

# The Basic Reproductive Number for Disease Systems with Multiple Coupled Heterogeneities

Alun L. Lloyd<sup>a,1,\*</sup>, Uriel Kitron<sup>b</sup>, T. Alex Perkins<sup>c</sup>, Gonzalo M.  
Vazquez-Prokopec<sup>b</sup>, Lance A. Waller<sup>d</sup>

<sup>a</sup>*Department of Mathematics and Biomathematics Graduate Program, North Carolina State University, Raleigh NC 27695, USA*

<sup>b</sup>*Department of Environmental Sciences, Emory University, Atlanta, GA 30322, USA*

<sup>c</sup>*Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN 46556, USA*

<sup>d</sup>*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](mailto:alun.lloyd@ncsu.edu)

<sup>1</sup>Keywords: Disease transmission model, heterogeneity, basic reproductive number, coupled heterogeneities

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 gonorrhoea 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 Benedictis 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

26 by a type  $j$  individual in an otherwise entirely susceptible population. Conse-  
27 quently, much attention has been directed towards those special cases of het-  
28 erogeneous transmission that lead to next generation matrices whose dominant  
29 eigenvalue is analytically tractable and hence for which the basic reproductive  
30 number can be calculated explicitly. In the context of spatial heterogeneity,  
31 these include symmetric spatial configurations such as equally-sized patches  
32 with all-to-all or nearest neighbor contacts (see, for example Lloyd and May  
33 (1996)).

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

$$R_0 = \frac{1}{n} \sum_{i=1}^n a_i b_i. \quad (1)$$

39 Using the result

$$E(XY) = E(X)E(Y) + \text{Cov}(X, Y) \quad (2)$$

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

$$\begin{aligned} R_0 &= \bar{a}\bar{b}(1 + r_{ab}CV_aCV_b) \\ &= R_0^{\text{hom}}(1 + r_{ab}CV_aCV_b). \end{aligned} \quad (3)$$

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

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

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

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

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

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

## 70 **2. Results**

71 We assume that there are  $n$  types of individuals, resulting from  $N$  different  
72 heterogeneities,  $N_1$  of which impact susceptibility and  $N_2$  of which impact in-  
73fectiousness. We further assume that the susceptibility of a type  $i$  individual

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

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

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

### 85 2.1. Analytic Results

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

$$\begin{aligned} E(X_1 X_2 X_3 X_4) &= E(X_1 X_2) E(X_3 X_4) + E(X_1 X_3) E(X_2 X_4) + \\ &E(X_1 X_4) E(X_2 X_3) - 2E(X_1) E(X_2) E(X_3) E(X_4). \end{aligned} \quad (6)$$

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

$$\begin{aligned} E(X_1 X_2 X_3 X_4) &= E(X_1) E(X_2) E(X_3) E(X_4) \times \\ &\{ (1 + r_{X_1, X_2} CV_{X_1} CV_{X_2}) (1 + r_{X_3, X_4} CV_{X_3} CV_{X_4}) + \\ &(1 + r_{X_1, X_3} CV_{X_1} CV_{X_3}) (1 + r_{X_2, X_4} CV_{X_2} CV_{X_4}) + \\ &(1 + r_{X_1, X_4} CV_{X_1} CV_{X_4}) (1 + r_{X_2, X_3} CV_{X_2} CV_{X_3}) - \\ &2 \}. \end{aligned} \quad (7)$$

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

95 For a set of  $N$  multivariate lognormally distributed random variables, prod-  
 96 uct moments are given by the formula (Kotz et al. (2000))

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

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

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

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

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

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

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

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

## 124 2.2. Numerical Results

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

$$R_0 = E(X_1)^2 E(X_3) \{1 + CV_{X_1}^2 + 2r_{X_1, X_3} CV_{X_1} CV_{X_3}\} \quad (11)$$

135 for bivariate normally distributed heterogeneities, and

$$R_0 = E(X_1)^2 E(X_3) \{1 + CV_{X_1}^2\} \{1 + r_{X_1, X_3} CV_{X_1} CV_{X_3}\}^2 \quad (12)$$

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

139 also agrees with the Koella formula. Furthermore, we see that when the cor-  
140 relation coefficient,  $r_{X_1, X_3}$ , between the two heterogeneities is zero, the two  
141 formulae are identical.

### 142 2.2.1. Bivariate Normal Distribution

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

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



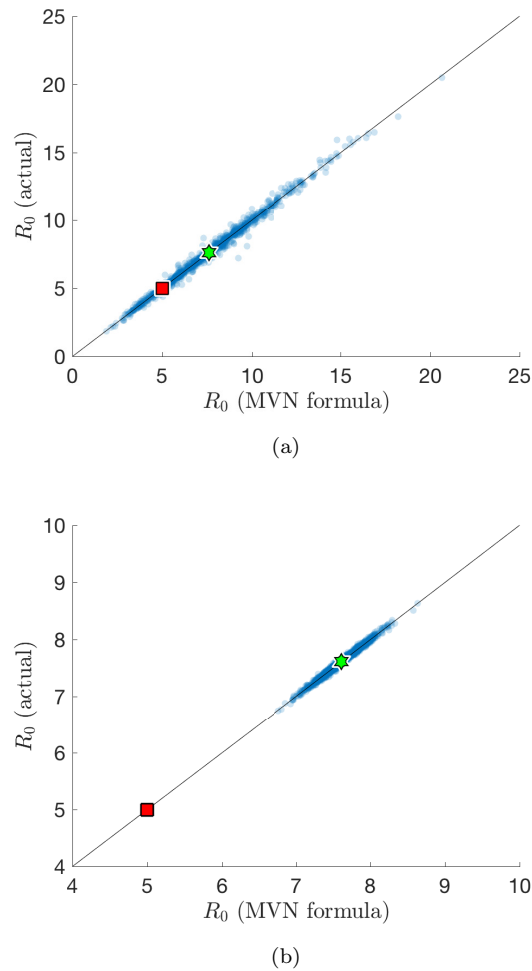


Figure 1: Comparison of  $R_0$  values calculated numerically (using eqn 5) with those predicted by the MVN formula (eqn 11) for heterogeneities distributed according to an MVN distribution. Each blue circle represents the values of  $R_0$  obtained for (panel a) a ten group model or (panel b) a thousand 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  $\text{Var}(X_1) = 0.2$  and  $\text{Var}(X_3) = 4$ , 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,  $R_0^{\text{hom}}$ , 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.

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

### 173 2.2.2. Bivariate Lognormal Distribution

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

179 In figure (2), the biting rates and durations of infectiousness are drawn from  
180 independent lognormal distributions, with means 1 and 3, and variances 1 and  
181 8, respectively. We compare the performance of the MVN formula (panel a)  
182 and MVLN formula (panel b) for a 1000-group setting (*i.e.* 1000 samples are  
183 drawn from the distributions). In this case, because there is zero correlation  
184 between the two heterogeneities, the population-level predictions of the MVN  
185 and MVLN formulae (*i.e.* the values obtained using the moments of the un-  
186 derlying distribution) are identical (shown by a yellow diamond). We see that  
187 the MVLN formula provides a better description ( $R^2 = 0.748$  using the MVN  
188 formula, while  $R^2 = 0.852$  using the MVLN formula), which is to be expected  
189 given that samples were drawn from an MVLN distribution.

190 Figure (3) explores a situation in which there is a positive correlation ( $r =$   
191 0.6) between the two components, *i.e.* the biting rate and duration of infectious-  
192 ness, of the bivariate lognormal distribution. We see that the MVLN formula  
193 performs well (panel b,  $R^2 = 0.852$ ). The positive correlation leads to the MVN  
194 formula underestimating  $R_0$  compared to the MVLN formula, for example as  
195 witnessed by the central estimates of  $R_0$  (yellow diamond: MVLN, green star:  
196 MVN). The MVN formula consistently provides a large underestimate of the  
197 true value of  $R_0$  (panel a), and its predictive ability is poor (the negative value  
198 of  $R^2$ , -0.596, indicates that the formula performs worse on these points than a

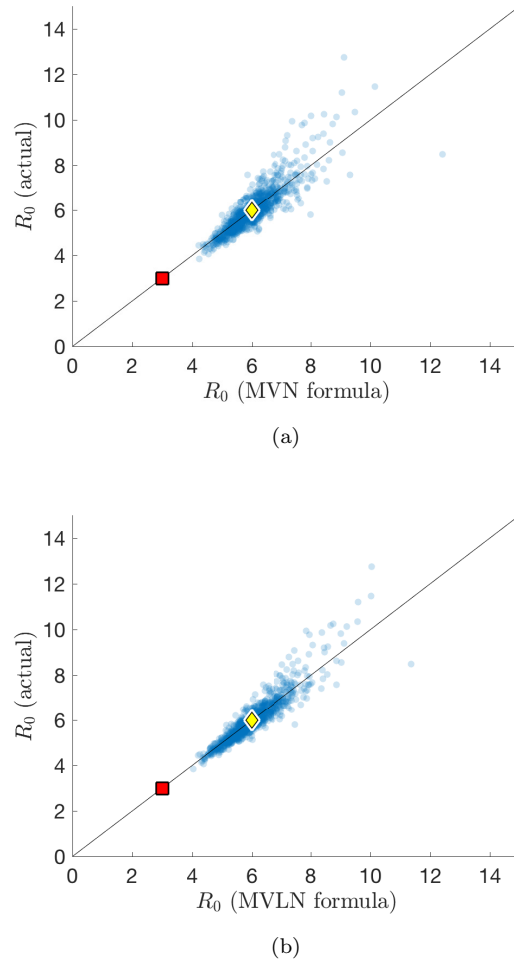


Figure 2: Comparison of  $R_0$  values calculated numerically (using eqn 5) with those predicted by the formula derived from the MVN distribution (eqn 11) (panel a) and those predicted by the formula derived from the MVLN distribution (eqn 12) (panel b). Each blue circle represents the values of  $R_0$  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,  $R_0^{\text{hom}}$ , 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.

199 constant predictor).

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

### 208 3. Discussion

209 In this paper we have shown that the well-known result for the impact of two  
210 coupled heterogeneities on the basic reproductive number of an epidemiological  
211 system under separable transmission does not have a general counterpart when  
212 there are three or more coupled heterogeneities. In the more general setting,  
213 the formula for the basic reproductive number depends on details of the joint  
214 distribution of the heterogeneities in a way that is quite different than in the  
215 setting with two heterogeneities. We were able to derive formulae that related  
216 the basic reproductive number to the magnitudes of the heterogeneities and  
217 their pairwise correlations for the special cases of multivariate normal and mul-  
218 tivariate lognormal distributions of heterogeneities. Under particular limiting  
219 cases (typically in the limit of low levels of heterogeneity), the two formulae give  
220 similar predictions. We showed that an earlier result in the literature (Koella,  
221 1991) should be viewed as an approximate result, although we noted that in ap-  
222 propriate limiting cases, the result agrees with our formula for the multivariate  
223 normal distribution.

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

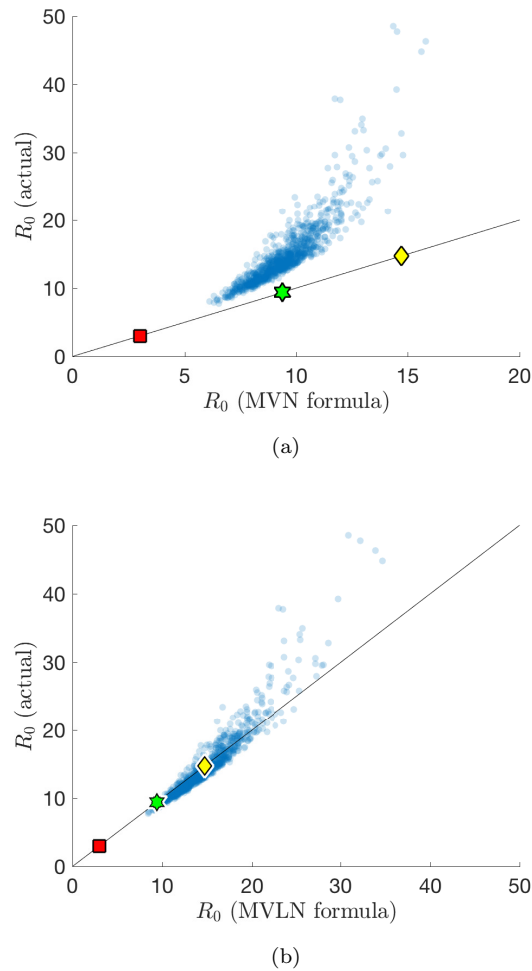


Figure 3: Comparison of  $R_0$  values calculated numerically (using eqn 5) with those predicted by the formula derived from the MVN distribution (eqn 11) (panel a) and those predicted by the formula derived from the MVLN distribution (eqn 12) (panel b). Details are as in Figure (2), except that here 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  $R_0$ .

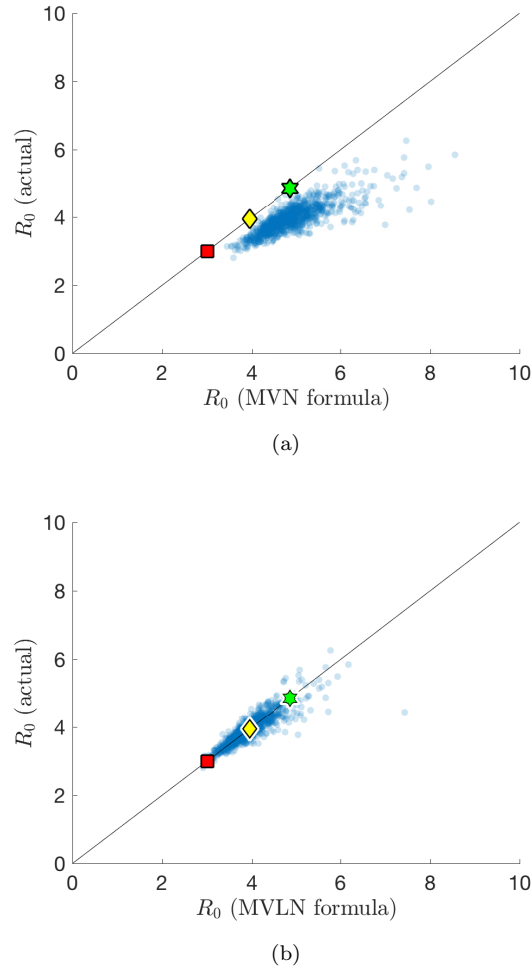


Figure 4: Comparison of  $R_0$  values calculated numerically (using eqn 5) with those predicted by the formula derived from the MVN distribution (eqn 11) (panel a) and those predicted by the formula derived from the MVLN distribution (eqn 12) (panel b). Details are as in Figures (2) and (3), except 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  $R_0$ .

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

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

#### 240 **Acknowledgments**

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

#### 244 **References**

- 245 Anderson, R. M., Medley, G. F., May, R. M., Johnson, A. M., 1986. A pre-  
246 liminary study of the transmission dynamics of the human immunodeficiency  
247 virus (HIV), the causative agent of AIDS. *IMA J. Math. Appl. Med. Biol.* 3,  
248 229–263.
- 249 Bär, W., Dittrich, F., 1971. Useful formula for moment computation of normal  
250 random variables with nonzero means. *IEEE Trans. Auto. Control* 16, 263–  
251 265.
- 252 Bendat, J. S., Piersol, A. G., 1966. *Random Data: Analysis and Measurement*  
253 *Procedures*. Wiley.

- 254 Carnevale, P., Frézil, J. L., Bosseno, M. F., Le Pont, F., Lancien, J., 1978.  
255 Etude de l'agressivité d'*Anopheles gambiae* A en fonction de l'âge et du sexe  
256 des sujets humains. Bull. WHO 56, 147–154.
- 257 De Benedictis, J., Chow-Shaffer, E., Costero, A., Clark, G. G., Edman, J. D.,  
258 Scott, T. W., 2003. Identification of the people from whom engorged *Aedes*  
259 *aegypti* took blood meals in Florida, Puerto Rico using polymerase chain  
260 reaction-based DNA profiling. Am. J. Trop. Med. Hyg. 68, 437–446.
- 261 Diekmann, O., Heesterbeek, J. A. P., 2000. Mathematical Epidemiology of In-  
262 fectious Diseases. John Wiley & Son, Chichester.
- 263 Dietz, K., 1980. Models for vector-borne parasitic diseases. In: Barigozzi, C.  
264 (Ed.), Vito Volterra Symposium on Mathematical Models in Biology. Vol. 39.  
265 Springer-Verlag, pp. 264–277.
- 266 Dye, C. M., Hasibeder, G., 1986. Population dynamics of mosquito-borne dis-  
267 ease: effects of flies which bite some people more frequently than others.  
268 Trans. Roy. Soc. Trop. Med. Hyg. 83, 69–77.
- 269 Gupta, S., Anderson, R. M., May, R. M., 1989. Networks of sexual contacts:  
270 implications for the pattern of spread of HIV. AIDS 3, 807–817.
- 271 Hethcote, H. W., Yorke, J. A., 1984. Gonorrhoea transmission dynamics and  
272 control. Vol. 56 of Lecture Notes in Biomathematics. Springer-Verlag, New  
273 York.
- 274 Isserlis, L., 1918. On a formula for the product-moment coefficient of any order  
275 of a normal frequency distribution in any number of variables. Biometrika 12,  
276 134–139.
- 277 Jacquez, J. A., Koopman, J., Simon, C., Sattenspiel, L., Perry, T., 1988. Mod-  
278 eling and analyzing HIV transmission: the effect of contact patterns. Math.  
279 Biosci. 92, 119–199.



- 280 Koella, J. C., 1991. On the use of mathematical models of malaria transmission.  
281 *Acta Tropica* 49, 1–25.
- 282 Kotz, S., Balakrishnan, N., Johnson, N. L., 2000. Continuous Multivariate Dis-  
283 tributions. Volume 1, Models and Applications, 2nd Edition. Wiley.
- 284 Liebman, K. A., Stoddard, S. T., Reiner, R. C., Perkins, T. A., Astete, H.,  
285 Sihuinchu, M., Halsey, E. S., Kochel, T. J., Morrison, A. C., Scott, T. W.,  
286 2014. Determinants of heterogeneous blood feeding patterns by *Aedes aegypti*  
287 in Iquitos, Peru. *PLoS Negl. Trop. Dis.* 8 (2), e2702.
- 288 Lloyd, A. L., May, R. M., 1996. Spatial heterogeneity in epidemic models. *J.*  
289 *theor. Biol.* 179, 1–11.
- 290 May, R. M., Anderson, R. M., 1987. Transmission dynamics of HIV infection.  
291 *Nature* 326, 137–142.
- 292 Nold, A., 1980. Heterogeneity in disease-transmission modeling. *Math. Biosci.*  
293 52, 227–240.
- 294 Paull, S. H., Song, S., McClure, K. M., Sackett, L. C., Kilpatrick, A. M., John-  
295 son, P. T., 2012. From superspreaders to disease hotspots: linking transmis-  
296 sion across hosts and space. *Front. Ecol. Environ.* 10, 75–82.
- 297 Song, I., Lee, S., 2015. Explicit formulae for product moments of multivariate  
298 Gaussian random variables. *Stat. Prob. Lett.* 100, 27–34.
- 299 Vazquez-Prokopec, G. M., Perkins, T. A., Waller, L. A., Lloyd, A. L., Reiner,  
300 R. C., Scott, T. W., Kitron, U., 2016. Coupled heterogeneities and their  
301 impact on parasite transmission and control. *Trends Parasitol.* 32, 356–367.