



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/241342/>

Version: Accepted Version

---

**Article:**

Marshall, M., Edwards, A.J., Pinnell, D. et al. (Accepted: 2026) The Wells–Riley model revisited II: parameter uncertainty and population heterogeneity. *Risk Analysis*. ISSN: 0272-4332 (In Press)

---

This is an author produced version of an article accepted for publication in *Risk Analysis*, made available via the University of Leeds Research Outputs Policy under the terms of the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

## ARTICLE TYPE

# The Wells–Riley model revisited II: parameter uncertainty and population heterogeneity

Marcus Marshall<sup>1</sup> | Alexander J. Edwards<sup>2</sup> | Dominique Pinnell<sup>1</sup> | Marco-Felipe King<sup>3</sup> | Catherine J. Noakes<sup>3</sup> | Nazif Elaldi<sup>4</sup> | Grant Lythe<sup>1</sup> | Martín López-García<sup>1</sup>

<sup>1</sup>School of Mathematics, University of Leeds, Leeds, UK

<sup>2</sup>School of Chemistry, University of Bristol, Bristol, UK

<sup>3</sup>School of Civil Engineering, University of Leeds, Leeds, UK

<sup>4</sup>Dept. of Infectious Diseases, Cumhuriyet University, Sivas, Turkey

**Correspondence**

Marcus Marshall,  
Email: ll17mjm@leeds.ac.uk  
Martín López-García,  
Email: m.lopezgarci@leeds.ac.uk

**Present address**

School of Mathematics, University of Leeds,  
LS2 9JT Leeds, UK

**Abstract**

In this work we revisit the Wells-Riley model, which has been widely used to estimate airborne infection risk in indoor settings. In particular, we consider a probabilistic (i.e., “stochastic”) framework of the Wells-Riley model which allows one to quantify infection risk in terms of the per-capita probability of infection for each susceptible individual, as well as the probability distribution of the number of infections (here referred to as “exposures”) during the indoor interaction. Directly extending the work by Edwards et al. (2024c), we consider here the situation where the main parameters in the Wells-Riley model (namely, the quanta generation rate  $q$ , the ventilation rate  $Q$ , the number of infectors  $I$  or the duration of the indoor interaction  $T$ ) may be random or uncertain. We show how, in this case, the per-capita infection risk  $P_{infection}$  becomes a random variable between 0 and 1, and compute its density function under some parametric assumptions. This allows for a comprehensive analytical quantification of uncertainty when dealing with heterogeneous populations, uncertain environmental conditions, or stochastic human behaviour. Our results reveal that infection risk can vary significantly depending on the distribution and variability of model parameters. In particular, using mean parameter values in the classical Wells-Riley model can lead to systematic inaccuracies: uncertainty in  $q$ ,  $T$  or  $b$  leads to infection risk overestimation, while environmental stochasticity (i.e. uncertainty in ventilation or removal rates) can lead to infection risk underestimation. We also investigate which parameter mainly drives the uncertainty in infection risk when model parameters are simultaneously random.

**KEYWORDS**

Airborne transmission, Wells-Riley model, Parameter uncertainty, Infection risk

## 1 | INTRODUCTION

Airborne transmission is a common mode of infection for various pathogens. It played a particularly significant role in the spread of COVID-19 during the pandemic that began in early 2020 (Karia et al. (2020)). This type of transmission occurs when a susceptible host breathes in small droplets or aerosols laden with an infectious pathogen, usually produced by an infector due to breathing, talking, coughing, sneezing or during aerosol-generating procedures. Unlike other routes of infection, such as fomite transmission which is driven by direct contact (e.g., touch) with a contaminated individual or surface, airborne transmission may infect individuals indirectly (e.g., shared air within indoor settings).

Long-range airborne transmission can be mitigated using ventilation strategies (Qian and Zheng (2018)) and through the use of personal protective equipment (PPE) such as masks (Wang et al. (2021)). During the early stages of the COVID-19 pandemic, a limited availability of medical masks resulted in illness amongst the frontline healthcare workforce (The Lancet (2020)). This led to staff shortages and put a strain on healthcare services worldwide (Xu et al. (2020)). It is therefore important to understand and assess the risk of airborne transmission particularly in healthcare settings. On the other hand, it is also important to analyse transmission dynamics in non-healthcare settings, especially in the case of large scale or super-spreading events, which can play a significant role in accelerating population level transmission during an epidemic. A paradigmatic example is the Skagit-County Choir rehearsal outbreak in Washington, USA during 2020, where an index case with COVID-19 was thought to have infected 53 of the 61 attendees, resulting in two

deaths (Hamner (2020)). The high number of exposures was likely linked to poor ventilation (Miller et al. (2021)) and the fact that individuals were singing, which is known to increase the volume of exhaled aerosols (Schijven et al. (2021)). This highlights the importance of understanding environmental and behavioral factors affecting airborne transmission risk.

To systematically evaluate infection risk, the Quantitative Microbial Risk Assessment (QMRA) framework can be used. We refer the reader to Haas et al. (2014) for a comprehensive and foundational overview of this methodology. QMRA aims to identify risk factors, and to link pathogen exposure to the probability of infection through the use of mathematical modeling. The use of mathematical models helps to identify the main factors contributing to infection risk, which may inform infection control strategies. Different models can be developed for various routes of transmission or scenarios. For example, QMRA models have been proposed for waterborne exposure (Haas et al. (1993), Cummins et al. (2010), World Health Organization (2016), Doménech et al. (2022), Paraskevopoulos et al. (2024)), fomite transmission (Nicas and Jones (2009), Sze-To et al. (2014), King et al. (2022), Abney et al. (2024), Bate et al. (2024)) and airborne transmission (Chen et al. (2006), Denpetkul et al. (2022), Miller et al. (2022), Bate et al. (2024)).

A particular QMRA framework to quantify airborne transmission risk is the Wells-Riley model. The Wells-Riley model uses the concept of a *quantum* of infectious pathogen which was defined by Wells (1955) as the amount of infectious pathogen required to infect a susceptible host upon inhalation. The model was formalised in its classical form by Riley et al. (1978), where it was applied to a case study involving a Measles outbreak within a Florida elementary school. The classical Wells-Riley model assumes well-mixed air, a steady-state concentration of pathogen in the air, and a precise knowledge of individual (breathing rate, infectiousness), behavioural (duration of the indoor interaction) and environmental (ventilation) conditions, in terms of specific values of model parameters. Extensions of the Wells-Riley model have been proposed to address some of its main limitations, such as spatial models linking Computational Fluid Dynamics approaches to the Wells-Riley model (Qian et al. (2009), Gupta et al. (2012)), adaptations to consider a transient concentration of pathogen in the air over time (Gammaitoni and Nucci (1997)), or an extended Wells-Riley model to leverage CO<sub>2</sub> measurements (Rudnick and Milton (2003)). In many real scenarios, it is challenging to determine precise values for the parameters of the Wells-Riley model, due to biological heterogeneity, stochastic human behaviour, fluctuating environmental conditions, or due to insufficient/incomplete observations. For example, emission rates or viral loads of individuals within a population may vary due to stage of infection, age, respiratory activity, and other behavioural or physiological factors (Buonanno et al. 2020b, Jones et al. 2021b, Euser et al. 2021, Jones et al. 2024, Aganovic et al. 2024, Sender et al. 2022, Schijven et al. 2021, Mikszewski et al. 2022); removal rates may vary substantially with changing weather, building operation, or room-level factors (Nazaroff 2021, Edwards et al. 2024a, Jones et al. 2021a, Henriques et al. 2025, Jones et al.

2025); the duration of indoor interactions is often uncertain or only partially observable (King et al. 2021, Edwards et al. 2024c). Furthermore, behavioural patterns such as mask wearing, movement within indoor environments, and variable contact durations introduce further heterogeneity (King et al. 2021, Huang et al. 2022). At the group level, this uncertainty can significantly affect infection risk estimates, producing highly skewed distributions of secondary infections even under identical environmental conditions (Iddon et al. 2022). They are often difficult to capture using fixed, deterministic parameter values.

Recent work has taken substantial steps to incorporate such uncertainty into airborne infection risk modelling. Monte Carlo and numerical approaches have been used to propagate variability in emission rates, breathing rates, removal processes, prevalence, and occupancy (Bate et al. 2024, Henriques et al. 2025, Iddon et al. 2022, Jones et al. 2021a 2025, Edwards et al. 2026). Other approaches have examined the role of uncertainty in viral load and emission using quanta-independent optimisation methods (Aganovic et al. 2024). These frameworks have informed building-design and mitigation guidance, including ASHRAE Standard 241 (Jones et al. 2025), by explicitly embedding parametric uncertainty into their risk-assessment methodologies. Together, these studies highlight that uncertainty is not an auxiliary component but a central feature of airborne transmission modelling, and that robust infection risk assessment requires methods capable of representing full distributions rather than only mean values.

The present study proposes a stochastic, analytical framework to incorporate uncertainty into the Wells-Riley model, extending work previously developed by Nicas (1996) and more recently by Edwards et al. (2024c). Nicas (1996) incorporated uncertainty into dose-response models by analysing Beta- and Gamma-distributed doses, while Edwards et al. (2024c) considered exponentially or Erlang-distributed quanta emission rates or exposure durations when revisiting the Wells-Riley model. Here, we extend this line of work by deriving analytic expressions for the *full* probability distribution of the per-capita infection risk when one or two model parameters are uncertain, rather than focusing solely on average risk or relying on Monte Carlo simulation. By treating key parameters as random variables, the per-capita infection risk itself becomes a random variable in  $(0, 1)$  with a well-defined density. Specifically, we consider Gamma-distributed parameters, which provide the flexibility to model unimodal non-negative distributions while retaining analytical tractability, and derive closed-form density functions as well as mean infection risks and exposure distributions. Our results extend those of Nicas (1996) and Edwards et al. (2024c) by analysing infection risk under broader classes of random parameters, considering pairs of parameters simultaneously, and incorporating random ventilation (a scenario not addressed in Edwards et al. (2024c)). This framework enables an analytically comprehensive characterisation of how uncertainty in emission rates, exposure times, and environmental conditions shapes the distribution of infection risk across different scenarios.

The rest of this paper is organised as follows: in Section 2, we introduce the classical Wells-Riley model and present its stochastic formulation, while incorporating uncertainty into model parameters (both in individual parameters and in pairs); the aim is to compute the density function of the per-capita infection risk, as well as the probability of observing exactly  $n$  exposures. In Section 3, we illustrate our analytical results by means of two case studies to investigate the impact of parameter uncertainty on infection risk. Section 4 contains a final discussion of our approach and findings.

## 2 | METHODOLOGY

The Wells-Riley model (Wells 1955, Riley et al. 1978) has been widely used to estimate airborne transmission risk during indoor interactions. In particular, for an indoor interaction between  $S$  susceptible and  $I$  infectious individuals during  $[0, T]$ , the Wells-Riley model estimates the *per-capita* infection risk (i.e., the probability of infection for each susceptible individual) as

$$P_{infection} = 1 - e^{-\frac{Ibq}{Q}T}, \quad (1)$$

where  $b$  is the pulmonary (breathing) rate [ $m^3 \cdot h^{-1}$ ],  $q$  is the quanta emission rate [ $quanta \cdot h^{-1}$ ] and  $Q$  is the ventilation rate [ $m^3 \cdot h^{-1}$ ]. In this equation, the ventilation rate  $Q$  can be replaced by a more general loss parameter which may account for pathogen decay or deposition; see Edwards et al. (2024c), Miller et al. (2021). It has recently been shown by Edwards et al. (2024c) that, since infection events occur independently of each other, the number of infections (which we will synonymously refer to as “exposures” since it is typically the case that these individuals become infected but not infectious during  $[0, T]$ , since  $T$  represents the duration of the indoor interaction which is typically in the order of minutes or hours rather than days) follows a binomial distribution

$$E \sim \text{Binomial}(S, P_{infection})$$

so that the probability of observing exactly  $n$  exposures during  $[0, T]$  is given by

$$\mathbb{P}(E = n) = \binom{S}{n} P_{infection}^n (1 - P_{infection})^{S-n}, \quad n \in \{0, 1, \dots, S\}.$$

The main assumptions behind the classical Wells-Riley model are:

- The air is well mixed, so that the concentration of pathogen in the air is spatially homogeneous.
- The concentration of pathogen in the air is in steady state during  $[0, T]$ , neglecting any transient dynamics before this steady-state concentration is actually reached. A transient version of the model was proposed by Gammaitoni and Nucci (1997); see Edwards et al. (2024c) for details.
- All individuals behave homogeneously, remain in the indoor environment for the duration of the interaction  $[0, T]$ , and the model parameters ( $I, b, q, T$ , and  $Q$ ) are known and fixed.

As described in Section 1, recent work in the literature has aimed to address some of these limitations. In particular, numerical approaches or Monte Carlo simulations have been leveraged to incorporate parametric uncertainty and population heterogeneity. On the other hand, a stochastic, mathematical formulation has been recently proposed by Edwards et al. (2024c) when revisiting the Wells-Riley model which also addresses some of these points. In particular, Edwards et al. (2024c) considered:

- The situation where the infector(s) leave(s) at time  $T$ , but the susceptible individuals remain in the room until  $T + t$  for some  $t \geq 0$ .
- The scenario where either  $b, q$  or  $T$  are random parameters, following either an Exponential or Erlang distribution. This could represent situations where there is population heterogeneity or to incorporate stochastic human behavior. Analytic results for the expected per-capita probability of infection and the number of exposures were obtained.

It is interesting to note, however, that if a parameter in  $\{I, b, q, T, Q\}$  is considered to be random, for example due to uncertainty, the per-capita infection risk  $P_{infection}$  given by Eq. (1) is also a random variable, defined in  $(0, 1)$  (i.e., representing the possible values that  $P_{infection}$  can take). That is, if a parameter such as  $q$  is random (i.e., infectiousness is unknown/stochastic), and follows a particular probability distribution, the per-capita infection risk for each susceptible individual,  $P_{infection}$ , would be uncertain and depend on the specific value of the random parameter  $q$  during the specific interaction. For each (random) value of  $q$ ,  $P_{infection}$  would take a different value in  $(0, 1)$ . Thus, Edwards et al. (2024c) provided an expected (i.e. average) value of this infection risk probability,  $\mathbb{E}[P_{infection}]$ , when either  $b, q$  or  $T$  are exponentially- or Erlang-distributed. Here, we propose to go beyond this approach and better capture the uncertainty in infection risk by considering  $P_{infection}$  as a random variable in  $(0, 1)$  and calculating its density function (instead of just its expected value). We consider here Gamma-distributed parameters, going beyond Exponential or Erlang distributions. The exponential distribution is obtained by setting the Gamma shape parameter to 1, whereas the Erlang distribution represents a Gamma distribution with integer shape parameter. Gamma distributions yield better parametric flexibility when modelling uni-modal parametric regimes in the non-negative line, while allowing for some analytical tractability in Section 2. We note that multi-modal regimes can also be considered through linear combinations of Gamma distributions (see Section 2.1.1 and Appendix A1.1). It may be possible to extend some of our results to alternative distributions, as discussed in Section 4. We also consider the possibility of up to two parameters being simultaneously unknown/uncertain rather than just one.

From now on, we classify parameters in the Wells-Riley model as *population-related* (i.e., mainly related to the individuals involved in the interaction and their behaviour:  $I, b, q, T$ ) or *environmental* ( $Q$ ). In Subsection 2.1, we present equations which define the density function

and mean value of the per-capita probability of infection  $P_{infection}$ , as well as the probability of observing exactly  $n$  exposures  $\mathbb{P}(E = n)$  when we have a single, random population-related parameter. Subsection 2.2 contains similar results when focusing on the environmental parameter  $Q$ , whereas in Subsection 2.3, we consider pairs of simultaneously random parameters.

## 2.1 | Uncertainty in population-related parameters

We consider here uncertainty in one of the population-related parameters  $\{l, b, q, T\}$ . Since these appear in the numerator of the exponential function in the per-capita infection risk probability in Equation (1), the analysis related to each of these parameters is completely symmetric, with the only difference being that  $l$  takes discrete values whereas parameters  $\{b, q, T\}$  take continuous ones.

### 2.1.1 | Uncertainty in breathing rate $b$ , quanta emission rate $q$ or exposure time $T$ .

We provide here results for the case where the quanta generation rate  $q$  is random. Since exactly the same arguments apply to parameters  $b$  and  $T$ , analogous results are omitted here. Uncertainty in  $q$  might be due to uncertainty in estimating its precise value, or alternatively represent population heterogeneity where different individuals in the population might have different quanta generation rates. In particular, we consider the situation where the quanta generation rate follows a Gamma distribution

$$q \sim \text{Gamma}(k, \lambda),$$

with shape parameter  $k > 0$  and rate parameter  $\lambda > 0$ . That is,  $q$  has probability density function

$$f_q(x) = \frac{\lambda^k x^{k-1} \exp(-\lambda x)}{\Gamma(k)}, \quad x > 0,$$

where  $\Gamma(k)$  is the Gamma function. We note that the particular case where  $k$  is a positive integer leads to an Exponential,  $q \sim \text{Exp}(\lambda)$ , if  $k = 1$  or Erlang,  $q \sim \text{Erlang}(k, \lambda)$ , if  $k = 2, 3, \dots$ , distribution. As discussed above, if  $q$  is random the per-capita infection risk

$$P_{infection} = 1 - e^{-\frac{lbq}{Q}T},$$

is also random. That is,  $P_{infection}$  is a random variable which can take different values in  $(0, 1)$  depending on the value of  $q$  that happens to "occur" during a specific indoor interaction. The aim here is to capture the uncertainty in  $P_{infection}$  caused by the randomness of  $q$ , by computing the probability density function of the random variable  $P_{infection}$ . For  $q \sim \text{Gamma}(k, \lambda)$ , this can be calculated as (see Appendix A1.1)

$$f_{P_{infection}}(p) = \left(\frac{\lambda Q}{lbT}\right)^k \frac{[-\ln(1-p)]^{k-1} (1-p)^{\frac{\lambda Q}{lbT}-1}}{\Gamma(k)}, \quad (2)$$

such that  $p \in (0, 1)$ . This density function allows one to estimate how likely different values of  $P_{infection}$  are within the interval  $(0, 1)$ , based on the distribution of  $q$ . We note that this expression is consistent with an analogous result obtained by Nicas (1996) in the context of dose-response models, when considering a Gamma-distributed dose. As expected, the density function in Equation (2) depends on parameters  $\{l, b, Q, T\}$  but does not depend on  $q$  itself, since it is not a fixed parameter value here but a random parameter instead. Thus, the density function of the per-capita infection risk probability depends instead on the extra parameters  $(k, \lambda)$  governing the distribution of  $q$ . From this, the "expected" or "average" per-capita infection risk is (see Appendix A1.1)

$$\mathbb{E}[P_{infection}] = 1 - \left(\frac{\lambda}{\lambda + \frac{lbT}{Q}}\right)^k. \quad (3)$$

This expression provides an *exact* average infection risk while accounting for the randomness or uncertainty in parameter  $q$ , whereas the full infection risk profile is captured by the density function of  $P_{infection}$  given by Equation (2).

It is interesting to explore what would happen if one considered the average quanta generation rate  $\bar{q} = \mathbb{E}[\text{Gamma}(k, \lambda)] = k/\lambda$  as a representative parameter value for  $q$ , and plugged it into the classical Wells-Riley approach to obtain a potentially representative per-capita infection risk probability

$$\bar{P}_{infection}^{WR} = 1 - e^{-\frac{lbT}{Q}\bar{q}} = 1 - e^{-\frac{lbT}{Q} \frac{k}{\lambda}}.$$

This is a typical approach in the literature when implementing the classical Wells-Riley framework to estimate infection risk under parameter uncertainty. Interestingly, it can be shown that  $\bar{P}_{infection}^{WR}$  will always overestimate the "true" mean  $\mathbb{E}[P_{infection}]$  in Equation (3) (which represents the *exact* average infection risk probability which fully accounts for the randomness in  $q$ ); that is,

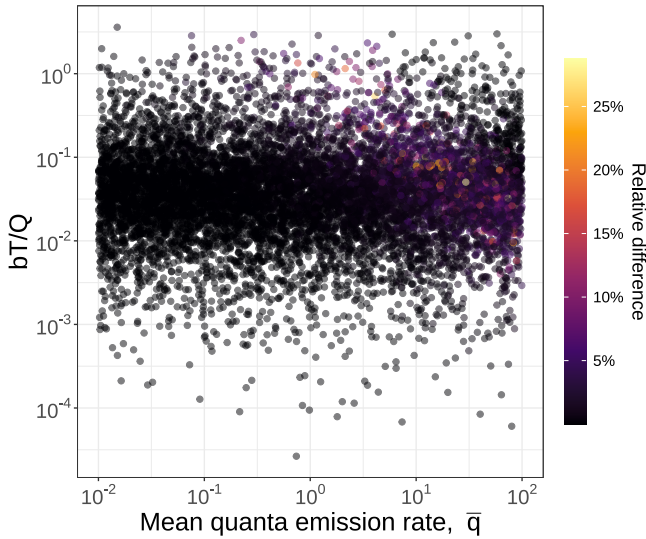
$$\bar{P}_{infection}^{WR} > \mathbb{E}[P_{infection}]$$

for all parameter regimes  $(l, b, T, Q, k, \lambda)$ . In fact, this holds for any general, positive-valued distribution of  $q$  (equivalently, of  $b$  or  $T$ ), and not just a Gamma distribution, since this relies only upon the concavity of the Wells-Riley equation and the use of Jensen's inequality; see Appendix A5.1.

For illustrative purposes, we plot in Figure 1 the relative difference,

$$\text{Relative difference (\%)} = \frac{\bar{P}_{infection}^{WR} - \mathbb{E}[P_{infection}]}{\mathbb{E}[P_{infection}]} \times 100,$$

between the classical approach,  $\bar{P}_{infection}^{WR}$ , and the analytical mean infection risk,  $\mathbb{E}[P_{infection}]$  in Eq. (3), for parameter values  $(k, \lambda, b, T, Q)$  randomly sampled within relatively wide ranges, and average quanta generation rates in  $\bar{q} \in [0.01, 100]$  quanta  $h^{-1}$  to represent a wide range of emission rates. As expected, strictly positive relative differences correspond to infection risk overestimation by the classical approach versus its exact average counterpart. The plot suggests that parametric



**FIGURE 1** Scatter plot for the relative difference between  $\bar{P}_{infection}^{WR}$  (classical Wells-Riley approach with constant, average quanta generation rate  $\bar{q} = k/\lambda$ ) and the exact analytical mean infection risk  $\mathbb{E}[P_{infection}]$  in Equation (3), versus  $\bar{q}$  and  $\frac{bT}{Q}$ . The shape and rate parameters  $(k, \lambda)$  of the Gamma distribution of  $q$ , as well as parameters  $(b, T, Q)$ , are randomly sampled within relatively wide ranges. In particular,  $\log_{10}(\bar{q}) \in [-2, 2]$ ,  $k \in [1, 20]$  (such that we set  $\lambda = k/\bar{q}$  and explore  $\bar{q} \in [0.01, 100]$  quanta  $h^{-1}$ ),  $b \in [0.2, 0.6]$   $m^3h^{-1}$ ,  $Q \in [10, 1000]$   $m^3h^{-1}$  (approximately  $ACH \in [0.1, 10]$  for a volume of  $V \approx 100$   $m^3$  (Edwards 2024b)), and  $T \in [0.1, 100]$  hours.

choices of  $(b, T, Q)$  (leading to relatively large values of  $e^{\frac{bT}{Q}}$ ), combined with sufficiently high  $\bar{q}$ , lead to larger relative differences. In particular, we observe overestimations of the analytical mean by the classical approach on the magnitude of 5–27% for a localised “strip” in the upper right region in Figure 1. Below and to the left of this strip, as well as above and to the right, we observe little to no relative difference. We note that these two darker regions correspond to a combination of low (below the strip) and high (above the strip) values of  $bT/Q$  and  $\bar{q}$ , resulting in both estimates,  $\bar{P}_{infection}^{WR}$  and  $\mathbb{E}[P_{infection}]$ , tending towards zero or one. The region of  $bT/Q$  at which overestimations occur are dependent on the value of  $\bar{q}$ . It is also interesting to note that nearby dots in Figure 1 can still have different colours indicating different relative differences. This is due to the particular distribution of  $q$  corresponding to those dots, which is determined by the specific values of  $(k, \lambda)$ . Wider Gamma distributions around  $\bar{q}$  (obtained for smaller values of  $k$ , with  $\lambda = k/\bar{q}$ ) translate to higher skewness and variance, representing higher uncertainty around  $\bar{q}$ . Overall, our results suggest that larger variance and skew of  $q$  lead to larger relative differences between  $\bar{P}_{infection}^{WR}$  and  $\mathbb{E}[P_{infection}]$ . In particular, the classical approach with average  $\bar{q}$  does not appropriately account for the possibility of small values of  $q$  happening with positive probability, which would lead to low infection risk in reality; thus,  $\bar{P}_{infection}^{WR}$  can overestimate the real average infection risk

in these situations, better captured by  $\mathbb{E}[P_{infection}]$ . It should be mentioned that for the values considered in Figure 1, relative differences for larger  $\bar{q}$  and  $bT/Q$  correspond to larger absolute differences in the infection risk estimates. The relative overestimations for different values of  $\bar{q}$  and  $bqT/Q$  have been analysed in further detail in Appendix A5.1. Since these infection risk estimates may be used in practice to propose or quantify the impact of specific mitigation strategies (Gammaitoni and Nucci 1997, Chen et al. 2006, Bazant and Bush 2021), our results highlight the importance of accounting for parametric uncertainty when estimating infection risk; in particular, implementing the classical Wells-Riley approach under average quanta generation rates seems to lead to overestimations of up to 27% for the parameter values considered in Figure 1.

Along with the infection risk density and mean infection risk (Equations (2) and (3)), the probability of observing exactly  $n$  exposures among  $S$  susceptible individuals during  $[0, T]$  is (see Appendix A1.1)

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left( \frac{\lambda}{(S-n+i)\frac{bT}{Q} + \lambda} \right)^k, \quad (4)$$

for  $n \in \{0, 1, \dots, S\}$ . We note that Equations (3) and (4) are consistent with those previously derived by Edwards et al. (2024c), whereas Equation (2) is a novel result which allows one to analytically capture the uncertainty that random  $q$  generates on the per-capita infection risk estimate  $P_{infection}$ . Random parameters  $b$  or  $T$  would lead to analogous results to those above, and are thus not reported explicitly here. It is also worth noting that the density function in Equation (2) is consistent with that derived by Nicas (1996) in the context of a dose-response model, where the dose is considered to be random instead of the specific parameters  $(b, q, T, Q)$  in the Wells-Riley framework.

Finally, we note that random  $q$  might also arise as a result of mitigations, human behaviour or population heterogeneity, and our mathematical formulation can be adapted to consider these situations. For example, one may consider that the infector wears a mask with probability  $w_{mask} \in [0, 1]$ , so that the quanta generation rate is

$$q_{eff} = q_{no\ mask}(w_{mask}(1 - \eta) + (1 - w_{mask})),$$

where  $\eta \in (0, 1)$  represents the reduction in quanta generation due to mask filtration efficacy, and  $q_{no\ mask}$  is the underlying quanta emission rate without mask wearing. For simplicity, we do not consider mask-wearing by susceptible individuals and focus instead on the impact of mask-wearing by the infector on infection risk. In this scenario, the Wells-Riley equation effectively becomes ((Gammaitoni and Nucci 1997, Bazant and Bush 2021, Rothamer et al. 2021, Huang et al. 2022))

$$P_{infection} = 1 - e^{-\frac{I b q_{eff} T}{Q}}.$$

One can still consider uncertainty in the underlying quanta emission rate,  $q_{no\ mask} \sim \text{Gamma}(k, \lambda)$ , while incorporating stochastic human behaviour through probability  $w_{mask}$ , so that the effective quanta emission

rate,  $q_{\text{eff}}$ , is a random variable with density function

$$f_{q_{\text{eff}}}(q) = w_{\text{mask}} \frac{1}{1-\eta} f_{\text{dno-mask}}\left(\frac{q}{1-\eta}\right) + (1-w_{\text{mask}}) f_{q_{\text{no-mask}}}(q).$$

Our results then can be naturally extended so that (see Appendix A1.1)

$$f_{P_{\text{infection}}}(p) = \frac{(-\ln(1-p))^{k-1}}{\Gamma(k)} \left(\frac{\lambda Q}{lbT}\right)^k \left( w_{\text{mask}} (1-\eta)^{-k} (1-p)^{\frac{\lambda Q}{lbT(1-\eta)}-1} + (1-w_{\text{mask}})(1-p)^{\frac{\lambda Q}{lbT}-1} \right), \quad \text{for } p \in (0, 1), \quad (5)$$

$$\mathbb{E}[P_{\text{infection}}] = 1 - \left[ w_{\text{mask}} \left( \frac{\lambda}{\lambda + \frac{lbT}{Q}(1-\eta)} \right)^k + (1-w_{\text{mask}}) \left( \frac{\lambda}{\lambda + \frac{lbT}{Q}} \right)^k \right], \quad (6)$$

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left[ w_{\text{mask}} \left( \frac{\lambda}{\lambda + (S-n+i) \frac{lbT}{Q}(1-\eta)} \right)^k + (1-w_{\text{mask}}) \left( \frac{\lambda}{\lambda + (S-n+i) \frac{lbT}{Q}} \right)^k \right], \quad \text{for } n = 0, 1, \dots, N. \quad (7)$$

A slightly different application arises when one aims to represent population heterogeneity, so that  $q \sim \text{Gamma}(k, \lambda)$  is a population-level description of infectiousness. However, since the term  $ql$  in Equation (1) represents the rate of quanta generation by all  $l$  infectious individuals during the indoor interaction, these individuals are all still assumed to generate quanta at the same rate  $q$ , which can be globally sampled from  $\text{Gamma}(k, \lambda)$ . This is appropriate when  $l = 1$ , so that the infector's infectiousness is unknown and sampled from a population-level distribution. On the other hand, for scenarios where more than one,  $l > 1$ , infectious individual is assumed to be present within the interaction, it is more appropriate to replace the term  $ql$  by  $\sum_{i=1}^l q_i$ , where  $q_i$  represents the quanta generation rate of each individual  $i$ , which can be sampled from the population-level distribution  $\text{Gamma}(k, \lambda)$ , so that  $q_i \sim \text{Gamma}(k, \lambda)$  are independent and identically distributed random variables. Although not the main focus of this paper, we generalise the results above to this situation in Appendix A4.

## 2.1.2 | Disease prevalence

While it is common to assume that the number of infectors  $l$  during the indoor interaction is known, this is not typically the case in reality when trying to assess risk in different settings *a priori*. In particular, by assuming that a single infector is present in the room (i.e., by setting  $l = 1$ ), the per-capita probability of infection may well be overestimated, since one of the main factors impacting infection risk is the probability of an infector being present in the room at all. Thus, it may be more appropriate to consider a disease prevalence  $\rho$ , understood here as the proportion of individuals in the population being infectious at any given time, so that (assuming independence) the number of infectious individuals in the room is random and follows a binomial distribution

$$l \sim \text{Bin}(N, \rho),$$

where  $N$  is the total number of individuals involved in the indoor interaction. The probability mass function of  $l$  is given by:

$$\mathbb{P}(l = i) = \binom{N}{i} \rho^i (1-\rho)^{N-i}, \quad i \in \{0, 1, \dots, N\}.$$

Importantly, this incorporates the possibility that there might not be infectors during the interaction ( $l = 0$ ), which occurs with probability  $(1-\rho)^N$ , or that there might be more than one ( $l > 1$ ). Since  $l$  is a discrete random variable with support  $\{0, 1, \dots, N\}$ , the per-capita infection risk  $P_{\text{infection}}$  is also a discrete random variable on this occasion. Thus, its probability mass function is (see Appendix A1.2)

$$f_{P_{\text{infection}}}(p) = \binom{N}{i} \rho^i (1-\rho)^{N-i} \quad (8)$$

with possible (positive probability) values  $p = 1 - e^{-\frac{bqT}{Q}i}$  where  $i \in \{0, 1, \dots, N\}$ . The expected per-capita infection risk is (see Appendix A1.2)

$$\mathbb{E}[P_{\text{infection}}] = 1 - \left[ \rho e^{-\frac{bqT}{Q}} + (1-\rho) \right]^N. \quad (9)$$

The probability that there are exactly  $n$  exposures is given in its raw form by (see Appendix A1.2)

$$\mathbb{P}(E = n) = \binom{N}{n} \sum_{i=0}^{N-n} \binom{N-n}{i} (1 - e^{-\frac{bqT}{Q}i})^n e^{-\frac{bqT}{Q}i(N-i-n)} \rho^i (1-\rho)^{N-i}, \quad (10)$$

for  $n = 0, 1, \dots, N$ .

## 2.2 | Uncertainty in ventilation rate

Here, we consider the situation where the ventilation rate  $Q$  is uncertain (i.e., random), for example, in situations where this may depend on external factors (e.g., weather) or human behaviour (e.g., windows opening). We consider a Gamma distribution

$$Q \sim \text{Gamma}(r, \gamma),$$

with shape parameter  $r > 0$  and rate parameter  $\gamma > 0$ . Since  $Q$  appears in the denominator (instead of the numerator) of the exponent in the per-capita infection risk probability in Equation (1), the results obtained in Section 2.1 for  $b, q$  and  $T$  do not directly apply here. Instead, for  $Q \sim \text{Gamma}(r, \gamma)$ , the density function of the per-capita infection risk  $P_{\text{infection}}$  is (see Appendix A2)

$$f_{P_{\text{infection}}}(p) = \frac{(\gamma lbqT)^r}{\Gamma(r)} \frac{e^{\frac{\gamma lbqT}{\ln(1-p)}}}{(1-p)^{-\ln(1-p)} [r+1]}, \quad p \in (0, 1). \quad (11)$$

As a result, the average per-capita infection risk is given by (see Appendix A2)

$$\mathbb{E}[P_{\text{infection}}] = 1 - \frac{2}{\Gamma(r)} (\gamma lbqT)^{r/2} K_r \left( 2\sqrt{\gamma lbqT} \right), \quad (12)$$

where  $K_\theta(\zeta)$  is the modified Bessel function of the second kind (NIST (2025a))

$$K_\theta(\zeta) = \frac{1}{2} \left( \frac{\zeta}{2} \right)^\theta \int_0^\infty u^{-(\theta+1)} e^{-\left(u - \frac{\zeta^2}{4u}\right)} du.$$

Finally, the probability of observing exactly  $n$  exposures is (see Appendix A2)

$$\mathbb{P}(E = n) = \frac{2}{\Gamma(r)} \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} a_i^{r/2} K_r(2\sqrt{a_i}) \quad (13)$$

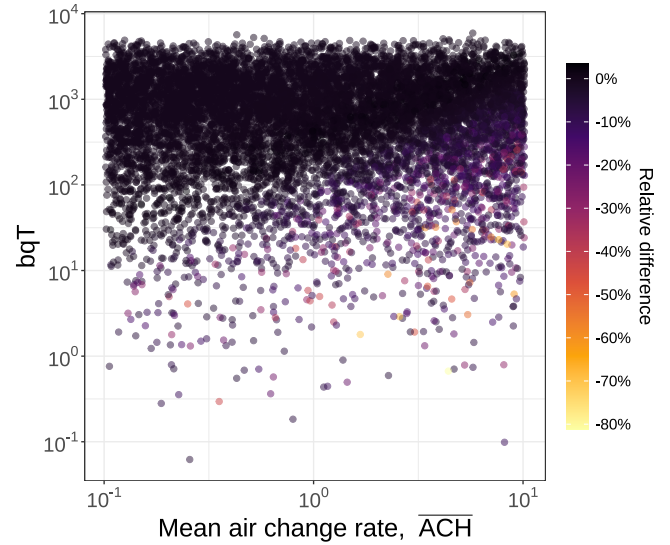
for  $n = 0, 1, \dots, S$ , where  $a_i = \gamma lbqT(S - n + i)$ . We note that for the special case  $n = S$ , the factor  $(S - n + i)$  equals 0 when  $i = 0$ . Since  $K_\theta(\zeta) \sim 2^{\theta-1} \Gamma(\theta) \zeta^{-\theta}$  when  $\zeta \rightarrow 0$  (NIST (2025b)), one has  $a_i^{r/2} K_r(2\sqrt{a_i}) \rightarrow \frac{1}{2} \Gamma(r)$  when  $(S - n + i) \rightarrow 0$  (meaning  $a_i \rightarrow 0$ ), so that in the  $i = 0$  term one gets  $\frac{2}{\Gamma(r)} \times \frac{1}{2} \Gamma(r) = 1$  (since the other elements in the term are also 1).

Similar to our analysis in Section 2.1 for random  $q$ , we can investigate here the relative difference between the infection risk estimated via the classical Wells-Riley approach

$$\bar{P}_{infection}^{WR} = 1 - e^{-\frac{lbqT}{\bar{Q}}},$$

with average ventilation rate  $\bar{Q} = r/\gamma$ , and the *exact* analytical average infection risk  $\mathbb{E}[P_{infection}]$  in Equation (12) (which accounts for the randomness in  $Q$ ). This relative difference is plotted in Figure 2. While considering random  $q$  systematically led to overestimating infection risk under the classical approach (due to Jensen's inequality), the fact that  $P_{infection}$  is not necessarily concave on  $Q$  means that it is not possible to determine *a priori* if the relative difference would be positive or negative depending on the particular values of  $(r, \gamma, b, q, T)$ . Still, we tend to see infection risk underestimation (i.e., negative relative difference) in the case of random ventilation for most of the parametric regimes considered in Figure 2. This is due to  $P_{infection}$  being convex on  $Q$  for the majority of parameter values considered. In fact, where there are overestimations (positive relative errors corresponding to black-coloured points), these are very close to zero, and occur for very small ventilation rates combined with high values of  $bqT$ , such that both the classical and analytical estimates each tend towards 1. Outside of this specific region, it appears as though the size of the relative difference is largely independent of the specific values of  $bqT$  and  $\overline{ACH}$ . In fact, as was the case for random quanta, it increases with higher variance and positive skew of the Gamma distribution of the ventilation rate (something we cannot see by plotting against  $\overline{ACH}$  alone). The magnitude of these inaccuracies are far larger in the case of random  $Q$  compared to that of random  $q$  (relative underestimations of 80% versus overestimations of 25% of the analytical  $\mathbb{E}[P_{infection}]$ , see Figure 1 and 2). This is because the infection risk does not vary so significantly in the region of  $\bar{q}$  as it does for  $\bar{Q}$ : variations in  $Q$  leading to small ventilation rates produce higher stochasticity in the infection risk than variation in  $q$ , since in the case of random  $Q$  scenarios with very low ventilation can lead to notably high infection risk. In Appendix A5.2, the size of the relative difference has been investigated further for the parameter space considered here.

Finally, we note that if a particular model parameter has a non-random component, our results can be easily adapted. For example, we explore in Section 3 the situation where natural ventilation in an indoor environment might be stochastic, whereas mechanical ventilation can be more precisely controlled or estimated. In this situation, one



**FIGURE 2** Scatter plot for the relative difference between  $\bar{P}_{infection}^{WR}$  (classical Wells-Riley approach with constant, average ventilation rate  $\bar{Q} = r/\gamma$ ) and the *exact* analytical mean infection risk  $\mathbb{E}[P_{infection}]$  in Equation (12), versus  $\bar{Q}$  and  $bqT$ . The shape and rate parameters  $(r, \gamma)$  of the Gamma distribution of  $Q$ , as well as parameters  $(b, q, T)$ , are randomly sampled within relatively wide ranges. In particular,  $\log(\bar{Q}) \in [1, 3]$ ,  $r \in [1, 20]$  (such that we set  $\gamma = r/\bar{Q}$  and explore  $\overline{ACH} \in [0.1, 10]$  for a room volume  $V \approx 100 \text{ m}^3$  (Edwards 2024b)),  $b \in [0.2, 0.6] \text{ m}^3 \text{ h}^{-1}$ ,  $q \in [0.01, 100] \text{ quanta h}^{-1}$ , and  $T \in [0.1, 100] \text{ hours}$ .

might define the ventilation rate as  $Q = Q_{nat} + Q_{mech}$  where  $Q_{nat} \sim \text{Gamma}(r, \gamma)$  and  $Q_{mech}$  is constant, so that

$$P_{infection} = 1 - e^{-\frac{lbqT}{Q_{nat} + Q_{mech}}}. \quad (14)$$

Our results can be extended to this situation, so that for example the probability density function of the infection risk is given as

$$f_{P_{infection}}(p) = \frac{\gamma^r \left( \frac{-lbqT}{\ln(1-p)} - Q_{mech} \right)^{r-1} e^{\frac{\gamma lbqT}{\ln(1-p)} + \gamma Q_{mech}} lbqT}{\Gamma(r)(1-p)(\ln(1-p))^2}, \quad p \in (0, 1); \quad (15)$$

see Appendix A2.1 for further details.

## 2.3 | Uncertainty in two model parameters

We now consider the situation where two independent parameters are simultaneously random. The aim is to investigate their compounding effect on infection risk. We can consider two distinct cases: two population-related parameters are random, or one population-related parameter is random as well as ventilation (the environmental parameter). Let us start with the former case where two population-related parameters  $\{b, q, T\}$  are random; say the quanta emission rate  $q$  and the exposure time  $T$  without any loss of generality (since equivalent

arguments would apply to  $\{b, q\}$  or  $\{b, T\}$ . In particular, we consider

$$\begin{aligned} q &\sim \text{Gamma}(k, \lambda), \\ T &\sim \text{Gamma}(\ell, \mu), \end{aligned}$$

for shapes  $k, \ell > 0$  and rates  $\lambda, \mu > 0$ . In this scenario, the density function of the per-capita infection risk  $P_{infection}$  is (see Appendix A3)

$$f_{P_{infection}}(p) = \frac{2(\lambda\mu)^{\frac{k+\ell}{2}} \left[ -\frac{1}{c} \ln(1-p) \right]^{\frac{k+\ell}{2}-1} K_{k-\ell} \left( 2\sqrt{d[-\ln(1-p)]} \right)}{c\Gamma(k)\Gamma(\ell)(1-p)} \quad (16)$$

with  $c = \frac{lb}{Q}$ ,  $d = \frac{\lambda\mu}{c}$  and for  $p \in (0, 1)$ . The average per-capita infection risk is given by (see Appendix A3)

$$\mathbb{E}[P_{infection}] = 1 - d^\ell U(\ell, \ell - k + 1, d) = 1 - d^k U(k, k - \ell + 1, d) \quad (17)$$

where  $U(a, b, z)$  is the confluent hypergeometric function of the second kind (NIST (2025c)), sometimes referred to as the *Tricomi* function

$$U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zv} v^{a-1} (1+v)^{b-a-1} dv.$$

The probability of observing exactly  $n$  exposures is given by (see Appendix A3)

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} y_i^\ell U(\ell, \ell - k + 1, y_i) \quad (18)$$

$$= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} y_i^k U(k, k - \ell + 1, y_i) \quad (19)$$

for  $n = 0, \dots, S$ , where  $y_i = \frac{d}{(S-n+i)}$ . We note here that when  $S-n+i = 0$  we use the limit of the Tricomi function (NIST (2025d))  $U(a, b, z) \sim z^{-a}$  when  $z \rightarrow \infty$ . Given that  $i = 0$  and  $S = n$  (where  $S - n + i = 0$ ), the first term in each sum would evaluate to  $y_0^\ell y_0^{-\ell} = y_0^k y_0^{-k} = 1$ . We also note that each of the two forms proposed for  $\mathbb{E}[P_{infection}]$  and  $\mathbb{P}(E = n)$  is derived by first solving the underlying double integral either with respect to  $q$  or  $T$ . The equivalence of these two forms is further validated by Kummer's transformation (NIST 2025e)  $U(a, b, z) = z^{1-b} U(1+a-b, 2-b, z)$ .

Finally, it is important to note that some of our results above simplify when one considers an Erlang-distributed parameter instead of Gamma (i.e., when the shape parameter is an integer). For example, Equations (17)-(19) may be written as finite sums if  $k \in \mathbb{N}$  (i.e.,  $q$  is Erlang-distributed):

$$\mathbb{E}[P_{infection}] = 1 - \frac{e^d}{(k-1)!} \sum_{i=0}^{k-1} \binom{k-1}{i} (-1)^i d^{i+\ell} \Gamma(k-\ell-i, d), \quad (20)$$

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n \sum_{j=0}^{k-1} \frac{(-1)^{i+j}}{(k-1)!} \binom{n}{i} \binom{k-1}{j} y_i^{\ell+j} e^{y_i} \Gamma(k-\ell-j, y_i), \quad (21)$$

for  $n = 0, \dots, S$ , where

$$\Gamma(s, x) = \int_x^\infty w^{s-1} e^{-w} dw$$

is the upper incomplete Gamma function (NIST 2025f). If  $T$  is Erlang-distributed instead of  $q$  (that is, if  $\ell \in \mathbb{N}$  instead of  $k$ ) one gets analogous results omitted here. Further simplified expressions are obtained if both parameters are exponentially distributed,  $k = \ell = 1$ , in which case

$$f_{P_{infection}}(p) = \frac{2d}{(1-p)} K_0 \left( 2\sqrt{d[-\ln(1-p)]} \right), \quad p \in (0, 1),$$

$$\mathbb{E}[P_{infection}] = 1 - de^d \Gamma(0, d),$$

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} y_i e^{y_i} \Gamma(0, y_i),$$

for  $n = 0, \dots, S$ .

Finally, we consider here the case where one population-related parameter from  $\{b, q, T\}$  is random alongside the ventilation rate  $Q$ . Without loss of generality we consider  $q$  to be random, in particular

$$\begin{aligned} q &\sim \text{Gamma}(k, \lambda), \\ Q &\sim \text{Gamma}(r, \gamma). \end{aligned}$$

The density function of the per-capita infection risk  $P_{infection}$  is (see Appendix A3)

$$f_{P_{infection}}(p) = \frac{\Gamma(k+r)}{\Gamma(k)\Gamma(r)} \frac{\lambda^k (\gamma lbT)^r [-\ln(1-p)]^{k-1}}{(1-p)([-\ln(1-p)]\lambda + \gamma lbT)^{k+r}} \quad (22)$$

for  $p \in (0, 1)$ . The expected per-capita infection risk is given by (see Appendix A3)

$$\mathbb{E}[P_{infection}] = 1 - \left( \frac{\gamma lbT}{\lambda} \right)^r \frac{\Gamma(k+r)}{\Gamma(r)} U\left(k+r, r+1, \frac{\gamma lbT}{\lambda}\right). \quad (23)$$

The probability of  $n$  exposures is (see Appendix A3)

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n \binom{n}{i} (-1)^i \alpha_i^r \frac{\Gamma(k+r)}{\Gamma(r)} U\left(k+r, r+1, \alpha_i\right), \quad (24)$$

for  $n = 0, \dots, S$ , where  $\alpha_i = \frac{\gamma lbT}{\lambda} (S - n + i)$ . If  $k, r \in \mathbb{N}$ , then Equations (23) and (24) simplify to

$$\mathbb{E}[P_{infection}] = 1 - \frac{e^{\frac{\gamma lbT}{\lambda}}}{(r-1)!} \sum_{i=0}^{r+k-1} \binom{r+k-1}{i} (-1)^i \left( \frac{\gamma lbT}{\lambda} \right)^i \Gamma(r-i, \frac{\gamma lbT}{\lambda}), \quad (25)$$

$$\mathbb{P}(E = n) = \binom{S}{n} \frac{1}{(r-1)!} \sum_{i=0}^n \sum_{j=0}^{r+k-1} (-1)^{i+j} \binom{n}{i} \binom{r+k-1}{j} \alpha_i^j e^{\alpha_i} \Gamma(r-j, \alpha_i), \quad (26)$$

for  $n = 0, \dots, S$ . Additionally, when  $k = r = 1$

$$f_{P_{infection}}(p) = \frac{\lambda \gamma lbT}{(1-p)([-\ln(1-p)]\lambda + \gamma lbT)^2}, \quad p \in (0, 1)$$

$$\mathbb{E}[P_{infection}] = e^{\frac{\gamma lbT}{\lambda}} \frac{\gamma lbT}{\lambda} \Gamma\left(0, \frac{\gamma lbT}{\lambda}\right),$$

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left( 1 - e^{\alpha_i} \alpha_i \Gamma(0, \alpha_i) \right),$$

for  $n = 0, \dots, S$ , where we have simplified Equations (25) and (26), along with the following property (NIST 2025g); if  $s \in \mathbb{N}$

$$\Gamma(s, x) = (s-1)! e^{-x} \sum_{i=0}^{s-1} \frac{x^i}{i!}.$$

If instead  $s = -n$  with  $n \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$  then (NIST (2025h))

$$\Gamma(-n, x) = \frac{(-1)^n}{n!} \left( \Gamma(0, x) - e^{-x} \sum_{i=0}^{n-1} \frac{(-1)^i i!}{x^{i+1}} \right).$$

Each of these may be used to compute values for Equations (20) & (21) and (25) & (26), instead of trying to evaluate multiple incomplete Gamma functions.

**TABLE 1** Table summarising analytical results from Section 2.

Scenarios	Infection probability density, $f_{P_{Infection}}(p)$ , for $p \in (0, 1)$ unless otherwise stated	Expected infection probability, $\mathbb{E}[P_{Infection}]$	Probability of $n$ exposures, $\mathbb{P}(E = n)$ , for $n = 0, \dots, S$ unless otherwise stated
Random quanta emission rate, $q \sim \text{Gamma}(k, \lambda)$ . Constant $(l, b, T, Q)$ .	$\left(\frac{\lambda Q}{lbT}\right)^k \frac{[-\ln(1-p)]^{k-1} (1-p)^{\frac{\lambda Q}{lbT}-1}}{\Gamma(k)}$	$1 - \left(\frac{\lambda}{\lambda + \frac{lbT}{Q}}\right)^k$	$\binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left(\frac{\lambda}{(S-n+i)\frac{lbT}{Q} + \lambda}\right)^k$
Random ventilation rate, $Q \sim \text{Gamma}(r, \gamma)$ . Constant $(l, b, T, q)$ .	$\frac{(\gamma lbqT)^r}{\Gamma(r)} \frac{e^{\frac{\gamma lbqT}{(1-p)[-\ln(1-p)]}}}{(1-p)^{r-1}}$	$1 - \frac{2}{\Gamma(r)} (\gamma lbqT)^{r/2} K_r\left(2\sqrt{\gamma lbqT}\right)$	$\frac{2}{\Gamma(r)} \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} a_i^{r/2} K_r\left(2\sqrt{a_i}\right),$ where $a_i = \gamma lbqT(S - n + i)$
Random quanta emission rate, $q \sim \text{Gamma}(k, \lambda)$ . and exposure time $T \sim \text{Gamma}(\ell, \mu)$ . Constant $(l, b, Q)$ .	$\frac{2(\lambda\mu)^{\frac{k+\ell}{2}} \left[-\frac{1}{c} \ln(1-p)\right]^{\frac{k+\ell}{2}-1} K_{k-\ell}\left(2\sqrt{d[-\ln(1-p)]}\right)}{c\Gamma(k)\Gamma(\ell)(1-p)},$ where $c = \frac{lb}{Q}$ , $d = \frac{\lambda\mu}{c}$	$1 - d^\ell U(\ell, \ell - k + 1, d) = 1 - d^k U(k, k - \ell + 1, d)$	$\binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} y_i^\ell U(\ell, \ell - k + 1, y_i)$ $= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} y_i^k U(k, k - \ell + 1, y_i),$ where $y_i = \frac{d}{(S-n+i)}$
Random quanta emission rate, $q \sim \text{Gamma}(k, \lambda)$ . and ventilation rate $Q \sim \text{Gamma}(r, \gamma)$ . Constant $(l, b, T)$ .	$\frac{\Gamma(k+r)}{\Gamma(k)\Gamma(r)} \frac{\lambda^k (\gamma lbT)^r [-\ln(1-p)]^{k-1}}{(1-p)^{r-1} [\lambda + \gamma lbT]^{k+r}}$	$1 - \left(\frac{\gamma lbT}{\lambda}\right)^r \frac{\Gamma(k+r)}{\Gamma(r)} U\left(k+r, r+1, \frac{\gamma lbT}{\lambda}\right)$	$\binom{S}{n} \sum_{i=0}^n \binom{n}{i} (-1)^i \alpha_i^r \frac{\Gamma(k+r)}{\Gamma(r)} U\left(k+r, r+1, \alpha_i\right),$ where $\alpha_i = \frac{\gamma lbT}{\lambda}(S - n + i)$
Random number of infectors, $l \sim \text{Binomial}(N, \rho)$ . Constant $(q, b, T, Q)$ .	$\begin{cases} \binom{N}{i} \rho^i (1-\rho)^{N-i}, & \text{if } p = 1 - e^{-\frac{bqT}{Q}i}, i = 0, \dots, N \\ 0, & \text{otherwise.} \end{cases}$	$1 - \left[\rho e^{-\frac{bqT}{Q}} + (1-\rho)\right]^N$	$\binom{N}{n} \sum_{i=0}^{N-n} \binom{N-n}{i} \rho^i (1-\rho)^{N-i} \left(1 - e^{-\frac{bqT}{Q}i}\right)^n e^{-\frac{bqT}{Q}i(N-i-n)},$ for $n = 0, \dots, N$

### 3 | RESULTS

In Section 2, we have extended the Wells-Riley methodology while considering Gamma-distributed parameters, computing the density function of the per-capita infection risk probability, its mean, and the probability distribution of the number of exposures. These analytical results are summarised in Table 1, where results for a random quanta emission rate,  $q$ , are analogous to those for a random breathing rate,  $b$ , or exposure time,  $T$ . In this Section, we aim to explore the impact that uncertainty in these parameter values has on infection risk analytically, by considering Gamma distributions for some of these parameters. While other distributions may lead to slightly better representations, our aim is not to find the best distribution in each scenario, but to illustrate instead our analytical results in Section 2, while showing the impact of incorporating uncertainty in parameter values. We discuss how some of our results may be generalised to alternative distributions in Section 4. In this Section, we consider different parametric regimes inspired by scenarios related to healthcare and hospitality. First, we outline in Section 3.1 some of these parametric choices.

#### 3.1 | Model parameters

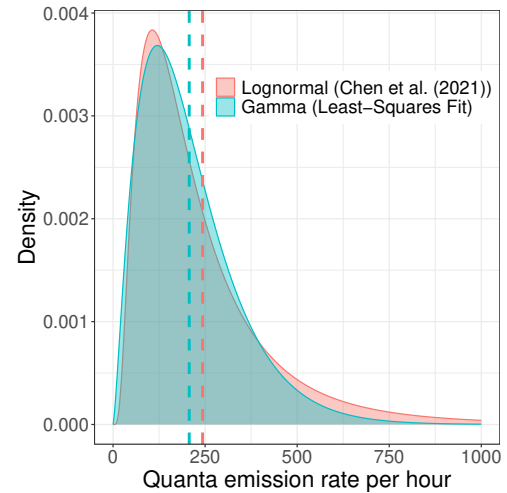
Unless stated otherwise, parameters  $b = 0.4 \text{ m}^3/\text{hour}$  (Pleil et al. (2021)) and  $l = 1$  are considered in most of our results. For the quanta emission rate of an infector, we note that quanta generation rates can vary widely across many different orders of magnitude depending on a multitude of factors, including: pathogen of interest (Mikszewski et al. 2022, Chen et al. 2006), differences in physical activity (Jones et al. 2024, Buonanno et al. 2020a, Aganovic et al. 2023), the stage of infection (Cheng et al. 2025, Ferretti et al. 2020, Sender et al. 2022, Wu et al. 2021, Euser et al. 2021) and individual heterogeneity such as age or gender (Jones et al. 2021b, Euser et al. 2021). For example, for SARS-CoV-2 the quanta emission rate across infectors has been described via a Lognormal distribution,  $LN(GM = 0.0092, GSD = 29)$ , by Jones et al. (2024), spanning several orders of magnitude across individuals. In particular, this distribution illustrates the significant variability across individuals, with its arithmetic mean  $\bar{q} = 2.70 \text{ quanta/hour}$  representing the 95<sup>th</sup> percentile of the distribution. Indeed, significantly larger estimates have been obtained in the literature when studying particular large outbreaks during the pandemic, such as the Skagit county choir (Miller et al. 2021) and the Diamond princess cruise ship (Chen et al. 2021), which would correspond to top percentiles of the Lognormal distribution above.

Since our aim is to highlight the impact of parameter uncertainty in those scenarios where infection risk is non-negligible, we focus here on scenarios where the infector is relatively highly infectious, corresponding to values near or above the mean value  $\bar{q} = 2.70 \text{ quanta/hour}$  estimated in Jones et al. (2024). In particular, values  $q = 1$  and  $q = 100 \text{ quanta/hour}$  are considered in some of our numerical results below. When aiming to illustrate the impact of uncertainty in  $q$ , we leverage instead the estimates provided by (Chen et al. 2021) for the Diamond

princess cruise ship; as such, the focus in this particular scenario in Section 3.2.2 is to consider uncertainty in  $q$  while focusing on a highly infectious individual, and our results in Figure 9 should be interpreted accordingly. In particular, to incorporate uncertainty while leveraging analytical results in Section 2, we use in Section 3.2.2 a Gamma distribution as an alternative to the Log-normal distribution estimated by (Chen et al. 2021),  $LN(184.64, 2.1)$ . Using the least squares method, we get

$$q \sim \text{Gamma}(\text{shape} = 2.39, \text{rate} = 0.012);$$

see Figure 3.

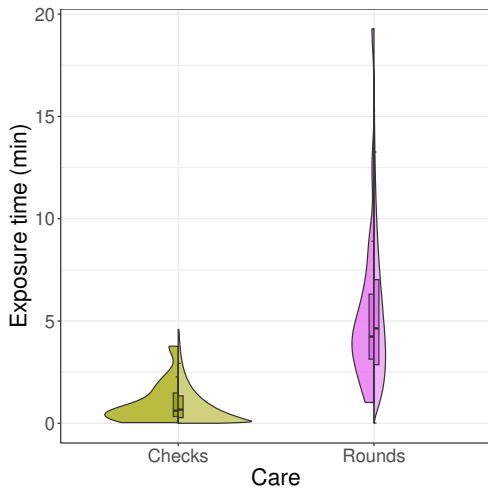


**FIGURE 3** Gamma distribution for  $q$  with shape  $k = 2.39$  and rate  $\lambda = 0.012$ , calibrated against the Lognormal distribution  $LN(GM = 184.64, ASD = 2.1)$  estimated by (Chen et al. 2021) for the Diamond princess cruise COVID-19 outbreak.

Finally, different parameter values (and distributions) for  $Q$  and  $T$  will be explored in Sections 3.2 and 3.3 when considering scenarios related to healthcare and hospitality. In particular, we consider in Section 3.2 a healthcare worker interacting with patients in a hospital ward, to illustrate the impact of uncertainty in ventilation, visit duration and infectiousness. In Section 3.3, we consider infection risk during meals, to illustrate the impact of random prevalence as well as uncertainty in two parameter values simultaneously.

##### 3.1.1 | Case Study 1: Healthcare visits

In Section 3.2, we consider a similar scenario to that considered by Edwards et al. (2024c) of a healthcare worker providing different types of care to patients in a hospital ward. In particular, and to illustrate how the duration of the indoor interaction may be uncertain, we leverage data from an observational study by King et al. (2021) which measured real and mock HCW visit times to a patient during various types of care (intravenous (IV) drip care, observational care and doctor's rounds). We



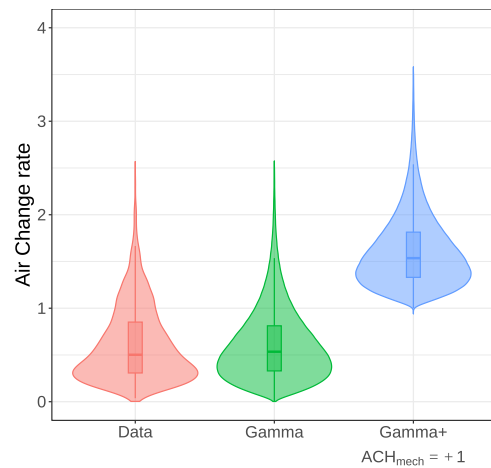
**FIGURE 4** Split violin plots showing measured visit times (left side; King et al. (2021)) with their best fitting Gamma distributions (right side) for each care type.

focus here on measurements corresponding to actual observational care for routine checks on patients (*Checks*) and doctor's rounds (*Rounds*), plotted in Figure 4. To account for this uncertainty when estimating infection risk while illustrating results in Section 2, we calibrate Gamma distributions to these two datasets by minimising the negative log likelihood. This leads to distributions (in units of *hours*)

$$\begin{aligned} T_{Checks} &\sim \text{Gamma}(\text{shape} = 0.99, \text{rate} = 60), \\ T_{Rounds} &\sim \text{Gamma}(\text{shape} = 2.51, \text{rate} = 28.1), \end{aligned}$$

which are plotted in Figure 4 against the corresponding datasets. In both cases, the distributions provide reasonable fits which allow one to capture the uncertainty around the corresponding mean visit duration and the unimodal shape of both datasets. We note that the duration of Checks  $T_{checks}$  has a narrower distribution (i.e. less uncertainty/variance) around  $\mathbb{E}[T_{Checks}] \approx 1 \text{ min}$ , whereas the duration of doctors' rounds is longer and more stochastic around its mean  $\mathbb{E}[T_{Rounds}] \approx 5.36 \text{ min}$ .

Finally, uncertainty in ventilation rate is incorporated in the health-care scenario by considering stochasticity in the natural ventilation rate as estimated by Edwards (2024b), while accounting for the possibility of some constant mechanical ventilation being present in the ward. In particular, we use natural ventilation rate estimates from Edwards (2024b), which accounted for weather factors such as wind direction, wind speed and outside temperature. A Gamma distribution is calibrated against this dataset by minimising the negative log likelihood. This resulted in an air change rate of  $ACH_{nat} \sim \text{Gamma}(2.52, 4.10)$  per hour, plotted in Figure 5 against the corresponding dataset. This distribution is able to capture both the unimodal shape of the dataset, with mode at around  $ACH = 0.3$ , the possibility of very low ACHs near zero, and the tail representing ACHs up to 2.6. For a typical hospital ward volume of



**FIGURE 5** Violin plots for the natural ventilation data (ACH in red, from Edwards (2024b)), and the calibrated gamma distribution  $ACH_{nat} \sim \text{Gamma}(2.52, 4.10)$ , with (blue,  $ACH_{mech} = +1$ ) and without (green) mechanical ventilation.

$V = 98.35 \text{ m}^3$  (Edwards 2024b), this leads to the ventilation rate

$$Q_{nat} \sim \text{Gamma}(2.52, 0.042) \text{ m}^3/\text{hour}.$$

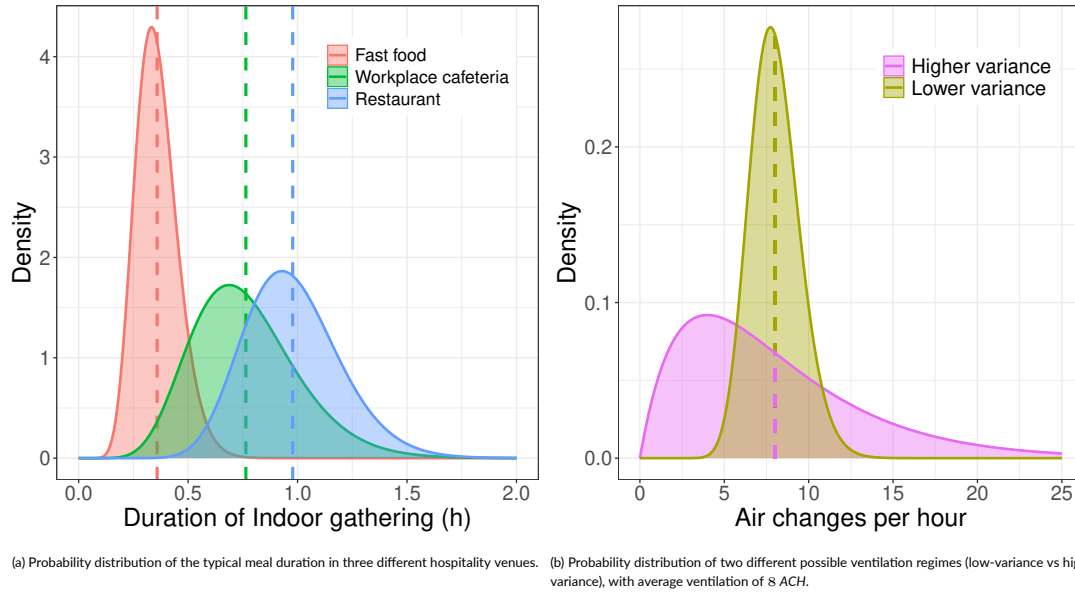
The possibility of adding some mechanical ventilation is depicted in Figure 5, which just shifts the Gamma distribution by a constant amount.

### 3.1.2 | Hospitality parameters

In Section 3.3, we consider a similar scenario to the one considered by Edwards et al. (2024c) to estimate infection risk during meals in different hospitality venues. In particular, Edwards et al. (2024c) investigated the effects on infection risk from different durations of indoor gathering for different group sizes during the following lunch time scenarios: (i) a fast-food restaurant, (ii) a workplace cafeteria and (iii) a moderately-priced standard restaurant. Edwards et al. (2024c) leveraged observational data from Bell and Pliner (2003) and considered Erlang distributions for the typical duration of a meal in each venue,

$$\begin{aligned} T_{Fast} &\sim \text{Erlang}(14, 39.06) \text{ hours}, \\ T_{Cafeteria} &\sim \text{Erlang}(10, 13.09) \text{ hours}, \\ T_{Restaurant} &\sim \text{Erlang}(20, 20.45) \text{ hours}. \end{aligned}$$

Each of these distributions along with their mean values are reported in Figure 6a. Since Erlang distributions correspond to Gamma distributions with integer shape parameter, and to make our results comparable to those by Edwards et al. (2024c), we consider the same distributions for our results in Section 3.3. We note that meal duration in the fast-food restaurant is typically shorter, with less uncertainty, whereas durations increase in workplace cafeterias and restaurants, leading also to higher



**FIGURE 6** Probability distributions of the duration of the meal ( $T$ ) and the ventilation rate  $Q$  in the hospitality scenario in Section 3.3 The vertical dashed lines represent the mean values of the distributions.

variability. We also note that the duration of these lunch times are longer and more varied than that of the visit durations in the healthcare setting in Figure 4.

Finally, and in order to illustrate in this scenario how uncertainty in two parameter values might simultaneously impact infection risk, we also consider in Section 3.3.1 two possible random ventilation rates corresponding to low variance and high variance around an average of 8 ACH (Repace 2000, Casals Ventilation n.d., Deutsches Institut für Normung 2019). In particular, and for purely illustrative purposes, we consider (see Figure 6b)

$$\begin{aligned} ACH_{highVar} &\sim \text{Gamma}(2, 0.25), \\ ACH_{lowVar} &\sim \text{Gamma}(30, 3.75), \end{aligned}$$

for our results in Figure 11. Importantly, both distributions have the same mean 8 ACH, but different variances ( $\text{Var}(ACH_{highVar}) \approx 32$  versus  $\text{Var}(ACH_{lowVar}) \approx 2$ ). Moreover, the ventilation regime with higher variance is more skewed towards smaller air change rates than the other ( $\text{Skew}(ACH_{highVar}) \approx 1.41$  versus  $\text{Skew}(ACH_{lowVar}) \approx 0.37$ ). A room volume of  $300 \text{ m}^3$  is used in Section 3.3 as in Edwards et al. (2024c), leading to ventilation rates

$$\begin{aligned} Q_{highVar} &\sim \text{Gamma}(2, 0.25/300), \\ Q_{lowVar} &\sim \text{Gamma}(30, 3.75/300). \end{aligned}$$

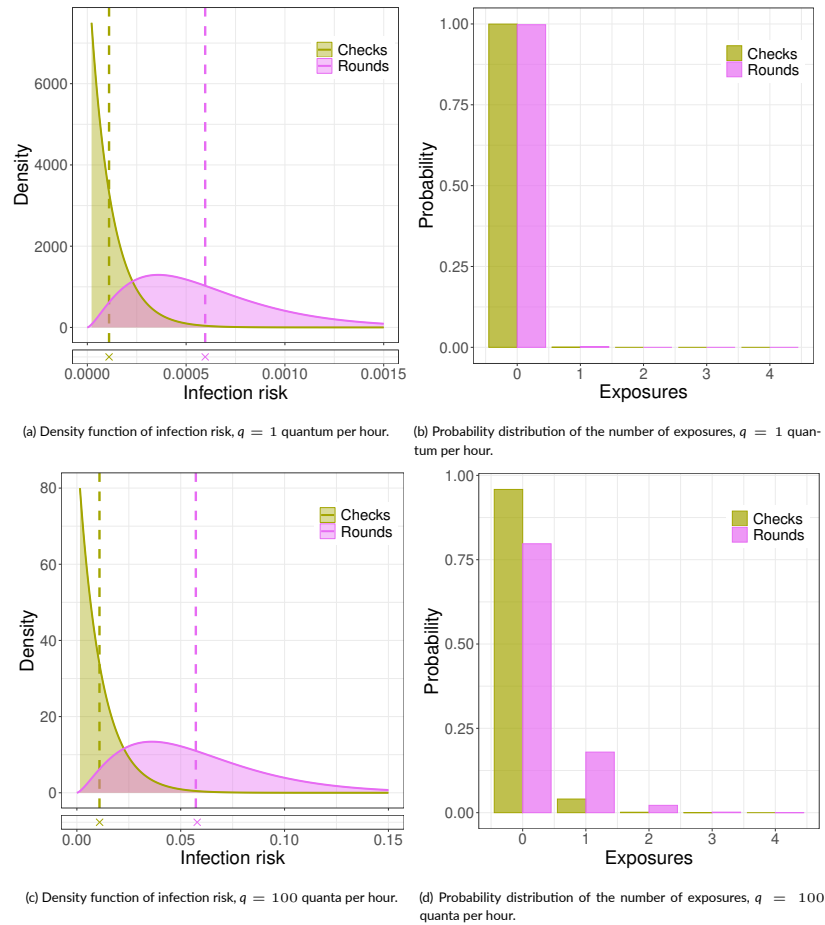
### 3.2 | Infection risk during healthcare

In this Section, we analyse infection risk for  $S = 4$  patients in a hospital room being visited by an infectious healthcare worker while accounting for uncertainty in visit times, infectiousness and ventilation rate, and by considering parametric regimes discussed in Section 3.1.1

#### 3.2.1 | Uncertainty in visit times

The Gamma distributions for  $T_{Checks}$  and  $T_{Rounds}$  (see Section 3.1.1) provide a reasonably good representation of the empirical observations by King et al. (2021) (Figure 4), and allow us to incorporate uncertainty in visit times while calculating infection risk via the extended Wells-Riley methodology proposed in Section 2. In particular, using these random visit times, we can use the analogous result to that in Equation (2) (just by swapping  $q$  with  $T$ , as well as their respective shape and rate parameters) to estimate the density function of the per-capita infection risk for each patient in the room. These density functions  $f_{P_{infection}}(p)$  of the per-capita infection risk probability, for Checks and Rounds, are reported in Figures 7a and 7c, for  $q = 1$  and 100 quanta per hour respectively, along with their expected values  $\mathbb{E}[P_{infection}]$  indicated by the vertical dashed lines.

First, we note that overall observational care (i.e., Checks) leads to lower infection risk compared to doctors' rounds (Rounds), which is to be expected given the shorter duration during Checks. Less visit duration uncertainty during Checks (i.e., a narrower Gamma distribution compared to that of Rounds) leads to less uncertainty in the corresponding infection risk probability (i.e., narrower density function  $f_{P_{infection}}(p)$  for Checks compared to Rounds). On the other hand, more uncertainty in the duration of doctors' rounds leads to higher uncertainty in the infection risk for susceptible patients in the room. This leads to a density function of infection risk around the mean values  $\mathbb{E}[P_{infection}] \approx 0.0006$  (for  $q = 1$  quantum per hour) and  $\mathbb{E}[P_{infection}] \approx 0.057$  (for  $q = 100$  quanta per hour) but showing larger variability for Rounds compared to Checks. For example, in the case of  $q = 100 \text{ quanta h}^{-1}$ , infection risk might be significantly higher (up to 0.15) for longer Rounds episodes or smaller (down to almost 0) for shorter ones, with a mode infection risk

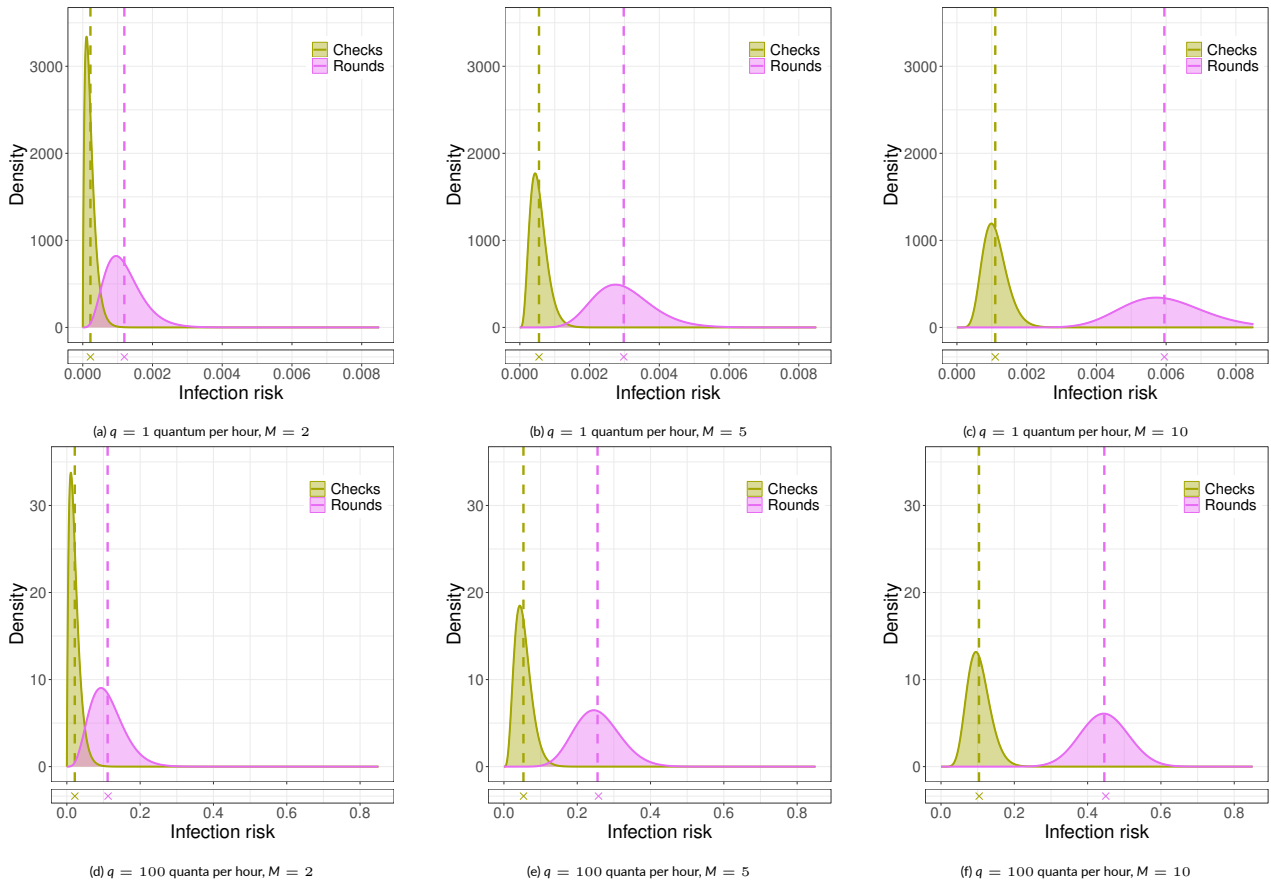


**FIGURE 7** (a),(c): Shaded density plots showing the distribution of the risk of infection with random visit times during each type of care (Equation (2) adapted to random  $T$ ), for quanta emission rates of  $q = 1$  and 100 quanta  $h^{-1}$ . (b),(d): Probability distributions of the number of exposures in the hospital ward scenario during each type of care calculated using Equation (4) adapted to random  $T$  (as seen in Table 1), for quanta emission rates of  $q = 1$  and 100 quanta  $h^{-1}$ .

probability of  $\text{Mode}(P_{infection}) \approx 0.04$ . We note that the shape of the per-capita infection risk density functions in Figure 7 is preserved across  $q = 1$  and 100 quanta per hour for each type of care. In fact, for a particular type of care, these densities scale linearly (by around  $10^2$ ) between  $q = 1$  quanta  $h^{-1}$  and  $q = 100$  quanta  $h^{-1}$ , since for small enough infection risk the Wells-Riley model equation becomes approximately linear on  $q$ . However, this is not the case for the  $n$ -exposure distributions plotted in Figures 7b and 7d, for  $n = 0, 1, 2, 3, 4$  for a room with  $S = 4$  susceptible patients. For  $q = 1$  quantum per hour, the per-capita infection risk is small enough so that we see no notable differences in the distribution of exposures between Rounds and Checks. On the other hand, for  $q = 100$  quanta per hour longer visit times for Rounds lead to a rightward-shift in the corresponding distribution, where the probability of at least 1 exposure is non-negligible.

It is interesting to compare our results, which allow us to better quantify uncertainty in infection risk, with the classical Wells-Riley approach where the visit time is considered to be known and fixed. In particular, we can compute the per-capita infection risk probability  $\bar{P}_{infection}^{WR} = 1 - e^{-\frac{bq}{Q}\bar{T}}$  while setting the visit time as the mean visit time for each type

of care in Figure 4; that is by setting  $\bar{T} = \mathbb{E}[T_{Checks}] \approx 1$  min for Checks and  $\bar{T} = \mathbb{E}[T_{Rounds}] \approx 5.36$  min for Rounds. This infection risk probability computed via the classical Wells-Riley approach is reported as a colored cross at the bottom of the density plots in Figure 7. In these scenarios, the classical Wells-Riley approach (using the mean visit duration) gives a very similar infection risk probability  $\bar{P}_{infection}^{WR}$  to the expected value  $\mathbb{E}[P_{infection}]$  (dashed vertical lines) calculated via our methodology (which summarises the uncertainty around  $T$  in an average infection risk value). For  $q = 1$  quanta  $h^{-1}$  (Figure 7a), we have  $bq/Q = 2/300$ , for which we know from Figure 1 (such that  $\bar{q}$  is replaced by  $\bar{T}$  and  $bT/Q$  is replaced by  $bq/Q$ ) overestimations do not occur in the region of  $\bar{T}_{Rounds} \approx 10^{-1}$ ,  $\bar{T}_{Checks} \approx 10^{-2}$  and  $bq/Q \approx 10^{-2}$ . However, there are other situations where using the classical Wells-Riley approach can lead to a more significant overestimation of infection risk as described in Subsection 2.1.1. By looking in Appendix A5.1 at Figure A1, we can see that for  $bq/Q$  being on the order of  $10^{-2}$ , we would need  $\bar{T} = 30$  hours before any noticeable relative difference can be seen. Similarly, for  $q = 100$  quanta  $h^{-1}$  we have  $bq/Q = 2/3$ , for which we would require  $\bar{T} = 1$  hour before overestimations start to occur.



**FIGURE 8** Density plots of the per-capita infection risk probability (Equation (A15)),  $f_{P_{infection}}(p)$ , for a susceptible patient being visited sequentially  $M \in \{2, 5, 10\}$  times by a HCW, for either *Checks* (Yellow) or *Rounds* (Pink). Vertical dashed lines represent the mean infection risk  $\mathbb{E}[P_{infection}]$  (Equation (A16)). The crosses represent the classical Wells-Riley approach average infection risk  $\bar{P}_{infection}^{WR} = 1 - e^{-\frac{pq}{\alpha} \bar{T}}$ , with the mean visit time  $\bar{T}$  as a constant value.

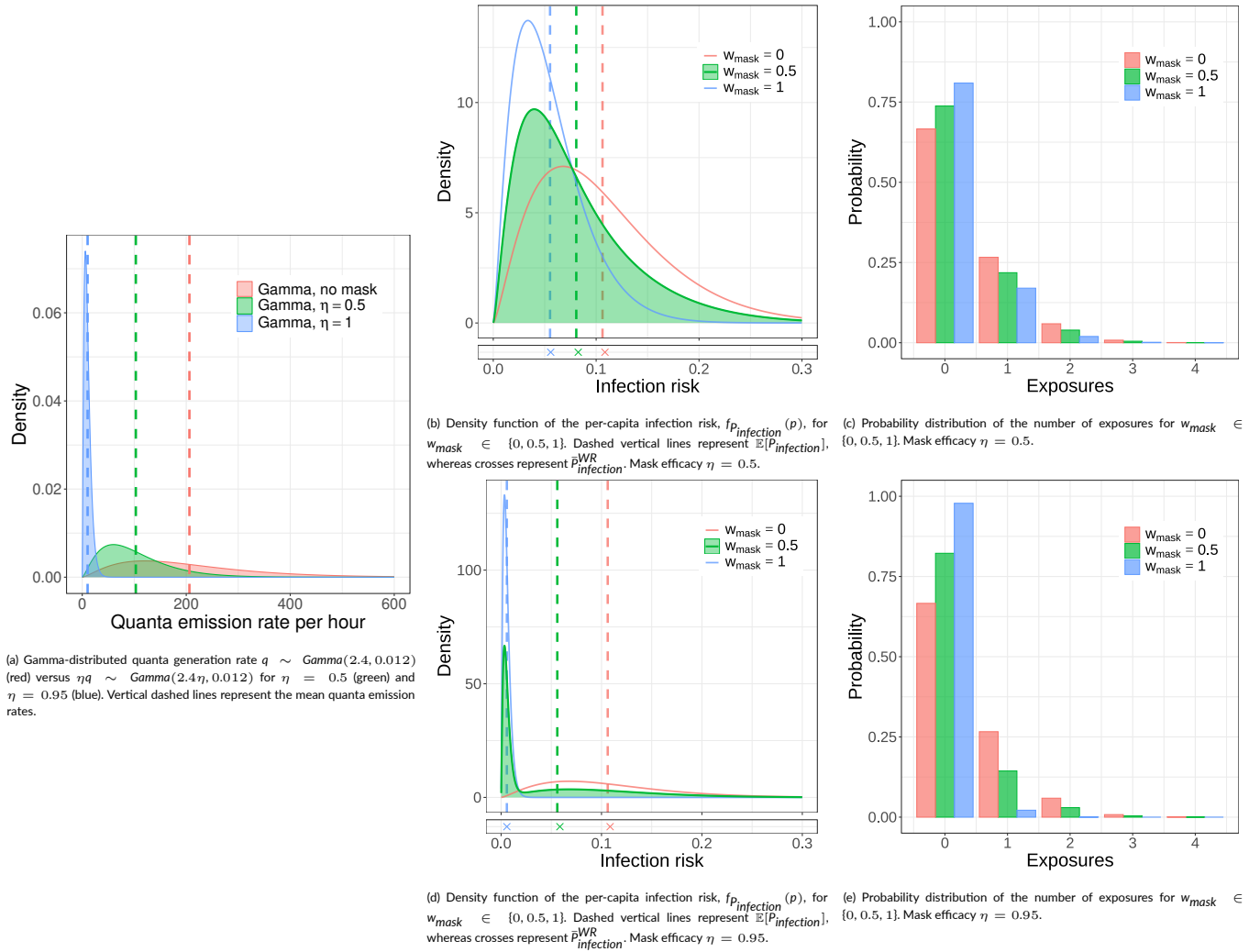
A particular feature of our approach (considering Gamma-distributed parameters) is that one can, for example, consider the impact that sequential visits have on infection risk. In particular, since the sum of independent Gamma distributions is a Gamma distribution, one can consider sequential visits from HCWs to a given room. For example, for  $M > 1$  sequential Checks (e.g., during a day), assuming that each observational care visit to the room is carried out independently with duration  $T_{Checks} \sim \text{Gamma}(0.99, 60)$ , the total time spent by the HCW in the room is  $T_{Total\_Checks} \sim \text{Gamma}(0.99M, 60)$ . Thus, due to the steady-state assumption in the Wells-Riley framework, the overall infection risk from  $M$  sequential visits (each of duration  $T_{Checks}$ ) is equal to the infection risk from a single visit with duration  $T_{Total\_Checks}$ , which can be analysed via our methodology in Section 2.1, since it is Gamma-distributed; see Appendix A4 for further details.

In Figure 8, we plot the density function of the infection risk probability for a susceptible patient from  $M \in \{2, 5, 10\}$  sequential HCW visits (either Checks or Rounds), for rates of  $q = 1$  and  $q = 100$  quanta per hour. As expected, an increasing number of visits from an infector to the room significantly increases infection risk in both care-type scenarios. It

is however interesting to see how an increasing number of visits compounds uncertainty in each visit duration (leading to higher uncertainty in infection risk), where even a narrow Gamma distribution for  $T_{Checks}$  leading to low risk uncertainty in Figures 7a and 7c now leads to notably higher uncertainty in infection risk for  $M = 2, 5$  and 10 sequential Checks in Figure 8. This highlights the importance of considering uncertainty in parameter values when estimating infection risk in these types of situations. As the number of sequential visits increases, the increase in infection probability between the cases  $q = 1$  and  $q = 100$  quanta per hour becomes non-linear; we still observe uncertainty around infection risk but the relative magnitude of this uncertainty is no longer the same between  $q = 1$  quanta  $h^{-1}$  and  $q = 100$  quanta  $h^{-1}$ ; this is particularly true for Rounds in Figures 8c and 8f.

### 3.2.2 | Uncertainty in infectiousness

Here, we analyse the impact of uncertainty in quanta generation rate by considering  $q \sim \text{Gamma}(2.4, 0.012)$  quanta per hour as in Figure 3, while illustrating results in Section 2.1.1 to incorporate the possibility of the infector wearing a mask with probability  $w_{mask}$ , which reduces the



**FIGURE 9** Gamma-distributed quanta generation rates depending on mask efficacy parameter  $\eta$ . The dashed lines represent the mean quanta emission rates. (b) and (d) Per-capita infection risk density curves (Equation (5)) for mask efficacies  $\eta = 0.5$  and  $0.95$  and probabilities of the infector wearing the mask  $w_{mask} \in \{0, 0.5, 1\}$ . Vertical dashed lines represent mean infection risk,  $\mathbb{E}[p_{infection}]$  (Equation (6)), and the crosses represent the classical Wells-Riley estimate  $\bar{p}_{infection}^{WR} = 1 - \exp(-\frac{IbT}{Q}\bar{q}_{eff})$ . (c) and (e) Probability distribution of the number of exposures (Equation (7)) for  $\eta = 0.5$  and  $0.95$  and probabilities of the infector wearing the mask  $w_{mask} \in \{0, 0.5, 1\}$ .

corresponding generation rate by a factor of  $\eta$ . As a result, randomness in infectiousness may be due to uncertainty in the quanta generation rate estimate (Figure 3) or due to human behaviour. In Figure 9, we explore two different values of mask efficacies of 50% ( $\eta = 0.5$ ) and 95% ( $\eta = 0.95$ ), to explore a wide range of values which can be affected by mask type and quality of fit (Davies et al. 2013, Wilson 2020, Bazant and Bush 2021, Rothamer et al. 2021, Huang et al. 2022). The impact of  $\eta$  on the corresponding Gamma-distributed quanta generation rate is plotted in Figure 9a. For these two values of  $\eta$ , we plot the infection risk densities in Figures 9b and 9d for probabilities  $w_{mask} \in \{0, 0.5, 1\}$  of the infector wearing the mask, as well as the probability distribution of the number of exposures in Figures 9c and 9e. We consider an interaction of  $T = 5$  minutes in all these scenarios. When focusing on the density curve of the per-capita infection risk in Figure 9b, one can note that  $w_{mask} = 1$  (infector wears mask, blue curve) generally leads to lower

risk compared to  $w_{mask} = 0$  (infector does not wear mask, red curve), as one would expect. Still, high uncertainty in these generation rates (variance of the corresponding Gamma distributions in Figure 9a) means that these red and blue density functions overlap. Thus, for  $w_{mask} = 1$ , one could still get relatively high values of  $p_{infection}$  (i.e., per-capita infection risk above 0.15 corresponding to the right tail of the blue density curve) while for  $w_{mask} = 0$  one could still get low infection risk (i.e., per-capita infection risk close to 0 in the left part of the red density curve), with non-negligible probabilities. Setting  $w_{mask} = 0.5$  (the infector has an equal chance of wearing or not wearing the mask) means that the density curve of the probability of infection shares characteristics with both of the underlying densities (for  $w_{mask} = 0$  and 1), which is particularly interesting in the case  $\eta = 0.95$  (Figure 9d). In this case, the high mask efficacy  $\eta$  reduces the effective quanta emission by enough (blue vs red in Figure 9a) that the underlying distributions have little overlap.

This highlights how stochasticity in human behaviour ( $w_{mask}$ ) can have a high impact on infection risk in this type of situation. In particular, the large value of  $\eta = 0.95$  leads to bi-modality in the resulting (green) density curve for the per-capita infection risk in Figure 9d, representing two significantly different per-capita infection risk regions depending on the individual wearing or not wearing the mask. In this situation, it is clear that the mean infection risk (dashed vertical green line in Figure 9d) or the single estimate (green cross) provided by the classical Wells-Riley approach with average quanta generation rate  $q_{eff}$  are not good representations of the true infection risk profile (given by the whole green density curve in Figure 9d). In these type of high-stochasticity high-impact scenarios, our results highlight the importance of considering this uncertainty when predicting risk, rather than relying on average/fixed parameter values or just focusing on average per-capita infection risk estimates.

Finally, the  $n$ -exposure distributions have been plotted in Figures 9c and 9e. The reduction of approximately 2% (8.3% to 5.9%) in the mean per-capita infection risk (for  $w_{mask} = 0.5$ ) between  $\eta = 0.5$  to  $\eta = 0.95$  leads to a slightly leftwards shift in the  $n$ -exposure distribution. The probability of at least one exposure occurring increases notably for smaller  $w_{mask}$  and  $\eta$ . For a larger value of  $\eta$  (0.95 compared to 0.5), the impact of mask-wearing becomes larger; this means that the  $n$ -exposure distributions are more different for different values of  $w_{mask}$  when  $\eta = 0.95$  compared to 0.5. That is, for decreasing efficacy values  $\eta \rightarrow 0$  (representing decreasing fit or mask quality), the differences between different values of  $w_{mask}$  would become less significant.

### 3.2.3 | Uncertainty in ventilation rate

In this section, we investigate the effects of random ventilation by considering the distributions in Figure 5. In particular, we plot in Figure 10 the per-capita infection risk density curves under only natural ventilation ( $Q_{mech} = 0$ ) or the presence of some mechanical ventilation ( $ACH_{mech} = +1$ ), while considering two visit durations ( $T = 5$  vs  $T = 15$  mins) and  $l \in \{1, 2\}$ . We consider the quanta generation rate  $q = 100$  quanta per hour here to represent high infectivity; a shift on infection risk similar to that observed in Figure 7 would be expected if one considered  $q = 1$  quanta  $h^{-1}$  instead.

As reported by Edwards et al. (2024a), reliance on natural ventilation alone can lead to high-risk situations during moments (or days) where the particular environmental conditions lead to significantly low ventilation rates in a given room, represented by values of  $Q_{nat}$  close to 0 in the corresponding Gamma distribution in Figure 5. Indeed, these values lead to long “high-risk” tails of the per-capita infection risk density curves, close to 40% in Figure 10a and close to 100% in Figure 10c, especially if  $l = 2$  or  $T = 15$  minutes. It is also interesting to note how the same level of uncertainty in  $Q$  (in terms of the same variance for the Gamma distributions in Figure 5) between the different scenarios considered (natural ventilation only in Figures 10a and 10c, vs additional mechanical ventilation in Figures 10b and 10d) can lead to substantially

different levels of uncertainty in the per-capita infection risk. In particular, this level of uncertainty in  $Q$  has a more significant impact when only natural ventilation is considered (which leads to some days, as described, with very low ventilation rates near 0). Once some mechanical ventilation ( $ACH_{mech} = +1$ ) is added, so that the distribution of  $Q$  is shifted in Figure 5, uncertainty in the per-capita probability of infection significantly decreases, consistent with findings in Edwards et al. (2024a). This highlights the importance that ensuring some ventilation in the room, even if not significant, may have on infection risk when the chances of having very low ventilation rates near 0 are non-negligible in that particular environment.

When considering uncertainty in  $Q$ , the classical Wells-Riley approach (which consists of computing  $\bar{P}_{infection}^{WR} = 1 - e^{-\frac{lqT}{\bar{Q}}}$ , with  $\bar{Q} = \bar{Q}_{nat} + Q_{mech}$  being the average total ventilation rate) typically leads to an underestimation of risk in Figure 10, consistent with our results in Figure 2. This is due to the fact that using the average ventilation rate partly neglects high-risk situations where, for example, reliance on natural ventilation leads to situations with very low ventilation rates on particular days, as discussed above; see also Appendix A5.2 for further details. Overall, our approach, where ventilation rates and infection risk are represented as random variables, allows one to capture the probability (and impact) of these events when estimating infection risk in the presence of uncertainty.

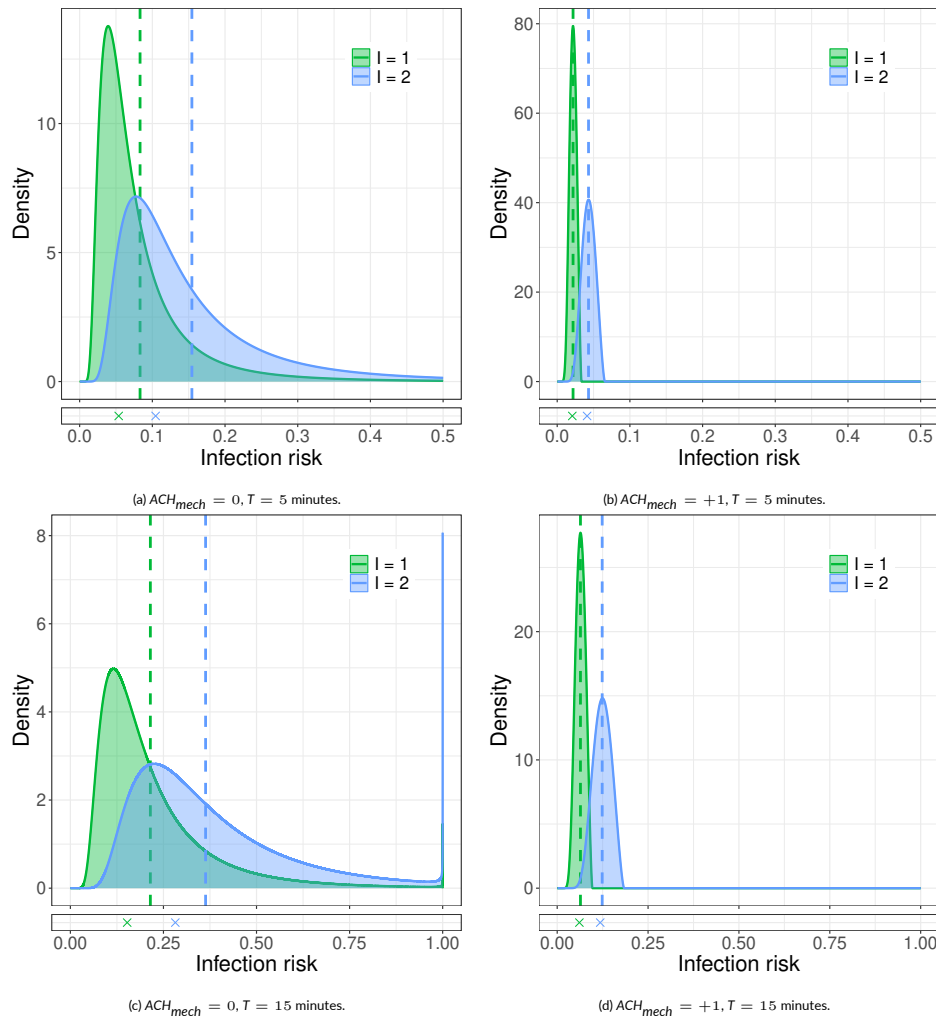
## 3.3 | Infection risk during meals

Here, we consider a similar scenario to the one considered by Edwards et al. (2024c) to estimate infection risk during meals in different hospitality venues, and while accounting for uncertainty in two parameter values simultaneously, or in disease prevalence.

### 3.3.1 | Uncertainty in meal duration and ventilation rate

We investigate here the impact of considering simultaneous uncertainty in  $T$  and  $Q$  in this scenario, while leveraging the Gamma distributions described in Section 3.1.2. Once again, we set  $q = 100$  quanta per hour here for illustrative purposes, corresponding to a highly infectious individual in the venue. We use Equations (22) and (23) (with  $q$  substituted by  $T$ ) to calculate the density curves and expected values of  $P_{infection}$  given our different scenarios and ventilation rates. These are all plotted in Figure 11. Within each subplot, we consider the situation where only one parameter is considered to be random (either  $T$  or  $Q$ ), while the other is set to the mean value of its corresponding Gamma distribution, or where both parameters are simultaneously random.

Overall, infection risk is lower in fast-food venues compared to workplace cafeterias or restaurants, due to shorter lunch durations. More variability in lunch duration in these latter venues also leads to higher variability in infection risk. Still, stochasticity in ventilation rate “dominates” (i.e. against randomness of  $T$ ) the stochasticity in infection risk

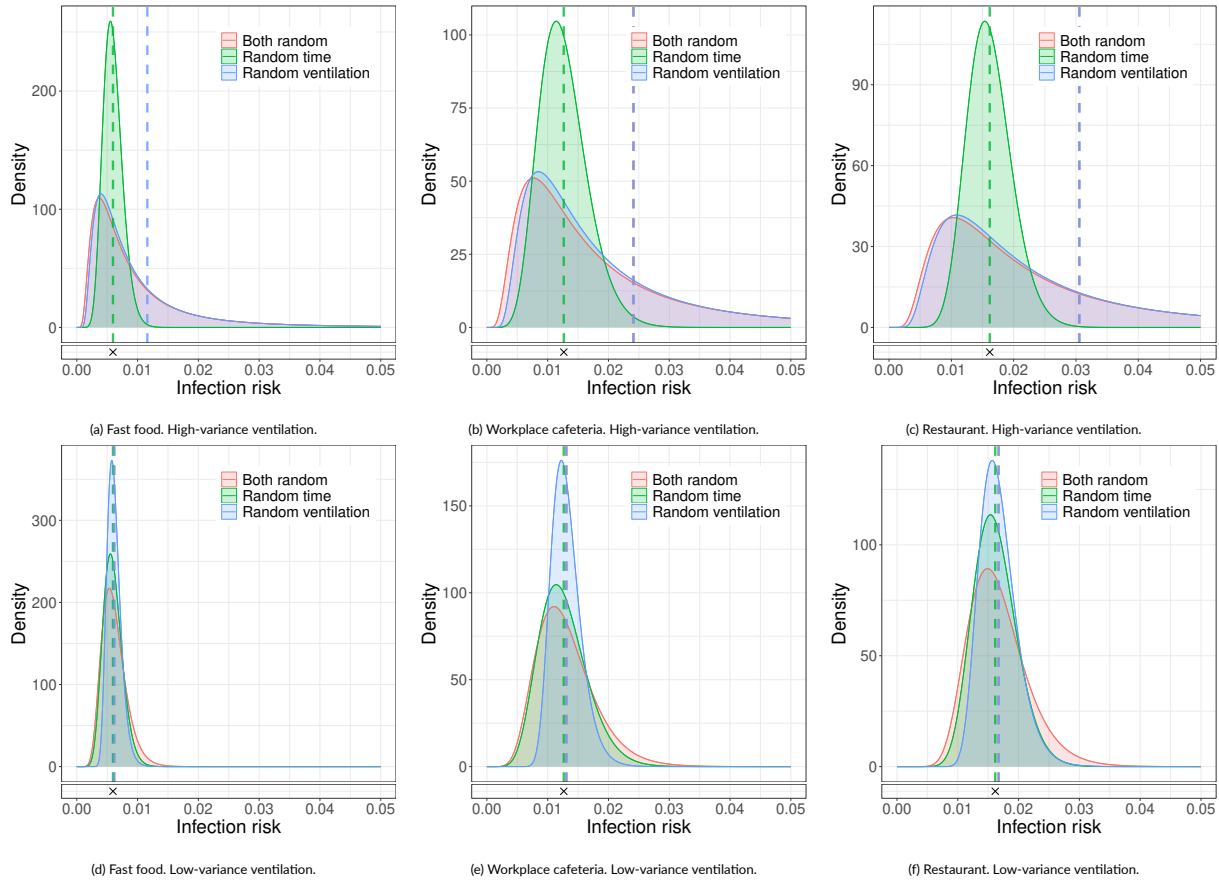


**FIGURE 10** Probability density curves of the per-capita infection risk, for natural ventilation only ( $ACH_{mech} = 0$ ) and added mechanical ventilation ( $ACH_{mech} = 1$ ), for exposure times  $T = 5$  and 15 minutes, a quanta emission rate of  $q = 100$  quanta per hour, and for  $l = 1$  or  $l = 2$  infectors. Dashed vertical lines represent the distribution mean,  $\mathbb{E}[P_{infection}]$ , and crosses represent the average value provided by classical Wells-Riley approach with average ventilation rate parameter,  $\bar{P}_{infection}^{WR} = 1 - \exp(-lbqT/(\bar{Q} + Q_{mech}))$ .

when the ventilation rate distribution has higher variance and a higher positive skewness (Figures 11a-11c), which can be noticed by the red and blue density curves almost overlapping in these scenarios. Once again, the distribution of the ventilation rate in the high-variance scenario in Figure 6b leads to low ventilation rates with high probability, even though the mean ventilation rate is still 8 ACH. This directly contributes to the long right tail of the density function of  $P_{infection}$  in Figure 11a-11c (red and blue curves), which is missing when only randomness in  $T$  is considered (green curve). We refer the reader to Appendix A6 where we investigate further the reasons for one parameter to dominate the shape of the density function of  $P_{infection}$  over the other, when both parameters are random. On the other hand, the low-variance ventilation rate scenario in Figure 6b leads to a very low probability of observing a ventilation rate below 5 ACH at any given time. In this case, randomness in  $T$  becomes more relevant when estimating the density function

of  $P_{infection}$ . This can be seen in Figures 11d-11f, where the red density curve (both parameters considered to be random), does not closely mimic any of the other two densities (blue and green). Interestingly, the red density curve always has larger variance compared to the green and blue ones. This highlights the compounding effect that having uncertainty in two parameter values simultaneously has on the uncertainty in the per-capita infection risk. That is, for example, if both  $T$  and  $Q$  are random, there will be scenarios where a particularly long lunch duration combined with a particularly low ventilation rate (sampled from the corresponding parameter distributions) could lead to increased infection risk, leading to the relatively longer right tail of the red density curve in Figures 11d-11f.

We can also compare our predictions with the per-capita infection risk estimated via the classical Wells-Riley approach when both parameters are set to be constant equal to their mean values, represented by the crosses in Figure 11. The most striking difference is that observed



**FIGURE 11** Density curves of the per-capita infection risk probability  $P_{infection}$  for different hospitality venues, given random  $T$  and  $Q$  (Equation (22)). For each scenario, we compare these density curves against the ones that one would obtain if one parameter is random and the other set to its mean value instead (Equation (2) and Equation (11)). Vertical dashed lines indicate density means (Equations (23), (3) and (12)). Crosses below the subplots indicate the classical Wells-Riley estimate  $\bar{P}_{infection}^{WR}$  which uses average parameter values instead. Three hospitality venues with either high or low variance ventilation rates from Figure 6b.

between these values (crosses, representing the classical Wells-Riley approach  $\bar{P}_{infection}^{WR}$ ) and the dashed vertical red and blue lines (which are on top of each other, representing analytical mean infection risks  $\mathbb{E}[P_{infection}]$ ) in Figures 11a-11c, where we observe underestimations by the classical approach of approximately 50% relative to the analytical means. This highlights how the classical (average) Wells-Riley approach would clearly underestimate infection risk if uncertainty in  $Q$  is not properly accounted for. On the other hand, the fact that the red and blue lines overlap, whereas the green vertical dashed line is significantly below and closer to the cross, means that in this type of scenario considering uncertainty in  $Q$  is enough to capture most of the real uncertainty in  $P_{infection}$ , whereas the uncertainty in  $T$  becomes less important here and could be ignored.

### 3.3.2 | Uncertainty in number of infectors: considering disease prevalence

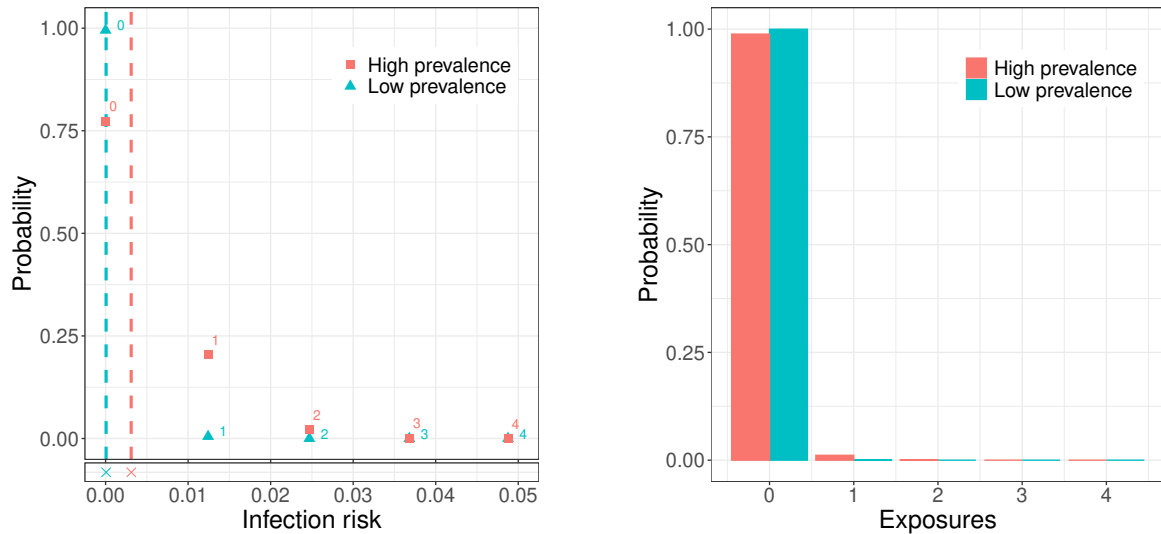
It is typical when assessing infection risk to assume that an infector is present. However, the probability of an infector being present is one

of the main factors influencing infection risk for susceptible individuals in reality. This can be incorporated, as discussed in Section 2.1.2, via a probability distribution for  $I$ . If we consider the population-level probability of being infectious to be  $\rho$ , and assuming independence, one can consider  $I \sim \text{Bin}(N, \rho)$  where  $N$  is the number of individuals involved in the indoor interaction. Looking at the trends for the percentage of positive COVID-19 tests during 2022 (Office for National Statistics 2023), we can see that they typically ranged between 1% and 8%. Thus, for illustrative purposes, we consider two potential values of  $\rho \in \{0.001, 0.05\}$  to represent low and high prevalence scenarios, respectively. We define two variables for the number of infectors

$$I_{low} \sim \text{Bin}(5, 0.001),$$

$$I_{high} \sim \text{Bin}(5, 0.05).$$

As an example, we will consider the workplace cafeteria scenario, fixing  $\bar{T} = 10/13.09 \approx 45 \text{ min}$  to be the corresponding mean lunch duration and a ventilation of 8 ACH. Equations (8) to (10) are used to calculate the probability distribution of the per-capita infection risk as well as



(a) Probability mass function of the per-capita infection risk (Equation (8)). The vertical dashed lines in Figure 12a represent the "true" mean of the per-capita infection risk (Equation (9)) and the crosses represent the infection risk estimated by the classical Wells-Riley approach,  $\bar{p}_{WR}^{infection}$ , which uses the mean number of infectors,  $I = \rho N$ , as a constant. Each specific mass point arises from each possible number of infectors being present, indicated by the number next to it.

(b) Probability distribution for the number of exposures (Equation (10)) given that the number of infectors is unknown,  $I \sim \text{Bin}(N, \rho)$ , for  $\rho = 0.1\%$  (blue) and  $\rho = 5\%$  (red) disease prevalence, and for a total number  $N = 5$  of individuals.

**FIGURE 12** Infection risk in the workplace cafeteria for  $\bar{T} = 45 \text{ min}$  and 8 ACH, with a random number of infectors  $I \sim \text{Bin}(N, \rho)$ . Disease prevalence  $\rho = 0.1\%$  (low, blue) and  $\rho = 5\%$  (high, red), and  $N = 5$  individuals in total.

the probability of observing  $n$  exposures. The per-capita infection risk distributions and their means have been plotted in Figure 12a, along with the classical Wells-Riley probability of infection calculated by fixing  $\bar{T} = N\rho$  as the mean number of infectors. We note that since  $I$  is a discrete random variable, the per-capita infection risk becomes now a discrete random variable as well, which is why we plot a probability mass function in Figure 12a instead of a density curve.

When we have a prevalence of  $\rho = 0.1\%$ , it is highly unlikely for anyone present at the lunch to be infectious, leading to the per-capita infection probability being 0 with a likelihood close to 100%. On the other hand, when the prevalence is increased to  $\rho = 5\%$ , the probability of at least one infectious individual being present becomes non-negligible ( $\mathbb{P}(I \geq 0) \approx 24\%$ ), which explains why the red probability mass function in Figure 12a attributes non-negligible likelihood to positive per-capita infection risks. Still, the magnitude of the per-capita infection risk remains low in all cases in Figure 12a (less than 5%), which is explained by the high ventilation rate considered. For both prevalence rates, the Wells-Riley approach seems to capture the density mean well, possibly because risks are low in this scenario.

In Figure 12b, we have plotted the probability distribution of the number of exposures within the group. The high probability of observing 0 exposures, even under high prevalence, highlights the importance of taking into account the probability of an infector being present when carrying out infection risk assessment (rather than assuming that an infector is in the room). This is directly related to the idea of implementing mitigations which may avoid an infector being present in the first place (e.g., avoiding contact if symptomatic), above other additional mitigations which might be considered.

## 4 | DISCUSSION

In this study, we revisited the classical Wells-Riley model by introducing a stochastic framework that analytically accounts for uncertainty in key parameters such as the quanta emission rate, ventilation rate, exposure time, and number of infectors. Unlike existing approaches in the literature which are typically based on numerical or Monte Carlo simulation methods, we propose a probabilistic, mathematical framework which extends the work by Nicas (1996) and Edwards et al. (2024c). In particular, by treating these parameters as random variables, we derived probability distributions for both the per-capita infection risk and the number of exposures during indoor interactions. When compared to a classical Wells-Riley model with average parameter values, our probabilistic approach allows for a more nuanced and realistic assessment of airborne infection risk, particularly in heterogeneous populations and variable environmental conditions. It is worth noting that there exist dose-response models which use an exponential-Poisson relationship to infection probability similar to that of the Wells-Riley model, and authors have previously incorporated uncertainty into the dose (Sze To and Chao (2010), Nicas (1996)). In particular, Nicas (1996) investigates Beta-distributed and Gamma-distributed doses, and derives infection probability distributions. Here, we instead consider uncertainty in each of the parameters which can impact the dose within the Wells-Riley framework.

Our findings demonstrate that the classical Wells-Riley model under average parameter values can lead to significant inaccuracies when estimating infection risk. A particularly interesting result is that generated in Section 2.1.1 which leverages Jensen's inequality: the per-capita

infection risk will be overestimated if one implements the classical Wells-Riley methodology under average parameter values for  $(p, q, T)$ , compared to computing the “true” average per-capita infection risk which incorporates the whole probability distribution of the corresponding parameter. On the other hand, considering an average ventilation rate typically leads to underestimating the true per-capita infection risk. This is particularly significant in scenarios where very poor ventilation can occur with positive probability; in these situations, considering an average ventilation rate neglects the possibility of these high-risk events happening. These insights underscore the importance of incorporating parameter uncertainty into infection risk assessments, especially in settings where precise measurements are unavailable or impractical. It also highlights the importance of incorporating low-probability high-impact events (e.g. scenarios with very low ventilation, the presence of a highly infectious individual or the possibility of this infector wearing a mask), which can be done via the analytical framework here proposed.

Overall, large uncertainty in parameter values leads to large uncertainty in the per-capita infection risk estimates. The nature of this resulting uncertainty can significantly change across scenarios though. For example, we illustrated in Figure 9d how, in some scenarios, bi-modality might arise where particular stochastic conditions or random human behaviour might lead to very different infection risk profiles depending on a particular random event (in this figure, the infector wearing or not a mask). In these situations, it is clear how considering either average parameter values or average infection risk estimates would not be representative or informative of the real infection risk profile outlined by the green density in Figure 9e. This has direct implications in terms of risk mitigation design, where for example here one would focus on decreasing risk in those scenarios corresponding to the second larger mode in Figure 9e (e.g. by increasing  $w_{mask}$ ) rather than trying to shift the whole probability density curve to the left via alternative mitigations.

Our results also highlight how, when considering two parameters being simultaneously random, uncertainty in one of these parameters can in some scenarios dominate the uncertainty in the corresponding per-capita infection risk, as explored in Figure 11. This has direct implications in terms of both incorporating this uncertainty into the risk estimates and considerations around data collection. That is, if uncertainty around the ventilation rate is expected to have the largest impact on the infection risk estimates, one would need to incorporate this uncertainty into the corresponding mathematical framework (as described in Section 2), as well rely on better measurements for this particular parameter. In these instances, variability in other parameters may be neglected if this variability is expected to be smaller or have a smaller impact on infection risk. Our methodology allows one to explore this in detail when carrying out infection risk assessment in different settings.

Proposing an analytical framework as the one in Section 2 allows one to carry out extensive sensitivity analysis under different parametric regimes, as well as to leverage analytical properties as we did with Jensen’s inequality in Section 2.1.1. Moreover, this analytical, probabilistic framework has the potential to be extended to additional scenarios

(e.g. alternative parametric distributions) or when relaxing particular assumptions (e.g. steady-state pathogen concentration in the air). While some of the closed formulae obtained in Section 2.1, such as Equations 2-11 and 15, can be readily applied, other results in this work depend on integral functions such as the modified Bessel or the Tricomi function. Still, these are well-known functions in the literature for which analytical properties and numerical implementations are readily available, and we leveraged these in some of our results in Section 3. Thus, our analytical results could be easily embedded into existing QMRA frameworks, allowing one to quickly test uncertainty in various scenarios without a major need in computing power. These analytical results could also be used directly in Bayesian frameworks when estimating model parameters from observations, since some of our results would arise when computing likelihood functions in these scenarios.

Despite the strengths described above, several limitations should be acknowledged. First, our approach is derived from the classical Wells-Riley model, which uses the steady-state assumption for the concentration of infectious pathogen in the air. For short exposure interactions combined with low ventilation, the steady-state assumption can overestimate infection risk, as shown in Edwards et al. (2023). This limitation has been addressed by Gammaitoni and Nucci (1997), who extended the Wells-Riley model to include transient concentration dynamics (exponential growth or decay) (see also Edwards et al. (2023), López-García et al. (2019)) Our approach could be strengthened by incorporating parameter uncertainty into that transient model.

Secondly, we have considered at most two (independent) parameters being simultaneously random when estimating infection risk. In reality, uncertainty might be present in more than two parameters, or some parameters may exhibit correlations (for example, between infectiousness and behaviour, or between ventilation patterns and occupancy). Although the methodology could in principle be extended to more than two random parameters, or pairs of correlated parameters (e.g. via joint distributions), the resulting analytical expressions may soon become intractable. In these situations, numerical or Monte Carlo simulation approaches as those existing in the literature (Bate et al. 2024, Henriques et al. 2025, Iddon et al. 2022, Jones et al. 2021a 2025, Edwards et al. 2026) would be preferred.

Thirdly, we have considered Gamma-distributed parameters in Section 2 due to their mathematical tractability and versatility, going beyond Exponential or Erlang-distributed ones as considered by Edwards et al. (2024c). The flexibility of Gamma distributions to represent different datasets has been shown in Section 3, where bi-modal parameter regimes or sequential indoor interactions could be accounted for by considering different combinations of Gamma distributions. However, there might be situations where different distributions might provide a better fit to data for a particular parameter, in which case our results would need to be extended, if possible. For example, quanta emission rates have been modelled as log-normal in the literature with variation across multiple orders of magnitude (Jones et al. 2024). Future work might be able to partially address this limitation. For example, the analytical mean infection risk in Equation (3) can be written in terms of the

Laplace transform of the random parameter. We recall that for a random variable  $X$  with density  $f_X(\cdot)$ , the Laplace-Stieltjes transform is defined as  $\mathcal{L}_{f_X}(s) = \int_0^\infty e^{-sx} f_X(x) dx$ . Then, the mean infection risk  $\mathbb{E}[P_{\text{infection}}]$  for a generally-distributed quanta emission rate  $q \sim G(\cdot)$  can be computed as

$$\mathbb{E}[P_{\text{infection}}] = 1 - \int_0^\infty e^{-\frac{lbqT}{Q}} f_G(q) dq = 1 - \mathcal{L}_{f_G}\left(\frac{lbT}{Q}\right),$$

which could be applied to any distribution with a well-defined Laplace-Stieltjes transform.

Future work could extend this framework by integrating more detailed behavioural and environmental data or exploring non-parametric or data-driven distributions. Moreover, incorporating real-time monitoring of indoor air quality and occupancy could enable dynamic risk assessments and inform adaptive mitigation strategies. Ultimately, our stochastic extension of the Wells–Riley model provides a valuable analytical tool for robust and precautionary infection risk estimation in diverse indoor settings under environmental and behavioural uncertainty, and population heterogeneity. These insights into the impact of parameter uncertainty in infection risk assessment have important implications for pandemic preparedness and resilience, particularly given the potential for emergence of novel airborne pathogens through reassortment and other evolutionary mechanisms. Segmented RNA viruses pose particular concern, as reassortment between co-infecting strains can generate progeny with altered transmission characteristics, exemplified by influenza pandemic strains (Neumann et al. 2009, Taubenberger and Kash 2010). Similarly, Bunyaviruses with tripartite segmented genomes demonstrate significant reassortment potential that could alter virulence and transmissibility (Briese et al. 2013, Rezelj et al. 2019). While most are vector-borne, some Bunyaviruses show person-to-person transmission capabilities (e.g., Andes Hantavirus via aerosols (Martinez et al. 2005) and SFTS virus through contact transmission (Gai et al. 2012)) suggesting biological potential for enhanced airborne transmission in reassortant strains. Our stochastic framework's ability to account for uncertainty in human behaviour and environmental conditions provides essential tools for rapidly assessing transmission risks from such emerging pathogens, where precise parameter estimates may be unavailable during initial outbreak responses.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## ACKNOWLEDGMENTS

The authors would like to thank the editor and two anonymous reviewers for their constructive comments, which significantly improved the manuscript.

## REFERENCES

- Abney, S.E., Higham, C.A., Wilson, A.M. et al. (2024) Transmission of Viruses from Restroom Use: A Quantitative Microbial Risk Assessment. *Food and Environmental Virology*, 16, 65–78. doi:10.1007/s12560-023-09580-1. URL <https://doi.org/10.1007/s12560-023-09580-1>
- Aganovic, A., Cao, G., Kurnitski, J. & Wargocki, P. (2023) New dose-response model and sars-cov-2 quanta emission rates for calculating the long-range airborne infection risk. *Building and Environment*, 228, 109924. doi:10.1016/j.buildenv.2022.109924, epub 2022 Dec 14. URL <https://doi.org/10.1016/j.buildenv.2022.109924>
- Aganovic, A., Kurnitski, J. & Wargocki, P. (2024) A quanta-independent approach for the assessment of strategies to reduce the risk of airborne infection. *Science of the Total Environment*, 927, 172278. doi:10.1016/j.scitotenv.2024.172278. URL <https://www.sciencedirect.com/science/article/pii/S0048969724024215?via%3Dihub>
- Aggarwal, C.C. & Reddy, C.K. (2014) *Data Clustering: Algorithms and Applications*, 1st Edition. : Chapman and Hall/CRC. URL <https://doi.org/10.1201/9781315373515>
- Bate, A.M., Miller, D., King, M.F., Moseley, K.A., Xu, J., Hall, I. et al. (2024) A quantitative microbial risk assessment approach to estimate exposure to SARS-CoV-2 on a bus. *Journal of Transport & Health*, 38, 101829. doi:<https://doi.org/10.1016/j.jth.2024.101829>. URL <https://www.sciencedirect.com/science/article/pii/S2214140524000756>
- Bazant, M.Z. & Bush, J.W.M. (2021) A guideline to limit indoor airborne transmission of covid-19. *Proceedings of the National Academy of Sciences*, 118(17), e2018995118. doi:10.1073/pnas.2018995118. URL <https://www.pnas.org/doi/abs/10.1073/pnas.2018995118>
- Bell, R. & Pliner, P.L. (2003) Time to eat: the relationship between the number of people eating and meal duration in three lunch settings. *Appetite*, 41(2), 215–218. doi:[https://doi.org/10.1016/S0195-6663\(03\)00109-0](https://doi.org/10.1016/S0195-6663(03)00109-0). URL <https://www.sciencedirect.com/science/article/pii/S0195666303001090>
- Briese, T., Calisher, C.H. & Higgs, S. (2013) Viruses of the family Bunyaviridae: are all available isolates reassortants? *Virology*, 446(1-2), 207–216. doi:10.1016/j.virol.2013.07.030. URL <https://www.sciencedirect.com/science/article/pii/S0042682213004509>
- Buonanno, G., Morawska, L. & Stabile, L. (2020a) Quantitative assessment of the risk of airborne transmission of SARS-CoV-2 infection: Prospective and retrospective applications. *Environment International*, 145, 106112. doi:[doi:10.1016/j.envint.2020.106112](https://doi.org/10.1016/j.envint.2020.106112). URL <https://www.sciencedirect.com/science/article/pii/S0160412020320675>
- Buonanno, G., Stabile, L. & Morawska, L. (2020b) Estimation of airborne viral emission: Quanta emission rate of sars-cov-2 for infection risk assessment. *Environment International*, 141, 105794. doi:<https://doi.org/10.1016/j.envint.2020.105794>. URL <https://www.sciencedirect.com/science/article/pii/S0160412020312800>
- Casals Ventilation (n.d.) *Air Change Rates in Typical Rooms and Buildings*. Accessed: 2025-06-18. URL <https://www.casals.com/en/ventilation-tools/calculation-help/air-change-rates-in-typical-rooms-and-buildings/>
- Chen, L., Ban, G., Long, E., Kalonji, G., Cheng, Z., Zhang, L. et al. (2021) Estimation of the SARS-CoV-2 transmission probability in confined traffic space and evaluation of the mitigation strategies. *Environmental Science and Pollution Research*, 28, 42204–42216. URL <https://link.springer.com/article/10.1007/s11356-021-13617-y>
- Chen, S.C., Chang, C.F. & Liao, C.M. (2006) Predictive models of control strategies involved in containing indoor airborne infections. *Indoor Air*, 16(6), 469–481. doi:<https://doi.org/10.1111/j.1600-0668.2006.00443.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0668.2006.00443.x>

- 668.2006.00443.x
- Cheng, P., Yen, H.L. & Li, Y. (2025) Infectious quanta generation rates of sars-cov-2 variants throughout the infection course. *Building and Environment*, 283, 113333. doi:https://doi.org/10.1016/j.buildenv.2025.113333. URL https://www.sciencedirect.com/science/article/pii/S0360132325008121
- Cummins, E., Kennedy, R. & Cormican, M. (2010) Quantitative risk assessment of Cryptosporidium in tap water in Ireland. *Science of the Total Environment*, 408(4), 740–753. URL https://doi.org/10.1016/j.scitotenv.2009.11.008
- Davies, A., Thompson, K.A., Giri, K., Kafatos, G., Walker, J. & Bennett, A. (2013) Testing the efficacy of homemade masks: Would they protect in an influenza pandemic? *Disaster Medicine and Public Health Preparedness*, 7(4), 413–418. doi:10.1017/dmp.2013.43. URL https://www.cambridge.org/core/journals/disaster-medicine-and-public-health-preparedness/article/testing-the-efficacy-of-homemade-masks-would-they-protect-in-an-influenza-pandemic/0921A05A69A9419C862FA2F35F819D55
- Denpetkul, T., Pumkaew, M., Sittipunsakda, O., Leangwutiwong, P., Mongkolsuk, S. & Sirikanchana, K. (2022) Effects of face masks and ventilation on the risk of SARS-CoV-2 respiratory transmission in public toilets: a quantitative microbial risk assessment. *Journal of Water and Health*, 20(2), 300–313. doi:10.2166/wh.2022.190. URL https://doi.org/10.2166/wh.2022.190
- Deutsches Institut für Normung (2019) *DIN 1946-6: Ventilation and air conditioning – Part 6: Ventilation for residential buildings – General requirements, requirements for measuring, execution and commissioning, maintenance*. German Institute for Standardization (DIN), dIN 1946-6:2019-12.
- Doménech, E., Martorell, S., Kombo-Mpindou, G., Macián-Cervera, J. & Escuder-Bueno, I. (2022) Risk assessment of Cryptosporidium intake in drinking water treatment plant by a combination of predictive models and event-tree and fault-tree techniques. *Science of The Total Environment*, 838, 156500. doi:https://doi.org/10.1016/j.scitotenv.2022.156500. URL https://www.sciencedirect.com/science/article/pii/S0048969722035975
- Edwards, A., King, M.F., López-García, M., Peckham, D. & Noakes, C. (2024a) Assessing the effects of transient weather conditions on airborne transmission risk in naturally ventilated hospitals. *Journal of Hospital Infection*, 148, 1–10. doi:https://doi.org/10.1016/j.jhin.2024.02.017. URL https://www.sciencedirect.com/science/article/pii/S01957670124000756
- Edwards, A.J. (2024b) Transient models to assess transmission and control of airborne infection risks in a respiratory ward, (PhD Thesis; Embargoed until 01/12/2025). URL https://etheses.whiterose.ac.uk/id/eprint/35876/
- Edwards, A.J., Benson, L., Guo, Z., López-García, M., Noakes, C.J., Peckham, D. et al. (2023) A mathematical model for assessing transient airborne infection risks in a multi-zone hospital ward. *Building and Environment*, 238, 110344. doi:https://doi.org/10.1016/j.buildenv.2023.110344. URL https://www.sciencedirect.com/science/article/pii/S0360132323003712
- Edwards, A.J., King, M.F., López-García, M., Peckham, D. & Noakes, C.J. (2026) The role of natural ventilation in long-range airborne transmission in a hospital respiratory ward: A monte carlo simulation. *Indoor Environments*, 3(1), 100153. doi:https://doi.org/10.1016/j.indenv.2026.100153. URL https://www.sciencedirect.com/science/article/pii/S2950362026000056
- Edwards, A.J., King, M.F., Noakes, C.J., Peckham, D. & López-García, M. (2024c) The Wells–Riley model revisited: Randomness, heterogeneity, and transient behaviours. *Risk Analysis*, 44(9), 2125–2147. doi:https://doi.org/10.1111/risa.14295. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/risa.14295
- Euser, S., Aronson, S., Manders, I., van Lelyveld, S., Herpers, B., Sinnige, J. et al. (2021) Sars-cov-2 viral-load distribution reveals that viral loads increase with age: a retrospective cross-sectional cohort study. *International Journal of Epidemiology*, 50(6), 1795–1803. doi:10.1093/ije/dyab145. URL https://doi.org/10.1093/ije/dyab145
- Ferretti, L., Ledda, A., Wymant, C., Zhao, L., Ledda, V., Abeler-Dörner, L. et al. (2020) The timing of covid-19 transmission. *MedRxiv*, 2020–09. URL https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v2
- Gai, Z., Liang, M., Zhang, Y., Zhang, S., Jin, C., Wang, S.W. et al. (2012) Person-to-person transmission of severe fever with thrombocytopenia syndrome bunyavirus through blood contact. *Clinical Infectious Diseases*, 54(2), 249–252. doi:10.1093/cid/cir776. URL https://academic.oup.com/cid/article/54/2/249/468741
- Gammaitoni, L. & Nucci, M.C. (1997) Using a mathematical model to evaluate the efficacy of TB control measures. *Emerging Infectious Diseases*, 3(3), 335–342. doi:10.3201/eid0303.970310. URL https://wwwnc.cdc.gov/eid/article/3/3/97-0310\_article
- Glasser, L., Kohl, K., Koutschan, C., Moll, V. & Straub, A. (2012) The integrals in Gradshteyn and Ryzhik. Part 22: Bessel-K functions. *Scientia. Series A: Mathematical Sciences. New Series*, 22.
- Gupta, J.K., Lin, C.H. & Chen, Q. (2012) Risk assessment of airborne infectious diseases in aircraft cabins. *Indoor Air*, 22(5), 388–395. doi:https://doi.org/10.1111/j.1600-0668.2012.00773.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0668.2012.00773.x
- Haas, C.N., Rose, J.B., Gerba, C. & Regli, S. (1993) Risk Assessment of Virus in Drinking Water. *Risk Analysis*, 13(5), 545–552. doi:https://doi.org/10.1111/j.1539-6924.1993.tb00013.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.1993.tb00013.x
- Haas, C.N., Rose, J.B. & Gerba, C.P. (2014) *Quantitative microbial risk assessment*. : John Wiley & Sons.
- Hamner, L. (2020) High SARS-CoV-2 attack rate following exposure at a choir practice—Skagit County, Washington, March 2020. *MMWR. Morbidity and mortality weekly report*, 69. URL https://www.cdc.gov/mmwr/volumes/69/wr/mm6919e6.htm
- Henriques, A., Jia, W., Aleixo, L., Mounet, N., Fontana, L., Simniceanu, A. et al. (2025) An integrated airborne transmission risk assessment model for respiratory viruses: short- and long-range contributions. *Journal of the Royal Society Interface*, 22(223), 20240740. doi:10.1098/rsif.2024.0740. URL https://royalsocietypublishing.org/rsif/article/22/223/20240740/90689/An-integrated-airborne-transmission-risk
- Huang, J., Jones, P. & He, X. (2022) Masks, ventilation and exposure time: A web-based calculator of indoor covid-19 infection risk. *Frontiers in Built Environment*, Volume 8 - 2022. doi:10.3389/fbuil.2022.986923. URL https://www.frontiersin.org/journals/built-environment/articles/10.3389/fbuil.2022.986923
- Ildon, C., Jones, B., Sharpe, P., Cevik, M. & Fitzgerald, S. (2022) A population framework for predicting the proportion of people infected by the far-field airborne transmission of sars-cov-2 indoors. *Building and Environment*, doi:10.1016/j.buildenv.2022.109309. URL https://www.sciencedirect.com/science/article/pii/S0360132322005431?via%3Dihub
- Jensen, J.L.W.V. (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1), 175–193.

- doi:10.1007/BF02418571.  
URL <https://doi.org/10.1007/BF02418571>
- Jones, B., Iddon, C. & Sherman, M. (2024) Quantifying quanta: Determining emission rates from clinical data. *Indoor Environments*, 1(3), 100025. doi:<https://doi.org/10.1016/j.indenv.2024.100025>. URL <https://www.sciencedirect.com/science/article/pii/S2950362024000225>
- Jones, B., Iddon, C., Zaatari, M., Wargocki, P. & Bruns, R. (2025) Risk modeling for ashrae standard 241-2023 – control of infectious aerosols. *Building and Environment*, 283, 113318. doi:10.1016/j.buildenv.2025.113318. URL <https://www.sciencedirect.com/science/article/pii/S036013232500798X?via%3Dihub>
- Jones, B., Sharpe, P., Iddon, C., Hathway, E.A., Noakes, C.J. & Fitzgerald, S. (2021a) Modelling uncertainty in the relative risk of exposure to the sars-cov-2 virus by airborne aerosol transmission in well mixed indoor air. *Building and Environment*, 191, 107617. doi:10.1016/j.buildenv.2021.107617. URL <https://www.sciencedirect.com/science/article/pii/S0360132321000305?via%3Dihub>
- Jones, T.C., Biele, G., Mühlemann, B., Veith, T., Schneider, J., Beheim-Schwarzbach, J. et al. (2021b) Estimating infectiousness throughout sars-cov-2 infection course. *Science*, 373(6551), eabi5273. doi:10.1126/science.abi5273. URL <https://www.science.org/doi/abs/10.1126/science.abi5273>
- Karia, R., Gupta, I., Khandait, H., Yadav, A. & Yadav, A. (2020) COVID-19 and its modes of transmission. *SN Comprehensive Clinical Medicine*, 2(10), 1798–1801. doi:10.1007/s42399-020-00498-4. URL <https://doi.org/10.1007/s42399-020-00498-4>
- King, M., Wilson, A., López-García, M., Proctor, J., Peckham, D., Clifton, I. et al. (2021) Why is mock care not a good proxy for predicting hand contamination during patient care? *Journal of Hospital Infection*, 109, 44–51. URL [https://www.journalofhospitalinfection.com/article/S0195-6701\(20\)30540-5/fulltext](https://www.journalofhospitalinfection.com/article/S0195-6701(20)30540-5/fulltext)
- King, M.F., Wilson, A.M., Weir, M.H., López-García, M., Proctor, J., Hiwar, W. et al. (2022) Modeling fomite-mediated SARS-CoV-2 exposure through personal protective equipment doffing in a hospital environment. *Indoor Air*, 32(1), e12938. doi:<https://doi.org/10.1111/ina.12938>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ina.12938>
- López-García, M., King, M.F. & Noakes, C.J. (2019) A Multicompartiment SIS Stochastic Model with Zonal Ventilation for the Spread of Nosocomial Infections: Detection, Outbreak Management, and Infection Control. *Risk Analysis*, 39(8), 1825–1842. doi:10.1111/risa.13300, epub 2019 Mar 29.
- Martinez, V.P., Bellomo, C., San Juan, J., Pinna, D., Forlenza, R., Elder, M. et al. (2005) Person-to-person transmission of Andes virus. *Emerging Infectious Diseases*, 11(12), 1848–1853. doi:10.3201/eid1112.050501. URL [https://wwwnc.cdc.gov/eid/article/11/12/05-0501\\_article](https://wwwnc.cdc.gov/eid/article/11/12/05-0501_article)
- McShane, E.J. (1937) Jensen's inequality. *Bulletin of the American Mathematical Society*, 43(8), 521 – 527.
- Mikszewski, A., Stabile, L., Buonanno, G. & Morawska, L. (2022) The airborne contagiousness of respiratory viruses: A comparative analysis and implications for mitigation. *Geoscience Frontiers*, 13(6), 101285. doi:<https://doi.org/10.1016/j.gsf.2021.101285>. URL <https://www.sciencedirect.com/science/article/pii/S1674987121001493>
- Miller, D., King, M.F., Nally, J., Drodge, J.R., Reeves, G.I., Bate, A.M. et al. (2022) Modeling the factors that influence exposure to SARS-CoV-2 on a subway train carriage. *Indoor Air*, 32(2), e12976. doi:<https://doi.org/10.1111/ina.12976>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ina.12976>
- 6
- Miller, S.L., Nazaroff, W.W., Jimenez, J.L., Boerstra, A., Buonanno, G., Dancer, S.J. et al. (2021) Transmission of SARS-CoV-2 by inhalation of respiratory aerosol in the Skagit Valley Chorale superspreading event. *Indoor air*, 31(2), 314–323. URL <https://onlinelibrary.wiley.com/doi/10.1111/ina.12751>
- Nazaroff, W.W. (2021) Residential air-change rates: A critical review. *Indoor Air*, 31(2), 282–313. URL <https://onlinelibrary.wiley.com/doi/10.1111/ina.12785>
- Neumann, G., Noda, T. & Kawaoka, Y. (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*, 459(7249), 931–939. doi:10.1038/nature08157. URL <https://www.nature.com/articles/nature08157>
- Nicas, M. (1996) An Analytical Framework for Relating Dose, Risk, and Incidence: An Application to Occupational Tuberculosis Infection. *Risk Analysis*, 16(4), 527–538. doi:<https://doi.org/10.1111/j.1539-6924.1996.tb01098.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.1996.tb01098.x>
- Nicas, M. & Jones, R.M. (2009) Relative Contributions of Four Exposure Pathways to Influenza Infection Risk. *Risk Analysis*, 29(9), 1292–1303. doi:<https://doi.org/10.1111/j.1539-6924.2009.01253.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.2009.01253.x>
- NIST (2025a) *DLMF, Equation 10.32.E10*. <https://dlmf.nist.gov/10.32#E10>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025b) *DLMF, Equation 10.30.E2*. <https://dlmf.nist.gov/10.30#E2>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025c) *DLMF, Equation 13.4.E4*. <https://dlmf.nist.gov/13.4#E4>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025d) *DLMF, Equation 13.2.6: Kummer's Equation*. <https://dlmf.nist.gov/13.2#E6>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025e) *DLMF, Equation 13.2.11: Kummer's Equation*. <https://dlmf.nist.gov/13.2#E11>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025f) *DLMF, Equation 8.2.E2*. <https://dlmf.nist.gov/8.2#E2>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025g) *DLMF, Equation 8.4.E8*. <https://dlmf.nist.gov/8.4#E8>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- NIST (2025h) *DLMF, Equation 8.4.E15*. <https://dlmf.nist.gov/8.4#E15>, national Institute of Standards and Technology, Release 1.2.4 (2025-03-15).
- Office for National Statistics (2023) *Coronavirus (COVID-19) Latest Insights*. Accessed: 2024-03-07. URL <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronavirus/covid19/latestinsights>
- Paraskevopoulos, S., Vrachimis, S., Kyriakou, M., Eliades, D.G., Smeets, P., Polycarpou, M. et al. (2024) Modeling the health impact of wastewater contamination events in drinking water networks. *Journal of Cleaner Production*, 479, 143997. doi:<https://doi.org/10.1016/j.jclepro.2024.143997>. URL <https://www.sciencedirect.com/science/article/pii/S0959652624034462>
- Pleil, J.D., Wallace, M.A.G., Davis, M.D. & Matty, C.M. (2021) The physics of human breathing: flow, timing, volume, and pressure parameters for normal, on-demand, and ventilator respiration. *Journal of breath research*, 15(4), 042002. URL <https://iopscience.iop.org/article/10.1088/1752-7163/ac2589>

- Qian, H., Li, Y., Nielsen, P.V. & Huang, X. (2009) Spatial distribution of infection risk of SARS transmission in a hospital ward. *Building and Environment*, 44(8), 1651–1658. doi:https://doi.org/10.1016/j.buildenv.2008.11.002. URL https://www.sciencedirect.com/science/article/pii/S0360132308002564
- Qian, H. & Zheng, X. (2018) Ventilation control for airborne transmission of human exhaled bio-aerosols in buildings. *Journal of Thoracic Disease*, 10(Suppl 19), S2295–S2304. doi:10.21037/jtd.2018.01.24. URL https://jtd.amegroups.org/article/view/18723/17349
- Repace, J. (2000) Can ventilation control secondhand smoke in the hospitality industry? *University of California, eScholarship*. URL https://escholarship.org/uc/item/1tt9k4xb
- Rezelj, V.V., Mottram, T.J., Hughes, J., Elliott, R.M., Kohl, A. & Brennan, B. (2019) M segment-based minigenomes and virus-like particle assays as an approach to assess the potential of tick-borne phlebovirus genome reassortment. *Journal of Virology*, 93(6), e02068–18. doi:10.1128/JVI.02068-18. URL https://journals.asm.org/doi/10.1128/jvi.02068-18
- Riley, E., Murphy, G. & Riley, R. (1978) Airborne spread of measles in a suburban elementary school. *American journal of epidemiology*, 107(5), 421–432. URL https://academic.oup.com/aje/article-abstract/107/5/421/58522?redirectedFrom=fulltext
- Rothamer, D.A., Sanders, S., Reindl, D. & Bertram, T.H. (2021) Strategies to minimize sars-cov-2 transmission in classroom settings: combined impacts of ventilation and mask effective filtration efficiency. *Science and Technology for the Built Environment*, 27(9), 1181–1203. doi:10.1080/23744731.2021.1944665. URL https://doi.org/10.1080/23744731.2021.1944665
- Rudnick, S. & Milton, D.K. (2003) Risk of indoor airborne infection transmission estimated from carbon dioxide concentration. *Indoor air*, 13(3). URL https://onlinelibrary.wiley.com/doi/abs/10.1034/j.1600-0668.2003.00189.x
- Schijven, J., Vermeulen, L.C., Swart, A., Meijer, A., Duizer, E. & de Roda Husman, A.M. (2021) Quantitative microbial risk assessment for airborne transmission of SARS-CoV-2 via breathing, speaking, singing, coughing, and sneezing. *Environmental health perspectives*, 129(4), 047002. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8016178/
- Sender, R., Bar-On, Y., Park, S.W., Noor, E., Dushoff, J. & Milo, R. (2022) The unmitigated profile of covid-19 infectiousness. *eLife*, 11, e79134. doi:10.7554/eLife.79134. URL https://doi.org/10.7554/eLife.79134
- Sze To, G.N. & Chao, C.Y.H. (2010) Review and comparison between the Wells–Riley and dose-response approaches to risk assessment of infectious respiratory diseases. *Indoor air*, 20(1), 2–16. URL https://doi.org/10.1111/j.1600-0668.2009.00621.x
- Sze-To, G.N., Yang, Y., Kwan, J.K., Yu, S.C. & Chao, C.Y. (2014) Effects of surface material, ventilation, and human behavior on indirect contact transmission risk of respiratory infection. *Risk analysis*, 34(5), 818–830. URL https://doi.org/10.1111/risa.12144
- Taubenberger, J.K. & Kash, J.C. (2010) Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host & Microbe*, 7(6), 440–451. doi:10.1016/j.chom.2010.05.009. URL https://www.cell.com/cell-host-microbe/fulltext/S1931-3128(10)00172-1
- The Lancet (2020) COVID-19: protecting health-care workers. *The Lancet*, 395(10228), 922. doi:10.1016/S0140-6736(20)30644-9, publisher: Elsevier. URL https://doi.org/10.1016/S0140-6736(20)30644-9
- Wang, Y., Deng, Z. & Shi, D. (2021) How effective is a mask in preventing COVID-19 infection? *Medical devices & sensors*, 4(1), e10163. URL https://onlinelibrary.wiley.com/doi/10.1002/mds3.10163
- Wells, W.F. (1955) *Airborne Contagion and Air Hygiene: An Ecological Study of Droplet Infections*. Cambridge, Massachusetts: Harvard University Press.
- Wilson, A.M. (2020) Covid-19 and use of non-traditional masks: how do various materials compare in reducing the risk of infection for mask wearers? *Journal of Hospital Infection*, 105(4), 640–642. doi:10.1016/j.jhin.2020.05.036. URL https://www.journalofhospitalinfection.com/article/S0195-6701(20)30276-0/fulltext
- World Health Organization (2016) *Quantitative Microbial Risk Assessment: Application for Water Safety Management*. Geneva, Switzerland: World Health Organization. URL https://www.who.int/publications/i/item/9789241565370
- Wu, P., Liu, F., Chang, Z., Lin, Y., Ren, M., Zheng, C. et al. (2021) Assessing asymptomatic, presymptomatic, and symptomatic transmission risk of severe acute respiratory syndrome coronavirus 2. *Clinical Infectious Diseases*, 73(6), e1314–e1320. doi:10.1093/cid/ciab271. URL https://doi.org/10.1093/cid/ciab271
- Xu, H., Intrator, O. & Bowblis, J.R. (2020) Shortages of staff in nursing homes during the COVID-19 pandemic: what are the driving factors? *Journal of the American Medical Directors Association*, 21(10), 1371–1377. URL https://www.jamda.com/article/S1525-8610(20)30691-5/fulltext



## APPENDIX

$$P_{infection} = 1 - e^{-\frac{bqT}{Q}}$$

when one (or two) of the parameters are considered random. We also estimate the probability of observing exactly  $n$  exposures during the indoor interaction. In particular, in Appendix A1, we focus on scenarios where a population parameter  $\{b, q, T, l\}$  is random. In Appendix A2, we consider the situation where the ventilation rate  $Q$  is random. In Appendix A3, we look at the scenario where a pair of parameters are simultaneously random.

First, let us consider one parameter  $X \in \{b, q, l, Q, T\}$  being random, and define  $g(X) = -\frac{bqT}{Q}$ , which is then a random variable. Then

$$P_{infection} = 1 - e^{g(X)} \Rightarrow X = g^{-1}(\ln(1 - P_{infection})).$$

We can express the cumulative distribution function of  $P_{infection}$  in terms of that of  $X$

$$F_{P_{infection}}(p) = F_X(g^{-1}(\ln(1 - p))).$$

Now, to obtain the density function of  $P_{infection}$ , we differentiate  $F_{P_{infection}}(p)$  with respect to  $p$

$$\frac{dF_{P_{infection}}(p)}{dp} = \frac{dF_X(g^{-1}(\ln(1 - p)))}{dp},$$

from which we obtain

$$f_{P_{infection}}(p) = f_X(g^{-1}(\ln(1-p))) \left| \frac{dx}{dp} \right|, \quad (A1)$$

where  $f_{P_{infection}}(\cdot)$  and  $f_X(\cdot)$  are the density functions of the random variables  $P_{infection}$  and  $X$ , respectively. This is known as the change of variables method. To find the expected infection risk, we compute:

$$\mathbb{E}[P_{infection}] = \int_0^1 p f_{P_{infection}}(p) dp = \int_0^\infty (1 - e^{g(x)}) f_X(x) dx. \quad (A2)$$

Finally, since the number of exposures  $E \sim \text{Bin}(S, P_{infection})$ , we have

$$\begin{aligned} \mathbb{P}(E = n) &= \int_0^\infty \binom{S}{n} (1 - e^{g(x)})^n (e^{g(x)})^{S-n} f_X(x) dx \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{(S-n+i)g(x)} f_X(x) dx, \end{aligned} \quad (A3)$$

for  $n = 0, \dots, S$ , where we have used the binomial expansion of  $(1 - e^{g(x)})^n$ .

## A1 UNCERTAINTY IN POPULATION PARAMETERS

### A1.1 Breathing rate, quanta emission rate or exposure time

We consider here a random quanta emission rate  $q \sim \text{Gamma}(k, \lambda)$  without any loss of generality, since other parameters  $T$  and  $b$  are symmetric in  $P_{infection}$ , so analogous results can be obtained. According to the arguments above, we have

$$g(q) = -\frac{lbT}{Q}q, \quad f_q(q) = \frac{\lambda^k q^{k-1} e^{-\lambda q}}{\Gamma(k)},$$

and since  $P_{infection} = 1 - e^{-\frac{lbT}{Q}q}$ , one gets  $q = g^{-1}(\ln(1 - P_{infection})) = \frac{-\ln(1 - P_{infection})Q}{lbT}$ . Thus,

$$\frac{dq}{dP_{infection}} = \frac{Q}{(1 - P_{infection})lbT}$$

and by Equation (A1), the density function of  $P_{infection}$  is

$$\begin{aligned} f_{P_{infection}}(p) &= \frac{\lambda^k \left(\frac{-\ln(1-p)Q}{lbT}\right)^{k-1} e^{\left(\frac{\ln(1-p)\lambda Q}{lbT}\right)} \frac{Q}{(1-p)lbT}}{\Gamma(k)} \\ &= \left(\frac{\lambda Q}{lbT}\right)^k \frac{[-\ln(1-p)]^{k-1} (1-p)^{\frac{\lambda Q}{lbT}-1}}{\Gamma(k)}, \end{aligned} \quad (A4)$$

for  $p \in (0, 1)$ . Using Equation (A2), we find the expectation of  $P_{infection}$  as

$$\begin{aligned} \mathbb{E}[P_{infection}] &= \int_0^\infty \left(1 - e^{-\frac{lbT}{Q}q}\right) \frac{\lambda^k q^{k-1} e^{-\lambda q}}{\Gamma(k)} dq \\ &= 1 - \frac{\lambda^k}{\Gamma(k)} \int_0^\infty e^{-q\left(\frac{lbT}{Q} + \lambda\right)} q^{k-1} dq \\ &= 1 - \left(\frac{\lambda}{\lambda + \frac{lbT}{Q}}\right)^k, \end{aligned} \quad (A5)$$

where we have used

$$\int_0^\infty u^b e^{-au} du = \frac{\Gamma(b+1)}{a^{b+1}}, \quad (*)$$

with  $a = \frac{lbT}{Q} + \lambda$  and  $b = k - 1$ . The probability of observing exactly  $n$  exposures is computed using Equation (A3) as

$$\begin{aligned} \mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{-q\left((S-n+i)\frac{lbT}{Q} + \lambda\right)} q^{k-1} \frac{\lambda^k}{\Gamma(k)} dq \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{\Gamma(k)}{\left((S-n+i)\frac{lbT}{Q} + \lambda\right)^k} \frac{\lambda^k}{\Gamma(k)} \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left(\frac{\lambda}{(S-n+i)\frac{lbT}{Q} + \lambda}\right)^k, \end{aligned} \quad (A6)$$

for  $n = 0, \dots, S$ , where we have again used Equation (\*), this time with  $a = (S-n+i)\frac{lbT}{Q} + \lambda$  and  $b = k - 1$ .

As described in Section 2.1.1, some of the uncertainty in  $q$  might arise from mitigations such as mask-wearing. In particular, given a mask wearing probability  $w_{mask}$ , and the mask's efficacy parameter  $\eta$ , the distribution for the effective quanta emission rate is given by

$$\begin{aligned} f_{q_{eff}}(q) &= \mathbb{P}(mask) f_{q_{eff}|mask}(q) + \mathbb{P}(no\ mask) f_{q_{eff}|no\ mask}(q) \\ &= w_{mask} \frac{1}{1-\eta} f_{q_{no\ mask}}\left(\frac{q}{1-\eta}\right) + (1-w_{mask}) f_{q_{no\ mask}}(q). \end{aligned}$$

with  $w_{mask} \in [0, 1]$ ,  $f_{q_{no\ mask}} \sim \text{Gamma}(k, \lambda)$  and where  $P_{infection} = 1 - e^{-\frac{lbT q_{eff}}{Q}}$ . In this scenario, our arguments above can be easily adapted. In particular, the density of  $P_{infection}$  is given by Equation (A1). By considering a linear combination for  $q_{eff}$ , only the density for the quanta emission rate changes, not the transformation to  $P_{infection}$ . Thus, we still have  $q = g^{-1}(\ln(1 - P_{infection})) = \frac{-\ln(1 - P_{infection})Q}{lbT}$ . This results in the following infection risk density

$$\begin{aligned} f_{P_{infection}}(p) &= \left(\frac{w_{mask}}{1-\eta} f_{q_{no\ mask}}\left(\frac{-\ln(1-p)Q}{lbT(1-\eta)}\right) + \right. \\ &\quad \left. \times (1-w_{mask}) f_{q_{no\ mask}}\left(\frac{-\ln(1-p)Q}{lbT}\right)\right) \left| \frac{dq}{dp} \right| \\ &= \frac{(-\ln(1-p))^{k-1}}{\Gamma(k)} \left(\frac{\lambda Q}{lbT}\right)^k \left(\frac{w_{mask}}{(1-\eta)^k} (1-p)^{\frac{\lambda Q}{lbT(1-\eta)}-1} \right. \\ &\quad \left. + (1-w_{mask})(1-p)^{\frac{\lambda Q}{lbT}-1}\right), \end{aligned} \quad (A7)$$

for  $p \in (0, 1)$ . The average infection risk is

$$\begin{aligned} \mathbb{E}[P_{infection}] &= \int_0^\infty \left(1 - e^{-\frac{lbT}{Q}q}\right) f_{q_{eff}}(q) dq \\ &= 1 - \left[ w_{mask} \left(\frac{\lambda}{\lambda + \frac{lbT}{Q}(1-\eta)}\right)^k \right. \\ &\quad \left. + (1-w_{mask}) \left(\frac{\lambda}{\lambda + \frac{lbT}{Q}}\right)^k \right], \end{aligned} \quad (A8)$$

and the probability of observing exactly  $n$  exposures is

$$\begin{aligned} \mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{(S-n+i)g(q)} f_{q_{\text{eff}}}(q) dq \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left[ w_{\text{mask}} \left( \frac{\lambda}{\lambda + (S-n+i) \frac{lbqT}{Q} (1-\eta)} \right)^k \right. \\ &\quad \left. + (1 - w_{\text{mask}}) \left( \frac{\lambda}{\lambda + (S-n+i) \frac{lbqT}{Q}} \right)^k \right]. \end{aligned} \quad (\text{A9})$$

for  $n \in \{0, 1, \dots, S\}$ . The results given by Equations (A7) - (A9) and their derivations are analogous to that of Equations (A4)-(A6), with the additional term and coefficients.

## A1.2 Disease prevalence

If the number of infectors is a discrete random variable  $I \sim \text{Bin}(N, \rho)$ , the corresponding per-capita probability of infection,  $P_{\text{infection}}$  is also a discrete random variable (i.e., will take a number of discrete values with different probabilities). In particular, since

$$P(I = i) = \binom{N}{i} \rho^i (1-\rho)^{N-i}, \quad i \in \{0, 1, \dots, N\},$$

the per-capita probability of infection  $P_{\text{infection}}$  takes the form of  $1 - e^{-\frac{baT}{Q} I}$  weighted with binomial probabilities  $P(I = i)$  for  $i \in \{0, 1, \dots, N\}$

$$f_{P_{\text{infection}}}(p) = \begin{cases} \binom{N}{i} \rho^i (1-\rho)^{N-i}, & \text{if } p = 1 - e^{-\frac{baT}{Q} i}, \\ 0, & \text{otherwise.} \end{cases}$$

The expectation of  $P_{\text{infection}}$  is characterised by:

$$\begin{aligned} \mathbb{E}[P_{\text{infection}}] &= \sum_{i=0}^N \binom{N}{i} \rho^i (1-\rho)^{N-i} \left( 1 - e^{-\frac{baT}{Q} i} \right) \\ &= 1 - \sum_{i=0}^N \left[ \binom{N}{i} \left( \rho e^{-\frac{baT}{Q}} \right)^i (1-\rho)^{N-i} \right] \\ &= 1 - \left[ \rho e^{-\frac{baT}{Q}} + (1-\rho) \right]^N. \end{aligned}$$

The probability of observing exactly  $n$  exposures is

$$\begin{aligned} \mathbb{P}(E = n) &= \sum_{i=0}^{N-n} \mathbb{P}(E = n \mid I = i) \mathbb{P}(I = i) \\ &= \binom{N}{n} \sum_{i=0}^{N-n} \binom{N-n}{i} \left( 1 - e^{-\frac{baT}{Q} i} \right)^n e^{-\frac{baT}{Q} i(N-n)} \rho^i (1-\rho)^{N-i}, \end{aligned}$$

for  $n = 0, 1, \dots, N$ .

## A2 UNCERTAINTY IN VENTILATION RATE

Let us consider  $Q \sim \text{Gamma}(r, \gamma)$  so that

$$g(Q) = -\frac{lbqT}{Q}, \quad f_Q(\omega) = \frac{\gamma^r \omega^{r-1} e^{-\gamma\omega}}{\Gamma(r)},$$

and since  $P_{\text{infection}} = 1 - e^{-\frac{lbqT}{Q}}$ , one gets  $Q = g^{-1}(\ln(1 - P_{\text{infection}})) = \frac{-lbqT}{\ln(1 - P_{\text{infection}})}$ . Thus,

$$\frac{dQ}{dP_{\text{infection}}} = \frac{-lbqT}{(1 - P_{\text{infection}})(\ln(1 - P_{\text{infection}}))^2}.$$

Despite this derivative being negative, we use its absolute value in calculating the density function of  $P_{\text{infection}}$ , as according to Equation (A1)

$$\begin{aligned} f_{P_{\text{infection}}}(p) &= \frac{\gamma^r \left( \frac{-lbqT}{\ln(1-p)} \right)^{r-1} e^{\frac{\gamma lbqT}{\ln(1-p)}}}{\Gamma(r)} \frac{lbqT}{(1-p)(\ln(1-p))^2} \\ &= \frac{(\gamma lbqT)^r}{\Gamma(r)} \frac{e^{\frac{\gamma lbqT}{\ln(1-p)}}}{(1-p)[-\ln(1-p)]^{r+1}}, \quad p \in (0, 1). \end{aligned}$$

Using Equation (A2), we find the expectation of  $P_{\text{infection}}$  as

$$\begin{aligned} \mathbb{E}[P_{\text{infection}}] &= \int_0^\infty \left( 1 - e^{-\frac{lbqT}{\omega}} \right) \frac{\gamma^r \omega^{r-1} e^{-\gamma\omega}}{\Gamma(r)} d\omega \\ &= 1 - \frac{\gamma^r}{\Gamma(r)} \int_0^\infty e^{-\gamma\omega - \frac{lbqT}{\omega}} \omega^{r-1} d\omega \\ &= 1 - \frac{1}{\Gamma(r)} \int_0^\infty e^{-u - \frac{\gamma lbqT}{u}} u^{r-1} du \\ &= 1 - \frac{2}{\Gamma(r)} (\gamma lbqT)^{r/2} K_r(2\sqrt{\gamma lbqT}) \end{aligned}$$

where we used the substitution  $u = \gamma\omega$  and (NIST (2025a))

$$K_\theta(\zeta) = \frac{1}{2} \left( \frac{\zeta}{2} \right)^\theta \int_0^\infty u^{-(\theta+1)} e^{-\left(u - \frac{\zeta^2}{4u}\right)} du, \quad (\dagger)$$

with  $\theta = -r$  and  $\zeta = 2\sqrt{\gamma lbqT}$ , alongside Proposition 2.1 from Glasser et al. (2012) which states  $K_{-\theta}(\zeta) = K_\theta(\zeta)$ . The probability of observing exactly  $n$  exposures is computed using Equation (A3) as

$$\begin{aligned} \mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{(S-n+i)g(\omega)} f_Q(\omega) d\omega \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{-\gamma\omega - (S-n+i) \frac{lbqT}{\omega}} \omega^{r-1} \frac{\gamma^r}{\Gamma(r)} d\omega \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{-u - (S-n+i) \frac{\gamma lbqT}{u}} u^{r-1} \frac{1}{\Gamma(r)} du \quad (\text{w/ Eq. } (\dagger)) \\ &= \frac{2}{\Gamma(r)} \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} a_i^{r/2} K_r(2\sqrt{a_i}), \end{aligned}$$

for  $n = 0, \dots, S$ , where  $a_i = \gamma lbqT(S - n + i)$  and  $u = \gamma\omega$ .

### A2.1 Accounting for mechanical ventilation

In Section 3.2.3, we look at increasing natural ventilation,  $Q$ , by a constant mechanical air change rate,  $Q_{\text{mech}} = \text{ACH}_{\text{mech}} \times V$  for mechanical air change rate  $\text{ACH}_{\text{mech}}$  and room volume  $V$ . To implement this in our results, we use the derivation from Appendix A2, instead setting  $g(Q) = -\frac{lbqT}{Q + Q_{\text{mech}}}$ . This means we are now considering  $P_{\text{infection}} = 1 - e^{-\frac{lbqT}{Q + Q_{\text{mech}}}}$  and  $Q = g^{-1}(\ln(1 - P_{\text{infection}})) = \frac{-lbqT}{\ln(1 - P_{\text{infection}})} - Q_{\text{mech}}$ . The new density

function for  $P_{infection}$  is

$$\hat{f}_{P_{infection}}(p) = \frac{\gamma^r \left( \frac{-lbqT}{\ln(1-p)} - Q_{mech} \right)^{r-1} e^{\frac{\gamma lbqT}{\ln(1-p)} + \gamma Q_{mech}} lbqT}{\Gamma(r)(1-p)(\ln(1-p))^2}, \quad p \in (0, 1). \quad (A10)$$

An analytical expression for the mean infection risk,  $\mathbb{E}[P_{infection}]$ , has not been obtained in this paper. The introduction of  $Q_{mech}$  increases the complexity of the underlying integral. The mean infection risk values for Figures 10b and 10d were obtained numerically instead.

### A3 UNCERTAINTY IN TWO MODEL PARAMETERS

Let us consider the situation here where two model parameters are random. Here, we consider two population-related parameters in  $\{b, q, T\}$  being (independently) random; say  $q \sim \text{Gamma}(k, \lambda)$  and  $T \sim \text{Gamma}(\ell, \mu)$  without any loss of generality. Given that  $q$  and  $T$  are independent, their joint density is  $f_{q,T}(q, t) = f_q(q)f_T(t)$ . Using the change of variables method, we can write

$$f_{Z,T}(z, t) = f_T(t)f_q(z/t) \frac{dq}{dz} = f_T(t)f_q(z/t) \frac{1}{t},$$

where  $Z = qT$ . We then integrate this joint density over possible values of  $T$  to get:

$$\begin{aligned} f_Z(z) &= \int_0^\infty t^{-1} f_T(t) f_q(z/t) dt = \int_0^\infty t^{-1} \frac{\mu^\ell t^{\ell-1} e^{-\mu t}}{\Gamma(\ell)} \frac{\lambda^k \left(\frac{z}{t}\right)^{k-1} e^{-\frac{\lambda z}{t}}}{\Gamma(k)} dt \\ &= \frac{\mu^k \lambda^k z^{k-1}}{\Gamma(\ell)\Gamma(k)} \int_0^\infty u^{\ell-k-1} e^{-u} \frac{(2\sqrt{\lambda\mu z})^2}{4u} du \\ &= \frac{2(\lambda\mu z)^{\frac{1}{2}(k+\ell)} K_{k-\ell}(2\sqrt{\lambda\mu z})}{z\Gamma(\ell)\Gamma(k)}, \end{aligned}$$

where we used the substitution  $u = \mu t$  and Equation (†) with  $r = k - \ell$  and  $\zeta = \sqrt{\lambda\mu z}$ . So, we have

$$g(Z) = -\frac{lb}{Q}Z, \quad f_Z(z) = \frac{2(\lambda\mu z)^{\frac{1}{2}(k+\ell)} K_{k-\ell}(2\sqrt{\lambda\mu z})}{z\Gamma(\ell)\Gamma(k)},$$

where  $P_{infection} = 1 - e^{-\frac{lb}{Q}Z}$  results in  $Z = g^{-1}(\ln(1 - P_{infection})) = \frac{-\ln(1 - P_{infection})Q}{lb}$ . Therefore,

$$\frac{dZ}{dP_{infection}} = \frac{Q}{lb(1 - P_{infection})}$$

and the resulting density function of  $P_{infection}$  is

$$\begin{aligned} \hat{f}_{P_{infection}}(p) &= \frac{2(\lambda\mu \left(\frac{-\ln(1-p)Q}{lb}\right)^{\frac{1}{2}(k+\ell)} K_{k-\ell}\left(2\sqrt{\lambda\mu \frac{-\ln(1-p)Q}{lb}}\right)}{-\frac{\ln(1-p)Q}{lb}\Gamma(\ell)\Gamma(k)} \frac{Q}{lb(1-p)} \\ &= \frac{2(\lambda\mu)^{\frac{k+\ell}{2}} \left[-\frac{1}{c}\ln(1-p)\right]^{\frac{k+\ell}{2}-1} K_{k-\ell}\left(2\sqrt{d[-\ln(1-p)]}\right)}{c\Gamma(k)\Gamma(\ell)(1-p)} \end{aligned}$$

with  $c = \frac{lb}{Q}$ ,  $d = \frac{\lambda\mu}{c}$  and for  $p \in (0, 1)$ . The expected per-capita infection risk is

$$\begin{aligned} \mathbb{E}[P_{infection}] &= \int_0^\infty \int_0^\infty \left(1 - e^{-\frac{lbqt}{Q}}\right) f_q(q) f_T(t) dq dt \\ &= 1 - \int_0^\infty \frac{\lambda^k \mu^\ell}{\Gamma(k)\Gamma(\ell)} q^{k-1} e^{-\lambda q} \frac{\Gamma(\ell)}{\left(\mu + \frac{lbq}{Q}\right)^\ell} dq \quad (\text{using Eq. } (\star)) \\ &= 1 - \left(\frac{\lambda\mu Q}{lb}\right)^k \frac{1}{\Gamma(k)} \int_0^\infty e^{-\frac{\lambda\mu Q}{lb}v} v^{k-1} (1+v)^{-\ell} dv \\ &= 1 - d^k U(k, k - \ell + 1, d), \end{aligned} \quad (A11)$$

where we have used  $v = \frac{lb}{\mu Q}q$ , and  $U(a, b, z)$  is the confluent hypergeometric function of the second kind (NIST (2025c)), sometimes referred to as the *Tricomi* function

$$U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zv} v^{a-1} (1+v)^{b-a-1} dv. \quad (\ddagger)$$

By solving the double integral with respect to  $q$  first, or using Kummer's transformation (NIST 2025e)

$$U(a, b, z) = z^{1-b} U(1+a-b, 2-b, z), \quad (\star\star)$$

an alternate (equivalent) form for the expected per-capita infection risk can be obtained:

$$\mathbb{E}[P_{infection}] = 1 - d^\ell U(\ell, \ell - k + 1, d).$$

The probability of observing exactly  $n$  exposures is given by

$$\begin{aligned} \mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty \int_0^\infty e^{-(S-n+i)\frac{lbqt}{Q}} f_q(q) f_T(t) dq dt, \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty \frac{\lambda^k \mu^\ell}{\Gamma(k)\Gamma(\ell)} q^{k-1} e^{-\lambda q} \\ &\quad \times \frac{\Gamma(\ell)}{\left(\mu + \frac{lb}{Q}(S-n+i)q\right)^\ell} dq, \quad (\text{using Eq. } (\star)) \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left(\frac{\lambda\mu Q}{lb(S-n+i)}\right)^k \frac{1}{\Gamma(k)} \\ &\quad \times \int_0^\infty e^{-\frac{\lambda\mu Q}{lb(S-n+i)}v} v^{k-1} (1+v)^{-\ell} dv \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \gamma_i^k U(k, k - \ell + 1, \gamma_i), \quad (\text{using Eq. } (\ddagger)) \end{aligned} \quad (A12)$$

for  $n = 0, \dots, S$ , where we have used the substitution  $v = \frac{lb}{\mu Q}q(S-n+i)$ , parameter  $\gamma_i = d/(S-n+i)$ . By solving the double integral with respect to  $q$  first, or using Kummer's transformation (Equation  $(\star\star)$ ) an alternate (equivalent) form for the  $n$ -exposure probability can be obtained

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \gamma_i^\ell U(\ell, \ell - k + 1, \gamma_i),$$

for  $n = 0, \dots, S$ . We note that if one of the parameters is Erlang-distributed, where the shape parameter is a positive integer, Equations (A11) and (A12) may be written as finite sums of upper incomplete gamma functions. For example, if  $k \in \mathbb{N}$  so that  $q$  is Erlang-distributed,

one gets

$$\begin{aligned}
\mathbb{E}[P] &= \int_0^\infty \int_0^\infty \left(1 - e^{-\frac{lbqt}{q}}\right) f_q(q) f_T(t) dq dt \\
&= \int_0^\infty \left(1 - \left(\frac{\mu}{\mu + \frac{lbq}{Q}}\right)^\ell\right) \frac{\lambda^k q^{k-1} e^{-\lambda q}}{(k-1)!} dq \\
&= 1 - \frac{\lambda^k \mu^\ell}{(k-1)!} \int_{\frac{\lambda\mu Q}{lb}}^\infty \frac{\left(w - \frac{\lambda\mu Q}{lb}\right)^{k-1} \lambda^{1-k} e^{-w + \frac{\lambda\mu Q}{lb}}}{w^\ell \left(\frac{lb}{\lambda Q}\right)^\ell} \frac{1}{\lambda} dw \\
&= 1 - \frac{e^{\frac{\lambda\mu Q}{lb}}}{(k-1)!} \sum_{i=0}^{\infty} \binom{k-1}{i} (-1)^i \left(\frac{\lambda\mu Q}{lb}\right)^{i+\ell} \int_{\frac{\lambda\mu Q}{lb}}^\infty w^{k-\ell-i-1} e^{-w} dw, \\
&= 1 - \frac{e^d}{(k-1)!} \sum_{i=0}^{k-1} \binom{k-1}{i} (-1)^i d^{i+\ell} \Gamma(k-\ell-i, d),
\end{aligned}$$

where  $w = \lambda q + \frac{\lambda\mu Q}{lb}$  and  $\Gamma(s, x)$  is the upper incomplete gamma function (NIST 2025f)

$$\Gamma(s, x) = \int_x^\infty w^{s-1} e^{-w} dw. \quad (\ddagger\ddagger)$$

Similarly, one gets

$$\mathbb{P}(E = n) = \binom{S}{n} \sum_{i=0}^n \sum_{j=0}^{k-1} \frac{(-1)^{i+j}}{(k-1)!} \binom{n}{i} \binom{k-1}{j} y_i^{\ell+j} e^{y_i} \Gamma(k-\ell-j, y_i),$$

for  $n = 0, \dots, S$ . Similar arguments apply when  $\ell \in \mathbb{N}$ , just by swapping  $k$  with  $\ell$  and  $\lambda$  for  $\mu$  in each derivation.

Finally, one can consider one random population-related parameter in  $\{b, q, T\}$  and one random ventilation parameter ( $Q$ ). Because  $Q$  appears on the denominator of the product of the exponent in the classical Wells-Riley model, new equations for  $P_{infection}$  arise. Let us assume that  $q$  is random without loss of generality, as the methods will be identical for  $b$  or  $T$ , so that  $q \sim \text{Gamma}(k, \lambda)$  and  $Q \sim \text{Gamma}(r, \gamma)$ . The quotient of two Gamma distributions is known to result in a Generalised Beta Prime distribution. In particular, for  $Y = q/Q$  we have  $Y \sim \beta'(k, r, 1, \frac{\gamma}{\lambda})$ , with density function

$$f_Y(y) = \frac{\lambda^k \gamma^r y^{k-1} \Gamma(k+r)}{\Gamma(k)\Gamma(r)(\lambda y + \gamma)^{k+r}}.$$

We now have,

$$g(Y) = -lbTY.$$

Moreover,  $Y = g^{-1}(\ln(1 - P_{infection})) = -\frac{\ln(1 - P_{infection})}{lbT}$ , which means

$$\frac{dY}{dP_{infection}} = \frac{1}{(1 - P_{infection})lbT}.$$

Using Equation (A1), we find the density

$$\begin{aligned}
f_{P_{infection}}(p) &= \frac{\lambda^k \gamma^r \left[-\frac{\ln(1-p)}{lbT}\right]^{k-1} \Gamma(k+r)}{\Gamma(k)\Gamma(r)(\lambda \left[-\frac{\ln(1-p)}{lbT}\right] + \gamma)^{k+r}} \frac{1}{(1-p)lbT} \\
&= \frac{\Gamma(k+r)}{\Gamma(k)\Gamma(r)} \frac{\lambda^k (\gamma lbT)^r [-\ln(1-p)]^{k-1}}{(1-p)([-\ln(1-p)]\lambda + \gamma lbT)^{k+r}},
\end{aligned}$$

for  $p \in (0, 1)$ . The expected per-capita infection risk is

$$\begin{aligned}
\mathbb{E}[P_{infection}] &= \int_0^\infty \int_0^\infty \left(1 - e^{-\frac{lbqt}{\omega}}\right) f_q(q) f_Q(\omega) dq d\omega \\
&= 1 - \int_0^\infty \frac{\lambda^k \gamma^r}{\Gamma(k)\Gamma(r)} \omega^{r-1} e^{-\gamma \omega} \frac{\Gamma(k)}{(\lambda + \frac{lbT}{\omega})^k} d\omega \quad (\text{using Eq. } (\star)) \\
&= 1 - \left(\frac{\gamma lbT}{\lambda}\right)^r \frac{1}{\Gamma(r)} \int_0^\infty e^{-\frac{\gamma lbT}{\lambda} v} v^{k+r-1} (1+v)^{-k} dv \\
&= 1 - \left(\frac{\gamma lbT}{\lambda}\right)^r \frac{\Gamma(k+r)}{\Gamma(r)} U\left(k+r, r+1, \frac{\gamma lbT}{\lambda}\right), \quad (\text{using Eq. } (\ddagger))
\end{aligned} \quad (\text{A13})$$

where we have used  $v = \frac{\lambda}{lbT} \omega$ . The probability of observing exactly  $n$  exposures is

$$\begin{aligned}
\mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty \int_0^\infty e^{-(S-n+i)\frac{lbT}{\omega}} f_q(q) f_Q(\omega) dq d\omega \\
&= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty \frac{\lambda^k \gamma^r}{\Gamma(k)\Gamma(r)} \omega^{r-1} \\
&\quad \times e^{-\gamma \omega} \frac{\Gamma(k)}{(\lambda + \frac{lbT}{\omega}(S-n+i))^k} d\omega, \quad (\text{using Eq. } (\star)) \\
&= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left(\frac{\gamma lbT}{\lambda}(S-n+i)\right)^r \frac{1}{\Gamma(r)} \\
&\quad \times \int_0^\infty e^{-\frac{\gamma lbT}{\lambda}(S-n+i)v} v^{k+r-1} (1+v)^{-k} dv \\
&= \binom{S}{n} \sum_{i=0}^n \binom{n}{i} (-1)^i \alpha_i^r \frac{\Gamma(k+r)}{\Gamma(r)} U\left(k+r, r+1, \alpha_i\right), \quad (\text{w/ Eq. } (\ddagger))
\end{aligned} \quad (\text{A14})$$

for  $n = 0, \dots, S$  and  $\alpha_i = \frac{\gamma lbT}{\lambda}(S-n+i)$ . Here we have used Equation  $(\star)$  and the substitution  $v = \frac{\lambda}{lbT(S-n+i)} \omega$ . Equations (A13) and (A14) have simplified forms when some of the parameters are Erlang-distributed. For example, if both  $k, r \in \mathbb{N}$ , we have

$$\begin{aligned}
\mathbb{E}[P_{infection}] &= \int_0^\infty \int_0^\infty \left(1 - e^{-\frac{lbqt}{\omega}}\right) f_q(q) f_Q(\omega) dq d\omega, \\
&= 1 - \frac{\lambda^k \gamma^r}{\Gamma(r)} \int_0^\infty \frac{\omega^{r-1} e^{-\gamma \omega}}{(\gamma \omega + \frac{\gamma lbT}{\lambda})^k \left(\frac{\lambda}{\gamma \omega}\right)^k} d\omega, \quad (\text{using Eq. } (\star)) \\
&= 1 - \frac{e^{\frac{\gamma lbT}{\lambda}}}{\Gamma(r)} \int_{\frac{\gamma lbT}{\lambda}}^\infty \left(w - \frac{\gamma lbT}{\lambda}\right)^{r+k-1} w^{-k} e^{-w} dw \\
&= 1 - \frac{e^{\frac{\gamma lbT}{\lambda}}}{\Gamma(r)} \sum_{i=0}^{r+k-1} \binom{r+k-1}{i} (-1)^i \left(\frac{\gamma lbT}{\lambda}\right)^i \int_{\frac{\gamma lbT}{\lambda}}^\infty w^{r-1-i} e^{-w} dw \\
&= 1 - \frac{e^{\frac{\gamma lbT}{\lambda}}}{(r-1)!} \sum_{i=0}^{r+k-1} \binom{r+k-1}{i} (-1)^i \left(\frac{\gamma lbT}{\lambda}\right)^i \Gamma\left(r-i, \frac{\gamma lbT}{\lambda}\right)
\end{aligned}$$

which uses the substitution  $w = \gamma\omega + \frac{\gamma lbT}{\lambda}$  and Equation (††). The probability of observing exactly  $n$  exposures simplifies to

$$\begin{aligned} \mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty \int_0^\infty e^{-(S-n+i)\frac{lbT}{\omega}} f_q(q) f_Q(\omega) dy d\omega \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty \frac{\lambda^k \gamma^r}{\Gamma(k)\Gamma(r)} \omega^{r-1} e^{-\gamma\omega} \\ &\quad \times \frac{\Gamma(k)}{(\lambda + \frac{lbT}{\omega}(S-n+i))^k} d\omega \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{\lambda^k \gamma^r}{\Gamma(r)} \int_0^\infty \omega^{r-1} (\gamma\omega + \frac{\gamma lbT}{\lambda}(S-n+i))^{-k} \\ &\quad \times (\gamma\omega)^k \lambda^{-k} e^{-\gamma\omega} d\omega \\ &= \binom{S}{n} \frac{1}{(r-1)!} \sum_{i=0}^n \sum_{j=0}^{r+k-1} (-1)^{i+j} \binom{n}{i} \binom{r+k-1}{j} \alpha_i^j e^{\alpha_i} \\ &\quad \times \int_{\alpha_i}^\infty w^{r-1-j} e^{-w} dw \\ &= \binom{S}{n} \frac{1}{(r-1)!} \sum_{i=0}^n \sum_{j=0}^{r+k-1} (-1)^{i+j} \binom{n}{i} \binom{r+k-1}{j} \alpha_i^j e^{\alpha_i} \Gamma(r-j, \alpha_i), \end{aligned}$$

for  $n = 0, \dots, S$ , where we used the substitution  $w = \gamma\omega + \alpha_i$  and Equation (††).

## A4 POPULATION HETEROGENEITY

When considering a random quanta generation rate  $q \sim \text{Gamma}(k, \lambda)$ , it is important to note that if  $l > 1$  and one still considers  $ql$  in the corresponding analysis, then the main assumption is that all infectors have the same quanta generation rate  $q$  (which is sampled *once* from the corresponding distribution  $\text{Gamma}(k, \lambda)$ ). This would be appropriate for situations where one wants to consider homogeneous infectiousness across infectors but uncertainty in the precise value of  $q$ . For situations where  $l > 1$ , and if the aim is to consider heterogeneity in infectiousness, it is more appropriate to consider that each individual  $i \in \{1, 2, \dots, l\}$  has quanta generation rate  $q_i \sim \text{Gamma}(k, \lambda)$ , where this Gamma distribution represents population-level heterogeneity in infectiousness. In this situation, the total quanta emission rate is given by the sum of rates from each individual:  $q_{\text{Total}} = \sum_{i=1}^l q_i$ . Since the sum of  $l$  independent Gamma distributions with shape  $k$  and rate  $\lambda$  is itself a Gamma distribution of shape  $kl$  and rate  $\lambda$ , our results can be easily generalised to this scenario; that is,  $q_{\text{Total}} \sim \text{Gamma}(lk, \lambda)$  and we consider  $P_{\text{infection}} = 1 - e^{-\frac{bT}{Q} q_{\text{Total}}}$ . In particular, one has

$$g(q_{\text{Total}}) = -\frac{bT}{Q} q_{\text{Total}}, \quad f_{q_{\text{Total}}}(q) = \frac{\lambda^{lk} q^{lk-1} e^{-\lambda q}}{\Gamma(lk)}.$$

Since  $P_{\text{infection}} = 1 - e^{-\frac{bT}{Q} q_{\text{Total}}}$  we have  $q_{\text{Total}} = \frac{-\ln(1-P_{\text{infection}})Q}{bT}$  and

$$\frac{dq_{\text{Total}}}{dP_{\text{infection}}} = \frac{Q}{(1-P_{\text{infection}})bT}.$$

Thus, Equation (A1) results in the density function of  $P_{\text{infection}}$

$$\begin{aligned} f_{P_{\text{infection}}}(p) &= \frac{\lambda^{lk} \left[ \frac{-\ln(1-p)Q}{bT} \right]^{lk-1} e^{-\lambda \left[ \frac{-\ln(1-p)Q}{bT} \right]}}{\Gamma(lk)} \frac{Q}{(1-p)bT} \\ &= \left( \frac{\lambda Q}{bT} \right)^{lk} \frac{(-\ln(1-p))^{lk-1} (1-p)^{\frac{\lambda Q}{bT}-1}}{\Gamma(lk)}, \quad p \in (0, 1). \end{aligned} \quad (\text{A15})$$

Using Equation (A2), we find the expected per-capita infection risk

$$\begin{aligned} \mathbb{E}[P_{\text{infection}}] &= \int_0^\infty \left( 1 - e^{-\frac{bT}{Q} q} \right) \frac{\lambda^{lk} q^{lk-1} e^{-\lambda q}}{\Gamma(lk)} dq \\ &= 1 - \frac{\lambda^{lk}}{\Gamma(lk)} \int_0^\infty e^{-q(\frac{bT}{Q} + \lambda)} q^{lk-1} dq = 1 - \left( \frac{\lambda}{\frac{bT}{Q} + \lambda} \right)^{lk}, \end{aligned} \quad (\text{A16})$$

where we have used Equation (\*). The probability of observing exactly  $n$  exposures is computed using Equation (A3) leading to

$$\begin{aligned} \mathbb{P}(E = n) &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \int_0^\infty e^{-q((S-n+i)\frac{bT}{Q} + \lambda)} q^{lk-1} \frac{\lambda^{lk}}{\Gamma(lk)} dq \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \frac{\Gamma(lk)}{\left( (S-n+i)\frac{bT}{Q} + \lambda \right)^{lk}} \frac{\lambda^{lk}}{\Gamma(lk)} \\ &= \binom{S}{n} \sum_{i=0}^n (-1)^i \binom{n}{i} \left( \frac{\lambda}{(S-n+i)\frac{bT}{Q} + \lambda} \right)^{lk}, \end{aligned} \quad (\text{A17})$$

for  $n = 0, \dots, S$ .

Due to the symmetry in the product in the equation for the classical Wells-Riley model, these results are analogous to the case of sequential visits, such that we have a sum of  $M$  independent randomly distributed exposure times,  $T_{\text{Total}} = \sum_{i=1}^M T_i$ , instead of a sum of quanta emission rates. This is the approach used in Section 3.2.1 and Figure 8 in particular.

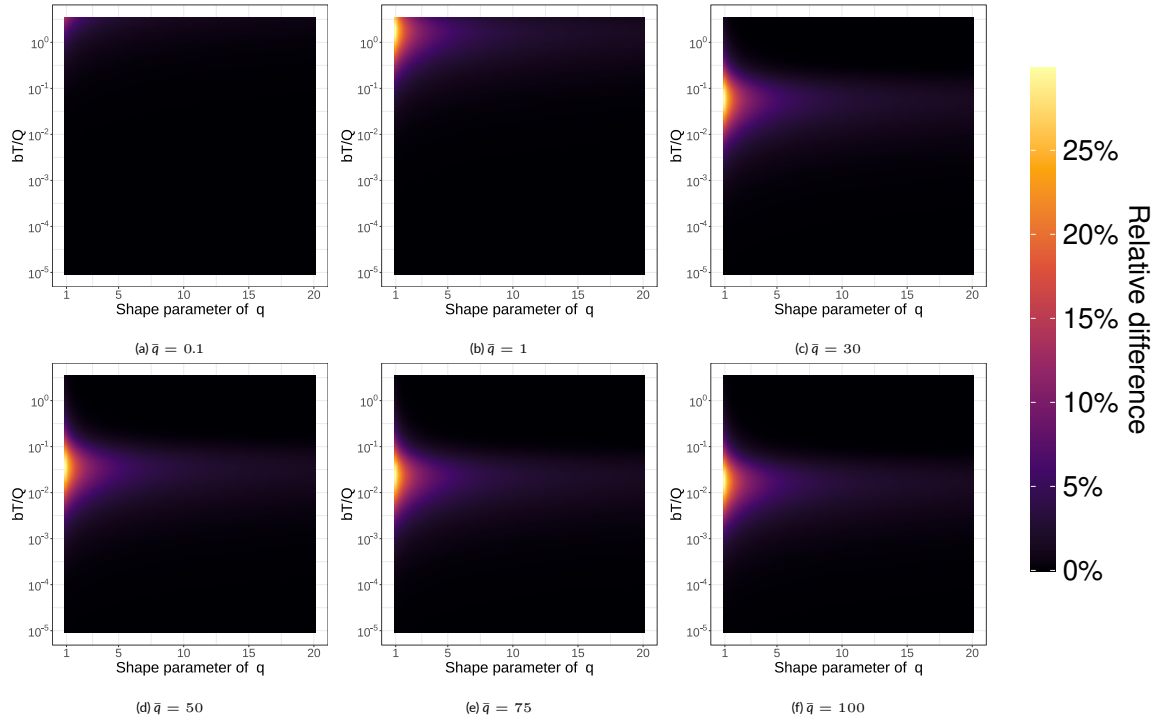
## A5 COMPARING TO THE CLASSICAL WELLS-RILEY APPROACH

Here, we aim to compare the “true” average risk  $\mathbb{E}[P_{\text{infection}}]$  estimated via our approach with the per capita infection risk estimated using the classical Wells-Riley methodology with average parameter value instead,  $\bar{P}_{\text{infection}}^{\text{WR}}$ .

### A5.1 Random population parameter

For a single random population parameter, one can estimate the average per capita infection risk  $\mathbb{E}[P_{\text{infection}}]$  using Equation (3). It is of interest to compare this “true” mean risk value with that obtained via the classical Wells-Riley approach in Equation (1) when using an average quanta emission rate  $\bar{q} = k/\lambda$ . In Section 2.1.1, we denote this by

$$\bar{P}_{\text{infection}}^{\text{WR}} = 1 - e^{-\frac{bT}{Q} \bar{q}}.$$



**FIGURE A1** Heatmaps for the relative difference between  $\bar{P}_{infection}^{WR}$  (classical Wells-Riley approach with constant, average quanta generation rate  $\bar{q} = k/\lambda$ ) and the analytical mean infection risk  $\mathbb{E}[P_{infection}]$  in Equation 3, versus  $k$ , for  $\bar{q} = 0.1, 1, 30, 50, 75, 100$ .

It is of particular interest to analyse the difference between the resulting average infection risk of each approach. We can denote this difference by

$$D(A, k, \lambda) = \bar{P}_{infection}^{WR} - \mathbb{E}[P_{infection}] = \left( \frac{\lambda}{\lambda + A} \right)^k - e^{-A \frac{k}{\lambda}}, \quad (\text{A18})$$

where  $A$  represents the combination of non-random parameters, such that  $g(X) = -\frac{bqT}{Q} = -AX$  for some random parameter  $X \in \{l, b, q, T\}$  (for example  $A = \frac{bT}{Q}$  for random  $q$ ), and  $(k, \lambda)$  represent the shape and rate parameters of the Gamma-distributed random parameter. Depending on the sign of  $D(A, k, \lambda)$ , we can make the following conclusions:

- $D(A, k, \lambda) > 0$ : The classical approach with average parameter value is overestimating the “true” mean infection risk.
- $D(A, k, \lambda) = 0$  (or  $\approx 0$ ): The classical approach with average parameter value is capturing the “true” mean infection risk.
- $D(A, k, \lambda) < 0$ : The classical approach with average parameter value is underestimating the “true” mean infection risk.

In the case of random  $b$ ,  $q$  or  $T$  one can use Jensen’s inequality to show that the classical approach ( $\bar{P}_{infection}^{WR} = 1 - e^{-A\bar{X}}$ ) will always overestimate the true mean infection risk,  $\mathbb{E}[P_{infection}]$ . Jensen’s inequality (Jensen (1906), McShane (1937)) states that, for a strictly concave function  $P(q)$  (sometimes referred to as “concave-down”) and a non-constant random variable  $q$ , the following inequality holds

$$\mathbb{E}[P(q)] < P(\mathbb{E}[q]).$$

The per-capita infection probability  $P_{infection}(q) = 1 - e^{-Aq}$  is a concave function of  $q$  since the second derivative is always negative,

$$\frac{d^2 P_{infection}}{dq^2} = -A^2 e^{-Aq} < 0.$$

So, we recognise  $P_{infection}(\mathbb{E}[q]) = \bar{P}_{infection}^{WR}$ , and  $\mathbb{E}[P(q)] = \mathbb{E}[P_{infection}]$ , to conclude that

$$D(A, k, \lambda) = \bar{P}_{infection}^{WR} - \mathbb{E}[P_{infection}] > 0$$

holds for all  $A > 0$  and  $k, \lambda$ . Furthermore, since  $P_{infection}$  is always concave this result holds for generally distributed  $b$ ,  $q$  or  $T$ . So, given  $A > 0$  and randomly distributed  $b$ ,  $q$  or  $T$ , the classical Wells-Riley approach (which uses the mean value of the random parameter) will always overestimate the true mean infection risk,  $\mathbb{E}[P_{infection}]$ .

If we fix  $A$  and the mean  $\bar{q}$ , then we can write  $D(A, k, \lambda) = D(k)$ , since  $\lambda = k/\bar{q}$  (now depends on  $k$ ). Then

$$D(k) = \bar{P}_{infection}^{WR} - \mathbb{E}[P_{infection}] = \left( \frac{k}{k + A\bar{q}} \right)^k - e^{-A\bar{q}}.$$

By taking the derivative with respect to  $k$ , we can find the value  $k_{max}$  at which we have the largest overestimation:

$$\frac{dD(k)}{dk} = \left( \frac{k}{k + A\bar{q}} \right)^k \left[ \ln \left( \frac{k}{k + A\bar{q}} \right) + 1 - \frac{k}{k + A\bar{q}} \right], \quad (\text{A19})$$

for which no finite  $k$  satisfies  $\frac{dD(k)}{dk} = 0$  (since we would need  $\frac{k}{k + A\bar{q}} = 1$  which holds in the limit of  $k \rightarrow \infty$ ). We also find that  $\frac{dD(k)}{dk} < 0$  for all finite  $k > 0$ , meaning that  $D(k)$  is monotonically decreasing. This means

that for fixed  $A$ , and fixed mean  $\bar{q}$ , the largest overestimation appears when  $k \rightarrow 0$  (i.e. for small values of the shape parameter in the Gamma distribution for  $q$ ). We can determine the maximum overestimation for any fixed mean  $\bar{q} = k/\lambda$ . Using a logarithmic argument, one can find  $\lim_{k \rightarrow 0^+} D(k) = 1 - e^{-A\bar{q}} = \bar{P}_{infection}^{WR}$ . This means that as the mean  $\bar{q}$  and  $A$  increase (decrease), the maximum possible overestimation increases (decreases), and this maximum occurs in the limit  $k \rightarrow 0$  (implying  $\lambda \rightarrow 0$  since  $\bar{q}$  is fixed). As  $k$  and  $\lambda$  both decrease such that the mean,  $\bar{q} = k/\lambda$ , is fixed, two things happen:

1. The variance,  $\frac{k}{\lambda^2} = \frac{\bar{q}}{\lambda}$ , of the random parameter increases.
2. The skewness,  $\frac{2}{\sqrt{k}}$ , increases.

We can conclude that large variance and positive skewness of a random population parameter leads to larger overestimations of the density mean by the classical Wells-Riley model (with mean parameter value), where all the other parameters are fixed. This could suggest that inaccuracies are caused by the fact that considering the mean value of our random population parameter neglects small values (with non-negligible probabilities) which significantly impact infection risk, particularly in the case of a high variance and positively skewed Gamma distribution. When we have a distribution that is highly concentrated around the mean with little skewness (when  $k$  is very high), then the overestimation becomes small (one can show that  $\lim_{k \rightarrow \infty} D(k) = 0$ ). It should be noted that the maximum overestimation is affected by the value of  $A$ , if  $A \ll 1$  ( $A \gg 1$ ), then the density mean,  $\mathbb{E}[P_{infection}]$ , and the classical Wells-Riley,  $\bar{P}_{infection}^{WR}$ , will both approach zero (one), where their difference disappears.

Let us consider some numerical experiments to explore this further. We focus here on the parameter  $q$  but a similar analysis would apply to  $T$  or  $b$  if they were Gamma-distributed. We consider the quanta generation rate to follow a Gamma distribution

$$q \sim \text{Gamma}(k, \lambda).$$

The aim is to explore how different combinations of  $(k, \lambda)$  impact on the per capita infection risk and the error of the classical approach relative to the analytical mean infection risk:  $D(A, k, \lambda)/\mathbb{E}[P_{infection}]$ . To this end, we uniformly sample  $b$ ,  $T$  and  $Q$ , as well as the shape,  $k$ , and rate,  $\lambda$  parameters of the Gamma distribution describing  $q$ . We use wide intervals that produce plausible values for the Wells-Riley parameters and a wide range of variances for  $q$ . In particular, we choose  $\bar{q} \in [0.01, 100]$  quanta  $h^{-1}$  and  $k \in [1, 20]$  (such that  $\lambda = k/\bar{q}$ ). Likewise, for the fixed parameters (sampled here but treated as fixed after the fact, as in Equation (3)) we consider:  $b \in [0.2, 0.6]$   $m^3 h^{-1}$ ,  $T \in [5 \text{ minutes}, 100 \text{ hours}]$ , and  $Q \in [10, 1000]$   $m^3 h^{-1}$  (which represents  $ACH \in [0.1, 100]$  in a typical ward (Edwards 2024b)), such that  $l = 1$ . These samples combine to make pairs of  $A = bT/Q$  and  $\bar{q} = k/\lambda$ , for which we plot the corresponding relative errors:  $D(A, k, \lambda)/\mathbb{E}[P_{infection}]$ , in Figure 1 within Section 2.1.1. We noticed that overestimations were observed for a 'strip' of values  $(\bar{q}, bT/Q)$  within the upper right region of Figure 1. Furthermore, we noticed variations points that were adjacent

to one another, suggesting that the size of the overestimations do not depend on  $\bar{q}$  and  $bT/Q$  alone. To investigate this, we plotted the relative difference whilst varying  $k$ , the shape parameter of Gamma-distributed  $q$ , and the value of  $bT/Q$  for different mean quanta emission rates  $\bar{q}$  in Figure A1. By observing the resulting pattern in the relative difference, we can see that (i) the size of the interval of  $bT/Q$  is about one order of magnitude on the log scale, and its position is directly dependent on the order of  $\bar{q}$  (ii) within this region where overestimations occur, the largest are observed when  $k$  is small (1 here); this agrees with our earlier intuition about large variation and positive skew of  $q$  leading to systematic inaccuracies. The dark regions in Figure A1 (indicating little to no relative difference) are those where the infection risk is either close to zero/one (where  $bT/Q$  is small/large) or where the shape parameter is so high that there is little uncertainty in the quanta emission rate  $q$  around its mean value  $\bar{q}$ . We identify the following criteria which lead to overestimations by the classical WR approach ( $\bar{P}_{infection}^{WR}$ ) of the analytical mean infection risk ( $\mathbb{E}[P_{infection}]$ )

1. A mean parameter value and  $A$  value ( $\bar{q} = k/\lambda$  and  $bT/Q$  in this case) that is neither too small nor too large, so as to avoid saturation of the infection risk at either 0 or 1.
2. A large variance and positive skewness for the distribution of the random parameter, such that smaller values (which now have sufficiently high probabilities of occurring) are neglected when considering the mean parameter value.

## A5.2 Random ventilation rate

In the case of a random ventilation rate, the classical Wells-Riley approach is now given by

$$\bar{P}_{infection}^{WR} = 1 - e^{-\frac{lbqT}{Q}},$$

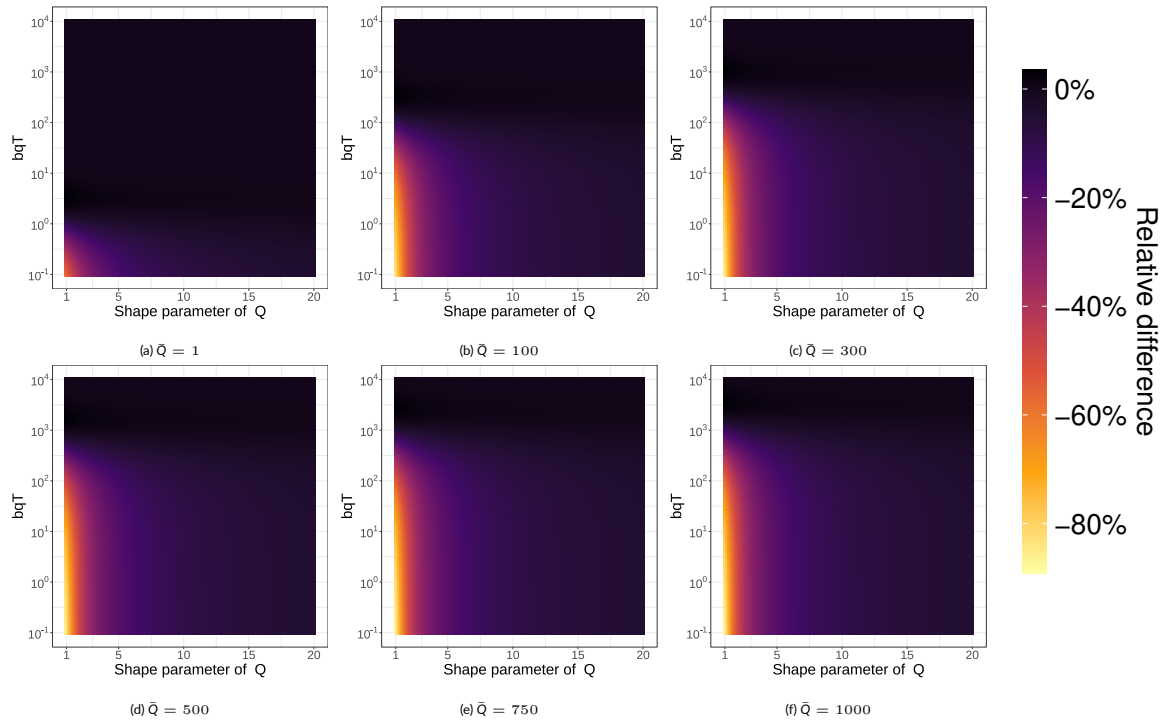
as in Section 2.2, where  $Q \sim \text{Gamma}(r, \gamma)$ , meaning  $\bar{Q} = \frac{r}{\gamma}$ . As such, the difference between the "true" mean infection risk and the mean calculated by the classical Wells-Riley approach with average ventilation rate, is given by

$$D(A, r, \gamma) = \frac{2}{\Gamma(r)} (\gamma A)^{r/2} K_r \left( 2\sqrt{\gamma A} \right) - e^{-A\frac{r}{\gamma}},$$

such that  $A = lbqT$ . The function  $P_{infection}(Q) = 1 - e^{-\frac{A}{Q}}$  is neither totally concave or convex. This means that we cannot explicitly apply Jensen's inequality here. However, we can find the inflection point of  $P_{infection}$ . The second derivative of  $P_{infection}$  with respect to  $Q$  is

$$\frac{d^2 P_{infection}(Q)}{dQ^2} = \frac{A}{Q^3} e^{-\frac{A}{Q}} \left( 2 - \frac{A}{Q} \right).$$

Setting this to zero and solving for  $Q$  tells us that the inflection point is  $Q^* = A/2$ . For  $Q \in (0, Q^*)$ , we have  $\frac{d^2 P_{infection}}{dQ^2} < 0$ , and for  $Q \in (Q^*, \infty)$  the opposite is true  $\frac{d^2 P_{infection}}{dQ^2} > 0$ . This means that there are two distinct regions:



**FIGURE A2** Heatmaps for the relative difference between  $\bar{P}_{infection}^{WR}$  (classical Wells-Riley approach with constant, average ventilation rate  $\bar{Q} = r/\gamma$ ) and the analytical mean infection risk  $\mathbb{E}[P_{infection}]$  in Equation 12, versus shape parameter  $r$ , for  $\bar{Q} = 1, 100, 300, 500, 750, 1000$ .

- Concave region:  $(0, Q^*)$ .
- Convex region:  $(Q^*, \infty)$ .

If the support for  $Q$  is contained exclusively within the convex (concave) region, then we can conclude, by Jensen's inequality, that the classical Wells-Riley approach would underestimate (overestimate) the true mean infection risk, given by  $\mathbb{E}[P_{infection}]$ .

Using  $b \in [0.2, 0.6] m^3 h^{-1}$ ,  $T \in [5 \text{ minutes}, 100 \text{ hours}]$  (as in Appendix A5.2),  $q \in [0.01, 100] \text{ quanta } h^{-1}$  (to represent values considered in Section 3), along with  $\bar{Q} \in [10, 1000] m^3 h^{-1}$  (to represent  $ACH \in [0.1, 10]$  for a typical ward (Edwards 2024b)) and  $r \in [1, 20]$  (such that  $\gamma = r/\bar{Q}$ ), we now numerically investigate possible inaccuracies in the case of random ventilation. Combining these samples, we obtain pairs of  $A = bqT$  and  $\bar{ACH} = \frac{\bar{Q}}{100}$  for which we plot the error of the classical Wells-Riley approach ( $\bar{P}_{infection}^{WR}$ ) relative to the analytical mean infection risk ( $\mathbb{E}[P_{infection}]$ ) in Figure 2 within Section 2.2. First, we notice that, unlike Figure 1, the range of inaccuracies spans both positive and negative values relating to both overestimations and underestimations (with overestimations being very small). For the samples that we generated, the overestimations seem to only occur for very large  $A = bqT$  values. Mathematically, we know that this can be caused by the entirety (or large enough portion) of the distribution for  $Q$  existing in the concave region of  $P_{infection}$ ,  $(0, Q^*)$  where  $Q^* = A/2$ . Intuitively, these overestimations, albeit small, may be due to the fact that using the mean value of  $Q$  in place of its distribution neglects higher ventilation rates. It is possible that when  $A = bqT$  is large enough, using the mean ventilation

rate results in saturation of  $P_{infection}$  near 1, but larger and statistically significant ventilation rates would outweigh this saturation if they were to occur. Therefore, in this case, we are neglecting scenarios in which the infection risk does not saturate at 1. This overestimation is likely smaller in magnitude compared to the possible underestimations under different parameter regimes because the Gamma distribution is positively skewed towards smaller values, so the effect of neglecting larger values is dampened as they have smaller probabilities of occurring. Even with a positively skewed distribution, there are scenarios in which the mean does not adequately capture the larger ventilation rates that result in smaller infection risks.

As with the case for random  $q$  (see Appendix A5.1), we observe local fluctuations of the relative difference where underestimations are observed in Figure 2, suggesting that  $\bar{ACH}$  and  $bqT$  alone do not determine the degree of underestimation. We again plot the relative difference for varying  $bqT$ , shape parameter of the ventilation rate,  $r$ , and for different mean ventilation rates  $\bar{Q}$  in Figure A2. First, we notice that a lower bound for the region of underestimations (non-black colour) is not observed for the parameter values considered. However, we know that a lower bound exists by using a limit approach which confirms  $\lim_{bqT \rightarrow 0} (\bar{P}_{infection}^{WR}) = \lim_{bqT \rightarrow 0} (\mathbb{E}[P_{infection}])$  (implying that  $D(A, r, \gamma) \rightarrow 0$ ). So, at least for these parameter values, the region at which we observe relative difference is magnitudes larger in the case of random ventilation than it is for random quanta. Furthermore, the degree of this relative difference is also much larger in the random ventilation case (over 80%), and occurs also for small values of the shape parameter,  $r$ ,

corresponding to high variation and positive skew in  $Q$ . We see examples of this when comparing the errors seen in Figures 7, 8, 9 to those in Figures 10a and 10c. Another example is the error produced by only considering random time (green curves) compared to only considering random ventilation (blue curves) for high variance ventilation in Figure 11. The reason for the higher degree of relative difference for random ventilation versus random quanta likely has to do with how sensitive the Wells-Riley model is to their respective fluctuations. That is, if  $q$  is random, then the classical approach does not vary as strongly in the region of  $\bar{q}$ , as it would in the region of  $\bar{Q}$  in the case of random  $Q$ . When  $\bar{Q}$  is random, the possibility of small removal rates causes the infection risk to increase notably, even sometimes to close to 1 (as observed in Figure 10b), whereas the same effect being achieved in the case of random quanta requires the possibility of very large  $q$ , which is not as likely in the case of a positively skewed distribution. As with the case for random quanta, when ventilation is random the following criteria are identified which are required for overestimations to occur

1. A mean value,  $\bar{Q}$ , that 'balances' the value of the other parameters,  $A = bqT$ , is necessary so as to avoid saturation of  $P_{infection}$  at 0 or 1.
2. A large enough variance and positive skewness for fixed mean  $\bar{Q}$ , such that  $r$  and  $\gamma$  are small, is necessary by which small values leading to significantly high infection risks are neglected.

## A6 IMPACT OF TWO RANDOM PARAMETERS ON INFECTION RISK: UNCERTAINTY DOMINANCE

In this section, the aim is to understand how simultaneous uncertainty in two different parameters may affect infection risk. In particular, there might be situations where even if both parameters are random, there is one that clearly dominates the impact on the per capita infection risk probability  $P_{infection}$ . Let us consider two random parameters  $X, Y \in \{b, q, T, Q, I\}$  and denote by  $P_{infection}^{X,Y}$  the per capita probability of infection when both parameters are considered to be random simultaneously. On the other hand, let us denote by  $P_{infection}^X$  the per capita probability of infection when  $X$  is considered to be random, but parameter  $Y$  is considered constant and set instead to its mean value  $\bar{Y}$ .

Using arguments from Section 2, we can compute the density functions of  $P_{infection}^{X,Y}$  and  $P_{infection}^X$ , denoted by  $f_{infection}^{P^{X,Y}}(p)$  and  $f_{infection}^{P^X}(p)$  respectively. Results in Section 3.3.1 (Figure 11) suggest that high variance and skewness in the distribution of the random parameters lead to larger propagation of uncertainty into the corresponding infection risk probability. More specifically, if the variance (and skewness) of one random parameter,  $X$ , is larger than that of another,  $Y$ , we expect this parameter to dominate the propagation of uncertainty into  $P_{infection}^{X,Y}$  and, thus,  $f_{infection}^{P^{X,Y}}(p)$ . This would mean that the density functions  $f_{infection}^{P^{X,Y}}(p)$  and  $f_{infection}^{P^X}(p)$  are more similar to each other, compared to  $f_{infection}^{P^Y}(p)$  (i.e. most of the uncertainty in infection risk comes from uncertainty in  $X$ ,

and not in  $Y$ ). To quantify this, we can make use of the Hellinger distance (Aggarwal and Reddy 2014). The Hellinger distance measures the difference between two distributions,  $f_X(x)$  and  $g_X(x)$  as

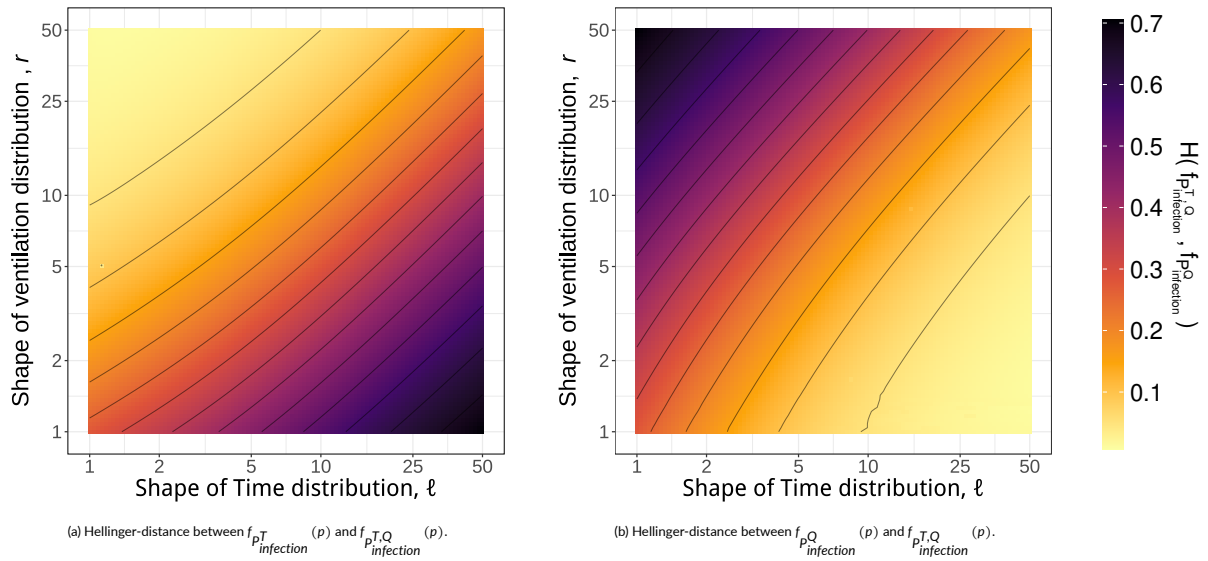
$$H^2(f_X, g_X) = \frac{1}{2} \int \left( \sqrt{f_X(x)} - \sqrt{g_X(x)} \right)^2 dx = 1 - \int \sqrt{f_X(x)g_X(x)} dx.$$

This gives a number between 0 and 1 where values closer to 0 represent more similarity between the distributions, whereas values close to 1 correspond to more different distributions.

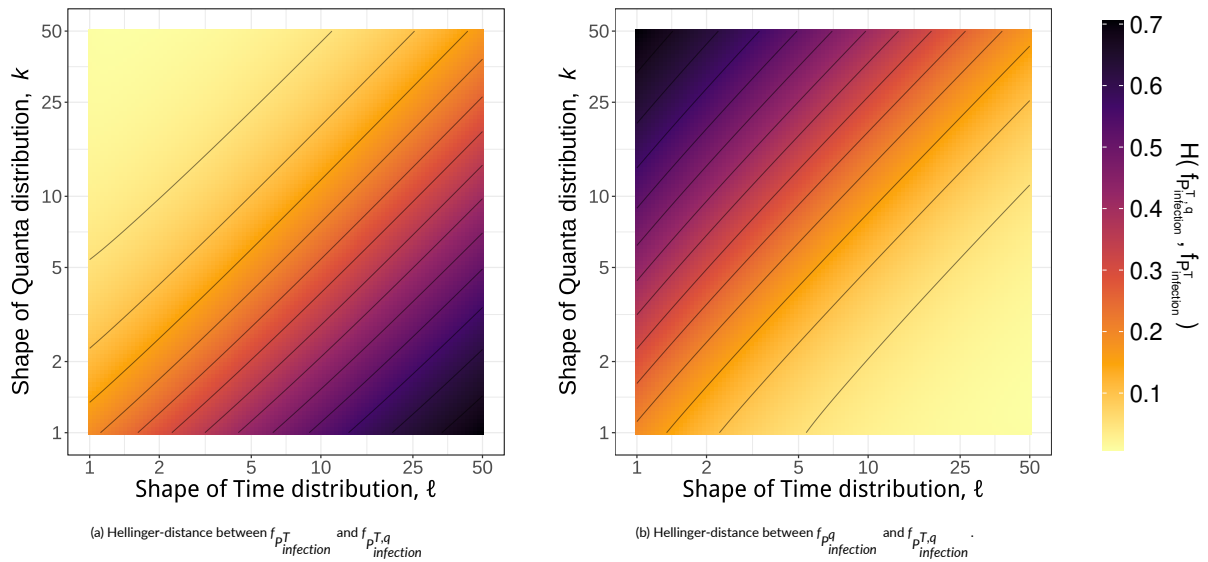
We focus here on pairs of parameters  $(q, T)$  and  $(T, Q)$  for illustrative purposes, with distributions  $q \sim \text{Gamma}(k, \lambda)$ ,  $T \sim \text{Gamma}(\ell, \mu)$  and  $Q \sim \text{Gamma}(r, \gamma)$  (when considered to be random). The aim is to study how randomness in these parameters translates into uncertainty in infection risk. In Figure A3 we plot the Hellinger distance between  $f_{infection}^{P^{T,Q}}(p)$  and  $f_{infection}^{P^T}(p)$  and between  $f_{infection}^{P^{T,Q}}(p)$  and  $f_{infection}^{P^Q}(p)$ . In Figure A4 we plot the Hellinger distance between  $f_{infection}^{P^{T,q}}(p)$  and  $f_{infection}^{P^T}(p)$  and between  $f_{infection}^{P^{T,q}}(p)$  and  $f_{infection}^{P^q}(p)$ . Both figures use the parameter regimes presented in Section 3.3.1. In both figures, for each Gamma-distributed random parameter we set the mean parameter value to a constant while changing the shape parameter of their Gamma distribution (and accordingly changing the rate parameter so that the mean parameter value remains the same). Lighter areas in these plots indicate smaller Hellinger distances, indicating that the particular parameter being looked at ( $T$  in Figures A3a and A4a,  $q$  in Figure A4b and  $Q$  in Figure A3b) is the one that dominates the propagation of uncertainty into the per capita infection risk. Within each plot, in Figures A3 and A4, the patterns support the idea that a smaller shape (therefore an increased variance and positive skewness) for one random parameter, and a larger shape for another random parameter, leads to dominance on the impact on infection risk by the former over the latter.

In the case that one parameter has a small shape value, linearly increasing the shape of the other random parameter diminishingly increases the likeness of the first parameter's corresponding infection risk density to that of the case that both parameters are random (shown by a diminishingly decreasing Hellinger distance). For example, we see that in the case of random  $T$  and  $Q$  (Figure A3), when the shape of  $T$  is  $\ell = 2$ , there is not much difference in the similarity of  $f_{infection}^{P^T}(p)$  (where the ventilation rate is considered constant and set to  $\bar{Q} = r/\gamma$ ) to  $f_{infection}^{P^{T,Q}}(p)$  when the shape of  $Q$  varies between  $r = 25$  and  $r = 50$ , indicated by unchanging colour and lack of contour lines in that region of Figure A3a. On the other hand, as  $r$  increases, we see a noticeable increase in the Hellinger distance,  $H(f_{infection}^{P^{T,Q}}(p), f_{infection}^{P^Q}(p))$ , in Figure A3b (indicated by darker regions). A similar behaviour can be seen when comparing Figures A4a and A4b. That is, for a small shape of  $T$  (for example,  $\ell = 2$ ), there is a negligible decrease in the Hellinger distance. One should expect that, when the shape of one parameter, say  $X$ , becomes increasingly large then the infection risk density corresponding to the other parameter,  $f_{infection}^{P^Y}(p)$ , should tend towards the distribution in which both are random,  $f_{infection}^{P^{X,Y}}(p)$ . This is because, whilst technically still a Gamma-distributed random parameter,  $X$  effectively becomes constant (random with asymptotically small variance).

The ranges of Hellinger distances are [0.012, 0.704] for Figure A3 and



**FIGURE A3** Heatmaps of the Hellinger distance between the density functions of the per capita infection risk probability when two parameters are random ( $f_{p_{infection}^{T,Q}}(p)$ ) vs. when only one is random ( $f_{p_{infection}^{T,Q}}(p)$  or  $f_{p_{infection}^{Q}}(p)$ ) and the other is set to its mean value instead ( $\bar{Q} = r/\gamma$  or  $\bar{T} = \ell/\mu$ ). All default parameter values are taken from the hospitality scenario (Section 3.3.1).



**FIGURE A4** Heatmaps of the Hellinger distance between the density functions of the per capita infection risk probability when two parameters are random ( $f_{p_{infection}^{T,q}}(p)$ ) vs. when only one is random ( $f_{p_{infection}^{T,q}}(p)$  or  $f_{p_{infection}^{q}}(p)$ ) and the other is set to its mean value instead ( $\bar{q} = k/\lambda$  or  $\bar{T} = \ell/\mu$ ). All default parameter values are taken from the hospitality scenario (Section 3.3.1).

[0.009, 0.704] for Figure A4. The minute difference in the lower bounds could be due to the fact that having a Gamma-distributed ventilation rate can lead to scenarios in which the rate is close to zero with positive probability, resulting in higher densities of high infection risk (as seen in Figures 10a and 10c), whereas this is not the case with Gamma-distributed  $T$  or  $q$  where the same effect would require  $T$  or  $q$  to be significantly large.