100152 (2022).
- 48 Lindsley *et al.* Efficacy of Portable Air Cleaners and Masking for Reducing Indoor Exposure to Simulated Exhaled SARS-CoV-2 Aerosols - United States, 2021. *MMWR Morb Mortal Wkly Rep.* **70**(27):972-976 (2021), doi:10.15585/mmwr.mm7027e1.
- 49 Dadras *et al.* The relationship between COVID-19 viral load and disease severity: A systematic review *Immun Inflamm Dis.* 10(3): e580 (2022), doi:10.1002/iid3.580