bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The Basic Reproductive Number for Disease Systems with Multiple Coupled Heterogeneities

Alun L. Lloyd[a,1,*], Uriel Kitron[b], T. Alex Perkins[c], Gonzalo M. Vazquez-Prokopec[b], Lance A. Waller[d]

[a] *Department of Mathematics and Biomathematics Graduate Program, North Carolina State University, Raleigh NC 27695, USA*
[b] *Department of Environmental Sciences, Emory University, Atlanta, GA 30322, USA*
[c] *Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, USA*
[d] *Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA*

Abstract

In mathematical epidemiology, a well-known formula describes the impact of heterogeneity on the basic reproductive number for situations in which transmission is separable and for which there is one source of variation in susceptibility and one source of variation in infectiousness. This formula is written in terms of the magnitudes of the heterogeneities, as quantified by their coefficients of variation, and the correlation between them. A natural question to ask is whether analogous results apply when there are multiple sources of variation in susceptibility and/or infectiousness. In this paper we demonstrate that under three or more coupled heterogeneities, the basic reproductive number depends on details of the distribution of the heterogeneities in a way that is not seen in the well-known simpler situation. We provide explicit results for the cases of multivariate normal and multivariate log-normal distributions, showing that the basic reproductive number can again be expressed in terms of the magnitudes of the heterogeneities and the pairwise correlations between them. The results, however, differ between the two multivariate distributions, demonstrating that no formula of this type applies generally when there are three or more

*Corresponding author; email alun_lloyd@ncsu.edu
[1] Keywords: Disease transmission model, heterogeneity, basic reproductive number, coupled heterogeneities

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

coupled heterogeneities. We see that the results are approximately equal when heterogeneities are relatively small and show that an earlier result in the literature (Koella, 1991) should be viewed in this light. We provide numerical illustrations of our results.

1. Introduction

The basic reproductive number, R_0, plays a crucial role in determining both whether a pathogen is able to spread and the strength of control measures needed to halt its spread. The simplest descriptions of R_0 assume simple transmission scenarios, such as perfect mixing of a population and homogeneity of the individuals in the population, e.g. in terms of their susceptibility and infectiousness. The inadequacies of such descriptions have long been realized and much attention has been directed towards understanding the impact of heterogeneities in transmission on the basic reproductive number. Early efforts included accounting for differing activity levels amongst the population and various mixing patterns of the population (e.g. proportionate/random mixing, assortative and disassortative mixing). Much of this work was prompted by the heterogeneities known to exist for the spread of sexually transmitted infections, notably gonorrhea and HIV (Nold (1980); Hethcote and Yorke (1984); Anderson et al. (1986); May and Anderson (1987); Jacquez et al. (1988); Gupta et al. (1989)). In the context of vector-borne diseases, it has long been realized that vectors' bites are not distributed uniformly across hosts; instead, there is a heterogeneity in hosts' attractiveness to vectors, with some individuals being disproportionately favored to receive bites (Carnevale et al. (1978); Dye and Hasibeder (1986); De Benedicitis et al. (2003); Liebman et al. (2014)).

A now standard argument shows that the basic reproductive number for a multi-type transmission system can be calculated as the dominant eigenvalue of the next generation matrix (Diekmann and Heesterbeek (2000)). For an n-type setting, the next generation matrix is an n by n non-negative matrix whose (i, j)'th entry gives the average number of secondary infections of type i caused

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

by a type j individual in an otherwise entirely susceptible population. Consequently, much attention has been directed towards those special cases of heterogeneous transmission that lead to next generation matrices whose dominant eigenvalue is analytically tractable and hence for which the basic reproductive number can be calculated explicitly. In the context of spatial heterogeneity, these include symmetric spatial configurations such as equally-sized patches with all-to-all or nearest neighbor contacts (see, for example Lloyd and May (1996)).

More generally, a commonly-studied situation involves separable transmission (Diekmann and Heesterbeek (2000)), where each group has a susceptibility, a_i and an infectiousness, b_i. In this case, the next generation matrix is of rank one, and, taking groups to be of equal sizes, has entries $a_i b_j/n$, and dominant eigenvalue

$$R_0 = \frac{1}{n}\sum_{i=1}^{n} a_i b_i. \tag{1}$$

Using the result

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) + \mathrm{Cov}(X,Y) \tag{2}$$

for the expectation of a product of random variables, eqn (1) can be rearranged into the following well-known formula (Dietz (1980); Dye and Hasibeder (1986)) that sheds insight into the impact of heterogeneity on R_0 in this separable setting:

$$\begin{aligned} R_0 &= \bar{a}\bar{b}\left(1 + r_{ab}\mathrm{CV}_a\mathrm{CV}_b\right) \\ &= R_0^{\mathrm{hom}}\left(1 + r_{ab}\mathrm{CV}_a\mathrm{CV}_b\right). \end{aligned} \tag{3}$$

Here, $\bar{a}$ and $\bar{b}$ denote the average values of a_i and b_i, r_{ab} denotes the Pearson product-moment correlation coefficient between the a_i and b_i, CV_a and CV_b denote the coefficients of variation (i.e. standard deviation divided by the mean) of a_i and b_i and R_0^{hom} denotes the value of R_0 that would be predicted if the heterogeneity was ignored, i.e. the average values of a_i and b_i were used. We

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

emphasize that these results are exact, holding for arbitrary distributions of the a_i and b_i.

Heterogeneity can inflate or deflate the value of R_0, depending on whether there is positive or negative correlation between susceptibility and infectivity across the groups (Dietz, 1980). In the special case where susceptibility and infectivity are proportional, e.g. for a situation such as differing activity levels or mosquito biting preferences where the heterogeneity impacts both susceptibility and infectiousness in the same way, the formula reduces to

$$\begin{aligned} R_0 &= R_0^{\text{hom}} \left(1 + \text{CV}_a^2\right) \\ &= R_0^{\text{hom}} \left(1 + \frac{\text{Var(a)}}{\bar{a}^2}\right). \end{aligned} \tag{4}$$

This formula has appeared in the literature numerous times in a number of different settings and guises (Dietz, 1980; Dye and Hasibeder, 1986; May and Anderson, 1987).

Particularly with the increasing realization that many systems are subject to multiple, often coupled, heterogeneities (Paull et al., 2012; Vazquez-Prokopec et al., 2016), an important question is whether results such as eqns. (3) and (4) generalize to situations in which there are more than two heterogeneities. In this paper, we show that the answer to this question is no: the effect of multiple interacting heterogeneities on the basic reproductive number depends on the details of the distributions of the heterogeneities, in contrast to what occurs in the two-heterogeneity setting. We provide results for both multivariate normal and multivariate log-normal distributions of heterogeneities and demonstrate that the two settings can give markedly different results.

2. Results

We assume that there are n types of individuals, resulting from N different heterogeneities, N_1 of which impact susceptibility and N_2 of which impact infectiousness. We further assume that the susceptibility of a type i individual

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

can be written as the product $x_i^1 x_i^2 \cdots x^{N_1}$, taken over the heterogeneities that impact susceptibility, and that the infectivity of a type j individual can similarly be written as $x_j^1 \cdots x_j^{N_2}$. Taking group sizes to be equal and assuming separable transmission, the entries of the next generation matrix will have the form $x_i^1 x_i^2 \ldots x_i^{N_1} x_j^1 \cdots x_j^{N_2}/n$, where $N_1 + N_2 = N$. This matrix is of rank one and has dominant eigenvalue given by

$$R_0 = \frac{1}{n} \sum_{i=1}^{n} x_i^1 x_i^2 \cdots x_i^N. \tag{5}$$

As explained above, the well-known result arises from the ability to express the expectation of the product of a pair of random variables in terms of their two expectations and their covariance. Extension of the result requires corresponding manipulations of expectations of products of three or more random variables—the so-called product moments of the joint distribution.

2.1. Analytic Results

In the case of a set of random variables whose joint distribution is multivariate normal, numerous authors have obtained results for product moments (see, for example, Isserlis (1918), Bendat and Piersol (1966), Bär and Dittrich (1971) and Song and Lee (2015)). For instance, in the four dimensional case we have (Bendat and Piersol (1966) and Bär and Dittrich (1971))

$$\begin{aligned} \mathrm{E}(X_1X_2X_3X_4) &= \mathrm{E}(X_1X_2)\mathrm{E}(X_3X_4) + \mathrm{E}(X_1X_3)\mathrm{E}(X_2X_4) + \\ &\quad \mathrm{E}(X_1X_4)\mathrm{E}(X_2X_3) - 2\mathrm{E}(X_1)\mathrm{E}(X_2)\mathrm{E}(X_3)\mathrm{E}(X_4). \end{aligned} \tag{6}$$

The expectations of pairwise products can be rewritten in the way described above to give

$$\begin{aligned} \mathrm{E}(X_1X_2X_3X_4) &= \mathrm{E}(X_1)\mathrm{E}(X_2)\mathrm{E}(X_3)\mathrm{E}(X_4)\times \\ &\quad \big\{ (1 + r_{X_1,X_2}\mathrm{CV}_{X_1}\mathrm{CV}_{X_2})(1 + r_{X_3,X_4}\mathrm{CV}_{X_3}\mathrm{CV}_{X_4}) + \\ &\quad (1 + r_{X_1,X_3}\mathrm{CV}_{X_1}\mathrm{CV}_{X_3})(1 + r_{X_2,X_4}\mathrm{CV}_{X_2}\mathrm{CV}_{X_4}) + \\ &\quad (1 + r_{X_1,X_4}\mathrm{CV}_{X_1}\mathrm{CV}_{X_4})(1 + r_{X_2,X_3}\mathrm{CV}_{X_2}\mathrm{CV}_{X_3}) - \\ &\quad 2\big\}. \end{aligned} \tag{7}$$

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We remark that the case of three random variables can be obtained by setting $X_4 = 1$.

For a set of N multivariate lognormally distributed random variables, product moments are given by the formula (Kotz et al. (2000))

$$\mathrm{E}\left(\prod_{j=1}^{N} X_j^{r_j}\right) = \exp\left(\boldsymbol{r}^{\mathrm{T}}\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{r}^{\mathrm{T}}V\boldsymbol{r}\right), \tag{8}$$

where $\boldsymbol{\xi}$ and V are the mean and variance of the corresponding multivariate normal distribution. Some simple manipulation leads to

$$\begin{aligned}\mathrm{E}(X_1X_2X_3X_4) \quad = \quad & \mathrm{E}(X_1)\mathrm{E}(X_2)\mathrm{E}(X_3)\mathrm{E}(X_4)\times \\ & (1 + r_{X_1,X_2}\mathrm{CV}_{X_1}\mathrm{CV}_{X_2})\,(1 + r_{X_1,X_3}\mathrm{CV}_{X_1}\mathrm{CV}_{X_3})\times \\ & (1 + r_{X_1,X_4}\mathrm{CV}_{X_1}\mathrm{CV}_{X_4})\,(1 + r_{X_2,X_3}\mathrm{CV}_{X_2}\mathrm{CV}_{X_3})\times \\ & (1 + r_{X_2,X_4}\mathrm{CV}_{X_2}\mathrm{CV}_{X_4})\,(1 + r_{X_3,X_4}\mathrm{CV}_{X_3}\mathrm{CV}_{X_4})\,. \end{aligned} \tag{9}$$

Given that eqns (7) and (9) differ, and that their reduced forms when $X_4 = 1$ also differ, we have shown that there is no general formula of this type for the basic reproductive number when there are three or more coupled heterogeneities. We do notice, however, that the two formulae give approximately equal results in the limit of small coefficients of variation, *i.e.* when one can ignore products involving two or more pairs of coefficients of variation.

The majority of papers in the literature that provide analytic results for the basic reproductive number under heterogeneity focus on at most two coupled heterogeneities. One notable exception is the work of Koella (1991), which provides—without proof or qualification for its applicability—the following formula for a vector-borne pathogen subject to heterogeneities in mosquito biting rate, a, human susceptibility, b, and duration of human infection, ρ

$$R_0 \quad = \quad R_0^{\mathrm{hom}}\left[1 + \frac{\mathrm{Var}(a)}{\bar{a}^2} + 2\frac{\mathrm{Cov}(a,b)}{\bar{a}\bar{b}} + 2\frac{\mathrm{Cov}(a,\rho)}{\bar{a}\bar{\rho}} + \frac{\mathrm{Cov}(\rho,b)}{\bar{\rho}\bar{b}}\right]. \tag{10}$$

Note that the single biological heterogeneity in biting rate impacts both infectiousness and susceptibility, resulting in it being treated as two perfectly correlated heterogeneities.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

We note that eqution (10) has no terms that involve products of pairs of covariances (or, in the language of the earlier formulae, correlation coefficients). As in the remark above comparing results between multivariate normal and lognormal distributions, this formula should, in general, be seen as an approximation that is likely most accurate when coefficients of variation are small (*i.e.* the heterogeneities are relatively minor). As a comment that is germane to a numerical example shown below, we remark that the Koella formula does agree with the result for the multivariate normal distribution, eqn (7), if the coefficient of variation describing either human susceptibility or the duration of human infection is equal to zero.

2.2. Numerical Results

We illustrate the above results using numerical simulation, allowing us to explore the differences between predictions made using the formulae for the two distributions and also using the formula in the small coefficient of variation limit. For concreteness, we place these simulations within the vector-host setting described by Koella (1991), but for simplicity we hold one of the factors constant. Specifically, hosts differ in their attractiveness to mosquitoes, impacting their susceptibility and infectiousness (thus treated as two perfectly correlated heterogeneities, X_1 and X_2, within our framework), and also in their durations of infection, X_3. Setting $X_4 = 1$ and taking $X_2{=}X_1$, we obtain the following two formulae:

$$R_0 = \mathrm{E}(X_1)^2\mathrm{E}(X_3)\big\{1 + \mathrm{CV}_{X_1}^2 + 2r_{X_1,X_3}\mathrm{CV}_{X_1}\mathrm{CV}_{X_3}\big\} \tag{11}$$

for bivariate normally distributed heterogeneities, and

$$R_0 = \mathrm{E}(X_1)^2\mathrm{E}(X_3)\big\{1 + \mathrm{CV}_{X_1}^2\big\}\big\{1 + r_{X_1,X_3}\mathrm{CV}_{X_1}\mathrm{CV}_{X_3}\big\}^2 \tag{12}$$

for bivariate lognormally distributed heterogeneities. We notice that in this reduced setting of $X_4 = 1$, the first of these formulae coincides with the small coefficient of variation limit of the two general formulae, and, as discussed above,

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

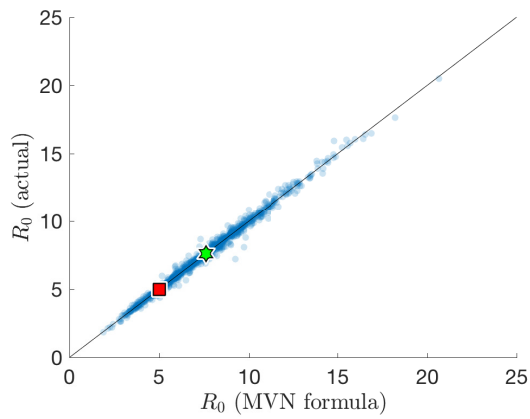
also agrees with the Koella formula. Furthermore, we see that when the correlation coefficient, r_{X_1,X_3}, between the two heterogeneities is zero, the two formulae are identical.

2.2.1. Bivariate Normal Distribution

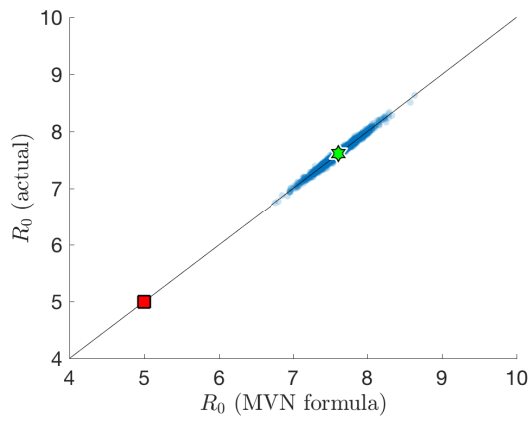
Figure (1) illustrates the performance of eqn (11) when the group attributes X_1 and X_3 are sampled from a bivariate normal distribution. Parameter values (given in the figure caption) were chosen for illustrative purposes and are not intended to represent a specific real-world infection. For each of a thousand replicates, either ten (panel a) or a thousand (panel b) pairs of values of biting rate and duration of human infection were sampled from a bivariate normal distribution. The actual R_0 value for each replicate, calculated from eqn (5), is plotted against the value of R_0 predicted for that replicate by the MVN formula, eqn (11). In addition, we show (using a red square) the R_0 value that would be predicted if there was no heterogeneity (*i.e.* X_1 and X_3 are set equal to their respective average values) and the value predicted by the MVN formula for the underlying MVN distribution (green star).

To aid comparison, the 45° diagonal line is shown on both plots: deviations from these lines represent deviations from the values predicted by the MVN formula. We quantitate these deviations by calculating the coefficient of determination, R^2, in a way that is familiar from regression theory. In both cases, the R^2 value is high, and approximately equal to 0.99. The numerically calculated values of R_0 fall in a cloud centered on the value predicted by the MVN formula, with the size of the cloud being smaller for the panel resulting from the larger number of groups. The deviations here result from sampling error, with samples drawn from the bivariate normal not being perfectly representative of the entire distribution. This effect is more pointed when there are fewer samples (*i.e.* fewer groups), with the cloud of points shrinking as the number of groups increases. In fact, since the basic reproductive number, eqn (5), is calculated as the average of a sequence of independent, identically distributed quantities, the central limit theorem can be used to quantify the variation seen about the

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



(a)



(b)

Figure 1: Comparison of R_0 values calculated numerically (using eqn 5) with those predicted by the MVN formula (eqn 11) for heterogeneities distributed according to an MVN distribution. Each blue circle represents the values of R_0 obtained for (panel a) a ten group model or (panel b) a thousand group model with pairs of values of biting rate, X_1, and average duration of human infection X_3 drawn from a bivariate normal distribution with means 1 and 5, respectively, variances $\mathrm{Var}(X_1) = 0.2$ and $\mathrm{Var}(X_3) = 4$, and correlation $r_{X_1,X_2} = 0.9$. The value of the susceptibility parameter was fixed at 1 for each group. The red square denotes the value of R_0 if there was no heterogeneity, *i.e.* obtained using the average values, R_0^{hom}, while the green star denotes the value of R_0 calculated using the MVN formula using the means, variances and covariances of the underlying bivariate normal distribution. As described in the text, the predictive ability of the MVN formula is measured by the coefficient of determination, R^2, and equals: (panel a) 0.991 and (panel b) 0.988. Note the different scales on the axes between the two panels.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

central predicted value. We remark that a corresponding figure generated in the case of two coupled heterogeneities would exhibit no deviation from the diagonal line as eqn (3) is exact: it does not rely on any distributional assumption of heterogeneities across groups.

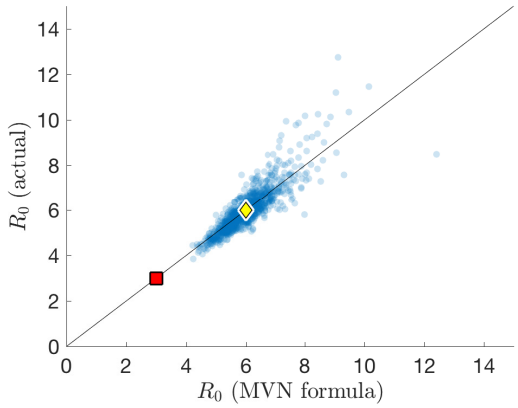
2.2.2. Bivariate Lognormal Distribution

Use of a multivariate lognormal distribution allows us to explore settings in which the components of transmission exhibit more severe heterogeneity and to assess the extent to which the impact of such heterogeneities are misrepresented by either the small coefficient of variation formula or by the formula that pertains in the multivariate normal case.

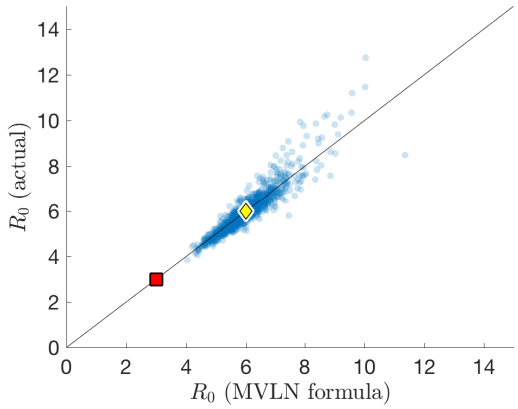
In figure (2), the biting rates and durations of infectiousness are drawn from independent lognormal distributions, with means 1 and 3, and variances 1 and 8, respectively. We compare the performance of the MVN formula (panel a) and MVLN formula (panel b) for a 1000-group setting (*i.e.* 1000 samples are drawn from the distributions). In this case, because there is zero correlation between the two heterogeneities, the population-level predictions of the MVN and MVLN formulae (*i.e.* the values obtained using the moments of the underlying distribution) are identical (shown by a yellow diamond). We see that the MVLN formula provides a better description ($R^2 = 0.748$ using the MVN formula, while $R^2 = 0.852$ using the MVLN formula), which is to be expected given that samples were drawn from an MVLN distribution.

Figure (3) explores a situation in which there is a positive correlation ($r = 0.6$) between the two components, *i.e.* the biting rate and duration of infectiousness, of the bivariate lognormal distribution. We see that the MVLN formula performs well (panel b, $R^2 = 0.852$). The positive correlation leads to the MVN formula underestimating R_0 compared to the MVLN formula, for example as witnessed by the central estimates of R_0 (yellow diamond: MVLN, green star: MVN). The MVN formula consistently provides a large underestimate of the true value of R_0 (panel a), and its predictive ability is poor (the negative value of R^2, -0.596, indicates that the formula performs worse on these points than a

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



(a)



(b)

Figure 2: Comparison of R_0 values calculated numerically (using eqn 5) with those predicted by the formula derived from the MVN distribution (eqn 11) (panel a) and those predicted by the formula derived from the MVLN distribution (eqn 12) (panel b). Each blue circle represents the values of R_0 obtained for a thousand group model with values of biting rate, X_1, and average duration of human infection, X_3, drawn from independent lognormal distributions (means 1 and 3, variances 1 and 8, respectively). The value of the susceptibility parameter was fixed at 1 for each group. The red square denotes the value of R_0 obtained using the average values, R_0^{hom}, while the yellow diamond denotes the value of R_0 calculated using the MVLN formula using the means, variances and covariances of the bivariate lognormal distribution. Because the distributions of X_1 and X_3 are assumed to be independent in this figure, the population-level predictions of the MVLN and MVN formulae are identical. The predictive ability of the MVLN formula (panel b, $R^2 = 0.852$) is greater than that of the MVN formula (panel a, $R^2 = 0.748$), as should be expected given that draws were made from an MVLN distribution.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

constant predictor).

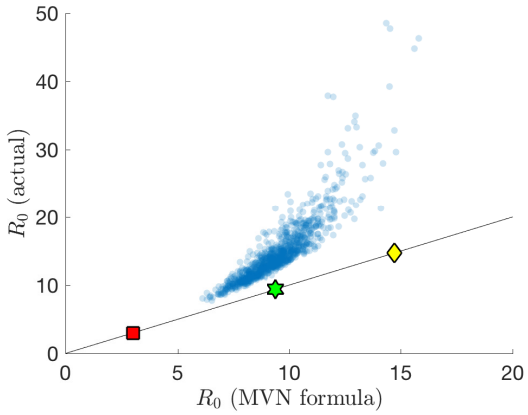
Finally, we consider a setting in which the two heterogeneities are negatively correlated, with $r = -0.2$. Figure (4) shows that in this case, the MVLN formula correctly predicts lower values of R_0 than does the MVN formula (*e.g.* compare the locations of the yellow diamond and the green star, obtained from MVLN and MVN formulae, respectively, using the moments of the underlying distribution). The MVLN formula provides reasonable predictions ($R^2 = 0.735$), whereas the MVN formula ($R^2 = -4.34$) consistently overestimates the value of R_0, and often by a considerable amount.

3. Discussion

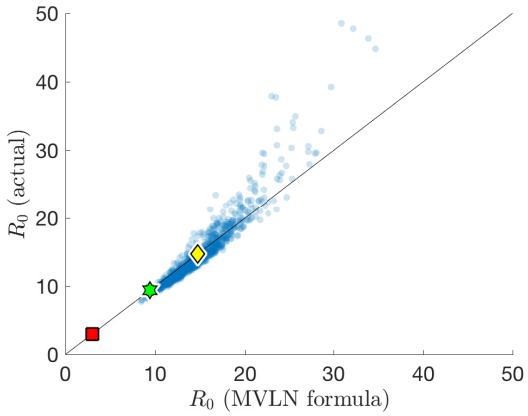
In this paper we have shown that the well-known result for the impact of two coupled heterogeneities on the basic reproductive number of an epidemiological system under separable transmission does not have a general counterpart when there are three or more coupled heterogeneities. In the more general setting, the formula for the basic reproductive number depends on details of the joint distribution of the heterogeneities in a way that is quite different than in the setting with two heterogeneities. We were able to derive formulae that related the basic reproductive number to the magnitudes of the heterogeneities and their pairwise correlations for the special cases of multivariate normal and multivariate lognormal distributions of heterogeneities. Under particular limiting cases (typically in the limit of low levels of heterogeneity), the two formulae give similar predictions. We showed that an earlier result in the literature (Koella, 1991) should be viewed as an approximate result, although we noted that in appropriate limiting cases, the result agrees with our formula for the multivariate normal distribution.

Given the reliance of the R_0 formulae obtained here on the joint distribution of the heterogeneities, our results are only exact as the number of groups in the multi-type model approaches infinity. For a finite number of groups, where the levels of the heterogeneities across groups are sampled from the underlying

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



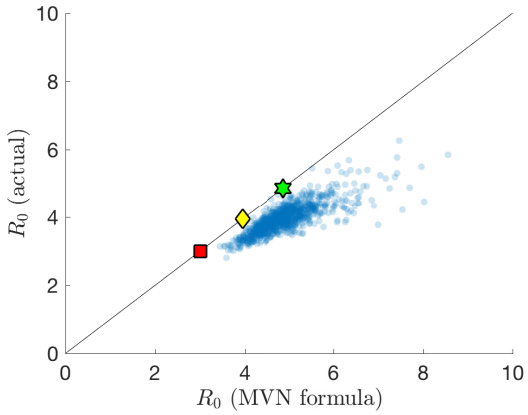
(a)



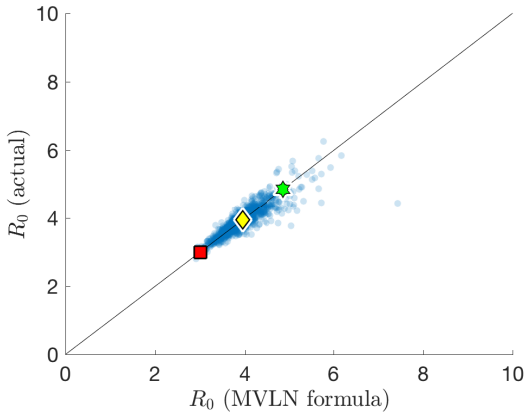
(b)

Figure 3: Comparison of R_0 values calculated numerically (using eqn 5) with those predicted by the formula derived from the MVN distribution (eqn 11) (panel a) and those predicted by the formula derived from the MVLN distribution (eqn 12) (panel b). Details are as in Figure (2), except that here there is a positive correlation, $r = 0.6$, between the two components of the bivariate log-normal distribution. In this case, the population-level predictions of the MVN and MVLN formulae (green star and yellow diamond, respectively) differ. The predictive ability of the MVLN formula (panel b, $R^2 = 0.828$) is greater than that of the MVN formula (panel a, $R^2 = -0.596$). The MVN formula performs worse than a constant predictor and consistently provides a large underestimate of the true value of R_0.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



(a)



(b)

Figure 4: Comparison of R_0 values calculated numerically (using eqn 5) with those predicted by the formula derived from the MVN distribution (eqn 11) (panel a) and those predicted by the formula derived from the MVLN distribution (eqn 12) (panel b). Details are as in Figures (2) and (3), except that here there is a negative correlation, $r = -0.2$, between the two components of the bivariate lognormal distribution. Again, the predictive ability of the MVLN formula (panel b, $R^2 = 0.721$) is greater than that of the MVN formula (panel a, $R^2 = -4.91$), with the latter performing worse than a constant predictor and consistently providing a large overestimate of the true value of R_0.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

distribution, the predictions made by these formulae are not perfect. This again is in contrast to the two heterogeneity setting, in which the well-known result is exact.

Although theoretical attention has typically focused on the two heterogeneity case, and this has provided much insight, heterogeneous transmission in the real world typically involves more than two factors (Paull et al. (2012); Vazquez-Prokopec et al. (2016)). As such, it is important to gain understanding of how multiple coupled heterogeneities impact transmission and the limitations of general results that can be obtained in such more realistic settings. This paper provides a theoretical step in that direction and aims to guide more detailed studies that involve numerical exploration of specific situations, yielding further insights into the epidemiological role of individual variability.

Acknowledgments

This work was supported by grants from the National Institutes of Health (P01-AI098670; all authors, and R01-AI091980; ALL) and the National Science Foundation (RTG/DMS1246991; ALL).

References

Anderson, R. M., Medley, G. F., May, R. M., Johnson, A. M., 1986. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. IMA J. Math. Appl. Med. Biol. 3, 229–263.

Bär, W., Dittrich, F., 1971. Useful formula for moment computation of normal random variables with nonzero means. IEEE Trans. Auto. Control 16, 263–265.

Bendat, J. S., Piersol, A. G., 1966. Random Data: Analysis and Measurement Procedures. Wiley.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Carnevale, P., Frézil, J. L., Bosseno, M. F., Le Pont, F., Lancien, J., 1978. Etude de l'agressivité d'*Anopheles gambiae* A en fonction de l'âge et du sexe des sujets humains. Bull. WHO 56, 147–154.

De Benedicitis, J., Chow-Shaffer, E., Costero, A., Clark, G. G., Edman, J. D., Scott, T. W., 2003. Identification of the people from whom engorded *Aedes aegypti* took blood meals in Florida, Puerto Rico using polymerase chain reaction-based DNA profiling. Am. J. Trop. Med. Hyg. 68, 437–446.

Diekmann, O., Heesterbeek, J. A. P., 2000. Mathematical Epidemiology of Infectious Diseases. John Wiley & Son, Chichester.

Dietz, K., 1980. Models for vector-borne parasitic diseases. In: Barigozzi, C. (Ed.), Vito Volterra Symposium on Mathematical Models in Biology. Vol. 39. Springer-Verlag, pp. 264–277.

Dye, C. M., Hasibeder, G., 1986. Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others. Trans. Roy. Soc. Trop. Med. Hyg. 83, 69–77.

Gupta, S., Anderson, R. M., May, R. M., 1989. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS 3, 807–817.

Hethcote, H. W., Yorke, J. A., 1984. Gonorrhea transmission dynamics and control. Vol. 56 of Lecture Notes in Biomathematics. Springer-Verlag, New York.

Isserlis, L., 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. Biometrika 12, 134–139.

Jacquez, J. A., Koopman, J., Simon, C., Sattenspiel, L., Perry, T., 1988. Modeling and analyzing HIV transmission: the effect of contact patterns. Math. Biosci. 92, 119–199.

bioRxiv preprint doi: https://doi.org/10.1101/220004; this version posted November 16, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Koella, J. C., 1991. On the use of mathematical models of malaria transmission. Acta Tropica 49, 1–25.

Kotz, S., Balakrishnan, N., Johnson, N. L., 2000. Continuous Multivariate Distributions. Volume 1, Models and Applications, 2nd Edition. Wiley.

Liebman, K. A., Stoddard, S. T., Reiner, R. C., Perkins, T. A., Astete, H., Sihuincha, M., Halsey, E. S., Kochel, T. J., Morrison, A. C., Scott, T. W., 2014. Determinants of heterogeneous blood feeding patterns by *Aedes aegypti* in Iquitos, Peru. PLoS Negl. Trop. Dis. 8 (2), e2702.

Lloyd, A. L., May, R. M., 1996. Spatial heterogeneity in epidemic models. J. theor. Biol. 179, 1–11.

May, R. M., Anderson, R. M., 1987. Transmission dynamics of HIV infection. Nature 326, 137–142.

Nold, A., 1980. Heterogeneity in disease-transmission modeling. Math. Biosci. 52, 227–240.

Paull, S. H., Song, S., McClure, K. M., Sackett, L. C., Kilpatrick, A. M., Johnson, P. T., 2012. From superspreaders to disease hotspots: linking transmission across hosts and space. Front. Ecol. Environ. 10, 75–82.

Song, I., Lee, S., 2015. Explicit formulae for product moments of multivariate Gaussian random variables. Stat. Prob. Lett. 100, 27–34.

Vazquez-Prokopec, G. M., Perkins, T. A., Waller, L. A., Lloyd, A. L., Reiner, R. C., Scott, T. W., Kitron, U., 2016. Coupled heterogeneities and their impact on parasite transmission and control. Trends Parasitol. 32, 356–367.