

Computational and statistical approaches
for quantifying the role of multi-scale
heterogeneity in *Leishmania*
transmission dynamics.

Samuel Wallace Carmichael

PHD

UNIVERSITY OF YORK
MATHEMATICS

SEPTEMBER 2023

Abstract

Leishmaniasis, a neglected tropical disease caused by infection with *Leishmania* parasites, affects millions of people annually across the globe. *Leishmania* transmission is facilitated by the sand fly vector, thus occurring across a range of climates with notable hotspots in Brazil and India. Its persistence despite ongoing eradication efforts underscores the importance of a complete understanding of the transmission dynamics in a range of environments. Developing this understanding requires tailored tools as the transmission dynamics are affected by heterogeneity at multiple scales, giving rise to a complex web of interactions. At the micro-scale, transmission is influenced by the heterogeneous parasite distributions of the host's skin as well as the complex parasite life cycle in the sand fly and its link to sand fly biting behaviour. We derive and parameterise a simple model incorporating these factors, finding that their interactions give rise to unexpected transmission opportunities. The communities in which leishmaniasis typically propagates are highly heterogeneous but also ideal candidates for deploying network models. We test analytic estimates for two epidemiologically relevant quantities, the R_0 (the average number of secondary infections caused by a single infected individual over their entire duration of infection) and the endemic equilibrium, in the context of heterogeneous networks. Although both prove to be unreliable for these structured communities, they have the potential to improve our understanding of when and where epidemics are likely to occur and be more severe. We also demonstrate the potential of using survival analysis to investigate medium and large-scale dynamics, first by confirming the role of distance in leishmaniasis transmissibility at the community level, and then by highlighting the role of social vulnerability in creating endemic hotspots. We then offer guidance for optimal application of survival analysis to future leishmaniasis research. Finally, these findings are synthesised with the wider literature to identify potential methodological improvements and further avenues of inquiry to further develop our knowledge of leishmaniasis transmission.

Contents

Abstract	2
Contents	2
List of Tables	6
List of Figures	10
List of Accompanying Material	20
Acknowledgements	21
Declarations	22
1 Introduction	23
1.1 Overview of Leishmaniasis	23
1.2 Leishmaniasis Modelling Efforts	25
1.3 Overview of Thesis	27
2 Variable bites and dynamic populations; new insights in <i>Leishmania</i> transmission.	30
2.1 Abstract	30
2.2 Author summary	31
2.3 Introduction	31
2.4 Model Details	33
2.4.1 Modelling Approach	33
2.4.2 Model Definitions	34
2.5 Results	35
2.5.1 Analytic results	37
2.5.2 Simulation study	41
2.6 Discussion	46
2.7 Conclusion	49
2.8 Materials and methods	49

3	The use of Survival Analysis in assessing the factors governing leishmaniasis transmission in empirical data sets.	50
3.1	Abstract	50
3.2	Introduction	50
3.3	Data	53
3.3.1	Dataset 1: Fulbaria thana cross-sectional study.	53
3.3.2	Dataset 2: Global leishmaniasis incidence database.	54
3.3.3	Dataset 3: Visceral leishmaniasis incidence rates in northeastern Brazil.	55
3.4	Survival Analysis Methods	56
3.5	Results	58
3.5.1	Case proximity in Fulbaria thana.	58
3.5.2	Wider applications of survival analysis	64
3.5.3	Limitations of survival analysis.	77
3.5.4	Considerations for a good survival analysis dataset.	81
3.6	Discussion	82
3.7	Conclusion	87
3.8	Materials and Methods	87
4	Applications and limitations of an analytic R_0 estimate for heterogeneous network models.	88
4.1	Abstract	88
4.2	Introduction	89
4.3	Derivation of the R_0 Estimate	90
4.3.1	Transmissibility	91
4.3.2	Expected Number of Contacts	92
4.4	Analytic Approximations of R_0 : Discrete Time	96
4.4.1	Fixed Transmission Probability	96
4.4.2	Uniform Random Infection Probability	97
4.4.3	Bimodal Infection Probability	97
4.5	Numerical Simulations	98
4.5.1	Network Structure and Degree Distributions	99
4.6	Results	101
4.7	Discussion	112
4.8	Conclusion	115
4.9	Materials and Methods	116

5	Heterogeneity in individual degree distributions allows spontaneous deviations from a stable equilibrium in <i>Leishmania</i> epidemics on networks.	117
5.1	Abstract	117
5.2	Introduction	118
5.3	Endemic Steady State, Analytic Predictions	120
5.4	Application to <i>Leishmania</i> Transmission	123
5.4.1	Network models in the context of leishmaniasis transmission.	123
5.4.2	Sample Networks and Trajectories	124
5.5	Results	125
5.6	Discussion	135
5.6.1	Insights from the SCPW	135
5.6.2	Implications for leishmaniasis modelling.	136
5.7	Conclusion	138
5.8	Materials and Methods	139
6	Discussion	140
6.1	Overview	140
6.2	Chapter 2	140
6.3	Chapter 3	141
6.4	Chapter 4	141
6.5	Chapter 5	142
6.6	Synthesis and Future Perspectives	143
6.6.1	Prospects at the Micro-Scale	144
6.6.2	Future Inquiries at Larger Scales	145
6.6.3	Closing Remarks	146
	References	147
	A Supplementary Information	192
	Appendices	192
A.1	Chapter 2 Supporting information	192
A.1.1	Population Dynamics	196
A.1.2	Supplementary Figures	200
A.2	Chapter 3 Supplementary Materials	206
A.3	Chapter 4 Supplementary Materials	208
A.3.1	Supplementary Table 1	208

A.3.2 Supplementary Methods 1: Analytic Approximations of R_0 in Continuous Time	210
--	-----

List of Tables

2.1	Table of default model parameter values.	35
3.1	Individuals in close proximity to an infected individual experience greater hazard of infection. Summary of hazard ratios, confidence intervals and p-values for each covariate fitted via univariate analysis.	59
3.2	Hazard of infection does not increase additively due to infections in multiple distance bands. Summary of hazard ratios, confidence intervals and p-values for each covariate and all interaction terms, fitted via multivariate analysis.	62
3.3	The relationship between infection proximity and hazard of infection is unmodified by incorporating an incubation period. Summary of hazard ratios, confidence intervals and p-values for each covariate fitted via univariate analysis, with the inclusion of a delay of 4 months.	63
3.4	The relationship between overall SVI and hazard of endemicity is linear. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.	68
3.5	The relationship between urban infrastructure SVI and hazard of endemicity is non-linear. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.	70
3.6	The relationship between human capital SVI and hazard of endemicity is non-linear. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.	70
3.7	The relationship between hazard of endemicity and income/work SVI is linear and positive. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.	73

3.8	The components of the SVI index are all positively correlated with each other. Pearson’s product moment correlation coefficient, with associated confidence interval and p-values, for all pairwise inter-component comparisons.	73
3.9	Bivariate analysis suggests that hazard of endemicity does not depend on proximity to an endemic municipality. Summary of hazard ratios, confidence intervals and associated p-values for both distance band covariates and their interaction term.	76
3.10	Endemic municipalities within 1 unit do not increase the hazard of endemicity. Summary of hazard ratios, confidence intervals and p-values for the nearest distance band.	76
3.11	Endemic municipalities between 1 and 1.5 units away do not increase the hazard of endemicity. Summary of hazard ratios, confidence intervals and p-values for the further distance band.	76
3.12	A poorly represented reference category prevents calculation of the Wald statistic. Summary of hazard ratios, coefficients and their standard errors, with the associated p-values for all SVI categories. The reference category is the “very low” SVI.	79
3.13	Choosing an alternative reference category mitigates the issues calculates the Wald statistic. Summary of hazard ratios, coefficients and their standard errors, with the associated p-values for all SVI categories. The reference category is the “very high” SVI.	80
3.14	Using alternative SVI categories allows for the calculation of the coefficients and associated Wald statistics. Summary of hazard ratios, coefficients and their standard errors, with the associated p-values for all SVI categories. The reference category is the “low” SVI.	81
5.1	Kolmogorov-Smirnov tests verify that very structured simulations experience more excursions than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with maximum transmission rate 0.015. Non-significant tests are excluded from this table.	129

5.2	Kolmogorov-Smirnov tests verify that very structured simulations experience more excursions than the unstructured simulations.	
	Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with moderate transmission rate 0.0115. Non-significant tests are excluded from this table.	130
5.3	Kolmogorov-Smirnov tests verify that very structured simulations experience more excursions than the unstructured simulations.	
	Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with low transmission rate 0.0087. Non-significant tests are excluded from this table.	132
5.4	Kolmogorov-Smirnov tests verify that the most structured simulations experience longer excursions on average than the unstructured simulations.	
	Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with max transmission rate 0.015. Non-significant tests are excluded from this table.	132
5.5	Kolmogorov-Smirnov tests verify that the most structured simulations experience longer excursions on average than the unstructured simulations.	
	Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with moderate transmission rate of 0.0115. Non-significant tests are excluded from this table.	132
5.6	Kolmogorov-Smirnov tests verify that the most structured simulations experience longer excursions on average than the unstructured simulations.	
	Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with low transmission rate of 0.0087. Non-significant tests are excluded from this table.	134
5.7	Kolmogorov-Smirnov tests verify that the excursions of the most structured simulations do not reach as far on average as those of the unstructured simulations.	
	Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with max transmission rate of 0.015. Non-significant tests are excluded from this table.	134

5.8	Kolmogorov-Smirnov tests verify that the excursions of the most structured simulations do not reach as far on average as those of the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with mid transmission rate of 0.0115. Non-significant tests are excluded from this table.	135
5.9	Kolmogorov-Smirnov tests verify that the excursions of the most structured simulations do not reach as far on average as those of the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with low transmission rate of 0.0087. Non-significant tests are excluded from this table.	136
A.1	Penalised spline fitting suggests a linear relationship between hazard of endemicity and overall SVI. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values, for penalised spline fitting with a threshold of 2 cases per 100000.	206
A.2	For the maximum endemic threshold, penalised spline fitting suggests a non-linear relationship. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.	208
A.3	Summary of parameter definitions.	210

List of Figures

2.1	Flowchart overview of the modelling approach. Two dynamic models, calibrated to replicate prior results, evaluate parasite population dynamics in the sand fly vector. These can be used as part of larger simulations to obtain insights into <i>Leishmania</i> transmission.	33
2.2	Replicating the results of [45] and [54]. A) Comparison of the numbers of metacyclics (top) and retroleptomonads (bottom) at specific days throughout the lifespan of the simulated flies. Blue represents flies that bite only at day 0, orange represents flies that take a subsequent blood meal at day 12. The two categories are combined prior to day 12. B) Number of simulated sand flies considered infectious at 7 days post-infection for RAG mice 10-18, parameterised according to Doehl <i>et al.</i> (see S1 Table)	36
2.3	Retroleptomonad dynamics dominate over skin heterogeneity and result in elevated mean R_0 values. Heatmaps of the mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top half) or 25% (bottom half) chance of biting an infected host. Note that each model utilises a different scale for clarity (see Fig A.15 for a version with unified scales, and Fig A.16 for a version of the top-right panel with greater sample size). Each tile represents 10,000 independent simulations.	43
2.4	Retroleptomonad dominance is dependent on having a sufficiently large maximum lifespan. A, B) Heatmaps of the mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 20 days (A) or 15 days (B). Crosses indicate the mean skin parasite burden and skin homogeneity (k) of various mice from [45]. C) Mean R_0 value against maximum lifespan for RAG mice 1-18 from Doehl <i>et al.</i> [45] (S1 Table). See Fig A.17 for a version where the heatmaps share a common z-scale.	44

3.1	Individuals sharing a household with an infected individual are more likely to become infected than those that do not. Cumulative hazard of infection for individuals with (red) or without (blue) an infected individual in their household.	59
3.2	A greater proportion of individuals sharing a household with an infected individual are infected at a given time than those who do not. Survival of individuals in households with (red) or without (blue) infections. Solid line is mean survival, dotted lines are +/- 2 s.e.	60
3.3	Individuals are infected at a greater rate if a prior infection exists within 25m than if one does not. Survival of individuals with (red) or without (blue) an infected between 1 and 25m away. Solid line is mean survival, dotted lines are +/- 2 s.e.	61
3.4	Individuals are infected at a greater rate if a prior infection exists between 25 and 50m away than if one does not. Survival of individuals with (red) or without (blue) an infected between 25 and 50m away. Solid line is mean survival, dotted lines are +/- 2 s.e.	61
3.5	Individuals experience similar hazards of infection irrespective of the presence of an infection between 50 and 75m away. Survival of individuals with (red) or without (blue) an infected between 50 and 75m away. Solid line is mean survival, dotted lines are +/- 2 s.e.	62
3.6	Individuals with an infected household member experience greater cumulative hazard of infection, even with a delay. Survival of individuals with (red) or without (blue) an infected household member, with a 4 month delay. Solid line is mean survival, dotted lines are +/- 2 s.e.	64
3.7	The cumulative hazard experienced by an individual is qualitatively similar with or without a delay. Survival of individuals with or without an infected household member, comparing the scenario with a 4 month delay with the scenario without. Solid line is mean survival, dotted lines are +/- 2 s.e.	65
3.8	Peaks in relative hazard of infection correspond to regions of emphasis in the literature, not necessarily areas of greater hazard. Relative hazard against latitude for cutaneous leishmaniasis. Solid line is mean hazard, dotted lines are +/- 2 s.e.	66
3.9	Distribution of overall SVI (top left), urban infrastructure SVI (top right), human capital SVI (bottom left) and income/work SVI (bottom right) values for all municipalities in northeastern Brazil.	68

3.10	The linear relationship remains visible, though weakened, under the increase endemic threshold. Relative hazard against overall SVI, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the overall SVI values for all municipalities included in the analysis (bottom).	69
3.11	The relative hazard of endemicity appears to peak at an SVI-UI of 0.6. Relative hazard against the urban infrastructure SVI component, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the urban infrastructure SVI values for all municipalities included in the analysis (bottom).	71
3.12	The relative hazard peaks at approximately 0.5 SVI-HC. Relative hazard against the human capital SVI component, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the human capital SVI values for all municipalities included in the analysis (bottom)	72
3.13	Relative hazard of endemicity increases linearly with the income/work SVI score. Relative hazard against the income/work SVI component, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the income/work SVI values for all municipalities included in the analysis (bottom).	74
3.14	The SVI component scores for all three components are positively correlated. Scatter plot showing pairwise comparisons of the three SVI components for all municipalities.	75
3.15	Cases in the Pigott <i>et al.</i> dataset are not distributed uniformly across the range of latitudes. Histogram of latitudes for all cutaneous leishmaniasis cases belonging to the “point” category.	78
3.16	Not all official categories of SVI are equally well represented. Histogram of municipality SVI values, under the categorisations scheme recommended by (Brasil2019).	79
3.17	Histogram of municipality SVI values, under our alternative categorisation scheme.	80
4.1	Sample network generated using the homogeneous degree distribution. Each node is an individual, edges represent viable routes of transmission.	99
4.2	Sample network generated using the geometric degree distribution. Each node is an individual, edges represent viable routes of transmission.	100

4.3	Sample networks generated for the fully structured (A) and fully random (B) versions of the community structured simulations. Each node is an individual, edges represent viable routes of transmission.	101
4.4	The estimate for R_0 is reasonable for the initial infections, but overestimates greatly for later generations. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.29) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation.	102
4.5	With uniformly distributed transmission rates the R_0 estimate still performs reasonably for generation 1, and overestimates thereafter. Mean number of secondary infections caused by infections of different generations (black) compared with the analytic estimate from Equation (4.30) (red), for a homogeneous degree distribution and uniformly distributed transmission rates. Whiskers indicate the mean across all replications +/- 1 standard deviation.	103
4.6	With a strictly bimodal distribution of transmission rates the R_0 estimate underestimates slightly for generation 1, and overestimates thereafter. Mean number of secondary infections caused by infections of different generations (black) compared with the analytic estimate from Equation (4.31) (red), for a homogeneous degree distribution and bimodally distributed transmission rates. Whiskers indicate the mean across all replications +/- 1 standard deviation.	104
4.7	For community structure networks, the R_0 estimate consistently overestimates for all generations. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation.	104
4.8	The R_0 estimate consistently overestimates for all generations, even once the community structure has been replaced with random connections. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation.	105

- 4.9 **With either uniformly distributed or bimodally distributed transmission rates, the R_0 estimate consistently overestimates.** Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation. 106
- 4.10 **The R_0 estimate consistently overestimates for all generations, for both high and low fixed transmission rates and a geometric degree distribution.** Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation. 106
- 4.11 **The R_0 estimate can successfully predict the epidemic threshold for a fixed transmission rate on a geometric degree distribution network.** A: Proportion of simulations which experience extinction within the first 120 timesteps for various transmission rates. Vertical grey line indicates the predicted epidemic threshold. B: Time to extinction for individual simulations for various transmission rates. Solid black line indicates the predicted epidemic threshold. Both A and B apply to simulations with fixed transmission rate and geometric degree distribution. 109
- 4.12 **The R_0 estimate can successfully predict the epidemic threshold for a fixed transmission rate on the "fully random" community structure networks.** A: Proportion of simulations which experience extinction within the first 120 timesteps for various transmission rates. Vertical grey line indicates the predicted epidemic threshold. B: Time to extinction for individual simulations for various transmission rates. Solid black line indicates the predicted epidemic threshold. Both A and B apply to simulations with fixed transmission rate and randomised community structure network. 110
- 4.13 **The R_0 estimate can still successfully predict the epidemic threshold for a fixed transmission rate on the "fully structured" community structure networks.** A: Proportion of simulations which experience extinction within the first 120 timesteps for various transmission rates. Vertical grey line indicates the predicted epidemic threshold. B: Time to extinction for individual simulations for various transmission rates. Solid black line indicates the predicted epidemic threshold. Both A and B apply to simulations with fixed transmission rate and community structure network. 111

5.1	Example infection trajectory (top) and sample network (bottom) for the fully random (left, structuredness = 0.0) and fully structured (right, structuredness = 1.0) simulation types, as detailed in Section 5.4.1. Each node in the network samples represents an individual.	125
5.2	Example infection trajectory (top) and sample network (bottom) for the moderate structure (left, structuredness = 0.5) and high structure (right, structuredness = 0.9) simulation types, as detailed in Section 5.4.1. Each node in the network samples represents an individual.	126
5.3	The analytic predictor has greater accuracy on simulations with lower structuredness. Average proportion of individuals infected at endemic equilibrium under various transmission rates. Output from numeric simulations of varying structuredness (see figure legend, Equation 5.17) is compared to the analytic estimate (red, dashed) provided by Equation 5.15.	126
5.4	The analytic predictor has greater accuracy on simulations with lower structuredness. Average proportion of individuals (plotted as a log scale) infected at endemic equilibrium under various transmission rates. Output from numeric simulations of varying structuredness (see figure legend, Equation 5.17) is compared to the analytic estimate (red, dashed) provided by Equation 5.15.	127
5.5	Sample excursion from a simulation with structuredness 0 and transmission rate 0.015. Excursion start/end times are denoted by the vertical red dashed lines, the endemic equilibrium is denoted by a horizontal dotted red line, and the threshold that must be exceeded to start an excursion is the solid blue line.	128
5.6	Highly structured simulations experience fewer excursions on average. Combined boxplot and violin plot of the number of excursions per simulation, grouped according to simulation structuredness. All simulations have the maximum transmission rate of 0.015 per timestep per neighbour.	128
5.7	Under a reduced transmission rate, highly structured simulations still experience fewer excursions on average. Combined boxplot and violin plot of the number of excursions per simulation, grouped according to simulation structuredness. All simulations have the moderate transmission rate of 0.0115 per timestep per neighbour.	129

5.8	Highly structured simulations experience fewer excursions on average, even at a low transmission rate. Combined boxplot and violin plot of the number of excursions per simulation, grouped according to simulation structuredness. All simulations have the low transmission rate of 0.0087 per timestep per neighbour.	130
5.9	Highly structured simulations display greater variation in excursion duration. Combined boxplot and violin plot of the average excursion duration per simulation, grouped according to simulation structuredness. All simulations have the maximum transmission rate of 0.015 per timestep per neighbour.	131
5.10	Under a reduced transmission rate, highly structured simulations still display greater variation in excursion duration. Combined boxplot and violin plot of the average excursion duration per simulation, grouped according to simulation structuredness. All simulations have the moderate transmission rate of 0.0115 per timestep per neighbour.	131
5.11	At the low transmission rate, highly structured simulations now display much lower variation in excursion duration. Combined boxplot and violin plot of the average excursion duration per simulation, grouped according to simulation structuredness. All simulations have the low transmission rate of 0.0087 per timestep per neighbour.	133
5.12	Highly structured simulations experience less severe excursions. Combined boxplot and violin plot of the average excursion distance per simulation, grouped according to simulation structuredness. All simulations have the maximum transmission rate of 0.015 per timestep per neighbour.	133
5.13	Under a reduced transmission rate, highly structured simulations experience less severe excursions. Combined boxplot and violin plot of the average excursion distance per simulation, grouped according to simulation structuredness. All simulations have the moderate transmission rate of 0.0115 per timestep per neighbour.	134
5.14	Even at the lowest transmission rate, highly structured simulations still experience less severe excursions. Combined boxplot and violin plot of the average excursion distance per simulation, grouped according to simulation structuredness. All simulations have the low transmission rate of 0.0087 per timestep per neighbour.	135

A.1	Distribution of amastigote numbers drawn from negative binomial distributions with varying skin homogeneity k : a heterogeneous environment ($k = 0.2, \text{top}$), a homogeneous environment ($k = 2.0, \text{bottom}$) and an intermediate environment ($k = 1.0, \text{middle}$). Extremely large values are depicted using crosses.	194
A.2	The proportion of different parasite life stages in the sand fly vector for the first 12 days post-infection under the system outlined in Model A (Eqns 2.1 - 2.3). An additional bite at day 6 is indicated by the vertical line.	196
A.3	The proportion of different parasite life stages in the sand fly vector for the first 12 days post-infection under the system outlined in Model B (Eqns 2.4 - 2.9). An additional bite at day 6 is indicated by the vertical line.	197
A.4	The number of parasites in each life stage in the sand fly vector for the first 12 days post-infection under the system outlined in Model A (Eqns 2.1 - 2.3). An additional bite at day 6 is indicated by the vertical line.	197
A.5	The number of parasites in each life stage in the sand fly vector for the first 6 days post-infection under the system outlined in Model B (Eqns 2.4 - 2.9).	198
A.6	The number of parasites in each life stage in the sand fly vector from day 6 to day 12 post-infection under the system outlined in Model B (Eqns 2.4 - 2.9).	198
A.7	Replicating the results of [54] (parasite proportions). Comparison of the proportions of metacyclics (top) and retroleptomonads (bottom) at specific days throughout the lifespan of the simulated flies. Blue represents flies that bite only at day 0, orange represents flies that bite at day 12. The two categories are combined prior to day 12.	200
A.8	Evaluating model robustness by randomising parameters. Number of metacyclics within the sand flies at specific days, with all parameters randomised prior to the start of each simulation. Parameters lie within 10% of the default value (Table 1). Blue represents flies that bite only a day 0, orange represents flies that bite at day 12.	200
A.9	Additional infected host proportions reflect the retroleptomonad dominance. Heatmaps of the Mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top row), 50% (second row), 25% (third row), and 10% (bottom row) chance of biting an infected host, with the smooth transmission threshold function.	201

A.10	Heatmap dynamics remain qualitatively similar under a binary transmission threshold. Heatmaps of the Mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top row), 50% (second row), 25% (third row), and 10% (bottom row) chance of biting an infected host, with the binary transmission threshold.	202
A.11	Reduced lifespan (20 days) dynamics remain qualitatively similar under a binary transmission threshold. Heatmap of the Mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 20 days, with the binary transmission threshold.	203
A.12	Reduced lifespan (15 days) dynamics remain qualitatively similar under a binary transmission threshold. Heatmap of the Mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 15 days, with the binary transmission threshold.	203
A.13	The inclusion of parasite induced mortality results in quantitative, but not qualitative, changes. Heatmaps of Mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top row) or 25% (bottom row) chance of biting an infected host, with a smooth transmission threshold. After infection, sand flies receive a 20% reduction to their remaining lifespan.	204
A.14	Removing crucial assumptions of the model has minimal influence. Mean R_0 against maximum lifespan for a representative subsample of RAG mice. A) Full model adapted from Fig 4c. B) Full model, but with no carrying capacity. C) Full model, but with additional small population sinks. D) Full model, but with larger population sinks.	204
A.15	Retroleptomonad dynamics dominate over skin heterogeneity and result in elevated mean R_0 values. Heatmaps of the mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top half) or 25% (bottom half) chance of biting an infected host. Each tile represents 10,000 independent simulations.	205
A.16	A variant of the top-right heatmap of Figure 2.3, with 40000 runs per tile instead of 10000 runs per tile.	205

A.17	Retroleptomonad dominance is dependent on having a sufficiently large maximum lifespan. A, B) Heatmaps of the mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 20 days (A) or 15 days (B). Crosses indicate the mean skin parasite burden and skin homogeneity (k) of various mice from [45]. C) Mean R_0 value against maximum lifespan for RAG mice 1-18 from Doehl <i>et al.</i> [45] (S1 Table).	206
A.18	Relative hazard of infection increases almost linearly with overall SVI. Relative hazard against overall SVI, with an endemic criterion of 2.0 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the overall SVI values for all municipalities included in the analysis (bottom).	207
A.19	The previous linear relationship remains possible despite the lack of significance in the spline fitting. Relative hazard against overall SVI, with an endemic criterion of 4.5 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the overall SVI values for all municipalities included in the analysis (bottom).	209

List of Accompanying Material

All MATLAB and R code used to replicate both the simulations and associated analysis and plotting can be found within my online GitHub repository (plain text version: <https://github.com/SamuelWCarmichael/ThesisSupplementaryCode>).

Acknowledgements

There are many to whom I would give thanks for supporting me as I worked on this thesis, probably far too many to name. I thank my supervisors for sharing their wisdom and experience, and giving me their time and patience. I thank my friends, who have supported me when I needed them; particularly Dylan, who has kept me afloat these last few years. I thank my family, without whom I would be lost. Thank you for putting up with me for all these years, and for coping with endless conversations of maths and epidemiology. Finally, I dedicate this work to my grandfather, whose conversations sparked my interest in science and set me on this path so many years ago.

To all those who have stood by me, both those I name and those I do not, thank you. Thank you from the bottom of my heart.

Declaration

I declare that this thesis is a presentation of original work and that, with the exception of Chapter 2, I am the sole author of its content. Chapter 2 (“Variable bites and dynamic populations; new insights in *Leishmania* transmission.”) is a collaborative piece of work that has been published. The author list is as follows: Samuel Carmichael, Ben Powell, Thomas Hoare, Pegine B. Walrad, and Jonathan W. Pitchford. It is published as titled in PLoS Neglected Tropical Diseases 15(1): e0009033, dated 25/01/2021. See also [172] for full details. I have also undertaken additional collaborative work that is not included within this thesis, but is referenced (see [314]). This has been published as “Spatial point pattern analysis identifies mechanisms shaping the skin parasite landscape in *Leishmania donovani* infection.” in ‘Frontiers in Immunology’ 12, dated 21/12/2021, with the following author list: Johannes S.P. Doehl, Helen Ashwin, Najmeeyah Brown, Audrey Romano, Samuel Carmichael, Jon W. Pitchford, and Paul M. Kaye. My contribution to this work was to review and advise upon the suitability of the statistical methods employed by the study. This thesis has not previously been submitted for an award at this, or any other, University. All sources used within are acknowledged as References.

Chapter 1

Introduction

1.1 Overview of Leishmaniasis

Leishmaniasis is one of a number of neglected tropical diseases [1], responsible for approximately 3.32 million disability-adjusted life years (DALYs) and more than 50,000 deaths [2] every year. Despite years of elimination efforts [3], from deploying insecticide-treated bednets (ITNs) [4] and indoor residual spraying [5] to on-going vaccine development [6], leishmaniasis remains endemic across the globe, from Brazil [7] and Ethiopia [8] to Iran [9] and India [10]. In affected regions it remains a scourge of vulnerable people, ensnaring vulnerable laborers (agricultural [11] or otherwise [12]) in a cycle of poverty [13, 14] and compounding the suffering experienced by refugees fleeing conflict zones [15, 16]. As its range shifts and expands with climate change [17], it remains imperative that we thoroughly understand the transmission dynamics of leishmaniasis.

Leishmaniasis is caused by protozoan parasites of the *Leishmania* genus, which can infect a broad array of different mammalian species [18] including domestic dogs [19] and wild mammals such as deer [20]. Transmission between hosts is facilitated by their vector, the sand fly [21] (although some studies have identified other potential vectors [22], none are officially recognised). There is a spectrum of different forms of leishmaniasis [23]. Visceral leishmaniasis is often asymptomatic [24] but symptomatic cases present a prolonged fever and enlarged organs [25] and bear a high mortality rate [26] especially if untreated or if co-infecting with HIV/AIDS [27]. Cutaneous leishmaniasis is more varied, but often manifests as slow-healing and painless skin ulcers [28] and is rarely fatal [29]. Post-kala-azar dermal leishmaniasis (PKDL) similarly forms skin lesions [30], but usually emerges after infection with visceral leishmaniasis [31] (see also [32]) and is thought to play a role in maintaining it [33]. The specific pathology of the disease is often determined by the *Leishmania* species responsible [34, 35]; thus, which form is more prominent in a given area is usually dictated by which species are prevalent there [36].

Leishmania parasites have a sophisticated lifecycle. In their mammalian hosts, the parasites persist in the form of amastigotes [37] that primarily infect macrophages [38] (but can infect other immune cells too [39]). Whilst discussion of the intricacies of this lifecycle stage are best left to others (see [40, 41] for example), it is worth noting that the resulting dissemination of the amastigotes throughout the skin [42] and blood [43] of the host facilitates their uptake by feeding sand flies [44, 45]. Sand flies primarily feed on sugar-rich plants [46], but females require a blood meal before oviposition can occur [47]. It is to this cycle of blood meals that the *Leishmania* parasites have attuned their lifecycle in the sand fly vector. Once the amastigote-infected macrophages are ingested by the sand fly the amastigotes differentiate into procyclic promastigotes [48] and, after a brief period of replication, subsequently differentiate into nectomonad promastigotes [49]. The nectomonads migrate through to the thoracic midgut [50] where they differentiate into leptomonad promastigotes. The leptomonads replicate further and migrate to the stomodeal valve [49] where they differentiate into either metacyclic promastigotes or hap-tomonad promastigotes. The metacyclics are especially mobile [49] and positioned such that when the sand fly takes another blood meal, they will be regurgitated into the host and can infect new macrophages (and differentiate back into amastigotes [49]). The hap-tomonads, on the other hand, are bound to the stomodeal valve. They are thought to be important to degrading the function of the valve [51] to enhance outward transmission [52] but remain relatively enigmatic [53]. Each time the sand fly takes a blood meal, any metacyclics that are not regurgitated undergo reverse metacyclogenesis and become retroleptomonads [54], which replicate rapidly for a few days before differentiating once more into metacyclics [54].

The behaviour of the sand flies has substantial influence over the transmission of leishmaniasis. Sand flies are a diverse group of roughly 1000 species [55] from 6 genera [21] found across a broad range of climates, from tropical forests [56] to arid farmland [57] and from desert caves [58] to more temperate climes [59]. Sand flies can also live in our settlements, in particular in the local vegetation [60] and around livestock pens [61]. Given the diversity of habitats, it is perhaps unsurprising that the different species can have very distinct blood feeding preferences: *Phlebotomus orientalis* has a diverse diet [62] but appears to prefer bovines, donkeys and humans [63, 64] whereas *P. perniciosus* displayed a preference for rabbits [65], and some species may choose a meal based more on host prevalence [66]. Indeed, not every species will feed upon humans and in the context of human leishmaniasis transmission focus is usually placed on a few genera: *Lutzomyia* [67] in the New World, and *Phlebotomus* [68] and *Sergentomyia* [65] in the Old World.

Feeding preference is not the only relevant behaviour: sand fly mobility is another

key factor. Sand flies display a preference for short-distance flight [69], usually remaining close to the ground [70] but are capable of higher, longer-distance flights to overcome natural barriers if necessary [71]. As a result, the sand fly population is heterogeneously distributed across a given region. Individuals cluster in proximity to their chosen blood meal source [72] if possible, as well as near sites that are suitable for oviposition [73, 74]. Daytime resting spots [75] are also typically hotspots: this includes vegetation [75, 76], caves or burrows [77, 78] and man-made structures such as latrines [79] and livestock shelters [61].

The traits discussed thus far do not act in isolation. Studies have demonstrated that the odour of *Leishmania*-infected canines is more attractive to female sand flies [80, 81] compared with that of their uninfected peers, and that this is also true for other mammals [82]. Combined with the propensity of sand flies to remain near blood sources [72], this creates regions of higher infection risk around reservoir hosts [78] and habitat fragments [83]. The parasites also influence the behaviour of the sand flies directly. Infected sand flies are more persistent in their attempts to feed [84] and, crucially, are more likely to bite a different host if disturbed [84]. The promastigote secretory gel (PSG) plug makes it more difficult to feed [49] and is regurgitated into the host along with a mixture of metacyclics and non-metacyclic parasite forms, thought to enhance transmission [85]. The PSG has also been reported to amplify the recruitment of macrophages to the site of a sand fly bite [86] and exacerbate cutaneous leishmaniasis in mice [87].

1.2 Leishmaniasis Modelling Efforts

The spread of *Leishmania* parasites is thus not only affected by their life cycle in both the host and vector, but also by the environment in which they exist and the specific actors involved. Our understanding of this array of underlying factors is fraught with lingering unknowns, and their many interactions create a complex and often poorly charted web that, along with the multi-scale nature of leishmaniasis transmission, renders capturing the dynamics in full an almost impossible challenge. Instead, studies have been inspired by (and limited to) individual facets of the greater problem. The aim of these studies is to employ a diverse arsenal of modelling techniques, informed by the work of empirical scientists, to identify specific pieces of the jigsaw of leishmaniasis transmission. Together, these studies have illuminated many aspects of the transmission dynamics, and we summarise some of the key outcomes of this endeavor below.

Considerable effort has been devoted to identifying and understanding regions of particularly intense leishmaniasis transmission. Recently, Karunaweera *et al.* [88] analysed the

incidence rates of cutaneous leishmaniasis rates in Sri Lanka, identifying two hotspots (areas of particularly high incidence rate) and a coldspot (an area of particularly low incidence rate). Studies have also used the Getis-Ord GI^* statistic to identify hot and coldspots of cutaneous leishmaniasis transmission in Iran [89] and Pakistan [90] and zoonotic visceral leishmaniasis transmission in China [91], and similar regions of higher-intensity transmission have been identified in Brazil [92] by examining the incidence rates in the mesoregions of Minas Gerais. These findings are complemented by studies of the associated risk factors. Recent work in Sri Lanka highlighted a link between precipitation and leishmaniasis transmission using a spatio-temporal regression-autoregression model [93], a finding reflected by studies using Bayesian geostatistical analysis [94, 95].

Modelling efforts have also sought to understand the distribution of the sand fly vectors. Ecological niche modelling (ENM) [96], a diverse tool used to predict species ranges [97] and distributions [98] (and particularly how they evolve over time [99]) using ecological niche theory [96, 100], has been often applied to sand fly distributions. Ferro *et al.* [101] found factors such as elevation and precipitation to be important to the distribution of several sand fly species in Colombia, and it has also provided evidence for *Lutzomyia intermedia* as the main cutaneous leishmaniasis vector for southeastern Brazil [102]. ENM has been used to evaluate how the ranges of different sand fly species will be affected by climate change, predicting their potential expansion further into North America [103] as well as in Morocco [104] and Central Europe [105]. Such models provide a means to identify populations that could become at risk from leishmaniasis [106], though they must be used carefully [107, 108] to avoid common mistakes [109] and ensure proper reproducibility [110].

Another valuable modelling tool is survival analysis. Survival analysis identifies risk factors that increase the likelihood of an event of interest happening in a given time period [111] by analysing the ‘event times’ of cohorts of individuals. It has seen a range of applications, from identifying risk factors of equine influenza transmission [112] to modelling *Aedes aegypti* lifespan [113] and the survival time of cystic fibrosis patients [114]. It has also been applied to leishmaniasis: survival analysis has evaluated the potential for thermotherapy to treat cutaneous leishmaniasis [115] and pentamidine in preventing relapses [116], highlighted the role of age [117] and arsenic exposure [118] in visceral leishmaniasis mortality, and investigated the emergence of PKDL after visceral leishmaniasis recovery [119]. Another less conventional application is to identify risk factors associated with the transmission of leishmaniasis. Previously, Coura-Vital *et al.* used survival analysis to link the presence of unplastered walls to canine visceral leishmaniasis transmission [120] and more recently Oliveira *et al.* highlighted the role of major roads in sand fly dispersal and the expansion of leishmaniasis in Brazil [121].

A more commonly used tool for identifying risk factors is logistic regression [122]. It has been applied in a range of contexts [123–125] including leishmaniasis transmission. Logistic regression models have identified a range of risk factors associated with leishmaniasis infection. Household proximity to forest [126] and other vegetation [127] increases the risk of infection for both human [128] and canine [129] leishmaniasis, as does the presence of poor quality [11, 130] or damaged [131, 132] walls. Agricultural work has also been highlighted as a risk factor [133] especially for migrant workers [134] or those who sleep in temporary shelters [135]. Farm work also appears to pose a particularly acute risk during the weeding and harvesting seasons [136, 137], perhaps due to increased exposure to sand fly resting places [138]. Logistic regression has found positive factors too, demonstrating the protective benefits of insecticide treated bed-nets [139] and properly screened windows [140], as well as the benefit of sleeping upstairs [141]. Not all factors are so simple, however: some studies have reported that owning livestock or dogs [142, 143] increased the risk of visceral leishmaniasis infection whereas others found livestock to have no such effect [144]. Studies have also suggested that the risk may be influenced by the type of animal present [127], and one study indicated that the presence of sheep could have a protective effect [145], contrasting with earlier work [143].

Other modelling techniques have also seen application to leishmaniasis. Chapman *et al.* [146] used spatiotemporal SEIR model parameterised using a Monte Carlo Markov Chain (MCMC) approach to evaluate how transmission is affected by distance between individuals and found that the transmission rate decreases rapidly with increasing distance [146], reflecting similar findings from studies using logistic regression [128, 147]. Doehl *et al.* [45] captured the heterogeneous distribution of *Leishmania* parasites in the skin of a mammalian host. Using a model motivated by earlier work on predator-prey encounter rates [148], they concluded that a model including heterogeneity at both micro-scales (within a given patch of skin) and macro-scales (across the whole skin of the host) most accurately predicted the outward transmissibility of the host to the sand flies [45]. Collectively, this array of different techniques has yielded insights into the many different facets of leishmaniasis transmission and pathology.

1.3 Overview of Thesis

Several recent experimental and theoretical developments have arisen which present new opportunities for leishmaniasis research. Serafim *et al.* identified a new form of parasite in the sand fly vector, the retroleptomonad promastigote, which emerges after subsequent blood meals, highlighting its role in amplifying the parasite burden of the sand fly [54]. In

a similar vein, Doehl *et al.* demonstrated the potential importance of skin heterogeneity in outward transmission [45]. At larger scales, Chapman *et al.* found that the distance between individual mammalian hosts has an important role in determining the transmission rate [146] and Ribeiro *et al.* illustrated the connection between leishmaniasis transmission in Brazil and regions of high social vulnerability [149]. Network models also present opportunities: in addition to existing R_0 estimates for heterogeneous networks [150], recent work by Corcoran and Hastings [151] found an analytic solution to the Super Compact Pairwise Model [152] of disease transmission on networks, which may have applications in leishmaniasis transmission. In this work, we build upon these findings to investigate the role of heterogeneity at multiple scales in leishmaniasis transmission.

The focus of Chapter 2 is on micro-scale heterogeneity: how the heterogeneous distribution of parasites in the skin of a mammalian host, identified by Doehl *et al.* [45], and the life cycle of the parasites in their sand fly vector (particularly the newly identified retroleptomonads [54]) interact to allow transmission even in contexts that were typically regarded as unfeasible. To assess the relative importance of these factors, a simple model is derived and compared with prior experimental results to ensure the parameterisation is appropriate, and then applied with numerical simulation to explore how transmission is influenced by different biting regimes and skin heterogeneity values.

Chapter 3 focuses on heterogeneity at larger scales, such as communities or regions. Survival analysis is applied to a range of publicly available datasets to assess its potential and limitations within the context of leishmaniasis. Echoing the work of Chapman *et al.* [146] at the community-scale, the role of distance to nearest infection is examined as well as the presence of an infection within the household. At regional scales, focus is instead placed on social vulnerability factors such as urban infrastructure and access to healthcare (inspired by the work of Ribeiro *et al.* [149]). Finally, at the largest scales the limits of survival analysis are highlighted using global case distributions.

In Chapter 4, the focus shifts towards the potential for network models in capturing and exploring the dynamics of leishmaniasis transmission. An analytic estimate for the basic reproduction rate (R_0) in the context of heterogeneous networks, presented by Lloyd and Valeika [150], is explored and specific estimates are derived for a range of transmission rate distributions of interest. The accuracy of these estimates is tested using numerical simulation, as well as the potential application of the R_0 value in identifying an epidemic threshold for the transmission rate.

Chapter 5 continues to explore the applications of network models, now in the context of the endemic equilibrium. A recent estimate for the endemic equilibrium of an SIS model on heterogeneous networks, derived by Corcoran and Hastings [151] is examined and its

accuracy tested under a range of networks designed to mimic a spectrum of different structures (from fully structured networks inspired by real-world communities to fully randomised configuration networks). The limitations of the estimate are discussed, and scenarios where it could be beneficial in restricting the transmission of leishmaniasis are highlighted.

Finally, Chapter 6 presents an overarching discussion and synthesis of the key findings of the preceding chapters, with the aim of grounding them in the wider research context. It also seeks to provide suggestions for the prioritisation of future research work and other efforts intended to minimise the harm, and maximise our understanding, of leishmaniasis transmission.

Chapter 2

Variable bites and dynamic populations; new insights in *Leishmania* transmission.

Variable bites and dynamic populations; new insights in *Leishmania* transmission

Samuel Carmichael¹, Ben Powell¹, Thomas Hoare², Pegine B. Walrad^{2*}, Jonathan W. Pitchford^{1,2*}

¹ Department of Mathematics, University of York, York, UK

² Department of Biology, University of York, York, UK

* Corresponding authors: jon.pitchford@york.ac.uk, pegine.walrad@york.ac.uk

2.1 Abstract

Leishmaniasis is a neglected tropical disease which kills an estimated 50,000 people each year, with its deadly impact confined mainly to lower to middle income countries. *Leishmania* parasites are transmitted to human hosts by sand fly vectors during blood feeding. Recent experimental work shows that transmission is modulated by the patchy landscape of infection in the host's skin, and the parasite population dynamics within the vector. Here we assimilate these new findings into a simple probabilistic model for disease transmission which replicates recent experimental results, and assesses their relative importance. The results of subsequent simulations, describing random parasite uptake and

dynamics across multiple blood meals, show that skin heterogeneity is important for transmission by short-lived flies, but that for longer-lived flies with multiple bites the population dynamics within the vector dominate transmission probability. Our results indicate that efforts to reduce fly lifespan beneath a threshold of around two weeks may be especially helpful in reducing disease transmission.

2.2 Author summary

Two recent discoveries hold particularly important ramifications for *Leishmania* transmission. First, parasites are heterogeneously distributed within the skin of an infected host. Second, the discovery of a new lifecycle stage known as the retroleptomonad promastigote changes the within-vector parasite dynamics. It is not yet known how these newly identified factors may interact to influence transmission. In this study, we design a tractable model for parasite population dynamics in the sand fly vector that consolidates these new results into a single system. We first demonstrate that our model can replicate established experimental results. We then interrogate this model, both analytically and numerically, to draw conclusions about *Leishmania* transmission in an ecological and epidemiological context. We conclude that the relative importance of the two focal factors depends critically on sand fly lifespan. In short-lived sand flies the heterogeneity in the number of parasites initially taken up by a sand fly is typically the crucial factor in *Leishmania* transmission, whereas for longer-lived sand flies the retroleptomonad lifecycle stage is likely to drive transmission. In a practical context these results suggest that minimising the expected sand fly lifespan could be an effective strategy to reduce *Leishmania* transmission.

2.3 Introduction

Leishmaniasis is caused by parasites of the *Leishmania* genus. Details of the infection depend on the particular species [153], but all species share the same general vector-borne lifecycle, with distinct and complex life cycle stages in the mammalian host and sand fly vector [48]. *Leishmania* parasites have two main morphological forms. Broadly speaking, amastigotes (ovoid, non-flagellated) dominate the mammalian stage of the lifecycle. Promastigotes (larger, flagellated) are found in the vector, and are divided into multiple developmental subclasses [49, 154].

Sand flies in natural settings are often opportunistic feeders, capable of feeding on a variety of mammalian and avian species [18, 65]. Mature female sand flies require a blood meal during each oviposition cycle. When an uninfected female sand fly bites an infected

mammal, it ingests amastigote-infected macrophages from the host's skin or blood [45]. Within the first few days, amastigotes differentiate into procyclic promastigotes, which are resistant to the digestive enzymes of the sand fly midgut [48]. Procyclics then exponentially replicate before differentiating into nectomonad promastigotes [49]. Nectomonads are able to migrate towards the thoracic midgut [48] and bind to the midgut epithelium [50] where they differentiate into leptomonad promastigotes [49]. Leptomonads are the second replicative stage, and migrate through the thoracic midgut to the stomodeal valve [49] where these differentiate into metacyclic promastigotes, the human-infectious stage. Metacyclics have a short cell body and long flagellum to enhance motility [49], and can be transmitted to a new host where they infect host macrophages via phagocytosis, which is a requirement for the parasite life cycle in the host (the full infection dynamics in the host are similarly complex [41,155], but are not relevant to this investigation which focuses on transmission potential from vector to host). Two recent key findings concerning details of *Leishmania* biology offer new insights into the possibility of understanding, and possibly controlling, the spread of the disease. They are described below.

Patchy landscape of infection in the host Transmission from host to vector occurs when a sand fly consumes a blood meal from an infected host. Doehl *et al.* [45] examined amastigote *Leishmania donovani* infections in immunodeficient mice. By evaluating the correlation of the sand fly parasite burden with multiple measures of host parasite burden, they showed first that the parasite load in mammalian host skin, rather than blood, is the major determinant of successful sand fly infection. They further found that skin parasite burden is highly variable within and between mammalian hosts and developed a modelling approach to investigate the consequences of this patchiness. For a host with a low mean parasite burden, a patchy skin landscape enhanced outward transmission (although the overall probability of successful transmission remained low), whereas for a host with a high parasite burden a homogenous distribution favoured transmission.

Retroleptomonads A new lifecycle stage was identified by Serafim *et al.* [54], the retroleptomonad promastigote [54]. When a sand fly with a mature (metacyclic enriched) infection takes another blood meal, the metacyclic stage can de-differentiate into a leptomonad-like stage, termed the retroleptomonad. These replicate for 3-4 days before differentiating back into metacyclics [54]. This serves to greatly amplify the parasite load prior to the next bite (4.5 fold increase in the number of metacyclics 18 days post infection in comparison to a sand fly that has fed only once) and thus increases the probability of disease transmission [54], a finding confirmed experimentally under laboratory conditions.

Doehl *et al.* [45] observed that often the sand flies would only carry a relatively small infection after a single feed, suggesting that perhaps sand flies may only be expected to

infect once they had taken 2 previous bites (and thus had their infection amplified via the the retroleptomonad stage [54]), but the correlation between these two mechanisms has not yet been fully explored.

The objective of the work presented here is to build a mathematical model to incorporate these new findings and assess the impact upon *Leishmania* transmission. A simple differential equation model, parameterised by data from [49], was developed to describe the population dynamics of nectomonad, leptomonad and metacyclic promastigote stages within the vector (Model A). This model was then refined by the addition of the retroleptomonad lifecycle stage, using data and observations from [54] (Model B). These models of population dynamics within the sand fly provide a framework for a series of stochastic simulations which describe the random processes of feeding and parasite ingestion across multiple blood meals. Such simulations allow the consequences of changes in disease prevalence at the epidemiological scale and the thresholds of disease transmission to be quantifiably predicted.

2.4 Model Details

2.4.1 Modelling Approach

The modelling strategy is summarised in Fig 2.1. First, we develop a simple, algebraically tractable and computationally efficient model for parasite population dynamics within a single infected sand fly, and then parameterise this model according to the available information. This model then forms a key ingredient in a series of larger stochastic simulations intended to extract useful details about the transmission of *Leishmania*.

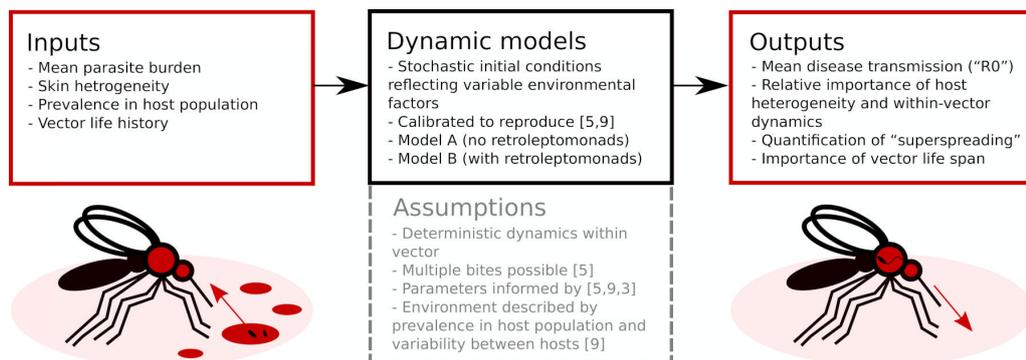


Figure 2.1. Flowchart overview of the modelling approach. Two dynamic models, calibrated to replicate prior results, evaluate parasite population dynamics in the sand fly vector. These can be used as part of larger simulations to obtain insights into *Leishmania* transmission.

In order to create a tractable model, several key assumptions are made. In addition to those represented in Fig 2.1, we also assume that differentiation between parasite life

cycle stages occurs at 100% efficiency and that there is a single globally applied sand fly carrying capacity of *Leishmania* parasites.

2.4.2 Model Definitions

Model A describes the dynamics of Nectomonads (N), Leptomonads (L) and Metacyclics (M) using a simple set of near-linear ordinary differential equations (ODEs),

$$\frac{dN}{dt} = -\alpha N \quad (2.1)$$

$$\frac{dL}{dt} = \alpha N + rL \left(1 - \frac{N + L + M}{C}\right) - sL \quad (2.2)$$

$$\frac{dM}{dt} = sL - uM \quad (2.3)$$

The assumptions are biologically parsimonious: N differentiate into L at rate α , L replicate at rate r (limited by a carrying capacity C) and differentiate to M at rate s , and M are also subject to mortality at rate u .

Model B extends Model A to incorporate the dynamics of the Retroleptomonads (R) [54] using two sets of near-linear ODEs. Initially, Model B consists of Equations 2.4-2.7:

$$\frac{dN}{dt} = -\alpha N \quad (2.4)$$

$$\frac{dL}{dt} = \alpha N + rL \left(1 - \frac{N + L + M + R}{C}\right) - sL \quad (2.5)$$

$$\frac{dM}{dt} = sL + vR - uM \quad (2.6)$$

$$\frac{dR}{dt} = qR \left(1 - \frac{N + L + M + R}{C}\right) - vR \quad (2.7)$$

In addition to the original assumptions, it is assumed that any existing R differentiate to M at rate v and replicate at rate q limited by carrying capacity C . For a four-day period after subsequent bites, we exchange Equations 2.6 and 2.7 for Equations 2.8 and 2.9 respectively:

$$\frac{dM}{dt} = sL - gM - uM \quad (2.8)$$

$$\frac{dR}{dt} = qR \left(1 - \frac{N + L + M + R}{C}\right) + gM \quad (2.9)$$

Now, M dedifferentiate to R at rate g and R no longer differentiate to M .

Parameterisation of Model A was performed using data obtained from Rogers *et al* [49]

(see S1 Method) but due to a lack of suitable data, it was not possible to perform similar parameter fitting for the new parameters in Model B.

Table 2.1 includes a summary of the default parameter values chosen.

Table 2.1. Table of default model parameter values.

Parameter	Name	Default Value	Source
α	Nectomonad differentiation rate	1.52	[A]
r	Leptomonad replication rate	1.45	[A]
s	Leptomonad differentiation rate	1.65	[A]
u	Metacyclic decline rate	1.61	[A]
C	Carrying capacity	$2 * 10^6$	[A]
v	Retroleptomonad differentiation rate	4.0	[B]
q	Retroleptomonad replication rate	3.5	[B]
g	Metacyclic dedifferentiation rate	4.0	[B]

All parameters and their default values. [A]: Values are derived from parameterisation based on data from Rogers *et al.* [49], see S1 Method. [B]: Parameter estimates chosen to ensure that simulated population dynamics are consistent with experimental observations from Serafim *et al.* [54].

For a more thorough examination of the parasite population dynamics see Section A.1.1, and for an implementation of the above models see Supplementary S1 Code.

2.5 Results

Replicating experimental results on sand fly feeding schedules and mammalian infection heterogeneity

In order to verify that our retroleptomonad-inclusive Model B is capable of replicating the experimental results observed by Serafim *et al.* [54], we ran a set of 20,000 Monte Carlo simulations designed to imitate their experimental setup. In this scenario, all flies take a bite at day 0 from an infected host. Half the flies take an additional bite at day 12 from an uninfected host, the other half take no subsequent bites. We fix the mean skin parasite burden to 2×10^6 and let the skin homogeneity $k = 2$ to mimic the blood source used by Serafim *et al.* After the initial bite, we take up a number of amastigotes determined using a negative binomial distribution with P (probability of a positive result) and R (number of required successes) such that:

$$R = \frac{\mu_N^2}{\sigma_N^2 - \mu_N} \qquad P = \frac{\mu_N}{\sigma_N^2}.$$

where:

$$\mu_N = P_B V_{BM} \qquad \sigma_N^2 = (\mu_N) \left(1 + \frac{\mu_N}{k}\right)$$

and we define P_B to be the mean skin parasite burden and V_{BM} to be the blood meal volume. The conversion of amastigotes to nectomonads is not covered directly by our models, but the number of nectomonads is approximately three times greater than the number of amastigotes [49] and thus we simply multiply the number of amastigotes by three. See Section S2 Method for further details and examples of the distribution of amastigotes under varying k values.

In this example, the initial number of nectomonads N_0 is determined using a distribution with mean μ and variance σ^2 :

$$\mu = 9,600 \qquad \sigma^2 = 46,108,800$$

and we assume that $V_{BM} = 1.6 \times 10^{-3}$ throughout.

Of particular interest are the numbers of metacyclics and retroleptomonads present in each fly throughout their adult lifespan. Fig 2.2A compares the numbers of metacyclics and retroleptomonads at each day sampled by Serafim *et al.*

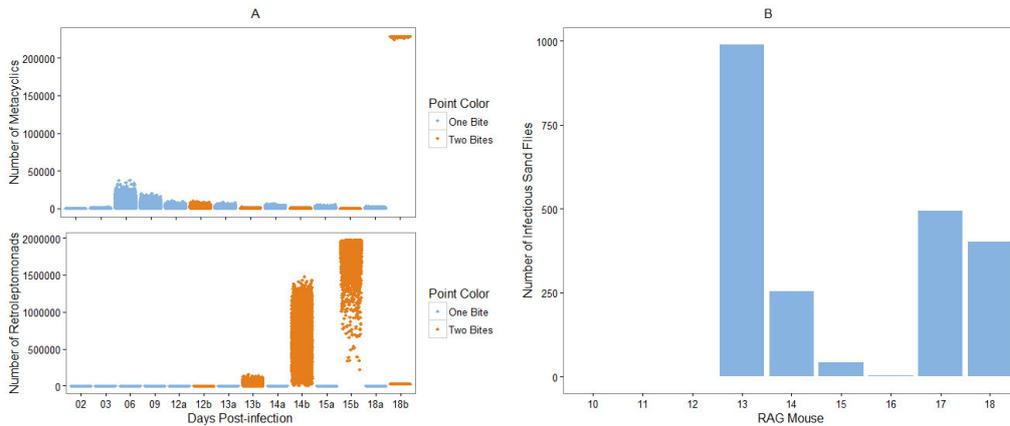


Figure 2.2. Replicating the results of [45] and [54]. A) Comparison of the numbers of metacyclics (top) and retroleptomonads (bottom) at specific days throughout the lifespan of the simulated flies. Blue represents flies that bite only at day 0, orange represents flies that take a subsequent blood meal at day 12. The two categories are combined prior to day 12. B) Number of simulated sand flies considered infectious at 7 days post-infection for RAG mice 10-18, parameterised according to Doehl *et al.* (see S1 Table)

Fig 2.2A reflects the qualitative dynamics observed in the experiments of Serafim *et al.* We observe a similar reduction in the number of metacyclics immediately after the bite at day 12 and a corresponding increase in the number of retroleptomonads over the same time period. Similar behaviour can be observed for the proportions of metacyclics and retroleptomonads (Fig A.7), and this behaviour is sufficiently robust to be observed even with parameter randomisation (Fig A.8).

We also wish to verify that our model can describe the role of heterogeneity in the skin parasite distribution as reported by Doehl *et al* [45]. To do so, we ran sets of 1000 Monte Carlo simulations for parameter combinations corresponding to mice 10-18 as calculated by Doehl *et al* (S1 Table). Each simulated fly fed on an infected host at $t = 0$. We then sampled the number of metacyclics in each fly after 7 days. Based on the work of Sadlova *et al.* [156], we consider a sand fly to be infectious if 500 metacyclics are present at day 7 post-infection. This is a distinct, but similar, approach to that of Doehl *et al.* [45] Whereas Doehl *et al.* predicted the number of flies with mature infections based upon amastigote uptake, we evaluate this number directly using a comparable threshold. Fig 2.2B compares the number of infectious sand flies for each mouse.

We observe that heavily infected mice, such as mouse 13, result in a large proportion, if not all, of the sand flies being mammalian-infectious at day 7 post-infection (S1 Table). Relatively smaller infections, such as those of mice 10 and 16, typically lead to negligibly-infectious sand flies. This matches the observations made by Doehl *et al* [45] and verifies that our model successfully captures the relationship between outward transmission potential and skin patchiness.

2.5.1 Analytic results

In this section we provide analytically-derived properties and consequences of simplified versions of our models. These serve to reinforce and validate the numerically derived behaviours discussed in Section 2.5.2 and to highlight the key processes driving transmission. In particular, we present expressions bounding implied disease transmission probabilities in a range of hypothetical scenarios.

In order to render it analytically tractable, it is necessary to make two simplifications to our model. Explicitly, we assume that 1) blood meals only occur at specific predetermined times, rather than at random gamma-distributed times as in the full model, 2) no sand fly mortality occurs during our simulations. This simplifies the probabilistic model such that the only random variables affecting the parasite transmission events are the initial number of parasites present in the sand fly, and the presence or absence of a second blood meal.

More specifically, we restrict our attention to scenarios in which a sand fly takes either two or three blood meals over a period of 12 days. In all scenarios let N_0 be the number of nectomonads present in the sand fly 4 days post-blood meal. We choose $t = 0$ such that each sand fly initially carries N_0 nectomonads. We also assume that the fly feeds on an uninfected host at time $t = 12$, when it deposits M_{12} parasites in the metacyclic life cycle stage. N_0 is considered a random variable. M_{12} is considered a deterministic function of N_0 , and so inherits probabilistic behaviour from this random variable. A transmission event is associated with the sand fly depositing a number of parasites (M_{12}) exceeding a threshold T . Thus transmission is also a random variable inheriting probabilistic behaviours from N_0 .

The scenarios we consider differ in terms of the occurrence of an additional blood meal from an uninfected host at time $t = 6$. In our model, this 2nd ingested blood meal triggers differentiation to the retroleptomonad lifecycle stage, associated replication and re-differentiation back to metacyclic stage, impacting the number of metacyclics that can be deposited at time $t = 12$.

Given that there are blood meals only at times 0 and 12, the structure of the model described in Section 2.4 is such that M_{12} is proportional to N_0 i.e.

$$M_{12} = C_2 N_0 \quad (2.10)$$

where C_2 is a constant derived by solving the system of equations in Section 2.4 (the system is far from carrying capacity so this assumption of linearity provides a good approximation). It is implicitly a function of the model's differentiation rate parameters and the time elapsed between blood meals.

If an additional blood meal at time $t = 6$ does occur, a different set of equations that involve the retroleptomonads (specifically Equations 2.4-2.9, which comprise Model B) is used to determine the resulting number of metacyclics at time $t = 12$. M_{12} is now determined by N_0 and a correspondingly different multiplicative constant

$$M_{12} = C_3 N_0 \quad (2.11)$$

Expressions (2.10) and (2.11) can be combined to give

$$M_{12} = C_3 N_0 \mathbf{1}_B + C_2 N_0 (1 - \mathbf{1}_B) \quad (2.12)$$

where $\mathbf{1}_B$ is an indicator function taking value one when the $t = 6$ blood meal occurs, and zero otherwise.

We can now, for instance, consider the expectation of M_{12}

$$\begin{aligned}\mathbb{E}(M_{12}) &= C_3 \mathbb{E}(N_0) \mathbb{E}(\mathbf{1}_B) + C_2 \mathbb{E}(N_0)(1 - \mathbb{E}(\mathbf{1}_B)) \\ &= [C_2 + (C_3 - C_2) \mathbb{E}(\mathbf{1}_B)] \mathbb{E}(N_0)\end{aligned}\tag{2.13}$$

which follows on the assumption that $\mathbf{1}_B$ and N_0 are considered probabilistically independent. Note that $\mathbb{E}(\mathbf{1}_B)$ is the probability that the blood meal bite takes place.

Eq (2.12) can also be used to produce an expression for the transmission probability at time $t = 12$

$$\begin{aligned}P(\text{Transmission}) &= P(M_{12} \geq T) \\ &= P(M_{12} \geq T | \text{second bite})P(\text{second bite}) \\ &\quad + P(M_{12} \geq T | \text{no second bite})P(\text{no second bite}) \\ &= P(N_0 \geq T/C_3) \mathbb{E}(\mathbf{1}_B) + P(N_0 \geq T/C_2)(1 - \mathbb{E}(\mathbf{1}_B))\end{aligned}\tag{2.14}$$

We will use Eq (2.14) to express how the variability in N_0 , which was the subject of interest in Doehl *et al.* [45], and the variability in the blood meal availability, which was the subject of interest in Serafim *et al.* [54], both contribute to the probability of disease transmission.

To help progress our arguments here we appeal to Chebyshev's inequality, which tells us that a random variable takes values close to its expectation with high probability, more precisely it says that the probability of the random variable being further than $w > 0$ standard deviations from the expectation is smaller than w^{-2} i.e.

$$P(|X - \mathbb{E}(X)| \geq w\sqrt{\text{var}(X)}) \leq 1/w^2\tag{2.15}$$

or equivalently

$$P(|X - \mathbb{E}(X)| \geq w) \leq \text{var}(X)/|w|_+^2\tag{2.16}$$

where we have introduced the rectifier function

$$|w|_+ = \begin{cases} w & w > 0 \\ 0 & w \leq 0 \end{cases}\tag{2.17}$$

In order to accommodate negative w .

In the case when there is no bite at time $t = 6$ Chebyshev's inequality allows us to put

an upper bound on the transmission probability

$$\begin{aligned}
P[\text{Transmission} \mid \text{no second bite}] &= P[M_{12} \geq T \mid \text{no second bite}] \\
&= P[C_2 N_0 \geq T] \\
&= P[N_0 - \mathbb{E}(N_0) \geq T/C_2 - \mathbb{E}(N_0)] \\
&\leq P[|N_0 - \mathbb{E}(N_0)| \geq T/C_2 - \mathbb{E}(N_0)] \\
&\leq \text{var}(N_0) / |T/C_2 - \mathbb{E}(N_0)|_+^2 \tag{2.18}
\end{aligned}$$

Such an upper bound is useful because it suggests ways the transmission probability can, in principle at least, be forced down. We could, for example, force down the variance of the number of parasites ingested at time $t = 0$. Alternatively, by decreasing the conversion rate from nectomonads at time $t = 0$ to metacyclics at time $t = 12$ we would decrease C_2 which also serves to bring down the upper bound.

Considering the average over cases in which the blood meal bite does and does not occur at time $t = 6$, Chebyshev's inequality leads us to an expression of the form

$$\begin{aligned}
P[\text{Transmission}] &= P[M_{12} \geq T] \\
&\leq \text{var}(N_0) \left(\frac{\rho}{|T/C_3 - \mathbb{E}(N_0)|_+^2} + \frac{1 - \rho}{|T/C_2 - \mathbb{E}(N_0)|_+^2} \right) \\
&\leq \text{var}(N_0) \frac{1}{|T/(\rho C_3 + (1 - \rho)C_2) - \mathbb{E}(N_0)|_+^2} \\
&= \text{var}(N_0) \frac{1}{|T'/C_2 - \mathbb{E}(N_0)|_+^2} \tag{2.19}
\end{aligned}$$

where the second line follows from Jensen's inequality and ρ is the probability that the blood meal bite occurs. Since $C_3 > C_2$, the second bite/retroleptomonad phenomenon effectively leads to a version of Eq (2.18) in which the transmission threshold has been lowered from T to

$$T' = T \times \frac{1}{1 + \rho(C_3/C_2 - 1)} \tag{2.20}$$

As well as providing quantitative predictions, this 'equivalent threshold' result is intended to provide another angle from which to interpret the significance of the retroleptomonad reproduction mechanism. Specifically, the retroleptomonads do not negate the capacity for skin heterogeneity to increase metacyclic numbers to transmission-sufficient levels for a subset of flies. Rather, they make these levels easier to attain. We see the effects of skin heterogeneity and the retroleptomonads act together to contribute to disease transmission.

An alternative expression linking the retroleptomonads to the transmission probability

follows from assuming that the number of metacyclics derived from retroleptomonads is very large relative to the transmission threshold (i.e. $C_3N_0 \gg T$). In this case we can consider the transmission probability, given the blood meal bite at $t = 6$, is close to one

$$P(M_{12}^* \geq T | \text{second bite}) \approx 1 \quad (2.21)$$

Then, using Chebyshev's Inequality we see that

$$\begin{aligned} P(M_{12}^* \geq T) &\leq \rho + (1 - \rho) \frac{\mu_{M^*}(1 + \mu_{M^*}/k)}{(T/C_2 - \mu_{M^*})^2} \\ &= \rho + (1 - \rho) \frac{\text{var}(N_0)}{(T/C_2 - \mathbb{E}(N_0))^2} \end{aligned} \quad (2.22)$$

where μ_{M^*} is the mean number of metacyclics present at day 12 post-infection.

This bound provides another way to assess the relative influences of key parameters on the probability of transmission. For cases in which the transmission threshold is high relative to the number of metacyclics produced without the retroleptomonads (i.e. $C_2N_0 \ll T$) and the blood meal bite probability ρ is reasonably large, the rightmost expression in Eq. (2.22) dominates. We then see the transmission probability reduced to the blood meal bite probability. When ρ is very small, however, the variance of N_0 , and the skin heterogeneity that drives it, becomes important again. In this case it is this heterogeneity that provides each sand fly with the greatest likelihood of depositing a sufficient number of *Leishmania* parasites at time $t = 12$ to cause transmission.

Our simplified model, via Eq. (2.22), re-frames the competing roles of the second blood meal and the skin heterogeneity in a mathematically precise way. The simulations and discussions below do the same at increasing levels of realism, but necessarily decreasing levels of mathematical formalism.

2.5.2 Simulation study

This simplified model is useful because it allows us to make analytical predictions about the behaviour of our system. However such predictions are useful only where their implications can be related to more sophisticated systems. Let us once more consider the full system for both models as originally defined (Model A: Eqs 1-3; Model B: Eqs 4-9). Each sexually mature female fly has a predetermined lifespan drawn from an exponential distribution with a mean and standard deviation of 13 days. These sand flies bite throughout their lives, with inter-bite times drawn from a gamma distribution of mean 6 days, standard deviation $\sqrt{3}$ days and with bite loads as previously defined (S2 Method). We also reinstate a 3-day delay before the emergence of nectomonads and assume that all sand flies are initially

uninfected.

We require a suitable metric to assess the infectiousness of *Leishmania* under a variety of mean skin parasite burden and k values. One such metric commonly used in epidemiology is the R_0 [157] defined as "the number of secondary infections generated from a single infected individual introduced into a susceptible population" [158]. As we do not explicitly model individual hosts, this measure is unsuitable. Let us instead consider a proxy value: mean sand fly transmission capacity (hereafter referred to as mean R_0), defined to be the average number of infections caused by a single sand fly. Though this is not strictly an R_0 value, higher mean R_0 values imply a higher R_0 value for the disease assuming that the number of sand flies biting a given infected host remains unchanged.

We determine that a transmission has occurred at a given bite using either a binary threshold or a smooth 'threshold function'. In the case of the binary threshold, we assume that if the number of metacyclics transferred (M_T) exceeds some fixed threshold T , an infection is guaranteed (and if not an infection never occurs). For the smooth 'threshold function', we assume the chance of infection P_T at a given bite depends on M_T such that:

$$P_T = 0.5(\tanh(0.015(M_T - 200)) + 1) \quad (2.23)$$

Unfortunately, little is known about the true threshold for infection with leishmaniasis, but our chosen function possesses useful traits. Even if very few metacyclics are transferred, the probability of transmission is non-zero (if low), allowing mildly infected sand flies to occasionally infect a mammalian host (which is important because sand flies have been shown to often transmit between 10 and 100 parasites [159]). The probability then rises rapidly, reaching a probability of 1 at a dose of 400 parasites. This means that very high doses of parasites will always infect the host.

Whilst the binary threshold is easier to relate to our analytical work it is very unlikely to be applicable to a real situation, especially as it disregards any nutritional or genetic variation between potential hosts. Thus, let us consider the smooth threshold function. Corresponding figures for the binary threshold function can be found in the supplementary information, and we observe qualitatively similar behaviour with both the binary and smooth thresholds.

We compare our two models' outputs for a range of different scenarios. Assume that some proportion of hosts is initially infected and that this proportion is fixed with no dependence on time or transmissions. Initially, we will consider two scenarios where our simulated flies bite at random from a population of hosts in which either 100%, or 25%,

of hosts are infected (see Fig 2.3; for further scenarios see Fig A.9 and for the binary threshold equivalent see Fig A.10).

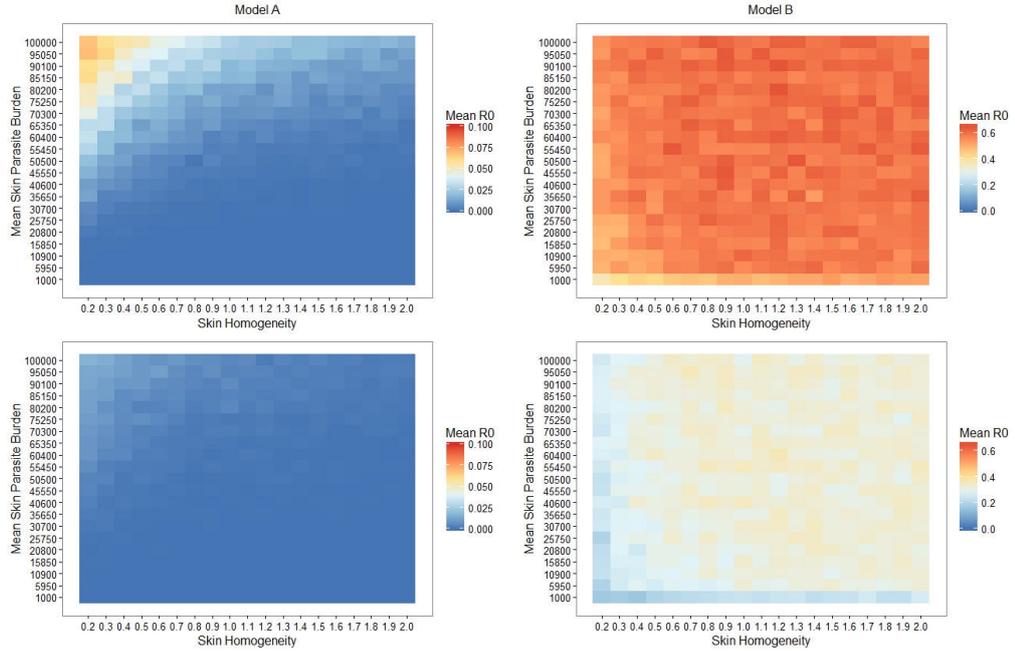


Figure 2.3. Retroleptomonad dynamics dominate over skin heterogeneity and result in elevated mean R_0 values. Heatmaps of the mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top half) or 25% (bottom half) chance of biting an infected host. Note that each model utilises a different scale for clarity (see Fig A.15 for a version with unified scales, and Fig A.16 for a version of the top-right panel with greater sample size). Each tile represents 10,000 independent simulations.

Although the simplest conclusion we can draw from these heatmaps is that introducing retroleptomonads increases our mean R_0 value, there are several other notable results. We observe that for Model A there is a peak in the mean R_0 value for low skin homogeneity and high mean skin parasite burden for both scenarios. Though our analytic approach does not deal directly with Model A, we could consider Model A to simply be the scenario where flies never take 3 blood meals (and thus where the retroleptomonad lifecycle stage has no significant role in day 12 transmission). In this context, we note that a low skin homogeneity increases the probability of transmission as some flies are able to ingest a sufficient number of parasites to become infectious by the next blood meal. In contrast, more homogeneous skin environments reduce the probability that any individual sand fly would ingest sufficient parasite numbers for strong transmission capacity. These findings support the prediction of Doehl *et al.* [45].

The peak is entirely absent from the corresponding heatmaps for Model B; instead we have a plateau spanning most of the parameter space with a slight decrease in mean R_0 for very low k values (i.e. very patchy environments). We note from our analytical section that as ρ (the chance of taking 3 bites) increases, k (skin homogeneity) has a

progressively reduced impact. Thus, given that ρ effectively remains constant (and non-zero) regardless of k one might anticipate that the mean R_0 would be independent of k . Similarly, considering the magnitude of the amplification of the metacyclics (Fig 2.2A) it is reasonable to expect that the mean skin parasite burden would be relatively unimportant. This does not hold for very low skin homogeneity and/or parasite burdens, because under these conditions it is possible that the sand fly may fail to be initially infected or may not remain infected by the time of their second blood meal. In such instances, the *Leishmania* parasite burden may not increase sufficiently for transmission despite the retroleptomonad-dependent population boost.

Accordingly, skin homogeneity has a particularly reduced role in very long lived sand flies that bite many times. In these flies, the number of metacyclics are repeatedly amplified, resulting in almost guaranteed parasite transmission to mammalian hosts at the third and subsequent blood meals for the majority, rendering such sand flies potential "super spreaders". To assess the impact of such flies, let us restrict the lifespans of the simulated flies to 20 days (Fig 2.4A, and see Fig A.11 for the binary threshold equivalent). Restrict-

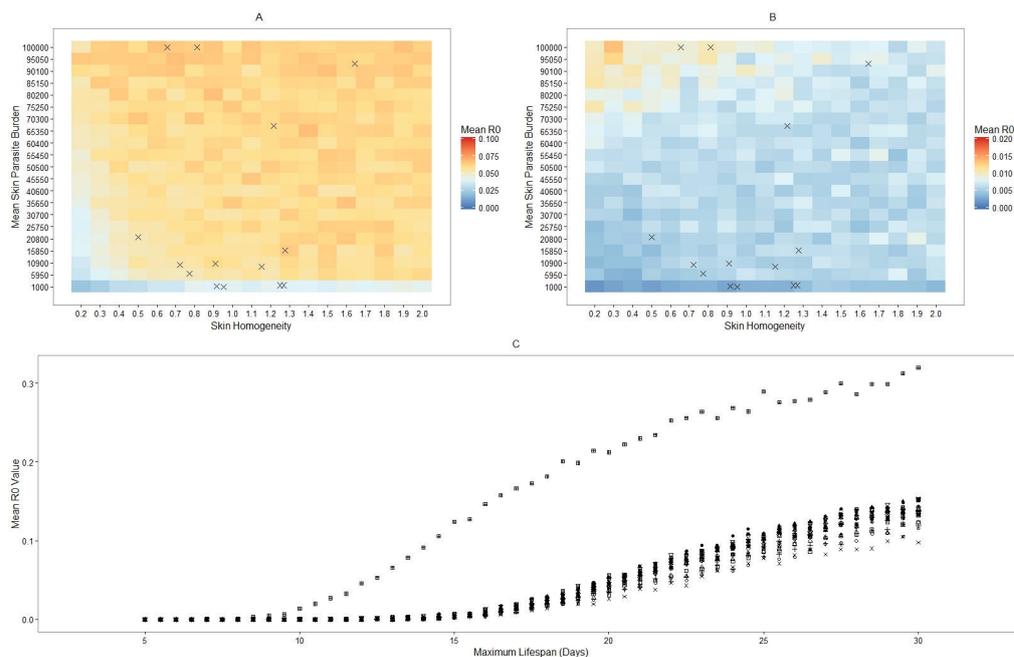


Figure 2.4. Retroleptomonad dominance is dependent on having a sufficiently large maximum lifespan. A, B) Heatmaps of the mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 20 days (A) or 15 days (B). Crosses indicate the mean skin parasite burden and skin homogeneity (k) of various mice from [45]. C) Mean R_0 value against maximum lifespan for RAG mice 1-18 from Doehl *et al.* [45] (S1 Table). See Fig A.17 for a version where the heatmaps share a common z-scale.

ing the lifespan of the flies to 20 days appears to have minimal effect on the influence of skin homogeneity, though a reduced plateau in mean R_0 value is achieved. This impact is predominantly due to the abbreviated capacity for metacyclic-enhancing blood meals in

female sand flies with reduced lifespans. It should be noted that with a mean inter-bite time of 6 days, it is not unlikely that a given individual could take 3 blood meals in 20 days.

We next consider a further restriction of the lifespan to 15 days (Fig 2.4B, and see Fig A.12 for the binary threshold equivalent). Under this new, harsher restriction we see that skin homogeneity has much stronger influence on the mean R_0 value. The peak observed in Model A is present again. The mean R_0 value does not drop to zero away from that peak, however. This is likely because some flies will still manage to bite three times and thus benefit from the retroleptomonad replicative cycle (this could also be interpreted as having a low, but non-zero, ρ and thus we would expect a similarly low but non-zero mean R_0).

Further simulations based on the Doehl *et al.* mice help elucidate the transition between these two states. Using the parameterisation for mice 1-18 from Doehl *et al.* [45] (S1 Table), we ran sets of 5,000 sand flies for each mouse for a range of different maximum lifespans and calculated the mean R_0 value for each set. We can then compare the trajectory taken by the mean R_0 value for each population of simulated sand flies as we increase the maximum lifespan (Fig 2.4C).

We note that the mean R_0 value increases with the maximum sand fly lifespan for all mice, especially once it exceeds 15 days, as anticipated from Fig 2.4A and 2.4B. As sand fly longevity increases it stimulates a smooth transition away from a patchiness-dominated scenario and towards a retroleptomonad-dominated scenario. Thus the conclusions of Doehl *et al* [45] do not hold for flies with unrestricted lifespans, but provide valuable insight into the transmission potential of shorter-lived sand fly populations. Reducing the maximum lifespan of the sand flies (and thus enlarging the shorter-lived portion) can have a tangible impact on the mean R_0 value.

It is important to consider the sensitivity of our conclusions to certain model assumptions. Firstly, we have not fully addressed the effect of *Leishmania* infection on the sand fly vector. It has been documented that sand flies experience a reduction in their lifespan when infected [84], although the effect is not yet fully understood. In S3 Method, we modify the model to incorporate a 20% reduction in sand fly lifespan once infected. Supplementary Fig A.13 demonstrates a quantitative reduction in mean R_0 but no qualitative changes to the behaviour of our system: we maintain the single peak exhibited by Model A, and the plateau of Model B. Though reduced, parasite infection and transmission dynamics are essentially unchanged.

We have also assumed that there exists a standard sand fly carrying capacity, suggesting a constant tolerance for infection by all parasite lifecycle stages. Supplementary Fig

A.14B shows the mean R_0 against maximum lifespan for a representative subsample of the RAG mice used by Doehl *et al.*, as in Fig 2.4C, but in simulations where no limit to population size is imposed. We note that the results are almost indistinguishable from those of the full system (Fig A.14A, Fig 2.4C). Our final sensitivity check removes the assumption of 100% efficiency in parasite differentiation. To represent this reduction in efficiency, we include a population sink at each lifecycle stage (see S3 Method for model specification and parameters). Supplementary Fig A.14C and Fig A.14D correspond to the small and large sinks, respectively. Although Supplementary Fig A.14D shows a marked decrease in mean R_0 , in all cases we still observe the same qualitative relationship between mean R_0 and maximum lifespan.

2.6 Discussion

We observe both numerically and analytically that the inclusion of retroleptomonads allows sand flies which take multiple bites to transfer more parasites on subsequent bites and thus be more effective at transmitting leishmaniasis, as anticipated by Serafim *et al* [54]. Less trivially, we also observe that the inclusion of retroleptomonad-dependent amplification in the model alters the relationship between the mean R_0 and skin homogeneity. In scenarios where the retroleptomonad life cycle stage is absent (Model A) or play a substantially reduced role (Fig 2.4B) we see a strong dependence on skin homogeneity, with patchy environments leading to more transmissions as some flies take up many parasites and can then cause infections, as predicted by Doehl *et al* [45]. In scenarios where retroleptomonads are more important however, we see the opposite: skin homogeneity is unimportant to the transmission of the disease, as even small numbers of parasites initially present can be amplified greatly.

This result may reduce the perceived importance of the predictions made by Doehl *et al.* [45], yet there are important considerations that highlight its relevance. Doehl *et al.* predicted that patchy skin distributions would enhance transmissions because sand flies could occasionally take up higher parasite loads and then can lead to increased sand fly and subsequent mammal infections. Homogeneous skin environments, on the other hand, would reduce the likelihood of the *Leishmania* parasite establishing an initial sand fly infection. While we observe the loss of the relationship between skin homogeneity and mean R_0 for the full system there are scenarios where it re-emerges. Flies with short lifespans (Fig 2.4B) cause more transmissions with patchy than even skin distributions. Such sand flies are unlikely to live long enough to bite three or more times and thus the parasite populations do not typically benefit from the amplification step of the retroleptomonad stage

in the model. This is reflected in our analyses. Consider the short-lifespan flies to have a low chance of taking three bites (IE a low ρ), then from Eq 2.22 we see that low k values increase the chance of transmission. Thus, there are conditions under which the scenario posed by Doehl *et al.* is relevant to the spread of the parasite. Perhaps an important caveat to the *in vivo* infection study is that immunodeficient mice from Doehl *et al.* may not properly represent a typical immunocompetent individual. While patchiness has not been reported in immunocompetent mice, the phenomenon of patchy skin parasite distributions remains applicable to clinically symptomatic Post-Kala Azar Dermal Leishmaniasis (PKDL) patients.

The extent to which our model's outcomes apply to parasite transmission in natural settings is uncertain. Multiple lab-based studies suggest that female sand flies have fairly short adult lifespans (<20 days) [160] with further reductions when infected [84]. Lab-based sand fly viability estimates are confounded by numerous challenges in maintaining sand fly colonies [47] and additional mortality associated with factors such as oviposition [161] and bacterial infection [162] that do not appear to impact wild populations as prominently. Release-recapture studies in natural settings suggest that flies may live much longer than in lab environments [163]. To address this uncertainty, we have incorporated parasite-induced mortality for an exemplar scenario to begin to assess its influence upon *Leishmania* transmission. Though this new addition did not alter the qualitative behaviour of this system for our exemplar scenario, we did observe a reduction in mean R_0 in all tested parameter combinations. This mean R_0 reduction will grow in magnitude for more severe lifespan reductions. We would also observe a loss of the plateau in Model B if the parasite-induced mortality was sufficiently severe to prevent the retroleptomonads from emerging. Such scenarios are, however, unlikely to be reasonable. In order to properly model the impact of parasite-induced mortality on the transmission potential of sand flies, it will be crucial for future studies to discern the true expected lifespan of wild sand flies and the full extent to which this lifespan is reduced by *Leishmania* parasite infection.

Transmission dynamics are further complicated by the feeding behaviour of the sand flies. We chose to model the time between subsequent blood meals (in days) using a gamma distribution of mean 6. Though this is a reasonable approximation for our model, in reality there is little information available about how often sand flies feed. It is likely that the feeding rate is linked to the oviposition cycle (given the dependence of oviposition on a blood meal) and the abundance of potential blood sources and promiscuous feeding behaviour exhibited by sand flies [65]. The scenario of regular feeds posed by Serafim *et al* [54] is a significant improvement upon theories which incorporate only a second blood meal at day 12. This seems appropriate for sand flies with abundant sources of blood

meals, yet it is not uniformly true for all populations. We also consider human populations with different proportions of initially infected hosts (P_i) including values such as 25% and 10% which are more applicable to populations where leishmaniasis is endemic [164, 165]. Although we observe that our results hold for such scenarios, we assume that hosts are evenly distributed throughout the populations and this is unlikely to be biologically accurate.

There is significant evidence that the behaviour of the sand flies is also altered once infected. A notable component of *Leishmania* infection known to alter sand fly behaviour is Promastigote Secretory Gel (PSG), a filamentous proteophosphoglycan-based gel secreted into the thoracic midgut and stomodeal valve [48, 49]. The occupation of the midgut by PSG causes the sand flies to feed ineffectively, taking smaller blood meals [49, 166] and demonstrating increased persistence when disturbed (with an increased likelihood of biting a second host after a disturbance) [84]. PSG also acts as a filter allowing only metacyclics to pass through [49], and impedes the unidirectional flow of blood through the stomodeal valve, causing the sand fly to regurgitate PSG and the parasites within it into the bite. This may amplify the number of infectious parasites transferred to a new host on a successful bite [49, 52]. Giraud *et al.* [85] recently investigated the complexity of this impact upon transmission. They reported that sand flies could regurgitate high "quality" (metacyclic-enriched) parasite doses even after multiple successive bites in a feed, likely due to PSG acting as a filter [49], but subsequent maintenance varies as the infection progresses in the fly. They also report that differences in dose quality have tangible impacts on the trajectory of the resulting infection in a mouse host, with lower quality bites often leading to larger, but less outwardly infectious lesions.

The interactions between PSG, fly feeding behaviour, and *Leishmania* population dynamics could have important implications for transmission. Sand flies that do feed on multiple hosts during a feed [84] could cause multiple infections given the enriched doses they may transmit, and the variable dose quality [85] may contribute to the emergence of variable patchiness in the skin of mammalian hosts observed by Doehl *et al* [45]. Although we model the regurgitation of parasites by increasing the number of transferred metacyclics for heavily infected flies [159], we do not directly model the PSG due to insufficient information regarding its production and how it interacts with the parasites in the midgut. Similarly the role of superspreading in *Leishmania* transmission, though beyond the scope of this study, may have significant implications for future models.

Another avenue of future enquiry that holds potential value relates to improving the parameterisation of our model. As the discovery of the retroleptomonad lifecycle stage is very recent [54] we have insufficient data to parameterise Model B with accuracy. Although

our chosen parameters are informed by the population graphs of Serafim *et al.* and we can demonstrate that our model produces similar behaviour to that of the experimental system, it would be preferable to have more data to base our parameters upon. Future studies may seek to improve the identification of retroleptomonads using transcriptomics tools as has been done for previous life cycle stages [167]. Alternatively, they may seek to provide more information about the two lifecycle stages we omit from our model, the amastigotes and procyclic promastigotes. Either of these options would greatly improve predictions from future models.

2.7 Conclusion

This work has produced a basic population dynamic model for nectomonad, leptomonad and metacyclic promastigotes and integrated the recently discovered retroleptomonad promastigote. This model can be further enhanced via the addition of missing life cycle stages or additional parameter to improve the fit. This provides a basic tool that can be expanded upon depending on the aims of a study. For example, a similar model may prove useful if modelling the impact of interventions on promastigote dynamics. Through using Monte Carlo Simulations, we have demonstrated that the addition of retroleptomonads to the model greatly enhances transmission from the second bite onwards. This could suggest that retroleptomonads are a good stage to target in control efforts, potentially through interventions that reduce the number of bites a sand fly takes. We have also demonstrated that skin parasite heterogeneity does have an impact on *Leishmania* transmission, although a much smaller impact than retroleptomonads. A patchy distribution slightly enhances transmission when retroleptomonads are not present (such as the first bite), but a non-patchy distribution enhances transmission when retroleptomonads develop.

2.8 Materials and methods

Model parameterisation was performed in RStudio v1.2.5019 (R version 3.6.1) with the digitize package [168] using data from [49] (see Supplementary Method S1 for full details). All Monte Carlo simulations were performed in MATLAB R2019b. Data analysis was performed in RStudio v1.2.5019 (R version 3.6.1).

Chapter 3

The use of Survival Analysis in assessing the factors governing leishmaniasis transmission in empirical data sets.

3.1 Abstract

Leishmaniasis demonstrates considerable spatial heterogeneity at large scales through the clustering of cases within both individual settlements and broader administrative regions. Previous studies have uncovered some of the factors underpinning the clustering, such as the presence of habitat fragments like parks in urban regions. However, our understanding remains incomplete. A valuable tool that could aid in clarifying this matter is survival analysis. This methodology has previously demonstrated links between the emergence of leishmaniasis and the presence of major transport routes. We examine three publicly available datasets of leishmaniasis cases at different spatial and temporal scales through the lens of survival analysis. We find that proximity to prior infections increases the likelihood of infection, and that poorer access to stable work and education increases the likelihood that the disease may become endemic. We also discuss the limitations of survival analysis, and explain how some of these limitations can be remedied as new data become available.

3.2 Introduction

The ongoing efforts to contain the spread of leishmaniasis depend upon a thorough understanding of its transmission, both within human populations and between different

mammalian hosts. Leishmaniasis is a vector-borne disease with a single vector: the sand fly. At first glance, transmission may appear quite simple: the sand fly takes up *Leishmania* parasites in a blood meal from an infected individual, and deposits them in a susceptible individual on a subsequent blood meal. In reality the transmission process is much more complex. There are three key aspects of transmission that we need to understand to properly grasp the spread of leishmaniasis.

Firstly, the uptake of the parasites from the host. It was often assumed that blood parasitemia was an ideal measure of host infectiousness [169] with bites being equally infectious for a given host [170], but more recent work has highlighted that skin parasite load is important to outward transmission [44,171]. In contrast to blood, the skin has a highly patchy distribution of parasites [45] resulting in bites taking up unequal quantities of parasites.

Second, the life cycle of the parasite. *Leishmania* parasites have a complex multi-stage life cycle in their sand fly vector that differs from that in their hosts. Once taken up as Amastigotes, the parasites must transition through at least 3 non-infectious forms to reach the infectious Metacyclic stage [48,49]. It was previously believed that the Metacyclic stage was the final life cycle stage, but recent work by Serafim *et al.* [54] identified another stage, the Retroleptomonad, responsible for amplifying the numbers of parasites available after subsequent blood meals. Thus, the timing of blood meals can be crucial to successful transmission and even sand flies with initially light infections can end up heavily infected [172].

Finally, the preferences and behaviour of the sand flies. Sand fly species have long been categorized as anthropophilic or zoophilic depending upon their feeding preferences [173], though which blood source is chosen depends on factors such as movement cost [174]. The combination of differing blood preferences [174], variable sand fly flight capabilities [68,69,175] and the necessity of finding a breeding site after each feed makes it hard to accurately model sand fly movements, especially with highly variable sand fly population sizes [176,177] and habitat preferences [178].

Capturing every facet within a single model is unfeasible. Other diseases, such as malaria, have seen more success in probing transmission dynamics [179,180] and the effects of key factors like climate [181] and immunity [182] using simpler approximations that avoid modelling the full intricacies of transmission. Similar models have seen success with leishmaniasis also, tackling areas such as zoonotic transmission [183], the relevance of asymptomatic individuals [184] and the effectiveness of intervention strategies [185]. A similarly focused approach will be beneficial to our understanding of leishmaniasis transmission.

One region of particular interest to modelling is the Indian Subcontinent (ISC). The ISC has historically been a highly endemic region for visceral leishmaniasis [3] with typically anthroponotic transmission. Emphasis has been placed on intervention strategies [185,202] since leishmaniasis was targeted for elimination in the ISC [3]. However, research has also highlighted the prevalence and perceptions of visceral leishmaniasis [147,203] and key risk factors underpinning its transmission [147]. More recently, Chapman *et al.* [146] used distance kernels fitted by a Markov-chain Monte Carlo (MCMC) algorithm to evaluate plausible relationships between infection risk and distance from existing infections, building upon earlier work of Bern *et al.* [147].

Although the MCMC method employed by Chapman *et al.* found considerable success, the approach has drawbacks. In many contexts MCMC methods are prohibitively expensive, requiring substantial computational power and time (though can be made more efficient, see [197,198] for example). Any chosen MCMC method must be tailored correctly to the context, otherwise they may not converge [199] or may suffer even greater inefficiencies [200]. The expertise needed to perform this tailoring often means it is more practical to use pre-built models instead, but these are often ‘black-box’ methods that can be difficult to interpret and optimise.

Survival analysis has the potential to be a simpler and computationally cheaper alternative, especially in the context of risk evaluation. Survival analysis has been applied to various situations such as equine influenza transmission [112], mosquito survival times [186] and mortality during a plague epidemic [187]. It has been used to explore the dispersal of *Lutzomyia longipalpis* through São Paulo State, Brazil [121] and the emergence of post kala-azar dermal leishmaniasis (PKDL) after a visceral leishmaniasis infection [119]. Packages for performing survival analysis are widely available and versatile (see, for example, the ‘survival’ package for RStudio [188]). Even older methods such as the Cox proportional hazards model [189] have seen a wide array of applications [190–192] and there has been much exploration of alterations to handle missing data [193] or large datasets [194], as well as other more recent alternative models [195,196]. The ease of choosing alternative methods and the flexibility of survival analysis packages renders it a useful, if underused tool.

Here we apply survival analysis to three data sets that focus on leishmaniasis prevalence in different regions and at multiple scales [147,149,204]. We evaluate the relationship between transmission risk and proximity to infected individuals with particular emphasis on the impact of sharing a household with an infected individual. We examine the link between social vulnerability and endemic visceral leishmaniasis in Brazil [149] and demonstrate how ‘individuals’ may refer to entities beyond single patients. Finally, we

also highlight and discuss the strengths and limitations of survival analysis, with a focus on identifying common pitfalls and strategies to improve the suitability of a dataset for use with survival analysis.

3.3 Data

The subsequent analysis focuses on three publicly available datasets of visceral and/or cutaneous leishmaniasis cases at different spatial and temporal scales. Though full details are provided elsewhere [147, 149, 201, 204, 205] we highlight the relevant methodology and recorded data here for clarity.

3.3.1 Dataset 1: Fulbaria thana cross-sectional study.

The primary dataset used by this study is a publicly available dataset [147] originally collected via cross-sectional surveys in Fulbaria thana, Mymensingh district, Bangladesh between January 2002 and June 2004 with additional details of historic cases of leishmaniasis dating back to 1998. The original studies applied logistic regression to this dataset [147, 201] and provide full details of the study design and data collection, we summarise the main details here.

The community is subdivided into nine sections known as “paras”, and the survey was restricted to the three paras with the highest incidence rate of visceral leishmaniasis over several years prior to the study. Only individuals who had lived within these paras for at least 6 months in the three years prior to the study were included. Where possible, the month of symptom onset, diagnosis, treatment, relapse and any relapse treatment were recorded for each incidence of visceral leishmaniasis (retrospectively for cases prior to 2002). Births and deaths were recorded by annual census. A Global Positioning System (GPS) was used to map all houses present in 2002 (accurate to +/- 10m), and all individuals were allocated the GPS positions of the household to which they belonged. These positions were used to calculate pairwise distances for each individual (with individuals in the same household considered to be 0m apart).

Information about potential risk factors such as bed net usage, presence of livestock, diet, house construction materials, as well as socio-economic factors such as income and land ownership were collected via surveys. Univariate analysis was initially used to identify significant risk factors, and multivariate models were constructed by stepwise addition of all factors that were significant in the univariate analysis. All models accounted for household clustering using Generalised Estimating Equations. Their key findings were that being aged between 3 and 45 years old and being in proximity to previous cases increased

the risk of developing visceral leishmaniasis, but that the risk was reduced by the presence of cattle and the consistent usage of bed nets (particularly during the summer). This is likely because both reduce human exposure to the sand flies: the bed nets physically prevent the sand flies from biting, and the cattle represent a more convenient and thus more appealing blood source for the sand flies [201].

For the dataset to be amenable to survival analysis, all visceral leishmaniasis cases need a symptom onset month. Cases with only an onset year were subsequently assigned a symptom onset month drawn from a uniform random variable such that it lies within the onset year but before any diagnosis/death month and after any birth month, where relevant (in line with the multiple imputation approach [206,207]). Cases recorded as prior infections (i.e. infections which occurred before 1999) were not included in the analysis, but were considered for the purposes of determining if a household contains an infection.

3.3.2 Dataset 2: Global leishmaniasis incidence database.

Section 3.5.2 provides two examples of the wider applications of survival analysis. The first part of Section 3.5.2 examines a database of global leishmaniasis cases (both cutaneous and visceral) assembled by Pigott *et al.* [204] for the purpose of producing global leishmaniasis distribution maps [205]. For full details of its creation see [204, 205], we summarise the key methods and findings.

The database contains a list of cutaneous and visceral leishmaniasis cases, as well as any reported instance of post Kala-azar dermal leishmaniasis (PKDL), between 1960 and 2012, derived from literature searches via PubMed and Web of Knowledge. Only autochthonous cases were included, with imported cases excluded if they could not be traced to their origin [204]. For each case, several details are recorded: the year of onset, the country in which it occurred, an associated latitude and longitude, and information about whether it was cutaneous or visceral leishmaniasis (PKDL cases were considered to be visceral) are the most relevant to our analysis. Additionally, each entry is categorised as either a ‘point’ (where a specific location was known) or a ‘polygon’ (where it is ascribed to a particular region and the coordinates are those of the centroid of the polygon encompassing that region).

The evidence of the presence of both cutaneous and visceral leishmaniasis cases was used to produce global distribution maps. Each map was subdivided into administrative regions, each of which was coloured according to the consensus on whether leishmaniasis was present in the region. In addition, boosted regression trees [208] incorporating pseudo-data [209] were used to predict the occurrence of leishmaniasis, and determine key risk factors. These methods identified that people living in peri-urban environments as well as

areas with greater vegetation cover were most at risk from both forms of leishmaniasis.

3.3.3 Dataset 3: Visceral leishmaniasis incidence rates in northeastern Brazil.

The second part of Section 3.5.2 focuses on visceral leishmaniasis incidence in the Brazilian Northeast region between 2000 and 2017, using a dataset originally assembled and analysed by Ribeiro *et al.* [149]. For full details consult the original paper [149], we summarise the key methods and findings below.

The database contains annual population and visceral leishmaniasis case counts for each municipality in the Brazilian Northeast region, along with x and y coordinates and the social vulnerability index (SVI), drawn from [210]. The SVI is represented by an overall SVI score and by its three components: urban infrastructure (access to basic sanitation, commuting times), human capital (access to education and childcare) and income/work (dependence on child/elderly labour, availability of permanent employment). For each, a higher SVI value represents greater vulnerability, with 0 being an ideal scenario and 1 being the worst possible scenario [210]. The overall SVI score is an average of the component scores.

Long-term changes to leishmaniasis incidence rates were identified using segmented linear regression with statistical significance identified via the Monte Carlo permutation test, following the methodology of prior studies [211, 212]. This identified a long-term increase in the proportion of municipalities infected with leishmaniasis but a reduction in the prevalence within the general population (alongside other trends specific to factors such as age and sex). Focus was also put on identifying spatial and spatiotemporal trends in visceral leishmaniasis transmission. The crude transmission rates were smoothed using the local Bayesian empirical method, and then spatial autocorrelation was identified using the global Moran's I index [213]. Subsequently, the local Moran's index (LISA) was used to identify spatial clusters of municipalities with high transmission rates. Together, these methods identified that although visceral leishmaniasis transmission is broadly distributed across the region, there are high risk clusters in six states.

Spatiotemporal cluster analysis was performed using Kulldorff's retrospective space-time scan, following prior studies [214–216]. This analysis identified 12 significant spatiotemporal clusters, with 7 clusters located in Bahia state, and the primary cluster distributed across the centre and west of the region. Finally, bivariate spatial cluster analysis, incorporating the Spearman's correlation test and bivariate LISA, was used to identify correlations between the SVI indices and leishmaniasis transmission [213]. This identified a number of significant clusters, in particular a large cluster of high SVI-high transmission

municipalities in the north west of the region and a number of low SVI-low transmission clusters in the north east of the region.

3.4 Survival Analysis Methods

In survival analysis, we are typically interested in how the ‘survival time’ (the time taken for an individual to experience an event of interest) of an individual is affected by various factors known as ‘risk factors’. To investigate this, we introduce two functions: the survivor function $S(t)$ and the corresponding hazard function $h(t)$. It is informative to discuss these functions briefly in general terms before introducing the specifics relevant to our case studies. For a more thorough overview, consider reading [111, 217, 218].

Definition 1 (Survivor Function). The survivor function $S(t)$ is the probability that an individual survives longer than time t . For individual survival time T :

$$S(t) = Pr(T > t).$$

The survivor function is also referred to as the survival function. A useful counterpart to this is the hazard function:

Definition 2 (Hazard Function). The hazard function $h(t)$ is the instantaneous potential for an individual to experience an event given that the individual has survived until time t . For individual survival time T :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | t \leq T)}{\Delta t}.$$

Thus, the hazard increases when an individual is more likely to experience an event of interest. It is also to introduce now the cumulative hazard $H(t)$:

Definition 3 (Cumulative Hazard). The cumulative hazard is the total hazard experienced by an individual up to time t :

$$H(t) = \int_0^t h(u) du.$$

These three functions are linked, and one can be used to derive the other two:

$$\begin{aligned}
h(t) &= \frac{-d \log(S(t))}{dt} \\
H(t) &= -\log(S(t)) \\
S(t) &= \exp(-H(t))
\end{aligned}$$

Usually we place emphasis on the hazard function because we are mainly interested in the hazard experienced by an individual over a given time period. The exact form taken by the hazard function depends on the context in which it is used, but it must possess the following characteristics:

- $h(t) \geq 0$
- $h(t)$ has no upper bound.

This study performs survival analysis with time-dependent covariates using the Cox proportional hazards model extended to incorporate both time-independent and time-dependent covariates. When the hazard depends on one or more covariates, we label it $h(X, t)$ instead of $h(t)$. For the extended Cox proportional hazards model, the hazard function takes the form:

$$h(X, t) = h_0(t) \exp \left(\sum_{i=1}^{\infty} \beta_i X_i + \sum_{j=1}^{\infty} \delta_j X_j(t) \right) \quad (3.1)$$

where β_i is a constant coefficient corresponding to time-independent covariate X_i , δ_j is a constant coefficient corresponding to time-dependent covariate X_j , and $h_0(t)$ is a generic base hazard function.

In some situations, one may wish to consider the possibility of interactions between factors. Such interactions terms are possible and reasonably intuitive; consider the following example with two interacting covariates:

$$h(X, t) = h_0(t) \exp (\beta_1 X_1 + \beta_2 X_2 + \gamma X_1 X_2) \quad (3.2)$$

where γ is the constant coefficient corresponding to the interaction between X_1 and X_2 .

All coefficients are fitted via maximum likelihood procedure. In the case of the Cox proportional hazards model, we maximise the partial likelihood L . For a system with k possible failure times, L is defined to be:

$$L = L_1 \times L_2 \times L_3 \times \dots \times L_k = \prod_{j=1}^k L_j \quad (3.3)$$

where L_j is the likelihood of failure occurring at time j , given the risk set $R(t_j)$, the set of individuals at risk of failure at time j [111]. Usually, the natural log of L is maximised by solving the system of equations:

$$\frac{\partial \log L}{\partial \beta_i} = 0 \tag{3.4}$$

for all parameters β_i . We start with an initial estimate of each β_i (which is often but not necessarily 0) and then solve the system of equations represented by Equation 3.4 by gradient descent.

Initial analysis is performed using univariate analysis, with the models taking the form:

$$h(X, t) = h_0(t) \exp(\beta X)$$

where X is a binary variable indicating the presence of an infected individual with a specified distance band, and β represents the associated log hazard ratio. Additionally, we utilise a multivariate model involving all factors identified as significant by the univariate analysis. This is an extended version of Equation 3.2, taking the form:

$$h(X, t) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \gamma_{12} X_1 X_2 + \gamma_{13} X_1 X_3 + \gamma_{23} X_2 X_3 + \gamma_{123} X_1 X_2 X_3)$$

where β_i is the log hazard ratio associated with variable X_i and γ_{ij} , for example, is associated with the interaction between variables X_i and X_j (in this context, the additional hazard of having an individual in both band i and j).

3.5 Results

3.5.1 Case proximity in Fulbaria thana.

We begin our analysis by focusing on the first dataset, the leishmaniasis case data from Fulbaria thana, Bangladesh (see Section 3.3.1). Initially we restrict our analysis to a simple scenario: a single binary indicator function (indicating the presence of an infection within a specified distance of an individual) as our only covariate. Following the lead of Bern *et al.* we consider the following distances: within-household, 0-25m, 25-50m, 50-75m and 75-100m. Table 3.1 contains the hazard ratios and associated P-values.

Firstly, consider the result for the presence of an infected household member. We find that having an infected individual in your household increases your hazard of infection

Factor	Hazard Ratio	95% CI	P-value
Infected household member	2.2355	(1.9039,2.6251)	< 0.001
Infection within 25m	1.5084	(1.316,1.7288)	0.00259
Infection between 25m and 50m	1.4832	(1.3006,1.6915)	0.00271
Infection between 50m and 75m	1.0973	(0.955,1.2608)	0.504
Infection between 75m and 100m	1.1536	(1.009,1.3189)	0.288

Table 3.1. Individuals in close proximity to an infected individual experience greater hazard of infection. Summary of hazard ratios, confidence intervals and p-values for each covariate fitted via univariate analysis.

approximately two-fold. We can see the effect more clearly by plotting the cumulative hazard for individuals with or without infected households (Figure 3.1). We notice that, as suggested by our hazard ratio, the cumulative hazard of infection for those in an infected household is approximately double that of individuals without an infection in their household. Although this figure corresponds well to the hazard function fitted by the Cox proportional hazards model, it is typically more informative to look at the survival of individuals instead (Figure 3.2).

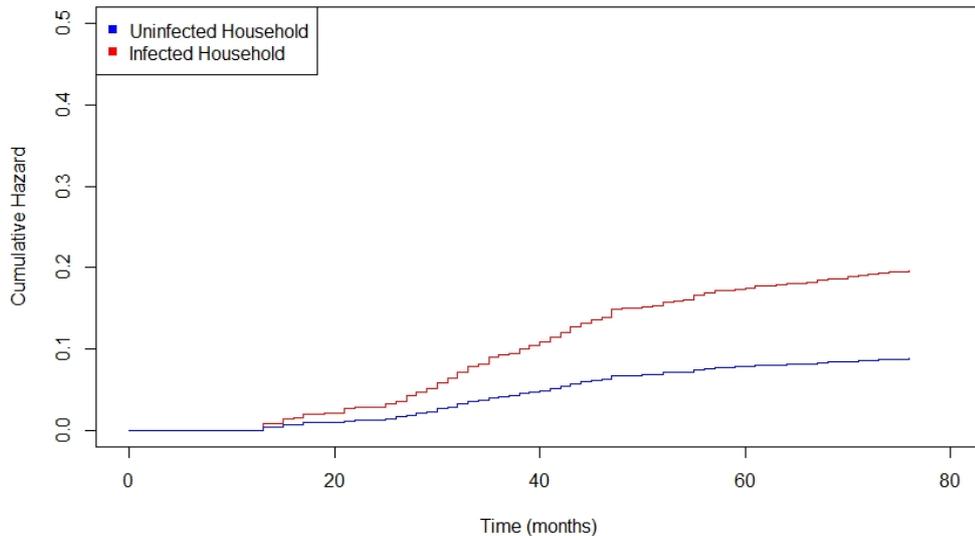


Figure 3.1. Individuals sharing a household with an infected individual are more likely to become infected than those that do not. Cumulative hazard of infection for individuals with (red) or without (blue) an infected individual in their household.

Figure 3.2 shows that individuals in households with an active infection experience reduced survival, that is to say that such individuals become infected faster than their peers in uninfected households. We next consider the same graph but for our second covariate,

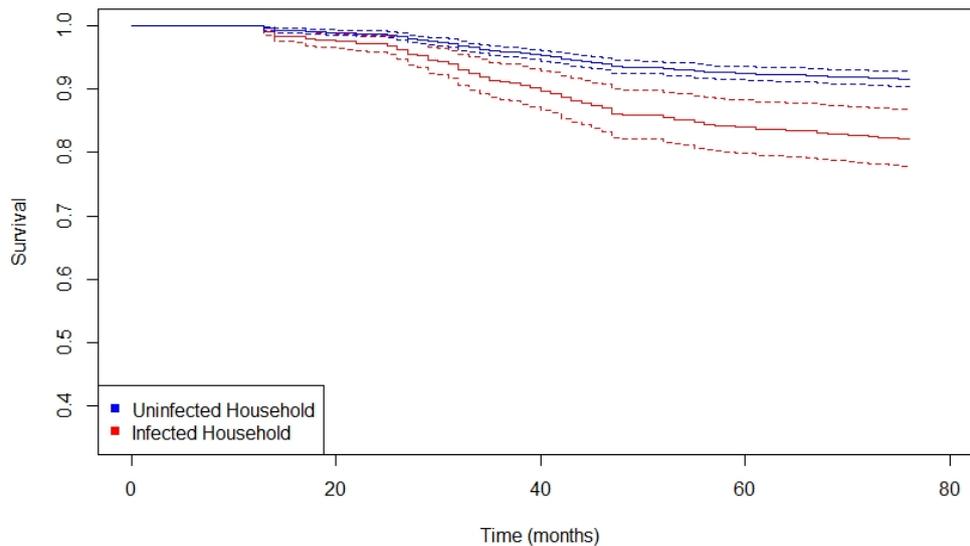


Figure 3.2. A greater proportion of individuals sharing a household with an infected individual are infected at a given time than those who do not. Survival of individuals in households with (red) or without (blue) infections. Solid line is mean survival, dotted lines are ± 2 s.e.

the presence of an infection between 1 and 25m away, resulting in Figure 3.3. Once again, we observe that survival decreases faster for individuals with an infection between 1 and 25m away, albeit at a slower rate than we observed when considering the presence of an infection in the household. This would be expected given the slightly reduced hazard ratio (approximately 1.5 versus approximately 2.2). Thus, we may conclude that although having an infection between 1 and 25m away increases your hazard of infection, it does not increase it as much as having an infection within your household.

Considering the final significant covariate, the presence of an infection between 25 and 50m away, leads to Figure 3.4. We once again see decreased survival for individuals with an infection present in our chosen distance band. In this case, we conclude that having an infection between 25 and 50m away has almost the same effect as having an infection between 1 and 25m away.

It would be informative to see how the survival curves differ if we consider a non-significant covariate, such as the presence of an infection between 50 and 75m away; the results of such an analysis are shown in Figure 3.5. In this case, we observe that although the survival curves differ slightly the mean survival for each consistently lies within 2

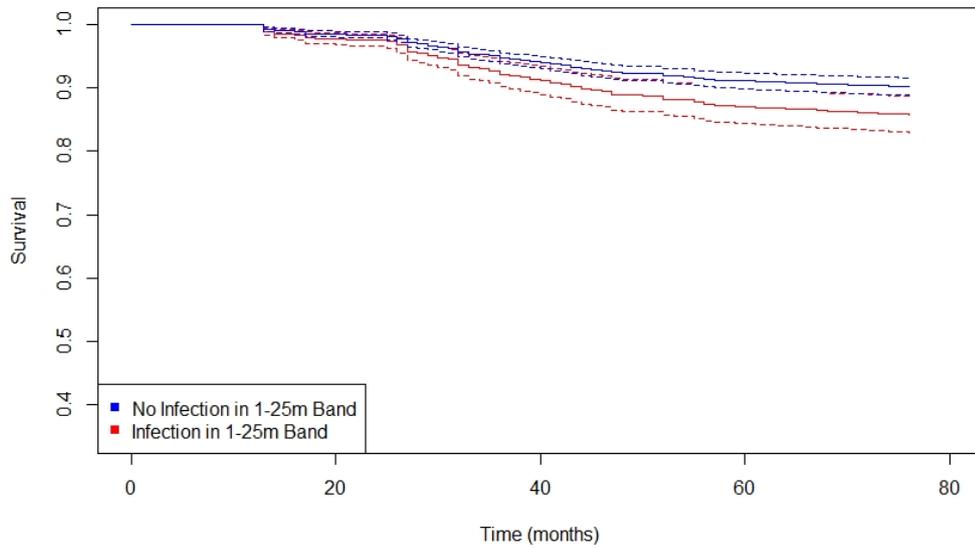


Figure 3.3. Individuals are infected at a greater rate if a prior infection exists within 25m than if one does not. Survival of individuals with (red) or without (blue) an infected between 1 and 25m away. Solid line is mean survival, dotted lines are ± 2 s.e.

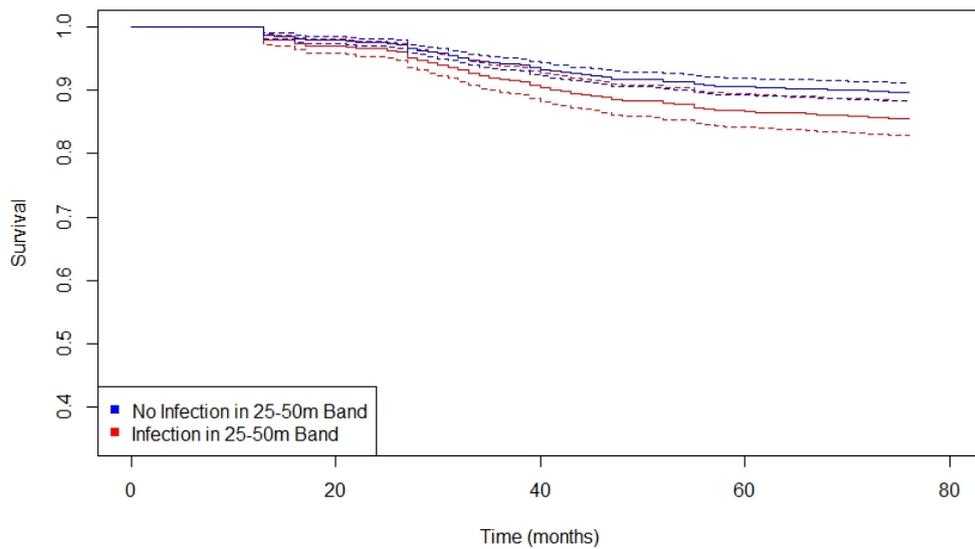


Figure 3.4. Individuals are infected at a greater rate if a prior infection exists between 25 and 50m away than if one does not. Survival of individuals with (red) or without (blue) an infected between 25 and 50m away. Solid line is mean survival, dotted lines are ± 2 s.e.

s.e. of the other, as would be expected from our p-values and hazard ratio (which is approximately 1). This strongly suggests that having an infection between 50 and 75m away has minimal impact on your hazard of infection.

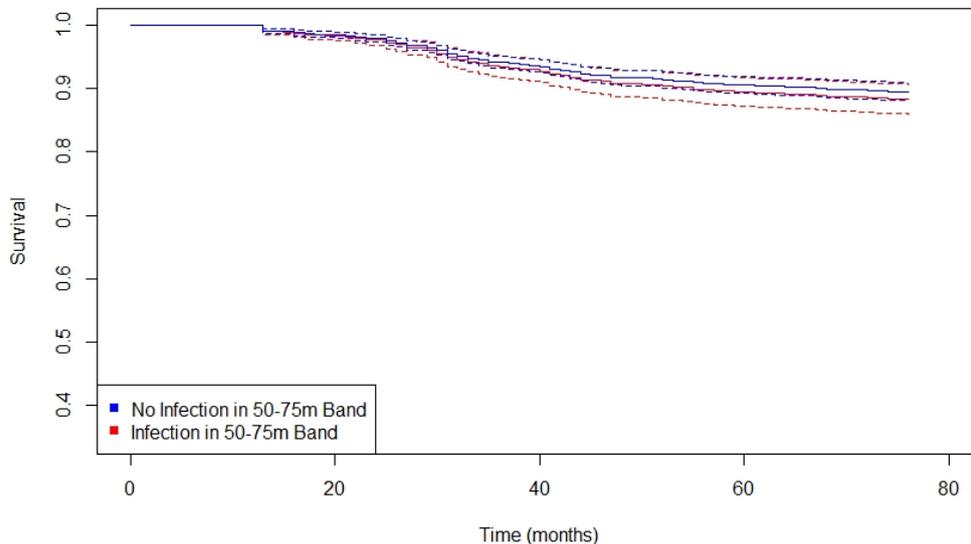


Figure 3.5. Individuals experience similar hazards of infection irrespective of the presence of an infection between 50 and 75m away. Survival of individuals with (red) or without (blue) an infected between 50 and 75m away. Solid line is mean survival, dotted lines are ± 2 s.e.

Our univariate analysis provides an informative way to interpret a set of models, allowing insights into risk factors of interest. However, it is important to consider the possibility of interactions between different factors. To do so, we now assemble a multivariate model using the covariates we identified as significant in our univariate analysis. Table 3.2 contains the final model hazard ratios and associated p-values:

Factor	Hazard Ratio	95% CI	P-value
Infected household member (A)	3.35	(2.6551,4.227)	< 0.001
Infection within 25m (B)	2.3579	(1.9915,2.7918)	< 0.001
Infection between 25m and 50m (C)	2.0822	(1.7819,2.4329)	< 0.001
Interaction term: AB	0.6737	(0.4572,0.9929)	0.30852
Interaction term: AC	0.5688	(0.4068,0.7953)	0.09232
Interaction term: BC	0.4832	(0.3758,0.6213)	0.00382
Interaction term: ABC	1.3428	(0.8053,2.2394)	0.5644

Table 3.2. Hazard of infection does not increase additively due to infections in multiple distance bands. Summary of hazard ratios, confidence intervals and p-values for each covariate and all interaction terms, fitted via multivariate analysis.

As with the univariate analysis we find that all three covariates increase the hazard of infection. However, we observe that the magnitude of this increase is greater for all covariates. This is because in univariate analysis we cannot control for the presence of other ‘confounding’ covariates of interest that could alter the hazard ratio [111]. For example, when we calculated the hazard ratio for the presence of an infected housemate in our original univariate analysis, the presence of an infection within 25m would have been a confounding covariate. The multivariate analysis allows us to control for these factors, and thus gives a more accurate (and in this case higher) estimate for the hazard ratio [111].

Additionally, we find that there exists an interaction between two of the covariates: the presence of an infection within 25m, and the presence of an infection between 25-50m away. As the interaction reduces the hazard of infection, we may conclude that the presence of an infection in both bands does not increase the hazard of infection that far beyond having an infection in only one of the bands.

Although we have thus far omitted any time delay from our analysis, it is important to note that leishmaniasis has an incubation period, often thought to be approximately 1-6 months in length [146, 147, 288]. An incubation period can be incorporated into the hazard function of the Cox proportional hazards model; the resulting hazard function for a system with a single covariate then takes the form:

$$h(X, t) = h_0(t) \exp \delta X(t - \tau) \tag{3.5}$$

where δ is the coefficient corresponding to covariate X , $h_0(t)$ is the base hazard function and τ is the length of the delay.

In order to assess the impact of such an incubation period, we now introduce a time-delay of 4 months to our model. The hazard ratios and associated p-values can be found in Table 3.3:

Factor	Hazard Ratio	95% CI	P-value
Infected household member	2.1734	(1.8281,2.5839)	< 0.001
Infection within 25m	1.5419	(1.3371,1.778)	0.00238
Infection between 25m and 50m	1.5107	(1.3142,1.7367)	0.00308
Infection between 50m and 75m	1.1347	(0.9822,1.3108)	0.3839
Infection between 75m and 100m	1.1492	(0.999,1.3221)	0.3234

Table 3.3. The relationship between infection proximity and hazard of infection is unmodified by incorporating an incubation period. Summary of hazard ratios, confidence intervals and p-values for each covariate fitted via univariate analysis, with the inclusion of a delay of 4 months.

We observe no changes to which covariates achieve significance, and only slight variations in the hazard ratios, with all new values remaining comfortably within the original 95% confidence intervals. Figure 3.6 shows the survival for individuals with or without an infected household member.

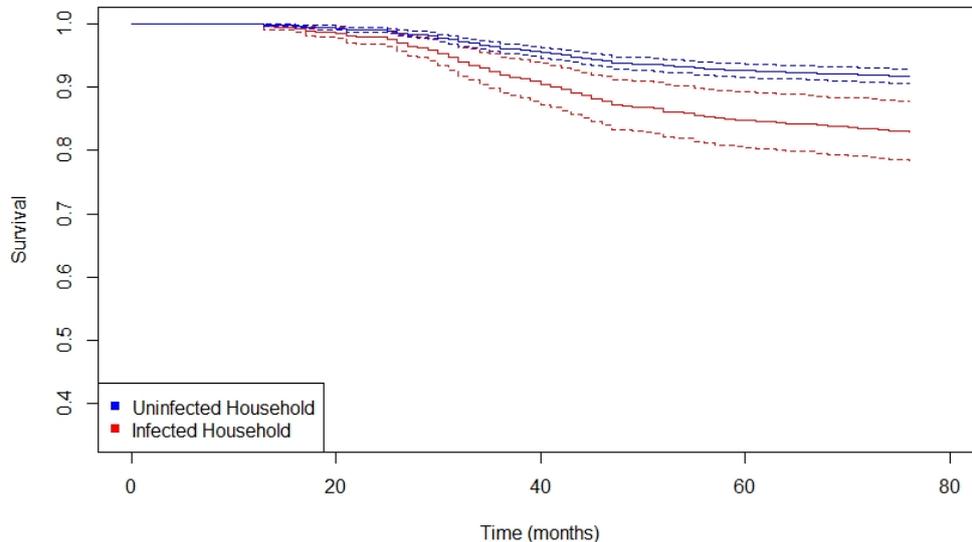


Figure 3.6. Individuals with an infected household member experience greater cumulative hazard of infection, even with a delay. Survival of individuals with (red) or without (blue) an infected household member, with a 4 month delay. Solid line is mean survival, dotted lines are ± 2 s.e.

Mirroring the results for the same covariate without the 4 month delay, we observe that individuals sharing a household with an infected individual experience reduced survival, and we observe that the mean survival for one category consistently falls outside of ± 2 s.e. of the other. If we compare this to the original version without the delay, Figure 3.7, we observe that the delay introduces minimal differences between the survival curves. The mean survival curves with the delay included lie comfortably within ± 2 s.e. of their counterparts without the delay. Thus, we may conclude that incorporating a delay into the model did not introduce any major changes to the hazard of infection, and can continue to use a delay of 0.

3.5.2 Wider applications of survival analysis

In order to demonstrate the variety of potential applications of survival analysis, we will now apply it to other datasets. Although it has seen similar application with other diseases

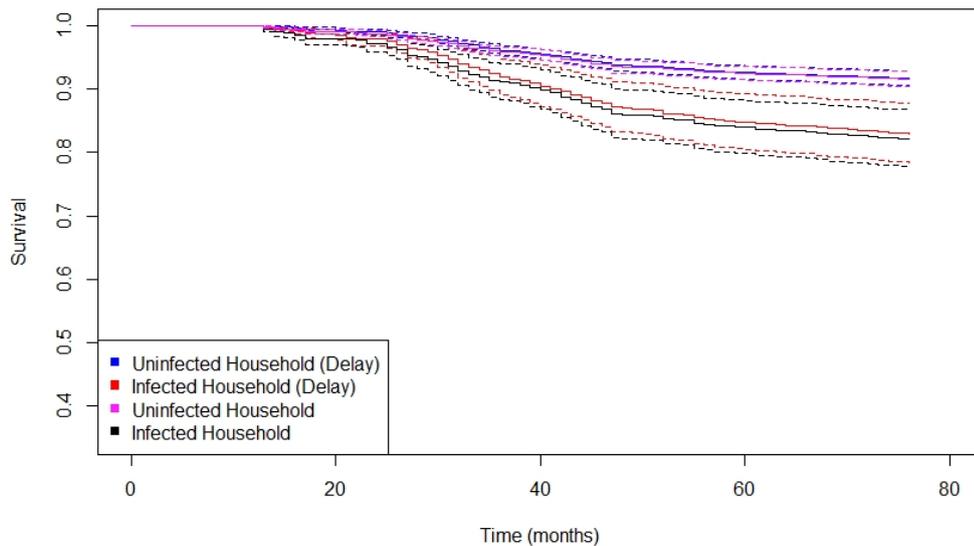


Figure 3.7. The cumulative hazard experienced by an individual is qualitatively similar with or without a delay. Survival of individuals with or without an infected household member, comparing the scenario with a 4 month delay with the scenario without. Solid line is mean survival, dotted lines are ± 2 s.e.

such as equine influenza [112], studies involving leishmaniasis have rarely used it (see [121] for one example). To address this, we now direct our attention to two other publicly available datasets. First we will demonstrate that pitfalls can accompany even a well-formatted dataset such as the global case database [204], and then we will show that datasets which initially appear ill-suited to survival analysis [149] can be effectively used if handled carefully. We will demonstrate that, with careful consideration, survival analysis is flexible enough to apply to a variety of datasets and yield meaningful insights.

Application 1: Global leishmaniasis incidence.

The first dataset we examine is the global leishmaniasis case database assembled by Pigott *et al.* [204] (see Section 3.3.2). For the purposes of broadening the scope of our survival analysis application, we will use analyse the cutaneous leishmaniasis data. We also restrict our analysis to include only cases of the ‘point’ category. For this analysis we will use the onset year as the survival time for each individual (but see Section 3.5.3) and their latitude as a covariate of interest. The latitude is a continuous variable, therefore we will fit a penalised spline (with the degrees of freedom determined using the AIC) to determine if there is a nonlinear relationship between it and the hazard of infection.

We first consider the latitude, with the results shown in Figure 3.8. This spline in-

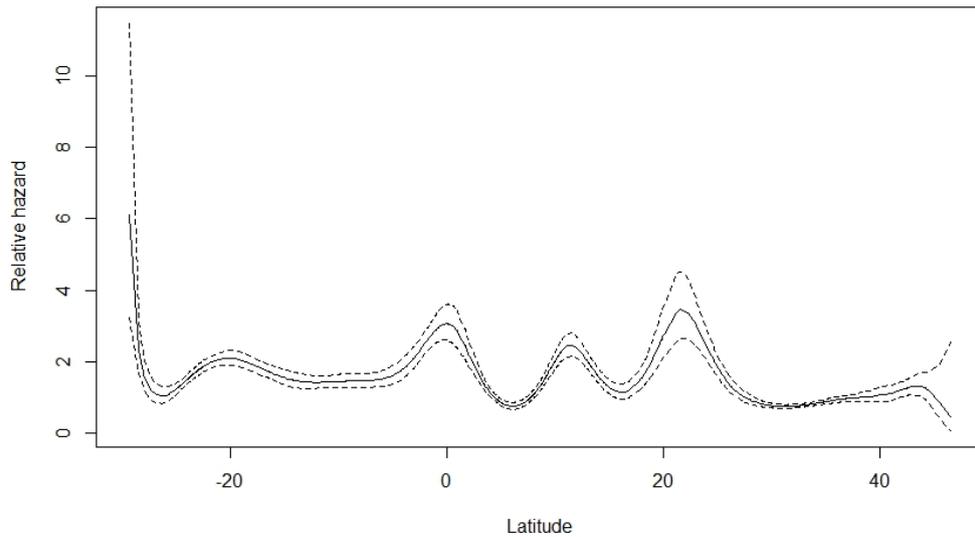


Figure 3.8. Peaks in relative hazard of infection correspond to regions of emphasis in the literature, not necessarily areas of greater hazard. Relative hazard against latitude for cutaneous leishmaniasis. Solid line is mean hazard, dotted lines are ± 2 s.e.

indicates a few peaks in relative hazard, the three most prominent located approximately at latitudes of 0, 13, and 22. We also note an increase at approximately -30, however this is accompanied by very large standard error (and can be attributed to a lack of data). These results must be interpreted with caution. Leishmaniasis is particularly prevalent in India [3], so the peak at a latitude of 20 would be expected. However, there is a trough located at approximately latitude 8 coinciding with Ethiopia, which is actually another known hotspot [295]. Hazard also remains low beyond 20 latitude, despite the presence of leishmaniasis in the Mediterranean [233]. Indeed, the peaks appear to coincide with older, well-documented hotspots.

This is in part due to our somewhat naive assumption that the onset year would make a viable event time (and thus would be suitable for determining the survival times). Unfortunately, this is not the case. The studies from which these cases were drawn did not occur in parallel. Thus, it would be inaccurate to take 1960 as the start time for every individual, and as such the majority of the survival times we used are artificially inflated. Alas, it is not possible to correct the start times using this dataset alone and we should instead consider the dataset unsuitable for survival analysis. This example hence serves to highlight that although a dataset may appear very amenable to survival analysis at first glance, care must be taken to assess its suitability before drawing conclusions. Perhaps future work could collect a larger, suitably formatted dataset and use it to create a map of global leishmaniasis infection risk (which could yield insights if combined with more a

more traditional map of disease prevalence).

Application 2: Endemicity in northeastern Brazil.

We now redirect our attention to another publicly available dataset, the leishmaniasis incidence rates for the municipalities of northeastern Brazil [149] (see Section 3.3.3). At a glance, this dataset may appear unsuitable for survival analysis as we do not track individual cases and survival times. In reality, this is not necessarily a barrier: we may consider the municipalities as individuals instead. For this to be possible, there are a couple of components we must suitably define.

Firstly, we require a suitable ‘failure criterion’ that we can use to determine the ‘survival time’ of each municipality. Naturally, the presence of an infection would make for a poor criterion since this is a region with high visceral leishmaniasis mortality [26], and thus many of the municipalities would fail at $t = 0$. Instead, we follow the example of Ribeiro *et al.* [149] and use the Brazilian government categories for leishmaniasis transmission as the basis of our criterion. For completeness, we shall consider a few viable thresholds and, in the absence of any noticeable differences, will then focus on a single one.

In 2019, the Brazilian government proposed a scheme to categorise municipalities into different intensities of visceral leishmaniasis transmission, based on the most recent 3-year average incidence rate for each municipality [219]. The categories are constructed as follows: sporadic transmission (< 2.4 cases per 100000 people), moderate transmission ($2.4 - 4.4$ cases per 100000 people), and intense transmission (≥ 4.4 cases per 100000 people). Informed by these categories, we use the following candidates for our threshold: 2.0 cases per 100000, 3.0 cases per 100000, and 4.5 cases per 100000 people (one within each category of transmission). To facilitate the use of a 3-year average, we now choose the reference point $t = 0$ to be 2003 (with the data for 2000, 2001, and 2002 used to calculate the first average), and omit any municipality for which the first 3-year average exceeds our chosen threshold. Throughout, we will refer to any municipality exceeding our threshold as ‘endemic’ and will typically refer to ‘hazard of endemicity’ rather than hazard of infection as we did previously.

We also need to choose suitable covariates for our analysis. For this analysis we shall initially focus on the SVI of the municipalities (we will consider covariates such a proximity to ‘endemic’ municipalities later). To ensure we capture the full picture, we will consider both the overall SVI and the individual components in separate analyses.

First we briefly examine the distribution of SVI values for the municipalities, shown in Figure 3.9. The overall SVI, the human capital component, and the income/work component all display a similar distribution: the vast majority of municipalities lie between

0.25 and 0.75, clustering strongly around 0.5, with a much smaller number of outliers. For the urban infrastructure component, the majority of municipalities lie between 0.0 and 0.5, with a long tail extending towards 1.0.

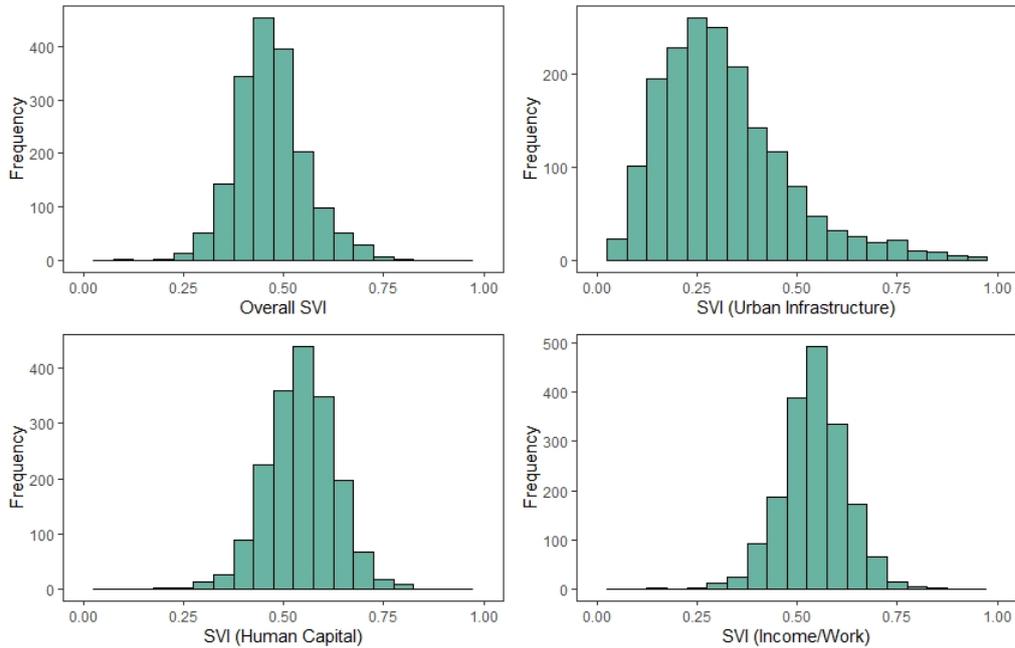


Figure 3.9. Distribution of overall SVI (top left), urban infrastructure SVI (top right), human capital SVI (bottom left) and income/work SVI (bottom right) values for all municipalities in northeastern Brazil.

We will consider the overall SVI for our initial analysis. SVI is a continuous variable, and although municipalities are categorised depending on their SVI value [210], we shall instead fit penalised splines (in the same manner as for the previous dataset) this time with a fixed 3 degrees of freedom. Throughout, we will consider a municipality to be endemic when it exceeds 3.0 cases per 100000 people. For details of the other endemic thresholds we consider see Section A.2.

Component	Coefficient	Standard Error	Chi Squared	P-Value
Linear	0.871	0.448	3.783	0.052
Non-linear	NA	NA	3.878	0.151
Likelihood Ratio Score:	9.95	P-Value:	0.02	

Table 3.4. The relationship between overall SVI and hazard of endemicity is linear. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.

The fitted coefficients and associated p-values for the spline can be found in Table 3.4. First, a comment on the p-values reported here. We may conclude from the likelihood

ratio test that there is a significant relationship between overall SVI and hazard, but we do not see this reflected in the other p-values. This is likely because the p-values for the linear and non-linear components are based on the Wald statistic instead. Since the likelihood ratio test is typically more reliable [111], and the p-value for the linear component is very nearly significant, we should accept the verdict of the LR test. Thus, we conclude that the hazard increases almost linearly with overall SVI.

If we examine the resulting spline (Figure 3.10a) we observe that the hazard of endemicity increases with overall SVI, although there is a slight decrease at the upper reaches of the SVI scale. The decrease coincides with a region of minimal data which is likely to have impacted the fitting of the spline (Figure 3.10b). Overall, we conclude that municipalities with greater social vulnerabilities are more likely to become endemic for visceral leishmaniasis within a given time period. Similar behaviour is observed for the other two thresholds (see Section A.2).

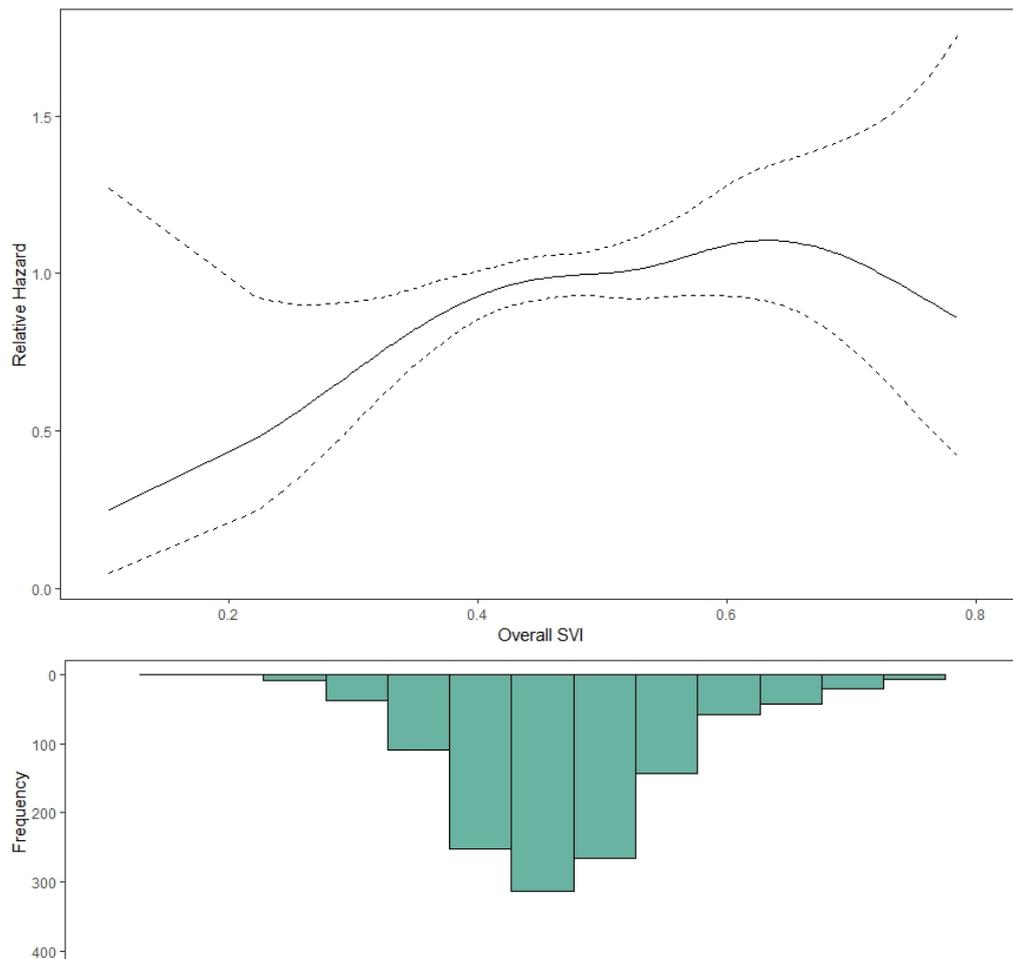


Figure 3.10. The linear relationship remains visible, though weakened, under the increase endemic threshold. Relative hazard against overall SVI, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are ± 2 s.e. (top) along with a histogram of the overall SVI values for all municipalities included in the analysis (bottom).

As the overall SVI is an average of its three components, there may be greater nuance present in the individual components than is captured by the overall SVI spline. Therefore, we will now consider each component individually, starting with the ‘urban infrastructure’ component. Table 3.5 shows the fitted coefficients and associated p-values. We conclude from this that there is a non-linear relationship between urban infrastructure SVI and the hazard of endemicity. For further details, we must examine the resulting spline (Figure 3.11a). Until approximately $SVI - UI = 0.6$, we observe that the hazard increases with the SVI-UI values. After that, it decreases sharply (with greatly increased standard error, as before, due to the lack of data). In fact, this pattern quite strikingly resembles the pattern for overall SVI.

Component	Coefficient	Standard Error	Chi Squared	P-Value
Linear	0.333	0.23	2.097	0.148
Non-linear	NA	NA	6.125	0.049
Likelihood Ratio Score:	9.75	P-Value:	0.02	

Table 3.5. The relationship between urban infrastructure SVI and hazard of endemicity is non-linear. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.

The second component of the SVI is the ‘human capital’ SVI: a measure of access to education and childcare, and for individuals to advance through and integrate with society. If we perform the same analysis as before (Table 3.6) we conclude that there is a non-linear relationship between SVI-HC and hazard of endemicity. If we examine the resulting spline (Figure 3.12a) we see the same qualitative relationship as before, though with a more definite peak at approximately 0.55 SVI-HC. We conclude that the hazard increases with increasing SVI-HC: municipalities with poorer access to childcare and education are more susceptible to the spread of visceral leishmaniasis.

Component	Coefficient	Standard Error	Chi Squared	P-Value
Linear	0.35	0.497	0.495	0.48
Non-linear	NA	NA	9.235	0.01
Likelihood Ratio Score:	12.4	P-Value:	0.006	

Table 3.6. The relationship between human capital SVI and hazard of endemicity is non-linear. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.

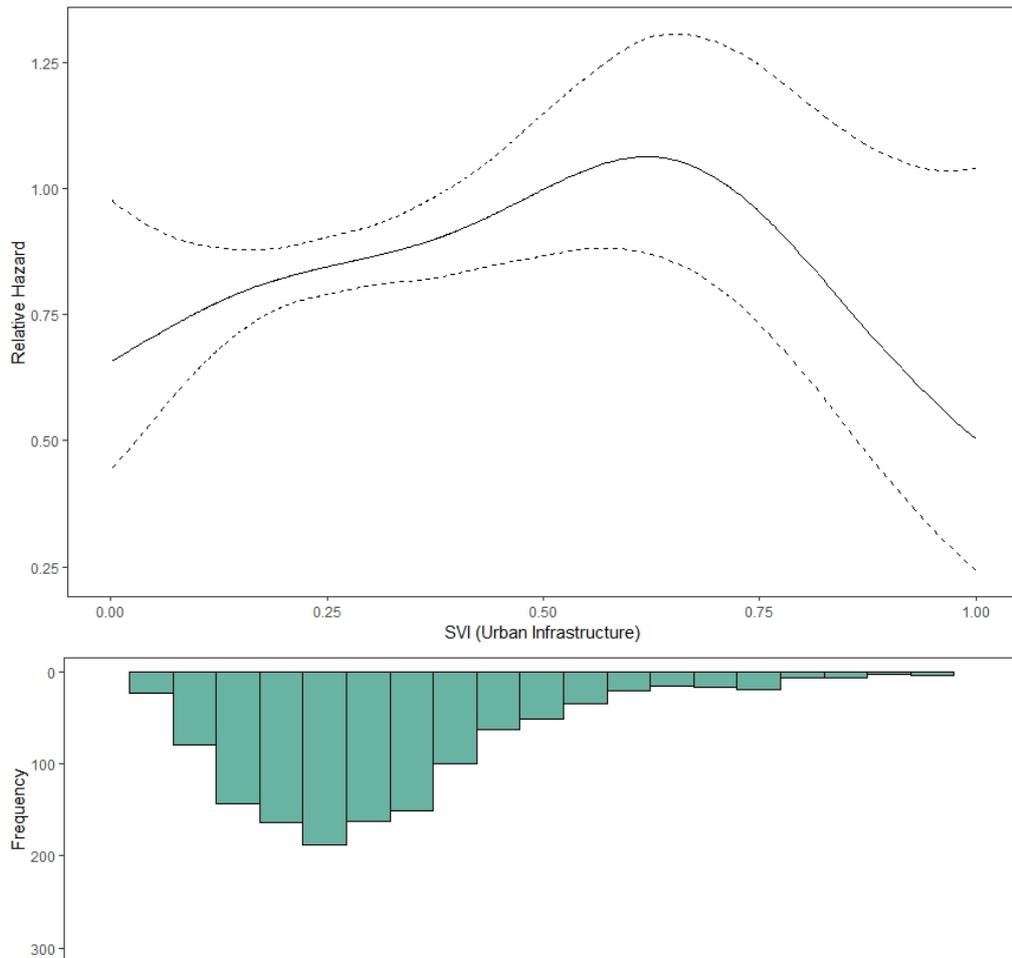


Figure 3.11. The relative hazard of endemicity appears to peak at an SVI-UI of 0.6. Relative hazard against the urban infrastructure SVI component, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are ± 2 s.e. (top) along with a histogram of the urban infrastructure SVI values for all municipalities included in the analysis (bottom).

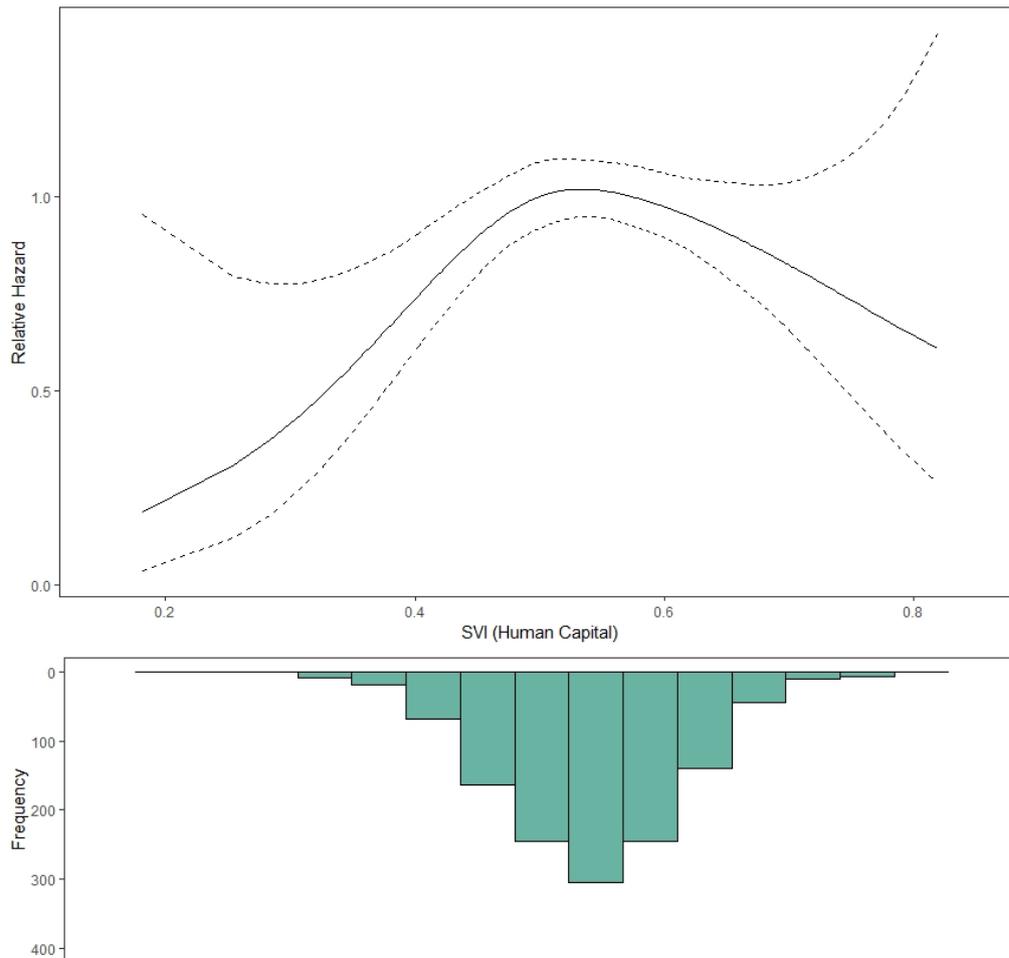


Figure 3.12. The relative hazard peaks at approximately 0.5 SVI-HC. Relative hazard against the human capital SVI component, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are ± 2 s.e. (top) along with a histogram of the human capital SVI values for all municipalities included in the analysis (bottom)

Finally, let us consider the ‘income/work’ component of SVI (Table 3.7). Unlike the previous two components, we would conclude that there is an entirely linear relationship between hazard of endemicity and SVI-I/W. We can see this more clearly in the resulting spline (Figure 3.13a). This spline shows a much simpler relationship than the previous ones. Hazard of endemicity increases almost linearly with SVI-I/W as anticipated (although we still observe the greatly increased standard error at each end of the parameter scale). Thus, we conclude that municipalities with poorer access to stable employment are more vulnerable to the spread of leishmaniasis.

Component	Coefficient	Standard Error	Chi Squared	P-Value
Linear	1.53	0.5	9.41	0.0022
Non-linear	NA	NA	1.18	0.5622
Likelihood Ratio Score:	11.6	P-Value:	0.009	

Table 3.7. The relationship between hazard of endemicity and income/work SVI is linear and positive. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.

It is important to establish whether there exist correlations between the different SVI components, as this could partially explain the similarities in the hazard of endemicity for each component. We test for such correlations between each pair of SVI components using Pearson’s product moment correlation coefficient (Table 3.8), which shows that all 3 components are positively correlated to a greater or lesser extent. This can also be observed if we plot the SVI component scores as a scatter plot (Figure 3.14).

Component 1	Component 2	Correlation	Confidence Interval	P-Value
SVI-UI	SVI-HC	0.4013	(0.3617, 0.4395)	< 0.0001
SVI-UI	SVI-I/W	0.2842	(0.241, 0.3262)	< 0.0001
SVI-HC	SVI-I/W	0.5468	(0.5134, 0.5784)	< 0.0001

Table 3.8. The components of the SVI index are all positively correlated with each other. Pearson’s product moment correlation coefficient, with associated confidence interval and p-values, for all pairwise inter-component comparisons.

Next, we will briefly focus our analysis on some time-dependent covariates. In a similar manner to our analysis of the Bern *et al.* dataset (Section 3.3.1), we will examine the effect

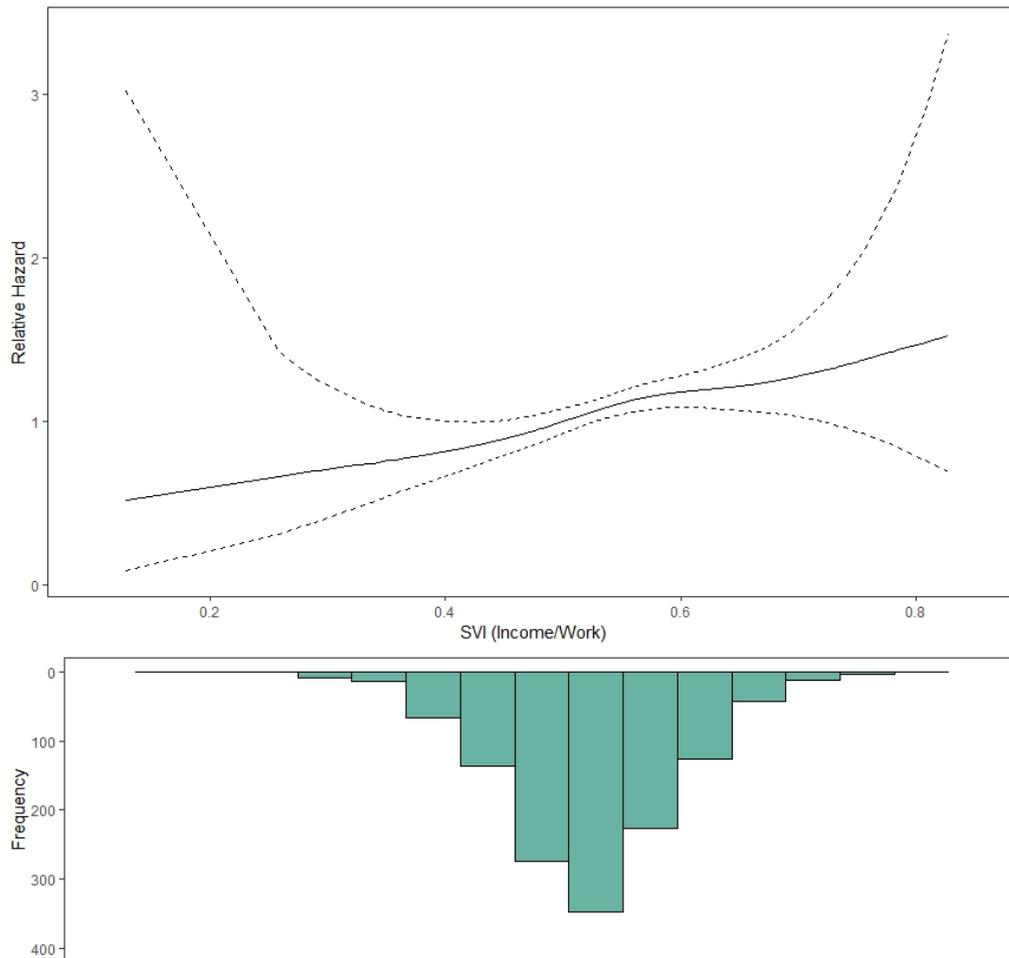


Figure 3.13. Relative hazard of endemicity increases linearly with the income/work SVI score. Relative hazard against the income/work SVI component, with an endemic criterion of 3.0 cases per 100000. Solid line is mean hazard, dotted lines are +/- 2 s.e. (top) along with a histogram of the income/work SVI values for all municipalities included in the analysis (bottom).

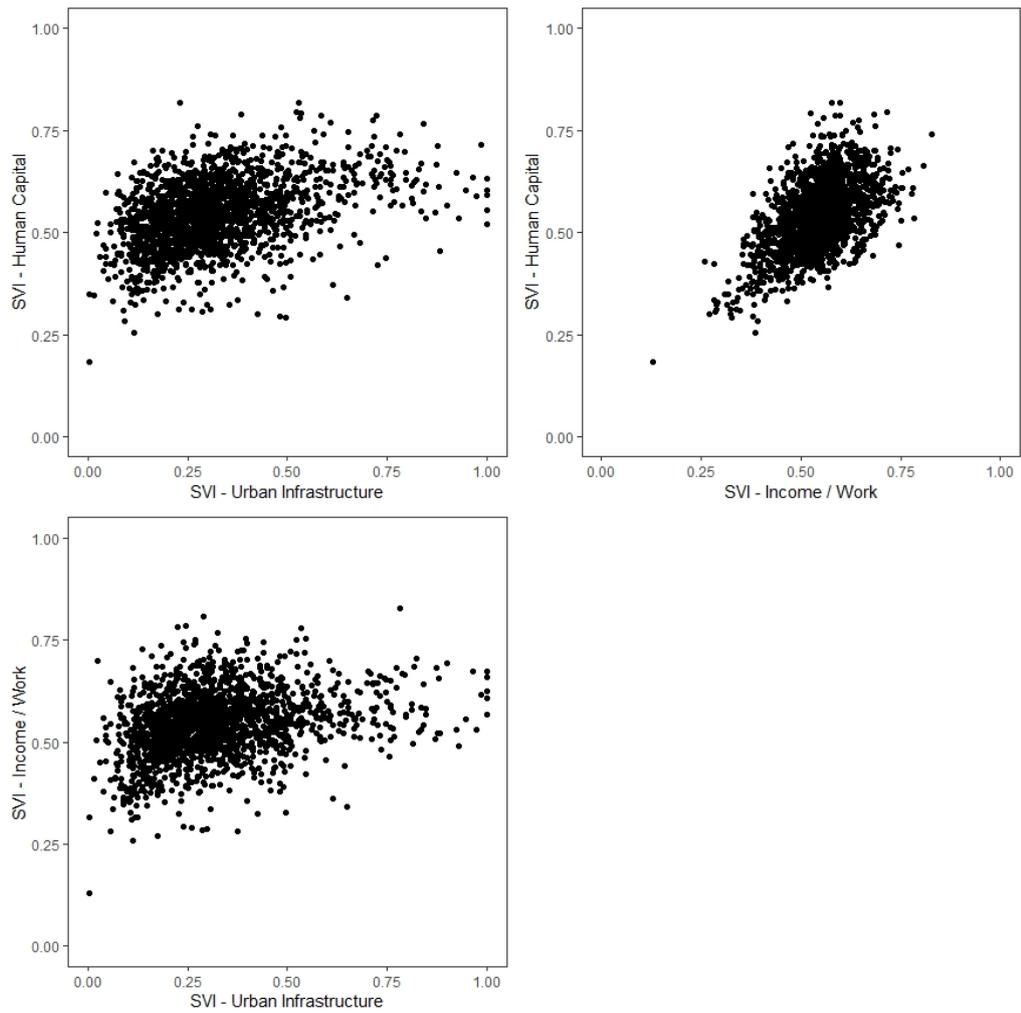


Figure 3.14. The SVI component scores for all three components are positively correlated. Scatter plot showing pairwise comparisons of the three SVI components for all municipalities.

that proximity to prior endemic municipalities has on the hazard of endemcity. We will use the latitude and longitude coordinates provided with the Ribeiro *et al.* dataset [149] to calculate the distance between the municipalities. As before, we will consider specific discrete bands (rather than a single ‘proximity’ covariate that is continuous). These bands must be sufficiently narrow that we avoid a scenario where all municipalities begin with an endemic municipality in both bands; we choose one band to be 0 - 1 unit of distance, and the other to be 1 - 1.5 units for this particular analysis.

In order to allow for an interaction between the two covariates, we will begin with a bivariate analysis (Table 3.9). We observe no significant relationship between either covariate and the hazard of endemcity. For completeness, we will also consider the covariates individually. First, the presence of an endemic municipality with 1 unit (Table 3.10), where as anticipated we observe that having an endemic municipality within 1 unit does not significantly affect the hazard experienced. Next, the presence of an endemic municipality between 1 and 1.5 units away (Table 3.11). This, too, remains non-significant.

Factor	Hazard Ratio	95% CI	P-value
Endemic municipality within 1 unit	2.5063	(0.8498,7.392)	0.396
Endemic municipality between 1 and 1.5 units	5.031	(1.7404,14.5429)	0.128
Interaction term	0.3008	(0.0964,0.939)	0.291
Likelihood Ratio Score:	4.66	P-Value:	0.1987

Table 3.9. Bivariate analysis suggests that hazard of endemcity does not depend on proximity to an endemic municipality. Summary of hazard ratios, confidence intervals and associated p-values for both distance band covariates and their interaction term.

Factor	Hazard Ratio	95% CI	P-value
Endemic municipality within 1 unit	1.06846	(0.7627,1.4967)	0.844
Likelihood Ratio Score:	0.04	P-Value:	0.8426

Table 3.10. Endemic municipalities within 1 unit do not increase the hazard of endemcity. Summary of hazard ratios, confidence intervals and p-values for the nearest distance band.

Factor	Hazard Ratio	95% CI	P-value
Endemic municipality between 1 and 1.5 units	1.8226	(1.2462,2.6658)	0.114
Likelihood Ratio Score:	3.06	P-Value:	0.08038

Table 3.11. Endemic municipalities between 1 and 1.5 units away do not increase the hazard of endemcity. Summary of hazard ratios, confidence intervals and p-values for the further distance band.

From these analyses, we may conclude that proximity to a municipality which we already consider endemic does not increase the hazard of a municipality becoming endemic. Though this may seem counter-intuitive, it may be due to the fact that many ‘susceptible’ municipalities already experience low levels of transmission at the start of the study period (insufficient cases to be considered endemic, but cases nonetheless). The proximity to an endemic municipality may be more important for establishing leishmaniasis cases in a naive municipality, but once a municipality is infected with leishmaniasis, the case numbers increase with minimal dependence on endemic neighbours. Alternatively, perhaps any effect of proximity to an endemic municipality is countered by human interventions taken as a response to the proximity. Future studies may be able to provide further illumination on this matter.

3.5.3 Limitations of survival analysis.

Our applications of survival analysis demonstrate the flexibility of the method. However they also highlight the limitations of the method. It is important to discuss these limitations in more detail, as an understanding of them can be crucial to properly applying the method. One limitation has been particularly evident in our examples thus far: the poor handling of sparse parameter ranges.

Consider once more the spline of relative hazard against latitude generated during our analysis of the Pigott *et al.* dataset (Figure 3.8). The spline had a notable region with particularly large standard error: latitude < -25 . If we examine the distribution of latitudes for recorded cases (Figure 3.15) we observe that the region with especially large standard error is also the region of the distribution containing the least data: few cases occurred with latitude < -25 . Interestingly, another less populated region of the data is located around latitude 20, for which we also observe increased standard error. This means we have a very small sample of survival times (especially since we do not track any individuals that ‘survived’ in this dataset) which causes problems for the fitting of the spline. For regions with many survival times, such as $-70 < x < -40$, we find that the standard error is very small. A similar pattern is observed with the Ribeiro *et al.* dataset. For regions of the SVI parameter space represented by very few municipalities (for example, $\text{SVI} < 0.2$ for the overall SVI, see Figure 3.9 for details), we also typically observed larger standard error in the fitted spline.

An alternative to fitting penalised splines is to subdivide the parameter space into discrete categories. This approach would be reasonable for the Ribeiro *et al.* dataset,

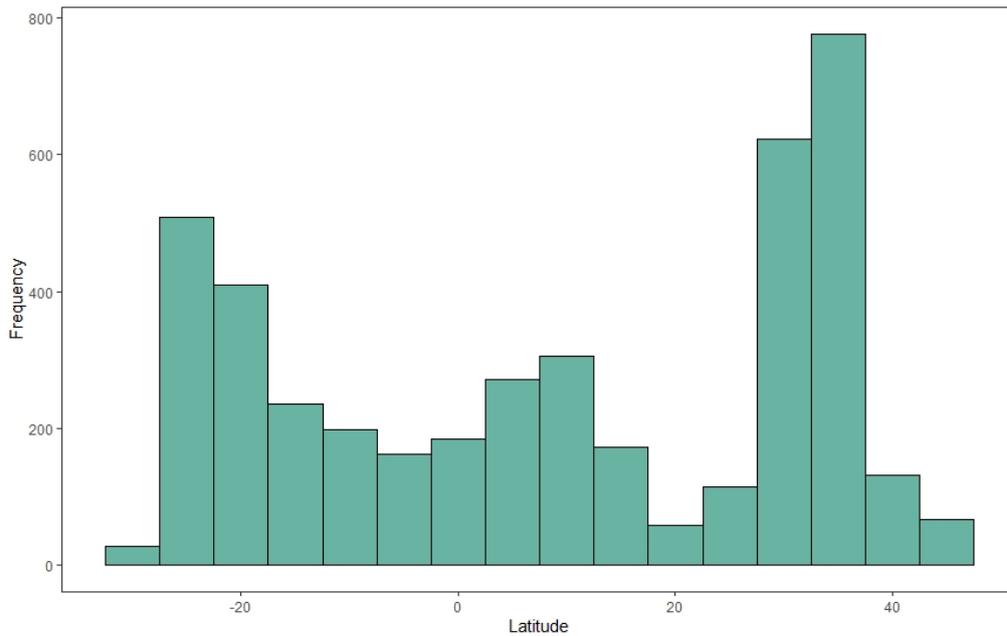


Figure 3.15. Cases in the Pigott *et al.* dataset are not distributed uniformly across the range of latitudes. Histogram of latitudes for all cutaneous leishmaniasis cases belonging to the “point” category.

for example, since the SVI of municipalities is typically categorised as very low (0.0-0.2), low (0.201-0.3), medium (0.301-0.4), high (0.401-0.5), and very high (> 0.5) [210]. Figure 3.16 shows how the distribution of municipalities would look under such a scheme. Unfortunately, we observe that the poorly represented region of the parameter space (SVI < 0.2) remains poorly represented even under the discrete system (as would be expected, since that region is assigned its own category).

To further illustrate this point, we repeat the survival analysis from Figure 3.10 using overall SVI as the only covariate, now discretised, and taking our failure criterion to be 3.0 cases per 100000 people as before. For the purposes of highlighting the model limits, we also take the “very low” SVI category as our reference category, resulting in Table 3.12. It is clear that we were unable to successfully fit the Cox PH model to this data. This is because our choice of reference category is the least represented category. We should instead use a more populated category, for example the “very high” category, as in Table 3.13. This greatly improves our fitted hazard ratios. However, we observe that the “very low” category still causes the same error as before, and thus we cannot obtain the proper hazard ratio.

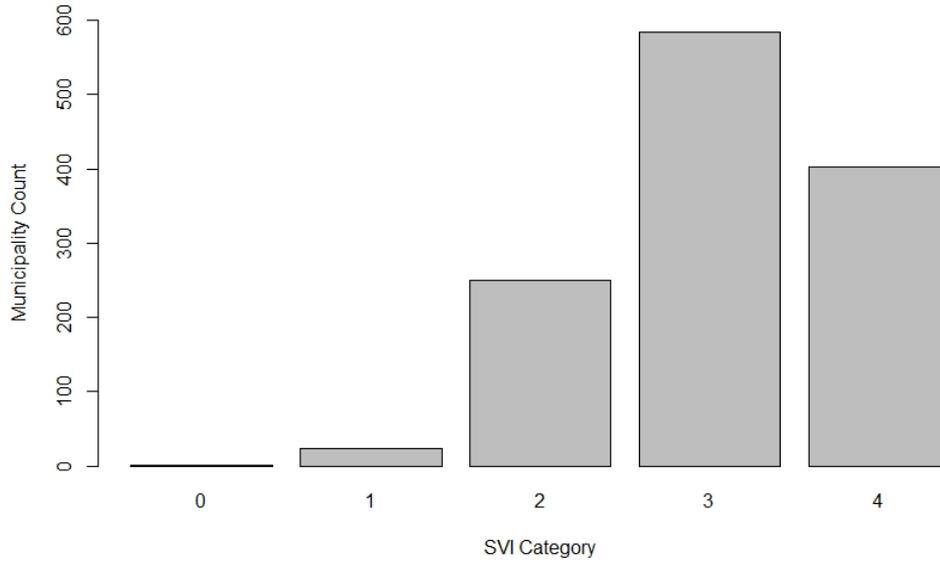


Figure 3.16. Not all official categories of SVI are equally well represented. Histogram of municipality SVI values, under the categorisations scheme recommended by (Brasil2019).

SVI Category	Hazard Ratio	Coefficient	Standard Error of Coefficient	P-value
Low SVI	227938.55	12.34	781.0	0.987
Medium SVI	394539.07	12.89	781.0	0.987
High SVI	461920.81	13.04	781.0	0.987
Very High SVI	470732.93	13.06	781.0	0.987

Table 3.12. A poorly represented reference category prevents calculation of the Wald statistic. Summary of hazard ratios, coefficients and their standard errors, with the associated p-values for all SVI categories. The reference category is the “very low” SVI.

Let us make an additional modification, and define different categories: for example, let us use low $([0.0, 0.3])$, medium $((0.3, 0.5])$, high $((0.5, 0.7])$ and very high $((0.7, 1.0])$. The distribution of municipalities is now shown in Figure 3.17. We find that although our low and very high categories are both less populated than the other two categories, they are much better than the original “very low” category. Repeating the survival analysis results in Table 3.14. We find that we are now free of the fitting error. Additionally, we now observe a significant effect: we observe that areas with high SVI values have a significantly higher hazard than those with low SVI, although this does not remain true for the municipalities with very high SVI values.

These issues with data quantity or handling are relevant beyond leishmaniasis. A recent study linking risk factors such as age to health outcomes for anal squamous cell

SVI Category	Hazard Ratio	Coefficient	Standard Error of Coefficient	P-value
High SVI	0.9799	-0.02026	0.08869	0.819
Medium SVI	0.8381	-0.1766	0.1136	0.12
Low SVI	0.4842	-0.7252	0.3841	0.059
Very Low SVI	2.124×10^{-6}	-13.06	781.0	0.987

Table 3.13. Choosing an alternative reference category mitigates the issues calculates the Wald statistic. Summary of hazard ratios, coefficients and their standard errors, with the associated p-values for all SVI categories. The reference category is the “very high” SVI.

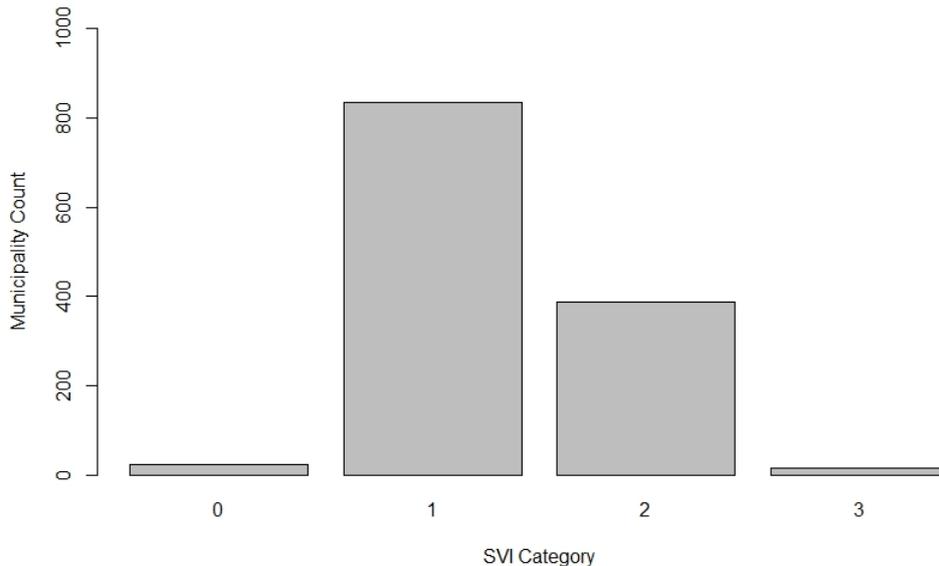


Figure 3.17. Histogram of municipality SVI values, under our alternative categorisation scheme.

carcinoma [220] drew criticism for having an insufficient number of events for the analysis performed [221]. Previous studies have highlighted the increasing risk of bias as the number of events per variable tested decreases [222] and have proposed a few target thresholds to mitigate this [223, 224], though employing them mandates estimating how many events one expects to observe prior to the study [225] and may be difficult for studies with smaller cohorts or infrequent events [220] for which it would be beneficial to consider alternatives to the Cox proportional hazards model [225]. The categorisation of continuous variables has similarly seen both applications and problems elsewhere [226], with poorly chosen categories shown to reduce the power of the test [227] especially if the choices were motivated by a desire to minimise the p-values produced [228], underscoring the importance of caution when choosing categories.

These examples demonstrate that although including under-represented regions of parameter space in a survival analysis can have detrimental effects of the fitting of models such as the Cox PH model (with both discrete and continuous covariates), it is possible to overcome some of the issues through careful handling of the parameters. Survival anal-

SVI Category	Hazard Ratio	Coefficient	Standard Error of Coefficient	P-value
Medium SVI	2.0349	0.7105	0.3811	0.0623
High SVI	2.2209	0.7979	0.3842	0.0378
Very High SVI	1.1673	0.1547	0.5855	0.7916

Table 3.14. Using alternative SVI categories allows for the calculation of the coefficients and associated Wald statistics. Summary of hazard ratios, coefficients and their standard errors, with the associated p-values for all SVI categories. The reference category is the “low” SVI.

ysis also has the ability to combine time-dependent covariates and the use of penalised splines such that we can assess continuous time-dependent covariates without needing the discrete bands we have used thus far. However, this approach is very sensitive to a lack of data: if any region of the parameter space is represented by very little data, it typically proves impossible to fit the penalised splines. Additionally, if the covariate changes too frequently, the time period is broken up into many small windows. This also prevents fitting coefficients to the data. We attempted to apply the method to both the original Bern et al. [147] and the Ribeiro et al. [149] sets, without success in either case.

3.5.4 Considerations for a good survival analysis dataset.

Using our examples, we have demonstrated the flexibility of survival analysis. However, we have also shown that it is a tool which must be applied with suitable consideration to avoid drawing misleading conclusions, and that has limitations that must be appropriately handled. In order to facilitate the application of survival analysis elsewhere, we conclude with a discussion of some of the factors that should be consider when choosing or collecting data for use with survival analysis.

The core feature of any dataset intended for use with survival analysis is the presence of survival times for all individuals. Although every individual must have a survival time, survival analysis is flexible in terms of how the survival time is formatted. Individuals are often right-censored (i.e. the study ended before an event occurred), but this is not the only acceptable format. As seen in the Bern *et al.* dataset [147], some individuals may join part way through the study (in this case, are born after the study began) and this can be handled by assigning them a start time. The main requirement is that the survival time accurately reflects the period in which we would have detected an event if it had occurred. Specifically, there are two main considerations. Firstly, the start time must represent the earliest time point that we could have detected an event (either the start of the study or the time the individual joined, whichever is most appropriate). Secondly, censoring should be non-informative (i.e. the cause of the censoring must not be related the event of interest). Informative censoring typically renders most standard survival analysis methods

invalid, and should be avoided if possible.

The consequences of using inappropriate survival times is best demonstrated by our analysis of the Pigott *et al.* dataset [204]: in the absence of start times, we naively assume all start times to be 1960. This means that infections occurring in 2000 have survival times of 40 years, when in fact some of the affected individuals may not have been alive in 1960, let alone monitored for infection. Thus, cases that occurred more recently have artificially increased survival times, which leads to misleading results.

Although survival analysis allows for considerable flexibility in choosing covariates of interest, there are a few important considerations. Deciding whether to use a discrete or continuous covariate is usually intuitive, though if the parameter space includes regions poorly represented in the data it may be preferable to use a discrete covariate even if one may initially wish to use a continuous one. Almost any factor can be chosen as a covariate, but there are pitfalls that one must be wary of. Firstly, the value of a covariate at a given time T cannot depend on information from after that time. It would be inappropriate to, for example, measure a time-dependent covariate at the end of the study and then use this value as a time-independent covariate, because this implies that the hazard experienced at the beginning of the study depends on something from the end of the study, which is not possible. Secondly, a covariate must not be a consequence of the event of interest happening. For example, imagine we have a treatment applied only to patients who are near-death. If we took death as the event of interest, and then considered whether this treatment had been applied as a covariate, we would find a significant relationship. We might then imply that administering the treatment made patients more likely to die, when in fact we are only administering the treatment because they are going to die: the event has already begun to happen.

3.6 Discussion

Using the Cox proportional hazards model, extended for use with time-dependent covariates, we have confirmed that individuals sharing a household with an existing leishmaniasis case experience increased hazard of infection. We also find that cases close to, but outside of, the household of an individual also increase the hazard of infection to a lesser extent. Additionally, our multivariate model suggests that once an infection is present within 25m of an individual, further infections beyond 25m away contribute minimally to their hazard of infection. Finally, we note that cases more than 50m away from an individual provide no significant contribution to their hazard of infection. We also demonstrate the extending the model fitting to incorporate a delay of 4 months does not result in meaningful changes.

Our survival analysis demonstrates that, whether we incorporate the incubation period of leishmaniasis or not, sharing a household with an infected individual increases the hazard of infection. This confirms the findings of previous studies. Bern *et al.* [147] used logistic regression to show individuals in the same house as an infected individual were much more likely to become infected than those whose nearest infected neighbour was more than 50m away. Similarly, Chapman *et al.* [146] reported that infected individuals exert greater infectious pressure on susceptible individuals in the same household, though were uncertain as to the true magnitude of the effect. Additionally, we find that infections which are close to an individual but not in the same household as them also increase the hazard of infection, mirroring the findings of Bern *et al.* and Chapman *et al.* [146,147].

These results are not unexpected, given the spatial clustering of cases typically observed in leishmaniasis outbreaks. Although we often consider clustering the form of large-scale hotspots such as those reported in China [91], Iran [89] and Sri Lanka [88], smaller-scale clustering is also observed. Clustering within districts in Sri Lanka has been documented [229] and modelled [93] highlighting that although some subdistricts are worse affected than others, cases are rarely evenly distributed. It has also been reported that clustering occurs in specific hamlets [12] at scales of a few hundred meters, showing a similar pattern to that seen in Fulbaria thana [146,147]. The distance-dependent hazard of infection observed both here and by Chapman *et al.* provides some clarity as to how such clustering emerges, at least at smaller scales.

The observed dependence of infection hazard on proximity to existing cases may arise from the behaviour of the sand flies. Sand flies have a limited flight range [69], and although they are capable of travelling long distance during dispersal [175] many travel much shorter distances [68], especially after feeding [175] or when suitable sources of blood are abundant [174]. Additionally, different species of sand fly have different habitat and resting spot preferences [77,230,231] leading to further small-scale heterogeneity in sand fly densities: for example, some sand flies prefer specific types of vegetation such as banana trees [232]. Some studies have reported clustering of cases due to proximity to sand fly habitat [233] and this may be further augmented by the potential for site and/or host loyalty in sand flies [234], which has long been known to contribute to maintaining other parasitic diseases such as malaria [235].

The proposed relationship between hazard of infection and distance appears to render it difficult to establish more distant infections (and thus new clusters). One possible compatible mechanism capable of generating new clusters is human migration. Some regions which have recently experienced new leishmaniasis epidemics contain a large number of migratory workers [12], which travel between their home village and potentially regions

where leishmaniasis is endemic. In Brazil, the spread of leishmaniasis has followed the movement of migratory construction workers [236]. Migration of infected people into regions such as Nepal [141], Turkey [237] and Ethiopia [238, 239] has also been proposed as a key aspect in the emergence of new epidemics, not to mention the role played by the movement of infected dogs, the presence of which is known to be a risk factor [240]. A related possibility is the migration of individuals with post kala-azar dermal leishmaniasis (PKDL). PKDL typically, but not always, emerges after recovery from visceral leishmaniasis [31], manifesting as a rash in otherwise healthy individuals. It is thought to be infectious towards sand flies [241] and an important leishmaniasis reservoir [242] (but see also [32]). Potentially, an otherwise healthy individual could move to a region with minimal leishmaniasis presence, develop PKDL, and become a reservoir capable of seeding a new cluster of leishmaniasis cases.

At larger scales we can draw useful conclusions about the likelihood that leishmaniasis becomes endemic in a given administrative region. We observe that, in the case of the northeast Brazil region, municipalities with higher social vulnerability are more vulnerable to the spread of leishmaniasis and more likely to experience endemic leishmaniasis. This conclusion is reflected across a variety of different measures of social vulnerability, from urban infrastructure to work/income stability and holds for a range of different endemicity thresholds informed by the governing policy of Brazil.

It is tempting to think it intuitive that increasingly developed urban infrastructure would reduce the likelihood of leishmaniasis becoming endemic in a municipality. Other vector-borne parasitic diseases such as malaria [243, 244] have shown greater reductions in urbanised regions than in their rural counterparts and it is possible that leishmaniasis experiences the same threat. However, the link between urbanisation and vector-borne disease transmission is far from straightforward. For example, urbanisation can facilitate the spread of dengue [245] by creating suitable habitats for species such as *Aedes aegypti* [246, 247]. Typically, where urbanisation is to the benefit of vector-borne diseases it is due to the associated high density of humans [248] which sustain higher vector densities (and the associated diseases).

In the case of leishmaniasis specifically, it is likely that the impact of urbanisation is due to the loss of suitable habitats for the sand fly vector. Human exploitation of naive wooded areas has been shown to increase the prevalence of cutaneous leishmaniasis [249] due to exposure to reservoirs and sand fly habitat, and clustering of cases around habitat fragments in urban areas has been reported [233], especially for those with preferred vegetation [232]. Livestock have also been associated with vector abundance and occurrence [250]. These factors are likely diminished by intensifying urban infrastructure. Additionally, the ur-

ban infrastructure component of the SVI score incorporates access to sanitation (and thus healthcare). Lack of access to sufficient sanitation and healthcare [251] and damaged or underdeveloped housing [130] have been demonstrated to be key risk factors in the spread of visceral leishmaniasis, and are greatly improved at the lower end of the SVI-UI scale.

The role of unstable employment, in the form of a dependence on temporary and potentially migratory labour (i.e. short-term farm work and other seasonal work), may also be important to maintaining leishmaniasis in a municipality. Such individuals have been found to be at greater risk of contracting leishmaniasis [131, 252] due to misconceptions about the disease [253] and living conditions which leave them exposed to sand flies [11] (but see also [254]). The reliance on transient jobs potentially forces individuals to work in multiple places over the course of a year, presenting more opportunities for them to be exposed to infected sand flies (particularly if their job places them in proximity to wild reservoir species [249]). Additionally, high SVI-I/W values can indicate greater dependence on child/elderly labour, thus exposing a greater proportion of the population to leishmaniasis. These age groups are known to be more likely to die from visceral leishmaniasis [255–257] and it is believed that younger individuals develop higher parasite burdens and more severe infections [258] than their older peers. Thus, having children exposed to leishmaniasis may have significant ramifications for transmission.

The enhancement of leishmaniasis transmission in regions of unstable employment mimics similar observations for diseases such as malaria. Employment-related human migration at various scales is known to be important [259] with seasonal work in areas of high vector density [260] and smaller-scale movement between outdoor areas [261] often highlighted as a gap in the targeting of malaria prevention strategies. Similarly, Chagas disease is known to have a higher burden in rural regions [262] particularly due to seasonal movement of workers [263] or activities such as hunting [264]. Although these diseases also maintain transmission in urban environments with more stable employment [265], the increased burden in rural communities highlights the impact of unstable employment.

Previous studies have highlighted the role that perceptions and knowledge of leishmaniasis may influence risk of transmission. In Ethiopia, migrant workers are aware of leishmaniasis from prior experience but have reservations about some preventative measures [253]. Elsewhere, lack of awareness and misconceptions about treatment efficacy contribute to poor uptake and adherence to treatments [266] and reliance upon potentially harmful alternative treatments [267]. The mental impacts of leishmaniasis [268] are known to be worsened in contexts with poorer access to education [269] due to a more intense social stigma associated with the disease. Similar issues with perceptions of risk and disease behaviour are documented globally [270]. Although in this case, we could not

confirm a definite relationship between hazard of endemicity and education access, there is precedent for it in the context of leishmaniasis.

Such a relationship is present for other vector-borne parasitic diseases too. Better access to education has enabled greater awareness of malaria even in areas in which it is in decline [271] leading to enhanced reporting of the disease [272] and better implementation of preventative measures [273, 274] and use of treatments [275]. In the case of Chagas disease, access to education have improved recognition of the link between the disease and the triatomine bug vector [276] but misconceptions regarding the treatment of the disease have negative impacts on the health outcomes for infected individuals [277].

Knowledge of the scales at which transmission operates will be crucial to our ongoing efforts to eliminate leishmaniasis. Elimination of leishmaniasis has long depended strongly on indoor residual spraying (IRS) to kill sand flies [278]. Recent studies have questioned the efficacy of IRS [279, 280] and a leading alternative, long-lasting insecticide-treated bednets (LLINs) [281, 282] (but see also [283]), potentially in part due to rising insecticide resistance in some sand fly species [284, 285]. While improvements [286, 287] are being developed it is possible to use our knowledge of transmission to enhance how we apply our existing methods by focusing their application appropriately.

Although it is tempting (especially in areas with limited access to preventative methods) to restrict application of IRS and LLINs to infected houses only, our findings suggest it is pertinent to expand the coverage to all houses within 50m of the infected household. In areas with vegetation known to support a high sand fly density, such as banana plants [232] or acacia trees [11, 143], there may be benefits from also applying insecticide carefully to the vegetation to eliminate, or reduce, it as a viable habitat fragment. This bears semblance to the hotspot-targeting approach used against malaria [289], which has proven successful in reducing transmission. Another similar approach would be to ensure that any accommodation within 50m of known sand fly habitat and/or reservoir species is treated with IRS, supplied with LLINs, and where possible) built to be appropriately resistant to sand fly colonisation [130, 290].

Additionally, interventions could be targeted at the scale of municipalities. Municipalities with high SVI scores are more susceptible to the spread of leishmaniasis, though the specific interventions must be tailored to local conditions. Regions with a high dependence on seasonal, especially migratory, workers would benefit from an increased focus on education campaigns (which have seen prior success at improving awareness of leishmaniasis [291]) and from making testing and treatment more readily available. The combination of improved health messages [292] and access to screening [293] have been highlighted as important to counter Chagas disease, and gaps in screening of canines for visceral leish-

maniasis is thought to have undermined control of the disease in Brazil [294]. By ensuring that seasonal workers have good access to testing and knowledge of effective treatment, we would be able to reduce the transmission of leishmaniasis with the municipality and reduce the risk of exporting cases to other municipalities. This may be especially important for situations where migration happens across a variety of distances [12, 88] and cases could be exported over greater distances.

There are a few useful avenues that future work in this area may choose to explore. One recurring factor we have encountered in this work is the potential role of employment type (and stability) in leishmaniasis transmission. A longer-term monitoring study of individuals in different types of employment could allow future analysis to reveal higher-risk jobs or industries, as well as further explore the relationship between job stability and leishmaniasis transmission. It would also be beneficial to be able to repeat the analysis we performed for Northeastern Brazil in other regions where leishmaniasis is endemic, such as the Indian Subcontinent or Ethiopia. This would allow us to test whether our conclusions remain applicable in other geographical contexts.

3.7 Conclusion

Through the application of survival analysis to publicly available datasets, we have demonstrated and discussed some of the strengths and weaknesses of the method in the context of understanding leishmaniasis transmission. We also identify some of the factors that underpin leishmaniasis transmission at different scales. We highlight that leishmaniasis spreads more readily to individuals within 50m of a prior case, especially if they share a household. We also highlight that larger-scale administrative regions can have varying levels of vulnerability to leishmaniasis depending on their urban infrastructure quality, access to education, and dependence on seasonal work. We use these identified behaviours to suggest possible methods of targeting our available intervention strategies to maximise their effectiveness in situations where they may not be universally applied due to factors such as cost.

3.8 Materials and Methods

All survival analysis was performed in RStudio v2022.07.1 Build 554 (R version 4.1.0) using the survival package [188] v3.4-0. All further analysis was also performed using RStudio v2022.07.1, Build 554 (R version 4.1.0).

Chapter 4

Applications and limitations of an analytic R_0 estimate for heterogeneous network models.

4.1 Abstract

Despite ongoing elimination efforts, leishmaniasis persists across the globe, both in endemic communities and by establishing new clusters of infections. Studies portray endemic communities as heterogeneous environments of mystery reservoirs, short-range sand fly vectors and varied habitat fragments; they are complex and only partially understood. The heterogeneous transmission landscape makes leishmaniasis an appealing target for the much-explored network models, often employed in contexts with structured communities such as those endemic for leishmaniasis. We replicate the derivation of a general analytic estimate for the basic reproduction number R_0 for heterogeneous networks. We produce context-specific estimates to assess its suitability for predicting the spread of leishmaniasis. We find that the estimate consistently overestimates the value of R_0 even on simpler heterogeneous networks. However, we find that the estimate can be repurposed to predict the epidemic threshold transmission rate with reasonable accuracy. We propose scenarios where this estimate could be beneficial for understanding leishmaniasis transmission as well as potential methods that could improve upon or be used in conjunction with the estimate in future.

4.2 Introduction

Leishmaniasis is a disease of surprisingly many forms [296–299] caused by parasites of the *Leishmania* genus. Despite its global burden [300, 301] it remains a neglected tropical disease, less studied than more prominent vector-borne diseases such as malaria. It is targeted for eradication [3], especially on the Indian Subcontinent, but despite concerted efforts to achieve this [302–304] it maintains a substantial presence in many countries across the world [9, 305–307], often in more vulnerable groups such as refugees [16, 308] and rural communities [309, 310]. Such groups often have restricted access to medical care [311, 312] resulting in worse health outcomes as a result of infection with leishmaniasis [313], making prediction and prevention an imperative.

Our knowledge of leishmaniasis transmission has advanced greatly in recent years. Studies have revealed that *Leishmania* parasites are distributed heterogeneously in the skin of mammalian hosts [314] and that skin parasites may be more important for outward transmission than those in the blood [45]. The life cycle of the parasite is influenced strongly by the biting behaviour of their sand fly vectors [54], but also influences the sand flies in return by encouraging more persistent biting [84]. These factors appear to interact to further enable transmission even in seemingly unlikely scenarios [172]. Larger scale transmission on the other hand is a diorama of endemic communities [93, 147] and newly established clusters [315], its range expanding with climate change [17, 106], infrastructure expansion [121] and intensifying land use [316] and sustained by habitat fragments [233] and partially-understood reservoirs [18, 242, 317]. Combined with factors such as housing material [130, 290] and working patterns [76], a complex, multifaceted portrait of leishmaniasis transmission emerges. Even if we overcome the remaining mysteries [53], capturing this in full would likely produce an entirely intractable system. Advancements are more easily made using abstracted systems and specific tools to glean particular insights about transmission. Tools like survival analysis [111] highlight the risk factors underpinning the large-scale spread of leishmaniasis [121], much as ecological niche modelling [96, 318] has charted the potential spread of its vectors [102, 319], and at smaller scales logistic regression [147, 201] has identified similar risk factors. Thus, we can avoid much of the complexity of the full system and yet still discern useful information.

An oft-applied metric that may be of use here is the ‘ R_0 ’ value, which represents the expected number of secondary infections caused by a typical infection in a naive population [320]. It is a convenient measure of how rapidly a disease is expected to spread, with higher R_0 values associated with faster transmission. The R_0 value has been estimated for a vast array of diseases [321–323] and circumstances [324, 325], seeing applications

in predicting the minimum vaccine uptake required to prevent an outbreak [326, 327] and identifying particularly vulnerable areas [328, 329]. Often, the R_0 value is calculated using data from outbreaks [330, 331], though this can sometimes produce widely varying estimates [320]. In the context of leishmaniasis, such calculations are likely to be stymied by the abundance of asymptomatic cases [332] and the ongoing uncertainty as to their role in transmission [146, 170, 317, 333]. This is further complicated by the restrictive conditions for transmission [146]. Any R_0 estimate must be properly tailored to this situation.

Given the structured nature of the communities that leishmaniasis spreads through, network models present an opportunity. Network models possess a variety of useful traits [334] and have seen applications in a wide array of epidemiological contexts. Modular networks have been used in the context of both human [335] and animal [336] social groups, bipartite networks are widely applied to host-vector dynamics [337, 338], and in some contexts even quite simple models can capture disease transmission through contact networks [339]. Important aspects of transmission have been explored using networks, ranging from the role of bridge nodes in inter-community transmission [340] and the impacts of modularity on the transmission of co-infections [341] to the variable effects of vaccination strategies [342] and the importance of social behaviours [343]. Despite this versatility, network models have seen minimal applications to leishmaniasis transmission thus far.

We derive a number of analytic estimates for the R_0 value of an outbreak on a heterogeneous network with a range of different transmission rate regimes and evaluate the accuracy and limitations of these estimates using numerical simulations. We also identify a potential application for the estimate in the context of predicting future leishmaniasis transmission.

4.3 Derivation of the R_0 Estimate

We will focus initially on following the steps of previous work by Diekmann and Heesterbeek [344] and Newman [345] to derive a general analytic form of our R_0 estimate which we will then define more explicitly for a specific set of scenarios. Specifically, we aim to derive:

$$R_0 = T \times \left(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}} \right) \quad (4.1)$$

as presented by Lloyd and Valeika [150], where:

$$T \in [0, 1]$$

$$\hat{k}, \hat{\hat{k}} \geq 0.$$

Here, T is the ‘transmissibility’, the average probability of transmission occurring between an infected individual and an uninfected individual [345], \hat{k} is the mean degree of individuals in the network, and $\hat{\hat{k}}$ is the variance of the degree distribution. This formula is best interpreted in two parts:

$$R_0 = T \times (\hat{k} - 1) + T \times \left(\frac{\hat{\hat{k}}}{\hat{k}} \right). \quad (4.2)$$

The first term is fairly straightforward: the transmissibility T multiplied by the average number of individuals that could be infected (one less than the mean degree \hat{k} , as for a fixed network all infected individuals have a single infected neighbour whom they were infected by). This term is reminiscent of older R_0 estimates for well-mixed populations [150]. The second term is less intuitive: a correction term that incorporates network heterogeneity [344] in the form of the variance of the degree distribution ($\hat{\hat{k}}$) allowing the estimate to be applied beyond simple homogeneous networks. Without this correction term, it is anticipated that the estimate would be an underestimate of the R_0 value in heterogeneous cases.

4.3.1 Transmissibility

The transmissibility of a disease, T , can be considered to be the average probability of transmission occurring between an infected individual and an uninfected individual [345]. Consider two individuals: an infected individual i and a susceptible individual j . The individuals have disease-causing contacts at rate r_{ij} per unit time and an individual remains infectious for a some time period τ_i .

For some small period of time δt , the probability of individual j not being infected is:

$$P = 1 - r_{ij}\delta t.$$

If we consider the infectious duration τ_i to be divided into many such ‘chunks’ of length δt , then the probability of j remaining uninfected is the probability that no infection occurs in *any* of the chunks:

$$P = (1 - r_{ij}\delta t)^{\frac{\tau_i}{\delta t}}.$$

Finally, if we take these chunks to be infinitesimally small, we find that:

$$1 - T_{ij} = \lim_{\delta t \rightarrow 0} (1 - r_{ij} \delta t)^{\frac{\tau_i}{\delta t}} = e^{-r_{ij} \tau_i} \quad (4.3)$$

where $1 - T_{ij}$ is the probability that individual j is *not* infected at any point during the infectious duration of i and where we assume continuous time. Thus, the probability of infection is:

$$T_{ij} = 1 - e^{-r_{ij} \tau_i} \quad (4.4)$$

We can derive a similar equation for discrete time. For this, we now define ρ_{ij} to be the probability that an infectious contact occurs between individual i and individual j in a single timestep and define λ_i to be the number of timesteps that individual i is infectious for (i.e. the infectious duration). Thus, T_{ij} now takes the form:

$$T_{ij} = 1 - (1 - \rho_{ij})^{\lambda_i} \quad (4.5)$$

If we once again consider ρ_{ij} and λ_i as i.i.d. according to some distributions $P(r)$ and $P(\lambda)$ respectively, we obtain:

$$T = \langle T_{ij} \rangle = 1 - \int_0^{\infty} d\rho \sum_{\lambda=0}^{\infty} P(\rho) P(\lambda) (1 - \rho)^\lambda \quad (4.6)$$

We will use this expression for transmissibility in the construction of a number of estimates for R_0 . For a summary of our notation, see Supplementary Table 1.

4.3.2 Expected Number of Contacts

In order to derive the rest of Equation (4.1) we follow the work of Diekmann and Heesterbeek [344]. They begin by introducing the concept of the “h-state”: information about an individual that is relevant to transmission but not directly related to the disease. The h-state can be discrete (for example, encoding details such as sex or species) or continuous (for traits such as age) as the situation demands. For clarity, we first consider a scenario with finitely many discrete h-states. In this case, Diekmann and Heesterbeek define k_{ij} to be the number of individuals with h-state i infected by a single individual with h-state j over their entire infectious duration. They also define the “next-generation matrix” to be:

$$K = (k_{ij}).$$

These are best illustrated with a simple example: consider a sexually transmitted infection in an entirely heterosexual population. Individuals have a h-state of one of two types, male

or female, and individuals of one type can only infect individuals of the other type. For simplicity, assume that each infected individual of either h-state can be expected to infect 3 people over their infectious duration. Then, the next generation matrix takes the form:

$$K = \begin{bmatrix} k_{mm} & k_{mf} \\ k_{fm} & k_{mm} \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}$$

This matrix can be used to predict how the number of infections will change from one generation to the next. Imagine a population (denoted ϕ) in which 6 males and 3 females are infected, s.t.:

$$\phi_1 = \begin{pmatrix} 6 \\ 3 \end{pmatrix}$$

which represents the initial infections. The number of infections in the next generation is then:

$$\phi_2 = K\phi_1 = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix} \begin{pmatrix} 6 \\ 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 18 \end{pmatrix}$$

where $K\phi_1$ denotes matrix multiplication.

For an arbitrary h-state space, which may have continuous or discrete elements, Diekmann and Heesterbeek [344] perform a similar derivation. Let Ω be the space containing all h-states. For some h-state η and region θ of h-state space, they define $\Lambda(\eta)(\theta)$ to be the expected number of individuals with h-state in θ caused by a single individual with h-state η over their entire duration of infection. They also define the number of cases with h-state in θ for a given generation to be some $m(\theta)$.

In this context, they define K to be the “next-generation operator” which once again describes the change in the number of infections between generations and thus:

$$(Km)(\theta) = \int_{\Omega} \Lambda(\eta)(\theta)m(d\eta).$$

In both contexts, the R_0 value is defined to be the spectral radius of K and although one can extract it directly (see, for example, [344,349] for details), Diekmann and Heesterbeek instead introduce a further assumption that simplifies the process: that the h-state of a new infection is independent of the h-state of the individual responsible for infecting them. In the contexts we have discussed, we thus have:

$$k_{ij} = a_i b_j$$

$$\Lambda(\eta)(\omega) = \alpha(\omega) b(\eta).$$

For conciseness, we focus on the more general context. Consider a generic eigenvector $v(\omega)$ upon which K operates. We thus have:

$$(Kv)(\omega) = \lambda v(\omega)$$

where λ is the corresponding eigenvalue. Therefore:

$$\begin{aligned} (Kv)(\omega) &= \int_{\Omega} \Lambda(\eta)(\omega) v(d\eta) = \lambda v(\omega) \\ &= \int_{\Omega} \alpha(\omega) b(\eta) v(d\eta) = \lambda v(\omega) \\ \alpha(\omega) \int_{\Omega} b(\eta) v(d\eta) &= \lambda v(\omega). \end{aligned}$$

This implies that $v(\omega) = \alpha(\omega)$ and thus:

$$\begin{aligned} v(\omega) \int_{\Omega} b(\eta) \alpha(d\eta) &= \lambda v(\omega) \\ \implies \lambda &= \int_{\Omega} b(\eta) \alpha(d\eta). \end{aligned}$$

Since λ is the only eigenvalue, we thus denote:

$$R_0 = \lambda = \int_{\Omega} b(\eta) \alpha(d\eta). \quad (4.7)$$

The final part of the derivation of Equation (4.1) is best demonstrated by following a final example from Diekmann and Heesterbeek [344]. Let all individuals have some contact rate ζ . We assume that any infected individual could infect any susceptible individual (rather than having distinct, incompatible subpopulations) and that ζ remains unchanged as a result of infection. Let $m(\theta)$ be the probability that, for any $\theta \subset \Omega$, an individual has $\zeta \in \theta$ as discussed above. Diekmann and Heesterbeek define:

$$\hat{\zeta} = \int_{\Omega} \zeta m(d\zeta) \quad (4.8)$$

to be the mean contact rate and:

$$\hat{\zeta} = \int_{\Omega} \zeta^2 m(d\zeta) - \hat{\zeta}^2 \quad (4.9)$$

to be the variance. Suppose we are interested in a formula for the number of new infections with h-state in θ in the next generation caused by an infected individual with h-state η ; that is to say, we wish to find an expression for $\Lambda(\eta)(\theta)$. Diekmann and Heesterbeek represent this by:

$$P(\text{transmission}) \times (\text{contact rate}) \times (\text{proportion contacts with } \zeta \in \theta) \quad (4.10)$$

They take the average probability of transmission to be the transmissibility of the disease T (see Section 4.3.1) and the average contact rate is simply $\hat{\zeta}$. The proportion of contacts with contact rate $\zeta \in \theta$ is given by:

$$\alpha(\theta) = \frac{\int_{\theta} \zeta m(d\zeta)}{\int_{\Omega} \zeta m(d\zeta)} = \frac{\int_{\theta} \zeta m(d\zeta)}{\hat{\zeta}} \quad (4.11)$$

Thus we have:

$$\Lambda(\eta)(\theta) = T \times \hat{\zeta} \times \alpha(\theta) \quad (4.12)$$

and if we denote $b(\zeta) = T\hat{\zeta}$ then:

$$\Lambda(\eta)(\theta) = b(\zeta)\alpha(\theta) \quad (4.13)$$

Recall Equation (4.7), we thus have:

$$R_0 = \int_{\Omega} b(\eta)\alpha(d\eta) \quad (4.14)$$

which is:

$$R_0 = \frac{T}{\hat{\zeta}} \int_{\Omega} \zeta^2 m(d\zeta) = T \times \left(\hat{\zeta} + \frac{\hat{\zeta}}{\hat{\zeta}} \right). \quad (4.15)$$

In the context of a network model, we can take the degree k of each individual to be their contact rate. We also note that each individual, with the exception of those introduced to a naive population, will have a single neighbour (the individual that infected them) that is not susceptible to infection because they are either still infected or have recovered and are now immune. Thus, we have:

$$R_0 = T \times \left(\hat{k} - 1 + \frac{\hat{k}}{\bar{k}} \right), \quad (4.16)$$

which is Equation (4.1) as originally desired. We now use this general form of R_0 to derive context-specific estimates for a selection of different transmission rate regimes. We will focus on estimates for the discrete time case, for continuous time counterparts see Supplementary Methods 1.

4.4 Analytic Approximations of R_0 : Discrete Time

Next, we will derive a set of R_0 estimates for a range of transmission probability regimes of interest. Specifically, we consider three scenarios: one in which all transmission probabilities are some fixed value ρ^* , one in which they are uniformly distributed between 0 and 1, and finally one in which they follow a strict bimodal distribution. Recall that our R_0 estimate is provided by Equation (4.1), with the transmissibility given by Equation (4.6). For simplicity we restrict our analysis to the case where the infectious duration is fixed, that is to say:

$$\lambda = \lambda^*$$

and thus:

$$T = 1 - \int_0^{\infty} d\rho P(\rho)(1 - \rho)^{\lambda^*}. \quad (4.17)$$

4.4.1 Fixed Transmission Probability

Consider the scenario where all connections between individuals have equal probability of transmitting the disease per timestep, such that

$$P(\rho) = \begin{cases} 1 & \rho = \rho^*, \\ 0 & \text{otherwise.} \end{cases} \quad (4.18)$$

Thus, we have:

$$T = 1 - (1 - \rho^*)^{\lambda^*}. \quad (4.19)$$

Substituting this into Equation (4.1) we have:

$$R_0 = (1 - (1 - \rho^*)^{\lambda^*}) \left(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}} \right). \quad (4.20)$$

4.4.2 Uniform Random Infection Probability

A more interesting scenario is provided by allowing the probability of infection for each connection to vary according to a uniform distribution with minimum value a and maximum value b . We continue to assume that $\lambda = \lambda^*$ as before and that ρ remains invariant in time. Thus, we have:

$$P(\rho) = \frac{1}{(b - a)} \quad (4.21)$$

and therefore:

$$\begin{aligned} T &= 1 - \int_a^b \frac{(1 - \rho)^{\lambda^*}}{(b - a)} d\rho \\ T &= 1 - \left[\frac{-(1 - \rho)^{\lambda^*+1}}{(\lambda^* + 1)(b - a)} \right]_a^b \\ \Rightarrow T &= 1 - \frac{(1 - a)^{\lambda^*+1} - (1 - b)^{\lambda^*+1}}{(\lambda^* + 1)(b - a)} \end{aligned}$$

Substituting this in Equation (4.1), as before, we have:

$$R_0 = \left(1 - \frac{(1 - a)^{\lambda^*+1} - (1 - b)^{\lambda^*+1}}{(\lambda^* + 1)(b - a)} \right) \left(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}} \right). \quad (4.22)$$

4.4.3 Bimodal Infection Probability

Consider a scenario where the probability of infection for the connections is distributed according to a strictly bimodal distribution s.t.:

$$P(\rho) = \begin{cases} p_1 & \rho = a \\ p_2 & \rho = b \\ 0 & \text{otherwise,} \end{cases}$$

where $a \ll b$ and $p_1 \gg p_2$. In this context, $P(\rho)$ is a probability mass function and we have:

$$T = 1 - \sum_{\rho \in v} P(\rho)(1 - \rho)^{\lambda^*} \quad (4.23)$$

where v is the set of possible values taken by ρ . We assume as before that $\lambda = \lambda^*$ is a fixed constant and that r is invariant in time. We now have:

$$T = 1 - p_1(1 - a)^{\lambda^*} - p_2(1 - b)^{\lambda^*} \quad (4.24)$$

and thus:

$$R_0 = (1 - p_1(1 - a)^{\lambda^*} - p_2(1 - b)^{\lambda^*}) \left(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}} \right). \quad (4.25)$$

4.5 Numerical Simulations

We next evaluate the accuracy of these analytic predictors using a set of numerical simulations for sets of different transmission rate distributions and individual degree distributions. Focus is placed on the discrete time context, but see the supplementary materials (Section S2) for the continuous time setting.

All simulations generate an independent network of 800 individuals, each with a degree that determines the number of individuals it can connect to (for details see Section 4.5.1). Every connection between two individuals is assigned a transmission rate drawn from an associated distribution (see Section 4.4). In the discrete time context, this transmission rate is the probability of transmission occurring along that edge per timestep (where one timestep is equivalent to a week of real-time). Each simulation begins with 20 initial infections distributed randomly across the population and runs for 120 timesteps, stopping prematurely if no infected individuals remain. At each timestep, all infected individuals check each connected susceptible individual to evaluate whether transmission occurs according to their edge-specific transmission rate. All infected individuals remain infected (and infectious) for a fixed duration of 40 timesteps (i.e. for our R_0 estimates we assume $\lambda^* = 40$) and become immune to reinfection upon recovering.

Throughout our analysis, we consider the R_0 to be the expected number of secondary infections caused by a single infected individual throughout their infectious duration. To aid in examining how the accuracy of our R_0 estimates changes over time, we divide infections into distinct generations. The initial infections are denoted ‘Generation 1’ and subsequent infections are always one generation higher (i.e. all individuals infected by the initial infections are generation 2 infections). For each simulation, we then calculate the mean number of secondary infections caused by an individual of each generation.

4.5.1 Network Structure and Degree Distributions

All networks generated for these simulations have a degree distribution that belongs to one of three different categories: homogeneous, geometric distribution, and community structure. The specifics are detailed below.

Homogeneous Network

The simplest network is fully homogeneous with all individuals generated with degree 5. Individuals are assigned ‘stubs’ according to this degree and these stubs are then paired up at random following a standard configuration model approach (see [334]). For an example network see Figure 4.1.

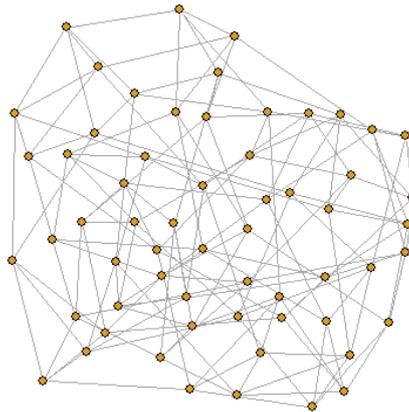


Figure 4.1. Sample network generated using the homogeneous degree distribution. Each node is an individual, edges represent viable routes of transmission.

Geometric Distribution Network

Individuals in this network type has a degree drawn from a geometric distribution with success probability p , where the degree of the individual is the number of trials recorded before the first success. Thus, the probability that an individual has degree k is:

$$Pr(X = k) = p(1 - p)^{k-1}, \quad (4.26)$$

with:

$$\text{Mean}(X) = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2}. \quad (4.27)$$

For these simulations, $p = 0.2577$. Thus, the degree distribution has mean 3.88 and variance 11.17. As in the homogeneous case, all individuals have stubs according to their degree and these stubs are paired randomly to produce the network. See Figure 4.2 for an example network.

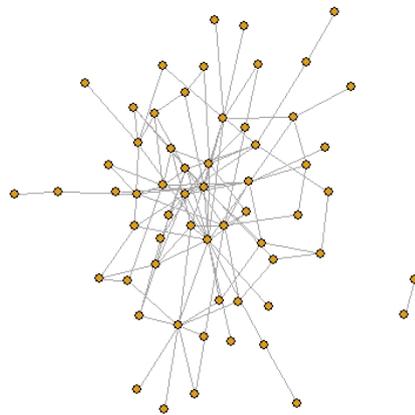


Figure 4.2. Sample network generated using the geometric degree distribution. Each node is an individual, edges represent viable routes of transmission.

Community Structure Network

The most sophisticated network used in these simulations replicates a real-world community. Individuals are each assigned to a household. Household sizes are drawn from a uniform distribution with maximum 6 and minimum 2. All individuals are connected to all other individuals in their household and two individuals in each household are chosen to connect to an adjacent household (such that each household is connected to two neighbouring households by a single connection). Since these networks possess structure beyond that provided by the degree distribution, we also generate a second version in which all the edges are removed and the resulting stubs are then randomised as in the previous networks. For sample networks see Figure 4.3.

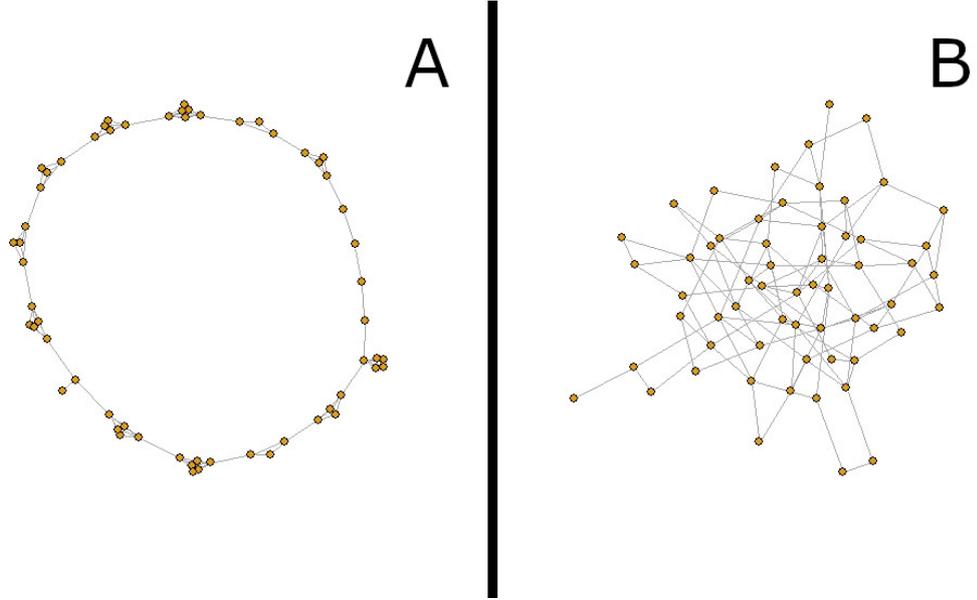


Figure 4.3. Sample networks generated for the fully structured (A) and fully random (B) versions of the community structured simulations. Each node is an individual, edges represent viable routes of transmission.

4.6 Results

It is prudent to begin by evaluating our analytic estimates for R_0 under a comparatively mundane set of conditions. We first consider the homogeneous case where all individuals possess identical degree. In its most generic form, the estimate for R_0 is reduced from:

$$R_0 = T \times \left(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}} \right)$$

to the more familiar:

$$R_0 = T \times (\hat{k} - 1). \quad (4.28)$$

In this homogeneous context, we note that the estimate for R_0 no longer depends on the variance of the degree distribution; indeed, it now depends only on the mean degree and the transmissibility T . We begin with the simplest regime of transmission rates: all edges possess an identical, fixed transmission rate. In this scenario, Equation (4.20) reduces to:

$$R_0 = (1 - (1 - \rho^*)^{\lambda^*})(\hat{k} - 1), \quad (4.29)$$

for a given transmission rate ρ^* and infectious duration λ^* . We first consider a transmission rate of 0.03 per neighbour, per timestep. If we examine the mean number of secondary infections for the first 8 generations of our simulations (Figure 4.4A) we observe that the approximation slightly underestimates for the initial infections (generation 1), but then overestimates for all subsequent generations (performing worse as the generation number

increases). We observe similar results for a reduced transmission rate of 0.015 (Figure 4.4B).

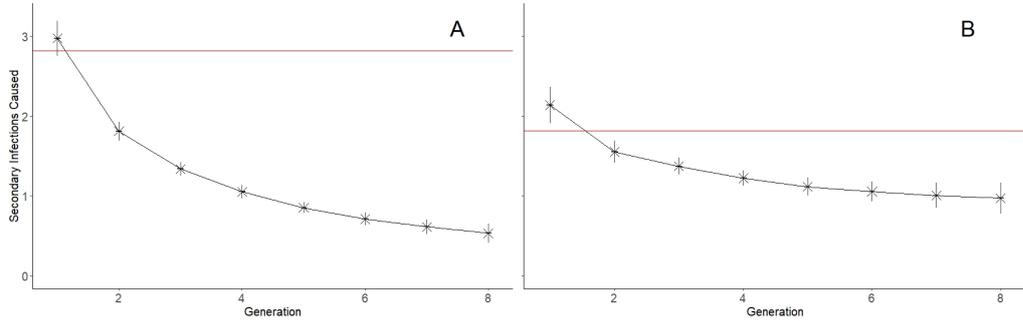


Figure 4.4. The estimate for R_0 is reasonable for the initial infections, but overestimates greatly for later generations. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.29) (red). Whiskers indicate the mean across all replications ± 1 standard deviation.

The decline in accuracy is not unexpected. This is an estimator for R_0 and thus only intended for the very beginning of an outbreak (when $t = 0$) and thus would be expected to perform poorly as the outbreak progresses. Despite this, the estimator performs reasonably well for the initial generation. The slight underestimation is also not entirely unexpected. In deriving Equation (4.28), we assume that all infections would have at least one infected neighbour (the person that infected them). This is accounted for by the use of $\hat{k} - 1$ rather than \hat{k} . However, this assumption does not hold for the initial infections.

Henceforth, focus will be placed on the higher transmission rate (see the supplementary figures for the lower-transmission counterparts). Consider now a scenario where transmission rates for each connection are drawn from a uniform random variable. We assume that the edges are not directional (i.e. transmission from individual i to individual j occurs at the same rate as from j to i). For our homogeneous network, Equation (4.22) reduces to:

$$R_0 = \left(1 - \frac{(1-a)^{\lambda^*+1} - (1-b)^{\lambda^*+1}}{(\lambda^*+1)(b-a)} \right) (\hat{k} - 1), \quad (4.30)$$

for minimum rate a and maximum rate b . To maintain a mean transmission rate of 0.03, we set $a = 0.02$ and $b = 0.04$. If we once again examine the mean number of secondary infections (Figure 4.5), we see similar results to those for the fixed transmission rate system. The estimate performs reasonably for the first generation but greatly overestimates for later generations of infections.

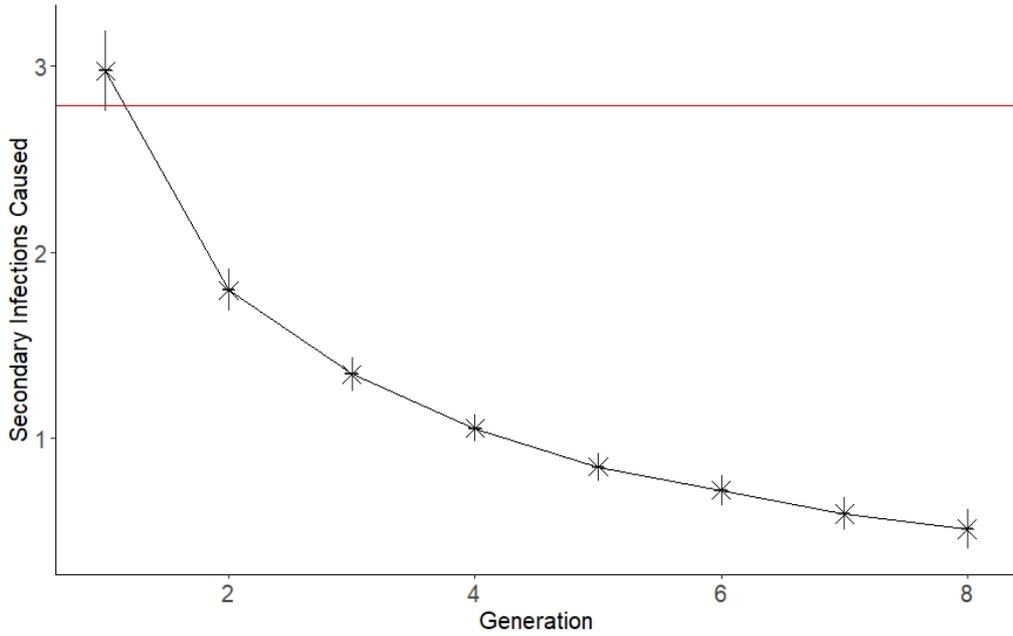


Figure 4.5. With uniformly distributed transmission rates the R_0 estimate still performs reasonably for generation 1, and overestimates thereafter. Mean number of secondary infections caused by infections of different generations (black) compared with the analytic estimate from Equation (4.30) (red), for a homogeneous degree distribution and uniformly distributed transmission rates. Whiskers indicate the mean across all replications ± 1 standard deviation.

Consider also the scenario where transmission rates are strictly bimodal; that is to say, most edges have low transmission rates but rarely an edge will have a much higher transmission rate. We once again assume transmission is symmetric for a given edge. For a homogeneous network, Equation (4.25) reduces to:

$$R_0 = (1 - p_1(1 - a)^{\lambda^*} - p_2(1 - b)^{\lambda^*})(\hat{k} - 1). \quad (4.31)$$

In order to maintain a mean transmission rate of 0.03, we choose our bimodal distribution to be:

$$P(\rho) = 0.98 \quad \rho = 0.0122 \quad (4.32)$$

$$P(\rho) = 0.02 \quad \rho = 0.9 \quad (4.33)$$

$$P(\rho) = 0 \quad \textit{otherwise.} \quad (4.34)$$

Examining the mean number of secondary cases once more (Figure 4.6) we observe a similar underestimation for the first generation but overestimation for subsequent generations as we did for the fixed and uniformly distributed transmission rate scenarios. We now examine a scenario that is more complex but more applicable to real-world scenarios: the community structure network.

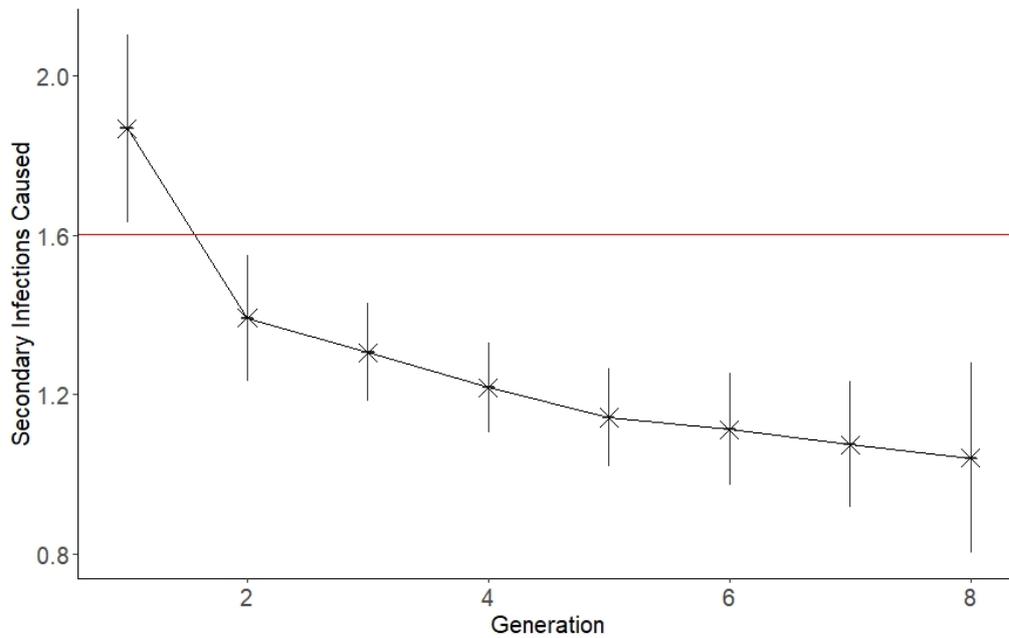


Figure 4.6. With a strictly bimodal distribution of transmission rates the R_0 estimate underestimates slightly for generation 1, and overestimates thereafter. Mean number of secondary infections caused by infections of different generations (black) compared with the analytic estimate from Equation (4.31) (red), for a homogeneous degree distribution and bimodally distributed transmission rates. Whiskers indicate the mean across all replications ± 1 standard deviation.

We begin with the fully structured simulations, as detailed in Section 4.5.1, and start with the most basic transmission regime as before: the fixed transmission rate. If we compare the mean secondary infections caused for the two different transmission rates (Figure 4.7), we observe remarkably similar results for both. This suggests that the limiting factor is the number of neighbours, rather than the transmission rate itself. More importantly, we also note that the analytic estimate (Equation (4.20)) greatly overestimates in both cases for all generations of infections.

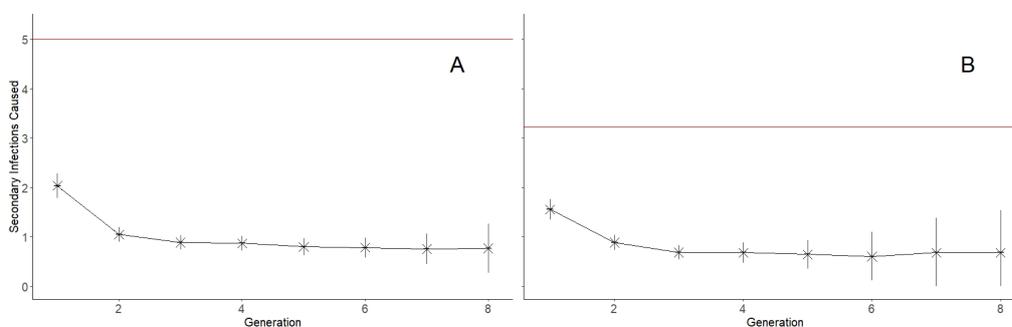


Figure 4.7. For community structure networks, the R_0 estimate consistently overestimates for all generations. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications ± 1 standard deviation.

It is possible that this poor performance is attributable to the highly structured nature of the simulations (it is well-documented that the R_0 is typically lower on structured networks [346] particularly if individuals are highly clustered [347, 348]). To evaluate this possibility, we instead examine their fully random counterparts. In these simulations the degree distribution is unchanged, but individuals are now randomly connected (see Section 4.5.1 and Figure 4.3). If we consider both transmission rates once more (Figure 4.8), we observe no real difference between the fully random and fully structured (Figure 4.7) simulations. Thus, the structured nature of the network appears to have no discernible impact on the performance of the analytic estimator.

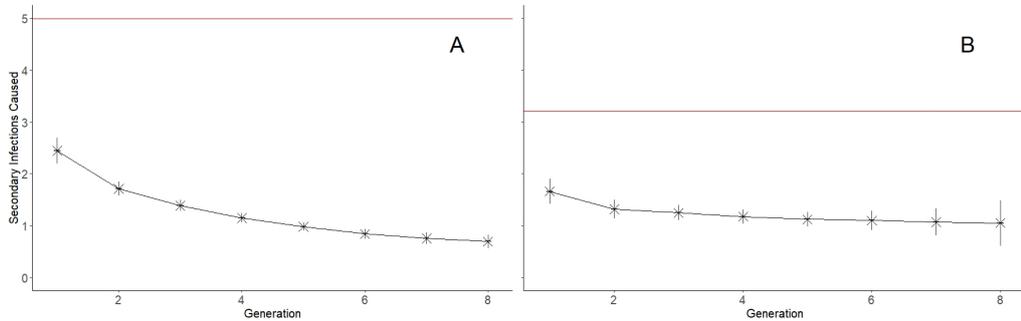


Figure 4.8. The R_0 estimate consistently overestimates for all generations, even once the community structure has been replaced with random connections. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications ± 1 standard deviation.

If we examine the other two transmission rate distributions for the fully structured simulations (Figure 4.9) we once again observe that the analytic estimate greatly overestimates the true number of secondary cases caused for all generations, although it performs slightly better for the bimodal distribution (Figure 4.9B) than for the uniform distribution (Figure 4.9A).

It would be informative to also examine a more conventional degree distribution and for this purpose we choose the geometric distribution (see Section 4.5.1). Returning once more to the fixed transmission rates (Figure 4.10) we note that the analytic estimate (Equation 4.20) still consistently overestimates the number of secondary infections caused for all generations and once again performs better for the lower transmission rate than the higher transmission rate.

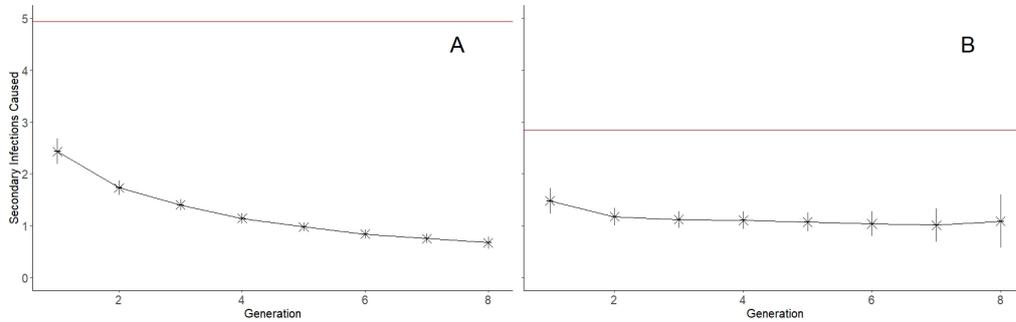


Figure 4.9. With either uniformly distributed or bimodally distributed transmission rates, the R_0 estimate consistently overestimates. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation.

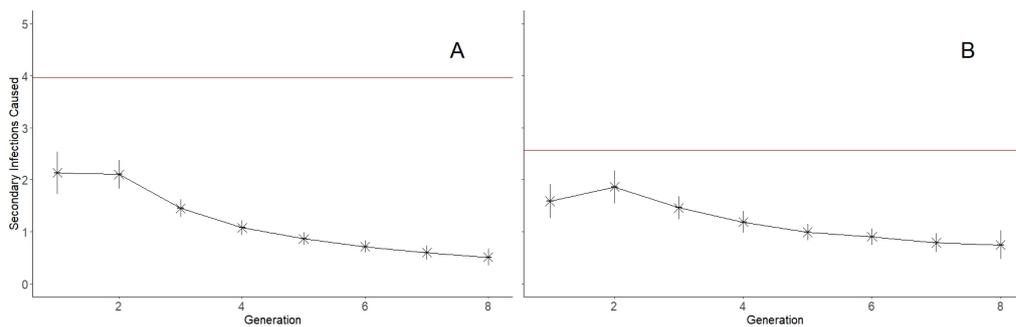


Figure 4.10. The R_0 estimate consistently overestimates for all generations, for both high and low fixed transmission rates and a geometric degree distribution. Mean number of secondary cases caused by infections of different generations (black) compared with the analytic estimate from Equation (4.20) (red). Whiskers indicate the mean across all replications +/- 1 standard deviation.

If we also examine the behaviour under the other two distributions (Supplementary Figure 1) we once again see this consistent overestimation, with better performance for the bimodal distribution than the uniform distribution. These results suggest that, regardless of the degree or transmission rate distribution chosen, the incorporation of the variance of the degree causes the estimate to overestimate the performance of the disease.

Nevertheless, there is another potential use of the analytic R_0 estimator: identifying the ‘epidemic threshold’. This threshold represents the minimum transmission rate for which one would expect the disease to spread successfully within the population. A simple requirement for this is:

$$R_0 \geq 1.$$

At the threshold we have $R_0 = 1$. If we consider the fixed transmission rate scenario and recall Equation (4.20), we have:

$$\begin{aligned} 1 &= (1 - (1 - \rho^*)^{\lambda^*}) \left(\hat{k} - 1 + \frac{\hat{k}}{\bar{k}} \right) \\ \frac{1}{\hat{k} - 1 + \frac{\hat{k}}{\bar{k}}} &= 1 - (1 - \rho^*)^{\lambda^*} \\ 1 - \rho^* &= \left(1 - \frac{1}{\hat{k} - 1 + \frac{\hat{k}}{\bar{k}}} \right)^{\frac{1}{\lambda^*}} \end{aligned}$$

Consider again the geometric degree distribution. We once again use a fixed infectious duration of 40 and recall that our degree distribution has (approximately) mean 3.88 and variance 11.17. Thus, we have:

$$\begin{aligned} \rho^* &\approx 1 - \left(1 - \frac{1}{5.76} \right)^{\frac{1}{40}} \\ &\approx 0.0048. \end{aligned}$$

It is important to note that this threshold does not represent a hard limit on transmission; that is to say, even if the transmission rate is below this value the disease is likely to persist in the population for a while. We would instead anticipate that such outbreaks would go extinct faster than those with transmission rates that are greater than the threshold. If we now examine a spectrum of transmission rates that spans both sides of the threshold we can

evaluate whether this expectation holds for our simulations. Consider first the proportion of simulations that go extinct within the first 120 timesteps (Figure 4.11A). We observe a marked difference between the transmission rates above and below the threshold, with a much greater proportion of lower transmission simulations going extinct (especially for the minimum value of 0.0015). Measuring the time to extinction (Figure 4.11B) reveals that simulations with lower transmission rates also go extinct earlier on average than those with transmission rates greater than the threshold.

We can calculate a similar threshold for the community structure simulations. It is trivial to demonstrate that we have:

$$\rho^* \approx 0.0038. \tag{4.35}$$

This threshold is identical for both the fully structured and fully randomised variants of the network because both share the same degree distribution. If we first examine the fully randomised community structure network (Figure 4.12) we observe a similarly marked difference between the above threshold and below threshold transmission rates. Once more, those below the threshold go extinct more often and faster than those above the threshold, also the transition appears more gradual than for the geometric degree distribution.

If we now examine the fully structured simulations (Figure 4.13), we can see the impact of the network structure. Although we do still observe a difference between the two halves, this difference is subdued in comparison; there is now a smooth transition from lower to higher transmission rates that renders that concept of a ‘threshold’ somewhat less applicable. This is likely due to the increase difficulty of maintaining the infection on a structured network, where the modular nature of the community creates additional opportunities for the disease to end up isolated with no further accessible susceptible individuals. Thus, we conclude that while the R_0 estimate is largely suitable for identifying the epidemic threshold rate of transmission for networks with heterogeneous degree distributions, caution must be taken if applying it to the most structured networks. For such networks, rather than having a threshold beyond which epidemics become likely (or even possible), there is instead a gradual increase in the likelihood of an epidemic over a broader range of transmission rates.

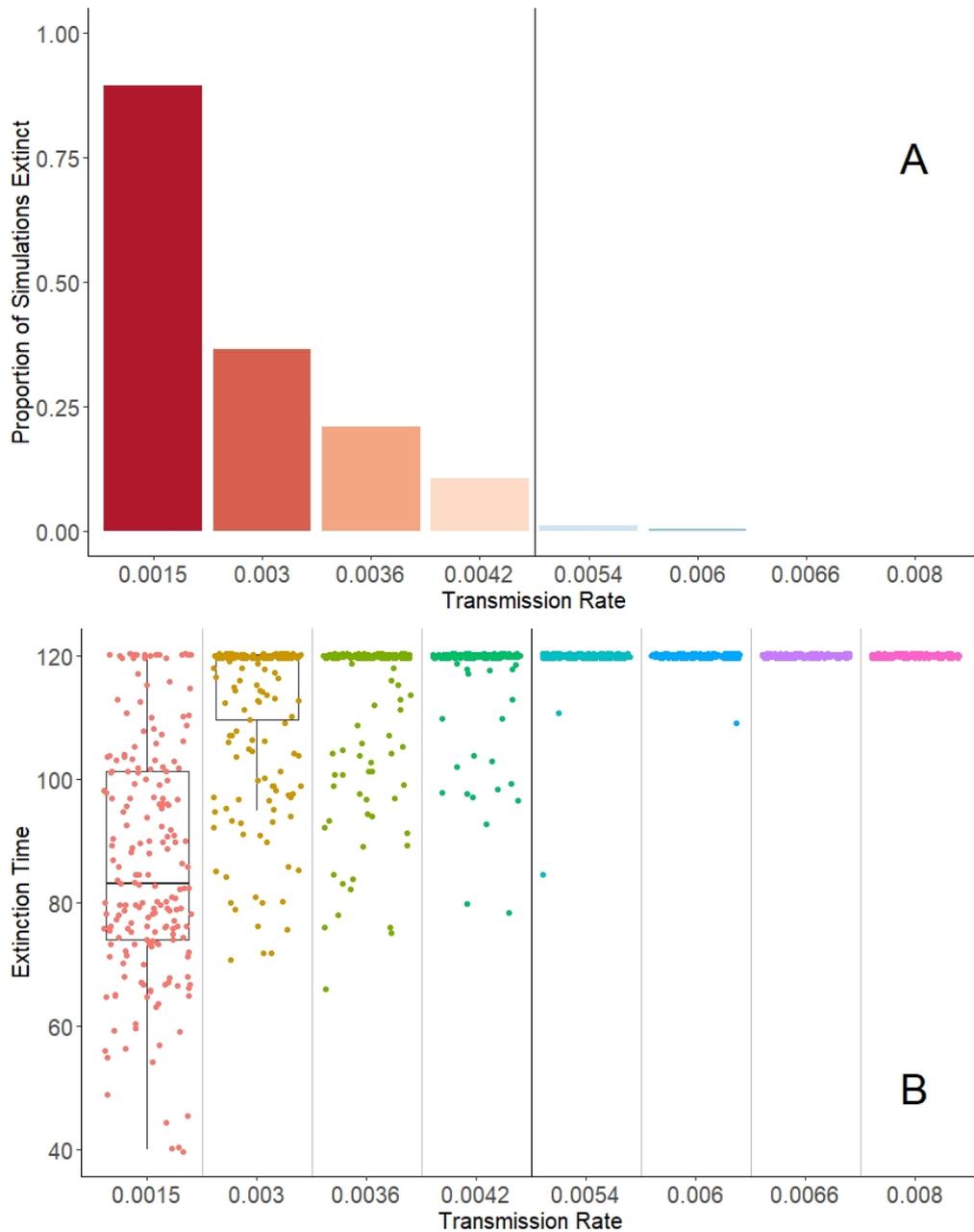


Figure 4.11. The R_0 estimate can successfully predict the epidemic threshold for a fixed transmission rate on a geometric degree distribution network. A: Proportion of simulations which experience extinction within the first 120 timesteps for various transmission rates. Vertical grey line indicates the predicted epidemic threshold. B: Time to extinction for individual simulations for various transmission rates. Solid black line indicates the predicted epidemic threshold. Both A and B apply to simulations with fixed transmission rate and geometric degree distribution.

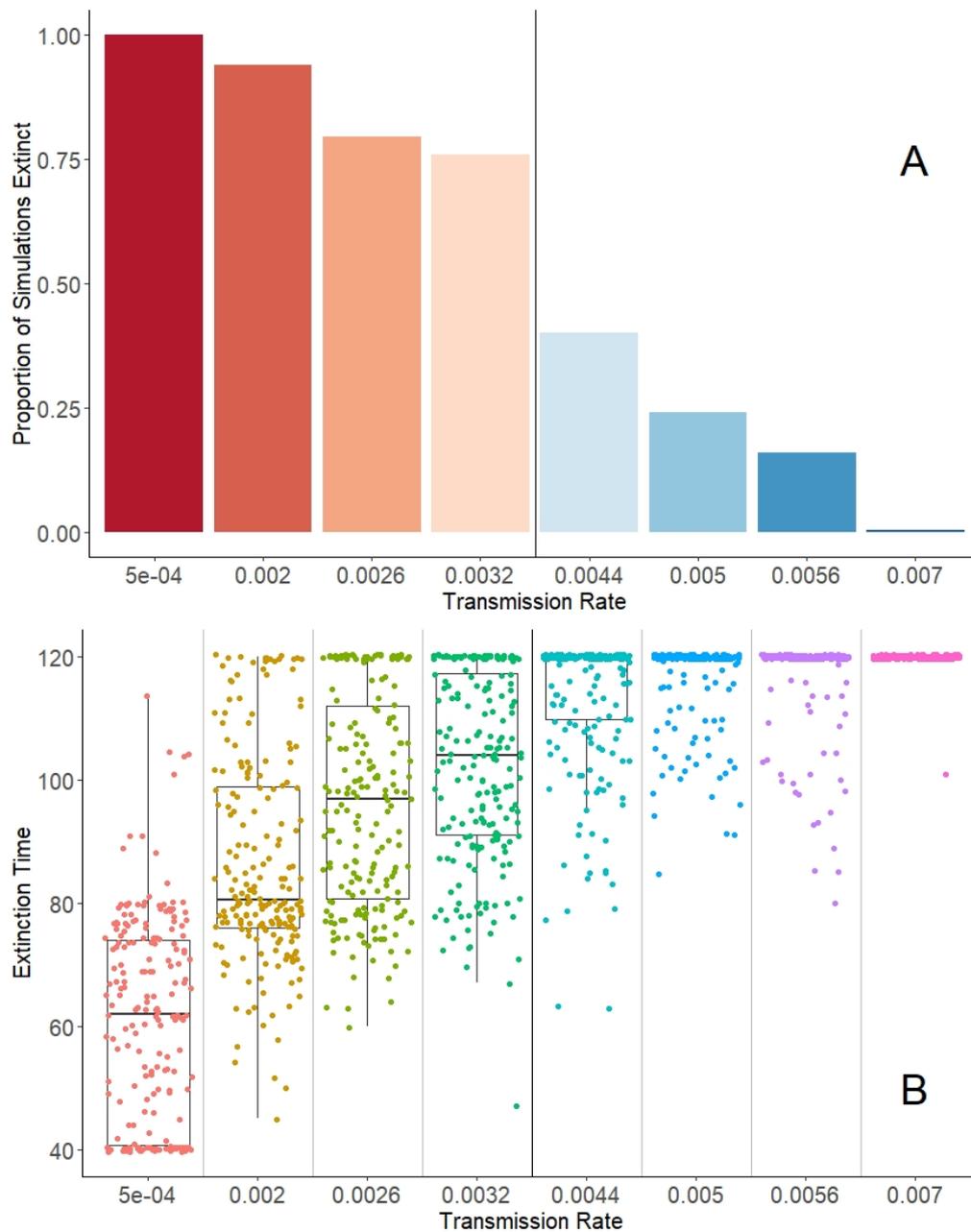


Figure 4.12. The R_0 estimate can successfully predict the epidemic threshold for a fixed transmission rate on the "fully random" community structure networks. A: Proportion of simulations which experience extinction within the first 120 timesteps for various transmission rates. Vertical grey line indicates the predicted epidemic threshold. B: Time to extinction for individual simulations for various transmission rates. Solid black line indicates the predicted epidemic threshold. Both A and B apply to simulations with fixed transmission rate and randomised community structure network.

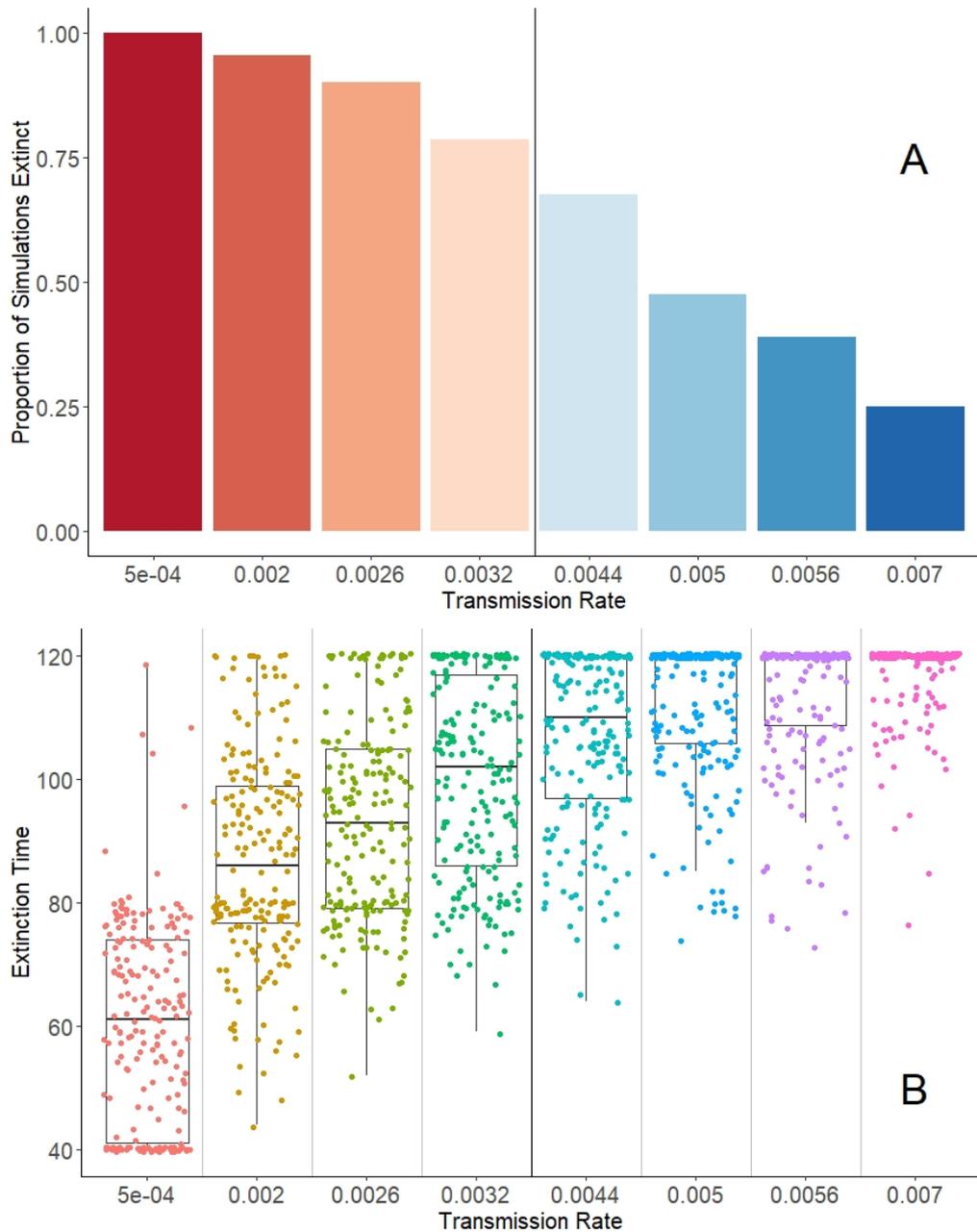


Figure 4.13. The R_0 estimate can still successfully predict the epidemic threshold for a fixed transmission rate on the "fully structured" community structure networks. A: Proportion of simulations which experience extinction within the first 120 timesteps for various transmission rates. Vertical grey line indicates the predicted epidemic threshold. B: Time to extinction for individual simulations for various transmission rates. Solid black line indicates the predicted epidemic threshold. Both A and B apply to simulations with fixed transmission rate and community structure network.

4.7 Discussion

Our numerical simulations suggest that our analytic R_0 estimators are poorly suited to predicting how a disease will spread on a heterogeneous network, even for the simplest transmission rate regimes (Figure 4.7) and well-known degree distributions (Figure 4.10), despite performing well for the initial infections on homogeneous networks (Figure 4.4). However, our results also suggest that it can be used to successfully predict the location of the epidemic threshold for heterogeneous networks (Figure 4.11), and thus is still a valuable tool for understanding disease transmission on such networks.

The extent to which our analytic estimate of R_0 overestimates for our simulations is striking. For example, in the context of the community structure simulations with the higher fixed transmission rate of 0.03 (Figure 4.7A), the analytic estimator predicted an R_0 of approximately 5. However, the mean degree for these simulations is 3.88 and thus it seems unreasonable to expect 5 secondary infections per infection on average, even if we assume a fully naive population (such that no two infections are neighbours). An explanation for this inaccuracy may be found in how the estimates were derived. Originally, Diekmann and Heesterbeek [344] used a measure of ‘activity’ ξ that, in the context of sexually transmitted infections, represented the number of contacts per unit time. In our derivation, we use the degree of the individual in the same manner as a measure of ‘activity’. A notable distinction between these approaches is that in our context, an individual is always connected to the same individuals and thus does not merely have k contacts per unit time, but rather the *same* k contacts every unit of time. In the original context, on the other hand, each contact is with a random individual (with more active individuals being more likely to be chosen).

One consequence of this is that the pool of accessible susceptible individuals is depleted potentially much faster in this context than in the free-mixing case. Consider, for example, the community structure: in each ‘household’ all individuals are interconnected. As a result, every individual infected by one member of the household is prevented from being infected by any other individual in the household. Thus, it is unlikely that any individual would cause as many secondary infections as one would expect from its degree; indeed, even if an individual did manage to do so, this would greatly reduce the number of infections that the other household members could cause. This is true more so for later generations of infections rather than for the initial infections (for whom the population is almost entirely naive). This effect has been documented in other analytic contexts [350, 351] as well as for diseases such as rabies [352]. It has important implications both to the spread of the disease [353] and to control measures such as contact tracing [354], and explains

why the analytic estimates derived here perform progressively worse over the course of the outbreak.

This effect cannot account for the overestimation of the R_0 value for the initial infections (for whom the population is almost entirely naive). Considering that the estimate does perform well for generation 1 of the homogeneous case (Figure 4.4), it seems likely that it is overvaluing the contributions of high-degree individuals to the R_0 value. This resembles previous findings that, without proper consideration, the contributions of highly active individuals to the R_0 value are overestimated [355]. In our context, although higher-degree individuals can cause more infections than lower-degree individuals, they are still restricted by their degree. Thus, so-called ‘superspreaders’ are only possible if there are individuals with suitably high degree present in the network. It is possible that in a more dynamic network, where the connections are shuffled periodically, the R_0 estimate would be more accurate.

On the other hand, the R_0 proved to be effective at predicting a suitably-defined epidemic threshold for both simpler degree distributions (Figure 4.11) and the community structured simulations (Figure 4.8), even when the network is highly structured (Figure 4.7). Care must be taken in defining and predicting epidemic thresholds: as Li et al. [356] remark and our results demonstrate, a disease can spread even if $R_0 < 1$ and may go extinct if $R_0 > 1$. This is why our epidemic threshold does not represent a hard limit on transmission, rather the approximate location of a ‘bifurcation’ of sorts in the behaviour of the system; as the transmission rate is reduced below it, the proportion of simulations that go extinct in a given time period increases rapidly.

If suitable care is taken in defining the threshold, it can prove a valuable tool. Previous studies have identified epidemic thresholds for a range of different systems [357–359], including bipartite networks [360] (particularly relevant to vector-borne diseases [361, 362]), as well as highlighting the vanishing of such thresholds for scale-free networks [363]. Network features such as the maximum degree [364], weighted edges [365] and degree heterogeneity [366] have all been found to be important to determining the epidemic threshold. Another priority has been to identify how the epidemic threshold is influenced by real-world factors such as behaviour. Studies have highlighted the importance of the movement of individual animals [367], the link between concurrent relationships and the spread of HIV [368] and the impact of behavioural changes in response to an emerging disease [369]. This interest is not restricted to disease outbreaks: misinformation, for example, has epidemics of its own [370] and given its role in hampering disease control [371], identifying methods to reduce the epidemic threshold for misinformation outbreaks has seen much interest [372], especially in the context of network segregation [373] and edge removal [374].

The primary aim of identifying the factors that influence the epidemic threshold is the same for both contexts: to devise methods (ideally, simple ones) of bringing transmission rates below the threshold and thus preventing an epidemic, for example by calculating the proportion of the population that needs to be vaccinated [375].

We can employ an epidemic threshold for leishmaniasis for a similar purpose. Previous studies have identified a wealth of potential risk factors (e.g. [147, 376] that could be incorporated into network models similar to those used here. The use of insecticide-treated nets (ITNs) [377] could be easily accounted for, with different scenarios of uptake or distribution represented by different transmission rate distributions. This could be used to evaluate the extent to which transmission between neighbours must be dampened by a net to prevent an epidemic, or perhaps what proportion of individuals need to use nets, providing another lens through which to examine the outstanding questions surrounding ITNs [4, 378] and to enhance the cost-effectiveness of insecticide-based approaches [282, 379]. Similarly, aspects such as the presence of habitat fragments [316, 380] and exposed resting spots [127] can be captured by suitably chosen network structures. Treatment of such areas to reduce the presence of sand flies could be represented via edge removal, allowing one to evaluate the extent to which the vector must be suppressed (see also [381]). Factors such as distance [146] may require more careful choices in terms of both transmission rates and degree distribution, but may yield insights in prioritising treatment or preventative methods and provide another alternative to expensive MCMC methods [198].

An estimate for the epidemic threshold may also prove useful in understanding the clustering of leishmaniasis. Leishmaniasis exhibits spatio-temporal clustering both at the community scale [12] and at the scale of larger provinces [382] (for both human [383] and zoonotic [91] leishmaniasis). Although studies have identified large-scale risk factors such as poor access to healthcare and stable employment [149], expansion of infrastructure (such as roads and pipelines) [121] and climate-related factors such as temperature [93], it remains non-trivial to identify specific vulnerable communities. Given the importance of smaller-scale factors such as sand fly population density [384] and housing conditions [131] it is likely that susceptibility to leishmaniasis invasion will vary from settlement to settlement, even under consistent large-scale conditions (for example, within a given municipality). If one could represent communities of interest as networks, informed by factors such as inter-household distances or sand fly densities, then the epidemic threshold estimate could be used to evaluate which communities are most likely to experience an epidemic in the event that leishmaniasis were to be introduced to the community (via human migration [242, 385], for instance). This would allow the prediction of potential

transmission hotspots, and thus the targeting of interventions, which would be particularly useful in areas with fewer resources for monitoring and suppressing leishmaniasis [386].

It is unfortunate that our R_0 estimates proved unsuitable beyond the initial infections. Leishmaniasis exists at an endemic state across many communities [387, 388]. The prevalence of the infection can vary greatly across different endemic communities: rural communities in Ghana [389] and Ethiopia [309] experience very high rates of leishmaniasis infection and communities in Bihar, India have been found to display moderate levels of post-kala-azar dermal leishmaniasis (PKDL) [390], whereas in southern France [391] and Spain [392] leishmaniasis persists at much lower levels (though can still cause acute outbreaks [393]). There remains much debate as to how leishmaniasis is maintained in the many endemic communities, despite concerted elimination efforts [394, 395]. Repeated spillover events from wild [396], synanthropic [18] and domestic [397] reservoir species are often cited as an important factor. Questions linger also on the importance of asymptomatic infections [398, 399] and cases of PKDL [242]. The conundrum is further complicated by the sometimes low infection prevalence within the sand fly vectors [400] and potential for other vector species [22]. If we are to fully eradicate leishmaniasis, it is clear we must further our knowledge of this endemic state and how it persists.

There are other more promising approaches to capturing the endemic state. Identifying endemic states for network models has a long history [401] in a range of contexts [402–404]. Previous work focusing on SIS dynamics [405] have yielded insights by identifying an endemic threshold [406] or using the quasi-stationary distribution for Markovian SIS dynamics [407] to link invasion probability to the endemic state. SIR models [408] are less studied; more prone to cycles of local extinction and reintroduction [409] than ending in an endemic state, but introducing population turnover or dynamic networks has yielded interesting results [410, 411]. Recently, an analytic estimate [151] of the endemic state of the Super Compact Pairwise model [152] has also been derived. Thus, there are a number of opportunities for further study of the endemic state of leishmaniasis that future studies could explore.

4.8 Conclusion

Although the analytic estimates for R_0 proved ineffective at predicting the true number of secondary cases caused by a single infection, they were not without application. We find that they can be repurposed to predict an epidemic threshold for the rate of transmission with reasonable accuracy (though it should be acknowledged that network structure still impacts their performance). In future, these thresholds could be used to identify neigh-

bourhoods that are particularly at risk of invasion, if suitable networks are constructed (ideally informed by data about transmission factors of interest, such as inter-household distance or vector densities).

4.9 Materials and Methods

All numerical simulations were performed using MATLAB R2021b (v9.11.0.1769968). All associated data analysis was performed using RStudio v2022.07.1, Build 554 (R version 4.1.0).

Chapter 5

Heterogeneity in individual degree distributions allows spontaneous deviations from a stable equilibrium in *Leishmania* epidemics on networks.

5.1 Abstract

Leishmaniasis is a neglected tropical disease caused by *Leishmania* parasites that is responsible for thousands of deaths every year in countries across the world. Despite advances in sand fly control and a better understanding of leishmaniasis transmission dynamics, full elimination remains an elusive goal in the many communities where the disease is endemic. For this goal to be achieved, we require better knowledge of how the disease persists at low levels. Network models represent a useful tool for approaching this matter, having seen prior success in modelling disease transmission in structured communities. We evaluate the efficacy of an analytic estimator for the endemic equilibrium of the Super Compact Pairwise model when applied to networks of varying structure. We also assess the transmission dynamics around this equilibrium and use this to highlight scenarios of particular concern for transmission of leishmaniasis, such as high-density, poor-quality housing or communities that depend strongly on agricultural work or seasonal migrant labour.

5.2 Introduction

The global burden of endemic diseases has garnered much attention over the last decade. This burden is perhaps most keenly felt in less economically developed regions. The impact of diseases such as cholera [412] and norovirus [413] is compounded further by a myriad of tropical diseases, ranging from well-known threats such as malaria [414,415] and dengue fever [416] to the more obscure lymphatic filariasis [417]. These impacts are far-reaching, with both health [412] and economic [418] implications. Of particular concern are the neglected tropical diseases [419–421]. Considerable effort has been devoted more recently to monitoring [422], diagnosing [423,424] and treating [425] them, often as part of elimination drives [426]. Despite this, our understanding of their transmission and the full extent of their impacts remains incomplete.

One such disease is leishmaniasis, which remains a disease with global presence [427], despite ongoing efforts to eliminate it. Countries such as Brazil [428] and India [429] bear much of the burden, though its reach is far wider [9,430]. Although progress towards the elimination targets has been made [395], due to intervention methods such as indoor residue spraying and insecticide-treated bed nets [431] and advancements in treatment and diagnosis [432], leishmaniasis continues to manifest as low-level transmission and sporadic novel outbreaks. Such outbreaks are present even in regions with minimal transmission, such as Spain [433].

Leishmaniasis transmission appears to be heterogeneous at multiple scales. At large scales, cases cluster in regions of greater social vulnerability [149] and areas with favourable conditions for the sand fly vectors [93], with expansion potentially facilitated by transport links [121] and urbanisation efforts [434]. Similar clustering of cases has long been reported at the scale of individual communities, too [147]. Early studies identified numerous risk factors associated with transmission [147,201], though a complete understanding of community transmission yet eludes us. A few factors have attracted particular attention. The first is the proximity to existing infections, oft-reported as having a strong effect on the likelihood of transmission [141,147,435], and this was further supported by recent modelling efforts [146]. A related factor is the role of reservoir species. Dogs are a well-studied reservoir species in urban environments [19] and the variety of viable mammalian hosts [18] has spurred efforts to identify other potential wild reservoir species [20]. The role of sand fly habitat fragments is also of interest [436], given the clustering of cases in proximity to them [233] and increasing land use intensification [316]. The core strand that connects these is proximity, perhaps unsurprising as sand flies are typically ineffective fliers [70] that prefer to travel only short distances [69,72].

This strand presents a challenge for modelling, but also an opportunity: to represent a community as a network of individuals connected based on factors such as distance. Network models [334] are well suited to capturing transmission dynamics that would be difficult to capture using traditional mass action models [437]. They have a rich history of applications to modelling disease transmission [438–440]. The well-known SIS [441] and SIR [442] models have translated well onto networks, and can be comfortably expanded to include aspects such as infection delays [443]. Networks need not be static [444] and such dynamic networks can capture changes in behaviour [445], particularly in terms of social contacts [351, 446]. Bipartite networks have been used to capture the dynamics of some vector-borne diseases [361], often in pursuit of an estimate for the basic reproduction number R_0 [362, 447]. Despite this success, network models have seen minimal application in the context of leishmaniasis.

Network structure has been explored through a variety of measures [448]. One measure with potential in modelling leishmaniasis is ‘modularity’ [449]. Modularity appears to be especially important in the context of social groups ranging from those of ants [450] to those of humans [451]. The impact of modularity is nuanced [452], especially when combined with other structural factors [453, 454], and has implications for the maintenance of endemic disease [343, 455]. Also of interest are ‘small-world’ networks, which are structured similarly to modular networks but also have long-range connections [438]. Relevant to disease transmission in both animal [336] and human [335, 456] populations, modelling efforts have shown small-world networks to be surprisingly vulnerable to disease transmission [336, 457]. Both have relevance in the context of leishmaniasis: many communities in which leishmaniasis is endemic [147] appear modular, but long-distance routine migration [12] opens the possibility for small-world behaviour. These networks can be captured by a spectrum of ‘structuredness’, with very structured modular networks at one end, fully random configuration networks (where all individuals are connected at random with no predetermined organisation beyond their degrees) at the other, and small-world networks falling somewhere between the two. Insights can be gleaned from assessing how different models perform across this spectrum.

We evaluate the performance of a recently derived analytic estimate [151] for the endemic equilibrium of the Super Compact Pairwise (SCPW) model [152] under a spectrum of differently structured networks. We compare the expected endemic equilibrium with the output of numeric simulations and then determine how epidemiologically relevant quantities are affected by the structuredness of the system.

5.3 Endemic Steady State, Analytic Predictions

Consider first the SIS Model, where each individual is either infected or susceptible and reverts to the susceptible state upon recovery. We assume that an infected individual infects a susceptible neighbour at some rate τ and recovers at some rate γ , both of which we initially assume are constant. We also assume that either there is no long-term immunity to the disease, or that the population turnover rate is sufficient to render any immunity negligible to the transmission dynamics.

One of the most tractable pairwise models is the Super Compact Pairwise (SCPW) Model, derived by Simon and Kiss [152]. It is defined as:

$$[\dot{S}] = \gamma[I] - \tau[SI], \quad (5.1)$$

$$[\dot{I}] = \tau[SI] - \gamma[I], \quad (5.2)$$

$$[\dot{SI}] = \gamma([II] - [SI]) - \tau[SI] + \tau[SI]([SS] - [SI])Q, \quad (5.3)$$

$$[\dot{SS}] = 2\gamma[SI] - 2\tau[SI][SS]Q, \quad (5.4)$$

$$[\dot{II}] = -2\gamma[II] + 2\tau[SI] + 2\tau[SI]^2Q \quad (5.5)$$

where:

$$Q = \frac{1}{n_S[S]} \left(\frac{\langle k^2 \rangle (\langle k^2 \rangle - \langle k \rangle n_S) + \langle k^3 \rangle (n_S - \langle k \rangle)}{n_S (\langle k^2 \rangle - \langle k \rangle^2)} - 1 \right), \quad n_S = \frac{[SI] + [SS]}{[S]}, \quad (5.6)$$

and $\langle k^n \rangle$ is the n^{th} moment of the degree distribution, τ the transmission rate and γ the recovery rate. In this notation, $[S]$ is the number of susceptible individuals, $[I]$ is the number of infected individuals, $[SI]$ is the number of edges connecting a susceptible individual to an infected individual, $[SS]$ is the number of edges connecting two susceptible individuals, and $[II]$ is the number of edges connecting two infected individuals. $\langle k^n \rangle$, and Q as a whole, emerge from the choice of closure at the level of triples used by Simon and Kiss [152]. The derivation is lengthy and beyond the scope of our work, but see [152] for full details and note also that they refer to $\langle k \rangle$ as n_1 , $\langle k^2 \rangle$ as n_2 and $\langle k^3 \rangle$ as n_3 throughout.

Simon and Kiss then compared the performance of the SCPW with that of both the original pairwise and compact pairwise models for an array of networks. The first set of networks chosen were bimodal networks in which each node was either of a low ($k_1 = 5$) or high ($k_2 = 35$) degree. Three different bimodal networks were considered: one with 10% k_1 nodes and 90% k_2 nodes, one with an equal proportion of k_1 and k_2 nodes, and one with 90% k_1 and 10% k_2 nodes. The SCPW was found to be almost identical to the compact pairwise model in all contexts, and performed much better than the original

pairwise model. The second set of networks examined were constructed such that the nodes follow a cutoff power law degree distribution with minimum degree k_{min} , maximum degree k_{max} and some power α . The probability of taking a given degree is:

$$p(k) = Ck^{-\alpha} \quad \forall k = k_{min}, k_{min} + 1, \dots, k_{max} \quad (5.7)$$

where:

$$\frac{1}{C} = \sum_{k=k_{min}}^{k_{max}} k^{-\alpha}. \quad (5.8)$$

Both a sparse ($k_{min} = 5, k_{max} = 30$) and dense ($k_{min} = 10, k_{max} = 140$) case were considered (with $\alpha = 2$ in both cases). In each case, the SCPW was found to provide excellent agreement with the compact pairwise model and much better performance than the original pairwise model.

Recently, Corcoran and Hastings [151] sought to identify implications for disease control by deriving an estimate for the endemic equilibrium of the SCPW and evaluating its sensitivity to network parameters such as the mean degree. First, they obtained an estimate for the threshold in transmission rate at which a stable endemic equilibrium emerges. By defining:

$$v = \frac{[S]}{N}, \quad w = \frac{[I]}{N}, \quad x = \frac{[SI]}{\langle k \rangle N}, \quad y = \frac{[SS]}{\langle k \rangle N}, \quad z = \frac{[II]}{\langle k \rangle N}$$

they converted Equations 5.1-5.5 into the following dimensionless form:

$$\dot{v} = w - \langle k \rangle \delta x, \quad (5.9)$$

$$\dot{w} = \langle k \rangle \delta x - w, \quad (5.10)$$

$$\dot{x} = z - (\delta + 1)x + \frac{\alpha \delta v x (y - x)}{\langle k \rangle (x + y)^2} + \frac{\beta \delta x (y - x)}{(x + y)}, \quad (5.11)$$

$$\dot{y} = 2x - \frac{2\alpha \delta v x y}{\langle k \rangle (x + y)^2} - \frac{2\beta \delta x y}{x + y}, \quad (5.12)$$

$$\dot{z} = -2z + 2\delta x + \frac{2\alpha \delta v x^2}{\langle k \rangle (x + y)^2} + \frac{2\beta \delta x^2}{x + y}, \quad (5.13)$$

where:

$$\delta = \frac{\tau}{\gamma}, \quad \alpha = \frac{\langle k^2 \rangle^2 - \langle k \rangle \langle k^3 \rangle}{\langle k^2 \rangle - \langle k \rangle^2}, \quad \beta = \frac{\langle k^3 \rangle - \langle k^2 \rangle \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle^2} - 1.$$

In this dimensionless form, w represents the proportion of individuals that are susceptible, v represents the proportion of individuals that are infected, x is the proportion of edges that connect a susceptible individual to an infected individual, y is the proportion of edges that connect two susceptible individuals and z is the proportion of edges that connect two infected individuals.

This gave rise to two key results. The first is the ‘endemic threshold’ δ_c . δ_c gives a measure of the minimum transmission rate required for a disease to become endemic within a population and thus persist long-term. In this context, δ_c is given by:

$$\delta_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \quad (5.14)$$

For all $\delta < \delta_c$ the only stable fixed point is the disease-free equilibrium (DFE) and the endemic equilibrium is unstable, whereas for $\delta > \delta_c$ the endemic equilibrium is stable and strictly positive (i.e. $w > 0$) and the DFE is unstable. Thus, a transcritical bifurcation occurs at δ_c : a stable and an unstable fixed point converge as δ increases, coalescing at δ_c and separating having swapped their stabilities (for further details see [458], for example). Although it was not possible to obtain an analytic estimate of the endemic equilibrium for all scenarios, Corcoran and Hastings [151] did find suitable estimates for two important cases. When $\delta \approx \delta_c$, the proportion of the population infected at endemic equilibrium (w^*) was:

$$w^* \approx \frac{\sigma\eta}{\lambda\sigma + \mu\delta_c + \mu - \delta_c} + O(\eta^2), \quad (5.15)$$

where:

$$\sigma = \langle k \rangle \delta_c, \quad \lambda = \frac{\alpha \delta_c}{\langle k \rangle}, \quad \mu = \beta \delta_c, \quad \eta = 1 - \frac{\delta_c}{\delta}.$$

For $\delta \gg \delta_c$, on the other hand, they found that:

$$w^* \approx 1 + \frac{\delta_c + \mu - \sigma}{\lambda\sigma} \epsilon + O(\epsilon^2), \quad (5.16)$$

where:

$$\epsilon = \frac{\delta_c}{\delta}.$$

They then tested the sensitivity of the endemic equilibrium to changes in $\langle k \rangle$ and $\langle k^2 \rangle$, with the aim of identifying conditions in which the endemic equilibrium could be reduced. In the case where $\delta \approx \delta_c$, they found that the endemic equilibrium could be reduced by increasing $\langle k \rangle$ or reducing $\langle k^2 \rangle$, whereas in the case where $\delta \gg \delta_c$ the opposite was true.

5.4 Application to *Leishmania* Transmission

5.4.1 Network models in the context of leishmaniasis transmission.

We build upon the work of Corcoran and Hastings [151] by evaluating the accuracy of their analytic estimates for the endemic equilibrium under different regimes of network ‘structuredness’. Initially, a fully structured network is generated. Individuals are assigned to a specific household and para (a collection of households that are clustered together). This approach is intended to mimic the structure of the community examined by Bern *et al.* [147]. In order to replicate conditions where transmission typically occurs over short distances [146] we initially connect each individual to all other individuals in their household. Within each household, two individuals will have a single connection to an individual in another household. All other individuals have no inter-household connections. This configuration of individuals is the ‘fully structured’ community.

In order to produce a spectrum of differently structured networks, we then remove different proportions of the edges in the network, leaving stubs for the individuals involved. These stubs are then reconnected at random, following standard procedure for a configuration model [334]. The proportion of edges randomised runs from the minimum of 0% (the fully structured community, resembling a highly modular network) to 100% (the completely random configuration model) in increments of 10%. For some example networks and an associated set of epidemic trajectories, see section 5.4.2. Throughout our analysis, we define ‘structuredness’ as follows:

$$\text{structuredness} = 1 - \rho, \quad \rho \in [0, 1], \quad (5.17)$$

where ρ is the proportion of edges randomised, such that a fully random network ($\rho = 1$) has a structuredness value of 0, and the fully structured network ($\rho = 0$) has a value of 1.

Each simulation uses an independent network. In each run, the population is constructed using 1200 individuals, divided into 3 ‘paras’ (collections of households) and into

households of varying sizes. Household size is drawn from a uniform random variable with maximum 6 and minimum 3 (the exception is the final household, which uses all remaining individuals once less than 6 remain to be allocated). Each simulation begins with 30 randomly chosen individuals as initial infections. The transmission dynamics are simulated in discrete time, where one timestep is equivalent to a week. The simulation then runs for 2000 timesteps, and we sample the final 500 timesteps to acquire the endemic equilibrium (with extinct simulations discarded). Each timestep, every infected individual has a fixed recovery chance and each neighbour has a fixed chance of being infected (a per-neighbour, per-timestep infection chance). All simulations use a fixed recovery rate of 0.025 per timestep for an average duration of infection of 40 weeks, informed by prior work indicating that outward transmission can occur months or even years after infection [459]. The per-neighbour, per-timestep transmission rates used here range from 0.008 to 0.015 in increments of 0.0007. The minimum value was chosen to be slightly higher than the minimal possible value for which the endemic equilibrium was stable (see Equations 5.9-5.14 for details). The maximum value was chosen to be small: although transmission is complex [45, 54, 146, 172] and there are many unknowns, it appears to be rare, with few sand flies found to be infected [460] and low leishmaniasis incidence rates [133, 461]. Finally, for our analysis we use sets of 50 independent simulations for each level of structuredness and each transmission rate.

5.4.2 Sample Networks and Trajectories

It is beneficial at this point to examine some sample networks and their corresponding disease transmission dynamics. The starkest differences exist between the two extremes of our spectrum. If we compare the fully random network (Figure 5.1A) to the fully structured network (Figure 5.1B) we see pronounced differences in both network layout and the associated transmission dynamics. The structured network is effectively a ring of small groups of well connected nodes. Such networks can be considered modular (see [449]). The unstructured network more closely resembles a single, large cluster of randomly connected individuals with minimal other organisational features. We also note that the structured simulations typically have fewer concurrent infections than their unstructured counterparts.

It is also informative to examine certain intermediate values. Consider another pair: one with 50% of edges randomised (Figure 5.2A) and one with 10% of edges randomised (Figure 5.2B). We observe that with 50% of edges randomised, we have mostly lost the

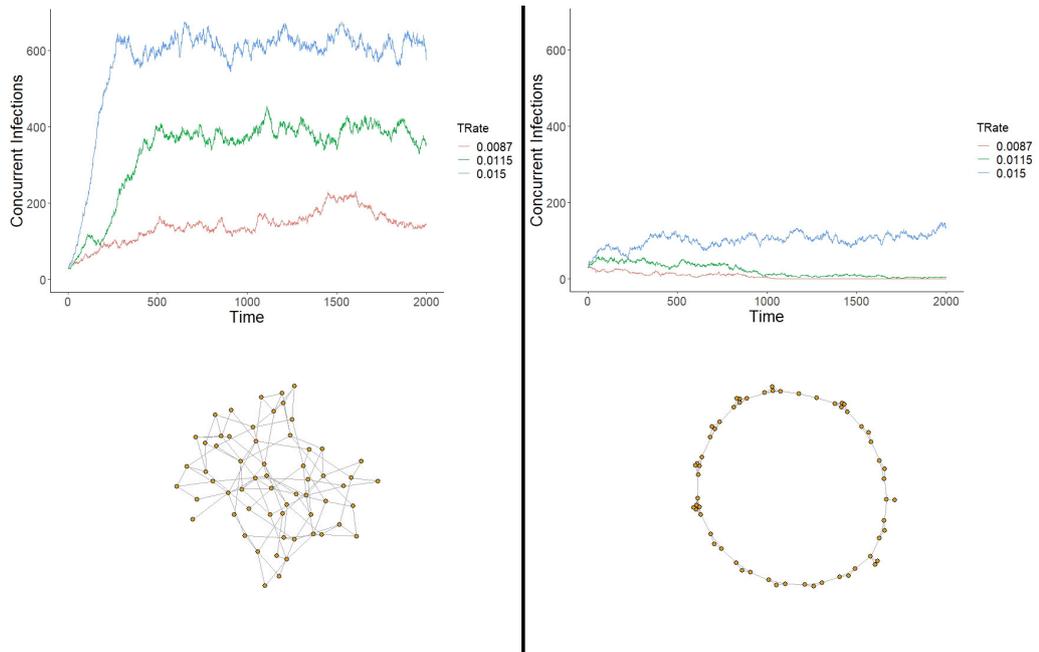


Figure 5.1. Example infection trajectory (top) and sample network (bottom) for the fully random (left, structuredness = 0.0) and fully structured (right, structuredness = 1.0) simulation types, as detailed in Section 5.4.1. Each node in the network samples represents an individual.

modular structure, though small clusters of interconnected individuals can still manifest. With 10% of edges randomised, the modular structure is still mostly visible, though some of the modules are now more firmly connected. We notice that, as in Figure 5.1, the more structured simulations have fewer concurrent infections than their more random counterparts.

5.5 Results

We first examine the mean endemic equilibrium values for our simulations for different structuredness values and transmission rates (Figure 5.3). We find that as the structuredness of the simulation increases, the analytic prediction (given by Equation 5.15) increasingly overestimates the proportion of infected individuals at endemic equilibrium. Although the endemic equilibrium of less structured simulations is predicted more accurately (particularly in the case of the fully random network), the analytic estimate is still an overestimate. We also see this reflected in Figure 5.4 which uses a log scale (allowing us to more easily examine the difference between the simulations at lower infection/recovery rates).

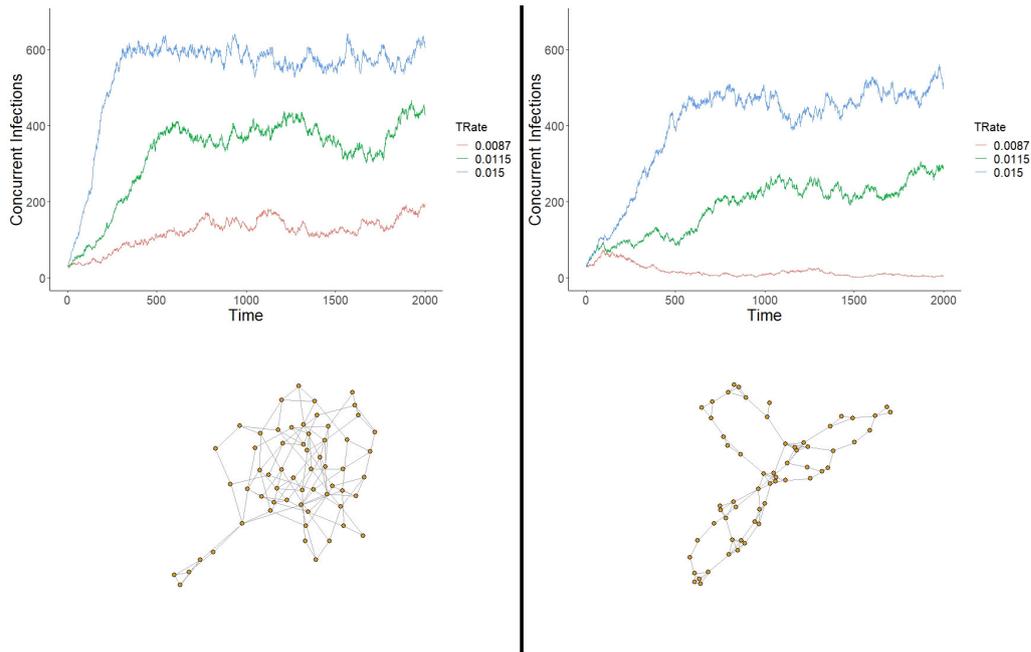


Figure 5.2. Example infection trajectory (top) and sample network (bottom) for the moderate structure (left, structuredness = 0.5) and high structure (right, structuredness = 0.9) simulation types, as detailed in Section 5.4.1. Each node in the network samples represents an individual.

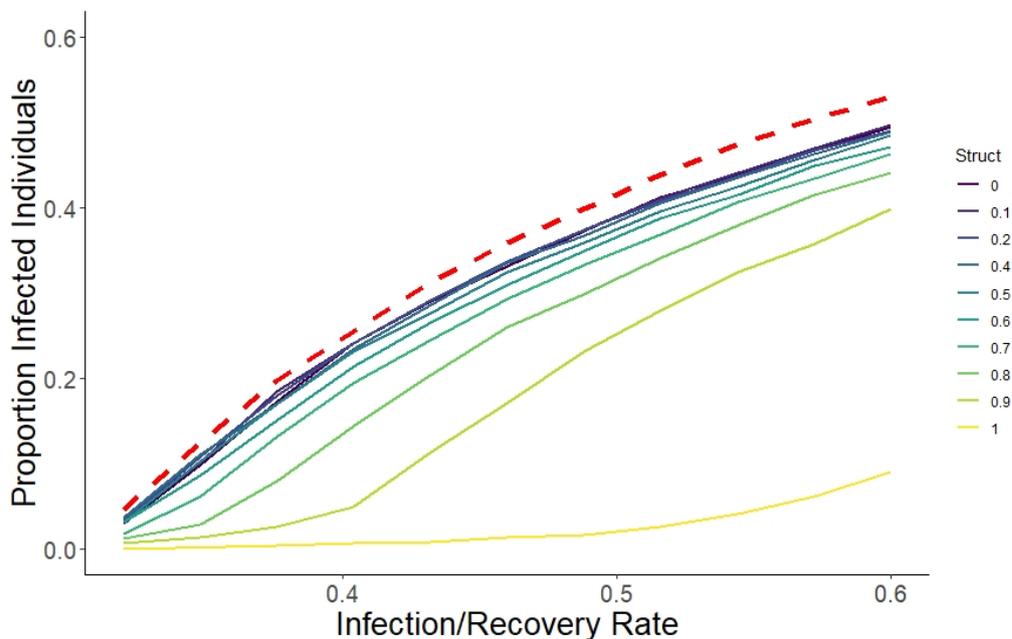


Figure 5.3. The analytic predictor has greater accuracy on simulations with lower structuredness. Average proportion of individuals infected at endemic equilibrium under various transmission rates. Output from numeric simulations of varying structuredness (see figure legend, Equation 5.17) is compared to the analytic estimate (red, dashed) provided by Equation 5.15.

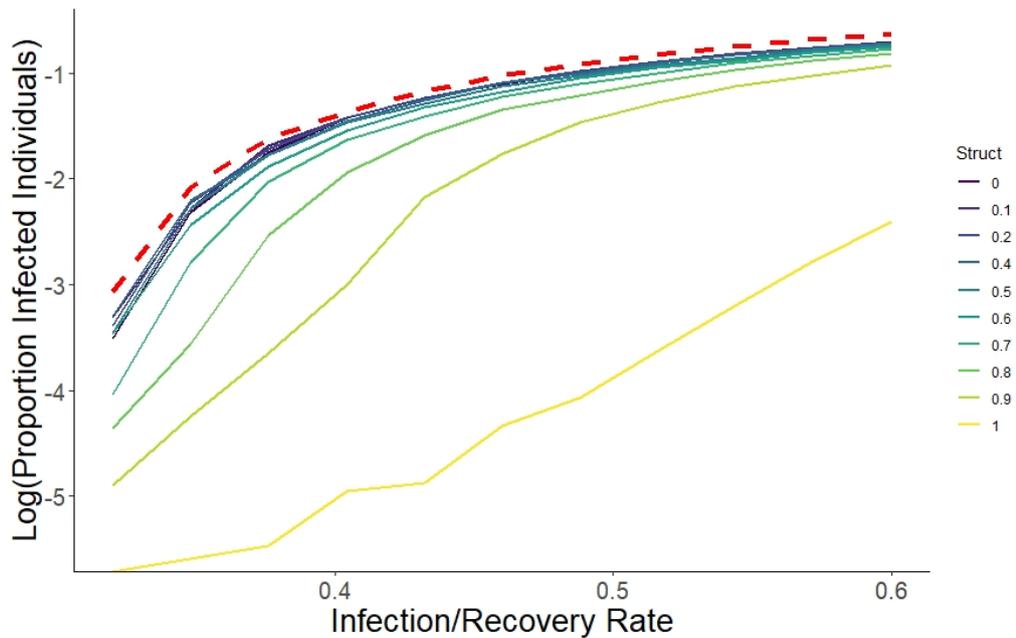


Figure 5.4. The analytic predictor has greater accuracy on simulations with lower structuredness. Average proportion of individuals (plotted as a log scale) infected at endemic equilibrium under various transmission rates. Output from numeric simulations of varying structuredness (see figure legend, Equation 5.17) is compared to the analytic estimate (red, dashed) provided by Equation 5.15.

Whilst the location of the endemic equilibrium provides valuable knowledge it is also beneficial to understand how robust the endemic equilibrium is (where we consider a robust equilibrium to be one where the number of infections typically remains close to the equilibrium with fewer, mostly short-lived, deviations). One measure of interest is the frequency at which the number of infections makes an ‘excursion’ away from the endemic equilibrium. We define an excursion to begin when the number of infections exceeds a specific threshold: 1 standard deviation higher than the endemic equilibrium. For an example of an excursion, see Figure 5.5. We consider now 3 different transmission rates: the maximum value of 0.015, the moderate value of 0.0115, and a low value of 0.0087.

We first examine the behaviour of the maximum transmission rate simulations. If we examine the number of excursions per simulation (Figure 5.6) we find that the most structured simulations have, on average, slightly fewer excursions than the unstructured simulations. We can further examine this by applying the two-sample Kolmogorov-Smirnov test. If we take the number of excursions for the unstructured simulation to be our reference, we can compare it to the distributions of the other structuredness values and identify any significant differences. The p-values for all significant tests are summarised in Table 5.1.

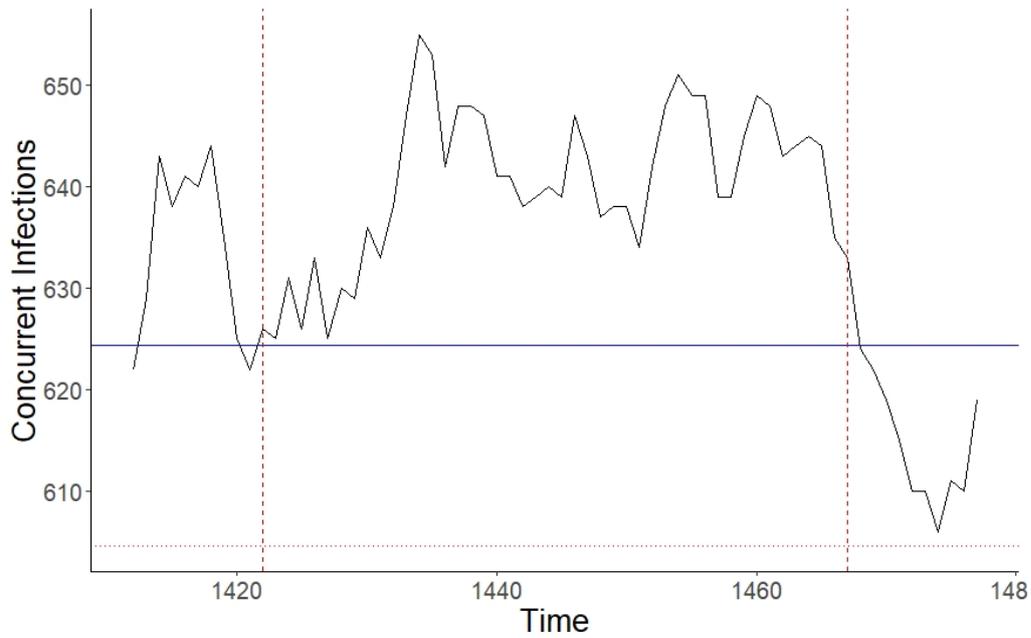


Figure 5.5. Sample excursion from a simulation with structuredness 0 and transmission rate 0.015. Excursion start/end times are denoted by the vertical red dashed lines, the endemic equilibrium is denoted by a horizontal dotted red line, and the threshold that must be exceeded to start an excursion is the solid blue line.

The K-S tests draw the same conclusion: highly structured simulations experience fewer excursions on average than their unstructured counterparts.

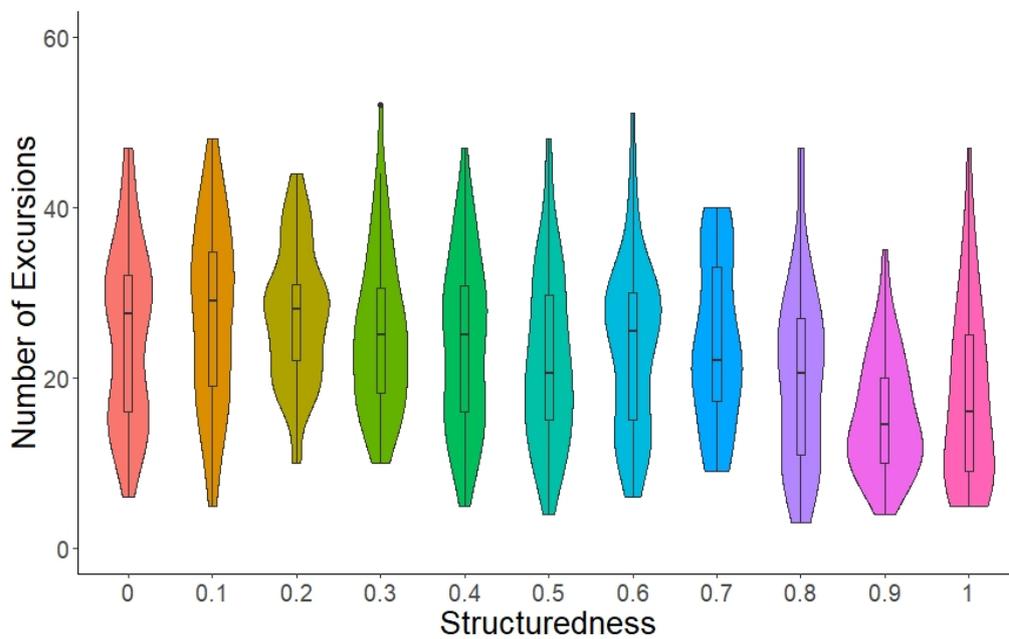


Figure 5.6. Highly structured simulations experience fewer excursions on average. Combined boxplot and violin plot of the number of excursions per simulation, grouped according to simulation structuredness. All simulations have the maximum transmission rate of 0.015 per timestep per neighbour.

Structuredness	P-value
1.0	0.01195
0.9	< 0.001
0.8	0.02222

Table 5.1. Kolmogorov-Smirnov tests verify that very structured simulations experience more excursions than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with maximum transmission rate 0.015. Non-significant tests are excluded from this table.

This is also visible for the moderate transmission rate (Figure 5.7). The most structured simulations once again have a lower number of excursions. This is confirmed by the K-S test (Table 5.2) which once again finds that structured simulations have significantly fewer excursions than the unstructured ones. The lowest transmission rate simulations also show this trend (Figure 5.8, Table 5.2), although many of the most structured simulations now go extinct.

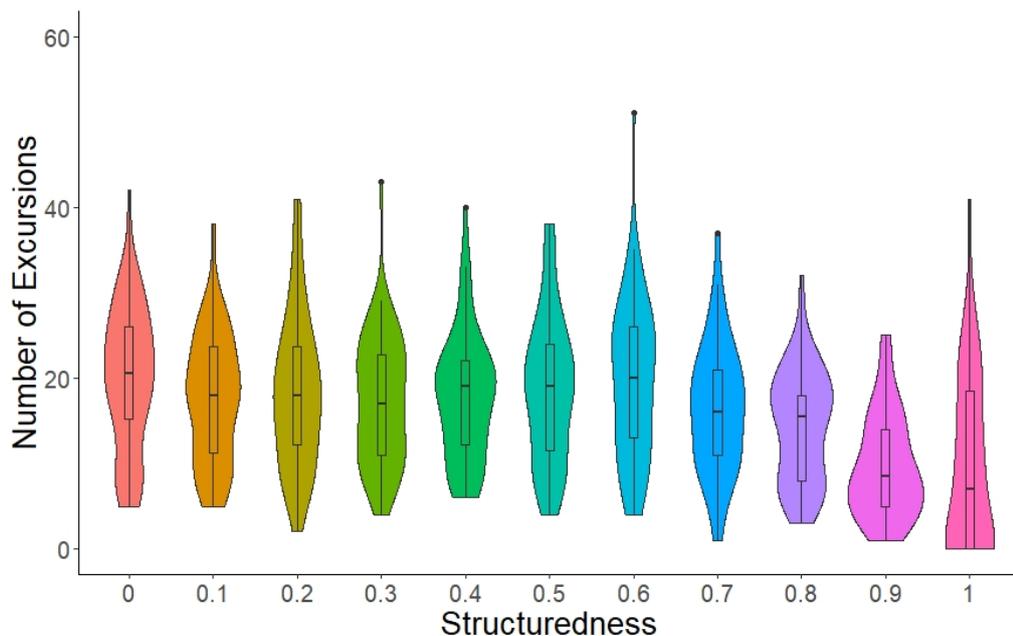


Figure 5.7. Under a reduced transmission rate, highly structured simulations still experience fewer excursions on average. Combined boxplot and violin plot of the number of excursions per simulation, grouped according to simulation structuredness. All simulations have the moderate transmission rate of 0.0115 per timestep per neighbour.

Structuredness	P-value
1.0	< 0.001
0.9	< 0.001
0.8	0.0062

Table 5.2. Kolmogorov-Smirnov tests verify that very structured simulations experience more excursions than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with moderate transmission rate 0.0115. Non-significant tests are excluded from this table.

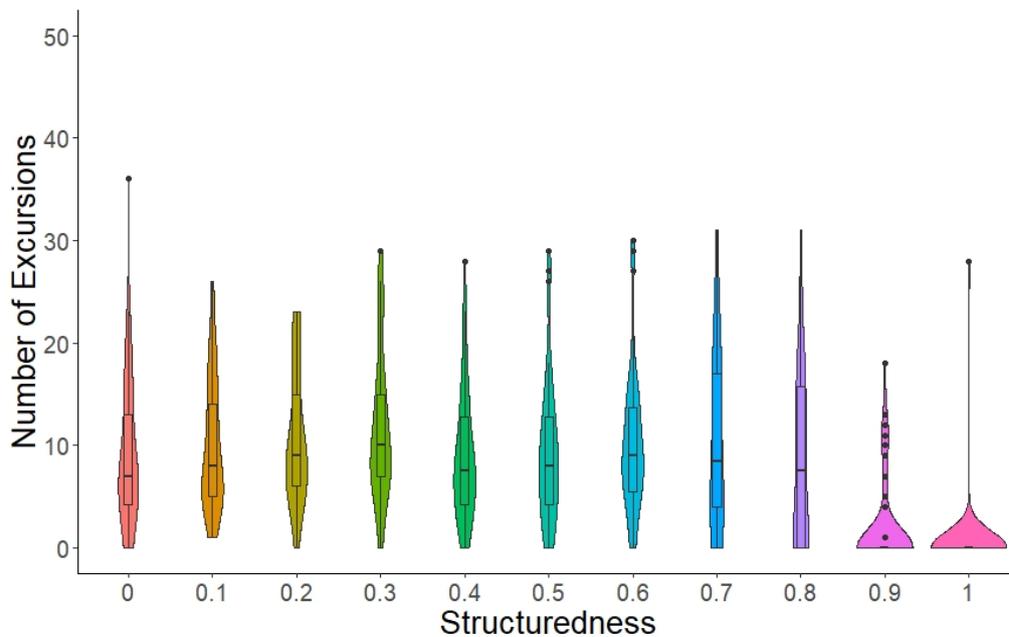


Figure 5.8. Highly structured simulations experience fewer excursions on average, even at a low transmission rate. Combined boxplot and violin plot of the number of excursions per simulation, grouped according to simulation structuredness. All simulations have the low transmission rate of 0.0087 per timestep per neighbour.

Another metric of interest is the average duration of the excursions; that is to say, the amount of time before the number of infected individuals falls below the threshold again. Once again, we examine the maximum transmission rate simulations first (Figure 5.9). We observe minimal change in the average excursion duration as the structuredness increases, however the high structuredness simulations display greater variation than the low structuredness ones. This is supported by the K-S tests (Table 5.4), which show that the most structured simulations are significantly different to the least structured ones.

At the moderate transmission rate (Figure 5.10) this trend persists, although it appears slightly less pronounced. The K-S testing (Table 5.5) once again confirms that the most structured simulations are significantly different to the least structured ones.

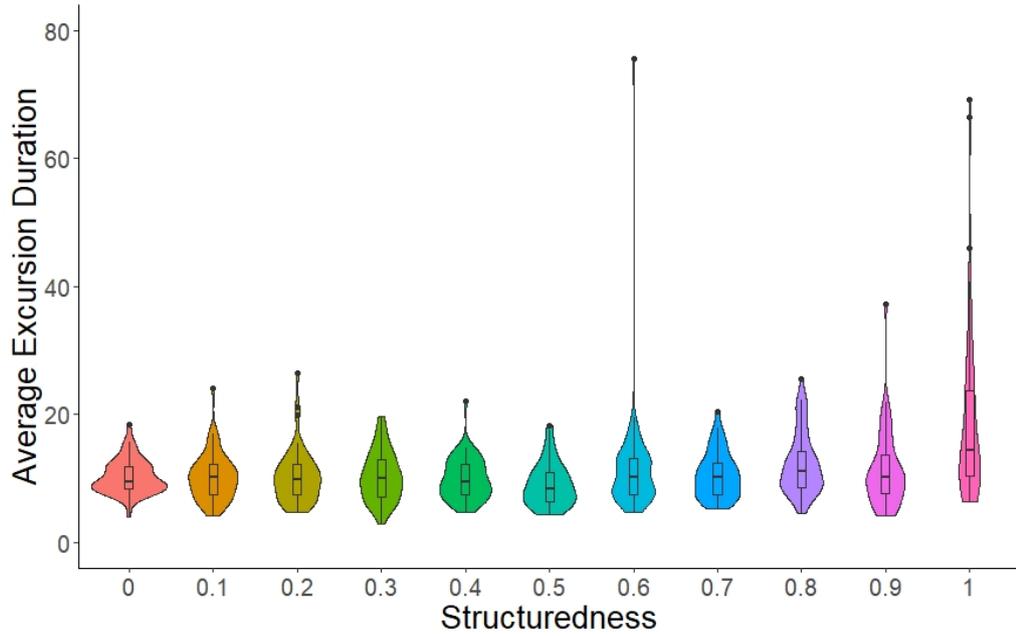


Figure 5.9. Highly structured simulations display greater variation in excursion duration. Combined boxplot and violin plot of the average excursion duration per simulation, grouped according to simulation structuredness. All simulations have the maximum transmission rate of 0.015 per timestep per neighbour.

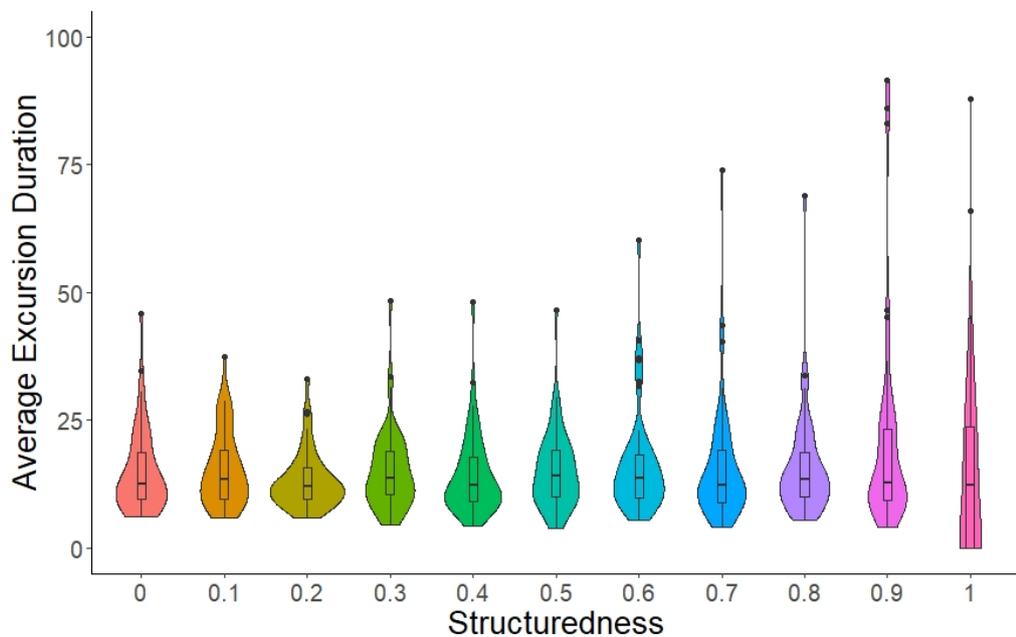


Figure 5.10. Under a reduced transmission rate, highly structured simulations still display greater variation in excursion duration. Combined boxplot and violin plot of the average excursion duration per simulation, grouped according to simulation structuredness. All simulations have the moderate transmission rate of 0.0115 per timestep per neighbour.

Structuredness	P-value
1.0	< 0.001
0.9	< 0.001

Table 5.3. Kolmogorov-Smirnov tests verify that very structured simulations experience more excursions than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with low transmission rate 0.0087. Non-significant tests are excluded from this table.

Structuredness	P-value
1.0	< 0.001

Table 5.4. Kolmogorov-Smirnov tests verify that the most structured simulations experience longer excursions on average than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with max transmission rate 0.015. Non-significant tests are excluded from this table.

Structuredness	P-value
1.0	< 0.001

Table 5.5. Kolmogorov-Smirnov tests verify that the most structured simulations experience longer excursions on average than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with moderate transmission rate of 0.0115. Non-significant tests are excluded from this table.

At the low transmission rate (Figure 5.11, Table 5.6), we observe a greatly reduced average excursion duration for the most structured simulations as well as much lower variation. This is likely partially due to the fact that the endemic equilibrium is only just stable.

Finally, it would be interesting to consider the size of the excursion; specifically, the distance between the maximum number of infections and the endemic equilibrium. Starting once more with the maximum transmission rate (Figure 5.12), we observe that the most structured simulations display a lower average excursion distance than their less structured counterparts. This is once again confirmed by the K-S tests (Table 5.7, which find that the 0.9 and 1.0 structuredness simulations are significantly different to the 0.0 structuredness ones. Similar behaviour is also observed for both the moderate transmission rate (Figure 5.13, Table 5.8) and the low transmission rate (Figure 5.14, Table 5.9).

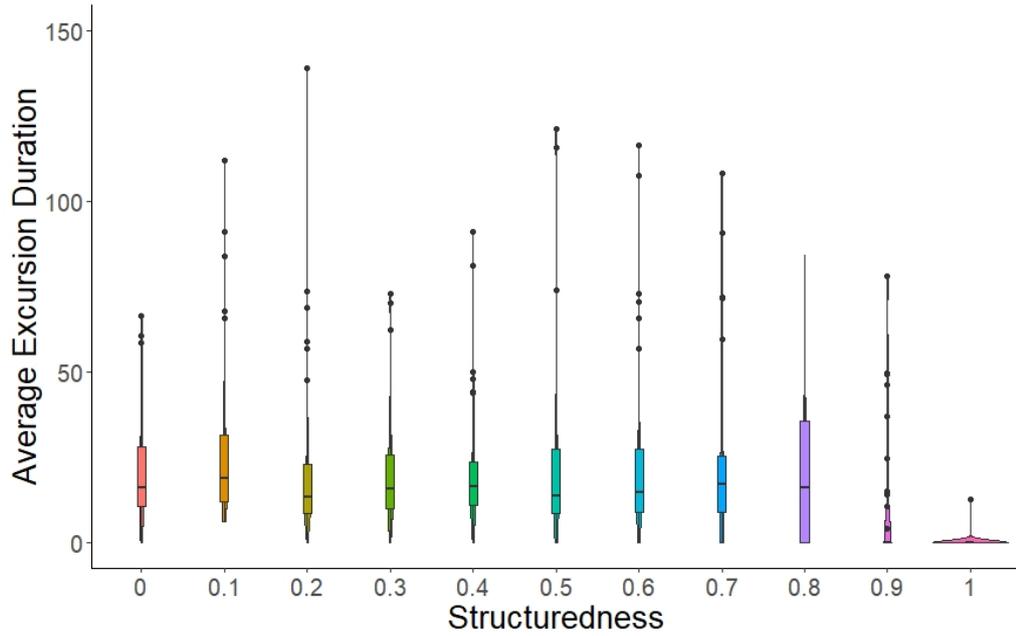


Figure 5.11. At the low transmission rate, highly structured simulations now display much lower variation in excursion duration. Combined boxplot and violin plot of the average excursion duration per simulation, grouped according to simulation structuredness. All simulations have the low transmission rate of 0.0087 per timestep per neighbour.

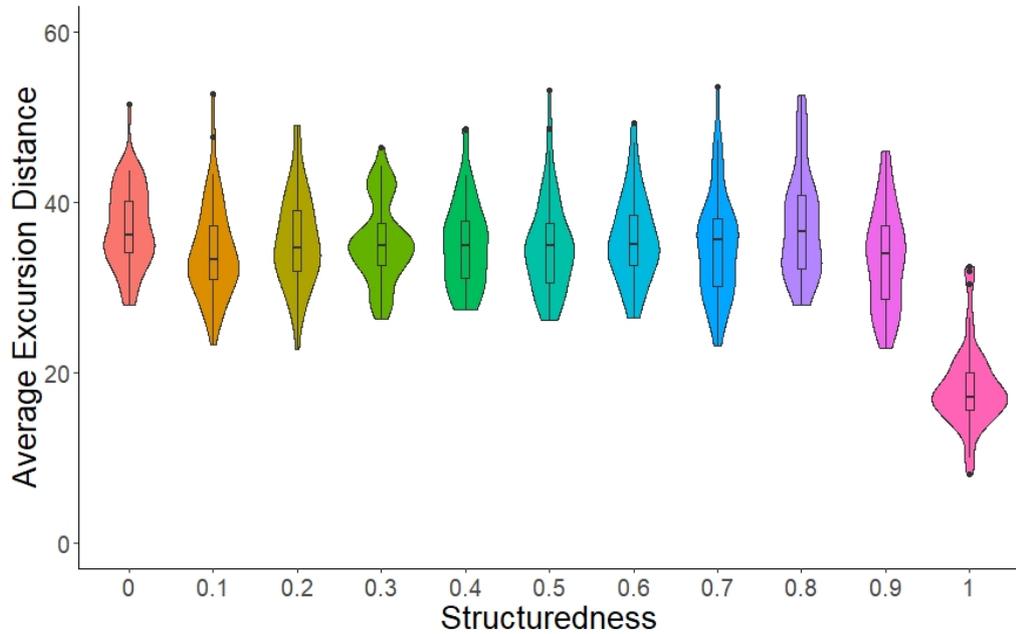


Figure 5.12. Highly structured simulations experience less severe excursions. Combined boxplot and violin plot of the average excursion distance per simulation, grouped according to simulation structuredness. All simulations have the maximum transmission rate of 0.015 per timestep per neighbour.

Structuredness	P-value
1.0	< 0.001
0.9	< 0.001

Table 5.6. Kolmogorov-Smirnov tests verify that the most structured simulations experience longer excursions on average than the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with low transmission rate of 0.0087. Non-significant tests are excluded from this table.

Structuredness	P-value
1.0	< 0.001
0.9	0.02171

Table 5.7. Kolmogorov-Smirnov tests verify that the excursions of the most structured simulations do not reach as far on average as those of the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with max transmission rate of 0.015. Non-significant tests are excluded from this table.

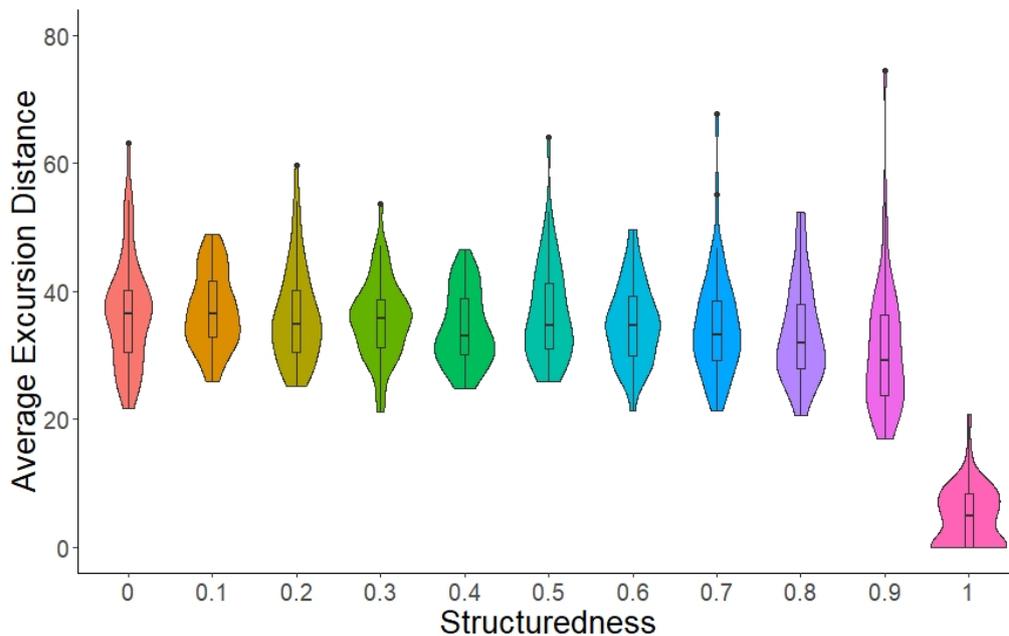


Figure 5.13. Under a reduced transmission rate, highly structured simulations experience less severe excursions. Combined boxplot and violin plot of the average excursion distance per simulation, grouped according to simulation structuredness. All simulations have the moderate transmission rate of 0.0115 per timestep per neighbour.

Structuredness	P-value
1.0	< 0.001
0.9	0.00584

Table 5.8. Kolmogorov-Smirnov tests verify that the excursions of the most structured simulations do not reach as far on average as those of the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with mid transmission rate of 0.0115. Non-significant tests are excluded from this table.

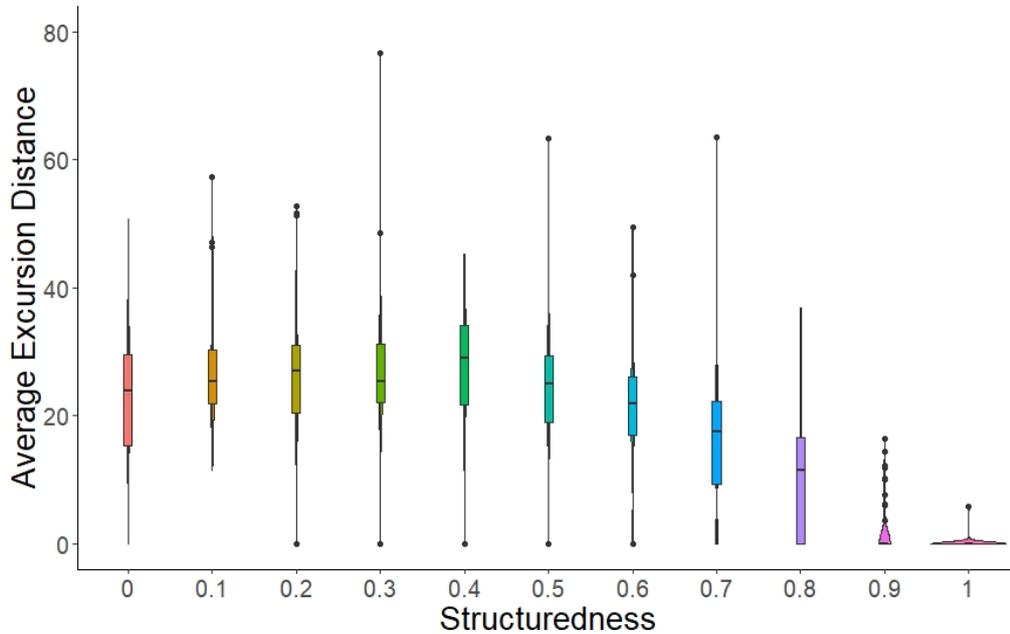


Figure 5.14. Even at the lowest transmission rate, highly structured simulations still experience less severe excursions. Combined boxplot and violin plot of the average excursion distance per simulation, grouped according to simulation structuredness. All simulations have the low transmission rate of 0.0087 per timestep per neighbour.

5.6 Discussion

5.6.1 Insights from the SCPW

We observe that the analytic estimate derived by Corcoran and Hastings [151] (Equation 5.15) is much less accurate for very structured networks than for traditional configuration models (Figure 5.3). We also observe that high structuredness considerably alters the behaviour of the system. The former result is not unexpected; it was anticipated that more structured networks would present issues for the SCPW [152] and thus our community networks may be among those for which the estimate may prove ineffective [151]. The presence of nonrandom elements in networks has long been an obstacle for mean-field models that requires specific efforts to overcome [462]. Nevertheless, it is informative to

Structuredness	P-value
1.0	< 0.001
0.9	< 0.001
0.8	< 0.001
0.7	0.00618

Table 5.9. Kolmogorov-Smirnov tests verify that the excursions of the most structured simulations do not reach as far on average as those of the unstructured simulations. Output from two-sample Kolmogorov-Smirnov tests comparing all structuredness values to the unstructured (structuredness = 0) category, with low transmission rate of 0.0087. Non-significant tests are excluded from this table.

see the magnitude to which this estimate over-predicts in contexts that mimic the real world.

Our results also suggest that highly structured communities would experience fewer (Figures 5.6-5.8) and gentler (Figures 5.12-5.14) outbreaks, but ones of less predictable duration (Figures 5.9-5.11). That we observe noticeable changes only once the simulations are heavily structured echoes previous work [463, 464]. In particular, Wren and Best [464] demonstrated that metrics such as the peak number of infected individuals and total proportion infected decreased on more structured networks, but that significant differences were only observed once the networks were sufficiently structured, particularly when the R_0 is high. Previous studies have also observed that highly ‘modular’ communities (that is, communities consisting of weakly connected groups of densely connected individuals) experience less severe peaks under SIR model dynamics [465], especially in comparison to more random alternatives such as Erdős-Rényi networks [466]. Similar results are also observed in the context of animal social groups, where more modular networks experience smaller outbreaks [467], but ones that are potentially longer [448]. This has been attributed to the difficulty of spreading between modules, which causes a disease to either go extinct quickly or slowly spread through a number of groups [465, 468].

5.6.2 Implications for leishmaniasis modelling.

Many of the communities in which leishmaniasis is endemic could be considered to have high modularity. Sand flies are typically nocturnal or crepuscular [469, 470] (though there are exceptions [471]), so most often feed on sleeping individuals. Their preference for shorter-distance flight [69] restricts their ability to travel between distant houses, thus creating a scenario where transmission can only easily be facilitated between members of the same house, or with very nearby houses. This dynamic is captured by our most structured simulations. Our findings suggest that real world communities are therefore conducive to low-level, long-term transmission. This may partially explain the limited success of the KalaNet trial [302]. Despite the effectiveness of insecticide-treated nets at

reducing visceral leishmaniasis incidence [4], they were insufficient to completely prevent transmission. Modelling efforts have suggested that elimination may require almost complete coverage using the nets [304] combined with other interventions [185]. The effort required is more understandable if this community structure is favourable for low-level transmission.

In contexts where individuals are either connected to many neighbours or have connections that span great distances across the network, our results suggest that outbreaks could be more frequent and severe, however. This would be expected in denser settlements with humid houses of low quality [290], which are more accommodating for sand flies and provide protection from indoor residue spraying [472], and larger household sizes [252]. An abundance of sand fly habitat [128] and other blood sources such as hyraxes [78,309] further amplify the issue by increasing the local sand fly density. Another relevant context would be non-agricultural industry that is dependent on migratory laborers. Recent studies have identified migratory work at brick kilns as a risk factor in leishmaniasis transmission [252]. Kumar *et al.* [12] suggested that this it was most likely a factor in developing the disease (rather than in transmission directly) due to the associated stress and poverty, given the need for damp organic-rich soils for sand fly oviposition [473], but brick kilns do display a surprisingly high sand fly species richness [474]. If laborers were to be housed in poor quality, high-density communal accommodation [130,140], with low bed-net usage due to a lack of affordable nets [475], limited access to education about leishmaniasis or difficulty using the bed-nets [253], then amplify the risk of leishmaniasis transmission (with implications for seeding outbreaks elsewhere [12,476]). Transmission in these urban contexts may be curbed effectively through application of indoor residue spraying around affected households [146], coupled with synthetic sand fly pheromone to increase efficiency [287] and improved access and utilisation of bednets [475].

A more difficult context is when there is repeated or prolonged outdoor contact between individuals of different households in settings where sand flies are present. Evening activity is associated with increased transmission risk [128] as the sand flies are more active [477], and similar transmission opportunities are presented during the day. Sand flies have daytime resting spots [478], particularly near vegetation [479] but also in shaded areas such as latrines [79]. Activity around these resting spots could disturb the sand flies, which may opportunistically seek a blood meal. Acacia trees, a known sand fly resting site [58,76], are associated with increased transmission [127] especially for people resting beneath them during the day [295]. Routine activity by a number of people could facilitate greater inter-household connectivity, and thus faster transmission. Though bednets would be largely ineffective in this context, targeted application of insecticide has previously

proven very effective at reducing sand fly numbers [480], despite concerns about insecticide resistance [481].

An important facet of transmission not captured here is the behaviour of the sand fly vector. Often, models have captured the role of the vector by using bipartite networks [361] or by handling numbers of individuals that bite and die at some fixed rates [482, 483]. Transmission between the sand fly and a mammalian host is very heterogeneous, however. Sand flies may have greatly differing parasite burdens depending on host skin parasite density [45] and feeding success [54], with direct consequences for their own infectiousness [172]. When coupled with altered feeding efficiency and tenacity [84, 484], and varied feeding preferences [65], this gives rise to a complex scenario where individual sand flies can have much greater importance to transmission than their peers. In the wild, sand flies appear rarely infected with *Leishmania* parasites [485, 486], emphasising the importance of rare individuals in transmission. Modelling this accurately was beyond the scope of this work.

In a similar vein, we also use an SIS model in which there is no long-lasting immunity. In reality, the interactions between *Leishmania* parasites and the host immune system are complex [487]. Visceral leishmaniasis sometimes transitions into post-kala-azar dermal leishmaniasis (PKDL) [488] which may be important in transmission [242], and individuals can relapse after treatment [489], especially if also infected with HIV [490]. There are also lingering questions about the importance and dynamics of asymptomatic individuals [32, 146, 398]. Capturing these dynamics would greatly reduce the tractability of the model, and given the magnitude of unknowns involved, is also best left to future studies.

5.7 Conclusion

Although the highly structured networks used in this study limited the efficacy of the analytic estimator derived by Corcoran and Hastings [151], the SCPW model has yielded insights. We highlight that very structured communities, for which the network displays high modularity, would be expected to experience milder but prolonged outbreaks. We use these findings to predict that rural communities with poorer-quality accommodation and a dependence on agricultural work, especially if combined with seasonal migrant labour, would experience more severe leishmaniasis transmission and thus which would benefit from tailored, priority responses and study.

5.8 Materials and Methods

All numerical simulations were performed using MATLAB R2021b (v9.11.0.1769968). All associated data analysis was performed using RStudio v2022.07.1, Build 554 (R version 4.1.0).

Chapter 6

Discussion

6.1 Overview

In this chapter, the key findings and implications of the preceding chapters are summarised. The overall outlook is synthesised in the context of the wider literature, and a range of future work is proposed with reference to recent work.

6.2 Chapter 2

Our numerical simulations revealed that the heterogeneous distribution of *Leishmania* parasites across the skin of a mammalian host [45] was particularly important for transmission in the context of short-lived sand flies, but that the life cycle of the parasites in their sand fly vector [54] was more important if the sand flies could live sufficiently long to bite multiple times. This was because the number of parasites in the sand fly increases dramatically after second and subsequent bites, allowing even lightly infected sand flies to accrue a significant parasite burden.

That subsequent bites are important to such an extent presents an opportunity: by reducing the sand fly lifespan to less than two weeks, it is possible to greatly inhibit the transmission of leishmaniasis. Although the true sand fly lifespan is uncertain [160] studies suggest that wild sand flies could reasonably exceed two weeks in age [163]. A range of methods to reduce the lifespan could be employed, from creating unfavourable conditions [47] in known sand fly hotspots [478] to using natural enemies [162] or insecticide treatment regimes tailored to regions and times of especially high sand fly densities [384, 480] to prevent the sand flies achieving their third or fourth bite. The findings also emphasised the importance of understanding the role of asymptomatic infections [184] or post-kala-azar dermal leishmaniasis (PKDL) cases [119] in maintaining leishmaniasis within a community. Their relevance remains unclear [32, 241] but our findings lent support to the suggestion

that they are a potential reservoir of leishmaniasis.

6.3 Chapter 3

We used survival analysis to identify a few important factors in leishmaniasis transmission. At the community-scale, we found that proximity to previous cases of leishmaniasis increased the risk of infection, especially if a susceptible individual shared a household with an infected individual. At larger scales, communities with poor access to stable employment and education were more vulnerable to leishmaniasis becoming endemic. We also highlighted the limitations of survival analysis, demonstrating how data that appears suitable for survival analysis may not necessarily yield meaningful insights and that if discrete categories of data are to be used, they must be chosen appropriately (with a suitable reference category).

The relationship we identified between case proximity and infection risk echoes prior work by Chapman et al. [146] but was found using a more approachable [188] and less computationally intense [198] method. It has implications for the control of leishmaniasis: it underscores the role of insecticide-treated bed nets (ITNs) [283] but also highlights the importance of extending preventative measures beyond the household (perhaps most relevant to indoor residue spraying (IRS) [278]). Such within-house measures have an important role, and we can improve [287] and more effectively deploy [289] them to further impede leishmaniasis transmission. Our findings at larger scales emphasised the importance of awareness campaigns [291] and medical screening [293, 294], as well as improved construction techniques [130] in preventing leishmaniasis from becoming endemic. A combination of techniques at different scales is thus essential to further our leishmaniasis eradication efforts.

6.4 Chapter 4

Following the work of previous studies [344, 345] we repeated the derivation of an analytic estimate of the basic reproduction number R_0 for heterogeneous networks, and used it to produce context-specific versions for a range of networks and transmission rate regimes. We used numerical simulation to test the accuracy of these analytic predictors, finding that they were largely ill-suited beyond the fully homogeneous context. However, we identified another potential application, finding that they were effective at predicting the approximate location of an ‘epidemic threshold’ for the transmission rate, beyond which an outbreak was likely to persist for several generations of infections.

Given the prominence of R_0 as a measure of disease transmission [491, 492], it is a shame

that these estimates proved unreliable. However, their potential in predicting whether a disease will establish itself in a naive community is still valuable, especially in contexts where local extinction of the disease is a significant risk [493] and thus transmission is maintained by invading new clusters [494, 495]. The establishment of new clusters of infections [91] is important both for maintaining leishmaniasis [233] and for expanding its range into new regions [93]. With the ability to predict vulnerable communities comes the opportunity to target interventions to maximise their impact, for example by prioritising improvements to healthcare access (both for treatment [496] and diagnosis [497]) or sanitation [149] in the most vulnerable areas. Such targeting is important to the on-going elimination efforts [459, 498], especially for regions with limited funding or fragile health systems [149, 496].

6.5 Chapter 5

We tested how the inclusion of network structure impacted the accuracy of a recently derived analytic estimate for the endemic equilibrium [151] of the Super Compact Pairwise (SCPW) model [152]. We demonstrated that as the network becomes more structured the estimate overestimated by a progressively greater amount, but was reasonably accurate for a fully random network. We then explored the dynamics of our SIS model on networks of varying ‘structuredness’, finding that epidemics on more structured networks experience rarer, less severe deviations from their endemic equilibrium, but that such excursions would display greater variation in their duration, than similar epidemics on fully random networks.

Leishmaniasis is endemic in many regions [9, 428, 429] and identifying tools to investigate its endemic state is an important part of the elimination effort. This particular approach may have proven ill-suited to a typical structured community [12, 147] but it could still yield insights for less structured scenarios such as shared sleeping areas [499] or communal latrines [79], especially those with poor construction materials [140] or in proximity to vegetation [126, 130]. Our results suggested that such situations would also permit more severe (though less long-lived) outbreaks, which could have ramifications for communities across a wide area [12] if the affected individuals are migrant workers or refugees [308] that could then disperse. By focusing our monitoring [497] efforts, as well as the deployment of bed nets [475] and infrastructure improvements [290], in and around these settings we have the potential to efficiently prevent multiple clusters of leishmaniasis infections from emerging.

6.6 Synthesis and Future Perspectives

Together, these findings indicate a nuanced relationship between leishmaniasis and heterogeneity at multiple scales. At large scales, factors such as urbanisation (Figure 3.11) and the types of employment available (Figure 3.13), as well as climate [93], create a patchwork of regions with differing susceptibilities [149] to leishmaniasis transmission. This heterogeneous landscape includes coldspots that hinder transmission [93], but also rural hotspots (Figure 3.11) where the mixing of reservoirs [18,500], exposed humans [137] and suitable vegetation [60,479] present the opportunity for effectively maintaining and spreading the disease. Such hotspots may be especially important to transmission if human migration enables longer-distance transmission [121] into other regions. The typically structured communities within which leishmaniasis persists are highly heterogeneous, with sand flies traveling short distances [69] and facilitating equally short-range transmission [146]. Individuals are thus able to infect only those close to them, giving rise to transmission networks akin to those we study in Chapters 4 and 5. Transmission appears to be hampered by such networks, but outbreaks may take longer to run their course (Figure 5.9) and rarer, highly connected individuals are potential superspreaders [501] that could cause many new infections. Even at the micro-scale, heterogeneity has a complex role: the heterogeneous skin parasite distribution [45] allows some flies to take up many parasites but also means that many flies will take up very few parasites and be less infectious (although minimally infected sand flies can still become very infectious if multiple bites are taken [54,172]). Thus, heterogeneity is both a blessing and a curse to leishmaniasis transmission; it presents barriers at multiple scales, but also creates valuable transmission opportunities in certain circumstances.

The findings also challenge some of the intuitive associations between ‘density’ and disease transmissibility. We would intuit that a high density of humans would benefit leishmaniasis transmission, as it does for many airborne human-to-human diseases [502,503]. Yet, Figure 3.11 implies that the densest environments, our urban centres, are less conducive to transmission. This is likely due to reduced sand fly population densities [504,505], which require an abundance of suitable soil and vegetation [74,75], and improved construction techniques that exclude sand flies from our buildings [290]. Thus, using human population density alone as a proxy for transmissibility would produce inaccurate expectations. A similar effect can be observed at the level of communities: a high density of adult sand flies is associated with increased transmission risk [384,484], but the distribution of sand flies throughout a settlement is also crucial. Transmission risk is elevated if sand fly-friendly habitat is located near human houses [506] (this motivates measures such

as indoor residual spraying [5]), thus settlements with similar average sand fly densities may experience different transmission rates depending on where the flies are located. Therefore, considering average sand fly population density alone is insufficient to capture the transmission dynamics, and it is important to also consider the location of the sand fly population hotspots. The role of ‘density’, be it of humans, vectors, or parasites, is scale-dependent and consideration of the specific context is vital to accurately capturing the transmission dynamics.

The overall picture that emerges is one of a disease for which transmission is slow and rare, but which can capitalise on any favourable situation that arises (as might be expected for a disease in a spatially structured host landscape with low turnover [507]). Yet, there remain many questions surrounding leishmaniasis transmission that require answers if we are to fully understand this picture.

6.6.1 Prospects at the Micro-Scale

The interplay of patchy skin parasite distributions and sand fly biting behaviour [172] presents a number of opportunities for further work, both in iterating upon the model and in informing other areas of leishmaniasis research. Our model omitted the role of promastigote secretory gel (PSG) [49] due to lingering uncertainty surrounding its production and interactions in the gut of the sand fly. However, it has a well-documented role in transmission [87, 166, 508] and given that it both influences [84] and is influenced by [509] sand fly feeding behaviour its inclusion in future models would be desirable, assuming sufficient supporting data could be obtained to parameterise it. Similarly, more data would be beneficial for improving our parameterisation of the retroleptomonad life stage. Despite much recent discussion of retroleptomonads [51, 510], questions remain: for example, what specifically causes the metacyclics to differentiate into retroleptomonads [54], and are there other factors that would cause the reproductive rate of the retroleptomonads to vary? Future studies that contribute to answering these questions, or simply provide more data to aid in parameterising the model, would be beneficial. Further iterations of our model may also incorporate the recent finding that patches of parasites in the skin cluster around older patches [314], which may have implications for parasite uptake.

It would also be beneficial to further investigate the life expectancy and biting rate of wild sand flies. Studies often take the typical lifespan to be approximately 14 days [172, 304], with biting occurring once every 5-6 days [511]. However, prior work provides variable estimates for lifespan [512, 513] and indicates that it may be temperature dependent [514]. *Leishmania* infection also appears detrimental to sand fly survival [84] yet conveys protection against certain pathogens that would infect the sand fly [515], so the

true impact on lifespan is unclear. Understanding this impact is important for determining whether reducing the sand fly lifespan is a viable control method. Reducing the lifespan could simply select for faster-replicating parasites, but this outcome is less likely if parasite replication imposes significant costs on the sand fly (or is otherwise constrained by some unknown factor). Blood meals are usually thought to occur once per oviposition cycle, but some species may feed more than once per cycle [516] and cycle length may depend on a range of factors including blood source [517] and temperature [514,518]. Future work that clarifies some of this uncertainty, even if only for a single sand fly/parasite species combination, would be valuable to understand the true impact of the retroleptomonads on transmission.

The importance of the retroleptomonad state is of relevance to other work, too. Studies have long partitioned humans into different groups based on infection stage [519] with varying outward infectiousness [170], using the resultant models to inform diagnosis [520] and elimination strategies [431,498]. Yet, even recent models incorporating PKDL treatment [304] and vaccination [521] assume all infected sand flies to be equally infectious towards humans. Our model suggests that future work would benefit from distinguishing between initially infected sand flies and those which have taken two or more bites (which are potentially much more infectious) to avoid underestimating their transmission potential. Further potential lies with the possible role of ‘superspreading’. It is often assumed that sand flies bite (and thus can infect) a single person when taking a blood meal. However, Rogers and Bates [84] previously found that infected sand flies were much more likely to bite another individual when disturbed while feeding than their uninfected peers. This raises the possibility for superspreading [501] events to occur, in which multiple individuals are bitten and potentially infected by a single, highly infectious fly during a blood meal (especially if the enlarged PSG plug [54] further lengthens the feeding process and makes disturbance more likely). As this would allow some flies to cause multiple infections without living an exceptionally long life, it would be enlightening for future work to investigate the potential impact of such events on leishmaniasis transmission.

6.6.2 Future Inquiries at Larger Scales

The R_0 estimate and associated epidemic threshold tested in Chapter 4 present two main avenues of prospective work. It is tempting to focus on correcting [522] the R_0 estimate given its prominence in the public eye, but since it can be misleading for structured populations [523] and is easy to mis-use [524], future studies are likely better served by other measures. Velázquez-Castro *et al.* [525] define the number of secondary cases caused by hosts and vectors to be ‘risk indices’ for each patch. Although these are components

of the R_0 for a vector-borne disease [362], and thus should be handled with appropriate caution, they could incorporate information about factors such as sand fly breeding/resting sites [73, 79], ITN usage [4] or sand fly densities [526, 527] for different neighbourhoods to produce an estimate of the vulnerability to leishmaniasis invasion. This would enable the tailoring of measures such as insecticide-spraying [5] to reduce the spread of leishmaniasis into new areas, and could serve as a model for use across the endemic regions.

The endemic equilibrium we tested in Chapter 5 proved unsuitable for use on structured networks, but it may be possible to introduce some form of correction term to account for these types of networks. Studies of disease transmission in animal groups, which can be highly structured [528], make use of a measure called Newman modularity [449, 528]. Perhaps it would be possible for future work to calculate a correction term that uses the Newman modularity of a network to adjust the endemic equilibrium, as one might correct an R_0 to account for bias [529] or factors such as population density [530], and thus improve its accuracy on these networks. Studies examining contexts in which population turnover is largely irrelevant may also want to examine alternative models. The SIS model for which the endemic equilibrium applies to contexts such as a migrant workforce [12], where infection may mean an individual leaves for treatment [253] and thus long-term immunity is not necessarily relevant. For other populations, an SIRS model would be more applicable (immunity is not thought to be permanent [521]). The SCPW forms of the SIS and SIR models have previously been constructed [152, 531] and perhaps the SIR model could be extended to account for temporary immunity, thus producing an SIRS model. If the resulting model is sufficiently tractable, an endemic equilibrium could be extracted from it as Corcoran and Hastings did for the SIS model [151].

6.6.3 Closing Remarks

Ultimately, advancements in these areas will provide a more complete understanding of leishmaniasis transmission. This will be an essential part of optimising our eradication efforts, which have already seen success: although not fully eradicated [307], the incidence rate has been greatly reduced [303] due to developments in treatment and preventative measures [10, 432]. With further research, we will be able to better target preventative measures [278, 283] and perhaps, eventually, vaccine deployment [532] to limit the spread of leishmaniasis into new areas [533] and finally reach the elusive eradication target. Only through increasing our knowledge and redoubling our efforts to eliminate leishmaniasis can we hope to break its grip on the communities it affects and render the *Leishmania* scourge a more manageable threat, or even a thing of the past.

Bibliography

- [1] Mathers, C.D., M. Ezzati, and A.D. Lopez. Measuring the burden of neglected tropical diseases: the global burden of disease framework. *PLoS Neg. Trop. Dis.* 2007; 1(2): e114.
- [2] Hotez, P.J., M. Alvarado, M.-G. Basáñez, I. Bolliger, R. Bourne, M. Boussinesq, S.J. Brooker, A.S. Brown, G Buckle, C.M. Budke, H. Carabin, L.E. Coffeng, E.M. Fèvre, T. Furst, Y.A. Halasa, R. Jasrasaria, N.E. Johns, J. Keiser, C.H. King, R. Lozano, M.E. Murdoch, S. O’Hanlon, S.D.S. Pion, R.L. Pullan, K.D. Ramaiah, T. Roberts, D.S. Shepard, J.L. Smith, W.A. Stolk, E.A. Undurraga, J. Utzinger, M. Wang, C.J.L. Murray, and M. Naghavi. The global burden of disease study 2010: Interpretation and implications for the neglected tropical diseases. *PLoS Neg. Trop. Dis.* 2014; 8(7): e2865.
- [3] World Health Organisation Global leishmaniasis update, 2006-2015: a turning point in leishmaniasis surveillance. *Weekly Epidemiological Record* 2017; 38(92): pp.557-572.
- [4] Chowdhury, R., V. Chowdhury, S. Faria, S. Akter, A.P. Dash, S.K. Bhattacharya, N.P. Maheswary, C. Bern, S. Akhter, J. Alvar, A. Kroeger, M. Boelaert, and Q. Banu. Effect of insecticide-treated bed nets on visceral leishmaniasis incidence in Bangladesh. A retrospective cohort analysis. *PLoS Neg. Trop. Dis.* 2019; 13(9): e0007724.
- [5] Faber, C., C.M. Quiñonez, O. Horstick, K.M. Rahman, and S. Runge-Ranzinger. Indoor residual spraying for the control of visceral leishmaniasis: A systematic review. *PLoS Neg. Trop. Dis.* 2022; 16(5): e0010391.
- [6] Gillespie, P.M., C.M. Beaumier, U. Strych, T. Hayward, P.J. Hotez, and M.E. Bottazzi. Status of vaccine research and development of vaccines for leishmaniasis. *Vaccine* 2016; 34(26): pp. 2992-2995.
- [7] Anversa, L., M.G.S. Tiburcio, V.B. Richini-Pereira, and L.E. Ramirez. Human leishmaniasis in Brazil: A general review. *Rev. Assoc. Med. Bras.* 2018; 64(3): pp. 281-289.
- [8] Haftom, M., P. Petrucka, K. Gemechu, J. Nesro, E. Amare, T. Hailu, Y. Ashebir, G. Gebreheat, H. Hagos, D. Gebremedhin, and A. Gebremariam. Prevalence and risk factors of human leishmaniasis in Ethiopia: A systematic review and meta-analysis. *Infect. Dis. Ther.* 2021; 10: pp. 47-60.

- [9] Rostamian, M., H. Bashiri, V. Yousefinejad, A. Bozorgomid, N. Sohrabi, S. Raeghi, M.T. Khodayari, K. Ghadiri, and S. Rezaeian. Prevalence of human visceral leishmaniasis in Iran: A systematic review and meta-analysis. *Comparative Immunology, Microbiology and Infectious Diseases* 2021; 75:101604.
- [10] Hirve, S., A. Kroeger, G. Matlashewski, D. Mondal, M.R. Banjara, P. Das, A. Benazir, B. Arana, and P. Olliaro. Towards elimination of visceral leishmaniasis in the Indian subcontinent: Translating research to practice to public health. *PLoS Neg. Trop. Dis.* 2017; 11(10): e0005889.
- [11] Argaw, D., A. Mulugeta, M. Herrero, N. Nombela, T. Teklu, T. Tefera, Z. Belew, J. Alvar, and C. Bern. Risk factors for visceral leishmaniasis among residents and migrants in Kafta-Humera, Ethiopia. *PLoS Neg. Trop. Dis.* 2013; 7(11): e2543.
- [12] Kumar, A., S. Saurabh, S. Jamil, and V. Kumar. Intensely clustered outbreak of visceral leishmaniasis (kala-azar) in a setting of seasonal migration in a village of Bihar, India. *BMC Infectious Diseases* 2020; 20:10.
- [13] Rijal, S., S. Koirala, P. van der Stuyft, and M. Boelaert. The economic burden of visceral leishmaniasis for households in Nepal. *Trans. R. Soc. Trop. Med. Hyg.* 2006; 100(9): pp. 838-841.
- [14] Okwor, I., and J. Uzonna. Social and economic burden of human leishmaniasis. *Am. J. Trop. Med. Hyg.* 2016; 94(3): pp. 489-493.
- [15] Saroufim, M., K. Charafeddine, G. Issa, H. Khalifeh, R.H. Habib, A. Berry, N. Ghosn, A. Rady, and I. Khalifeh. Ongoing epidemic of cutaneous leishmaniasis among Syrian refugees, Lebanon. *Emerg. Infect. Dis.* 2014; 20(10): pp. 1712-1715.
- [16] Al-Salem, W.S., D.M. Pigott, K. Subramaniam, L.R. Haines, L. Kelly-Hope, D.H. Molyneux, S.I. Hay, and A. Acosta-Serrano. Cutaneous leishmaniasis and conflict in Syria. *Emerg. Infect. Dis.* 2016; 22(5): pp. 931-933.
- [17] Carvalho, B.M., E.F. Rangel, P.D. Ready, and M.M. Vale. Ecological niche modelling predicts southward expansion of *Lutzomyia (Nyssomyia) flaviscutellata* (Diptera: Psychodidae: Phlebotominae), vector of *Leishmania (Leishmania) amazonensis* in South America, under climate change. *PLoS ONE* 2015; 10(11): e0143282.
- [18] Roque, A.L.R., and A.M. Jansen. Wild and synanthropic reservoirs of *Leishmania* species in the Americas. *Int. J. Parasitol.: Parasites and Wildlife* 2014; 3: pp. 251-262.
- [19] Travi, B.L., A. Cordeira-da-Silva, F. Dantas-Torres, and G. Miró. Canine visceral leishmaniasis: diagnosis and management of the reservoir living among us. *PLoS Neg. Trop. Dis.* 2018; 12(1): e0006082.
- [20] Magri, A., C. Bianchi, L. Chmelovà, M. Caffara, R. Galuppi, M. Fioravanti, V. Yurchenko, and A.Y. Kostygov. Roe deer (*Capreolus capreolus*) are a novel potential reservoir for human visceral leishmaniasis in the Emilia-Romagna region of north-eastern Italy. *Int. J. Parasitol.* 2022; 52: pp. 745-750.

- [21] Akhoundi, M., K. Kuhls, A. Cannet, J. Votýpka, P. Marty, P. Delaunay, and D. Sereno. A historical overview of the classification, evolution and dispersion of *Leishmania* parasites and sandflies. *PLoS Neg. Trop. Dis.* 2016; 10(3): e0004349.
- [22] Becvar, T., B. Vojtkova, P. Siriyasatien, J. Votýpka, D. Modry, P. Jahn, P. Bates, S. Carpenter, P. Volf, and J. Sadlova. Experimental transmission of *Leishmania (Mundinia)* parasites by biting midges (Diptera: Ceratopogonidae). *PLoS Pathog.* 2021; 17(6): e1009654.
- [23] Volpedo, G., T. Pacheco-Fernandez, E.A. Holcomb, N. Cipriano, B. Cox, and A.R. Satoskar. Mechanisms of immunopathogenesis in cutaneous leishmaniasis and post-kala-azar dermal leishmaniasis (PKDL). *Front. Cell. Infect. Microbiol.* 2021; 11: 685296.
- [24] Chakravarty, J., E. Hasker, S. Kansal, O.P. Singh, P. Malaviya, A.K. Singh, A. Chourasia, T. Singh, M. Sudarshan, A.P. Singh, B. Singh, R.P. Singh, B. Ostyn, M. Fakiola, A. Picado, J. Menten, J.M. Blackwell, M.E. Wilson, D. Sacks, M. Boelaert, and S. Sundar. Determinants for progression from asymptomatic infection to symptomatic visceral leishmaniasis: A cohort study. *PLoS Neg. Trop. Dis.* 2019; 13(3): e0007216.
- [25] Chapman, L.A.C., L. Dyson, O. Courtenay, R. Chowdhury, C. Bern, G.F. Medley, and T.D. Hollingsworth. Quantification of the natural history of visceral leishmaniasis and consequences for control. *Parasites and Vectors* 2015; 8: 521.
- [26] Martins-Melo, F.R., M. da S. Lima, A.N. Ramos Jr., C.H. Alencar, and J. Heukelbach. Mortality and case fatality due to visceral leishmaniasis in Brazil: A nationwide analysis of epidemiology, trends and spatial patterns. *PLoS ONE* 2014; 9(4): e93770.
- [27] Druzian, A.F., A.S. de Souza, D.N. de Campos, J. Croda, M.G. Higa Jr., M.E.C. Dorval, M.A. Pompilio, P.A. de Oliveira, and A.M.M. Paniago. Risk factors for death from visceral leishmaniasis in an urban area of Brazil. *PLoS Neg. Trop. Dis.* 2015; 9(8): e0003982.
- [28] Scorza, B.M., E.M. Carvalho, and M.E. Wilson. Cutaneous manifestations of human and murine leishmaniasis. *Int. J. Mol. Sci.* 2017; 18(6): 1296.
- [29] Reithinger, R., J.-C. Dujardin, H. Louzir, C. Pirmez, B. Alexander, and S. Brooker. Cutaneous leishmaniasis. *The Lancet Infectious Diseases* 2007; 7(9): pp. 581-596.
- [30] Zijlstra, E.E., F. Alves, S. Rijal, B. Arana, and J. Alvar. Post-kala-azar dermal leishmaniasis in the Indian subcontinent: A threat to the south-east Asia kala-azar elimination programme. *PLoS Neg. Trop. Dis.* 2017; 11(11): e0005877.
- [31] Zijlstra, E.E., A.M. Musa, E.A.G. Khalil, I.M. El Hassan, and A.M. El Hassan. Post-kala-azar dermal leishmaniasis. *The Lancet Infectious Diseases* 2003; 3(2): pp. 87-98.
- [32] Das, V.N.R., R.N. Pandey, N.A. Siddiqui, L.A.C. Chapman, V. Kumar, K. Pandey, G. Matlashewski, and P. Das. Longitudinal study of transmission in households with

visceral leishmaniasis, asymptomatic infections and PKDL in highly endemic villages in Bihar, India. *PLoS Neg. Trop. Dis.* 2016; 10(12): e0005196.

- [33] Gedda, M.R., B. Singh, D. Kumar, A.K. Singh, P. Madhukar, S. Upadhyay, O.P. Singh, and S. Sundar. Post-kala-azar dermal leishmaniasis: A threat to elimination program. *PLoS Neg. Trop. Dis.* 2020; 14(7): e0008221.
- [34] Bern, C., J.H. Maguire, and J. Alvar. Complexities of assessing the disease burden attributable to leishmaniasis. *PLoS Neg. Trop. Dis.* 2008; 2(10): e313.
- [35] McIlwee, B.E., S.E. Weis, and G.A. Hosler. Incidence of endemic human cutaneous leishmaniasis in the United States. *JAMA Dermatol.* 2018; 154(9): pp. 1032-1039.
- [36] Thakur, L., K.K. Singh, V. Shanker, A. Negi, A. Jain, G. Matlashewski, and M. Jain. Atypical leishmaniasis: A global perspective with emphasis on the Indian subcontinent. *PLoS Neg. Trop. Dis.* 2018; 12(9): e0006659.
- [37] Beattie, L., and P.M. Kaye. *Leishmania*-host interactions: what has imaging taught us? *Cellular Microbiology* 2011; 13(11): pp. 1659-1667.
- [38] Naderer, T., and M.J. McConville. The *Leishmania*-macrophage interaction: A metabolic perspective. *Cellular Microbiology* 2007; 10(2): pp. 301-308.
- [39] Laskay, T., G. van Zandbergen, and W. Solbach. Neutrophil granulocytes - Trojan horses for *Leishmania major* and other intracellular microbes? *Trends in Microbiology* 2003; 11(5): pp. 210-214.
- [40] McMahon-Pratt, D., and J. Alexander. Does the *Leishmania major* paradigm of pathogenesis and protection hold for New World cutaneous leishmaniasis or the visceral disease? *Immunological Reviews* 2004; 201(1): pp. 206-224.
- [41] Kima, P.E. The amastigote forms of *Leishmania* are experts at exploiting host cell processes to establish infection and persist. *Int. J. Parasitol.* 2007; 37: pp. 1087-1096.
- [42] Romero, I., J. Téllez, Y. Suárez, M. Cardona, R. Figueroa, A. Zelazny, and N.G. Saravia. Viability and burden of *Leishmania* in extralesional sites during human dermal leishmaniasis. *PLoS Neg. Trop. Dis.* 2010; 4(9): e819.
- [43] Konno, H., N. Yokoyama, Y. Tamura, K. Aoshima, R. Nakao, M. Takiguchi, and K. Katakura. An experimental challenge model for *Leishmania donovani* in beagle dogs, showing a similar pattern of parasite burden in the peripheral blood and liver. *Parasitology Research* 2022; 121: pp. 3569-3579.
- [44] Courtenay, O., C. Carson, L. Calvo-Bado, L.M. Garcez, and R.J. Quinnell. Heterogeneities in *Leishmania infantum* infection: Using skin parasite burdens to identify highly infectious dogs. *PLoS Neg. Trop. Dis.* 2014; 8(1): e2583.
- [45] Doehl, J.S.P., Z. Bright, S. Dey, H. Davies, J. Magson, N. Brown, A. Romano, J.E. Dalton, A.I. Pinto, J.W. Pitchford and P.M. Kaye. Skin parasite landscape determines host infectiousness in visceral leishmaniasis. *Nat. Commun.* 2017; 8(1): 57.

- [46] Schlein, Y., and R.L. Jacobson. High nocturnal CO₂ emanation guides the sand fly *Phlebotomus papatasi* to sugar-rich plants. *Physiological Entomology* 2008; 33(4): pp. 353-359.
- [47] Volf, P., and V. Volfova. Establishment and maintenance of sand fly colonies. *J. Vector Ecol.* 2011; 36(s1): pp. S1-S9.
- [48] Kamhawi, S. Plebotomine sand flies and *Leishmania* parasites: friends or foes? *Trends Parasitol.* 2006; 22(9): pp. 439-445.
- [49] Rogers, M.E., M.L. Chance, and P.A. Bates. The role of promastigote secretory gel in the origin and transmission of the infective stage of *Leishmania mexicana* by the sandfly *Lutzomyia longipalpis*. *Parasitology* 2002; 124(5): pp. 495-507.
- [50] Wilson, R., M.D. Bates, A. Dostalova, L. Jecna, R.J. Dillon, P. Volf, and P.A. Bates. Stage-specific adhesion of *Leishmania* promastigotes to sand fly midguts assessed using an improved comparative binding assay. *PLoS Neg. Trop. Dis.* 2010; 4(9): e816.
- [51] Cecílio, P., A. Cordeiro-da-Silva, and F. Oliveira. Sand flies: Basic information on the vectors of leishmaniasis and their interactions with *Leishmania* parasites. *Commun. Biol.* 2022; 5: 305.
- [52] Volf, P., M. Hajmova, J. Sadlova, and J. Votypka. Blocked stomodeal valve of the insect vector: Similar mechanics of transmission in two trypanosomatid models. *Int. J. Parasitol.* 2004; 34(11): pp. 1221-1227.
- [53] Sunter, J., and K. Gull. Shape, form, function and *Leishmania* pathogenicity: From textbook descriptions to biological understanding. *Open Biol.* 2017; 7(9): 170165.
- [54] Serafim, T.D., I.V. Coutinho-Abreu, F. Oliveira, C. Meneses, S. Kamhawi, and J.G. Valenzuela. Sequential blood meals promote *Leishmania* replication and reverse metacyclogenesis augmenting vector infectivity. *Nat. Microbiol.* 2018; 3: pp. 548-555.
- [55] Shimabukuro, P.H.F., A.J. de Andrade, and E.A.B. Galati. Checklist of American sand flies (*Diptera, Psychodidae, Phlebotominae*): genera, species and their distribution. *Zookeys* 2017; 660: pp. 67-106.
- [56] Alencar, R.B., R.G. de Queiroz, and T.V. Barrett. Breeding sites of phlebotomine sand flies (Diptera: Psychodidae) and efficiency of extraction techniques for immature stages in terra-firme forest in Amazonas State, Brazil. *Acta Tropica* 2011; 118(3): pp. 204-208.
- [57] Barhoumi, W., W.A. Qualls, R.S. Archer, D.O. Fuller, I. Chelbi, S. Cherni, M. Derbali, K.L. Arheart, E. Zhioua, and J.C. Beier. Irrigation in the arid regions of Tunisia impacts the abundance and apparent density of sand fly vectors of *Leishmania infantum*. *Acta Tropica* 2015; 141A: pp. 73-78.
- [58] Moncaz, A., R. Faiman, O. Kirstein, and A. Warburg. Breeding sites of *Phlebotomus sergenti*, the sand fly vector of cutaneous leishmaniasis in the Judean Desert. *PLoS Neg. Trop. Dis.* 2012; 6(7): e1725.

- [59] Oerther, S., H. Jost, A. Heitmann, R. Lühken, A. Krüger, I. Steinhausen, C. Brinker, S. Lorentz, M. Marx, J. Schmidt-Chanasit, T. Naucke, and N. Becker. Phlebotomine sand flies in Southwest Germany: an update with records in new locations. *Parasites and Vectors* 2020; 13: 173.
- [60] Lima, L.H.G. de M., M.R. Mesquita, L. Skrip, M.T. de S. Freitas, V.C. Silva, O.D. Kirstein, I. Abassi, A. Warburg, V. de Q. Balbino, and C.H.N. Costa. DNA barcode for the identification of the sand fly *Lutzomyia longipalpis* plant feeding preferences in a tropical urban environment. *Scientific Reports* 2016; 6: 29742.
- [61] Filho, A.A.P., M. da C.A. Bandeira, R.S. Fonteles, J.L.P. Moraes, C.R.G. Lopes, M.N. Melo, and J.M.M. Rebêlo. An ecological study of sand flies (Diptera: Psychodidae) in the vicinity of Lençóis Maranhenses National Park, Maranhão, Brazil. *Parasites and Vectors* 2015; 8: 442.
- [62] Elaagip, A., A. Ahmed, M.D. Wilson, D.A. Boakye, and M.M.A. Hamid. Studies of host preferences of wild-caught *Phlebotomus orientalis* and *Ph. papatasi* vectors of leishmaniasis in Sudan. *PLoS ONE* 2020; 15(7): e0236253.
- [63] Gebresilassie, A., I. Abbasi, E. Aklilu, S. Yared, O.D. Kirstein, A. Moncaz, H. Tekie, M. Balkew, A. Warburg, A. Hailu, and T. Gebre-Michael. Host-feeding preference of *Phlebotomus orientalis* (Diptera: Psychodidae) in an endemic focus of visceral leishmaniasis in northern Ethiopia. *Parasites and Vectors* 2015; 8: 270.
- [64] Yared, S., A. Gebresilassie, I. Abbasi, E. Aklilu, O.D. Kirstein, M. Balkew, A.S. Brown, R.M. Clouse, A. Warburg, A. Hailu, and T. Gebre-Michael. A molecular analysis of sand fly blood meals in a visceral leishmaniasis endemic region of north-western Ethiopia reveals a complex host-vector system. *Heliyon* 2019; 5(7): e02132.
- [65] Abbate, J.M., C. Maia, A. Pereira, F. Arfuso, G. Gaglio, M. Rizzo, G. Caracappa, G. Marino, M. Pollmeier, S. Giannetto, and E. Brianti. Identification of trypanosomatids and blood feeding preferences of phlebotomine sand fly species common in Sicily, Southern Italy. *PLoS ONE* 2020; 15(3): e0229536.
- [66] Bongiorno, G., A. Habluetzel, C. Khoury, and M. Maroli. Host preferences of phlebotomine sand flies at a hypoendemic focus of canine leishmaniasis in central Italy. *Acta Tropica* 2003; 88(2): pp. 109-116.
- [67] Casanova, C., F.E. Colla-Jacques, J.G.C. Hamilton, R.P. Brazil, and J.J. Shaw. Distribution of *Lutzomyia longipalpis* chemotype populations in São Paulo State, Brazil. *PLoS Neg. Trop. Dis.* 2015; 9(3): e0003620.
- [68] Orshan, L., S. Elbaz, Y. Ben-Ari, F. Akad, O. Afik, I. Ben-Avi, D. Dias, D. Ish-Shalom, L. Studentsky, and I. Zonstein. Distribution and dispersal of *Phlebotomus papatasi* (Diptera: Psychodidae) in a zoonotic cutaneous leishmaniasis focus, the Northern Negev, Israel. *PLoS Neg. Trop. Dis.* 2016; 10(7): e0004819.
- [69] Galvis-Ovallos, F., C. Casanova, D.P. Bergamaschi, and E.A.B. Galati. A field study of the survival and dispersal pattern of *Lutzomyia longipalpis* in an endemic area of visceral leishmaniasis in Brazil. *PLoS Neg. Trop. Dis.* 2018; 12(4): e0006333.

- [70] Faiman, R., O. Kirstein, A. Moncaz, H. Guetta, A. Warburg. Studies on the flight patterns of foraging sand flies. *Acta Tropica* 2011; 120:1-2, pp. 110-114.
- [71] Tonelli, G.B., C. Binder, C. Margonari, and J.D.A. Filho. Sand fly behaviour: Much more than weak flying. *Mem. Inst. Oswaldo Cruz* 2021; 116: e210230.
- [72] Pérez-Cutillas, P., C. Muñoz, J. M.-D. Puente, J. Figuerola, R. Navarro, M. Ortuño, L.J. Bernal, J. Ortiz, R.C. Soriguer, and E. Berriatua. A spatial ecology study in a high-diversity host community to understand blood-feeding behaviour in *Phlebotomus* sandfly vectors of *Leishmania*. *Medical and Veterinary Entomology* 2020; 34: pp. 164-174.
- [73] Vivero, R.J., C. Torres-Gutierrez, E.E. Bejarano, H.C. Peña, L.G. Estrada, F. Florez, E. Ortega, Y. Aparicio, and C.E. Muskus. Study on natural breeding sites of sand flies (Diptera: Phlebotominae) in areas of *Leishmania* transmission in Colombia. *Parasites and Vectors* 2015; 8: 116.
- [74] Khan, K., N.H. Khan, F. Anwar, I. Ullah, N. Badshah, I Irfan, K. Iqbal, I Shah, S.T. Aziz, M.S.N. Khan, and S.U. Shah. Characterisation of sand fly breeding sites in district Malakand, Khyber Pakhtunkhwa, Pakistan, and evaluation of risk factors for cutaneous leishmaniasis in the region. *Zoonoses and Public Health* 2021; 69(1): pp. 33-45.
- [75] Poché, R.M., R. Garlapati, D.-E.A. Elnaiem, D. Perry, and D. Poché. The role of Palmyra palm trees (*Borassus flabellifer*) and sand fly distribution in northeastern India. *J. Vector Ecol.* 2012; 37(1): pp. 148-153.
- [76] Elnaiem, D.-E. A. Ecology and control of the sand fly vectors of *Leishmania donovani* in East Africa, with special emphasis on *Phlebotomus orientalis*. *J. Vector Ecol.* 2011; 36(s1): pp. S23-S31.
- [77] Pareyn, M., E. van den Bosch, N. Girma, N. van Houtte, S. van Dongen, G. van der Auwera, F. Massebo, S. Shibrú, and H. Leirs. Ecology and seasonality of sandflies and potential reservoirs of cutaneous leishmaniasis in Ochollo, a hotspot in southern Ethiopia. *PLoS Neg. Trop. Dis.* 2019; 13(8):e0007667.
- [78] Salah, I., I. Abbasi, A. Warburg, N. Davidovitch, and B. Kotler. Ecology of leishmaniasis in an urbanised landscape: Relationship of sand fly densities, and *Leishmania tropica* infection rates with reservoir host colonies. *Acta Tropica* 2020; 204: 105332.
- [79] Wijerathna, T., and N. Gunathilaka. Diurnal adult resting sites and breeding habitats of phlebotomine sand flies in cutaneous leishmaniasis endemic areas Kurunegala District, Sri Lanka. *Parasites and Vectors* 2020; 13:284.
- [80] Chelbi, I., K. Maghraoui, S. Zhioua, S. Cherni, I. Labidi, A. Satoskar, J.G.C. Hamilton, and E. Zhioua. Enhanced attraction of sand fly vectors of *Leishmania infantum* to dogs infected with zoonotic visceral leishmaniasis. *PLoS Neg. Trop. Dis.* 2021; 15(7): e0009647.

- [81] Staniek, M.E., and J.G.C. Hamilton. Odour of domestic dogs infected with *Leishmania infantum* is attractive to female but male sand flies: Evidence for parasite manipulation. PLoS Pathog. 2021; 17(3): e1009354.
- [82] Nevatte, T.M., R.D. Ward, L. Sedda, and J.G.C. Hamilton. After infection with *Leishmania infantum*, Golden Hamsters (*Mesocricetus auratus*) become more attractive to female sand flies (*Lutzomyia longipalpis*). Scientific Reports 2017; 7: 6104.
- [83] Prestes-Carneiro, L.E., L.A.F. Daniel, L.C. Almeida, L.Z. D’Andrea, A.G. Vieira, I.R. Anjolete, L. André, and E.F. Flores. Spatiotemporal analysis and environmental risk factors of visceral leishmaniasis in an urban setting in São Paulo State, Brazil. Parasites and Vectors 2019; 12: 251.
- [84] Rogers, M.E., and P.A. Bates. *Leishmania* manipulation of sand fly feeding behaviour results in enhanced transmission. PLoS Pathog. 2007; 3(6): e91.
- [85] Giraud, E., O. Martin, L. Yakob, and M. Rogers. Quantifying *Leishmania* metacyclic promastigotes from individual sandfly bites reveals the efficacy of vector transmission. Commun. Biol. 2019; 2: 84.
- [86] Rogers, M., P. Kropf, B.-S. Choi, R. Dillon, M. Podinovskaia, P. Bates, and I. Muller. Proteophosphoglycans regurgitated by *Leishmania*-infected sand flies target the L-Arginine metabolism of host macrophages to promote parasite survival. PLoS Pathog. 2009; 5(8): e1000555.
- [87] Giraud, E., M. Svobodová, I. Muller, P. Volf, and M.E. Rogers. Promastigote secretory gel from natural and unnatural sand fly vectors exacerbate *Leishmania major* and *Leishmania tropica* cutaneous leishmaniasis in mice. Parasitology 2019; 146(Special Issue 14): pp. 1796-1802.
- [88] Karunaweera, N.D., S. Ginige, S. Senanayake, H. Silva, N. Manamperi, N. Samaranyake, Y. Siriwardana, D. Gamage, U. Senerath, and G. Zhou. Spatial epidemiological trends and hotspots of leishmaniasis, Sri Lanka, 2001-2018. Emerging Infections Diseases 2020; 26(1)
- [89] Holakouie-Naieni, K., E. Mostafavi, A.D. Bolorani, M. Mohebali, R. Pakzad. Spatial modelling of cutaneous leishmaniasis in Iran from 1983-2013. Acta Tropical 2017; 166: pp. 67-73
- [90] Zeb, I., N.A. Qureshi, N. Shaheen, M.I. Zafar, A. Ali, A. Hamid, S.A.A. Shah, and A. Ashraf. Spatiotemporal patterns of cutaneous leishmaniasis in the district upper and lower Dir, Khyber Pakhtunkhwa, Pakistan: A GIS-based spatial approaches. Acta Tropica 2021; 217: 105861.
- [91] Hao, Y., X. Hu, Y. Gong, J. Xue, Z. Zhou, Y. Li, Q. Wang, Y. Zhang, S. Li. Spatiotemporal clustering of mountain-type zoonotic visceral leishmaniasis in China between 2015 and 2019. PLoS Neg. Trop. Dis. 2021; 15(3):e0009152.
- [92] da Silva, T.A.M., W. Coura-Vital, D.S. Barbosa, C.S.F. Oiko, M.H.F. Morais, B.D. Tourinho, D.P.O. de Melo, I.A. Reis, and M. Carneiro. Spatial and temporal trends

- of visceral leishmaniasis by mesoregion in a southeastern state of Brazil, 2002-2013. *PLoS Neg. Trop. Dis.* 2017; 11(10): e0005950.
- [93] Karunaweera, N.D., S Senanayake, S. Ginige, H. Silva, N. Manamperi, N. Samaranyake, R. Dewasurendra, P. Karunanayake, D. Gamage, N. de Silva, U. Senarath, and G. Zhou. Spatiotemporal distribution of cutaneous leishmaniasis in Sri Lanka and future case burden estimates. *PLoS Neg. Trop. Dis.* 2021; 15(4):e0009346.
- [94] Karagiannis-Voules, D.-A., R.G.C. Scholte, L.H. Guimarães, J. Utzinger, and P. Vounatsou. Bayesian geostatistical modeling of leishmaniasis incidence in Brazil. *PLoS Neg. Trop. Dis.* 2013; 7(5): e2213.
- [95] Ullah, W., T.-Y. Yen, S. Niaz, N. Nasreen, Y.-F. Tsai, R.I. Rodriguez-Vivas, A. Khan, and K.-H. Tsai. Distribution and risk of cutaneous leishmaniasis in Khyber Pakhtunkhwa, Pakistan. *Trop. Med. Infect. Dis.* 2023; 8(2): 128.
- [96] Sillero, N. What does ecological modelling model? A proposed classification of ecological nice models based on their underlying methods. *Ecological Modelling* 2021; 222(8): pp. 1343-1346.
- [97] Kearney, M., and W. Porter. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters* 2009; 12(4): pp. 334-350.
- [98] Guisan, A., and W. Thuiller. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 2005; 8(9): pp. 993-1009.
- [99] Cunze, S., and S. Klimpel. From the Balkan towards Western Europe: Range expansion of the golden jackal (*Canis aureus*) - A climatic niche modeling approach. *Ecology and Evolution* 2022; 12(7): e9141.
- [100] Soberón, J., and A.T. Peterson. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biod. Inform.* 2005; 2: pp. 1-10.
- [101] Ferro, C., M. López, P. Fuya, L. Lugo, J.M. Cordovez, and C. González. Spatial distribution of sand fly vectors and eco-epidemiology of cutaneous leishmaniasis transmission in Colombia. *PLoS ONE* 2015; 10(10): e0139931.
- [102] Meneguzzi, V.C., C.B. dos Santos, G.R. Leite, B. Fux, and A. Falqueto. Environmental niche modelling of phlebotomine sand flies and cutaneous leishmaniasis identifies *Lutzomyia intermedia* as the main vector species in Southeastern Brazil. *PLoS ONE* 2016; 11(10): e0164580.
- [103] Moo-Llanes, D., C.N. Ibarra-Cerdeña, E.A. Rebollar-Téllez, S. Ibáñez-Bernal, C. González, and J.M. Ramsey. Current and future niche of North and Central American sand flies (Diptera: Psychodidae) in climate change scenarios. *PLoS Neg. Trop. Dis.* 2013; 7(9): e2421.
- [104] Daoudi, M., A. Outammassine, M. Amane, M. Hafidi, S. Boussaa, and A. Boumezough. Climate change influences on the potential distribution of the sand fly *Phlebotomus sergenti*, vector of *Leishmania tropica* in Morocco. *Acta Parasitologica* 2022; 67: pp. 858-866.

- [105] Fischer, D., P. Moeller, S.M. Thomas, T.J. Naucke, and C. Beierkuhnlein. Combining climatic projections and dispersal ability: A method for estimating the responses of sand fly vector species to climate change. *PLoS Neg. Trop. Dis.* 2011; 5(11): e1407.
- [106] González, C., O. Wang, S.E. Strutz, C. González-Salazar, V. Sánchez-Cordero, and S. Sarkar. Climate change and risk of leishmaniasis in North America: Predictions from ecological niche models of vector and reservoir species. *PLoS Neg. Trop. Dis.* 2010; 4(1): e585.
- [107] Lozier, J.D., P. Aniello, and M.J. Hickerson. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *Journal of Biogeography* 2009; 36(9): pp. 1623-1627.
- [108] Owens, H.L., L.P. Campbell, L.L. Dornak, E.E. Saupe, N. Barve, J. Soberón, K. Ingenloff, A. Lora-Noriega, C.M. Hensz, C.E. Myers, and A.T. Peterson. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* 2013; 263: pp. 10-18.
- [109] Sillero, N., and A.M. Barbosa. Common mistakes in ecological niche models. *International Journal of Geographical Information Science* 2019; 35(2): pp. 213-226.
- [110] Feng, X., D.S. Park, C. Walker, A.T. Peterson, C. Merow, and M. Papeş. A checklist for maximising reproducibility of ecological niche models. *Nat. Ecol. Evol.* 2019; 3: pp. 1382-1395.
- [111] Kleinbaum, D.G., and M. Klein. *Survival Analysis: A self learning text.* Second Edition Springer, 2005; New York.
- [112] Firestone, S.M., N. Cogger, M.P. Ward, J-A. L.M.L. Toribio, B.J. Moloney, and N.K. Dhand. The influence of meteorology on the spread of influenza: survival analysis of an equine influenza (A/H3N8) outbreak. *PLoS ONE* 2012; 7(4):e35284.
- [113] Velasco, H., H. Laniado, M.Toro, A. Catano-López, V. Leiva, and Y. Lio. Modeling the risk of infectious diseases transmitted by *Aedes aegypti* using survival and aging statistical analysis with a case study in Colombia. *Mathematics* 2021; 9(13): 1488.
- [114] Assael, B.M., C. Castellani, M.B. Ocampo, P. Iansa, A. Callegaro, and M.G. Valsecchi. Epidemiology and survival analysis of cystic fibrosis in an area of intense neonatal screening over 30 years. *American Journal of Epidemiology* 2002; 156(5): pp. 397-401.
- [115] Reithinger, R., M. Mohsen, M. Wahid, M. Bismullah, R.J. Quinnell, C.R. Davies, J. Kolaczinski, and J.R. David. Efficacy of thermotherapy to treat cutaneous leishmaniasis caused by *Leishmania tropica* in Kabul, Afghanistan: A randomized, controlled trial. *Clinical Infectious Diseases* 2005; 40(8): pp. 1148-1155.
- [116] Diro, E., K. Ritmeijer, M. Boelaert, F. Alves, R. Mohammed, C. Abongomera, R. Ravinetto, M.D. Crop, H. Fikre, C. Adera, R. Colebunders, H. van Loen, J. Menten, L. Lynen, A. Hailu, and J. van Griensven. Use of pentamidine as secondary prophylaxis to prevent visceral leishmaniasis relapse in HIV infected patients, the

- first twelve months of a prospective cohort study. *PLoS Neg. Trop. Dis.* 2015; 9(10): e0004087.
- [117] Maia-Elkhoury, A.N.S., G.A.S. Romero, S.Y.O.B. Valadas, M.L. Sousa-Gomez, J.A.L. Lindoso, E. Cupolillo, J.A. Ruiz-Postigo, D. Argaw, and M.J. Sanchez-Vazquez. Premature deaths by visceral leishmaniasis in Brazil investigated through a cohort study: A challenging opportunity? *PLoS Neg. Trop. Dis.* 2019; 13(12): e0007841.
- [118] Perry, M.R., V.K. Prajapati, J. Menten, A. Raab, J. Feldmann, D. Chakraborti, S. Sundar, A.H. Fairlamb, M. Boelaert, and A. Picado. Arsenic exposure and outcomes of antimonial treatment in visceral leishmaniasis patients in Bihar, India: A retrospective cohort study. *PLoS Neg. Trop. Dis.* 2015; 9(3): e0003518.
- [119] Uranw, S., B. Ostyn, A. Rijal, S. Devkota, B. Khanal, J. Mentel, M. Boelaert, and S. Rijal. Post-kala-azar dermal leishmaniasis in Nepal: A retrospective cohort study (2000-2010). *PLoS Neg. Trop. Dis.* 2011; 5(12): e1433.
- [120] Coura-Vital, W., A.B. Reis, S.L. Braga, B.M. Roatt, R.D. de O. Aguiar-Soares, M.J. Marques, V.M. Veloso, and M. Carneiro. Canine visceral leishmaniasis: Incidence and risk factors for infection in a cohort study in Brazil. *Veterinary Parasitology* 2013; 197(3-4): pp. 411-417.
- [121] Oliveira, A.M., R.V.M. López, M.R. Dibo, L.A.C. Rodas, M.M. Guirado, and F. Chiaravalloti-Neto. Dispersion of *Lutzomyia longipalpis* and expansion of visceral leishmaniasis in São Paulo State, Brazil: identification of associated factors through survival analysis. *Parasites and Vectors* 2018; 11:503.
- [122] Hosmer Jr., D.W., S. Lemeshow, and R.X. Sturdivant. *Applied Logistic Regression* (Third Edition) John Wiley and Sons 2013; Hoboken, New Jersey.
- [123] Das, S., and R.M. Rahman. Application of ordinal logistic regression analysis in determining risk factors of child malnutrition in Bangladesh. *Nutrition Journal* 2011; 10: 124.
- [124] Lacerda, E.M., K. Geraghty, C.C. Kingdon, L. Palla, and L. Nacul. A logistic regression analysis of risk factors in ME/CFS pathogenesis. *BMC Neurology* 2019; 19: 275.
- [125] Maja, T.F., and D. Maposa. An investigation of risk factors associated with tuberculosis transmission in South Africa using logistic regression model. *Infect. Dis. Rep.* 2022; 14(4): pp. 609-620.
- [126] Feliciangeli, M.D., O. Delgado, B. Suarez, and A. Bravo. *Leishmania* and sand flies: proximity to woodland as a risk factor for infection in a rural focus for visceral leishmaniasis in west central Venezuela. *Tropical Medicine and International Health* 2006; 11(12): pp. 1785-1791.
- [127] Nackers, F., Y.K. Mueller, N. Salih, M.S. Elhag, M.E. Elbadawi, O. Hammam, A. Mumina, A.A. Atia, J.-F. Etard, K. Ritmeijer, and F. Chappuis. Determinants of

visceral leishmaniasis: A case-control study in Gedaref State, Sudan. *PLoS Neg. Trop. Dis.* 2015; 9(11): e0004187.

- [128] Ngere, I., W.G. Boru, A. Isack, J. Muiruri, M. Obonyo, S. Matendehero, and Z. Gura. Burden and risk factors of cutaneous leishmaniasis in a peri-urban settlement in Kenya, 2016. *PLoS ONE* 2020; 15(1): e0227697.
- [129] Rombolà, P., G. Barlozzari, A. Carvelli, M. Scarpulla, F. Iacoponi, and G. Macrì. Seroprevalence and risk factors associated with exposure to *Leishmania infantum* in dogs, in an endemic Mediterranean region. *PLoS ONE* 2021; 16(1): e0244923.
- [130] Younis, L.G., A. Kroeger, A.B. Joshi, M.L. Das, M. Omer, V.K. Singh, C.K. Gurung, and M.R. Banjara. Housing structure including the surrounding environment as a risk factor for visceral leishmaniasis transmission in Nepal. *PLoS Neg. Trop. Dis.* 2020; 14(3): e0008132.
- [131] Yared, S., K. Deribe, A. Gebreselassie, W. Lemma, E. Akililu, O.D. Kirstein, M. Balkew, A. Warburg, T. Gebre-Michael, and A. Hailu. Risk factors of visceral leishmaniasis: a case control study in north-western Ethiopia. *Parasites and Vectors* 2014; 7: 470.
- [132] Eshetu, B., and H. Mamo. Cutaneous leishmaniasis in north-central Ethiopia: trend, clinical forms, geographic distribution, and determinants. *Tropical Medicine and Health* 2020; 48: 39.
- [133] Weigle, K.A., C. Santrich, F. Martinez, L. Valderrama, and N.G. Saravia. Epidemiology of cutaneous leishmaniasis in Colombia: Environmental and behavioural risk factors for infection, clinical manifestations, and pathogenicity. *The Journal of Infectious Diseases* 1993; 168(3): pp. 709-714.
- [134] Rostami, M.N., A. Saghatipour, and E. Vesali. A newly emerged cutaneous leishmaniasis focus in central Iran. *International Journal of Infectious Diseases* 2013; 17(12): pp. e1198-e1206.
- [135] Davies, C.R., E.A. Llanos-Cuentas, P. Campos, J. Monge, P. Villaseca, and C. Dye. Cutaneous leishmaniasis in the Peruvian Andes: risk factors identified from a village cohort study. *Am. J. Trop. Med. Hyg.* 1997; 56(1): pp. 85-95.
- [136] Lemma, W., H. Tekie, S. Yared, M. Balkew, T. Gebre-Michael, A. Warburg, and A. Hailu. Sero-prevalence of *Leishmania donovani* infection in labour migrants and entomological risk factors in extra-domestic habitats of Kafta-Humera lowlands - kala-azar endemic areas in the northwest Ethiopia. *BMC Infectious Diseases* 2015; 15: 99.
- [137] Tedla, D.G., F.H. Bariagabr, and H.H. Abreha. Incidence and trends of leishmaniasis and its risk factors in Humera, Western Tigray. *Journal of Parasitology Research* 2018; 8463097.

- [138] Poché, D., R. Garlapati, K. Ingenloff, J. Remmers, and R. Poché. Bionomics of phlebotomine sand flies from three villages in Bihar, India. *Journal of Vector Ecology* 2011; 36(s1): pp. S106-S117.
- [139] Bern, C., A.B. Joshi, S.N. Jha, M.L. Das, A. Hightower, G.D. Thakur, and M.B. Bista. Factors associated with visceral leishmaniasis in Nepal: Bed-net use is strongly protected. *Am. J. Trop. Med. Hyg.* 2000; 63(3,4): pp. 184-188.
- [140] Reithinger, R., M. Mohsen, and T. Leslie. Risk factors for anthroponotic cutaneous leishmaniasis at the household level in Kabul, Afghanistan. *PLoS Neg. Trop. Dis.* 2010; 4(3): e639.
- [141] Uranw, S., E. Hasker, L. Roy, F. Meheus, M.L. Das, N.R. Bhattarai, S. Rijal, and M. Boelaert. An outbreak investigation of visceral leishmaniasis among residents of Dharan town, eastern Nepal, evidence for urban transmission of *Leishmania donovani*. *BMC Infectious Diseases* 2013; 13: 21.
- [142] Votýpka, J., O.E. Kasap, P. Volf, P. Kodym, and B. Alten. Risk factors for cutaneous leishmaniasis in Cukurova region, Turkey. *Trans. R. Soc. Trop. Med. Hyg.* 2012; 106(3): pp. 186-190.
- [143] Custodio, E., E. Gadisa, L. Sordo, I. Cruz, J. Moreno, J. Nieto, C. Chicharro, A. Aseffa, Z. Abraham, T. Hailu, and C. Cañavate. Factors associated with *Leishmania* asymptomatic infection: Results from a cross-sectional survey in highland Northern Ethiopia. *PLoS Neg. Trop. Dis.* 2012; 6(9): e1813.
- [144] Singh, S.P., E. Hasker, A. Picado, K. Gidwani, P. Malaviya, R.P. Singh, M. Boelaert, and S. Sundar. Risk factors for visceral leishmaniasis in India: further evidence on the role of domestic animals. *Tropical Medicine and International Health* 2010; 15(S2): pp. 29-35.
- [145] Al-Ashwal, M.A., W.M. Atroosh, A.H. Al-Adhroey, A.A. Al-Subbary, L. Yee-Ling, and H.M. Al-Mekhlafi. A disfiguring neglected tropical disease sweeps war-torn Yemen: a community-based study of prevalence and risk factors of cutaneous leishmaniasis among rural communities in the western highlands. *Trans. R. Soc. Trop. Med. Hyg.* 2023; trad044.
- [146] Chapman, L.A.C., C.P. Jewell, S.E.F. Spencer, L. Pellis, S. Datta, R. Chowdhury, C. Bern, G.F. Medley, and T.D. Hollingsworth. The role of case proximity in the transmission of visceral leishmaniasis in a highly endemic village in Bangladesh. *PLoS Neg. Trop. Dis.* 2018; 12(10):e0006453.
- [147] Bern, C., A.W. Hightower, R. Chowdhury, M. Ali, J. Amann, Y. Wagatsuma, R. Haque, K. Kurkjian, L.E. Vaz, M. Begum, T. Akter, C.B. Cetre-Sossah, I.B. Ahluwalia, E. Dotson, W.E. Secor, R.F. Breiman, and J.H. Maguire. Risk factors for kala-azar in Bangladesh. *Emerg. Infect. Dis.* 2005; 11:5, pp.655-662.
- [148] Pitchford, J.W., and J. Brindley. Prey patchiness, predator survival, and fish recruitment. *Bull. Math. Biol.* 2001; 63: pp. 527-546.

- [149] Riberio, C.J.N., A.D. dos Santos, S.V.M.A. Lima, E.R. da Silva, B.V.S. Ribeiro, A.M. Duque, M.V.S. Peixoto, P.L. dos Santos, I.M. de Oliveira, M.W. Lipscomb, K.C.G.M. de Araújo, and t.R. de Moura. Space-time risk cluster of visceral leishmaniasis in Brazilian endemic region with high social vulnerability: an ecological time series study. *PLoS Neg. Trop. Dis.* 2021; 15(1): e0009006.
- [150] Lloyd, A.L., and S. Valeika. Network models in epidemiology: An overview *Network* 2005; 18:38.
- [151] Corcoran, C., and A. Hastings. A low-dimensional network model for an SIS epidemic: Analysis of the super compact pairwise model. *Bulletin of Mathematical Biology* 2021; 83:77.
- [152] Simon, P.L., and I.Z. Kiss. Super compact pairwise model for SIS epidemic on heterogeneous networks. *Journal of Complex Networks* 2016; 4, pp.187-200.
- [153] Bañuls AL, Bastien P, Pomares C , Arevalo J, Fisa R, Hide, M. Clinical pleiomorphism in human leishmaniasis, with special mention of asymptomatic infection. *Clin. Microbiol. Infect.* 2011 Oct; 17(10):1451-1461.
- [154] Bates PA. Revisiting leishmania's life cycle *Nat. Microbiol.* 2018; 3:529-530
- [155] Handman E, Bullen DVR. Interaction of *Leishmania* with the host macrophage. *Trends Parasitol.* 2002 Aug; 18(8):332-334
- [156] Sadlova, J, Myskova, J, Lestinova, T, Votypka, J, Yeo, M, Volf, P. *Leishmania donovani* development in *Phlebotomus argentipes*: comparison of promastigote- and amastigote-initiated infections. *Parasitology* 2017; 144:403-410
- [157] Massad E. The elimination of Chagas' disease from Brazil. *Epidemiol. Infect.* 2008 Sep; 136(9): 1153-1164
- [158] Luz PM, Struchiner CJ, Galvani AP. Modeling transmission dynamics and control of vector-borne neglected tropical diseases. *PLoS Negl. Trop. Dis.* 2010 Oct; 4(10): e761
- [159] Kimblin, N., N. Peters, A. Debrabant, N. Secundino, J. Egen, P. Lawyer, M.P. Fay, S. Kamhawi, and D. Sacks. Quantification of the infectious dose of *Leishmania major* transmitted to the skin by single sand flies. *PNAS* 2008; 105(29): pp. 10125-10130.
- [160] Abdel-Hamid YM. Life table characteristics of the female sand fly, *Phlebotomus papatasi* (Scopoli) (Diptera: Psychodidae) under three food regimes. *J. Vector Borne Dis.* 2012 Dec; 49(4):226-229
- [161] Killick-Kendrick R. Recent advances and outstanding problems in the biology of Phlebotomine sand flies: a review *Acta Trop.* 1978; 35(4): 297-313
- [162] Volf P, Kiewegová A, Nemeč A. Bacterial colonisation of the gut of *Phlebotomus duboscqi* (Diptera: Psychodidae): transtadial passage and the role of female diet. *Folia Parasitol. (Praha)* 2002; 49:73-77

- [163] Killick-Kendrick R, Rioux JA. Mark-release-recapture of sand flies fed on leishmanial dogs: the natural life-cycle of *Leishmania infantum* in *Phlebotomus ariasi*. *Parassitologia* 2002; 44(1-2): 67-71
- [164] Shokri A, Fakhhar M, Teshnizi SH. Canine visceral leishmaniasis in Iran: A systematic review and meta-analysis. *Acta Trop.* 2017 Jan; 165:76-89
- [165] Assefa A. Leishmaniasis in Ethiopia: A systematic review and meta-analysis of prevalence in animals and humans. *Heliyon.* 2018 Aug; 4(8):e00723
- [166] Rogers ME. The role of leishmania proteophosphoglycans in sand fly transmission and infection of the Mammalian host *Front. Microbiol.* 2012 Jun; 3: 223
- [167] Inbar E, Hughitt VK, Dillon LAL, Ghosh K, El-Sayed NM, Sacks DL. The Transcriptome of *Leishmania* major Developmental Stages in Their Natural Sand Fly Vector. *MBio.* 2017 Apr; 8(2)
- [168] Poisot T. The digitize package: extracting numerical data from scatterplots. *R J.* 2011; 3(1):25-26
- [169] Miller, E., A. Warburg, I. Novikov, A. Hailu, P. Volf, V. Seblova, and A. Huppert. Quantifying the contributions of hosts with different parasite concentrations to the transmission of visceral leishmaniasis in Ethiopia. *PLoS Neg Trop Dis* 2014; 8(10): e3288.
- [170] Stauch, A., R.R. Sarkar, A. Picado, B. Ostyn, S. Sundar, S. Rijal, M. Boelaert, J-C. Dujardin, and H-P. Duerr. Visceral leishmaniasis in the Indian subcontinent: modelling epidemiology and control. *PLoS Neg Trop Dis* 2011; 5(11): e1405.
- [171] Borja, L.S., O.M.F. de Sousa, M. da Silva Solcà, L.A. Bastos, M. Bordoni, J.T. Magalhães, D.F. Larangeira, S.M. Barrouin-Melo, D.B.M. Fraga, and P.S.T. Veras. Parasite load in the blood and skin of dogs naturally infected by *Leishmania infantum* is correlated with their capacity to infect sand fly vectors. *Veterinary Parasitology* 2016; 229: pp. 110-117.
- [172] Carmichael, S., B. Powell, T. Hoare, P.B. Walrad, J.W. Pitchford. Variable bites and dynamic populations: new insights in *Leishmania* transmission. *PLoS Neg Trop Dis* 2021; 15(1): e0009033.
- [173] Lewis, D.J. Phlebotomid sandflies. *Bulletin of the World Health Organisation* 1971; 44: pp. 531-555.
- [174] Pérez-Cutillas, P., C. Muñoz, J.M-D. Puente, J. Figuerola, R. Navarro, M. Ortuño, L.J. Bernal, J. Ortiz, R.C. Soriguer, and E. Berriatua. A spatial ecology study in a high-diversity host community to understand blood-feeding behaviour in *Phlebotomus* sandfly vectors of *Leishmania*. *Medical and Veterinary Entomology* 2020; 34: pp.164-174.
- [175] Killick-Kendrick, R., J-A. Rioux, M. Bailly, M.W. Guy, T.J. Wilkes, F.M. Guy, I. Davidson, R. Knechtli, R.D. Ward, E. Guilvard, J. Perieres, and H. Dubois. *Ecology*

of leishmaniasis in the south of France. Dispersal of *Phlebotomus ariasi* Tonnoir, 1921 as a factor in the spread of visceral leishmaniasis in the Cévennes. *Ann. Parasitol. Hum. Comp.* 1984; 59:6, pp. 555-572.

- [176] Wasserberg, G., I. Yarom, and A. Warburg. Seasonal abundance patterns of the sandfly *Phlebotomus papatasi* in climatically distinct foci of cutaneous leishmaniasis in Israeli deserts. *Medical and Veterinary Entomology* 2003; 17:4, 452-456.
- [177] Risueño, J., C. Muñoz, P. Pérez-Cutillas, E. Goyena, M. González, M. Ortuño, L.J. Bernal, J. Ortiz, B. Alten, and E. Berriatua. Understanding *Phlebotomus perniciosus* abundance in south-east Spain: assessing the role of environmental and anthropic factors. *Parasites and Vectors* 2017; 10:189.
- [178] Tabbabi, A., J. Ghrab, K. Aoun, P.D. Ready, and A. Bouratbine. Habitats of the sandfly vectors of *Leishmania tropica* and *L. major* in a mixed focus of cutaneous leishmaniasis in southeast Tunisia. *Acta Tropica* 2011; 119: pp. 131-137.
- [179] Feng, X., S. Ruan, Z. Teng, and K. Wang. Stability and backward bifurcation in a malaria transmission model with applications to the control of malaria in China. *Mathematical Biosciences* 2015; 266: pp.52-64.
- [180] Wang, X., and X-Q. Zhao. A malaria transmission model with temperature-dependent incubation period. *Bull. Math. Biol.* 2017; 79: pp.1155-1182.
- [181] Craig, M.H., R.W. Snow, and D. le Sueur. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today* 1999; 15(3): pp.105-111.
- [182] J.A.N. Filipe, E.M. Riley, C.J. Drakeley, C.J. Sutherland, and A.C. Ghani. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. *PLoS Comput. Biol.* 2007; 3(12):e255.
- [183] Bi, K., Y. Chen, S. Zhao, D. Ben-Arieh, and C-H.(J.) Wu. A new zoonotic visceral leishmaniasis dynamic transmission model with age-structure. *Chaos, Solitons and Fractals* 2020; 133:109622
- [184] Le Rutte, E.A., L.A.C. Chapman, L.E. Coffeng, S. Jervis, E.C. Hasker, S. Dwivedi, M. Karthick, A. Das, T. Mahapatra, I. Chaudhuri, M.C. Boelaert, G.F. Medley, S. Srikantiah, D.T. Hollingsworth, and S.J. de Vlas. Elimination of visceral leishmaniasis in the Indian Subcontinent: a comparison of predictions from three transmission models. *Epidemics* 2017; 18: pp.67-80
- [185] Stauch, A., H-P. Duerr, A. Picado, B. Ostyn, S. Sundar, S. Rijal, M. Boelaert, J-C. Dujardin, and M. Eichner. Model-based investigations of different vector-related intervention strategies to eliminate visceral leishmaniasis on the Indian Subcontinent. *PLoS Neg. Trop. Dis.* 2014; 8(4):e2810
- [186] Schmidt, C.A., G. Comeau, A.J. Monaghan, D.J. Williamson, and K.C. Ernst. Effects of desiccation stress on adult female longevity in *Aedes aegypti* and *Ae. albopictus*.

tus (Diptera: Culicidae): results of a systematic review and pooled survival analysis. *Parasites and Vectors* 2018; 11:267.

- [187] Alfani, G., and M. Bonetti. A survival analysis of the last great European plagues: the case of Nonantola (Northern Italy) in 1630. *Population Studies* 2019; 73(1): pp.101-118.
- [188] Therneau, T. A package for survival analysis in R. Accessed 14/02/2020. Available from <https://cran.r-project.org/web/packages/survival/index.html>
- [189] Cox, D.R. Regression models and life-tables. *Journal of the Royal Statistical Society, Ser. B* 1972; 34(2): pp. 187-202.
- [190] Mittal, S., D. Madigan, R.S. Burd, and M.A. Suchard. High-dimensional, massive sample-size Cox proportional hazards regression for survival analysis. *Biostatistics* 2014; 15(2): pp. 207-221.
- [191] Bustan, M.N., Arman, M.K. Aidid, F.A. Gobel, and Syamsidar. Cox proportional hazard survival analysis to inpatient breast cancer cases. *J. Phys.: Conf. Ser.* 2018; 1028:012230.
- [192] Fox, J., and S. Weisberg. *An R companion to applied regression*. Third Edition. Sage 2019; Thousand Oaks, CA.
- [193] Ali, A.M.G., S.-J. Dawson, F.M. Blows, E.Provenzano, I.O. Ellis, L. Baglietto, D. Huntsman, C. Caldas, and P.D. Pharoah. Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer. *British Journal of Cancer* 2011; 104: pp. 693-699.
- [194] Tarkhan, A., and N. Simon. An online framework for survival analysis: reframing Cox proportional hazards model for large data sets and neural networks. *Biostatistics* 2022; kxac039.
- [195] Orbe, J., E. Ferreira, and V. Núñez-Antón. Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine* 2002; 21(22): pp. 3493-3510.
- [196] Sparapani, R.A., B.R. Logan, R.E. McCulloch, and P.W. Laud. Nonparametric survival analysis using Bayesian additive regression trees (BART). *Statistics in Medicine* 2016; 35(16): pp. 2741-2753.
- [197] Lakner, C., P. van der Mark, J.P. Huelsenbeck, B. Larget, and F. Ronquist. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. *Systematic Biology* 2008; 57:1, pp. 86-103.
- [198] Turek, D., P. da Valpine, C.J. Paciorek and C. Anderson-Bergman. Automated parameter blocking for efficient Markov chain Monte Carlo sampling. *Bayesian Analysis* 2017; 12:2, pp. 465-490.

- [199] Mossel, E., and E. Vigoda. Limitations of Markov chain Monte Carlo algorithms for Bayesian inference of phylogeny. *The Annals of Applied Probability* 2006; 16:4, pp. 2215-2234.
- [200] Dobson, P., I. Fursov, G. Lord, and M. Ottobre. Reversible and non-reversible Markov chain Monte Carlo algorithms for reservoir simulation problems. *Computational Geosciences* 2020; 24, pp. 1301-1313.
- [201] Bern, C., R. Haque, R. Chowdhury, M. Ali, K. Kurkjian, L.E. Vaz, J. Amann, M.A. Wahed, Y. Wagatsuma, R.F. Breiman, J. Williamson, W.E. Secor, and J.H. Maguire. The epidemiology of visceral leishmaniasis and asymptomatic leishmanial infection in a highly endemic Bangladeshi village. *Am. J. Trop. Med. Hyg.* 2007; 76:5, pp.909-914
- [202] Picado, A., A.P. Dash, S. Bhattacharya, and M. Boelaert. Vector control interventions for visceral leishmaniasis elimination initiative in South Asia, 2005-2010. *Indian J. Med. Res.* 2012; 136(1): pp.22-31.
- [203] Ahluwalia, I.B., C. Bern, C. Costa, T. Akter, R. Chowdhury, M. Ali, D. Alam, E. Kenah, j. Amann, M. Islam, Y. Wagatsuma, R. Haque, R.F. Breiman, and J.H. Maguire. Visceral leishmaniasis: consequences of a neglected tropical disease in a Bangladeshi community. *Am. J. Trop. Med. Hyg.* 2003; 69(6): pp. 624-628.
- [204] Pigott, D.M., N. Golding, J.P. Messina, K.E. Battle, K.A. Duda, Y. Balard, P. Bastien, F. Pratlong, J.S. Brownstein, C.C. Freifeld, S.R. Mekaru, L.C. Madoff, D.B. George, M.F. Myers, and S.I. Hay. Global database of leishmaniasis occurrence locations, 1960-2012. *Scientific Data* 2014; 1:140036
- [205] Pigott, D.M., S. Bhatt, N. Golding, K.A. Duda, K.E. Battle, O.J. Brady, J.P. Messina, Y. Balard, P. Bastien, F. Pratlong, J.S. Brownstein, C.C. Freifeld, S.R. Mekaru, P.W. Gething, D.B. George, M.F. Myers, R. Reithinger, and S.I. Hay. Global distribution maps of the leishmaniasis. *eLife* 2014; 3:e02851.
- [206] Harel, O., and X.-H. Zhou. Multiple imputation: review of theory, implementation and software. *Statistics in Medicine* 2007; 26(16): pp. 3057-3077.
- [207] Kenward, M.G., and J. Carpenter. Multiple imputation: current perspectives. *Statistical Methods in Medical Research* 2007; 16(3): pp. 199-218.
- [208] Bhatt, S., P.W. Gething, O.J. Brady, J.P. Messina, A.W. Farlow, C.L. Moyes, J.M. Drake, J.S. Brownstein, A.G. Hoen, O. Sankoh, M.F. Myers, D.B. George, T. Jaenisch, G.R.W. Wint, C.P. Simmons, T.W. Scott, J.J. Farrar, and S.I. Hay. The global distribution and burden of dengue. *Nature* 2013; 496: pp. 504-507.
- [209] Elith, J., J.R. Leathwick, and T. Hastie. A working guide to boosted regression trees. *Journal of Animal Ecology* 2008; 77(4): pp. 802-813.
- [210] IPEA. Atlas da vulnerabilidade social nos municípios brasileiros. Costa, M.A., Marguti, B.O., editors. Instituto de Pesquisa Econômica Aplicada, Brasília: IPEA; 2015.

- [211] Kim, H.J., M.P. Fay, E.J. Feuer, and D.N. Midthune. Permutation tests for joinpoint regression with applications to cancer rates. *Stat. Med.* 2000; 19(3): pp. 335-351.
- [212] de Albuquerque, M.A.C., D.M. Dias, L.T. Vieira, C.A. Lima, and A.M. da Silva. Mortality trends for neglected tropical diseases in the state of Sergipe, Brazil, 1980-2013. *Infect. Dis. Poverty* 2017; 6(1): 20.
- [213] de Freitas Rocha, A.T., G.M. de Espindola, M.R.A. Soares, J. de R. de Sousa Rocha, and C.H.D. Costa. Visceral leishmaniasis and vulnerability conditions in an endemic urban area of northeastern Brazil. *Trans. R. Soc. Trop. Med. Hyg.* 2018; 112(7): pp. 317-325.
- [214] Kulldorff, M. A spatial scan statistic. *Commun. Stat. Theory Methods* 1997; 26(6): pp. 1481-1496.
- [215] Bermudi, P.M.M., M.M. Guirado, L.A.C. Rodas, M.R. Dibo, F. Chiaravalloti-Neto. Spatio-temporal analysis of the occurrence of human visceral leishmaniasis in Araçatuba, State of São Paulo, Brazil. *Rev. Soc. Bras. Med. Trop.* 2018; 51(4): pp. 452-460.
- [216] de Souza, C.D.F., V.S. Rocha, N.F. Santos, T.C. Leal, J.P.S. de Paiva, C.C.C. Oliveira, P.R.S. Martins-Filho, M.A.F.M. Magalhães, L.E. Cuevas, and V.S. Santos. Spatial clustering, social vulnerability, and risk of leprosy in an endemic area in Northeast Brazil: An ecological study. *J. Eur. Acad. Dermatol. Venereol.* 2019; 33(8): pp. 1581-1590.
- [217] Harrell, F.E. Regression modeling studies with applications to linear models, logistic regression, and survival analysis. Springer, New York 2001.
- [218] Emmert-Streib, F., and M. Dehmer. Introduction to survival analysis in practice. *Mach. Learn. Knowl. Extra.* 2019; 1(3): pp. 1013-1038.
- [219] Brasil. Guide to Health Surveillance. 3rd ed. Brasília; Ministério da Saúde. 2019.
- [220] Zhu, X., S. Jamshed, J. Zou, A. Azar, X. Meng, V. Bathini, K. Dresser, C. Strock, B. Yalamarti, M. Yang, K. Tomaszewicz, G. Tjionas, M.C. Mochel, L. Hutchinson, and J.R. Bledsoe. Molecular and immunophenotypic characterization of anal squamous cell carcinoma reveals distinct clinicopathologic groups associated with HPV and *TP53* mutation status. *Mod. Pathol.* 2021; 34(5): pp. 1017-1030.
- [221] de Cap, M.R., and S. Kaul. Limitations of multivariate survival analysis. *Mod. Pathol.* 2022; 35: 993.
- [222] Peduzzi, P., J. Concato, A.R. Feinstein, and T.R. Holford. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *J. Clin. Epidemiol.* 1995; 48(12): pp. 1503-1510.
- [223] Ogundimu, E.O., D.G. Altman, and G.S. Collins. Adequate sample size for developing prediction models is not simply related to events per variable. *J. Clin. Epidemiol.* 2016; 76: pp. 175-182.

- [224] Austin, P.C., A. Allignol, and J.P. Fine. The number of primary events per variable affects estimation of the subdistribution hazard competing risks model. *J. Clin. Epidemiol.* 2017; 83: pp. 75-84.
- [225] Schober, P., and T.R. Vetter. Survival analysis and interpretation of time-to-event data: The tortoise and the hare. *Anesth. Analg.* 2018; 127(3): pp. 792-798.
- [226] Clark, T.G., M.J. Bradburn, S.B. Love, and D.G. Altman. Survival analysis part IV: Further concepts and methods in survival analysis. *British Journal of Cancer* 2003; 89: pp. 781-786.
- [227] Schmoor, C., and M. Schumacher. Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Statistics in Medicine* 1997; 16(3): pp. 225-237.
- [228] Altman, D.G. Suboptimal analysis using ‘optimal’ cutpoints. *British Journal of Cancer* 1998; 78: pp. 556-557.
- [229] Kariyawasam, K.K.G.D.U.L., C.S. Edirisuriya, U. Senerath, D. Hensmen, H.V.Y.D. Siriwardana, and N.D. Karunaweera. Characterisation of cutaneous leishmaniasis in Matara district, southern sri Lanka: evidence for case clustering. *Pathogens and Global Health* 2015; 109(7): pp.336-343.
- [230] Basimike, M., M.J. Mutinga, and R. Kumar. Habitat preferences and seasonal variations of phlebotomine sandflies (Diptera, Psychodidae) in Marigat area, Baringo district, Kenya. *International Journal of Tropical Insect Science* 1992; 13: pp.307-314.
- [231] Elnaiem, D.A., H.K. Hassan, and R.D. Ward. Phlebotomine sandflies in a focus of visceral leishmaniasis in a border area of eastern Sudan. *Annals of Tropical Medicine and Parasitology* 1997; 91(3): pp.307-318.
- [232] Poché, D.M., R.M. Poché, S. Mukherjee, G.A. Frankowiak, L.N. Briley, D.J. Somers, and R.B. Garlapati. Phlebotomine sandfly ecology on the Indian subcontinent: does village vegetation play a role in sandfly distribution in Bihar, India? *Medical and Veterinary Entomology* 2017; 31(2): pp.207-213.
- [233] Gomez-Barroso, D., Z. Herrador, J.V. san Martín, A. Gherasim, M. Aguado, A. Romero-Maté, L. Molina, P. Aparicio, and A. Benito. Spatial distribution and cluster analysis of leishmaniasis outbreak in the south-western Madrid region, Spain, September 2009 - April 2013. *Euro Surveill.* 2015; 20(7):21037
- [234] Campbell-Lendrum, D.H., S.P. Brandhão-Filho, P.D. Ready and C.R. Davies. Host and/or site loyalty of *Lutzomyia whitmani* (Diptera: Psychodidae) in Brazil. *Medical and Veterinary Entomology* 2002; 13(2): pp.209-211.
- [235] Hasibeder, G., and C. Dye. Population dynamics of mosquito-borne disease: persistence in a completely heterogeneous environment. *Theoretical Population Biology* 1988; 33: 31-53.

- [236] Antonialli, S.A.C., T.G. Torres, A.C.P. Filho, and J.E. Tolezano. Spatial analysis of American visceral leishmaniasis in Mato Grosso do Sul State, Central Brazil. *Journal of Infection* 2007; 54: 509-514.
- [237] Eroglu, F., and O. Ozgoztasi. The increase in neglected cutaneous leishmaniasis in Gaziantep province of Turkey after mass human migration. *Acta Tropica* 2019; 192: pp.138-143.
- [238] Marlet, M.V.L., D.K. Sang, K. Ritmeijer, R.O. Muga, J. Onsongo, R.N. Davidson. Emergence or re-emergence of visceral leishmaniasis in areas of Somalia, northeastern Kenya, and southeastern Ethiopia in 2000-2001. *Trans Roy Soc Trop Med and Hyg* 2003; 97(5): pp.515-518.
- [239] Leta, S., T.H.T. Dao, F. Mesele, and G. Alemayehu. Visceral leishmaniasis in Ethiopia: an evolving disease. *PLoS Neg. Trop. Dis.* 2014; 8(9):e3131.
- [240] Sevá, A. da P., L. Mao, F. Galvis-Ovallos, J.M.T. Lima, and D. Valle. Risk analysis and prediction of visceral leishmaniasis dispersion in São Paulo state, Brazil. *PLoS Neg. Trop. Dis.* 2017; 11(2):e0005353.
- [241] Mondal, D., C. Bern, D. Ghosh, M. Rashid, R. Molina, R. Chowdhury, R. Nath, P. Ghosh, L.A.C. Chpaman, A. Alim, G. Bilbe, and J. Alvar. Quantifying the infectiousness of post-kala-azar dermal leishmaniasis towards sand flies. *Clinical Infectious Diseases* 2019; 69(2): pp. 251-258.
- [242] Le Rutte, E.A., E.E. Jijlstra, and S.J. de Vlas. Post-kala-azar dermal leishmaniasis as a reservoir for visceral leishmaniasis transmission. *Trends in Parasitology* 2019; 35(8): pp. 590-592.
- [243] Hay, S.I., C.A. Guerra, A.J. Tatem, P.M. Atkinson, and R.W. Snow. Urbanization, malaria transmission, and disease burden in Africa. *Nature Reviews Microbiology* 2005; 3, pp. 81-90.
- [244] Tatem, A.J., P.W. Gething, D.L. Smith, and S.I. Hay. Urbanization and the global malaria recession. *Malaria Journal* 2013; 12:133.
- [245] Robert, M.A., R.C. Christofferson, N.J.B. Silva, C. Vasquez, C.N. Mores, H.J. Wearing. Modeling mosquito-borne disease spread in U.S. urban areas: the case of dengue in Miami. *PLoS ONE* 2016; 11:8, e0161365.
- [246] Wilke A.B.B., C. Chase, C. Vasquez, A. Carvajal, J. Medina, W.D. Petrie, and J.C. Beier. Urbanization creates diverse aquatic habitats for immature mosquitoes in urban areas. *Scientific Reports* 2019; 9:15335.
- [247] Wilke, A.B.B., C. Vasquez, A. Carvajal, M. Moreno, D.O. Fuller, G. Cardenas, W.D. Petrie, and J.C. Beier. Urbanization favors the proliferation of *Aedes aegypti* and *Culex quinquefasciatus* in urban areas of Miami-Dade County, Florida. *Scientific Reports* 2021; 11:22989.

- [248] Kolimenakis, A., S. Heinz, M.L. Wilson, V. Winkler, L. Yakob, A. Michaelakis, D. Papachristos, C. Richardson, and O. Horstick. The role of urbanisation in the spread of *Aedes* mosquitoes and the diseases they transmit - A systematic review. *PLoS Neg. Trop. Dis.* 2021; 15(9): e0009631.
- [249] Chavy, A., A.F.D. Nava, S.L.B. Luz, J.D.Ramírez, G. Herrera, T.V. dos Santos, M. Ginouves, M. Demar, G. Prévot, J-F. Guégan, and B. de Thoisy. Ecological niche modelling for predicting the risk of cutaneous leishmaniasis in the Neotropical moist forest biome. *PLoS Neg. Trop. Dis.* 2019; 13(8): e0007629.
- [250] Berrozpe, P., D. Lamattina, M.S. Santini, A.V. Araujo, M.E. Utgés, and O.D. Salomón. Environmental suitability for *Lutzomyia longipalpis* in a subtropical city with a recently established visceral leishmaniasis transmission cycle, Argentina. *Mem. Inst. Oswaldo Cruz* 2017; 112(10).
- [251] Coulborn, R.M., T.G. Gebrehiwot, M. Schneider, S. Gerstl, C. Adera, M. Herrero, K. Porten, M. den Boer, K. Ritmeijer, J. Alvar, A. Hassen, and A. Mulugeta. Barriers to access to visceral leishmaniasis diagnosis and care among seasonal mobile workers in Western Tigray, Northern Ethiopia: A qualitative study. *PLoS Neg. Trop. Dis.* 2018; 12(11): e0006778.
- [252] Priyamvada, K., J. Bindroo, M.P. Sharma, L.A.C. Chapman, P. Dubey, T. Mahapatra, A.W. Hightower, C. Bern, and S. Srikantiah. Visceral leishmaniasis outbreaks in Bihar: community-level investigations in the context of elimination of kala-azar as a public health problem. *Parasites and Vectors* 2021; 14:52.
- [253] Berhe, R., M. Spigt, F. Schneider, L. Paintain, C. Adera, A. Nigusie, Z. Gizaw, Y.A. Tesfaye, D-E.A. Elnaiem, and M. Alemayehu. Understanding the risk perception of visceral leishmaniasis exposure and the acceptability of sandfly protection measures among migrant workers in the lowlands of Northwest Ethiopia: a health belief model perspective. *BMC Public Health* 2022; 22:989.
- [254] George, S., T.M. Joy, A. Kumar, K.N. Panicker, L.S. George, M. Raj, K. Leelamoni, and P. Nair. Prevalence of neglected tropical diseases (leishmaniasis and lymphatic filariasis) and malaria among a migrant labour settlement in Kerala, India. *Journal of Immigrant and Minority Health* 2019; 21, pp. 563-569.
- [255] Sampaio, M.J.A.dQ., N.V. Cavalcanti, J.G.B. Alves, M.J.C.F. Filho, and J.B. Correia. Risk factors for death in children with visceral leishmaniasis. *PLoS Neg. Trop. Dis.* 2010; 4(11): e877.
- [256] Coura-Vital, W., V.E.M. de Araújo, I.A. Reis, F.F. Amancio, A.B. Reis, and M. Carneiro. Prognostic factors and scoring system for death from visceral leishmaniasis: An historical cohort study in Brazil. *PLoS Neg. Trop. Dis.* 2014; 8(12): e3374.
- [257] Kämink, S.S., S.M. Collin, T. Harrison, F. Gatluak, A.W. Mullahzada, and K. Ritmeijer. A clinical severity scoring system for visceral leishmaniasis in immunocompetent patients in South Sudan. *PLoS Neg. Trop. Dis.* 2017; 11(10): e0005921.

- [258] Müller, I., A. Hailu, B-S. Choi, T. Abebe, J.M. Fuentes, M. Munder, M. Modolell, and P. Kropf. Age-related alteration of arginase activity impacts on severity of leishmaniasis. *PLoS Neg. Trop. Dis.* 2008; 2(5): e235.
- [259] Martens, P., and L. Hall. Malaria on the move: human population movement and malaria transmission. *Emerg. Infect. Dis.* 2000; 6:2, pp. 103-109.
- [260] Sevilla-Casas, E. Human mobility and malaria risk in the Naya river basin of Colombia. *Soc. Sci. Med.* 1993; 37:9, pp. 1155-1167.
- [261] Rodríguez-rodríguez, D., M. Katusele, A. Auwun, M. Marem, L.J. Robinson, M. Laman, M.W. Hetzel, and J. Pulford. Human behaviour, livelihood, and malaria transmission in two sites of Papua new Guinea. *The Journal of Infectious Diseases* 2021; 223:S2, S171-186.
- [262] Rosecrans, K., G. Cruz-Martin, A. King, E. Dumonteil. Opportunities for improved Chagas disease vector control based on knowledge, attitudes and practices of communities in the Yucatan Peninsula, Mexico. *PLoS Neg. Trop. Dis.* 2014; 8(3): e2763.
- [263] Bayer, A.M., G.C. Hunter, R.H. Gilman, J.G.C. del Carpio, C. Naquira, C. Bern, and M.Z. Levy. Chagas disease, migration and community settlement patterns in Arequipa, Peru. *PLoS neg. Trop Dis.* 2009; 3(12): e567.
- [264] Lynn, M.K., B.H. Bossak, P.A. Sandifer, A. Watson, and M.S. Nolan. Contemporary autochthonous human Chagas disease in the USA. *Acta Tropica* 2020; 205:105361.
- [265] El Sayed, B.B., D.E. Arnot, M.M. Mukhtar, O.Z. Baraka, A.A. Dafalla, D.E.A. Elnaiem, A.H.D. Nugud. A study of the urban malaria transmission problem in Khartoum. *Acta Tropica* 2000; 75:2, pp. 163-171.
- [266] Bautista-Gomez, M.M., J. Doerfler, M. dM. Castro. Barriers to cutaneous leishmaniasis care faced by indigenous communities of rural areas in Colombia: A qualitative study. *BMC Infectious Diseases* 2022; 22:302.
- [267] Ramdas, S. Cruel disease, cruel medicine: Self-treatment of cutaneous leishmaniasis with harmful chemical substances in Suriname. *Soc. Sci. Med.* 2012; 75:6, pp. 1097-1105.
- [268] Pires, M., B. Wright, P.M. Kaye, V. da Conceição, and R.C. Churchill. The impact of leishmaniasis on mental health and psychosocial well-being: A systematic review. *PLoS ONE* 2019; 14(10): e0223313.
- [269] Chahed, M.K., H. Bellali, S.B. Jemaa, and T. Bellaj. Psychological and psychosocial consequences of zoonotic cutaneous leishmaniasis among women in Tunisia: Preliminary findings from an exploratory study. *PLoS Neg. Trop. Dis.* 2016; 10(10): e0005090.
- [270] Alidosti, M., Z. Heidari, H. Shahnazi, and F. Zamani-Alavijeh. Behaviours and perceptions related to cutaneous leishmaniasis in endemic areas of the world: A review. *Acta Tropica* 2021; 223:106090.

- [271] Bauch, J.A., J.J. Gu, M. Msellem, A. Mårtensson, A.S. Ali, R. Gosling, and K.A. Baltzell. Perception of malaria risk in a setting of reduced malaria transmission: A qualitative study in Zanzibar. *Malaria Journal* 2013; 12:75.
- [272] Mboera, L.E.G., E.H. Shayo, K.P. Senkoro, S.F. Rumisha, M.R.S. Mlozi, and B.K. Mayala. Knowledge, perceptions and practices of farming communities on linkages between malaria and agriculture in Mvomero District, Tanzania. *Acta Tropica* 2010; 113:2, pp. 139-144.
- [273] Atieli, H.E., G. Zhou, Y. Afrane, M-C. Lee, I. Mwanzo, A.K. Githeko, and G. Yan. Insecticide-treated net (ITN) ownership, usage, and malaria transmission in the highlands of western Kenya. *Parasites and Vectors* 2011; 4:113.
- [274] Spjeldnæs, A.O., A.Y. Kitua, and B. Blomberg. Education and knowledge helps combating malaria, but not *degedege*: A cross-sectional study in Rufiji, Tanzania. *Malaria Journal* 2014; 13:200.
- [275] Dike, N., O. Onwujekwe, J. Ojukwu, A. Ikeme, B. Uzochukwu, and E. Shu. Influence of education and knowledge on perceptions and practices to control malaria in Southeast Nigeria. *Soc. Sci. Med.* 2006; 63:1, pp. 103-106.
- [276] Salm, A., and J. Gertsch. Cultural perception of triatomine bugs and Chagas disease in Bolivia: a cross-sectional field study. *Parasites and Vectors* 2019; 12:291.
- [277] Parisi, S., M. Navarro, J.D.D. Plessis, J.P. Shock, B.A. Michel, M.L. Espinoza, C. Terán, N.A.C. Tapia, K. Oltmanns, A.B. Mora, C.S. Irala, A.A.R. Rojas, G. Rubilar, T. Zoller, M. Pritsch. “We have already heard that the treatment doesn’t do anything, so why should we take it?”: A mixed method perspective on Chagas disease knowledge, attitudes, prevention and treatment behaviour in the Bolivian Chaco. *PLoS Neg. Trop. Dis.* 2020; 14(10): e0008752.
- [278] Faraj, C., E.B. Adlaoui, S. Ouahabi, M. Elkohli, M. Elrhazi, L. Laqraa, and B. Ameer. Field evaluation of alphacypermethrin in indoor residual spraying for leishmaniasis control in an endemic area, northern Morocco. *Parasites and Vectors* 2013; 6:354.
- [279] Coleman, M., G.M. Foster, R. Deb, R.P. Singh, H.M. Ismail, P. Shivam, A.K. Ghosh, S. Dunkley, V. Kumar, M. Coleman, J. Hemingway, M.J.I. Paine, and P. Das. DDT-based indoor residual spraying suboptimal for visceral leishmaniasis elimination in India. *PNAS* 2015; 112(28): pp.8573-8578.
- [280] Mandal, R., V. Kumar, S. Kesari, and P. Das. Assessing the combined effects of household type and insecticide effectiveness for kala-azar vector control using indoor residual spraying: a case study from north Bihar, India. *Parasites and Vector* 2019; 12:409.
- [281] Gidwani, K., A. Picado, S. Rijal, S.P. Singh, L. Roy, V. Volfova, E.W. Andersen, S. Uranw, B. Ostyn, M. Sudarshan, J. Chakravarty, P. Volf, S. Sundar, M. Boelaert, and M.E. Rogers. Serological markers of sand fly exposure to evaluate insecticidal nets

- against visceral leishmaniasis in India and Nepal: a cluster-randomised trial. *PLoS Neg. Trop. Dis.* 2011; 5(9):e1296
- [282] Faraj, C., J.Yukich, E.B. Adlaoui, R. Wahabi, A.P. Mnzava, M. Kaddaf, A.L.E. Idrissi, B. Ameer, and I. Kleinschmidt. Effectiveness and cost of insecticide-treated bednets and indoor residual spraying for the control of cutaneous leishmaniasis: a cluster-randomised control trial in Morocco. *Am. J. Trop. Med. Hyg.* 2016; 94(3): pp.679-685.
- [283] Coulibaly, C.A., B. Traore, A. Dicko, S. Samake, I. Sissoko, J.M. Anderson, J. Valenzuela, S.F. Traore, O. Faye, S. Kamhawi, F. Oliveira, and S. Doumbia. Impact of insecticide-treated bednets and indoor residual spraying in controlling populations of *Phlebotomus duboscqi*, the vector of *Leishmania major* in Central Mali. *Parasites and Vectors* 2018; 11:345.
- [284] Hassan, M.M., S.O. Widaa, O. M. Osman, M.S.M. Numiary, M.A. Ibrahim, and H.M. Abushama. Insecticide resistance in the sand fly, *Phlebotomus papatasi* from Khartoum State, Sudan. *Parasites and Vectors* 2012; 5:46.
- [285] Karakus, M., B. Gocmen, and Y. Özbel. Insecticide susceptibility status of wild-caught sand fly populations collected from two leishmaniasis endemic areas in western Turkey. *J. Arthropod-Borne Dis.* 2017; 11(1): pp. 86-94.
- [286] Saurabh, S., and R.K. Yadav. Recalibrating of insecticide-measuring containers could prevent underdosing of insecticide during indoor residual spraying for visceral leishmaniasis (kala-azar) elimination in India. *Trans. Roy. Soc. Trop. Med. and Hyg.* 2020; 114(7): pp.549-552.
- [287] Gonçalves, R., C.F. de Souza, R.B. Rontani, A. Pereira, K.B. Farnes, E.E. Gorsich, R.A. Silva, R.P. Brazil, J.G.C. Hamilton, and O. Courtenay. Community deployment of a synthetic pheromone of the sand fly *Lutzomyia longipalpis* co-located with insecticide reduces vector abundance in treated and neighbouring untreated houses: implications for control of *Leishmania infantum*. *PLoS Neg. Trop. Dis.* 2021; 15(2):e0009080
- [288] Rees, P.H., and P.A. Kager. Visceral leishmaniasis and post-Kala-azar dermal leishmaniasis. In: W. Peters, R. Killick-Kendrick, editors. *The leishmaniasis in biology and medicine. Vol II.* London: Academic Press, 1987. pp. 584-615.
- [289] Bousema, T., J.T. Griffin, R.W. Sauerwein, D.L. Smith, T.S. Churcher, W. Takken, A. Ghani, C. Drakeley, and R. Gosling. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med.* 2012; 9(1): e1001165.
- [290] Calderon-Anyosa, R., C. Galvez-Petzoldt, P.J. Garcia, and C.P. Carcamo. Housing characteristics and leishmaniasis: A systematic review. *Am. J. Trop. Med. Hyg.* 2018; 99:6, pp. 1547-1554.
- [291] López-Perea, N., L. Sordo, E. Gadisa, I. Cruz, T. Hailu, J. Moreno, A. Aseffa, C. Cañavate, and E. Custodio. Knowledge, attitudes and practices related to visceral

- leishmaniasis in rural communities of Amhara State: A longitudinal study in North-west Ethiopia. *PLoS Neg. Trop. Dis.* 2014; 8(4): e2799.
- [292] Patterson, N.M., B.R. Bates, A.E. Chadwick, C. Nieto-Sanchez, and M.J. Grijalva. Using the health belief model to identify communication opportunities to prevent Chagas disease in Southern Ecuador. *PLoS Neg. Trop. Dis.* 2018; 12(9): e0006841.
- [293] West, H.M., C.E. Milliren, O. Vragovic, J.R. Köhler, and C. Yarrington. Perceived barriers to Chagas disease screening among a diverse group of prenatal care providers. *PLoS ONE* 2021; 16(2): e0246783.
- [294] Sousa-Paula, L.C.d., L.G. da Silva, K.G.d.S. Sales, and F. Dantas-Torres. Failure of the dog culling strategy in controlling human visceral leishmaniasis in Brazil: A screening coverage issue? *PLoS Neg. Trop. Dis.* 2019; 13(6): e0007553.
- [295] Bashaye, S., N. Nombela, D. Argaw, A. Mulugeta, M. Herrero, J. Nieto, C. Chicharro, C. Cañavate, P. Aparicio, I.D. Vélez, J. Alvar, and C. Bern. Risk factors for visceral leishmaniasis in a new epidemic site in Amhara Region, Ethiopia. *Am. J. Trop. Med. Hyg.* 2009; 81(1): pp. 34-39.
- [296] Chappuis, F., S. Sundar, A. Hailu, H. Ghalib, S. Rijal, R.W. Peeling, J. Alvar, and M. Boelaert. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? *Nature Reviews Microbiology* 2007; 5: pp. 873-882.
- [297] Ives, A., C. Ronet, F. Prevel, G. Ruzzante, S. Fuertes-Marraco, F. Schutz, H. Zangger, M. Revaz-Breton, L.-F. Lye, S.M. Hickerson, S.M. Beverley, H. Acha-Orbea, P. Launois, N. Fasel, and S. Masina. *Leishmania* RNA virus controls the severity of mucocutaneous leishmaniasis. *Science* 2011; 331(6018): pp. 775-778.
- [298] Bailey, F., K. Mondragon-Shem, P. Hotez, J.A. Ruiz-Postigo, W. Al-Salem, A. Acosta-Serrano, and D.H. Molyneux. A new perspective on cutaneous leishmaniasis - Implications for global prevalence and burden of disease estimates. *PLoS Neg. Trop. Dis.* 2017; 11(8): e0005739.
- [299] Hasker, E., P. Malaviya, V.K. Scholar, P. de Koning, O.P. Singh, S. Kansal, K. Cloots, M. Boelaert, and S. Sundar. Post kala azar dermal leishmaniasis and leprosy prevalence and distribution in the Muzaffarpur health and demographic surveillance site. *PLoS Neg. Trop. Dis.* 2019; 13(10): e0007798.
- [300] Bezerra, J.M.T., V.E.M. de Araújo, D.S. Barbosa, F.R. Martins-Melo, G.L. Werneck, and M. Carneiro. Burden of leishmaniasis in Brazil and federated units, 1990-2016: Finds from Global Burden of Disease Study 2016. *PLoS Neg. Trop. Dis.* 2018; 12(9): e0006697.
- [301] Mohan, S., P. Revill, S. Malvoti, M. Malhame, M. Sculpher, and P.M. Kaye. Estimating the global demand curve for a leishmaniasis vaccine: A generalisable approach based on global burden of disease estimates. *PLoS Neg. Trop. Dis.* 2022; 16(6): e0010471.

- [302] Picado, A., M.L. Das, V. Kumar, S. Kesari, D.S. Dinesh, L. Roy, S. Rijal, P. Das, M. Rowland, S. Sundar, M. Coosemans, M. Boelaert, and C.R. Davies. Effect of village-wide use of long-lasting insecticidal nets on visceral leishmaniasis vectors in India and Nepal: A cluster randomised trial. *PLoS Neg. Trop. Dis.* 2010; 4(1): e587.
- [303] Cloots, K., S. Uranw, B. Ostyn, N.R. Bhattarai, E. Le Rutte, B. Khanal, A. Picado, F. Chappuis, E. Hasker, P. Karki, S. Rijal, and M. Boelaert. Impact of the visceral leishmaniasis elimination initiative on *Leishmania donovani* transmission in Nepal: a 10-year repeat survey. *The Lancet Global Health* 2020; 8(2): pp. e237-e243.
- [304] Fortunato, A.K., C.P. Glasser, J.A. Watson, Y. Lu, J. Rychtář, and D. Taylor. Mathematical modelling of the use of insecticide-treated nets for elimination of visceral leishmaniasis in Bihar, India. *R. Soc. Open Sci.* 2021; 8:201960.
- [305] de Arruda, R.M.F., D.T. Cardoso, R.G. Teixeira-Neto, D.S. Barbosa, R.K. Ferraz, M.H.F. Morais, V.S. Belo, and E.S. da Silva. Space-time analysis of the incidence of human visceral leishmaniasis (VL) and prevalence of canine VL in a municipality of southeastern Brazil: Identification of priority areas for surveillance and control. *Acta Tropica* 2019; 197: 105052.
- [306] Ibarra-Meneses, A.V., E. Carrillo, J. Nieto, C. Sánchez, S. Ortega, A. Estirado, P.L. Zamalloa, J. C. Sanz, L. García-Comas, M. Ordobás, and J. Moreno. Prevalence of asymptomatic *Leishmania* infection and associated risk factors, after an outbreak in south-western Madrid region, Spain, 2015. *Eurosurveillance* 2019; 24(22): pii=1800379.
- [307] Mahajan, R., S.I. Owen, S. Kumar, K. Pandey, S. Kazmi, V. Kumar, E.R. Adams, A. Harshana, and S. Burza. Prevalence and determinants of asymptomatic *Leishmania* infection in HIV-infected individuals living within visceral leishmaniasis endemic areas of Bihar, India. *PLoS Neg. Trop. Dis.* 2022; 16(8): e0010718.
- [308] Du, R., P.J. Hotez, W.S. Al-Salem, and A. Acosta-Serrano. Old world cutaneous leishmaniasis and refugee crises in the Middle East and North Africa. *PLoS Neg. Trop. Dis.* 2016; 10(5): e0004545.
- [309] Yohannes, M., Z. Abebe, and E. Boelee. Prevalence and environmental determinants of cutaneous leishmaniasis in rural communities in Tigray, northern Ethiopia. *PLoS Neg. Trop. Dis.* 2019; 13(9): e0007722.
- [310] Rubiano, L., N.D.E. Alexander, R.M. Castillo, A.J. Martínez, J.A.G. Luna, J.D. Arango, L. Vargas, P. Madriñán, L.-R. Hurtado, Y. Orobio, C.A. Rojas, H. del Corral, A. Navarro, N.G. Saravia, and E. Aronoff-Spencer. Adaptation and performance of a mobile application for early detection of cutaneous leishmaniasis. *PLoS Neg. Trop. Dis.* 2021; 15(2): e0008989.
- [311] Eid, D., M.S. Sebastian, A.-K. Hurtig, and I. Golicolea. Leishmaniasis patients' pilgrimage to access health care in rural Bolivia: a qualitative study using human rights to health approach. *BMC International Health and Human Rights* 2019; 19:12.

- [312] Bautista-Gomez, M.M., J. Doerfler, and M. del M. Castro. Barriers to cutaneous leishmaniasis care faced by indigenous communities of rural areas in Colombia: a qualitative study. *BMC Infectious Diseases* 2021; 22:302.
- [313] Singh, O.M., B. Singh, J. Chakravarty, and S. Sundar. Current challenges in treatment options for visceral leishmaniasis in India: a public health perspective. *Infect. Dis. Poverty* 2016; 5:19.
- [314] Doehl, J.S.P., H. Ashwin, N. Brown, A. Romano, S. Carmichael, J.W. Pitchford, and P.M. Kaye. Spatial point pattern analysis identifies mechanisms shaping the skin parasite landscape in *Leishmania donovani* infection. *Front. Immunol.* 2021; 12: 795554.
- [315] Li, Y., Z. Luo, Y. Hao, Y. Zhang, L. Yang, Z. Li, Z. Zhou, and S. Li. Epidemiological features and spatio-temporal clustering of visceral leishmaniasis in mainland China from 2019 to 2021. *Frontiers in Microbiology* 2022; 13: 959901.
- [316] de Oca-Aguilar, A.C.M., E.A. Rebollar-Telléz, E.I. Sosa-Bibiano, K.B. López-Avila, J.R. Torres-Castro, and E.N. Loría-Cervera. Effect of land use change on the Phlebotomine sand fly assemblages in an emergent focus of cutaneous leishmaniasis in Yucatan, Mexico. *Acta Tropica* 2022; 235:106628.
- [317] Molina, R., M. Jiménez, J. García-Martínez, J.V.S. Martín, E. Carillo, C. Sánchez, J. Moreno, F. Alves, and J. Alvar. Role of asymptomatic and symptomatic humans as reservoirs of visceral leishmaniasis in a Mediterranean context. *PLoS Neg. Trop. Dis.* 2020; 14(4): e0008253.
- [318] Melo-Merino, S.M., H. Reyes-Bonilla, and A. Lira-Noriega. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecological Modelling* 2020; 415: 108837.
- [319] Rodgers, M. de S.M., M.E. Bavia, E.O.L. Fonseca, B.O. Cova, M.N.N. Silva, D.D.M.T. Carneiro, L.L. Cardim, and J.B. Malone. Ecological niche models for sand fly species and predicted distribution of *Lutzomyia longipalpis* (Diptera: Psychodidae) and visceral leishmaniasis in Bahia state, Brazil. *Environmental Monitoring and Assessment* 2019; 191:331.
- [320] Guerra, F.M., S. Bolotin, G. Lim, J. Heffernan, S.L. Deeks, Y. Li, and N.S. Crowcroft. The basic reproduction number (R_0) of measles: a systematic review. *The Lancet Infectious Diseases* 2017; 17(12): pp. e420-e428.
- [321] Tesla, B., L.R. Demakovsky, H.S. Packiam, E.A. Mordecai, A.D. Rodríguez, M.H. Bonds, M.A. Brindley, and C.C. Murdock. Estimating the effects of variation in viremia on mosquito susceptibility, infectiousness, and R_0 of Zika in *Aedes aegypti*. *PLoS Neg. Trop. Dis.* 2018; 12(8): e0006733.
- [322] Ke, R., E. Romero-Severson, S. Sanche, and N. Hengartner. Estimating the reproductive number R_0 of SARS-CoV-2 in the United States and eight European countries and implications for vaccination. *Journal of Theoretical Biology* 2021; 517: 110621.

- [323] Branda, F., M. Pierini, and S. Mazzoli. Monkeypox: Early estimation of basic reproduction number R_0 in Europe. *Journal of Medical Virology* 2022; 95(1): e28270.
- [324] Britton, T. Epidemics in heterogeneous communities: estimation of R_0 and secure vaccination coverage. *Journal of the Royal Statistical Society: Series B (Methodological)* 2002; 63(4): pp. 705-715.
- [325] Barongo, M.B., K. Ståhl, B. Bett, R.P. Bishop, E.M. Fèvre, T. Aliro, E. Okoth, C. Masembe, D. Knobel, and A. Ssematimba. Estimating the basic reproductive number (R_0) for African Swine Fever Virus (ASFV) transmission between pig herds in Uganda. *PLoS ONE* 2015; 10(5): e0125842.
- [326] Anderson, R.M., and R.M. May. Vaccination and herd immunity to infectious diseases. *Nature* 1985; 318: pp.323-329.
- [327] Metcalf, C.J.E., J. Lessler, P. Klepac, F. Cutts, and B.T. Grenfell. Impact of birth rate, seasonality and transmission rate on minimum levels of coverage needed for rubella vaccination. *Epidemiology and Infection* 2012; 140(12): pp. 2290-2301.
- [328] Tildesley, M.J., and M.J. Keeling. Is R_0 a good predictor of final epidemic size: Foot-and-mouth disease in the UK. *Journal of Theoretical Biology* 2009; 258(4): pp. 623-629.
- [329] Becker, A.D., R.B. Birger, A. Teillant, P.A. Gastanaduy, G.S. Wallace, and B.T. Grenfell. Estimating enhanced prevaccination transmission hotspots in the context of cross-scale dynamics. *PNAS* 2016; 113(51): pp. 14595-14600.
- [330] Mukandavire, Z., S. Liao, J. Wang, H. Gaff, D.L. Smith, and J.G. Morris Jr. Estimating the reproductive numbers for the 2008-2009 cholera outbreaks in Zimbabwe. *PNAS* 2011; 108(21): pp. 8767-8772.
- [331] Nikbakht, R., M.R. Baneshi, A. Bahrapour, and A. Hosseinnataj. Comparison of methods to estimate basic reproduction number (R_0) of influenza, using Canada 2009 and 2017-18 A(H1N1) data. *J. Res. Med. Sci.* 2019; 24:67.
- [332] Ibarra-Menesis, A.V., A. Corbeil, V. Wagner, C. Onwuchekwa, and C. Fernandez-Prada. Identification of asymptomatic *Leishmania* infections: a scoping review. *Parasites and Vectors* 2022; 15:5.
- [333] Das, S., G. Matlashewski, G.S. Bhunia, S. Kesari, and P. Das. Asymptomatic *Leishmania* infections in northern India: a threat for the elimination program? *Trans. R. Soc. Trop. Med. Hyg.* 2014; 108(11): pp. 679-684.
- [334] Newman, M.E.J. The structure and function of complex networks. *SIAM Review* 2002; 45(2): pp. 167-256.
- [335] Weeden, K.A., and B. Cornwell. The small-world network of college classes: implications for epidemic spread on a university campus. *Sociological Science* 2020; 7: pp. 222-241.

- [336] Craft, M.E., E. Volz, C. Packer, and L.A. Meyers. Disease transmission in territorial populations: the small-world network of Serengeti lions. *J. R. Soc. Interface* 2010; 8(59): pp. 776-786.
- [337] Zhang, R. Global dynamic analysis of a model for vector-borne diseases on bipartite networks. *Physica A: Statistical Mechanics and its Applications* 2020; 545(1): 123813.
- [338] Nunn, C.L., A.Q. Vining, D. Chakraborty, M.H. Reiskind, and H.S. Young. Effects of host extinction and vector preferences on vector-borne disease risk in phylogenetically structured host-vector communities. *PLoS ONE* 2021; 16(8): e0256456.
- [339] Salines, M., M. Andraud, and N. Rose. Pig movements in France: Designing network models fitting the transmission route of pathogens. *PLoS ONE* 2017; 12(10): e0185858
- [340] Valdez, L.D., L.A. Braunstein, and S. Havlin. Epidemic spreading on modular networks: The fear to declare a pandemic. *Phys. Rev. E* 2020; 101(3): 032309.
- [341] Hébert-Dufresne, L., and B.M. Althouse. Complex dynamics of synergistic coinfections on realistically clustered networks. *PNAS* 2015; 112(33): 10551-10556.
- [342] Parousis-Orthodoxou, K.J., M.M. Stamos, and D.S. Vlachos. Global vaccination strategies in modular networks. *J. Phys.: Conf. Ser.* 2013; 410: 012079.
- [343] Silk, M.J., D.J. Hodgson, C. Rozins, D.P. Croft, R.J. Delahay, M. Boots, and R.A. McDonald. Integrating social behaviour, demography and disease dynamics in network models: applications to disease management in declining wildlife populations. *Phil. Trans. R. Soc. B.* 2019; 374: 20180211.
- [344] Diekmann, O., and J.A.P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases* Wiley 2001; Chichester, West Sussex, England.
- [345] Newman, M.E.J. Spread of epidemic diseases on networks *Physical Review E* 2002; 66:016128
- [346] Keeling, M.J. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B* 1999; 266(1421): pp. 859-867.
- [347] Eames, K.T.D., and M.J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *PNAS* 2002; 99(20): pp. 13330-13335.
- [348] Eames, K.T.D., and M.J. Keeling. Contact tracing and disease control. *Proc. R. Soc. Lond. B* 2003; 270(1533): pp. 2565-2571.
- [349] Baca er, N. Approximation of the basic reproduction number R_0 for vector-borne diseases with a periodic vector population. *Bulletin of Mathematical Biology* 2007; 69: pp. 1067-1091.
- [350] Lange, A., and N.M. Ferguson. Antigenic diversity, transmission mechanisms, and the evolution of pathogens. *PLoS Comput. Biol.* 2009; 5(10): e1000536.
- [351] Husein, I., H. Mawengkang, S. Suwilo, and Mardinarsih. Modeling the transmission of infectious disease in a dynamic network. *J. Phys.: Conf. Ser.* 2019; 1255: 012052.

- [352] Mancy, R., M. Rajeev, A. Lugelo, K. Brunker, S. Cleaveland, E.A. Ferguson, K. Hotopp, R. Kazwala, M. Magoto, K. Rysava, D.T. Haydon, and K. Hampson. Rabies shows how scale of transmission can enable acute infections to persist at low prevalence. *Science* 2022; 376(6592): pp. 512-516.
- [353] Chowell, G., C. Viboud, L. Simonsen, and S.M. Moghadas. Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *J. R. Soc. Interface* 2016; 13: 20160659.
- [354] Kiss, I.V., D.M. Green, and R.R. Kao. Infectious disease control using contact tracing in random and scale-free networks. *J. R. Soc. Interface* 2006; 3(6): pp. 55-62.
- [355] Britton, T., M.K. Nordvik, and F. Liljeros. Modelling sexually transmitted infections: The effect of partnership activity and number of partners on R_0 . *Theoretical Population Biology* 2007; 72(3): pp. 389-399.
- [356] Li, J., D. Blakeley, and R.J. Smith. The failure of R_0 . *Computational and Mathematic Methods in Medicine* 2011; 527610.
- [357] Eguíluz, V.M., and K. Klemm. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.* 2002; 89: 108701.
- [358] Olinky, R., and L. Stone. Unexpected epidemic thresholds in heterogeneous networks: The role of disease transmission. *Phys. Rev. E* 2004; 70: 030902.
- [359] Volz, E., and L.A. Meyers. Epidemic thresholds in dynamic contact networks. *J. R. Soc. Interface* 2008; 6: pp. 233-241.
- [360] Hernández, D.G., and S. Risau-Gusman. Epidemic thresholds for bipartite networks. *Phys. Rev. E* 2013; 88: 052801.
- [361] Bisanzio, D., L. Bertolotti, L. Tomassone, G. Amore, C. Ragagli, A. Mannelli, M. Giacobini, and P. Provero. Modeling the spread of vector-borne diseases on bipartite networks. *PLoS ONE* 2010; 5(11): e13796.
- [362] Johnstone-Robertson, S.P., M.A. Diuk-Wasser, and S.A. Davis. Incorporating tick feeding behaviour into R_0 for tick-borne pathogens. *Theor. Popul. Biol.* 2020; 131: pp. 25-37.
- [363] Boguñá, M., R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Phys. Rev. Lett.* 2003; 90: 028701.
- [364] Castellano, C., and R. Pastor-Satorras. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* 2010; 105: 218701.
- [365] Wu, Q., and F. Zhang. Threshold conditions for SIS epidemic models on edge-weighted networks. *Physica A: Statistical Mechanics and its Applications* 2016; 453: pp. 77-83.
- [366] Rizi, A.K., A. Faqeeh, A. Badie-Modiri, and M. Kivela. Epidemic spreading and digital contact tracing: Effects of heterogeneous mixing and quarantine failures. *Phys. Rev. E* 2022; 105: 044313.

- [367] Fofana, A.M., and A. Hurford. Mechanistic movement models to understand epidemic spread. *Phil. Trans. R. Soc. B* 2017; 372: 20160086.
- [368] Leung, K.Y., and M. Kretzschmar. Concurrency can drive an HIV epidemic by moving R_0 across the epidemic threshold. *AIDS* 2015; 29(9): pp. 1097-1103.
- [369] Sahneh, F.D., F.N. Chowdhury, and C.M. Scoglio. On the existence of a threshold for preventative behavioural responses to suppress epidemic spreading. *Scientific Reports* 2012; 2: 632.
- [370] Gradoń, K.T., J.A. Hołyst, W.R. Moy, J. Sienkiewicz, and K. Suchecki. Countering misinformation: A multidisciplinary approach. *Big Data and Society* 2021; 8(1): pp. 1-14.
- [371] Sontag, A., T. Rogers, and C.A. Yates. Misinformation can prevent the suppression of epidemics. *J. R. Soc. Interface* 2022; 19: 20210668.
- [372] Wang, W., Y. Ma, T. Wu, Y. Dai, X. Chen, and L.A. Braunstein. Containing misinformation spreading in temporal social networks. *Chaos* 2019; 29: 123131.
- [373] Tambuscio, M., D.F.M. Oliveira, G.L. Ciampaglia, and G. Ruffo. Network segregation in a model of misinformation and fact-checking. *Journal of Computational Social Science* 2018; 1: pp. 261-275.
- [374] Sulaimany, S., and A. Mafakheri. Reversibility of link prediction and its application to epidemic mitigation. *Scientific Reports* 2022; 12: 20880.
- [375] Hill, A.N., and I.M. Longini Jr. The critical vaccination fraction for heterogeneous epidemic models. *Mathematical Biosciences* 2003; 181(1): pp. 85-106.
- [376] Oryan, A., and M. Akbari. Worldwide risk factors in leishmaniasis. *Asian Pacific Journal of Tropical Medicine* 2016; 9(10): pp. 925-932.
- [377] Mondal, D., M.M. Huda, M.K. Karmoker, D. Ghosh, G. Matlashewski, S.G. Nabi, and A. Kroeger. Reducing visceral leishmaniasis by insecticide impregnation of bed-nets, Bangladesh. *Emerg. Infect. Dis.* 2013; 19(7): pp. 1131-1134.
- [378] Montenegro-Quiñonez, C.A., C. Buhler, O. Horstick, S. Runge-Ranzinger, K.M. Rahman. Efficacy and community-effectiveness of insecticide-treated nets for the control of visceral leishmaniasis: A systematic review. *PLoS Neg. Trop. Dis.* 2022; 16(3): e0010196.
- [379] Mondal, D., M.L. Das, V. Kumar, M.M. Huda, P. Das, D. Ghosh, J. Priyanka, G. Matlashewski, A. Kroeger, A. Uphill-Brown, and R. Chowdhury. Efficacy, safety and cost of insecticide-treated wall lining, insecticide treated bed nets and indoor wall wash with lime for visceral leishmaniasis vector control in the Indian Sub-Continent: A multi-country cluster randomized controlled trial. *PLoS Neg. Trop. Dis.* 2016; 10(8): e0004932.

- [380] Hassaballa, I.B., B. Torto, C.L. Sole, and D.P. Tchouassi. Exploring the influence of different habitats and their volatile chemistry in modulating sand fly population structure in a leishmaniasis endemic foci, Kenya. *PLoS Neg. Trop. Dis.* 2021; 15(2): e0009062.
- [381] Baca er, N., and S. Guernaoui. The epidemic threshold of vector-borne diseases with seasonality. *Journal of Mathematical Biology* 2006; 53: pp. 421-436.
- [382] Rodríguez, E.-M., F. Díaz, and M.-V. Pérez. Spatio-temporal clustering of American cutaneous leishmaniasis in a rural municipality of Venezuela. *Epidemics* 2013; 5(1): pp. 11-19.
- [383] Schriefer, A., L.H. Guimarães, P.R.L. Machado, M. Lessa, H.A. Lesso, E. Lago, G. Ritt, A. Góes-Neto, A.L.F. Schriefer, L.E. Riley, and E.M. Carvalho. Geographic clustering of leishmaniasis in northeastern Brazil. *Emerg. Infect. Dis.* 2009; 15(6): pp. 871-876.
- [384] Alten, B., C. Maia, M.O. Afonso, L. Campino, M. Jiménez, E. González, R. Molina, A.L. Bañuls, J. Prudhomme, B. Vergnes, C. Toty, C. Cassan, N. Rahola, M. Thierry, D. Sereno, G. Bongiorno, R. Bianchi, C. Khoury, N. Tsigotakis, E. Dokianakis, M. Antoniou, V. Christodoulou, A. Mazeris, M. Karakus, Y. Ozbel, S.K. Arserim, O.E. Kasap, F. Gunay, G. Oguz, S. Kaynas, N. Tsertsvadze, E. Giorgobiani, M. Gramiccia, P. Volf, and L. Gradoni. Seasonal dynamics of Phlebotomine sand fly species proven vectors of Mediterranean leishmaniasis caused by *Leishmania infantum*. *PLoS Neg. Trop. Dis.* 2016; 10(2): e0004458.
- [385] Brooker, S., N. Mohammed, K. Adil, S. Agha, R. Reithinger, M. Rowland, I. Ali, and J. Kolaczinski. Leishmaniasis in refugee and local Pakistani populations. *Emerg. Infect. Dis.* 2004; 10(9): pp. 1681-1684.
- [386] Magalhães, A.R., C.T. Codeço, J.-C. Svenning, L.E. Escobar, P. van de Vuurst, and T. Gonçalves-Souza. Neglected tropical diseases risk correlates with poverty and early ecosystem destruction. *Infectious Diseases of Poverty* 2023; 12: 32.
- [387] Kuhls, K., L. Keilonat, S. Ochsenreither, M. Schaar, C. Schweynoch, W. Presber, and G. Schonian. Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis. *Microbes and Infection* 2007; 9(3): pp. 334-343.
- [388] Bhattarai, N.R., M.L. Das, S. Rijal, G. van der Auwera, A. Picado, B. Khanal, L. Roy, N. Speybroeck, D. Berkvens, C.R. Davies, M. Coosemans, M. Boelaert, and J.-C. Dujardin. Natural infection of *Phlebotomus argentipes* with *Leishmania* and other trypanosomatids in a visceral leishmaniasis endemic region of Nepal. *Trans. R. Soc. Trop. Med. Hyg.* 2009; 103(11): pp. 1087-1092.
- [389] Akuffo, R., M. Wilson, B. Sarfo, N. Attram, M.-T. Mosore, C. Yeboah, I. Cruz, J.-A. Ruiz-Postigo, D. Boakye, J. Moreno, and F. Anto. Prevalence of *Leishmania* infection in three communities of Oit Region, Ghana. *PLoS Neg. Trop. Dis.* 2021; 15(5): e0009413.

- [390] Singh, R.P., A. Picado, S. Alam, E. Hasker, S.P. Singh, B. Ostyn, F. Chappuis, S. Sundar, and M. Boelaert. Post-kala-azar dermal leishmaniasis in visceral leishmaniasis-endemic communities in Bihar, India. *Tropical Medicine and International Health* 2012; 17(11): pp. 1345-1348.
- [391] Abass, E., C. Kang, F. Martinkovic, S.J. Semião-Santos, S. Sundar, P. Walden, R. Piarroux, A. el Harith, M. Lohoff, and U. Steinhoff. Heterogeneity of *Leishmania donovani* parasites complicates diagnosis of visceral leishmaniasis: Comparison of different serological tests in three endemic regions. *PLoS ONE* 2015; 10(3): 116408.
- [392] Herrador, Z., A. Gherasim, B.C. Jimenez, M. Granados, J.V.S. Martín, and P. Aparicio. Epidemiological changes in leishmaniasis in Spain according to hospitalization-based records, 1997-2011: Raising awareness towards leishmaniasis in non-HIV patients. *PLoS Neg. Trop. Dis.* 2015; 9(3): e0003594.
- [393] Arce, A., A. Estirado, M. Ordobas, S. Sevilla, N. García, L. Moratilla, S. de la Fuente, A.M. Martínez, A.M. Pérez, E. Aránguez, A. Iriso, O. Sevillano, J. Bernal, and F. Vilas. Re-emergence of leishmaniasis in Spain: community outbreak in Madrid, Spain, 2009 to 2012. *Eurosurveillance* 2013; 18(30): 20546.
- [394] Picado, A., B. Ostyn, S. Rijal, S. Sundar, S.P. Singh, F. Chappuis, M.L. Das, B. Khanal, K. Gidwani, E. Hasker, J.-C. Dujardin, V. Vanlerberghe, J. Menten, M. Coosemans, and M. Boelaert. Long-lasting insecticidal nets to prevent visceral leishmaniasis in the Indian Subcontinent; Methodological lessons learned from a cluster randomised controlled trial. *PLoS Neg. Trop. Dis.* 2015; 9(4): e0003597.
- [395] Singh, O.P., and S. Sundar. Visceral leishmaniasis elimination in India: Progress and the road ahead. *Expert Review of Anti-infective Therapy* 2022; 20(11): pp. 1381-1388.
- [396] Sosa-Bibiano, E.I., L.A. Sánchez-Martínez, K.B. López-Ávila, J.B. Chablé-Santos, J.R. Torres-Castro, E.A. Fernández-Figueroa, C. Rangel-Escareño, E.N. Loría-Cervera. *Leishmania (Leishmania) mexicana* infection in wild rodents in from an emergent focus of cutaneous leishmaniasis in Yucatan, Mexico. *Journal of Tropical Medicine* 2022; 8392005.
- [397] Kushwaha, A.K., B.M. Scorza, O.P. Singh, E. Rowton, P. Lawyer, S. Sundar, and C.A. Petersen. Domestic mammals as reservoirs for *Leishmania donovani* on the Indian Subcontinent: Possibility and consequences on elimination. *Transboundary and Emerging Diseases* 2022; 69(2): pp. 268-277.
- [398] Michel, G., C. Pomares, B. Ferrua, and P. Marty. Importance of worldwide asymptomatic carriers of *Leishmania infantum (L. chagasi)* in human. *Acta Tropica* 2011; 119(2-3): pp. 69-75.
- [399] Mannan, S.B., H. Elhadad, T.T.H. Loc, M. Sadik, M.Y.F. Mohamed, N.H. Nam, N.D. Thuong, B.-L. Hoang-Trong, N.T.M. Duc, A.N. Hoang, K.M. Elhusseiny, L.H.N. Minh, T.T.H. Quynh, T.L.B. Nghia, Y.M. Nhu, T.M. Tieu, K. Hirayama, N.T. Huy,

- and S. Hamano. Prevalence and associated factors of asymptomatic leishmaniasis: a systematic review and meta-analysis. *Parasitology International* 2021; 81: 102229.
- [400] González, E., M. Jiménez, S. Hernández, I. Martín-Martín, and R. Molina. Phlebotomine sand fly survey in the focus of leishmaniasis in Madrid, Spain, 2012-2014: seasonal dynamics, *Leishmania infantum* infection rates, and blood meal preferences. *Parasites and Vectors* 2017; 10: 368.
- [401] Pastor-Satorras, R., and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* 2001; 63: 066117.
- [402] Billings, L., W.M. Spears, and I.B. Schwartz. A unified prediction of computer virus spread in connected networks. *Physics Letters A* 2002; 297(3-4): pp. 261-266.
- [403] Mishra, B.K., and N. Keshri. Mathematical model on the transmission of worms in wireless sensor network. *Applied Mathematical Modelling* 2013; 37(6): pp. 4103-4111.
- [404] Wang, Z., X. Rui, G. Yuan, J. Cui, and T. Hadzibeganovic. Endemic information-contagion outbreaks in complex networks with potential spreaders based recurrent-state transmission dynamics. *Physica A: Statistical Mechanics and its Applications* 2021; 573(1): 125907.
- [405] Mei, W., S. Mohagheghi, S. Zampieri, and F. Bullo. On the dynamics of deterministic epidemic propagation of networks. *Annual Reviews in Control* 2017; 44: pp. 116-128.
- [406] Takaguchi, T., and R. Lambiotte. Sufficient conditions of endemic threshold on metapopulation networks. *Journal of Theoretical Biology* 2015; 380: pp. 134-143.
- [407] Wilkinson, R.R., and K.J. Sharkey. An exact relationship between invasion probability and endemic prevalence for Markovian SIS dynamics on networks. *PLoS ONE* 2013; 8(7): e69028.
- [408] Dottori, M., and G. Fabricius. SIR model on a dynamical network and the endemic state of an infectious disease. *Physica A: Statistical Mechanics and its Applications* 2015; 434: pp. 25-35.
- [409] Althouse, B.M., J. Lessler, A.A. Sall, M. Diallo, K.A. Hanley, D.M. Watts, S.C. Weaver, and D.A.T. Cummings. Synchrony of sylvatic dengue isolations: A multi-host, multi-vector SIR model of dengue virus transmission in Senegal. *PLoS Neg. Trop. Dis.* 2012; 6(11): e1928.
- [410] Schaubert, E.M., and R.S. Ostfeld. Modeling the effects of reservoir competence decay and demographic turnover in Lyme disease ecology. *Ecological Applications* 2002; 12(4): pp. 1142-1162.
- [411] Rocha, L.E.C., and V.D. Blondel. Bursts of vertex activation and epidemics in evolving networks. *PLoS Comput. Biol.* 2013; 9(3): e1002974.
- [412] Ali, M., A.R. Nelson, A.L. Lopez, and D.A. Sack. Updated global burden of cholera in endemic countries. *PLoS Neg. Trop. Dis.* 2015; 9(6): e0003832.

- [413] Kirk, M.D., S.M. Pires, R.E. Black, M. Caipo, J.A. Crump, B. Devleeschauwer, D. Döpfer, A. Fazil, C.L. Fischer-Walker, T. Hald, A.J. Hall, K.H. Keddy, R.J. Lake, C.F. Lanata, P.R. Torgerson, A.H. Havelaar, and F.J. Angulo. World Health Organisation estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: A data synthesis. *PLoS Medicine* 2015; 12(12): e1001940.
- [414] Gebremeskel, A.A., and H.E. Krogstad. Mathematical modelling of Malaria transmission. *American Journal of Applied Mathematics* 2015; 3(2): pp. 36-46.
- [415] Pradhan, M.M., and P.K. Meherda. Malaria elimination drive in Odisha: Hope for halting the transmission. *J. Vector Borne Dis.* 2019; 56(1): pp. 53-53.
- [416] Tian, N., J.-X. Zheng, Z.-Y. Guo, L.-H. Li, S. Xia, S. Lv, and X.-N. Zhou. Dengue incidence trends and its burden in major endemic regions from 1990 to 2019. *Trop. Med. Infect. Dis.* 2022; 7(8): 180.
- [417] Ramaiah, K.D., and E.A. Ottesen. Progress and impact of 13 years of the global programme to eliminate lymphatic filariasis on reducing the burden of filarial disease. *PLoS Neg. Trop. Dis.* 8(11): e3319.
- [418] Lee, B.Y., K.M. Bacon, M.E. Bottazzi, and P.J. Hotez. Global economic burden of Chagas disease: A computation simulation model. *The Lancet Infectious Diseases* 2013; 13(4): pp. 342-348.
- [419] Nakagawa, J., J.P. Ehrenberg, J. Nealon, T. Fürst, P. Aratchige, G. Gonzales, C. Chanthavisouk, L.M. Hernandez, T. Fengthong, J. Utzinger, P. Steinmann. Towards effective prevention and control of helminth neglected tropical diseases in the Western Pacific Region through multi-disease and multi-sectoral interventions. *Acta Tropica* 2015; 141(B): pp. 407-418.
- [420] Herricks, J.R., P.J. Hotez, V. Wanga, L.E. Coffeng, J.A. Haagsma, M.-G. Basáñez, G. Buckle, C.M. Budke, H. Carabin, E.M. Fèvre, T. Fürst, Y.A. Halasa, C.H. King, M.E. Murdoch, K.D. Ramaiah, D.S. Shepard, W.A. Stolk, E.A. Undurraga, J.D. Stanaway, M. Naghavi, and C.J.L. Murray. The global burden of disease study 2013: What does it mean for NTDs? *PLoS Neg. Trop. Dis.* 2017; 11(8): e0005424.
- [421] Barogui, Y.T., G. Diez, E. Anagonou, R.C. Johnson, I.C. Gomido, H. Amoukpo, Z.S. Bachirou, J.G. Houezo, R. Saizonou, G.E. Sopoh, and K. Asiedu. Integrated approach in the control and management of skin neglected tropical diseases in Lalo, Benin. *PLoS Neg. Trop. Dis.* 2018; 12(6): e0006584.
- [422] Hürlimann, E., N. Schur, K. Boutzika, A.-S. Stensgaard, M.L. de Himpsl, K. Ziegelbauer, N. Laizer, L. Camenzind, A.D. Pasquale, U.F. Ekpo, C. Simoonga, G. Mushinge, C.F.L. Saarnak, J. Utzinger, T.K. Kristensen, and P. Vounatsou. Toward an open-access global database for mapping, control and surveillance of neglected tropical diseases. *PLoS Neg. Trop. Dis.* 2011; 5(12): e1404.
- [423] Gass, K. Time for a diagnostic sea-change: Rethinking neglected tropical disease diagnostics to achieve elimination. *PLoS Neg. Trop. Dis.* 2020; 14(12): e0008933.

- [424] Bharadwaj, M., M. Bengtson, M. Golverdingen, L. Waling, and C. Dekker. Diagnosing point-of-care diagnostics for neglected tropical diseases. *PLoS Neg. Trop. Dis.* 2021; 15(6): e0009405.
- [425] Ciciriello, A.M., J.K. Fairley, E. Cooke, P.M. Emerson, P.J. Hooper, B. Bolton, G. LaCon, D.G. Addiss. Safety of integrated preventative chemotherapy for neglected tropical diseases. *PLoS Neg. Trop. Dis.* 2022; 16(9): e0010700.
- [426] Yang, G.-J., L. Liu, H.-R. Zhu, S.M. Griffiths, M. Tanner, R. Bergquist, J. Utzinger, and X.-N. Zhou. China's sustained drive to eliminate neglected tropical diseases. *The Lancet Infectious Diseases* 2014; 14(9): pp. 881-892.
- [427] Wamai, R.G., J. Kahn, J. McGloin, and G. Ziaggi. Visceral leishmaniasis: A global overview. *J. Glob. Health Sci.* 2020; 2:1, e3.
- [428] Lima, I.D., A.L.M. Lima, C. de O. Mendes-Aguiar, J.F.V. Coutinho, M.E. Wilson, R.D. Pearson, J.W. Queiroz, and S.M.B. Jeronimo. Changing demographics of visceral leishmaniasis in north-east Brazil: Lessons for the future. *PLoS Neg. Trop. Dis.* 2018; 12(3):e0006164.
- [429] Jervis, S., L.A.C. Chapman, S. Dwivedi, M. Karthick, A. Das, E.A. Le Rutte, O. Courtenay, G.F. Medley, I. Banerjee, T. Mahapatra, I. Chaudhury, S. Srikantiah, and T.D. Hollingsworth. Variations in visceral leishmaniasis burden, mortality and the pathway to care in Bihar, India. *Parasites and Vectors* 2017; 10:601.
- [430] Sunyoto, T., K. Verdonck, S. el Safi, J. Potet, A. Picado, and M. Boelaert. Uncharted territory of the epidemiological burden of cutaneous leishmaniasis in sub-Saharan Africa - A systematic review. *PLoS Neg. Trop. Dis.* 2018; 12(10):e0006914.
- [431] Le Rutte, E.A., L.E. Coffeng, D.M. Bontje, E.C. Hasker, J.A.R. Postigo, D. Argaw, M.C. Boelaert, and S.J. de Vlas. Feasibility of eliminating visceral leishmaniasis from the Indian subcontinent: explorations with a set of deterministic age-structured transmission models. *Parasites and Vectors* 2016; 9:24.
- [432] A. Selvapandiyam, S.L. Croft, S. Rijal, H.L. Nakhasi, and N.K. Ganguly. Innovations for the elimination and control of visceral leishmaniasis. *PLoS Neg. Trop. Dis.* 2019; 13(9):e0007616.
- [433] Mubayi, A., M. Paredes, and J. Ospina. A comparative assessment of epidemiologically different cutaneous leishmaniasis outbreaks in Madrid, Spain and Tolima, Colombia: An estimation of the reproduction number via a mathematical model. *Trop. Med. Infect. Dis.* 2018; 3(2):43.
- [434] Furtado, A.S., F.B.B. de F. Nunes, A.M. dos Santos and A. de J.M. Caldas. Space-time analysis of visceral leishmaniasis in the state of Maranhão, Brazil. *Ciênc. Saúde. Colet.* 2015; 20(12): pp. 3935-3942.
- [435] Picado, A., B. Ostry, S.P. Singh, S. Uranw, E. Hasker, S. Rijal, S. Sundar, M. Boelaert, and F. Chappuis. Risk factors for visceral leishmaniasis and asymptomatic *Leishmania donovani* infection in India and Nepal. *PLoS ONE* 2014; 9(1): e87641.

- [436] Curi, N.H. de A., A.M. de O. Paschoal, R.L. Massara, A.P. Marcelino, A.A. Ribeiro, M. Passamani, G.R. Demétrio, and A.G. Chiarello. Factors associated with the seroprevalence of leishmaniasis in dogs living around Atlantic forest fragments. *PLoS ONE* 2014; 9(8): e104003.
- [437] Scarpino, S.V., A. Allard, and L. Hébert-Dufresne. The effect of a prudent adaptive behaviour on disease transmission. *Nature Physics* 2016; 12: pp. 1042-1046.
- [438] Watts, D.J., and S.H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature* 1998; 393: pp. 440-442.
- [439] Eubank, S., H. Guclu, V.S.A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature* 2004; 429: pp. 180-184.
- [440] Danon, L., A.P. Ford, T. House, C.P. Jewell, M.J. Keeling, G.O. Roberts, J.V. Ross, and M.C. Vernon. Networks and the epidemiology of infectious diseases. *Interdisciplinary Perspectives on Infectious Diseases* 2011; 284909.
- [441] Keeling, M.J., T. House, A.J. Cooper, and L. Pellis. Systematic approximations to susceptible-infectious-susceptible dynamics on networks. *PLoS Comput. Biol.* 2016; 12(12): e1005296.
- [442] Stolerman, L.M., D. Coombs, and S. Boatto. SIR-Network model and its application to dengue fever. *SIAM Journal on Applied Mathematics* 2015; 75(6): pp. 2581-2609.
- [443] Xia, C., L. Wang, S. Sun, and J. Wang. An SIR model with infection delay and propagation vector in complex networks. *Nonlinear Dynamics* 2012; 69: pp. 927-934.
- [444] Kiss, I.Z., L. Berthouze, T.J. Taylor, and P.L. Simon. Modelling approaches for simple dynamic networks and applications to disease transmission models. *Proc. R. Soc. A* 2012; 468: pp.1332-1355.
- [445] Pipatsart, N., W. Triampo, and C. Modchang. Stochastic models of emerging infectious disease transmission on adaptive random networks. *Computational and Mathematical Methods in Medicine* 2017; 2403851.
- [446] Danon, L., T.A. House, J.M. Read, and M.J. Keeling. Social encounter networks: Collective properties and disease transmission. *J. R. Soc. Interface* 2012; 9: pp. 2826-2833.
- [447] Bichara, D., and C. Castillo-Chavez. Vector-borne diseases models with residence times - A lagrangian perspective. *Mathematical Biosciences* 2016; 281: pp. 128-138.
- [448] Sah, P., J. Mann, and S. Bansal. Disease implications of animal social network structure: A synthesis across social systems. *J. Anim. Ecol.* 2017; 87: pp. 546-558.
- [449] Newman, M.E.J. Modularity and community structure in networks. *PNAS* 2006; 103(23): pp.8577-8582.

- [450] Stroeymeyt, N., A.V. Grasse, A. Crespi, D.P. Mersch, S. Cremer, and L. Keller. Social network plasticity decreases disease transmission in a eusocial insect. *Science* 2018; 362(6417): pp. 941-945.
- [451] Meyers, L.A., B. Pourbohloul, M.E.J. Newman, D.M. Skowronski, and R.C. Brunham. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology* 2005; 232: pp. 71-81.
- [452] Romano, V., M. Shen, J. Pansanel, A.J.J. MacIntosh, and C. Sueur. Social transmission in networks: global efficiency peaks with intermediate levels of modularity. *Behavioural Ecology and Sociobiology* 2018; 72:154.
- [453] Valverde, S., B. Vidiella, R. Montañez, A. Fraile, S. Sacristán, and F. García-Arenal. Coexistence of nestedness and modularity in host-pathogen infection networks. *Nature Ecology and Evolution* 2020; 4: pp. 568-577.
- [454] Laker, P.R., W. Hoppitt, M. Weiss, and J.R. Madden. The modularity of a social group does not affect the transmission speed of a novel, socially learned behaviour, or the formation of local variants. *Proc. R. Soc. B.* 2021; 288: 20202614.
- [455] Silk, M.J, and N.H. Fefferman. The role of social structure and dynamics in maintenance of endemic disease. *Behavioural Ecology and Sociobiology* 2021; 75: 122.
- [456] Salathé, M., M. Kazandjieva, J.W. Lee, P. Levis, M.W. Feldmann, and J.H. Jones. A high-resolution human contact network for infectious disease transmission. *PNAS* 2010; 107(51): 22020-22025.
- [457] Moore, C., and M.E.J. Newman. Epidemics and percolation in small-world networks. *Phys. Rev. E* 2000; 61(5): 5678.
- [458] Strogatz, S.H. *Nonlinear dynamics and chaos*, 2nd Edition. Westview Press 2015; Boulder, Colorado, United States.
- [459] Chapman, L.A.C., S.E.F. Spencer, T.M. Pollington, C.P. Jewell, D. Mondal, J. Alvar, T.D. Hollingsworth, M.M. Cameron, C. Bern, and G.F. Medley. Inferring transmission trees to guide targeting of interventions against visceral leishmaniasis and post-kala-azar dermal leishmaniasis. *PNAS* 2020; 117(41): pp. 25742-25750.
- [460] Travi, B.L., J. Montoya, Y. Solarte, L. Lozano, and C. Jaramillo. Leishmaniasis in Colombia. 1. Studies on the phlebotomine fauna associated with endemic foci in the Pacific Coast region. *Am. J. Trop. Med. Hyg.* 1988; 39(3): pp. 261-266.
- [461] Quinnell, R.J., O. Courtenay, L. Garcez, and C. Dye. The epidemiology of canine leishmaniasis: transmission rates estimated from a cohort study in Amazonian Brazil. *Parasitology* 1997; 115(2): pp. 143-156.
- [462] Givan, O., N. Schwartz, A. Cygelberg, and L. Stone. Predicting epidemic thresholds on complex networks: limitations of mean-field approaches. *Journal of Theoretical Biology* 2011; 288: pp. 21-28.

- [463] Webb, S.D., M.J. Keeling, and M. Boots. Host-parasite interactions between the local and the mean-field: How and when does spatial population structure matter? *Journal of Theoretical Biology* 2007; 249(1): pp. 140-152.
- [464] Wren, L., and A. Best. How local interactions impact the dynamics of an epidemic. *Bulletin of Mathematical Biology* 2021; 83:124.
- [465] Salathé, M., and J.H. Jones. Dynamics and control of diseases in networks with community structure. *PLoS Comput. Biol.* 2010; 6(4): e1000736.
- [466] d'Andrea, V., R. Gallotti, N. Castaldo, and M. De Domenico. Individual risk perception and empirical social structures shape the dynamics of infectious disease outbreaks. *PLoS Comput. Biol.* 2022; 18(2): e1009760.
- [467] Rozins, C., M.J. Silk, D.P. Croft, R.J. Delahay, D.J. Hodgson, R.A. McDonald, N. Weber, and M. Boots. Social structure contains epidemics and regulates individual roles in disease transmission in a group-living mammal. *Ecology and Evolution* 2018; 8: pp. 12044-12055.
- [468] Evans, J.C., D.J. Hodgson, N.J. Boogert, and M.J. Silk. Group size and modularity interact to shape the spread of infection and information through animal societies. *Behavioural Ecology and Sociobiology* 2021; 75:163.
- [469] Teodoro, U., D. Alberton, J.B. Kühn, E.S. dos Santos, D.R. dos Santos, A.R. dos Santos, O. de Oliveira, T.G. Silveira, and M.V. Lonardoni. Ecology of *Lutzomyia (Nyssomyia) whitmani* in an urban area in Maringá, Paraná, Brazil. *Revista de Saude Publica* 2003; 37(5): pp. 651-656.
- [470] Souza, N.A., C.A. Andrade-Coelho, A.A. Peixoto, E.F. Rangel. Nocturnal activity rhythms of *Lutzomyia intermedia* and *Lutzomyia whitmani* (Diptera: Psychodidae) in a transmission area of American cutaneous leishmaniasis in Rio de Janeiro state, Brazil. *Journal of Medical Entomology* 2005; 42(6): pp. 986-992.
- [471] Campos, A.M., C.L.C. dos Santos, R. Stumpp, L.H.D. da Silva, R.A. Maia, A.P. Paglia, and J.D.A. Filho. Photoperiod differences in sand fly (Diptera: Psychodidae) species richness and abundance in caves in Minas Gerais state, Brazil. *Journal of Medical Entomology* 2017; 54(1): pp. 100-105.
- [472] Chaves, L.F., J.E. Calzada, C. Rigg, A. Valderrama, N.L. Gottdenker, and A. Saldaña. Leishmaniasis sand fly vector density reduction is less marked in destitute housing after insecticide thermal fogging. *Parasites and Vectors* 2013; 6:164.
- [473] Kumar, V., A. Rama, S. Kesari, G.S. Bhunia, D.S. Dinesh, and P. Das. Oviposition behaviour of *Phlebotomus argentipes* - A laboratory-based study. *Mem. Inst. Oswaldo Cruz* 2013; 108(8): pp. 1065-1067.
- [474] Jalali, H., S.H. Nikookar, N. Hosseini-Vasoukolaei, E. Jahanifard, A.A. Enayati, F. Motevalli-Haghi, and J. Yazdani-Charati. Ecology of sand flies (Diptera: Psychodidae, Phlebotominae) in Jajarm County, an area with high risk of cutaneous leishmaniasis, in North Khorasan, Iran. *BMC Zoology* 2020; 7:14.

- [475] Vanlerberghe, V., S.P. Singh, I.S. Paudel, B. Ostyn, A. Picado, A. Sánchez, S. Rijal, S. Sundar, C. Davies, and M. Boelaert. Determinants of bednet ownership and use in visceral leishmaniasis-endemic areas of the Indian subcontinent. *Tropical Medicine and International Health* 2010; 15(1): pp.60-67.
- [476] Chowdhury, R., V. Kumar, D. Mondal, M.L. Das, P. Das, A.P. Dash, and A. Kroeger. Implication of vector characteristics of *Phlebotomus argentipes* in the kala-azar elimination programme in the Indian subcontinent. *Pathogens and Global Health* 2016; 110(3): pp. 87-96.
- [477] Pareyn, M., A. Kochora, L. van Rooy, N. Eligo, B.V. Broecke, N. Girma, B. Merdekios, T. Wegayehu, L. Maes, G. Caljon, B. Lindtjørn, H. Leirs, and F. Massebo. Feeding behaviour and activity of *Phlebotomus pedifer* and potential reservoir hosts of *Leishmania aethiopica* in southwestern Ethiopia. *PLoS Neg. Trop. Dis.* 2020; 14(3): e0007947.
- [478] Hogsette, J.A., H.A. Hanafi, U.R. Bernier, D.L. Kline, E.Y. Fawaz, B.D. Furman, and D.F. Hoel. Discovery of diurnal resting sites of phlebotomine sand flies in a village in southern Egypt. *J. of the American Mosquito Control Association* 2008; 24(4): pp.601-603.
- [479] Müller, G.C., V.D. Kravchenko, L. Rybalov, J.C. Beier, and Y. Schlein. Characteristics of resting habitats of adult *Phlebotomus papatasi* in Neot Hakikar, an oasis south of the Dead Sea. *Journal of Vector Ecology* 2011; 36(s1): pp. S179-S186.
- [480] Elnaiem, D.-E.A., O. Dakein, A.M.-A. Alawad, B. Alsharif, A. Khogali, T. Jibreel, O.F. Osman, H. Has'san, A.M. Atia, M. Elhag, M. den Boer, K. Ritmeijer, C. Bern, J. Alvar, N. Khalid, and O. Courtenay. Outdoor residual insecticide spraying (ODRS), a new approach for the control of exophilic vectors of human visceral leishmaniasis: *Plebotomus orientalis* in East Africa. *PLoS Neg. Trop. Dis.* 2020; 14(10): e0008774.
- [481] Balaska, S., E.A. Fotakis, A. Chaskopoulou, and J. Vontas. Chemical control and insecticide resistance status of sand fly vectors worldwide. *PLoS Neg. Trop. Dis.* 2021; 15(8): e0009586.
- [482] Iggidr, A., G. Sallet, and M.O. Souza. On the dynamics of a class of multi-group models for vector-borne diseases. *J. Math. Anal. Appl.* 2016; 441: pp. 723-743.
- [483] Zhao, R., and Q. Liu. Dynamical behaviour and optimal control of a vector-borne diseases model on bipartite networks. *Applied Mathematic Modelling* 2022; 102: 540-563.
- [484] Courtenay, O., N.C. Peters, M.E. Rogers, and C. Bern. Combining epidemiology with basic biology of sand flies, parasites and hosts to inform leishmaniasis transmission dynamics and control. *PLoS Pathog.* 2017; 13(10): e1006571.
- [485] Tiwary, P., D. Kumar, M. Mishra, R.P. Singh, M. Rai, and S. Sundar. Seasonal variation in the prevalence of sand flies infected with *Leishmania donovani*. *PLoS ONE* 2013; 8(4): e61370.

- [486] de Ávila, M.M., A.F. Brilhante, C.F. de Souza, P.D. Bevilacqua, E.A.B. Galati, and R.P. Brazil. Ecology, feeding and natural infection by *Leishmania* spp. of phlebotomine sand flies in an area of high incidence of American tegumentary leishmaniasis in the municipality of Rio Branco, Acre, Brazil. *Parasites and Vectors* 2018; 11:64.
- [487] Ikeogu, N.M., G.N. Akaluka, C.A. Edechi, E.S. Salako, C. Onyilagha, A.F. Barazandeh, and J.E. Uzonna. *Leishmania* immunity: Advancing immunotherapy and vaccine development. *Microorganisms* 2020; 8(8): 1201.
- [488] Zijlstra, E.E. The immunology of post-kala-azar dermal leishmaniasis (PKDL). *Parasites and Vectors* 2016; 9: 464.
- [489] Gorski, S., S.M. Collin, K. Ritmeijer, K. Keus, F. Gatluak, M. Mueller, and R.N. Davidson. Visceral leishmaniasis relapse in Southern Sudan (1999-2007): A retrospective study of risk factors and trends. *PLoS Neg. Trop. Dis.* 2010; 4(6): e705.
- [490] Cota, G.F., M.R. de Sousa, and A. Rabello. Predictors of visceral leishmaniasis relapse in HIV-infected patients: A systematic review. *PLoS Neg. Trop. Dis.* 2011; 5(6): e1153.
- [491] Johnstone, G.L., D.L. Smith, and D.A. Fidock. Malaria's missing number: Calculating the human component of R_0 by a within-host mechanistic model of *Plasmodium falciparum* infection and transmission. *PLoS Comput. Biol.* 2013; 9(4): e1003025.
- [492] Sy, K.T.L., L.F. White, and B.E. Nichols. Population density and basic reproductive number of COVID-19 across United States counties. *PLoS ONE* 2021; 16(4): e0249271.
- [493] Nieddu, G.T., L. Billings, J.H. Kaufman, E. Forgoston, and S. Bianco. Extinction pathways and outbreak vulnerability in a stochastic Ebola model. *J. R. Soc. Interface* 2017; 14(127): 20160847.
- [494] Hagenaars, T.J., C.A. Donnelly, and N.M. Ferguson. Spatial heterogeneity and the persistence of infectious diseases. *Journal of Theoretical Biology* 2004; 229(3): pp. 349-359.
- [495] Jeong, J., and H. McCallum. The persistence of a SIR disease in a metapopulation: Hendra virus epidemics in Australian black flying foxes (*Pteropus alecto*). *Australian Journal of Zoology* 2021; 69(1): pp. 1-11.
- [496] Sunyoto, T., J. Potet, and M. Boelaert. Visceral leishmaniasis in Somalia: A review of epidemiology and access to care. *PLoS Neg. Trop. Dis.* 2017; 11(3): e0005231.
- [497] Das, A., M. Karthick, S. Dwivedi, I. Banerjee, T. Mahapatra, S. Srikantiah, and I. Chaudhury. Epidemiologic correlates of mortality among symptomatic visceral leishmaniasis cases: Findings from situation assessment in high endemic foci in India. *PLoS Neg. Trop. Dis.* 2016; 10(11): e0005150.

- [498] Le Rutte, E.A., L.A.C. Chapman, L.E. Coffeng, J.A. Ruiz-Postigo, P.L. Olliaro, E.R. Adams, E.C. Hasker, M.C. Boelaert, T.D. Hollingsworth, G.F. Medley, and S.J. de Vlas. Policy recommendations from transmission modeling for the elimination of visceral leishmaniasis in the Indian Subcontinent. *Clinical Infectious Diseases* 2018; 66(suppl. 4): pp. S301-S308.
- [499] Benabid, M., J. Ghrab, A. Rhim, R. Ben-romdhane, K. Aoun, and A. Bouratbine. Temporal dynamics and *Leishmania infantum* infection prevalence of *Phlebotomus perniciosus* (Diptera: Phlebotominae) in highly endemic areas of visceral leishmaniasis in Tunisia. *PLoS ONE* 2017; 12(9): e0184700.
- [500] Talmi-Frank, D., C.L. Jaffe, A. Nasereddin, A. Warburg, R. King, M. Svobodova, O. Peleg, and G. Baneth. *Leishmania tropica* in rock hyraxes (*Procapra capensis*) in a focus of human cutaneous leishmaniasis. *Am. J. Trop. Med. Hyg.* 2010; 82(5): pp. 814-818.
- [501] Lloyd-Smith, J.O., S.J. Schreiber, P.E. Kopp, and W.M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005; 438: pp. 355-359.
- [502] Zhang, N., H. Huang, B. Su, X. Ma, and Y. Li. A human behaviour integrated hierarchical model of airborne disease transmission in a large city. *Building and Environment* 2018; 127: pp. 211-220.
- [503] Jamal, Y., M. Gangwar, M. Usmani, A.E. Adams, C.-Y. Wu, T.H. Nguyen, R. Colwell, and A. Jutla. Identification of thresholds on population density for understanding transmission of COVID-19. *GeoHealth* 2022; 6(9): e2021GH000449.
- [504] Tarallo, V.D., F. Dantas-Torres, R.P. Lia, and D. Otranto. Phlebotomine sand fly population dynamics in a leishmaniasis endemic peri-urban area in southern Italy. *Acta Tropica* 2010; 116(3): pp. 227-234.
- [505] Miranda, D.E. de O., K.G. da S. Sales, M.A. da G. Faustino, L.C. Alves, S.P. Brandão-Filho, F. Dantas-Torres, and G.A. de Carvalho. Ecology of sand flies in a low-density residential rural area, with mixed forest/agricultural exploitation, in north-eastern Brazil. *Acta Tropica* 2015; 146: p. 89-94.
- [506] Bern, C., O. Courtenay, and J. Alvar. Of cattle, sand flies and men: A systematic review of risk factor analyses for South Asian visceral leishmaniasis and implications for elimination. *PLoS Neg. Trop. Dis.* 2010; 4(2): e599.
- [507] Lion, S., and M. Boots. Are parasites “prudent” in space? *Ecology Letters* 2010; 13: pp. 1245-1255.
- [508] Rogers, M.E., K. Corware, I. Muller, and P.A. Bates. *Leishmania infantum* proteophosphoglycans regurgitated by the bite of its natural sand fly vector, *Lutzomyia longipalpis*, promote parasite establishment in mouse skin and skin-distant tissues. *Microbes and Infection* 2010; 12(11): pp. 875-879.
- [509] Hall, M.J.R., D. Ghosh, D. Martín-Vega, B. Clark, I. Clatworthy, R.A. Cheke, and M.E. Rogers. Micro-CT visualization of a promastigote secretory gel (PSG) and

parasite plug in the digestive tract of the sand fly *Lutzomyia longipalpis* infected with *Leishmania mexicana*. PLoS Neg. Trop. Dis. 2021; 15(8): e0009682.

- [510] Serafim, T.D., I.V. Coutinho-Abreu, R. Dey, R. Kissinger, J.G. Valenzuela, F. Oliveira, and S. Kamhawi. Leishmaniasis: the act of transmission. Trends in Parasitology 2021; 37(11): pp. 976-987.
- [511] Ready, P.D. Biology of phlebotomine sand flies as vectors of disease agents. Annu. Rev. Entomol. 2013; 58: pp. 227-250.
- [512] Schlein, Y., and R.L. Jacobson. Sugar meals and longevity of the sandfly *Phlebotomus papatasi* in an arid focus of *Leishmania major* in the Jordan Valley. Medical and Veterinary Entomology 2002; 13(1): pp. 65-71.
- [513] Hurwitz, I., H. Hillesland, A. Fieck, P. Das, and R. Durvasula. The paratransgenic sand fly: A platform for control of *Leishmania* transmission. Parasites and Vectors 2011; 4: 82.
- [514] Kasap, O.E., and B. Alten. Comparative demography of the sand fly *Phlebotomus papatasi* (Diptera: Psychodidae) at constant temperatures. J. Vec. Ecol. 2006; 31(2): pp. 378-385.
- [515] Sant'Anna, M.R.V., H. Diaz-Albiter, K. Aguiar-Martins, W.S. Al Salem, R.R. Cavalcante, V.M. Dillon, P.A. Bates, F.A. Genta, and R.J. Dillon. Colonisation resistance in the sand fly gut: *Leishmania* protects *Lutzomyia longipalpis* from bacterial infection. Parasites and Vectors 2014; 7: 329.
- [516] Chaniotis, B., I. Spyridaki, E. Scoulika, and M. Antoniou. Colonization of *Phlebotomus neglectus* (Diptera: Psychodidae), the major vector of visceral leishmaniasis in Greece. J. Med. Entomol. 2000; 37(3): pp. 346-348.
- [517] Noguera, P., M. Rondón, and E. Nieves. Effect of blood source on the survival and fecundity of the sandfly *Lutzomyia ovallesi* Ortis (Diptera: Psychodidae), vector of *Leishmania*. Biomédica 2006; 26(Supl.1): pp. 57-63.
- [518] Erguler, K., I. Pontiki, G. Zittis, Y. Proestos, V. Christodoulou, N. Tsirigotakis, M. Antoniou, O.E. Kasap, B. Alten, and J. Lelieveld. A climate-driven and field data-assimilated population dynamics model of sand flies. Scientific Reports 2019; 9: 2469.
- [519] Elmojtaba, I.M., J.Y.T. Mugisha, and M.H.A. Hashim. Mathematical analysis of the dynamics of visceral leishmaniasis in the Sudan. Applied Mathematics and Computation 217; pp. 2567-2578.
- [520] Medley, G.F., T.D. Hollingsworth, P.L. Olliaro, and E.R. Adams. Health-seeking behaviour, diagnostics and transmission dynamics in the control of visceral leishmaniasis in the Indian subcontinent. Nature 2015; 528: pp. S102-S108.
- [521] Le Rutte, E.A., L.E. Coffeng, S. Malvoti, P.M. Kaye, and S.J. de Vlas. The potential impact of human visceral leishmaniasis vaccines on population incidence. PLoS Neg. Trop. Dis. 2020; 14(7): e0008468.

- [522] Blumberg, S., and J.O. Lloyd-Smith. Comparing methods for estimating R_0 from the size distribution of subcritical transmission chains. *Epidemics* 2013; 5(3): pp. 131-145.
- [523] Cross, P.C., J.O. Lloyd-Smith, P.L.F. Johnson, and W.M. Getz. Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecology Letters* 2005; 8(6): pp. 587-595.
- [524] Delamater, P.L., E.J. Street, T.F. Leslie, Y.T. Yang, and K.H. Jacobsen. Complexity of the basic reproduction number (R_0). *Emerg. Infect. Dis.* 2019; 25(1): pp. 1-4.
- [525] Velázquez-Castro, J., A. Anzo-Hernández, B. Bonilla-Capilla, M. Soto-Bajo, and A. Fraguera-Collar. Vector-borne disease risk indexes in spatially structured populations. *PLoS Neg. Trop. Dis.* 2018; 12(2): e0006234.
- [526] Boussaa, S., S. Guernaoui, B. Pesson, and A. Boumezzough. Seasonal fluctuations of phlebotomine sand fly populations (Diptera: Psychodidae) in the urban area of Marrakech, Morocco. *Acta Tropica* 2005; 95(2): pp. 86-91.
- [527] Ramos, W.R., J.F. Medeiros, G.R. Julião, C.M. Ríos-Velásquez, E.F. Marialva, S.J.M. Desmoulière, S.L.B. Luz, and F.A.C. Pessoa. Anthropogenic effects on sand fly (Diptera: Psychodidae) abundance and diversity in an Amazonian rural settlement, Brazil. *Acta Tropica* 2014; 139: pp. 44-52.
- [528] Sah, P., S.T. Leu, P.C. Cross, P.J. Hudson, and S. Bansal. Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *PNAS* 2017; 114(16): pp. 4165-4170.
- [529] Ferrari, M.J., O.N. Bjørnstad, and A.P. Dobson. Estimation and inference of R_0 of an infectious pathogen by a removal method. *Mathematical Biosciences* 2005; 198(1): pp. 14-26.
- [530] Ives, A.R., and C. Bozzuto. Estimating and explaining the spread of COVID-19 at the county level in the USA. *Commun. Biol.* 2021; 4: 60.
- [531] Kiss, I.Z., J.C. Miller, and P.L. Simon. *Mathematics of epidemics on networks*. Springer International Publishing, 2017; Switzerland.
- [532] Volpedo, G., R.H. Huston, E.A. Holcomb, T. Pacheco-Fernandez, S. Gannavaram, P. Bhattacharya, H.L. Nakhasi, and A.R. Satoskar. From infection to vaccination: Reviewing the global burden, history of vaccine development, and recurring challenges in global leishmaniasis protection. *Expert Review of Vaccines* 2021; 20(11): pp. 1431-1446.
- [533] Srisuton, P., A. Phumee, S. Sunantaraporn, R. Boonserm, S. Sor-suwan, N. Brownell, T. Pengsakul, and P. Siriyasatien. Detection of *Leishmania* and *Trypanosoma* DNA in field-caught sand flies from endemic and non-endemic areas of leishmaniasis in Southern Thailand. *Insects* 2019; 10(8): 238.

Appendix A

Supplementary Information

A.1 Chapter 2 Supporting information

S1 Table RAG Mouse Parameter Combinations The skin heterogeneity and mean skin parasite burden values for RAG mice 1-18 used throughout our simulations, as originally calculated by Doehl *et al.*. [A]: Values derived from Doehl *et al.* [45].

Mouse	K Value	Mean Skin Parasite Burden	Source
RAG 1	1.2697	$1 * 10^{3.2}$	[A]
RAG 2	0.9156	$1 * 10^{3.1}$	[A]
RAG 3	1.2474	$1 * 10^{3.2}$	[A]
RAG 4	3.296	$1 * 10^3$	[A]
RAG 5	0.9523	$1 * 10^3$	[A]
RAG 6	0.7229	$1 * 10^4$	[A]
RAG 7	0.6561	$1 * 10^5$	[A]
RAG 8	0.8124	$1 * 10^5$	[A]
RAG 9	0.7222	$1 * 10^4$	[A]
RAG 10	0.9081	$1.057 * 10^4$	[A]
RAG 11	0.7723	$6.2977 * 10^3$	[A]
RAG 12	1.1523	$9.424 * 10^3$	[A]
RAG 13	1.0874	$8.6053 * 10^6$	[A]
RAG 14	1.2146	$6.7356 * 10^4$	[A]
RAG 15	0.5017	$2.1384 * 10^4$	[A]
RAG 16	1.2758	$1.6032 * 10^4$	[A]
RAG 17	1.2462	$1.1973 * 10^5$	[A]
RAG 18	1.644	$9.315 * 10^4$	[A]

S1 Method Parameterisation of Model A

The basic model (Model A) produced in this study focuses on three promastigotes stages, nectomonads, leptomonads and metacyclics. Differential equation based models were produced based on the lifecycle described by Rogers *et al* [49]. This method assumes

that the parameters used for rates are all constant. Parameterisation was achieved by fitting data from Rogers *et al* to the model. Data was collected from Figure 1 of this paper using the digitize function in R as this data wasn't readily available [49,168]. The digitize function is used to manually collected data from plots. The data collected were the total number of parasites in the sand fly and the percentage of each promastigote stage (nectomonad, leptomonad, metacyclic) present in the sand fly over a course of 10 days. This was then used to calculate the number of each promastigote stage. This data was then exported into MATLAB where the function "lsqcurvefit" was used to produce the best fitting parameter values. The quality of fit was assessed via an R^2 value, which defines the proportion of variation that is explained by a model. A high R^2 is indicative of a good fit where as a low R^2 is indicative of a poor fit.

S2 Method Bite Mechanics

In this section, we explore a few examples of the negative binomial distribution that determines the uptake of amastigotes during a blood meal. For convenience we repeat the definition of the negative binomial distribution from the main text before we begin the example.

Bite loads were generated from a negative binomial distribution using the 'nbinrnd' function in MATLAB. This function outputs a random value from a negative binomial distribution. This takes the following inputs: P (Probability of a positive result) and R (the number of successes required). R and P are defined as follows:

$$R = \frac{\mu_N^2}{\sigma_N^2 - \mu_N} \qquad P = \frac{\mu_N}{\sigma_N^2}.$$

where:

$$\mu_N = P_B V_{BM} \qquad \sigma_N^2 = (\mu_N) \left(1 + \frac{\mu_N}{k}\right)$$

as outlined in the main text. For this example, we assume a fixed mean skin parasite burden $P_B = 10^5$ and blood meal volume $V_{BM} = 1.6 * 10^{-3}$ and consider three different k values: $k = 0.2$ corresponding to a heterogeneous distribution, $k = 2.0$ corresponding to a homogeneous distribution, and an intermediate value of $k = 1.0$ (Figure A.1).

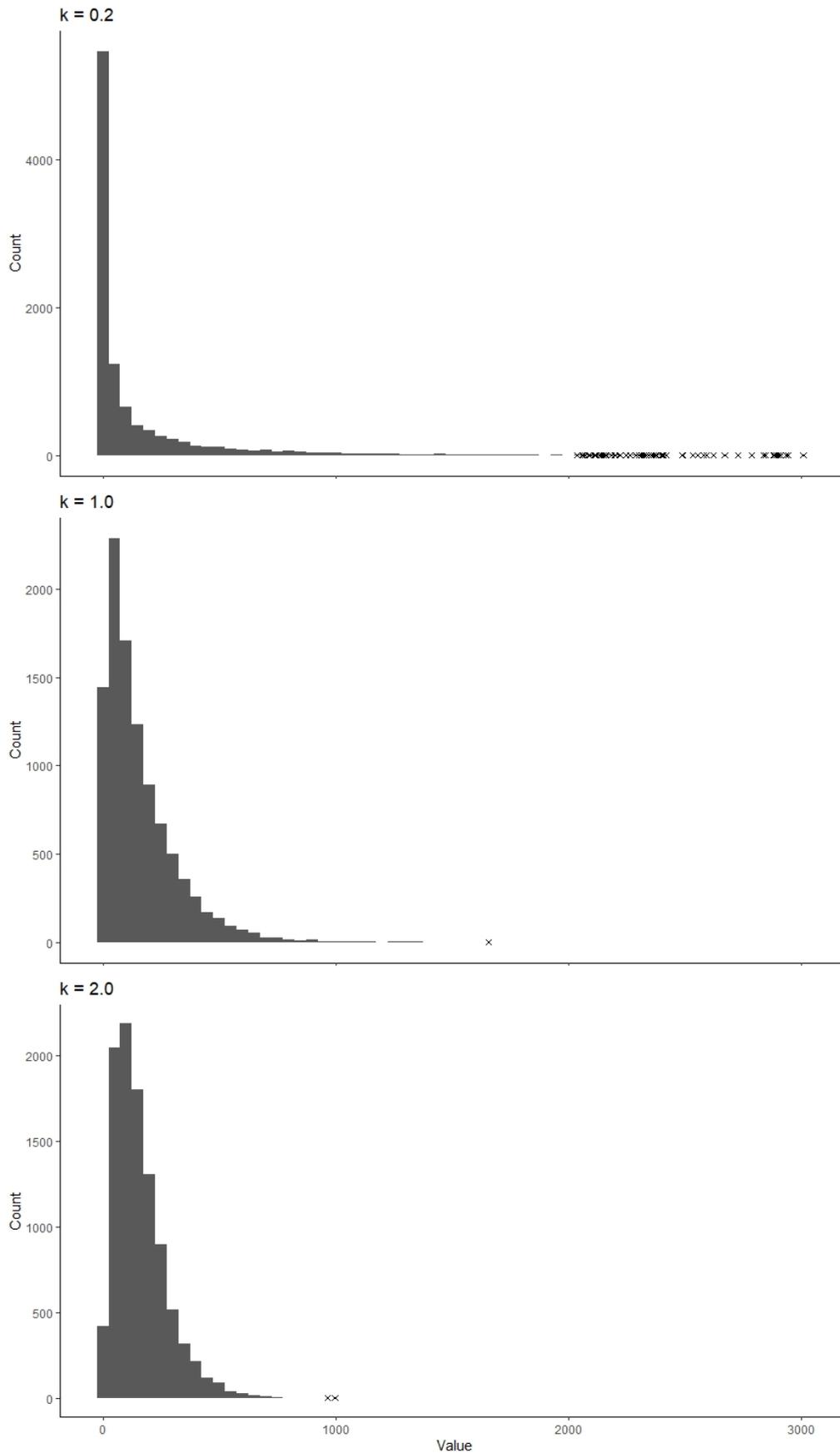


Figure A.1. Distribution of amastigote numbers drawn from negative binomial distributions with varying skin homogeneity k : a heterogeneous environment ($k = 0.2$, top), a homogeneous environment ($k = 2.0$, bottom) and an intermediate environment ($k = 1.0$, middle). Extremely large values are depicted using crosses.

S3 Method Population Sink Mechanics

In order to assess the sensitivity of our models to the assumption that we have 100% efficiency in converting between lifecycle stages, we present an adapted form of the model which incorporates a population sink at each lifecycle stage. For simplicity we assume that the population sinks are some constant death rate per parasite per day (or equivalently a loss of parasite differentiation capacity) and that this rate does not vary between lifecycle stages. With the inclusion of the population sinks, Model A is now:

$$\frac{dN}{dt} = -\alpha N - \gamma N \quad (\text{A.1})$$

$$\frac{dL}{dt} = \alpha N + rL \left(1 - \frac{N + L + M}{C}\right) - sL - \gamma L \quad (\text{A.2})$$

$$\frac{dM}{dt} = sL - uM - \gamma M \quad (\text{A.3})$$

where we assume γ to be constant. Model B can be modified similarly. Normal mode is now:

$$\frac{dN}{dt} = -\alpha N - \gamma N \quad (\text{A.4})$$

$$\frac{dL}{dt} = \alpha N + rL \left(1 - \frac{N + L + M + R}{C}\right) - sL - \gamma L \quad (\text{A.5})$$

$$\frac{dM}{dt} = sL + vR - uM - \gamma M \quad (\text{A.6})$$

$$\frac{dR}{dt} = qR \left(1 - \frac{N + L + M + R}{C}\right) - vR - \gamma R \quad (\text{A.7})$$

and dedifferentiation mode is:

$$\frac{dM}{dt} = sL - gM - uM - \gamma M \quad (\text{A.8})$$

$$\frac{dR}{dt} = qR \left(1 - \frac{N + L + M + R}{C}\right) + gM - \gamma R \quad (\text{A.9})$$

We shall consider two different exemplar values of γ . To represent a small population sink, we consider $\gamma = 0.05$. For a larger sink, we choose $\gamma = 0.15$.

A.1.1 Population Dynamics

In order to understand the population dynamics for Model A (Eqns 2.1-2.3) and Model B (Eqns 2.4 - 2.9), we consider a simple example. We examine a single fly with initially contains 200 nectomonads, and takes a blood meal from an uninfected host at day 6 (such that no further parasites are introduced to the population). We focus our attention initially on the proportion of the parasites in each category. Figure A.2 shows the dynamics for Model A:

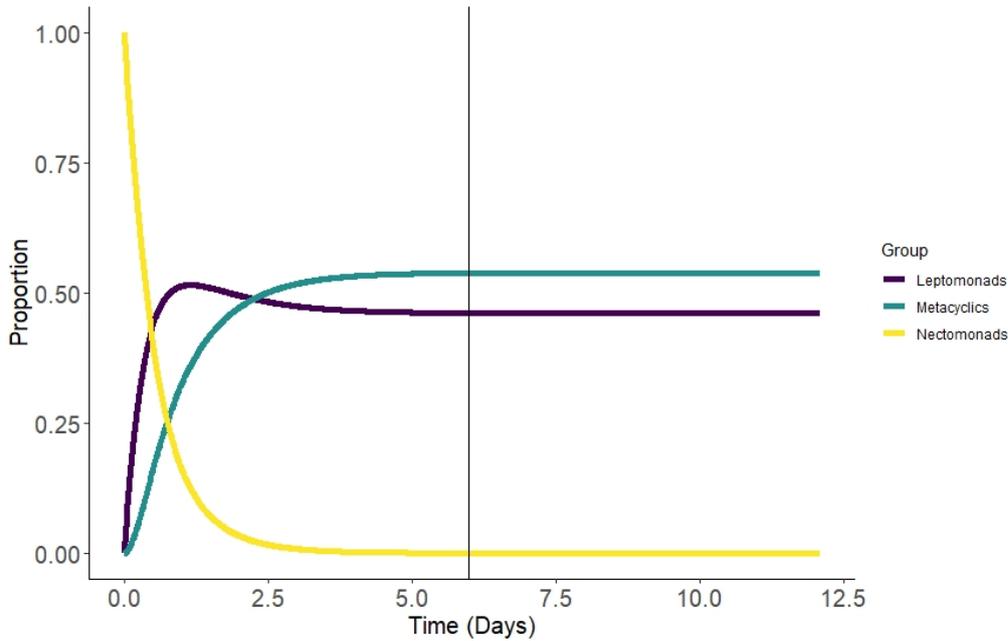


Figure A.2. The proportion of different parasite life stages in the sand fly vector for the first 12 days post-infection under the system outlined in Model A (Eqns 2.1 - 2.3). An additional bite at day 6 is indicated by the vertical line.

The initial peak of nectomonads decreases over time as they differentiate into leptomonads (and then into metacyclics). If we now examine the dynamics for Model B (Figure A.3):

The primary difference between the two models is that in Model B we see a dramatic shift to a population dominated by retroleptomonads at day 12, which lasts for 4 days until they begin differentiating back into metacyclics. We can glean further insights into the population dynamics by examining the total numbers of parasites of each life stage. If we once again examine the behaviour of Model A (Figure A.4):

It is worth noting that the number of metacyclics and leptomonads decrease steadily after an initial peak and this is because the leptomonads differentiate into metacyclics, the population of which then declines slowly with time. We can now consider the dynamics in Model B. The pre- and post-bite dynamics in Model B are very different, and for ease

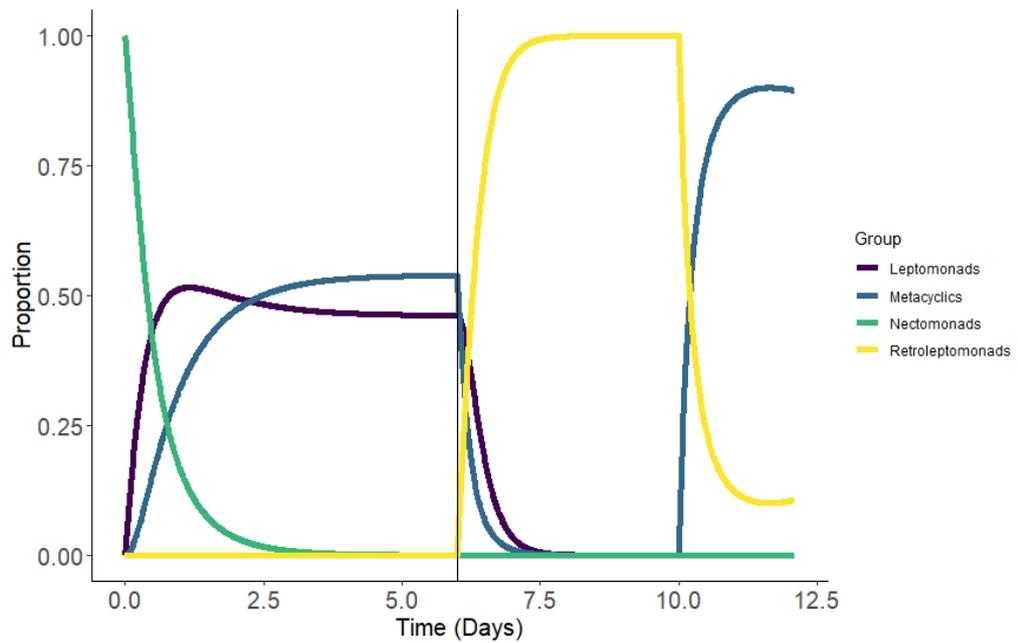


Figure A.3. The proportion of different parasite life stages in the sand fly vector for the first 12 days post-infection under the system outlined in Model B (Eqns 2.4 - 2.9). An additional bite at day 6 is indicated by the vertical line.

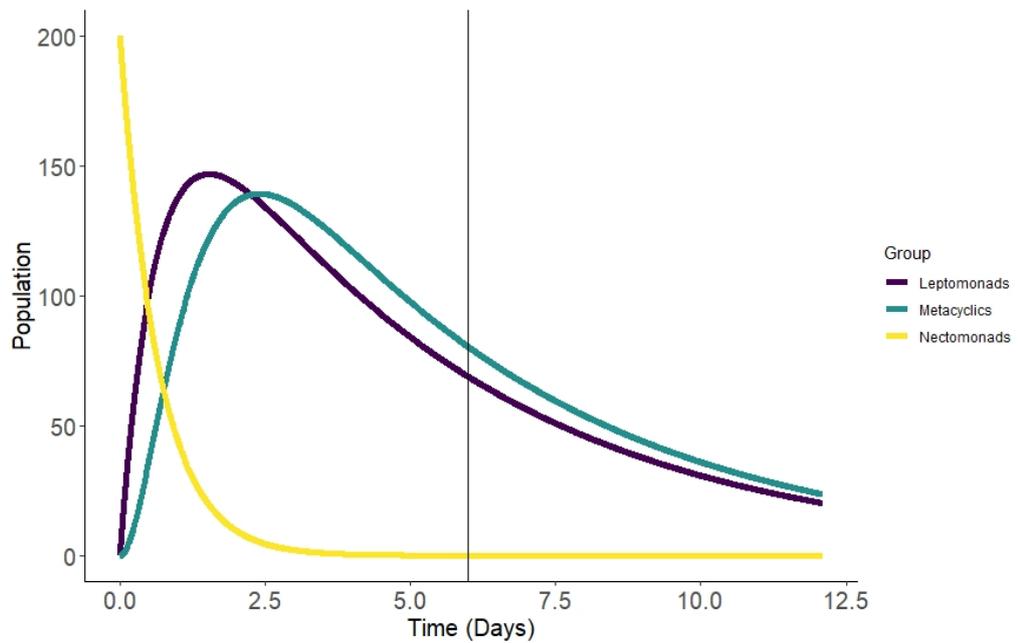


Figure A.4. The number of parasites in each life stage in the sand fly vector for the first 12 days post-infection under the system outlined in Model A (Eqns 2.1 - 2.3). An additional bite at day 6 is indicated by the vertical line.

of viewing are separated into different graphs. First, consider the dynamics for the first 6 days post-infection (Figure A.5):

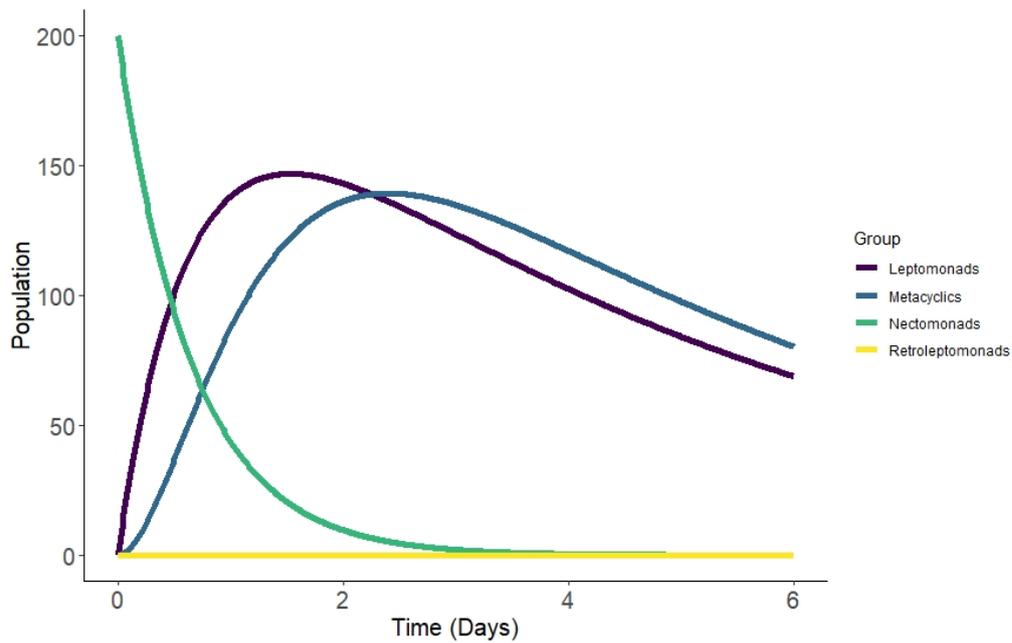


Figure A.5. The number of parasites in each life stage in the sand fly vector for the first 6 days post-infection under the system outlined in Model B (Eqns 2.4 - 2.9).

This behaviour is identical to that of Model A (Figure A.4). The distinction between the models becomes apparent if we consider the post-bite dynamics (Figure A.6):

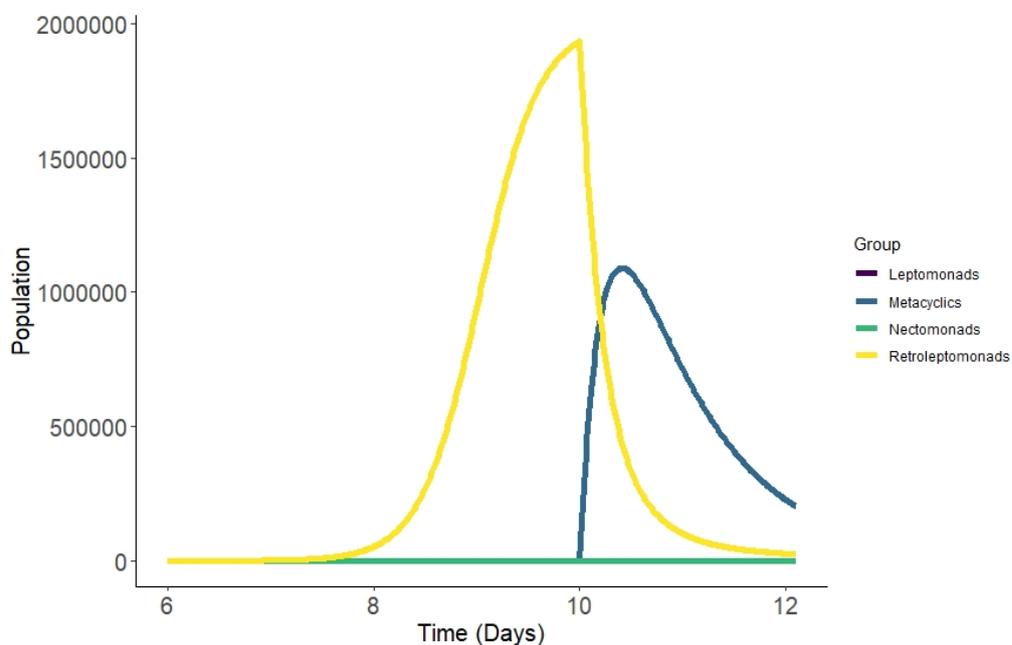


Figure A.6. The number of parasites in each life stage in the sand fly vector from day 6 to day 12 post-infection under the system outlined in Model B (Eqns 2.4 - 2.9).

The rapid replication of the newly-emerged retroleptomonads greatly increases the population of parasites in the sand fly, and the resulting number of metacyclics by day 12 is much higher in Model B than it was in Model A.

S1 Code

Supplementary Code. All MATLAB and R code comprising our implementation of the models and simulations used in this investigation.

A.1.2 Supplementary Figures

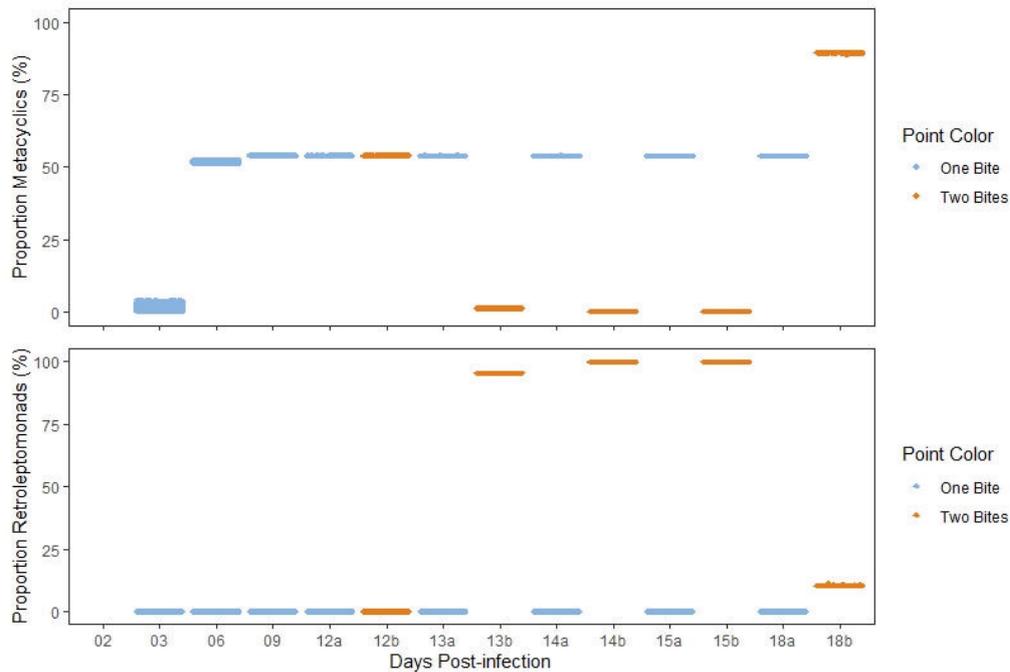


Figure A.7. Replicating the results of [54] (parasite proportions). Comparison of the proportions of metacyclics (top) and retroptomonads (bottom) at specific days throughout the lifespan of the simulated flies. Blue represents flies that bite only at day 0, orange represents flies that bite at day 12. The two categories are combined prior to day 12.

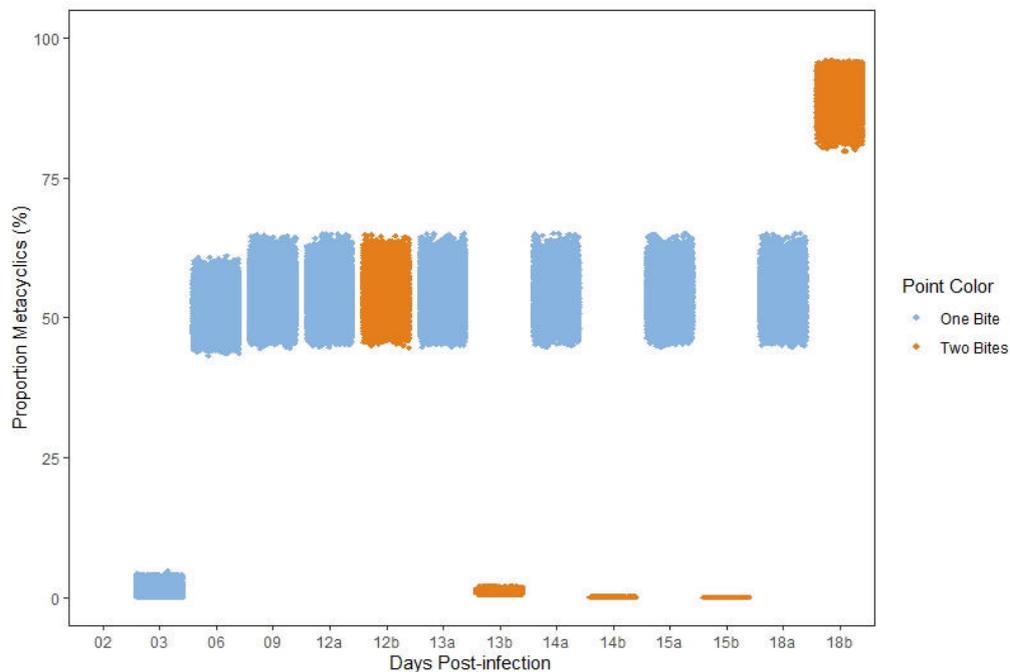


Figure A.8. Evaluating model robustness by randomising parameters. Number of metacyclics within the sand flies at specific days, with all parameters randomised prior to the start of each simulation. Parameters lie within 10% of the default value (Table 1). Blue represents flies that bite only a day 0, orange represents flies that bite at day 12.

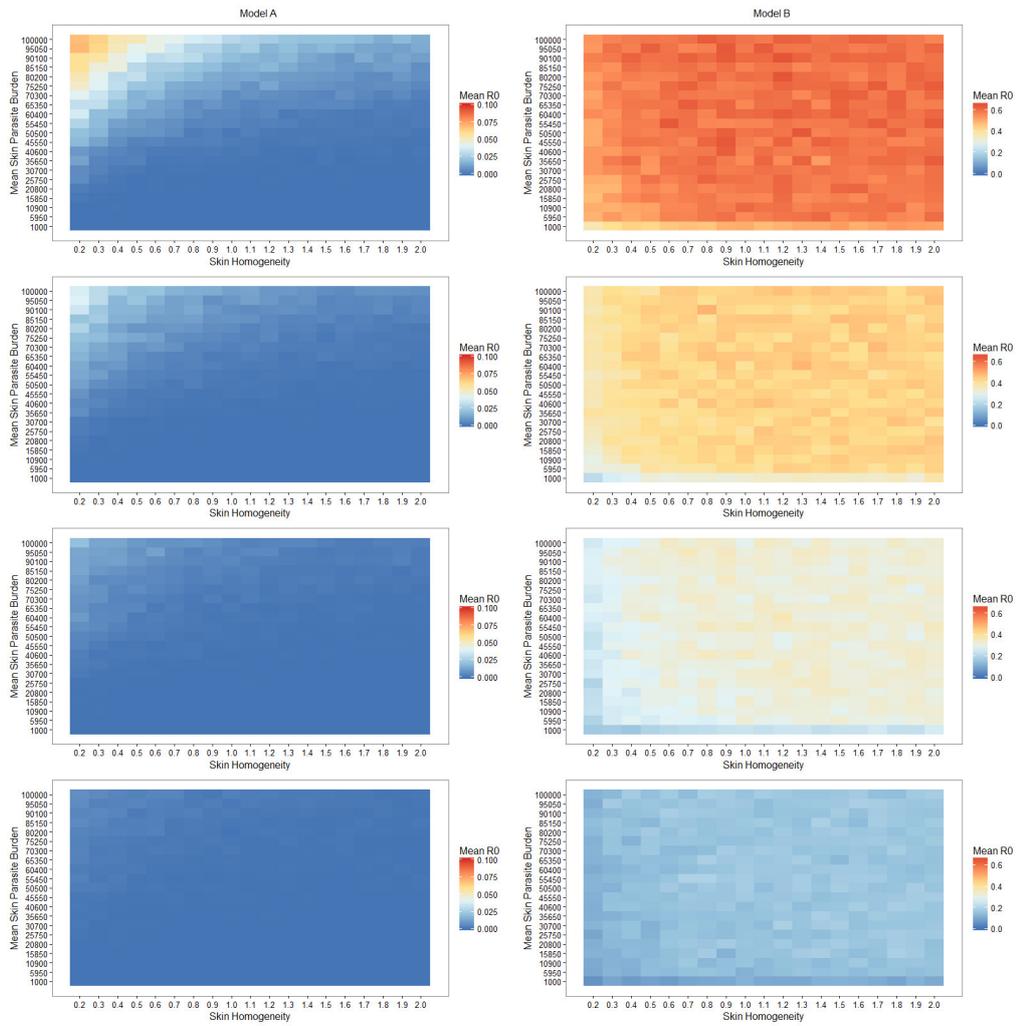


Figure A.9. Additional infected host proportions reflect the retroleptomonad dominance. Heatmaps of the Mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top row), 50% (second row), 25% (third row), and 10% (bottom row) chance of biting an infected host, with the smooth transmission threshold function.

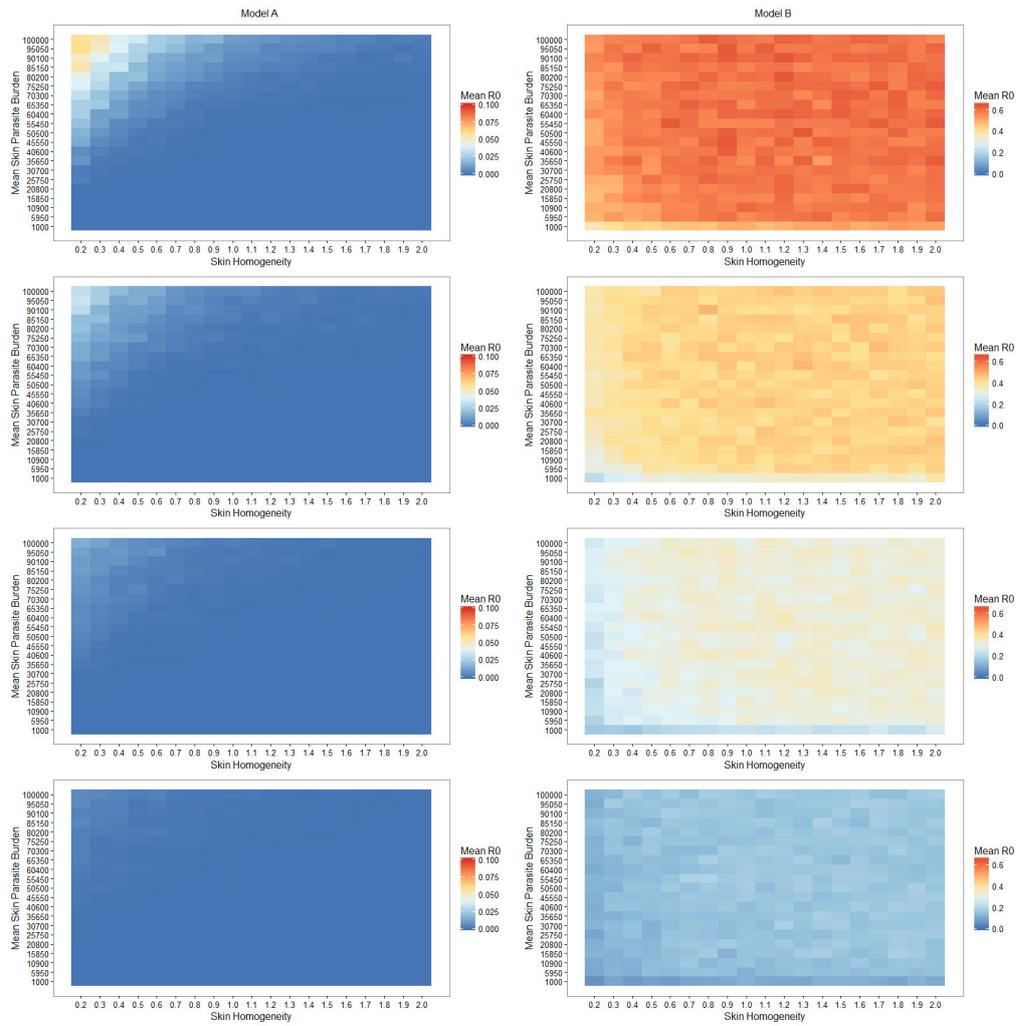


Figure A.10. Heatmap dynamics remain qualitatively similar under a binary transmission threshold. Heatmaps of the Mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top row), 50% (second row), 25% (third row), and 10% (bottom row) chance of biting an infected host, with the binary transmission threshold.

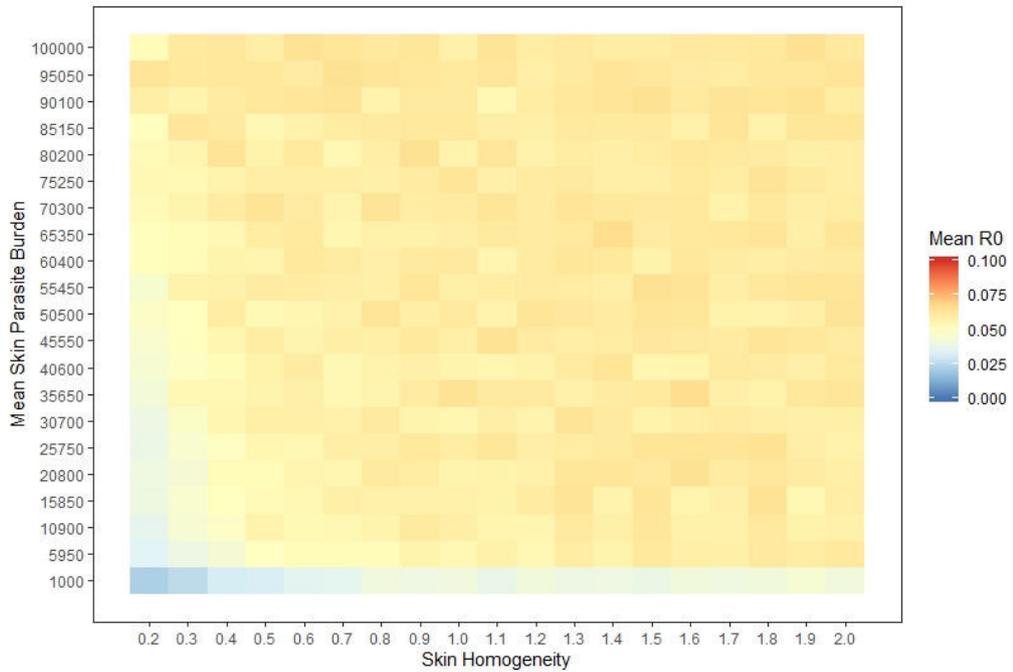


Figure A.11. Reduced lifespan (20 days) dynamics remain qualitatively similar under a binary transmission threshold. Heatmap of the Mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 20 days, with the binary transmission threshold.

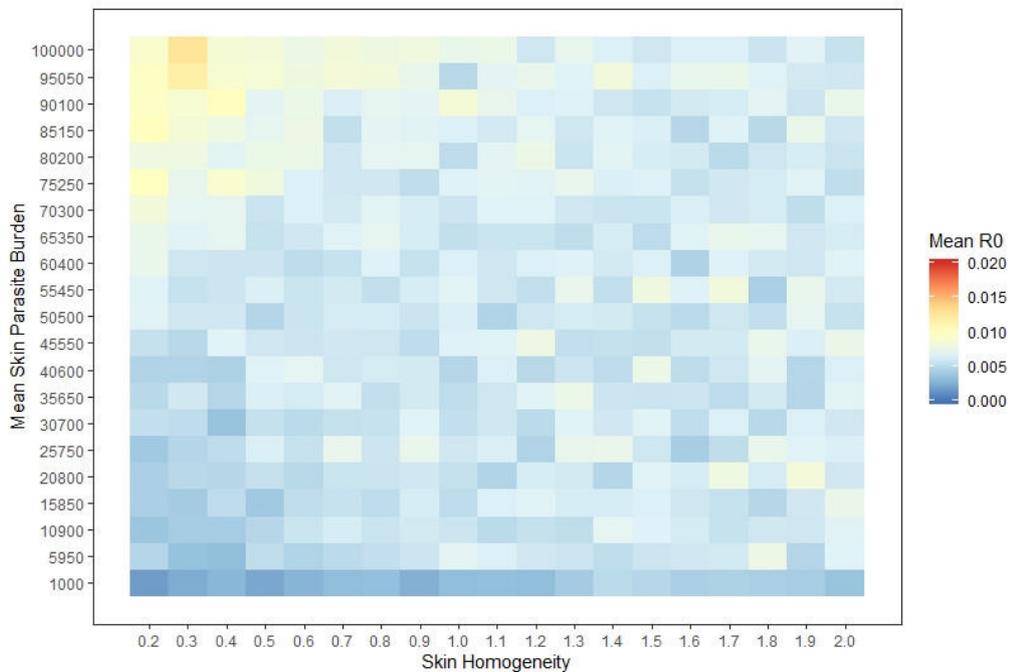


Figure A.12. Reduced lifespan (15 days) dynamics remain qualitatively similar under a binary transmission threshold. Heatmap of the Mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 15 days, with the binary transmission threshold.

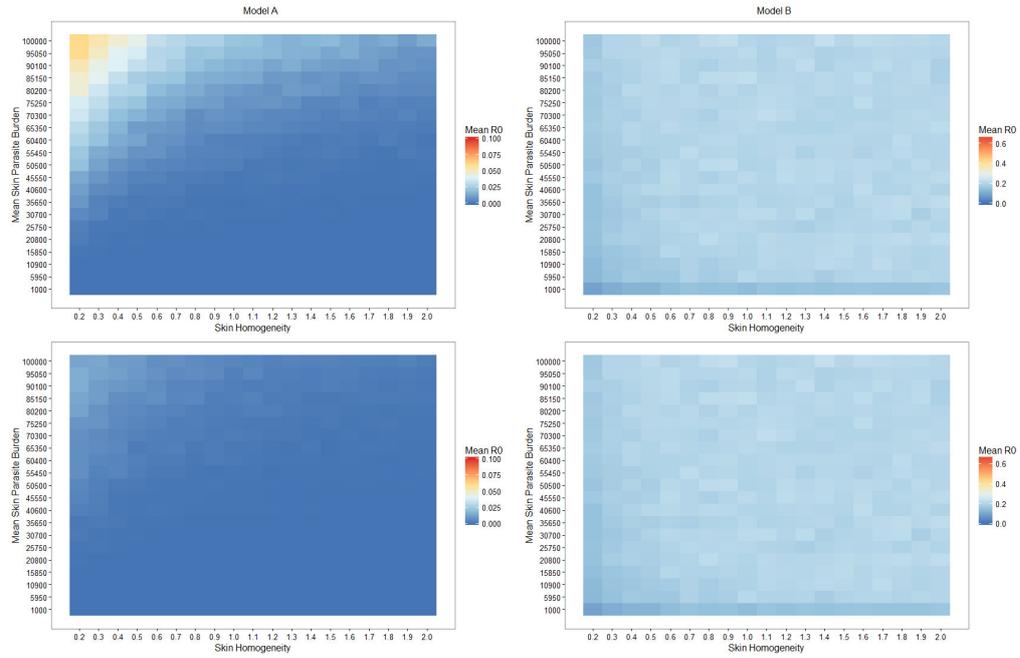


Figure A.13. The inclusion of parasite induced mortality results in quantitative, but not qualitative, changes. Heatmaps of Mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top row) or 25% (bottom row) chance of biting an infected host, with a smooth transmission threshold. After infection, sand flies receive a 20% reduction to their remaining lifespan.

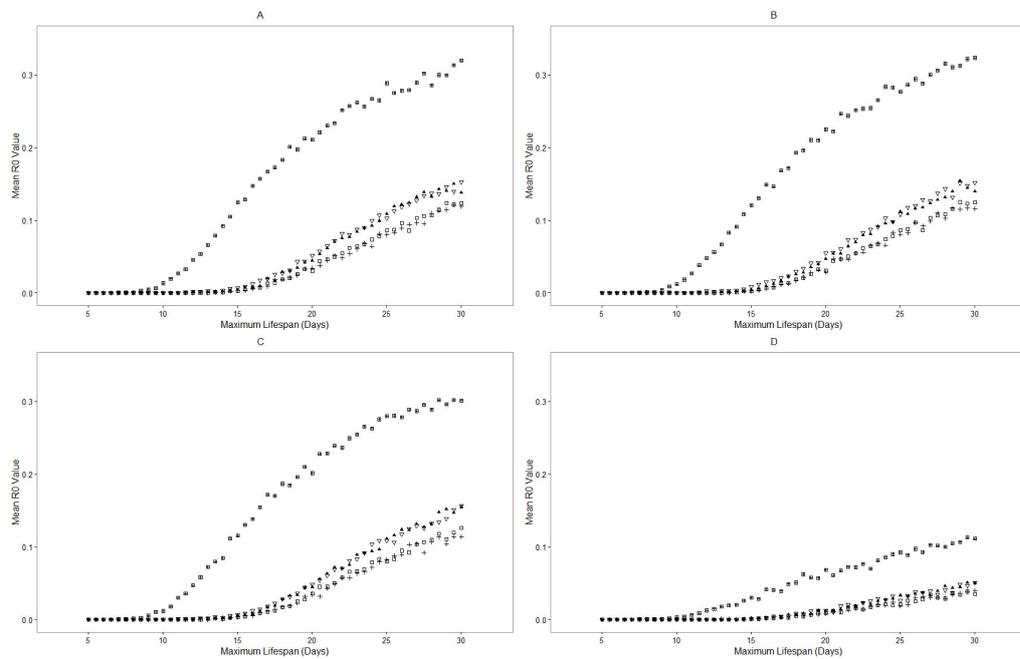


Figure A.14. Removing crucial assumptions of the model has minimal influence. Mean R_0 against maximum lifespan for a representative subsample of RAG mice. A) Full model adapted from Fig 4c. B) Full model, but with no carrying capacity. C) Full model, but with additional small population sinks. D) Full model, but with larger population sinks.

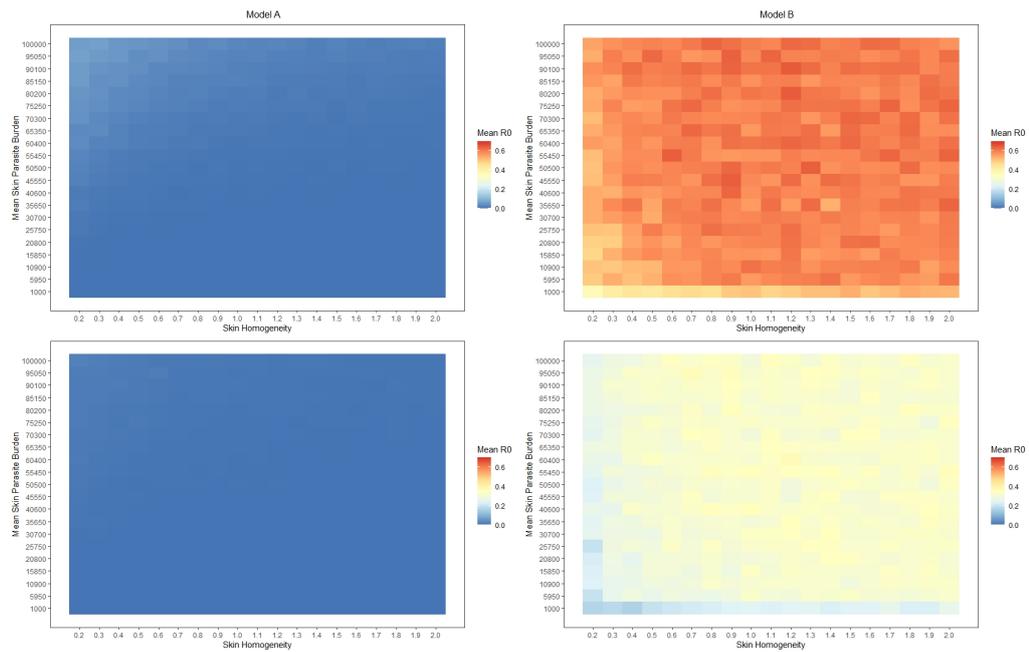


Figure A.15. Retroleptomonad dynamics dominate over skin heterogeneity and result in elevated mean R_0 values. Heatmaps of the mean R_0 for simulated sand flies for both Model A (left half) and B (right half) with 100% (top half) or 25% (bottom half) chance of biting an infected host. Each tile represents 10,000 independent simulations.

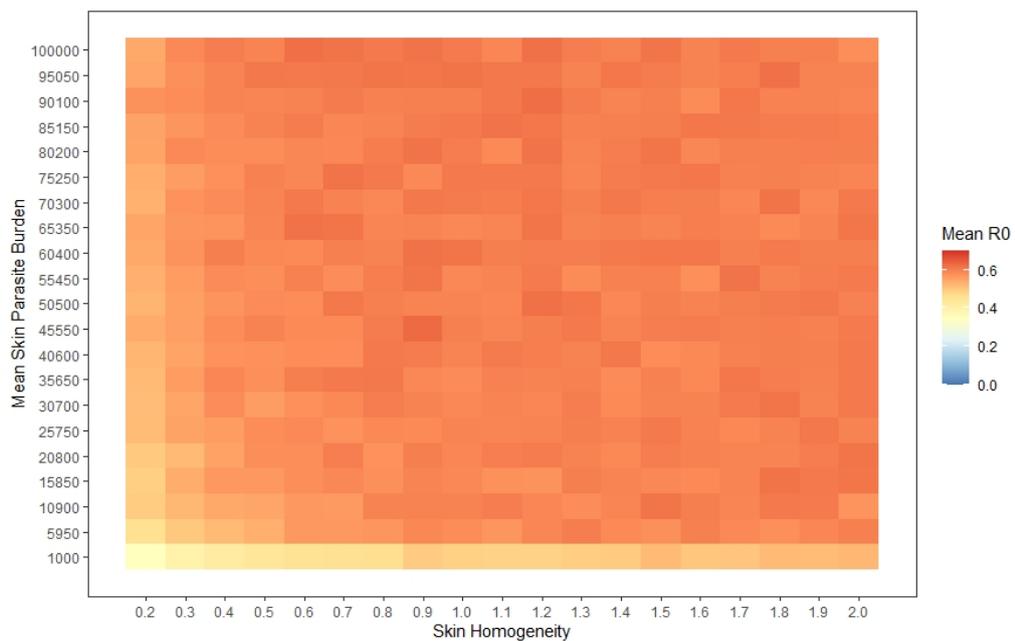


Figure A.16. A variant of the top-right heatmap of Figure 2.3, with 40000 runs per tile instead of 10000 runs per tile.

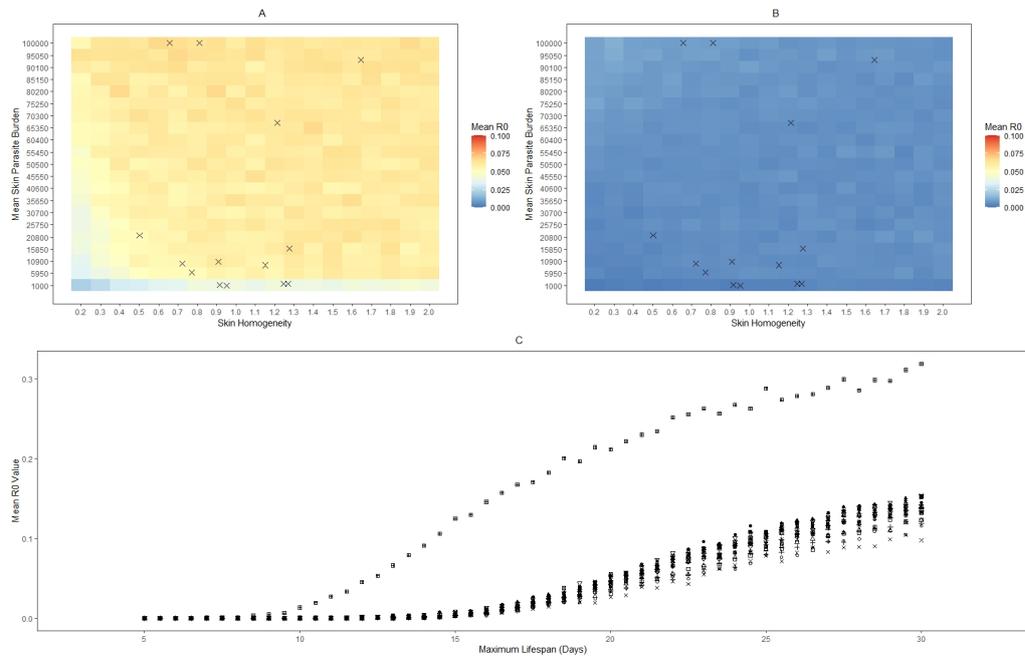


Figure A.17. Retroleptomonad dominance is dependent on having a sufficiently large maximum lifespan. A, B) Heatmaps of the mean R_0 for simulated sand flies in Model B with 100% chance of biting an infected host and with lifespans restricted to 20 days (A) or 15 days (B). Crosses indicate the mean skin parasite burden and skin homogeneity (k) of various mice from [45]. C) Mean R_0 value against maximum lifespan for RAG mice 1-18 from Doehl *et al.* [45] (S1 Table).

A.2 Chapter 3 Supplementary Materials

First, let us consider a municipality to be endemic when it exceeds 2.0 cases per 100000 on average for the previous 3 years. Table A.1 shows the fitted coefficient and p-values for the spline.

Component	Coefficient	Standard Error	Chi Squared	P-Value
Linear	1.13	0.431	6.862	0.0088
Non-linear	NA	NA	2.966	0.2375
Likelihood Ratio Score:	12.1	P-Value:	0.007	

Table A.1. Penalised spline fitting suggests a linear relationship between hazard of endemicity and overall SVI. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values, for penalised spline fitting with a threshold of 2 cases per 100000.

From the coefficient of the linear component of the spline, we would expect the hazard of endemicity to increase with the SVI of the municipality. This may be complicated by the non-linear component; however, we find this component to be non-significant and conclude that the relationship is purely linear.

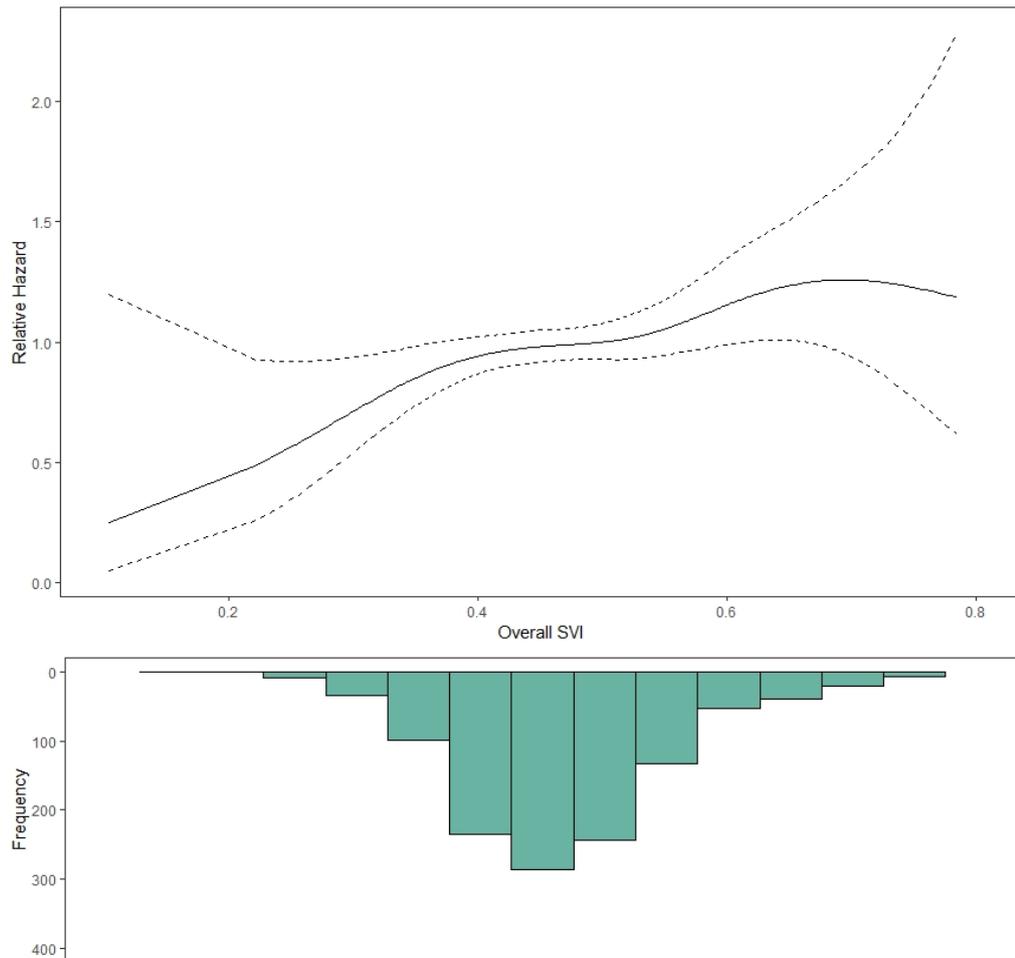


Figure A.18. Relative hazard of infection increases almost linearly with overall SVI. Relative hazard against overall SVI, with an endemic criterion of 2.0 cases per 100000. Solid line is mean hazard, dotted lines are ± 2 s.e. (top) along with a histogram of the overall SVI values for all municipalities included in the analysis (bottom).

The resulting spline is depicted in Figure A.18a. As anticipated, we see that the hazard of endemicity increases with overall SVI. The spline is almost linear, with minor deviations occurring at each end. This, along with the increased standard error, arises due to there being few municipalities with very high or low SVI values (Figure A.18b) which impacts the fitting of the spline. The overall conclusion remains unchanged: municipalities with greater social vulnerability are more like to become endemic in a given time period.

Repeating this analysis for the last threshold, 4.5 cases per 100000, results in Table A.2). As with the threshold of 3.0, the likelihood ratio test finds a significant relationship. This time, we conclude that the linear component is not significant (i.e. that the relationship is non-linear).

Component	Coefficient	Standard Error	Chi Squared	P-Value
Linear	0.674	0.473	2.032	0.154
Non-linear	NA	NA	5.257	0.075
Likelihood Ratio Score:	9.75	P-Value:	0.02	

Table A.2. For the maximum endemic threshold, penalised spline fitting suggests a non-linear relationship. Summary of spline components: coefficients and standard error (where applicable), chi-squared value and associated p-values.

We can gain a greater insight from examining the spline, as shown in Figure A.19a. This spline resembles the previous two splines: generally an increase in hazard with overall SVI, with a slight reduction at very large SVI values, and much larger standard error at the far ends of the parameter space. The main difference we observe here is that the spline more closely resembles a plateau for $SVI > 0.6$. Thus, although we can no longer conclude that the relationship would be sufficiently represented by a linear term in the model, the overall qualitative relationship between hazard and overall SVI remains the same. Given the similarities between the thresholds, we henceforth restrict our analysis to the threshold of 3.0 cases per 100000 people.

A.3 Chapter 4 Supplementary Materials

A.3.1 Supplementary Table 1

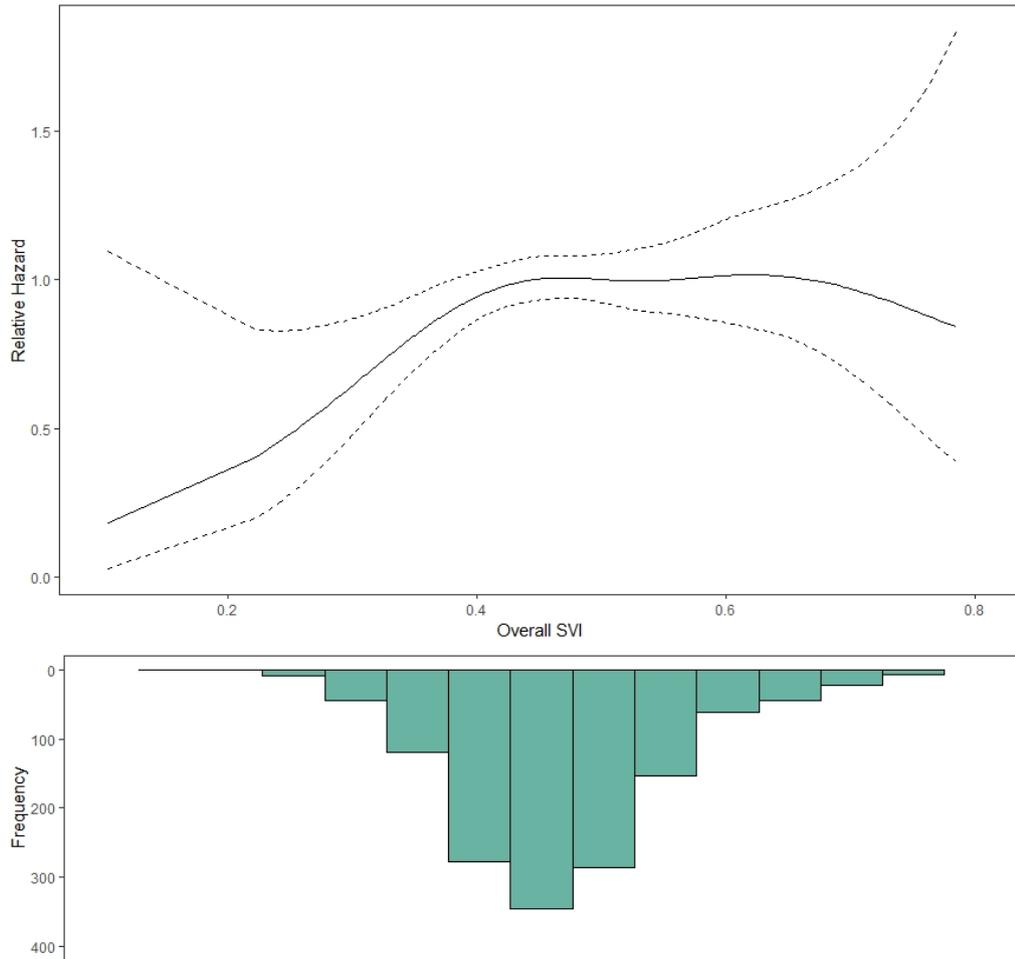


Figure A.19. The previous linear relationship remains possible despite the lack of significance in the spline fitting. Relative hazard against overall SVI, with an endemic criterion of 4.5 cases per 100000. Solid line is mean hazard, dotted lines are ± 2 s.e. (top) along with a histogram of the overall SVI values for all municipalities included in the analysis (bottom).

Table A.3. Summary of parameter definitions.

Parameter	Symbol
Transmissibility	T
Infectious Duration (Continuous Time)	τ
Infectious Contact Rate	r
Infectious Duration (Discrete Time)	λ
Per-timestep Infection Probability	ρ

A.3.2 Supplementary Methods 1: Analytic Approximations of R_0 in Continuous Time

In continuous time, we use a different form of T:

$$T = 1 - \int_0^\infty \int_0^\infty dr d\tau P(r)P(\tau) \exp^{-r\tau}. \quad (\text{A.10})$$

If we once again assume that the duration of infection τ is a constant τ^* as before, we have:

$$T = 1 - \int_0^\infty P(r) e^{-r\tau^*} dr. \quad (\text{A.11})$$

Fixed Contact Rate

For the fixed contact rate scenario, we assume r to be some fixed value for all individuals.

This simplifies T to:

$$T = 1 - e^{-r\tau^*} \quad (\text{A.12})$$

As in the discrete time case we have:

$$R_0 = T(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}}) \quad (\text{A.13})$$

Thus, for a fixed contact rate:

$$R_0 = (1 - e^{-r\tau^*})(\hat{k} - 1 + \frac{\hat{k}}{\hat{k}}) \quad (\text{A.14})$$

Uniform Random Contact Rate (Time Invariant)

We assume that r is distributed according to a random uniform variable with minimum value a and maximum value b . Thus, we have:

$$P(r) = \frac{1}{b-a} \quad \forall r \in [a, b] \quad (\text{A.15})$$

and therefore:

$$\begin{aligned} T &= 1 - \int_0^\infty P(r)e^{-r\tau^*} dr \\ &= 1 - \int_a^b \frac{e^{-r\tau^*}}{b-a} dr \\ &= 1 - \left[\frac{-e^{-r\tau^*}}{\tau^*(b-a)} \right]_a^b \\ &= 1 - \left(\frac{e^{-a\tau^*}}{\tau^*(b-a)} - \frac{e^{-b\tau^*}}{\tau^*(b-a)} \right) \\ \Rightarrow T &= 1 - \left(\frac{e^{-a\tau^*} - e^{-b\tau^*}}{\tau^*(b-a)} \right) \end{aligned}$$

Finally, we have:

$$R_0 = \left(1 - \left(\frac{e^{-a\tau^*} - e^{-b\tau^*}}{\tau^*(b-a)} \right) \right) \left(\hat{k} - 1 + \frac{\hat{k}}{k} \right) \quad (\text{A.16})$$

Bimodal Contact Rate

We once again consider the bimodal case where:

$$\begin{aligned} P(r) &= p_1 & r &= a \\ P(r) &= p_2 & r &= b \\ P(r) &= 0 & & \textit{otherwise.} \end{aligned}$$

We assume that $a \ll b$ and $p_1 \gg p_2$ as before. Thus:

$$\begin{aligned} T &= 1 - \sum_0^\infty P(r)e^{-r\tau^*} \\ \Rightarrow T &= 1 - (p_1 e^{-a\tau^*} + p_2 e^{-b\tau^*}) \end{aligned}$$

Therefore we have:

$$R_0 = (1 - (p_1 e^{-a\tau^*} + p_2 e^{-b\tau^*})) \left(\hat{k} - 1 + \frac{\hat{k}}{k} \right) \quad (\text{A.17})$$

Power Law Infection Probability

We now assume that r is distributed according to the Pareto distribution with minimum r_m s.t.:

$$P(r) = \frac{\alpha r_m^\alpha}{r^{\alpha+1}} \quad \forall r \geq r_m \quad (\text{A.18})$$

We consider first the case where $\alpha = 2$. Substituting this into our equation for T we have:

$$\begin{aligned} T &= 1 - \int_0^\infty P(r) e^{-r\tau^*} dr \\ &= 1 - \int_{r_m}^\infty \frac{2r_m^2}{r^3} e^{-r\tau^*} dr \\ &= 1 - \left(\left[\frac{-r_m^2 e^{-r\tau^*}}{r^2} \right]_{r_m}^\infty - \int_{r_m}^\infty \frac{r_m^2}{\tau^* r^2} e^{-r\tau^*} dr \right) \\ &= 1 - e^{-r_m \tau^*} - \left(\left[\frac{-r_m^2 e^{-r\tau^*}}{\tau^* r} \right]_{r_m}^\infty - \int_{r_m}^\infty \frac{r_m^2}{\tau^* 2r} e^{-r\tau^*} dr \right) \\ &= 1 - e^{-r_m \tau^*} - \frac{r_m}{\tau^*} e^{-r_m \tau^*} - \left(\left[\frac{r_m^2 \ln(r) e^{-r\tau^*}}{\tau^* 2} \right]_{r_m}^\infty + \int_{r_m}^\infty \frac{r_m^2}{\tau^* 3} \ln(r) e^{-r\tau^*} \right). \end{aligned}$$

If we now assume that r_m is small, we thus have:

$$\begin{aligned} \int_{r_m}^\infty \frac{r_m^2}{\tau^* 3} \ln(r) e^{-r\tau^*} &= \frac{r_m^2 e^{\tau^*}}{\tau^* 3} \int_{r_m}^\infty \ln(r) e^{-r} \\ \int_{r_m}^\infty \ln(r) e^{-r} &\approx \int_0^\infty \ln(r) e^{-r\tau^*} = -\gamma \\ \Rightarrow \int_{r_m}^\infty \frac{r_m^2}{\tau^* 3} \ln(r) e^{-r\tau^*} &\approx -\frac{r_m^2 e^{\tau^*} \gamma}{\tau^* 3} \end{aligned}$$

where γ is Euler's Number. Substituting this into our previous equation we have:

$$T \approx 1 - e^{-r_m \tau^*} - \frac{r_m}{\tau^*} e^{-r_m \tau^*} - \frac{r_m^2}{\tau^* 2} \ln(r_m) e^{-r_m \tau^*} + \frac{r_m^2 \gamma e^{\tau^*}}{\tau^* 3}. \quad (\text{A.19})$$