

**Supplementary Material for
A coupled Computational Fluid Dynamics and Wells-Riley model to predict COVID-19
infection probability for passengers on long-distance trains**

Zhaozhi Wang, Edwin R Galea*, Angus Grandison, John Ewer, Fuchen Jia

**Fire Safety Engineering Group, University of Greenwich, Old Royal Naval College,
30 Park Row, Greenwich, LONDON SE10 9LS, UK**

***Corresponding Author: E.R.Galea@gre.ac.uk**

Note in this supplementary material document, reference to [S1] will be abbreviated to Main Paper or MP. Furthermore, all figures, equations, tables and sections from the MP will be prefixed by MP when cited in this document. Similarly, sections, figure and tables within this supplementary material document are referred to as SM section S2.1 or SM Figure S2.

S1 Calculations of proximity-index-patient based probability and saloon average probability

S1.1 Seating location of index patients

In the supplementary material [S2] to Hu et al. [S3], the number of confirmed SARS-CoV-2 cases (227), and the number of susceptibles seated at various distances from the index patients, are presented in Table S1 of Hu et al [S2]. This suggests that there were 2605, 1996, 1845, 1825 and 1028 susceptible passengers seated in the same row but adjacent to the index patient: two-, three-, four- and five- seats away from the index patient, respectively. In their analysis, the cabin aisle is treated as a seat when the distance between a susceptible and an index patient is measured. Let the number of index patients at Seat A, B, C, D, and F be a , b , c , d , and f respectively. There are 2605 susceptibles within the same row and seated adjacent to the index patients. Then, if the index patient is seated in the:

- A seat, there is one adjacent seat (B).
- B seat, there are 2 adjacent seats (A and C).
- C seat, there is only one adjacent seat (B; note D is not adjacent to C).
- D seat, there is only one adjacent seat (F; note C is not adjacent to D).
- F seat, there is only one adjacent seat (D).

As there are 2605 susceptibles adjacent to index patients, this means:

$$a + 2b + c + d + f = 2605$$

There are 1028 susceptibles in the same row of the index patients but seated five-seats away. Then if the index patient is seated in the:

- A seat, there is only one location 5 columns from A (F).
- B, C, or D seats, there are no locations 5 columns from these seats.
- F seat, there is only one location 5 columns from F (A).

$$a + f = 1028$$

Similarly, the following set of equations can be derived:

$$a + 2b + c + d + f = 2605$$

$$\begin{array}{rclclcl}
a & + & & 2c & + & d & = & 1996 \\
& & b & + & c & + & d & + & f & = & 1845 \\
a & + & b & + & & & d & + & f & = & 1825 \\
a & + & & & & & & & f & = & 1028
\end{array}$$

The approximate integer solution of the above equations is $a=480$; $b=280$; $c=500$; $d=517$; $f=548$. This results in the following:

- SEAT A: 480 index patients located on this seat.
- SEAT B: 280 index patients located on this seat.
- SEAT C: 500 index patients located on this seat.
- SEAT D: 517 index patients located on this seat.
- SEAT F: 548 index patients located on this seat.

The solution assumes that all the seats were occupied. This assumption is probably valid as the time period for the data collection is during the peak travel period corresponding to Chinese New Year. The total number of index patients is given by $a + b + c + d + f = 2325$. However, the paper reported that there were 2334 index patients on the G-train. Furthermore, it is reported that there were 234 secondary case on the G-train, but from Table S1 of Hu et al. [S2], there are only 227 listed. Also, there are 71,531 close contacts listed in Table S1 Hu et al. [S2], but it is reported that there were 72,093 close contacts. It is therefore assumed that the other index patients, secondary cases and close contacts were not travelling within the carriages used in the analysis.

S1.2 Proximity based probabilities

Hu et al. [S3] proposed infection probabilities (IPs) based on proximity to the index patient. The MP [S1] attempts to predict these probabilities using coupled Wells-Riley and computational fluid dynamics modelling (WR-CFD). For the WR-CFD IPs to equate to the IPs presented in Hu et al. [S3], five proximity-based IPs must be assessed. These are the IPs associated with the following situations: adjacent in the same row, not adjacent but in the same row, one row away, 2 rows away and 3 rows away from the index patient.

It is assumed that the modelled index locations 6A, 6B, 6C, 6D, and 6F, are representative of all seats in seat columns A, B, C, D, and F respectively (see SM section S1.3). This assumption implies that the average probability within one particular row will be the same for any other row.

Adjacent:

The average IP due to being adjacent to an index case is given by:

$$P_{average}^{adjacent} = W_B P_{B \rightarrow A} + (W_A P_{A \rightarrow B} + W_C P_{C \rightarrow B}) + W_B P_{B \rightarrow C} + W_F P_{F \rightarrow D} + W_D P_{D \rightarrow F}$$

$$W_X = x / (a + 2b + c + d + f)$$

Where $x=a, b, c, d, f$; $P_{x \rightarrow y}$ is the probability of an index case in location X infecting a susceptible at Y and W_X is the relative weight of the index at location X (See Table S1).

On same row but not adjacent:

$$P_{average}^{non-adjacent} = (W_C P_{C \rightarrow A} + W_D P_{D \rightarrow A} + W_F P_{F \rightarrow A}) + (W_D P_{D \rightarrow B} + W_F P_{F \rightarrow B}) \\ + (W_A P_{A \rightarrow C} + W_D P_{D \rightarrow C} + W_F P_{F \rightarrow C}) + (W_A P_{A \rightarrow D} + W_B P_{B \rightarrow D} + W_C P_{C \rightarrow D}) \\ + (W_A P_{A \rightarrow F} + W_B P_{B \rightarrow F} + W_C P_{C \rightarrow F})$$

$$W_x = x / (3a + 2b + 3c + 3d + 3f)$$

Being *n* rows away:

$$P_{average}^{n-row} = \sum_{\dot{Y}=A..F} \sum_{X=A..F} W_X P_{X \rightarrow \dot{Y}}$$

$$W_x = 0.1x / (a + b + c + d + f)$$

Where \dot{Y} is a seat *n* rows (ahead and behind) away from the row of index cases.

Table S1. Weights of individual probabilities for calculation of proximity-based probabilities.

	Weights				
	W _A	W _B	W _C	W _D	W _F
Adjacent	0.184	0.107	0.192	0.198	0.210
Same row as the index patient but not adjacent	0.072	0.042	0.075	0.077	0.082
One, or two, or three rows from the index patient	0.021	0.012	0.022	0.022	0.024

S1.3 Independence of seat row

Throughout the analysis in the MP it is implicitly assumed that the seats in row 6 are representative of all the rows. To verify this assumption, additional CFD simulations were conducted, with the index patient located in row 13. Two index seating locations were selected, one with the index patient located in the D seat and another with the index patient located in the F seat.

As in the MP Section 4, the IPs were determined for an 8-hour exposure assuming that the index patient is located in the identified seats for the following five spatial separations from the index patient: i.e. adjacent seat; same seat row excluding adjacent; one-; two-; and three-seat rows away. The resulting IPs for a lone index patient located in seat 6D or 13D, and seat 6F or 13F are presented in Table S2.

Table S2: Predicted IPs during 8-hour exposure for an index patient in either 6D or 13D and 6F or 13F.

IP (%)	Index patient seat and row location			
	6D	13D	6F	13F
Adjacent	10.6	10.9	16	20.1
Same row	0.5	0.5	0.5	0.5
1 row away	2.84	2.88	3.63	3.59
2 rows away	1.37	1.45	1.65	1.67
3 rows away	0.81	0.95	0.92	1.05

As can be seen, the IPs are very similar irrespective of the seat row, resulting in similar conclusions concerning susceptibility of surrounding passengers. This supports the assumption that results are independent of seat row. However, it should be noted that the ends of the saloon are likely to produce regions of stagnant air potentially increasing the IP for those seated there, especially if the index patient is also located in the end region.

S2 Quanta generation rate

Within the analysis presented in the MP, the IP for COVID-19 on G-trains is dependent on the quanta generation rate, which is unknown. It is thought to be dependent on the variant of the SARS-CoV-2 virus causing the infection, the degree of infection of the index patient and the nature of the activity they are involved in, e.g. breathing while resting, breathing while involved in light/heavy activity, vocalisation volume, etc. As a result, a wide variety of quanta generation rates are suggested or derived from data for various COVID-19 transmission events reported in the literature [S4-S9] varying from the order 10^0 to 10^2 quanta/h. Passengers seated on a train are assumed to be at a low activity level, generally in the 'rest state'. Furthermore, quanta generation rates based on observational data must take into account the time and location from when they are derived as the virus variant involved will have an impact on quanta generation rates and hence suitability for use in simulations. The analysis in the MP refers to infection incidents in the beginning stages of the pandemic as do the sources quoted for the quanta generation rates.

The quanta generation rate used in the IP analysis in the MP is 14 quanta/h. This value is based on the minimum quanta generation rate determined in a study of several COVID-19 infection events in public spaces such as classrooms within China [S4]. Further support for the use of this quanta generation rate is derived from an analysis of epidemiological data of COVID-19 cases and close contacts among 9,265 airline passengers on 291 airplanes conducted by Hu et al. [S10]. This work suggests that the seat adjacent to the index patients had the highest risk, with an IP of 5.7%-14.4% (average of 9.2%), for 177 flights departing from Wuhan (China) with an average travel time of 2 hours. A theoretical study of inflight transmission of COVID-19 based on experimental dispersion data suggests that maximum calculated IPs, of between 4.5% and 16.5%, are produced in a 2-hour flight with quanta generation rates of 5 quanta/h and 20 quanta/h, respectively [S11]. This suggests that quanta generation rates of 5-20 quanta/h may be representative of those at the time of the investigated inflight transmission events [S10]. As the investigated inflight transmission events [S10] occurred over the same time period as the train transmission events [S3], the identified quanta generation rates of 5-20 quanta/h may also be appropriate for the investigated train transmission events. However, it is noted that the identified quanta generation rates assume that disease transmission is solely due to aerosols. Contact transmission and droplet transmission are not considered. Therefore, if disease transmission were actually caused by a combination of these other modes, the identified quanta generation rates would be less. There is no definitive evidence showing that aerosol/airborne transmission was the only cause of the investigated inflight infections [S10] but sources, such as the Centers for Disease Control (CDC) in the USA, now suggest that aerosol transmission is likely to be the dominant form of transmission [S12, S13] with fomite transmission representing less than a 1 in 10,000 risk of infection [S12, S14, S15]. Furthermore, the statistical inflight IPs are based on the assumption that the index patient and secondary infections were independent travellers [S10].

Further evidence supporting 14 quanta/h generation rate is derived from Buonanno et al. [S5]. They defined a number of quanta emission rate frequencies for breathing (resting), speaking (light activity), singing/speaking loudly (light activity). The speaking and speaking loudly

emission rates are adjusted for a resting breathing rate by multiplying by a scaling factor of 0.3551 (resting breathing rate (0.49 m³/h) divided by the light activity breathing rate (1.38 m³/h)). The adjusted rates are presented in Table S3.

Table S3. The percentile distribution of quanta emission rates (quanta/h) for breathing, speaking, and singing as defined by Buonanno et al. and adjustments for speaking distributions while at rest.

Quanta emission rate type	5 th	25 th	50 th	75 th	90 th	95 th	99 th
Breathing	0.024	0.12	0.37	1.1	3.1	5.7	17
Speak(original)	0.32	1.6	5.0	15	42	76	240
Speaking*	0.11	0.57	1.8	5.3	14	26	85
Singing/speaking loudly (original)	2.11	10	32	98	270	490	1500
Speaking* (loud)	0.75	3.6	11.4	34.8	95.9	174	532

*the original values were based on light activity; these have been adjusted for a resting breathing rate.

From this study, there is a wide range of quanta generation rates while at rest, which vary from 0.024 quanta/h to 532 quanta/h (see Table S3). The 14 quanta/h quanta generation rate adopted in [S1] is:

- Between the 95th and 99th percentiles of breathing at rest,
- Representative of the 90th percentile while normal speaking at rest,
- Between the 50th and 75th percentile while speaking loudly at rest.

Finally, it is also noted that the quanta generation rates, reviewed here, are specific to early variants of SARS-CoV-2. Newly reported variants of SARS-CoV-2, such as the B.1.1.7 mutation [S16] – which until early 2021 was the dominant strain of the virus in the UK [S17], and reported to have been detected in 55 other countries [S18] around that time, and the B.1.351 variant first detected in South Africa, are reported to be more transmissible [S17]. The B.1.1.7 variant is reported to be between 40% and 70% more transmissible than the original variant [S16]. At the time of writing (June 2021), the dominant strain of the virus within the UK is the B.1.617.2 variant first detected in India [S19] and this is reported to be between 26% and 115% more transmissible than the B.1.1.7 variant [S20, S21] with PHE suggesting an average of 64% more transmissible [S21]. Quanta generation rates associated with these strains are likely to be significantly greater than previously reported for the original SARS-CoV-2 variant. It is noted that Burridge et.al. have suggested that scaling the quanta generation rate by the transmission increase factor is one approach of dealing with more transmissible variants [S22].

S3 Derivation of Equation (9) in the MP

Here MP-equation (9) used for the calculation of the quanta concentration at the inlet of the saloon ventilation system at steady state is derived. According to the quanta transport paths in MP-Fig. 4, the quanta concentration at the inlet is:

$$C = \frac{\alpha y(1-\eta)}{q} \quad (S1)$$

This assumes that q is the quanta generation rate from an index patient. At steady state (i.e. total quanta rate into the saloon = total quanta rate out of the saloon), it will be the case that:

$$\alpha y(1 - \eta) + q = y \quad (S2)$$

From Equation (S2),

$$y = \frac{q}{1 - \alpha + \alpha \eta} \quad (S3)$$

Combining Equations (S1) and (S3), results in Equation (S4), which is MP-equation (9):

$$C = \frac{\alpha q(1 - \eta)}{(1 - \alpha + \eta \alpha) Q} \quad (S4)$$

S4 Mesh sensitivity analysis

Before undertaking a mesh sensitivity analysis, an appropriate time step size is determined. Using a mesh budget of 2,245,320 (378×60×99) cells, three different time step sizes were considered, 0.25 s, 0.5 s and 1.0 s. It is noted that at the quasi steady state, the simulations using the three time step sizes produce almost identical quanta concentration distributions in the horizontal cutting plane at nose height (see Fig. S1). However, given the high flow speeds generated by the ventilation, the 0.5 s time step was adopted for the analysis presented in this paper to improve computational stability.

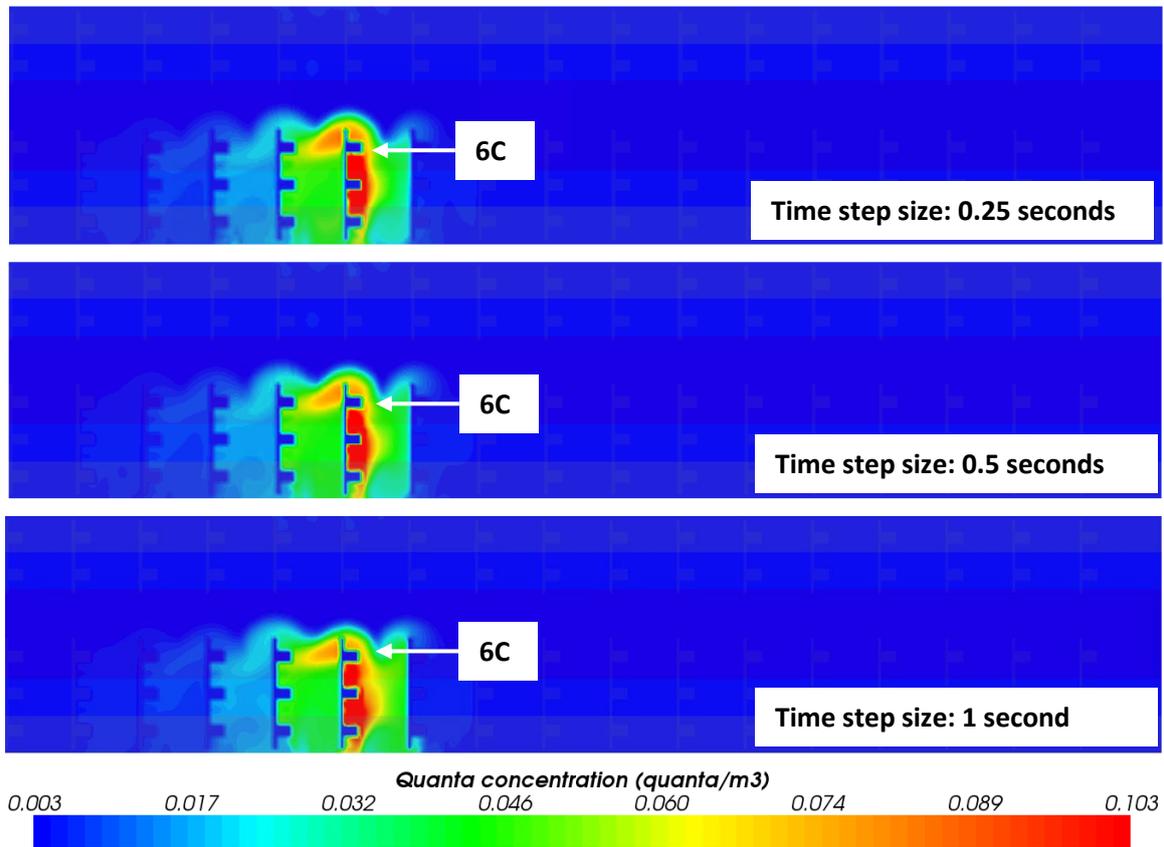
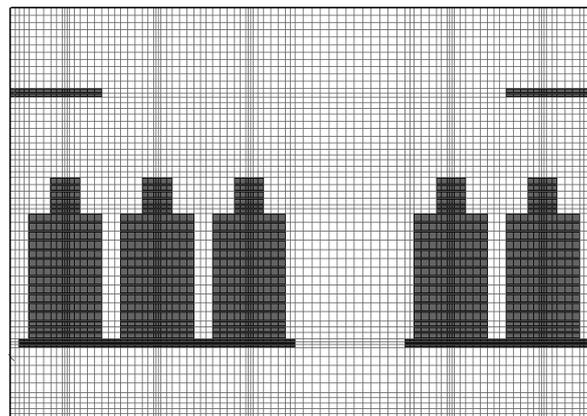
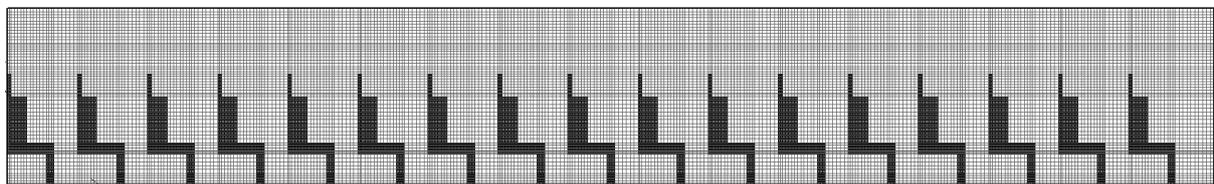


Fig. 1 Predicted quanta concentrations (quanta/m³) in a horizontal cutting plane at nose height produced using three computational time steps (0.25 s, 0.5 s and 1.0 s).

To assess the mesh sensitivity of the CFD simulations, three mesh density cases were considered for ventilation scenario 1: coarse, with 1,095,444 ($378 \times 46 \times 63$) cells; medium with 2,245,320 ($378 \times 60 \times 99$) cells and fine, with 5,091,240 ($399 \times 88 \times 145$) cells. Fig. S2 shows the mesh distribution within the computational domain for the medium mesh density case. A structured mesh was used to model the interior of the saloon. The mesh was divided into several blocks of cells describing various components within the geometry e.g. seats, passengers, etc. The cell distribution within each block of cells was uniform. A number of preliminary simulations were undertaken to establish that good convergence could be achieved for the structured mesh when relatively fine meshes were used in blocks around passengers' mouth/nose; around the seat bottom and seat back. The cell sizes in the medium mesh density case (the mesh used in the simulations presented in the paper) are between 0.013 m and 0.05 m.



(a) Cross section view



(b) Longitudinal view

Fig. S2 Saloon mesh distribution for the medium mesh density case (a) Cross section view and (b) Longitudinal view.

The additional cells introduced into the medium and fine mesh cases are primarily distributed through the width and height of the saloon. The mesh sensitivity analysis focuses on the variation of the flow pattern, quanta distribution, velocity profile and IP as the mesh density increases. As seen in Fig. S3(a-c), the three different mesh resolutions produce quite similar air circulation flows. Furthermore, the simulated quanta concentration distributions are similar, as shown in the three mesh cases (see Fig. S4(a), Fig. S4(b) and MP Fig. 8).

The longitudinal-, vertical- and transverse- velocity components at the horizontal line through the target volume of seat 6A, 6B, 6C, 6D and 6F are depicted in Fig. S5. The velocity along the longitudinal and vertical directions for the three mesh cases are quite similar (see Fig. S5(a) and S5(b)). However, there are relatively large differences in transverse velocity at 0.75 m and 2.75 m locations among the three mesh cases (see Fig. S5(c)). These two locations are in the vicinity of the vortices of the recirculation regions as seen in Fig. S3. Clearly, the medium

mesh provides a closer agreement to the structure and location of these vortices as predicted by the fine mesh than that provided by the coarse mesh.

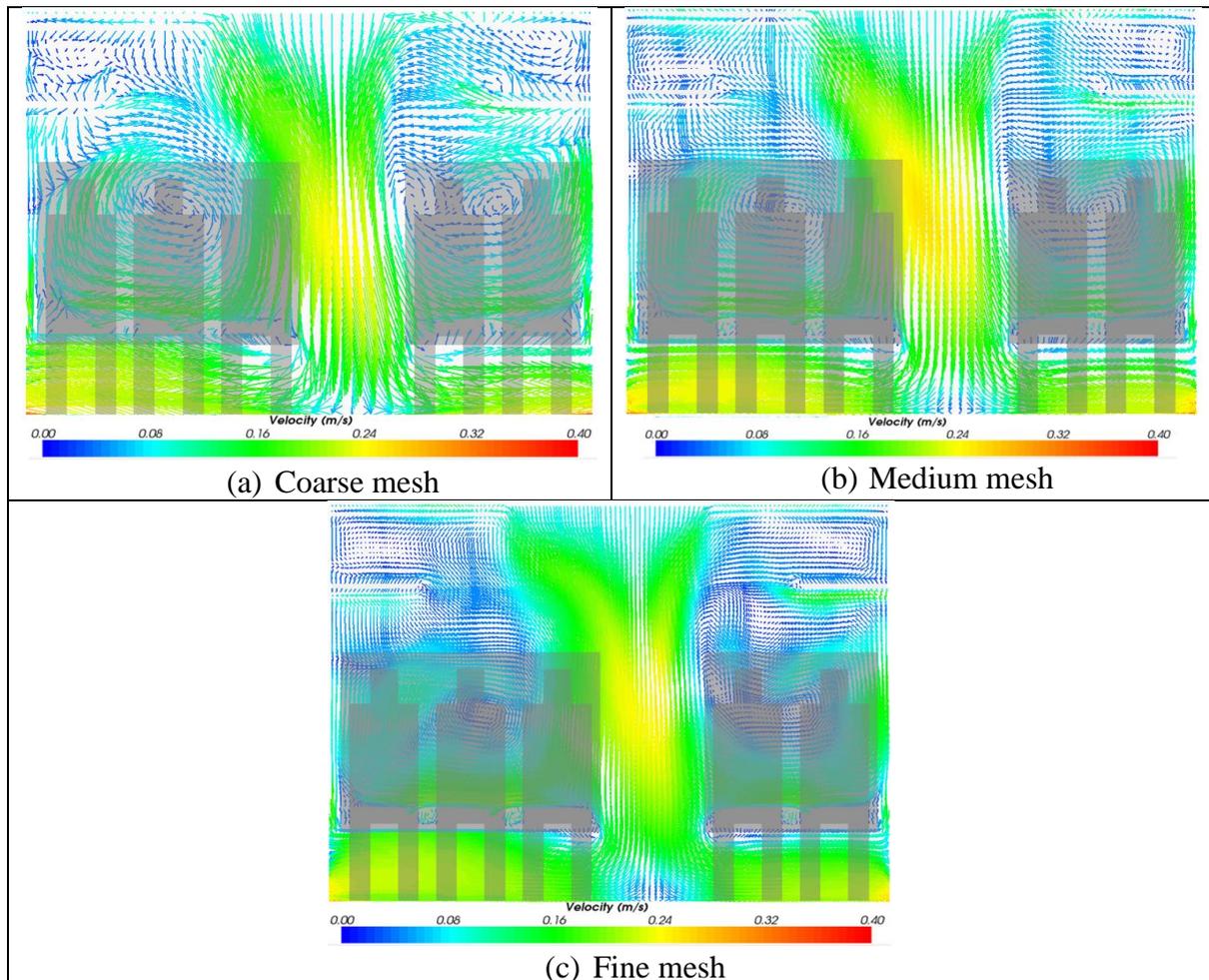


Fig. S3. Predicted velocity distribution in the vertical plane 0.1 m from the face of passengers in seat row 6 for the (a) coarse, (b) medium and (c) fine mesh cases.

The estimated IPs based on the CFD analysis for the A and C seat columns are presented in Fig. S6(a) and Fig. S6(b) respectively, while the estimated IP along seat row 5 and seat row 6 are presented in Fig. S6(c) and Fig. S6(d) respectively. In these cases, the index patient is located in seat 6C. As can be seen from these figures, the predicted IP curves follow the same trends and with similar values for the three mesh cases. It is noted that the coarse mesh produces a poorer representation of the predicted IP in Row 5 (see Fig. S6(c)), the row immediately behind the row containing the index patient.

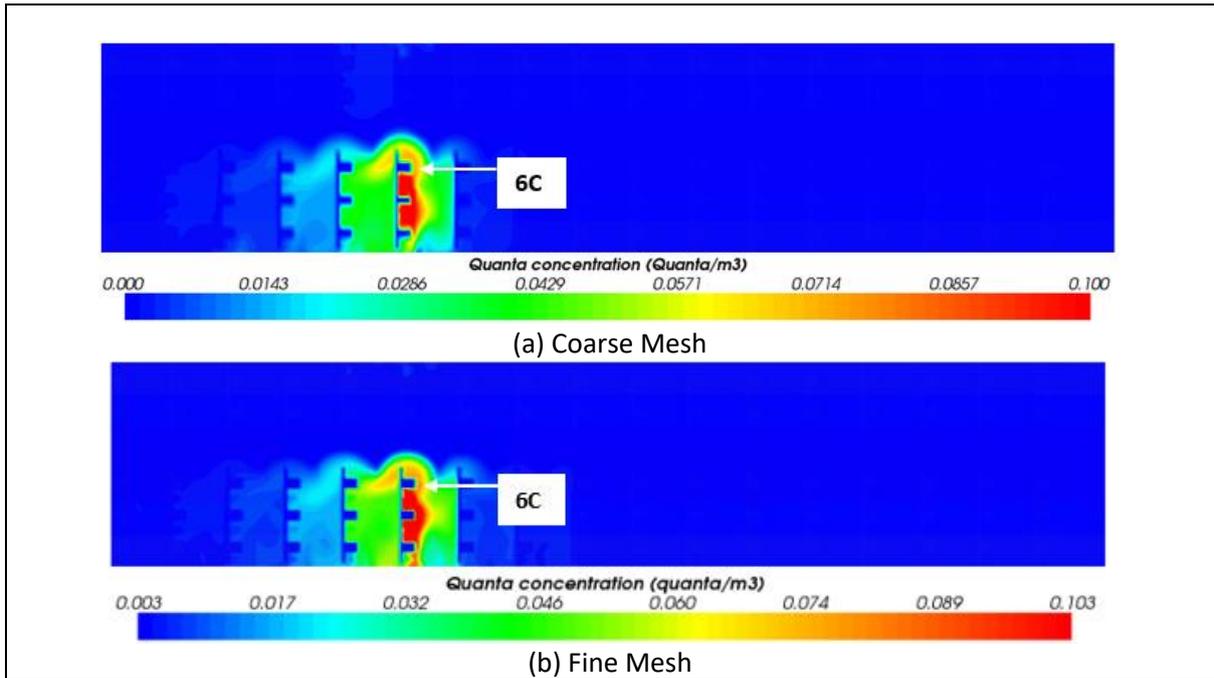
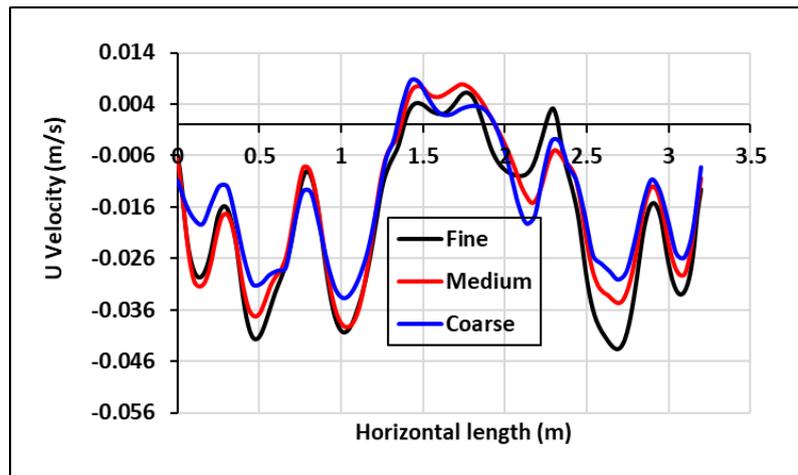
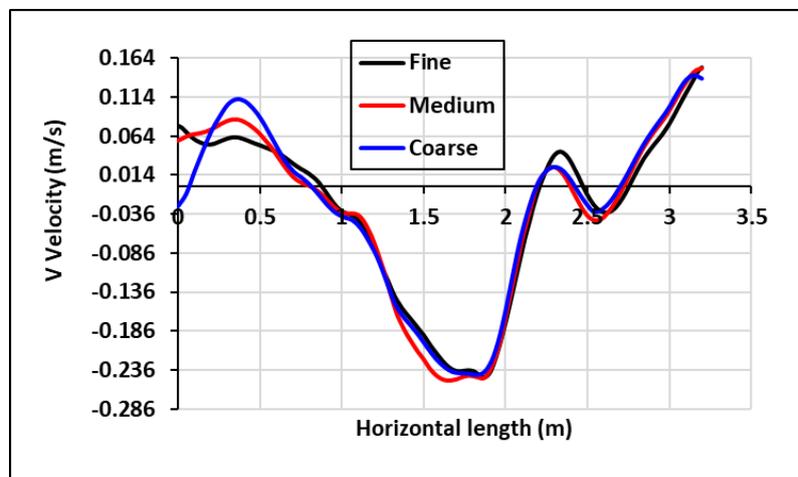


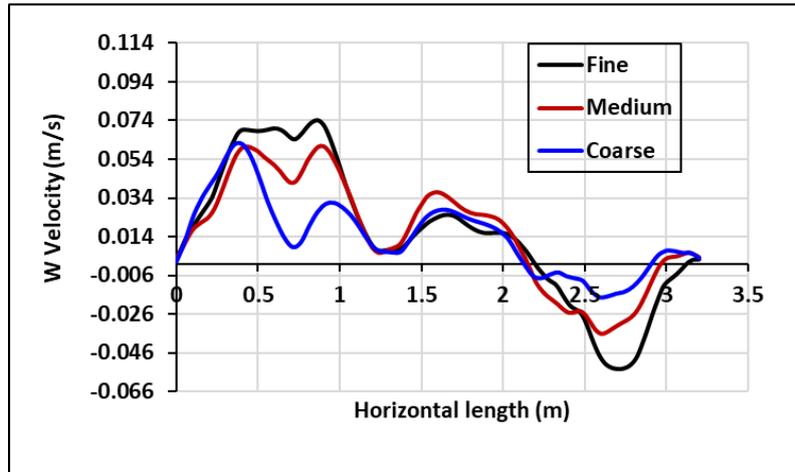
Fig. S4. Predicted quanta concentrations (quanta/m³) in a horizontal cutting plane at nose height for the (a) coarse and (b) fine mesh cases.



(a) Longitudinal velocity (U velocity)



(b) Vertical velocity (V velocity)



(c) Transverse velocity (W velocity)

Fig. S5. Velocities along a transverse line passing through the target volumes in Row 6 for the fine, medium and coarse mesh cases in the (a) longitudinal; (b) vertical and (c) transverse coordinate directions.

As in the MP Section 4, the IP for an 8-hour exposure assuming that the index patient is located in seat 6C for the following five spatial separations from the index patient: i.e. adjacent seat; same seat row excluding the adjacent seat; one-; two-; and three- seat rows away were determined using each of the three mesh refinements (see Table S4). As can be seen, all three mesh refinements produce similar values, with the medium mesh generally producing values closer to those produced by the fine mesh.

Table S4: IPs during 8-hour exposure determined for the three mesh refinements (Index 6C).

IP (%)	Mesh		
	Coarse	Medium	Fine
Adjacent	24.9	25.3	28.4
Same row	4.67	4.47	4.34
1-row	3.00	3.22	3.42
2-row	1.40	1.29	1.44
3-row	0.79	0.80	0.81

While all three computational meshes produced results with similar flow patterns, quanta concentration distributions and calculated IPs, the absolute values for the predicted IP produced by the medium mesh were more consistent with that produced by the finer mesh. Thus, while all three meshes would lead to the same broad conclusions, particularly in terms of relative effectiveness of the mitigation strategies explored, the medium mesh was considered appropriate for the study presented in the MP.

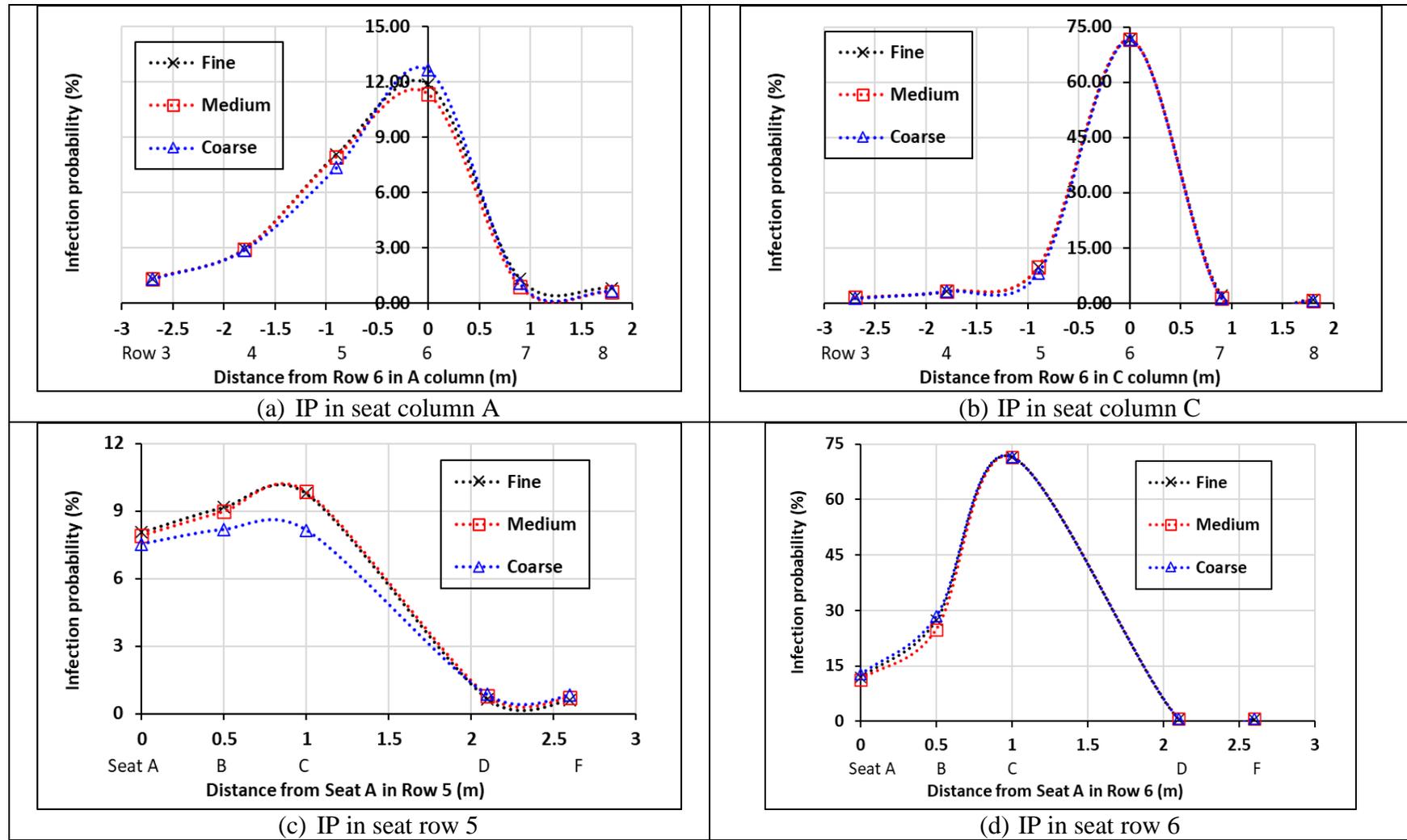


Fig. S6. Estimated IPs measured from Row 6 for, (a) locations along seat column A, and (b) locations along seat column C; and at locations measured from the A seat in (c) Row 5 and (d) Row 6. Note, the index patient is located in seat 6C.

S5 Temperature and air flow

The nature of the predicted temperature and flow field are briefly discussed in this section. As noted in MP-Section 6, the precise nature of the G-train ventilation ducts are unknown and the representation of the inlet/outlet ducts in this analysis is an extreme simplification of the actual situation on G-trains. Thus, the results presented in the paper should not be taken as an accurate representation of the situation on the actual trains. At best, the results are indicative of the differences that may occur for the different types of ventilation configuration. However, it is important to consider the nature of the flow dynamics that result from the different ventilation conditions examined as these can explain the broad differences in quanta dispersion observed in the various scenarios.

The body heat released from each of the simulated passengers generates a buoyancy driven thermal plume air flow around each passenger (see Fig. S7). However, the air injected from the side wall inlets, in ventilation Scenario 2, prevents the relatively warm rising air to spread past the luggage racks to the ceiling. However, as seen in Scenario 1, with ventilation inlets located in the ceiling above the aisle, the warm air plumes – rising from the passengers – can spread past the luggage racks to the ceiling.

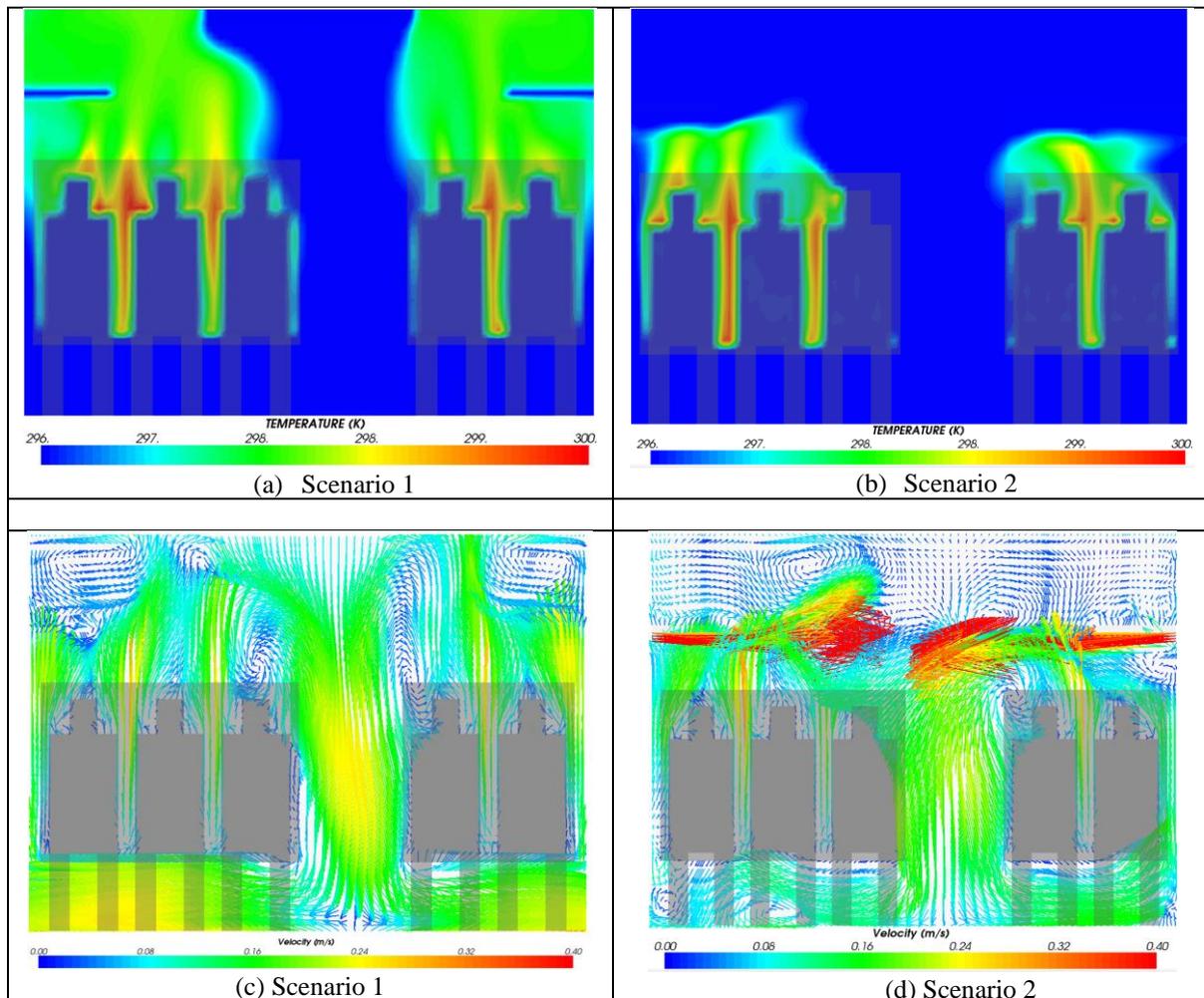
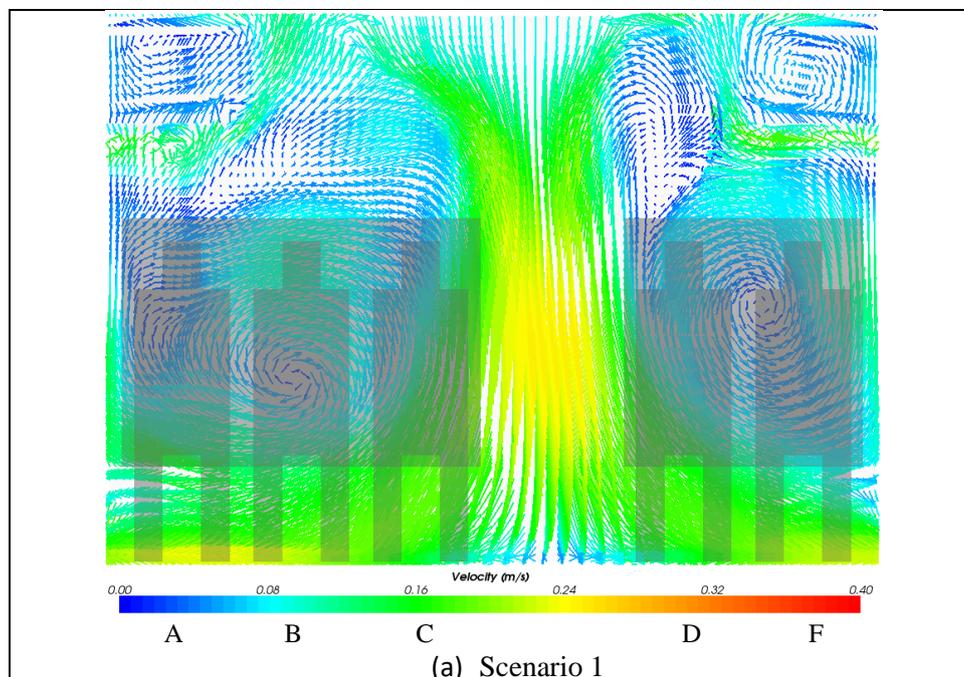


Fig. S7. Steady state temperature ((a) and (b)) and velocity distributions ((c) and (d)) in the vertical plane passing through the passengers at seat row 6 for ventilation Scenario 1 ((a) and (c)) and Scenario 2 ((b) and (d)) with Index patient located at 6C.

The flow fields for both scenarios are asymmetrical, primarily driven by the asymmetries in the saloon geometry (see Fig. S8, Fig. S9 and Fig. S3(b)). In Scenario 1, with a strong ceiling air injection vent, the downward flow in the aisle essentially splits the saloon into two sections – behaving like an air curtain along the saloon aisle (see Fig. S3(b) and Fig. S8(a)). On the right side (i.e. seats D and F), when looking down the saloon from front to back, there is an anticlockwise recirculation region over the passenger seated in seat D and seat F (see Fig. S3(b)).

If an index patient is seated in window seat F, the released quanta will be carried towards the face of the passenger in seat D. However, if an index patient is seated in aisle seat D, the released quanta follow two predominant flow paths, with part of the flow following a slanting upwards jet leaning towards the aisle, which then turns down towards the floor of the aisle. The second flow path follows an anticlockwise recirculation carrying a proportion of the emitted quanta towards the face of the passenger in window seat F. This explains why the IP for the susceptible in seat 6D is greater than that for the susceptible located in seat 6F (See MP-Section 5.2 Table 2d and Table 2e).

On the left side (seats A, B and C), there is a clockwise recirculation region over the three passengers approximately centred over seat B (see Fig. S3(b)). If the index patient is seated in seat C, the released quanta will flow towards the face of the passenger in seat B and then on towards the passenger in seat A due to the clockwise circulation flow. However, the passenger located in seat B experiences a stronger flow (centre of recirculation region offset towards seat C) and so has a higher IP (25.3%) than the passenger located in seat A (12.1%) (See MP-Section 5.2 Table 2c). If the index patient is in seat A, the released quanta follows a slanting upwards jet over the head of the passenger in seat B and onto the passenger in seat C.



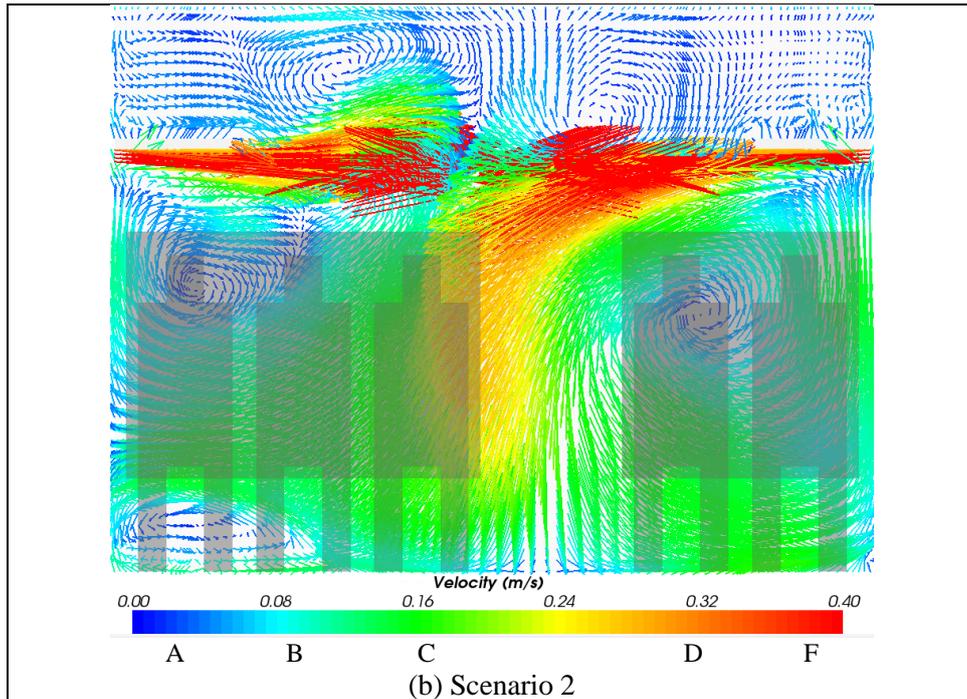


Fig. S8. Velocity distribution in the vertical plane 0.1 m behind the seat back of row 6 in (a) ventilation Scenario 1 and (b) Scenario 2 with Index patient located at 6C.

This results in the passenger in seat C having a slightly higher IP (14.7%) than the passenger in seat B (11.0%) (See MP-Section 5.2 Table 2a). Based on the recirculation flow in Fig. S3(b), if the index patient is located in seat B, then the passenger located in seat C will have a higher IP (15.8%) than the passenger in seat A (6.7%), due to the nature of the clockwise recirculation flow (See MP-Section 5.2 Table 2b). Thus, the pattern of air flows established in the saloon, due to the ceiling inlet, explains the nature of the IP distribution and, in particular, the location of the seats with maximum IP and why index patients located in seat C produce the highest maximum IP.

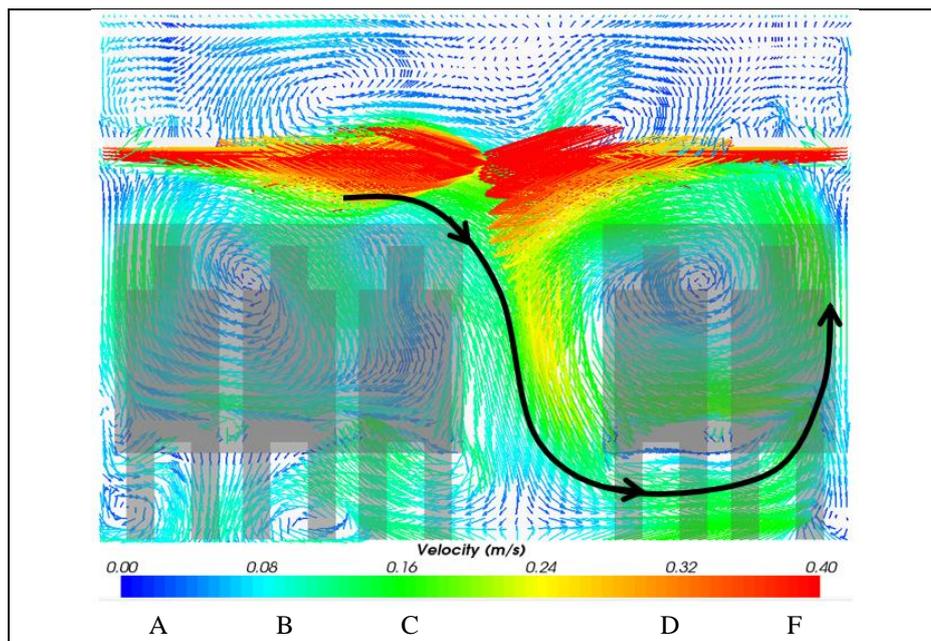


Fig. S9. Velocity distribution in the vertical plane 0.1 m from the face of passengers in seat row 6 in Scenario 2 with Index patients located at 6C.

In contrast to Scenario 1, the side wall inlets in Scenario 2 cause the flow to cross the aisle, causing released quanta to be distributed to both sides of the saloon. As seen in Fig. S8(b), 0.1 m behind the seat back in Row 6, the jet flow from the side wall on the two-seat block side crosses the aisle and is directed down towards the seat base of the three-seat block. In contrast, Fig. S9 depicts the flow in the vertical plane at 0.1 m from the passenger face in seat Row 6 (0.3 m ahead of the seat back in Row 6). In this plane, the flow crosses the aisle from the three-seat block side to the two-seat block following the approximate path indicated by the black arrows. Scenario 2 produces a clockwise recirculation region just over seat A and seat B. If an index patient is seated at seat A, the small clockwise recirculation flow directs the released quanta towards the face of the passenger in seat B. However, if an index patient is seated in seat C, the released quanta will be initially directed towards the lap of the passenger in seat C and then across to seat A and upwards. This results in the passenger in seat A having a slightly higher IP (13.2%) than the passenger in seat B (12.1%) (See MP-Section 5.1 Table 2h).

As in Scenario 1, on the two-seat block side, there is an anticlockwise recirculation region over the passenger seated in seat D and seat F (see Fig. S9) with its centre slightly off-set towards seat D. If an index patient is seated in window seat F, the released quanta will initially be carried upwards and then down towards the face of the passenger at seat D. However, if an index patient is seated in aisle seat D, the released quanta are initially directed downwards and then diverted upwards due to the recirculation. As the centre of recirculation region is offset towards seat D, only part of the emitted quanta are carried towards the face of the passenger in window seat F. This explains why the IP for the susceptible in seat 6D (19.0%) is greater than that for the susceptible located in seat 6F (12.9%) (See MP-Section 5.1, Table 2i and 2j). Therefore, for ventilation Scenario 2, the worst index locations are window seats A and F (See MP-Section 5.3).

Presented in Fig. S10 are the longitudinal flow fields for a vertical plane passing through the centre of passengers located in the F seats and streamlines originating from the aerosol source of the index patient seated at 6F in ventilation Scenario 1 and Scenario 2. Clearly, the longitudinal flow speed between the top of the passenger heads and the saloon luggage racks in Scenario 1 (see Fig. S10(a)), is significantly larger than that in Scenario 2 (see Fig. S10(b)). It should be noted that the high flow speed under the luggage rack (red arrows) in Scenario 2 (see Fig. S10(b)) are in the transverse direction. This results in a greater longitudinal dispersion of quanta in Scenario 1, compared to Scenario 2, along the two-seat block as seen in MP-Fig. 8 and the streamlines (see Fig. S10(c) and Fig. S10(d)). The relatively smaller longitudinal dispersion of quanta in Scenario 2 is the result of the high air flow injected from the side walls. The different longitudinal dispersions between the two-seat block and three-seat block in Scenario 1 are due to the different volumes of the two spaces generated by the air curtain along the aisle formed by the top injected air. The influence of the air curtain causes the flow in the smaller cross section space (the two-seat block section) to move faster longitudinally than in the larger space that includes the three-seat block.

It is noted that the nature of the flows, in both Scenario 1 and 2, are expected to be strongly dependent on the precise design of the ventilation outlets – which often have vanes, grilles or decorative covers and to a lesser extent, the mesh resolution (see SM Section S4).

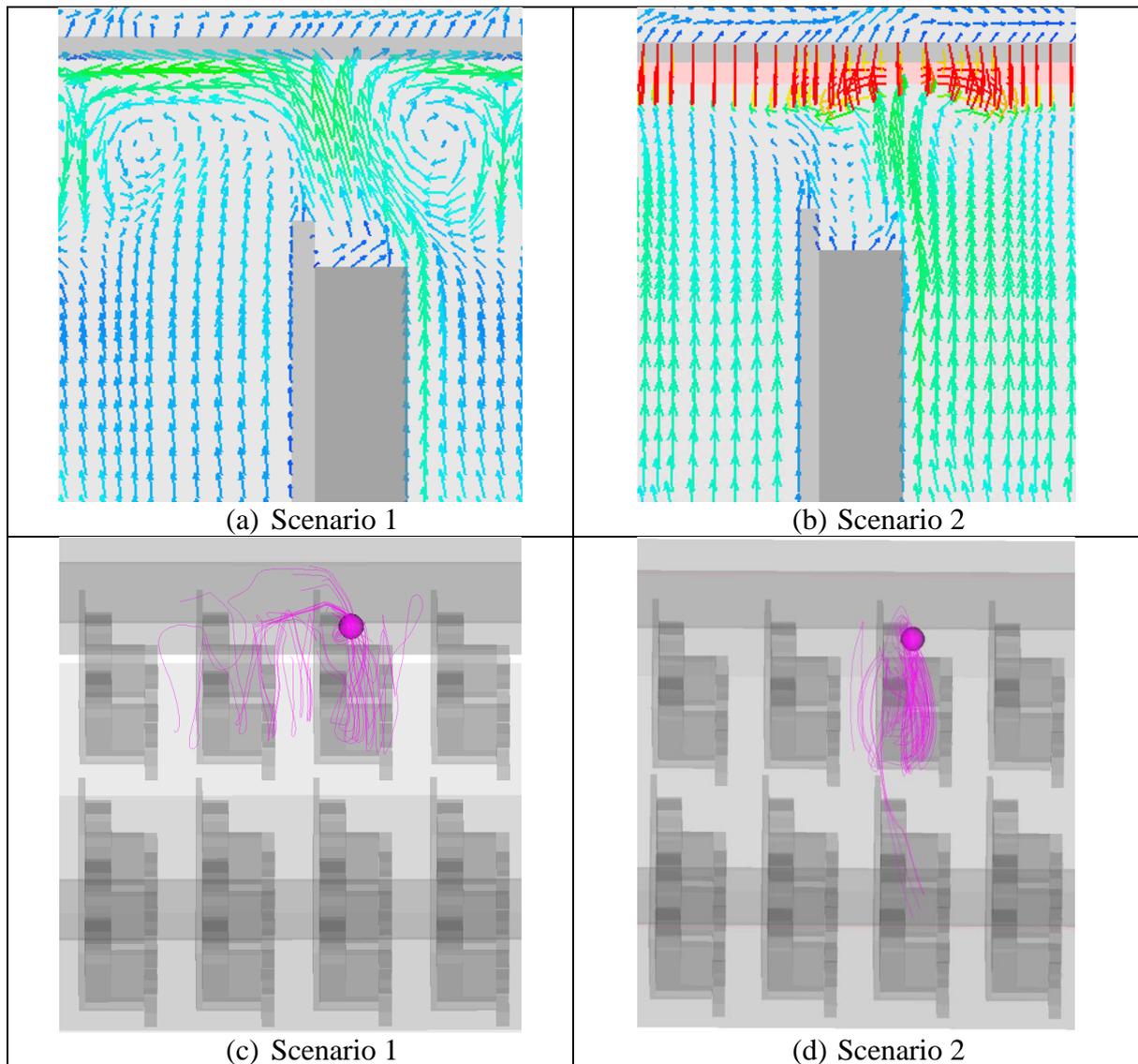


Fig. S10. Velocity vectors on a vertical plane passing through index patient in seat 6F for (a) ventilation Scenario 1 and (b) Scenario 2 and streamlines from the source of index patient for (c) Scenario 1 and (d) Scenario 2.

S6 Non-uniform quanta distribution

The quanta distribution throughout the saloon resulting from the recirculation produced by the ventilation system is non-uniform. While this is dependent on the seating location of the index source and the nature of the ventilation system, there appear to be general trends in the quanta distribution. As seen in MP-Fig. 8, very high quanta concentrations occur up to two seat rows behind the index patient, with lower but still high concentrations up to one seat row ahead of the index patient. The quanta distribution drops off significantly outside these regions.

The core assumption in the application of the Wells-Riley Model (WRM), when applied to the entire saloon, is that the volume of space is well-mixed resulting in a uniform quanta distribution. The CFD analysis indicates that this is an invalid assumption for train saloons and as a result, hot-spots of IP will occur. The impact of face coverings was introduced into the

coupled Wells Riley CFD (WR-CFD) model through MP-equation (8), which predicts IP based on local quanta concentration. Similarly, the impact of face coverings can be incorporated within the general WRM (MP-equation (1)) as follows:

$$p = 1 - (1 - x)^2 e^{-Iqvt/Q} - x(1 - x)(e^{-(1-a)Iqvt/Q} + e^{-Iq(1-b)vt/Q}) - x^2 e^{-(1-a)Iq(1-b)vt/Q} \quad (S5)$$

where x is the proportion of the population who are likely to wear face coverings (i.e. the probability of a passenger wearing face covering is x). If the usual “well mixed” WRM (Equation (S5)) is applied to the passenger area of the saloon, for Scenario 1, the ventilation rate for the passenger area is 5500 m³/h including fresh air rate of 1881 m³/h and 3619 m³/h recycled air (corresponding to the ventilation rate of 6200 m³/h including a fresh air rate of 2120 m³/h for the whole carriage). As the filtration efficiency is assumed to be 20%, the equivalent fresh air rate is approximately 2605 m³/h. Similarly, the equivalent fresh air rate for Scenario 2 is 2109 m³/h. Using Equation (S5), the predicted IPs as a function of time, for the seating area of the saloon, are presented in Fig. S11. It is thus not surprising that the predicted IPs for an 8-hour exposure are only 0.9% for ventilation Scenario 1 and 1.1% for ventilation Scenario 2 using the WRM, which are much lower than the reported data at locations within three seat rows from the index patient.

The failure of the WRM to accurately represent IP is due to the poor validity of the uniform quanta distribution assumption. Indeed, as depicted by MP-Fig. 8, the quanta distribution in the steady state is extremely complex and far from uniform. As a result, simply applying the WRM to the entire saloon fails to assess significant localised IP. For complex environments, such as a train saloon with asymmetries in configuration and complex ventilation systems, a well-mixed uniform state is an invalid assumption. Wells-Riley applications that predict low IPs for complex spaces such as classrooms, aircraft cabins, etc., as in [S4], need to be carefully reconsidered as they are likely to fail to identify high IPs in the vicinity of the index patients. Thus, use of the WRM for these types of environments, in particular to assess IP on a seat by seat basis, and thereby identify potential mitigation strategies, is inappropriate.

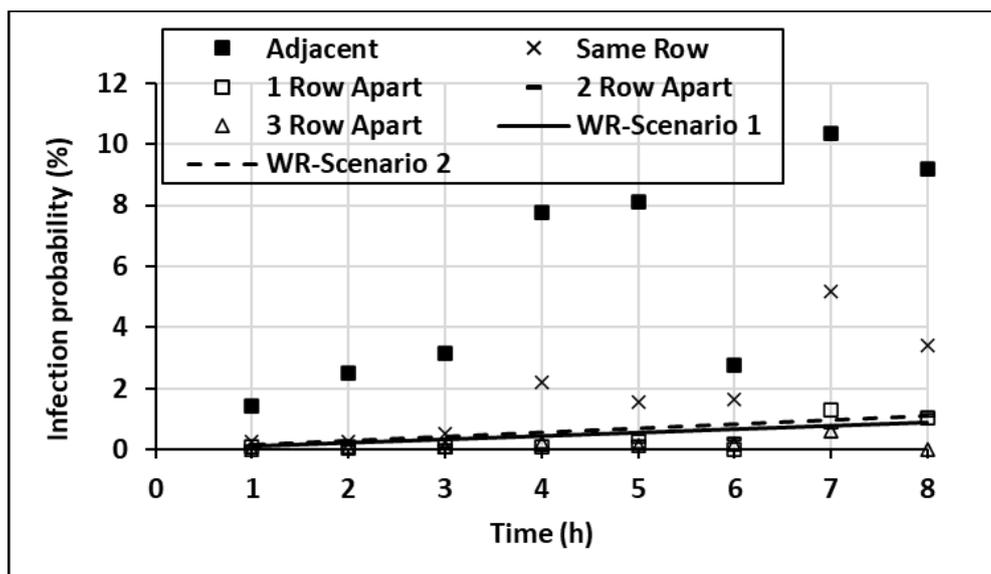


Fig. S11. Reported (symbols) [S3] and WRM predicted IPs (solid and dashed lines) for the seating area of the saloon.

S7 Impact of face coverings

S7.1 Usage of face coverings during the study period

While no mention of the use of face coverings was made in Hu et al. [S3], it is common practice in China to wear face coverings when ill with a respiratory ailment. It is thus likely that some people would have been wearing face coverings during the data collection period. According to the database [S23], there were 396 confirmed COVID-19 infections on G-trains before the 25th January 2020. Most of these passengers travelled on G-trains originating from Wuhan, China. Among the 396 cases, approximate 58% occurred from the 20th to the 23rd January 2020, during which it was common for people to wear face coverings at Wuhan Rail station [S24].

Video footage of people queuing to board trains in Wuhan Rail station on the 21st January 2020 show many (if not most) wearing face coverings [S24]. Thus, in this study it is assumed that 40% of the population on the G-train were wearing face coverings. Given the variability in types of masks worn and also on how well they are worn, in this study it is assumed that the face coverings have filtration efficiencies of 50% and 30% for index and susceptible passengers, respectively [S25].

S7.2 Infection probability with deterministic face covering state

The estimated average IP assumes that 40% of the passengers wear face coverings. For example, the average IP for an 8-hour exposure, for a susceptible in seat 6B in Scenario 1 with the index patient seated in 6C, is 25.3% (See MP-Table 2c).

However, for a given passenger, their face covering state is deterministic. There are four different deterministic face covering possibilities when estimating the IP of the passenger in seat 6B:

- Both index patient 6C and susceptible 6B do not wear face coverings;
- Index patient 6C wears a face covering but susceptible 6B does not;
- Index patient 6C does not wear a face covering but susceptible 6B does;
- Both index patient 6C and susceptible 6B wear face coverings.

The IPs for the above face covering combinations can be calculated using MP-equations (4)-(7), given the value of c determined from the CFD simulation for seat 6B. For the four different face covering states, the IPs of the susceptible, located in seat 6B, are 34.7%, 19.0%, 25.6% and 13.7%, respectively. Therefore, the IP for the susceptible seated in seat 6B ranges from 13.7% to 34.5% and is dependent on the face covering state (of the susceptible in seat 6B and the index patient in seat 6C), while the average probability calculated using MP-equation (8) is 25.3%, assuming that 40% of the passengers wear face coverings.

If either the index patient seated in 6C or the susceptible seated in 6B does not wear a face covering, the IP of the susceptible seated in 6B increases by a relative amount of at least 39% compared with the IP when both wear face coverings. If both do not wear face coverings, the IP increases by a relative amount of 152%. This analysis demonstrates that wearing a face covering has a significant impact on reducing infection transmission of COVID-19 on long train journeys.

S7.3 Impact of higher efficiency face coverings

As highlighted in the previous discussion, the IP of the susceptible located in seat 6B in Scenario 1 is 13.7%, when both susceptible in seat 6B and the index patient in seat 6C wear a face covering. This is based on mask filtration efficiencies of 0.5 and 0.3 for index and

susceptible respectively. These mask filtration efficiencies are appropriate for surgical masks. If a higher efficiency mask, such as an N95 mask worn correctly, filtration effectiveness for both index and susceptible could approach 0.9. The impact that this would have on IPs can be estimated by taking $a=b=0.9$ in MP-equation (7) and given the value of c determined from the CFD simulation for seat location 6B. Using these values, the IP becomes 0.4% during an 8-hour exposure. Compared with the IP of 13.7%, the higher quality face mask reduces the IP by a relative improvement of 97%. Furthermore, if 90% of passengers wear face masks with an efficiency of 90%, the number of expected secondary infections is reduced by at least 95% in Scenarios 1 and 2 (See MP-Table 4).

Thus, a significant reduction (95%) in IP, for all passengers, can be achieved if 90% of passengers correctly wear high efficiency face coverings for long distance travel on trains.

References

- [S1] Wang, Z., Galea E.R., Grandison A, Ewer J, Jia, F., A coupled Computational Fluid Dynamics and Wells-Riley model to predict COVID-19 infection probability for passengers on long-distance trains, accepted for publication by Safety Science, 1st November 2021.
- [S2] Hu M, Lin H, Wang J, et al., The risk of COVID-19 transmission in train passengers: an epidemiological and modelling study, *Clinical Infectious Diseases*, Vol 72, No. 4, 2021, pp 604-610. <https://doi.org/10.1093/cid/ciaa1057> (Supplementary Material).
- [S3] Hu M, Lin H, Wang J, et al., The risk of COVID-19 transmission in train passengers: an epidemiological and modelling study, *Clinical Infectious Diseases*, Vol 72, No. 4, 2021, pp 604-610. <https://doi.org/10.1093/cid/ciaa1057>
- [S4] Dai H, Zhao B, Association of the infection probability of COVID-19 with ventilation rates in confined spaces, *Build Simul*, Vol 13, 2020, pp 1321–1327. doi:10.1007/s12273-020-0703-5.
- [S5] Buonanno G, Morawska L, Stabile L, Quantitative assessment of the risk of airborne transmission of SARS-CoV-2 infection: Prospective and retrospective applications *Environment International*, Environment International, Vol. 145, 2020.
- [S6] Buonanno G, Stabile L, Morawska L, Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment, *Environment International*, Vol 141, 2020. <https://doi.org/10.1016/j.envint.2020.105794>
- [S7] Hota B, Stein B, Lin M. et al., Estimate of airborne transmission of SARS-CoV-2 using real time tracking of health care workers, *medRxiv*, 2020. <https://doi.org/10.1101/2020.07.15.20154567>.
- [S8] Marcus LJ, et al., Assessment of Risks of SARS-CoV-2 Transmission During Air Travel and Non-Pharmaceutical Interventions to Reduce Risk, Phase One Report: Gate-to-Gate Travel Onboard Aircraft. 2020, Faculty and Scientists at the Harvard T.H. Chan School of Public Health. National Preparedness Leadership Initiative, Cambridge, MA 02238, USA. <https://cdn1.sph.harvard.edu/wp-content/uploads/sites/2443/2020/10/HSPH-APHI-Phase-One-Report.pdf>.

- [S9] Miller SL, Nazaroff WW, Jimenez JL, et al., Transmission of SARS-CoV-2 by inhalation of respiratory, *Indoor Air*, Vol 31, No. 2, 2020, pp 314-323. <https://doi.org/10.1111/ina.12751>.
- [S10] Hu M, Wang JF, Lin H et al. Transmission risk of SARS-CoV-2 on airplanes and high-speed trains, *medRxiv 2020.12.21.20248383*, doi.org/10.1101/2020.12.21.20248383.
- [S11] Wang Z, Galea ER, Grandison A, et al., Inflight transmission of COVID-19 based on experimental aerosol dispersion data, *Journal of Travel Medicine*, Vol 28, No. 4, 2021. <https://doi.org/10.1093/jtm/taab023>
- [S12] CDC, Science Brief: SARS-CoV-2 and Surface (Fomite) Transmission for Indoor Community Environments. <https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/surface-transmission.html>, 5 April 2021. Accessed on 8th June 2021.
- [S13] CDC, Scientific Brief: SARS-CoV-2 Transmission, <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html> , 7 May 2021. Accessed on 8 June 2021.
- [S14] Meyerowitz EA, Richterman A, Gandhi RT, et al., Transmission of SARS-CoV-2: a review of viral, host, and environmental factors, *Annals of internal medicine*, 2020. <https://doi.org/10.7326/M20-5008>
- [S15] Kampf G, Brüggemann Y, Kaba H, et al., Potential sources, modes of transmission and effectiveness of prevention measures against SARS-CoV-2, *Journal of Hospital Infection*, Vol 106, No. 4, 2020, pp 687-697. <https://doi.org/10.1016/j.jhin.2020.09.022>
- [S16] Volz E, Mishra S, Chand M, et al., Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data, medRxiv preprint doi: <https://doi.org/10.1101/2020.12.30.20249034>; posted January 4, 2021.
- [S17] Galloway SE, Paul P, MacCannell DR, et al., Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. *MMWR Morb Mortal Wkly Rep.* ePub: 15 January 2021. DOI: <http://dx.doi.org/10.15585/mmwr.mm7003e2external> icon.
- [S18] Grinch (global report investigating novel coronavirus haplotypes) https://cov-lineages.org/global_report_B.1.1.7.html, Accessed 17 January 2021.
- [S19] SARS-CoV-2 variants of concern and variants under investigation in England Technical briefing 14 3 June 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/991343/Variants_of_Concern_VOC_Technical_Briefing_14.pdf
- [S20] Dagpunar J, Interim estimates of increased transmissibility, growth rate, and reproduction number of the Covid-19 B.1.617.2 variant of concern in the United Kingdom, medRxiv preprint, 2021 doi: <https://doi.org/10.1101/2021.06.03.21258293>

- [S21] Allen H, Vusirikala A, Flannagan J, Increased household transmission of COVID-19 cases associated with SARS-CoV-2 Variant of Concern B.1.617.2: a national case control study.
<https://khub.net/documents/135939561/405676950/Increased+Household+Transmission+of+COVID-19+Cases+--+national+case+study.pdf/7f7764fb-ecb0-da31-77b3-b1a8ef7be9aa>.
Accessed on 22nd, Jun 2021.
- [S22] Burridge HC, Fan S, Jones RL et al., Predictive and retrospective modelling of airborne infection risk using monitored carbon dioxide, 2021. eprint arXiv:2009.02999.
- [S23] <http://tj.bendibao.com/news/2020129/86594.shtm>. Accessed 6 September 2020.
- [S24] Youtube video clip, https://www.youtube.com/watch?v=625Yh_LO0YY. Accessed 31 12 2020.
- [S25] Davies A, Thompson K, Giri K, et al., Testing and efficacy of homemade masks: Would they protect in an influenza pandemic?, *Disaster Med Public Health Prep*, Vol. 7, No. 4, 2013, pp 413-418.