INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

DISSERTATION

CONTRIBUTIONS TO THE MATHEMATICAL THEORY OF EPIDEMIC DYNAMICS

Submitted by

Craig Willard McCarty

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

for the Degree Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 1999

UMI Number: 9950899

UMI Microform9950899

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

June 4, 1999

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY **CRAIG WILLARD MCCARTY** ENTITLED **CONTRIBUTIONS TO THE MATHEMATICAL THEORY OF EPIDEMIC DYNAMICS** BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Bruce a. Wunder

Paul W. Miller

Kenneth P. Burnham

Advisor

Dan Bradley

Department Head

ABSTRACT OF DISSERTATION

CONTRIBUTIONS TO THE MATHEMATICAL THEORY

OF EPIDEMIC DYNAMICS

I derive a novel model of disease transmission in dynamic host populations ("model MM") from mechanistic assumptions. This model addresses several concepts simultaneously: (1) infection of susceptible individuals occurs through 2 mechanisms, contact with "point" sources of infectious material resulting in a constant risk, and contact with infectious individuals resulting in additional risk that varies with the size of the infectious population; (2) that transmission may cross species boundaries; (3) that population size can effect risk; and, (4) that individual covariates affect risk. I demonstrate this model's theoretical application by forecasting possible outcomes of a bovine tuberculosis (*Mycobacterium bovis*) epidemic in a white-tailed deer (*Odocoileus virginianus*) population using data from a recent epidemic in Michigan. I conclude that the best use of epidemic forecasting exercises is to identify critical gaps in knowledge for future research.

Next, I propose 50 additional candidate models of the epidemic process. I present the structure of these models by first defining 4 general classes of epidemic models: multinomial, structured multinomial, nested logistic regression, and biological-mechanistic. I then challenge these models using 2 classic data sets of binomial chain epidemic data from the

literature (measles, "common cold") and evaluate the bias-variance tradeoff of each model/data set combination using Akaike's Information Criterion (AIC). Model selection results are not consistent between data sets and no universal model emerges. This conclusion raises questions about the rational of invoking popular epidemic models in attempts to explain or predict specific host/parasite dynamics. I then use Monte Carlo methods to evaluate the ability of AIC and it's Bayesian compliment, BIC, to detect model MM as the generating process of data of the same sample sizes and similar rates of infection as found in these 2 data sets. BIC outperformed AIC when the generating process was simple and in the candidate set of models. I conclude that sample sizes were insufficient to strongly rule out model MM as the underlying the generating process.

Finally, I summerize the literature on epidemic dynamics. I conclude that many fundamental questions remain open to debate. In response, I make several recommendations to direct future research.

Craig Willard McCarty
Graduate Degree Program in Ecology
Colorado State University
Fort Collins, CO 80523
Fall 1999

ACKNOWLEDGMENTS

A large number of people made significant contributions to this dissertation either directly by comment on its content and text or indirectly by guiding the attitude and intellect of the author; some did both. Unfortunately, I can only name and briefly describe the contributions of a fraction of these within the time and space limitations of this summary. I apologize to those who, through my own failing, were not recalled at the time of this writing.

First, I recognize the contributions of my advisor: Professor Kenneth P. Burnham. Ken was responsible for providing the academic and intellectual environment that made this work possible; without Ken, this dissertation would not exist.

Second, I recognize the contributions of my dissertation committee: Professor Bruce A. Wunder, Professor Paul W. Mielke, Jr., and Dr. Michael W. Miller. Bruce repeatedly challenged me as a student and provided numerous examples of the critical analysis of biological problems. This philosophy, demonstrated weekly at his graduate student meetings, encouraged the development of my own philosophy of biomathematics that I hope is apparent throughout this dissertation: "Always apply the mathematics to the biology and never the biology to the mathematics." Paul Mielke deserves recognition as the most positive attitude of my 8 year career as a graduate student. Paul also deserves special thanks for successfully contributing to the education of two generations of McCarty's. Finally, but not least, Mike

Miller's imagination was responsible for the initiation of this project and his patience and resourcefulness for its completion. Mike is also one of my best friends.

Third, I thank all those whose skill and patience as teachers are responsible for my training as an ecologist and mathematician. For my quantitative training, I especially thank Professors Richard A. Davis, Paul W. Mielke, Jr., Hari K. Iyer, Kenneth P. Burnham, Duane C. Boes, Alan J. Branford, Mohammed M. Siddiqui, S. Nandagopalan (Nandu), Yi-Ching Yao, David C. Bowden, Jennifer A. Hoeting, Frederick A. Johnson and Bennet Manvel. From Professors Mowafak D. Salmon and John S. Reif I gained a valuable appreciation for mathematical epidemiology. For contributing to my education as an ecologist, I specifically thank Professors David R. Anderson, Gary C. White, James A. Bailey, John E. Gross, Patricia L. Kennedy, Debra P. Coffin, and A. William Alldredge.

Fourth, I recognize the support of my friends. I especially thank Mary M. Conner, Sara J. Oyler-McCance, Gail S. Olson, Kirsta L. Scherif-Nelson, Dr. William L. Thompson, Dr. Alan B. Franklin, and Dr. Tanya M. Shenk. Also, I thank all of the patient graduate students who endured my weekly math/stat. lectures. Of these, several deserve special recognition for giving me positive memories that will not soon fade: Lindsey Christensen, Bradford B. Gersey, Melissa N. Howard, Kris L. Metzger, Stewart W. Breck, and Geneva W. Chung.

Fifth, I recognize the computer scientists responsible for the Linux operating system and the Linux documentation project (http://metalab.unc.edu/LDP); while their contributions were not direct, they nevertheless made this work run smoothly and shielded me from the "dark side". I especially thank Linus Torvalds for fathering Linux; the computing world

definitely needs more people of his philosophy and character. Also, I thank S. I. Feldman, and P. J. Weinberger for their FORTRAN to C translator (A portable FORTRAN 77 compiler *in* Unix time sharing systems programers manual, 10th ed., AT&T Bell Laboratories, 1990) and the unknown individuals who ported it to Linux.

Sixth, I acknowledge my funding sources: The Colorado Division of Wildlife, The College of Veterinary Medicine and Biomedical Sciences, and AWM Enterprises. Special thanks go to several employees of the Colorado Division of Wildlife who provided me with secondary employment and valuable practical experience: Dr. Daniel L. Baker, David J. Freddy, Dale F. Reed, Dr. Margaret A. Wild, and Nancy Wild-McEwen.

Finally, I thank my wife, Kimberly Paull McCarty, for her support and tolerance even when I did not deserve it. She gives meaning to my life.

Craig Willard McCarty
Saint George's University School of Medicine
Grenada, West Indies
April 10, 1999

To my parents, Robert Darrel and Liesolotte McCarty; my wife, Kimberly Paull McCarty; and my sisters, Sandra Lee Casselberry, and Roxie Gene Leitner.

TABLE OF CONTENTS

INTRODUCTION

The goal of this dissertation is to present a critical analysis of the classic epidemic problem: "How do we model the transition from the susceptible to the infected state?" Three general concepts are pervasive throughout as I address this question from different perspectives in successive chapters. First, I emphasize the distinction between mechanistic and statistical models. Second, I emphasize the relationship of data to model projection, forecasting, and model fitting and selection (i.e., respectively, describing the logical consequences of a model, predicting the future, and selecting the "best" approximating model for a data set). Third, I emphasize the importance of entertaining multiple *a priori* models and hypotheses.

In chapter 1, I derive a novel model of disease transmission in dynamic host populations and demonstrate its application by forecasting possible outcomes of a bovine tuberculosis (*Mycobacterium bovis*) epidemic in a white-tailed deer (*Odocoileus virginianus*) population. My approach is mechanistic and bases disease transmission on the probability of each susceptible individual becoming infected per unit time, and affords the flexibility necessary to model epidemics in dynamic wildlife populations. I conclude that perhaps the best use of epidemic modeling exercises is to uncover areas of uncertainty in our knowledge and suggest testable hypotheses.

In chapter 2, I define and compare 4 classes of epidemic models (41 models in total) using Akaike's information criteria (AIC)-based model selection and 2 data sets (measles,

"common cold") from the literature. The 4 classes included multinomial, structured multinomial, nested logistic regression, and biological-mechanistic models. In all, 52 possible model-data set combinations are fit and AICc values calculated. Model selection results are not consistent between data sets and no universal model emerges; only six models were highly plausible ($\Delta AICc \leq 4$) candidates for at least one of the data sets. Although I detected some evidence of underlying structure, neither data set suggests that the number of infectious or susceptible individuals, population size, or time strongly affected epidemic dynamics. My findings underscore the need for replicated data and experimental evidence to support selection of any particular epidemic model. The results of my analyses raise questions about invoking popular epidemic models in attempts to explain or predict specific host/parasite dynamics.

In chapter 3, I use Monte Carlo methods to evaluate the ability of AIC and it's Baysian complement, BIC, to select the model introduced in chapter 1 as the generating process of epidemic binomial chain data of the same sample sizes and similar rates of infection as found in the 2 data sets analyzed in chapter 2. As expected, BIC selected the generating process more frequently than AIC in all scenarios in which the generating process was simple and in the set of candidate models. My results do not support the notion suggested in chapter 2 that the selection of a biologically unreasonable model by AIC as the best approximating model for the one of these data sets ("common cold") is an unfortunate artifact of sample size.

I conclude the dissertation by discussing the complications of quantifying risk in epidemic models and reviewing current literature (emphasis on 1/1/98 - 4/30/99). Concepts addressed include spatial and temporal factors, individual and population heterogeneity, transmission, environmental effects, simultaneous epidemics, and multiple avenues of risk.

I provide many examples of each factor. I conclude that despite the volume of literature on epidemic modeling, many fundamental questions remain open to debate and many potentially fruitful avenues of expansion are open. In response, I make several suggestions to direct future research in this area: (1) whenever possible, research should emphasize experimental challenges of current and new theories of epidemic dynamics under the paradigm of multiple working hypotheses; (2) when true experiments are not possible, quasi-experiments and correlative studies should adopt information theory based analyses and entertain several *a priori* models; (3) epidemic modeling and the concept of "epidemic" should be expanded to include the propagation of other phenomenon through populations such as ideas, fads, religions, computer viruses, product specific accidents, cancer, poisoning, substance abuse, mental illness, eating disorders, child abuse, and domestic violence; and, (4) several principles of ecology deserve more attention as factors in epidemic dynamics: convergent and divergent evolution, interspecific competition, fragmentation, metapopulations, source and sink populations, and island biogeography.

CHAPTER 1

A Versatile Model of Disease Transmission Applied to Forecasting Bovine Tuberculosis Dynamics in White-tailed Deer

SUMMARY

We derived a model of disease transmission in dynamic host populations and demonstrated its application in forecasting possible outcomes of a bovine tuberculosis (*Mycobacterium bovis*) epidemic in a white-tailed deer (*Odocoileus virginianus*) population. Our approach was mechanistic, based disease transmission on the probability of each susceptible individual becoming infected per unit time, and afforded the flexibility necessary to model epidemics in dynamic wildlife populations. We applied this approach to a sex- and age-structured deer population model. Our model predicted that tuberculosis prevalence in a white-tailed deer population could rise from about 3% to about 21% over 25 years, and that neither lowered deer survival nor lowered transmission would be completely effective in eliminating disease from the population. Maternal transmission appeared unimportant to modeled tuberculosis dynamics; in contrast, disease was not maintained for more than15 years in models lacking lateral transmission.

Key Words: Bovine tuberculosis; epidemic modeling; modeling; epidemiology; white-tailed; deer *Mycobacterium bovis*; *Odocoileus virginianus*.

1.1. Introduction

Modeling is an important tool for describing and predicting the dynamics and outcomes of a wide variety of ecological processes, including infectious disease transmission (Starfield and Bleloch, 1986; Grenfell and Dobson, 1995). Epidemic models have been used to investigate the role of disease in population processes, to compare disease management strategies, and to assess risk of disease transmission within and among species (Anderson and May, 1979, 1982; Grenfell and Dobson, 1995; Barlow, 1996).

Dynamic wildlife populations present unique challenges to traditional epidemic modeling approaches. A combination of features, including environmental or demographic stochastisity, nonconstant survivorship, differential susceptibility or infectiousness, individual covariates, and density-dependence may be required to realistically portray epidemics in free-ranging wildlife populations. The differential equation approach to epidemic modeling (e.g., Anderson and May, 1979) becomes mathematically intractable when such features are incorporated (e.g., Bailey 1975; Heesterbeek and Roberts, 1995). Similarly, the Reed-Frost approach is limited by parameters that are conditional on a fixed population size throughout the duration of the epidemic (Bailey, 1975; Becker, 1981); the latter assumption may be appropriate in stable human populations, but is of limited utility in applications to fluctuating wildlife populations. In addition to these difficulties, few epidemic models can be readily parameterized from data typically provided by wildlife studies. Consequently, models depicting chronic epidemics spanning many generations of wildlife hosts are relatively uncommon (Barlow, 1995). Despite these challenges, wider availability of epidemic models could be useful to those responsible for managing wildlife disease problems.

Herein, we describe a versatile mechanistic model of disease transmission and demonstrate its application in forecasting possible outcomes of a bovine tuberculosis *(Mycobacterium bovis)* epidemic in a dynamic white-tailed deer (*Odocoileus virginianus*) population.

1.2. General Mathematical Approach

We represent the transition from the susceptible to the infected state with a simple mechanical model. Two assumptions are necessary for this model to apply. First, we assume the infectious subpopulation produces some number of infectious contacts per member (i) per unit time (j), denoted as $\beta_{i,j}$. We define an infectious contact as any interaction between an infectious individual and any other individual that would result in disease transmission if the other individual were susceptible. Second, we assume all individuals within the host population have equal or known probabilities of contacting any infectious individual per unit time. The probability of any one susceptible individual being infected by one or more of the I members of the infectious subpopulation per unit time ($P_{(S-I)}$) is calculated as

$$P_{(S-I)} = 1-(1-P)^{I}, \tag{1.1}$$

where P is the probability of any one susceptible individual becoming infected by receiving one or more infectious contacts from a single infectious individual per unit time. We calculate the latter probability as:

$$P = 1-\left(1-\frac{1}{N}\right)^{\beta}, \tag{1.2}$$

where β is the number of infectious contacts per infectious individual per unit time and N is

the total population size. Here, (1-1/N) is the probability of the susceptible individual not receiving a single infectious contact from that single infectious individual. Substituting equation 1.2 into equation 1.1 yields the probability of a susceptible individual becoming infected per unit time:

$$P_{(S-I)} = 1 - \left(1 - \frac{1}{N}\right)^{I\beta}. \qquad (1.3)$$

Our model has two key features. First, disease transmission is driven by the number of potential infectious contacts made by an infectious individual during a given time step (β). Because these contacts are randomly allocated across the population, some may be "wasted" on individuals that are immune or are already infected. This parameter is measurable. Moreover, β itself can be modeled to reflect changes in the nature of interactions within the host population, between host species, and with the environment. Second, this model recognizes that host population size also influences the probability of interactions between individuals during a given time step. By incorporating total population size into estimating the probability that a susceptible individual will become infected, this model allows populations to fluctuate without compromising baseline assumptions about disease transmission. It follows that three factors act in concert to affect the probability that any susceptible individual becomes infected during any given period of time: the number of infectious individuals, the number of infectious contacts each is capable of producing, and the size of the population where potential interactions may occur.

1.3. Application to Bovine Tuberculosis Epidemiology

In 1994, tuberculosis was first detected among free-ranging white-tailed deer in northeastern Michigan; subsequent investigations revealed that tuberculosis was well-established in the affected deer population (Schmitt et al., 1997). Because this epidemic potentially threatened local cattle herds and presented significant obstacles to long-term wildlife resource management, Michigan's Departments of Agriculture and Natural Resources (MDNR) requested an assessment of various associated risks (United States Department of Agriculture, unpubl. report). Using the mathematical approach outlined above, we constructed the following epidemic model as a foundation for the quantitative components of subsequent risk assessments.

1.4. Modeling Methods

We modeled epidemic and host population dynamics as a multivariate Markov process (Sharpe, 1988), with each dimension representing the number of individuals in a particular sex, age, and health class. Transitions between consecutive vectors were governed by a series of simple rules determined by population dynamics of the host species (analogous to the difference equations employed in matrix population models; Caswell, 1989) and host-parasite interactions.

Initially, we constructed a deterministic white-tailed deer population and tuberculosis epidemic model generally fitted to trends observed in the affected deer population in Michigan over the last 40 years (Fig. 1.1). This preliminary model served two purposes. First, it ensured reasonable representation of historic trends in population and tuberculosis dynamics observed in the affected white-tailed deer population. Second, it provided estimated

transmission coefficients for use in subsequent projections of epidemic trends. Our model featured two sexes (male, female), two age classes (fawn, adult), and four health states (susceptible, first and second year of incubation, infectious), and operated on an annual time step. Within each time step, fawns were recruited, maternal and lateral tuberculosis transmission occurred, yearlings and adults of all health states were removed via hunting and natural mortality, and the disease progressed in incubating animals. Yearling and adult survival rates were fixed (male = 0.45, female = 0.75) to produce male:female ratios (about 1 male:2 females before annual mortality and about 1 male:4 females after mortality) that matched field observations (MDNR, unpubl. data). To mimic processes driving white-tailed deer population dynamics (e.g., McCullough, 1979), fawn recruitment rates (the combination of fawn birth and survival rates) were adjusted annually (range 0.35 to 0.7) to fit the simulated population to historic trends (range about 7,600 to about 20,000 deer; Fig. 1.1)(MDNR, unpubl. data). We assumed recruitment and adult survival rates were reduced by both hunting and natural mortality, but their effects were combined in the model; we further assumed no compensation in recruitment to counterbalance reduced survival (e.g., Bartmann et al., 1992). Because average deer life spans were relatively short under these survival regimes (about 2 years for males, about 4 years for females), and because severe, disseminated tuberculosis was rarely observed in affected individual white-tailed deer (Schmitt et al., 1997), we assumed tuberculosis did not affect recruitment or survival.

Health states reflected exposure to and duration of infection with *M. bovis*. We modeled the transition from the susceptible to the infected state using the mechanical model described in equation 1.3. We assumed all deer in the population were equally susceptible to tuberculosis prior to infection. However, based on sex-related differences in prevalence

initially observed in the affected population (males = 0.08, females = 0.02; S. Schmitt, pers. comm.), we assumed that males were somehow more likely than females to contact an infectious individual at each time step. We allowed fawns born to infectious females additional opportunity to become infected via maternal transmission (T_{MAT}).

Once susceptible deer became infected, we assumed a 2 year incubation period before infected deer could transmit *M. bovis* to other deer; based on descriptions of lesion size and distribution in white-tailed deer (Schmitt et al., 1997), it appeared to us that deer in early stages of infection would be unlikely to shed infectious doses of *M. bovis*. During the second year of incubation, we assumed tuberculosis infections would be detectable by conventional postmortem diagnostic approaches (e.g., Schmitt et al., 1997). Infectious deer remained infectious for life, each transmitting disease at set rates expressed as infectious contacts per time step between infectious individuals and females and fawns (β_F) and between infectious individuals and adult males (β_M). Females were additionally capable of transmitting infection to their fawns at a fixed rate ($T_{MAT} = 0.25$). We assumed no immunity to or recovery from tuberculosis once deer were infected.

At each time step, the probability of a susceptible deer entering the first year of incubation was determined by equation 1.3; the total number of first year incubators was the sum of deer newly infected via lateral transmission and via maternal transmission. The probabilities of deer moving from the first to second year of incubation and from the second year of incubation to the infectious class were sole functions of sex- and age-specific survival rates; consequently, the overall probability of infected deer surviving to become infectious was about 0.20 for males and about 0.56 for females. The model tracked total numbers of deer in each health state at each time step, and also calculated prevalence estimates: we

defined the numerator for prevalence as the total number of infected deer detectable at postmortem exam (= second year incubating + infectious) to allow comparison to field data, and used total population less fawns as the denominator. The model also tracked the total number of infected deer (= first year incubating + second year incubating + infectious).

Initial conditions were set such that population size was about 15,000 deer and increasing. We apportioned the initial population using observed sex ratio data (about 33% males:67% females before annual mortality; MDNR, unpubl. data) and adjusted fawn recruitment rates to increase and decrease population size to track historic trends (MDNR, unpubl. data). We then added one infectious female to the population, and fit the model to estimate transmission coefficients necessary to yield observed prevalences of about 2.4% in ≥ 1-year-old females and 8.1% in ≥ 1-year-old males (3.1% overall for ≥ 1-year-old deer) 40 years later (Fig. 1.1). These transmission coefficients (β_F = 0.5, β_M = 8.1) were used as baseline epidemic conditions in our projection models.

Next, we allowed stochastic processes to influence the dynamics of the population and the epidemic to explore the range of possible future trajectories. Initial conditions for forecasting were taken from year 40 (= 1995) of the deterministic model. We used Monte Carlo methods (Ross, 1997) to investigate the range of possible outcomes and to derive variance estimates of relevant functions of the population vector (e.g., yearly prevalence and probability of transition from susceptible to infected). Alternative management scenarios also were modeled. As a baseline for comparisons, we left parameters describing population and epidemic dynamics unmodified from the deterministic model and projected progress of tuberculosis in the simulated herd for 30 years. We then considered the effects of two types of management: the first reduced survival rates, and the second reduced transmission; we

examined these alone and in combination.

We also ran simulations to examine model sensitivity and effects of temporal variation on model outcomes. Because the adult survivorship of deer in North America is relatively stable but recruitment rates vary widely across time both within and among populations (McCullough, 1979; Bartmann et al., 1992), we examined effects of temporal variation in recruitment for each scenario by modeling yearly recruitment as a Beta random variable (Ross, 1997) with mean = 0.48 and 80% of the probability mass between 0.20 and 0.75. Further, scenarios were examined with and without maternal transmission.

For each case of each scenario, we generated 1,000 possible sample paths. FORTRAN source code for these calculations is available from the authors upon request (see appendix for a more quantitative description).

1.5. Modeling Results

Under our deterministic model's survival and transmission assumptions, the basic reproductive ratio (R_0; Heesterbeek and Roberts, 1995) of tuberculosis in white-tailed deer was >1.0. Consequently, our stochastic model predicted that, given the foregoing assumptions, there is a high likelihood that tuberculosis will persist in the infected white-tailed deer population in Michigan. Model outcomes also illustrated the potential effects of prospective management interventions on the predicted trajectory of this epidemic (Figs. 1.2-1.3). In general, reducing survival was predicted to be less effective in reducing prevalence than reducing transmission until the lower threshold for population persistence was reached (Fig. 1.2); reducing both survival and transmission appeared more effective than either approach alone. For example, the model predicted that prevalence may rise to about 21%

over a 25 years period in the absence of management intervention (Fig. 1.3a). Reducing adult survival by 10% projected drastic population declines over 25 years, from 7,926 to 1,105 deer on average (standard deviation = 71; range 886 to 1,293,) (Fig. 1.3b). Lowering adult survival by 10% appeared slightly less effective in reducing predicted prevalence than lowering transmission coefficients by 10%, but lowering survival reduced the average number of infected deer to about one fifth the average number predicted by lowering transmission (Figs. 1.3b, 1.3c). The model predicted that reducing both survival and transmission by 10% would diminish both prevalence and total numbers of infected animals (Fig. 1.3d). Reducing survival beyond 70% of historic rates eliminated both the epidemic and the population by year 25 (Fig. 1.2).

Temporal variation in recruitment increased the uncertainty of observed prevalence dramatically: standard deviations increased 3.8- to 4.5-fold. However, the epidemic still persisted in all sample paths of most scenarios. Because of low recruitment rates and the long incubation period of bovine tuberculosis, maternal transmission was not a significant factor in modeled epidemics: reducing maternal transmission to zero had virtually no effect on model outcomes. In contrast, tuberculosis was not maintained for more than 15 years in simulated populations in the absence of lateral transmission.

1.6. Discussion

We believe it useful to distinguish two broad categories of modeling: model fitting and selection, and forecasting. We consider fitting and selection as a branch of statistics concerned with the efficient use of data when inferences are limited to the process that created those data. Here, the usual goal is to select a model, from among candidate models,

that achieves the best compromise between bias and precision (Burnham and Anderson 1992). In contrast, forecasting is predicting outcomes in the absence of data generated from the process being investigated (Caswell, 1989). Thus, all models of future events (in our context, modeling the possible trajectories of epidemics) are exercises in forecasting. The foregoing description of our deer-tuberculosis model illustrates several important facets of a general approach to epidemic modeling: model selection, fitting, and forecasting. As a final step, we encourage careful appraisal of model assumptions and outcomes. We believe this last step to be perhaps the most useful product of any modeling exercise.

Our deer-tuberculosis model illustrates the importance of appraising model outcomes. For example, our model predicted that tuberculosis is likely to persist in the affected deer population for >25 years under virtually any plausible management strategy and that even a substantial increase in deer harvest is unlikely to eliminate tuberculosis without imperiling long-term population survival. Although the model forecast combined reductions in transmission and survival as the most effective management strategy, no proven methods for such intervention have been identified. It is conceivable that severely reducing adult survival (e.g., >10%) will also reduce tuberculosis transmission, but we can offer no unequivocal support for this prediction. Also, because the effects of increased harvest on deer social structure are largely unknown, reducing the overall number of deer may not reduce average group size or social interactions per unit time; in the absence of such reductions, culling may be relatively ineffective as a disease management tool (Barlow, 1996). Clearly, management experiments designed to test these predictions are warranted.

Assumptions about disease transmission also deserve retrospective scrutiny, as our deer-tuberculosis model illustrates. As in all epidemic models, prevalence dynamics in our

model were driven by transmission coefficients, the true values of which remain unknown. We estimated these coefficients by fitting the deterministic model to our and others' perceptions of both population and disease trends. One perception influencing transmission coefficients was that of the epidemic's historic duration. We were forced to choose a somewhat arbitrary date for the epidemic's origin because no one actually knows when bovine tuberculosis was first introduced into the deer population in Michigan, or by what means that introduction occurred. Available information strongly suggested that *M. bovis* had been enzootic in the affected deer population for at least 20 years: in 1975, a seemingly isolated case of bovine tuberculosis was diagnosed in a hunter-killed deer from this same population (Schmitt et al., 1997). We believed it highly unlikely that this was the affected population's first tuberculosis case. Several plausible sources of infection for free-ranging deer existed in Michigan about four decades ago: bovine tuberculosis occurred in cattle, and in captive exotic and white-tailed deer in Michigan in the 1950's (Ferris et al., 1961, Towar et al., 1965). Consequently, we believed 1955 (40-years-ago) was a reasonable but conservative estimation of the epidemic's origin. Observed data generally support this assumption. A more recent introduction (e.g., 5- to 10-years-ago) would have required considerably higher transmission rates to elevate prevalence to levels observed in 1995 (Fig. 1.4). Using such transmission rates, our model projected that prevalence should now be doubling annually; the latter has not been observed (Schmitt et al., 1997; MDNR, 1998)(Fig. 1.3). Although more distant introductions (e.g., 50- to 70-years-ago) seemed feasible (Fig. 1.4), such scenarios might have tended to underestimate transmission rates and subsequently overestimate potential efficacy of management interventions. Because these assumptions have such profound effects on epidemic dynamics, experimental data on or independent estimates of tuberculosis

transmission among white-tailed deer would greatly enhance the reliability of future modeling efforts.

The violation of many other less obvious model assumptions also could affect forecasts. For example, although possible (Morris and Pfeiffer, 1995), for lack of evidence we assumed that no secondary infectious reservoir existed or will exist among sympatric species. The presence of a secondary reservoir could cause our model to underestimate future prevalence and probability of persistence. Similarly, an environmental source of infection (e.g., contaminated food sources; Schmitt et al., 1997) also could cause the model to underestimate prevalence and persistence in the future. To address such possibilities, equation 1.3 could be modified to include a parameter representing constant background risk posed by an additional source of infection:

$$P_{(S-I)} = 1 - \alpha\left(1 - \frac{1}{N}\right)^{I\beta}, \qquad (1.4)$$

where α is the probability of <u>not</u> becoming infected via contact with the environmental source of infection during each time step (appendix; note that when $\beta = 0$, all transmission comes from environmental sources). It is also conceivable that β is density-dependent. For lack of evidence, our model assumed tuberculosis transmission coefficients were density-independent. As a result, the probability of disease persistence may have been overestimated.

Assumptions about population structure and white-tailed deer behavior could also affect model outcomes. Based on local perceptions and anecdotal evidence, we modeled the study population as closed. Emigrating infected deer, possibly searching for new sources of food after cessation of supplemental feeding, could cause new epidemics in neighboring populations. Also, because white-tailed deer groups are matriarchal (McCullough, 1979;

Lagory, 1986; Nelson and Mech, 1987), estimates of transmission parameters made under a random mixing assumption may be unsuitable for forecasting. Disruption of matriarchal groups via increased harvest or other processes could affect tuberculosis transmission in ways not accounted for in our model, causing the model to either over- or underestimate prevalence trends. Moreover, lack of spatial structure in our model may have oversimplified disease dynamics by neglecting potential heterogeneity of tuberculosis distribution and prevalence (e.g., Cheeseman et al., 1981; Barlow, 1991); as a result, culling focused in high prevalence areas could be more effective than predicted by our model. Despite its potential shortcomings, we believe this model has at least served to raise questions about tuberculosis epidemiology in white-tailed deer that may stimulate further empirical investigations.

In conclusion, we suggest equation 1.3 as a plausible general model of the transition from the susceptible to the infected state. This model offers several desirable features. For example, β has a mechanistic interpretation: in equation 1.3, β is the mean number of infectious contacts between infectious individuals and any other member of the host population per unit time. It follows that values for β could be measured via controlled experiments. Unlike the transmission parameter of the Reed-Frost and Greenwood models, β is independent of population size. However, by holding N constant across time steps, equation 1.3 simplifies to the Reed-Frost and Greenwood models (Fine, 1977), the only epidemic models to date that have been subjected to rigorous model fitting and selection (Bailey 1975; Becker, 1981). In the context of individual-based models (DeAngelis and Gross, 1992), equation 1.3 can be interpreted as the probability of a susceptible individual receiving one or more of the available infectious contacts ($I\beta$) per time step. This model can be adapted from discrete to continuous time, although the resulting differential equations are

somewhat intractable (appendix). Finally, we believe our approach addresses at least two of the needs identified in Barlow's (1995) critical review of wildlife disease modeling: equation 1.3 is relatively simple, and it allows reconciliation of deterministic and stochastic models.

ACKNOWLEDGMENTS (*chapter 1*)

This work was supported by Federal Aid in Wildlife Restoration Project W-153-R. We thank S. Schmitt, E. Carlson, and T. Carlson for access to unpublished MDNR data and assistance with model development; B. Corso, B. Wagner, D. Norden, H. S. Hurd, C. Bruning-Fann, L. Paisley, O. Williams, and M. Chaddock also provided useful suggestions during the modeling process. J. Gross, T. Shenk, R. Bartmann, V. Apanius, and two anonymous reviewers provided valuable reviews of earlier manuscript drafts.

REFERENCES (*chapter 1*)

Anderson, R. M. and May, R. M. (1979). *Population biology of infectious diseases: Part I. Nature* **280**, 361-367.

Anderson, R. M. and May, R. M. (1982). *Population biology of infectious diseases*, Life sciences research report # 25. Berlin: Springer-Verlag, Berlin.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*, 2nd Edition. London: Charles Griffin & Co.

Barlow, N. D. (1991). A spatially aggregated disease/host model for bovine Tb in New Zealand possum populations. *Journal of Applied Ecology* **28**, 777-793.

Barlow, N. D. (1995). Critical evaluation of wildlife disease models. In *Ecology of infectious disease in natural populations*, B. T. Grenfell and A. P. Dobson (eds), 230-259. Cambridge: Cambridge University Press.

Barlow, N. D. (1996). The ecology of wildlife disease control: simple models revisited. *Journal of Applied Ecology* **33**, 303-314.

Bartmann, R. M., White, G. C. and Carpenter, L. H. (1992). Compensatory mortality in a Colorado mule deer population. *Wildlife Monographs* **121**.

Becker, N. (1981). A general chain binomial model for infectious diseases. *Biometrics* **37**, 251-258.

Burnham, K. P. and Anderson, D. R. (1992). Data-based selection of an appropriate biological model: The key to modern data analysis. In *Wildlife 2001: Populations*, D. R. McCullough and R. H. Barret, (eds). London: Elsevier Science Publishers.

Caswell, H. (1989). *Matrix population models, construction, analysis, and interpretation*. Sunderland, Massachusetts: Sinauer Associates.

Cheeseman, C. L., Jones,G. W., Gallagher, J., and Mallinson, P. J. (1981). The population structure, density and prevalence of tuberculosis (*Mycobacterium bovis*) in badgers (*Meles meles*) from four areas of south-west England. *Journal of Applied Ecology* **18**, 795-804.

Deangelis, D. L. and Gross, L. J. (1992). *Individual based models and approaches in ecology: Populations, and ecosystems*. New York: Chapman Hall.

Ferris, D. H.,. Beamer, P. D, Alberts, J. O., and Trainer, D. (1961). Tuberculosis in transported deer. *Journal of the American Veterinary Medical Association* **138**, 326-328.

Fine, P. E. M. (1977). A commentary on the mechanical analogue to the Reed-Frost epidemic model. *Journal of Epidemiology* **106**, 87-100.

Grenfell, B. T. and Dobson, A. P. (eds). (1995). *Ecology of infectious disease in natural populations*. Cambridge: Cambridge University Press.

Heesterbeek, J. A. P. and Roberts, M. G. (1995). Mathematical models for microparasites of wildlife. In *Ecology of infectious disease in natural populations*, B. T. Grenfell and A. P. Dobson (eds). Cambridge: Cambridge University Press.

Lagory, K. E. (1986). Habitat, group size, and the behavior of white-tailed deer. *Behavior* **98**, 168-179.

MCCullough, D. R. (1979). *The George Reserve deer herd, population ecology of a K-selected species*. Ann Arbor: University of Michigan Press.

Michigan Department of Natural Resources. (1998). *Northeast Michigan surveillance activities for bovine tuberculosis in the livestock and free-ranging deer populations. Update: January 20, 1998*. Michigan Department of Natural Resources, Wildlife Division, East Lansing.

Morris, R. S. and Pfeiffer, D. U. (1995). Directions and issues in bovine tuberculosis epidemiology and control in New Zealand. *New Zealand Veterinary Journal* **43**, 256-265.

Nelson, M. E. and Mech, L. D. (1987). Demes within a northeastern Minnesota deer population. In *Mammalian dispersal patterns: The effects of social structure on population genetics*, B. D. Chepko-Sade and Z. T. Halpin, (eds). Chicago: University of Chicago Press.

Ross, S. (1997). *Simulation*, 2nd Edition. San Diego: Academic Press.

Schmitt, S. M, Fitzgerald, S. D., Cooley, T. M., Bruning-Fann, C. S., Sullivan, L., Berry, D., Carlson, T., Minnis, R. B., Payeur, J. B., and Sikarskie, J. (1997). Bovine tuberculosis in free-ranging white-tailed deer from Michigan. *Journal of Wildlife Diseases* **33**, 749-758.

Sharpe, M. (1988). General theory of Markov processes. Boston: Boston Academic Press.

Starfield, A. M. and Bleloch, A. L. (1986). Building models for conservation and wildlife management. New York: Macmillan Publishing Company.

Towar, D. R., Scott, R. M., and Goyings, (1965). Tuberculosis in a captive deer herd. American Journal of Veterinary Research **26**, 339-346.

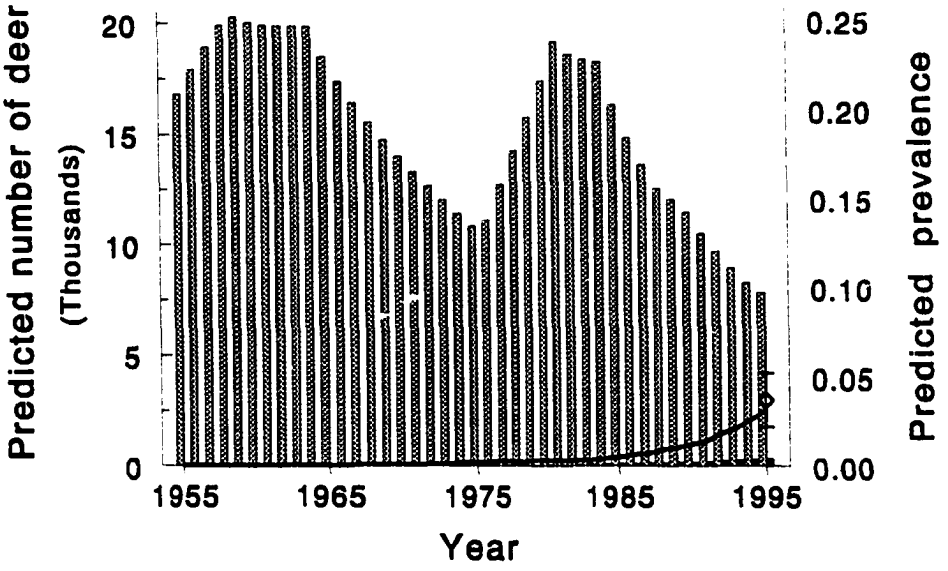


Figure 1.1. Our deterministic model simulated historic trends in deer population (left axis) and tuberculosis (right axis) dynamics. Transmission coefficients ($\beta_F = 0.5$, $\beta_M = 8.1$) were selected to yield a predicted prevalence of about 3.1% 40 years after introduction of one infected female. Gray bars are total numbers of deer predicted, black bars are numbers of infected deer predicted, and the line represents simulated prevalence. The diamond is an independent prevalence estimate derived from field data (Schmitt et al., 1997; MDNR, 1998); capped bars bound the estimate's 95% confidence interval.

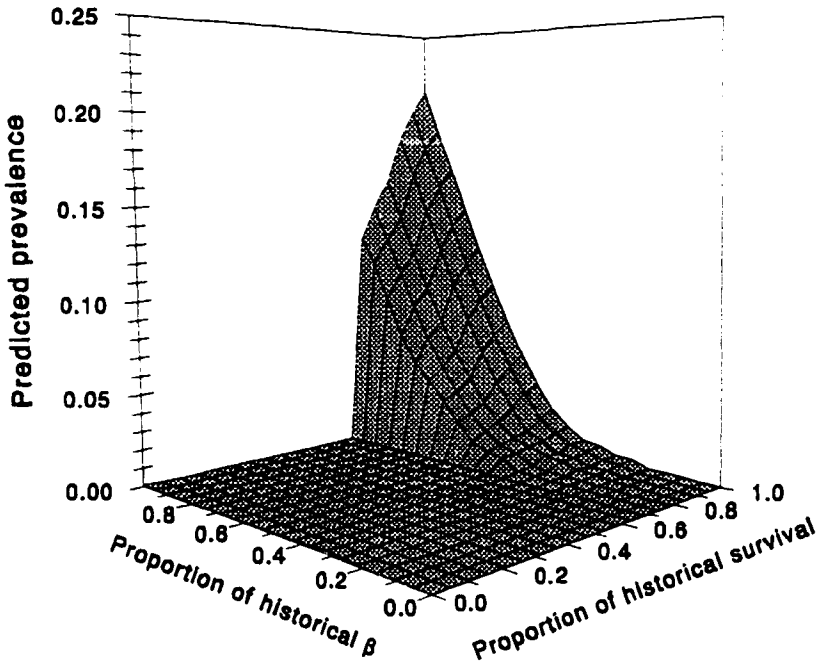


Figure 1.2. The effects of two alternative management approaches on tuberculosis prevalence were examined using our stochastic model. In general, the model predicted that lowering survival would be proportionately less effective in reducing predicted prevalence than lowering transmission until the threshold for population persistence was reached; lowering both survival and transmission appeared more effective than either approach alone.

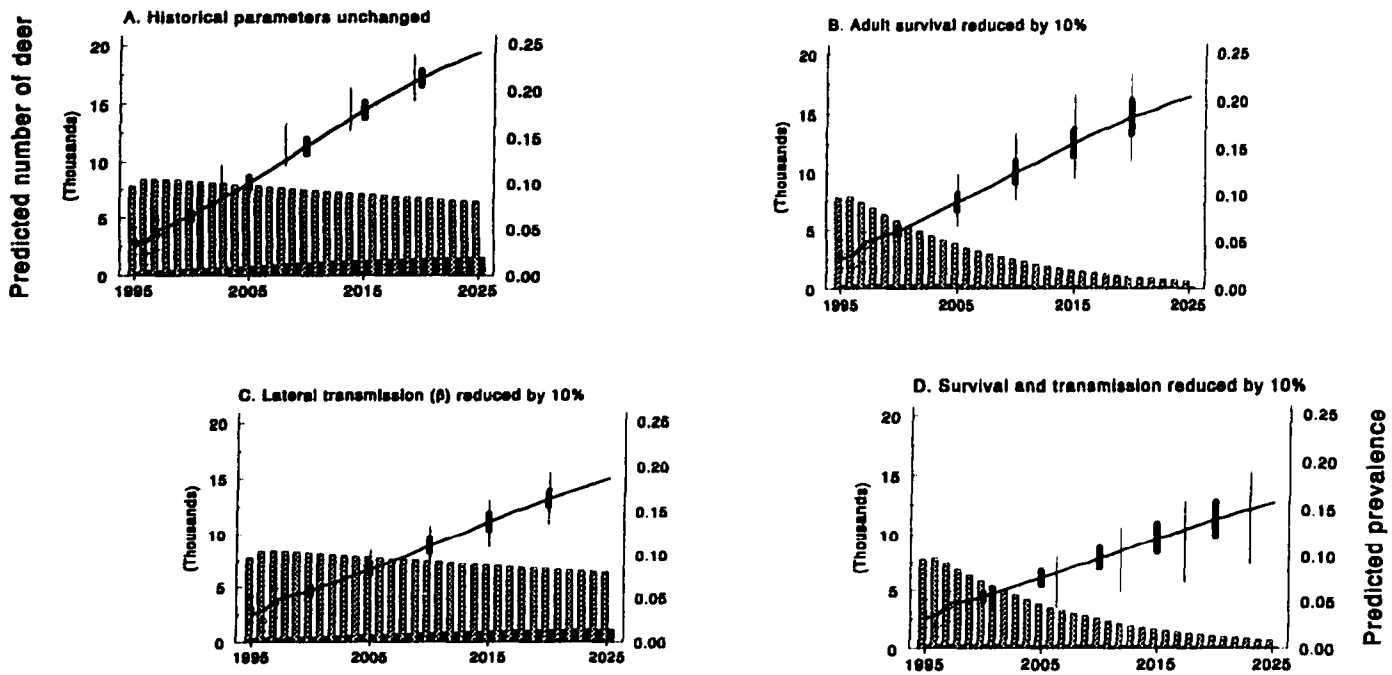


Figure 1.3. (a) Models with unaltered parameters predicted that tuberculosis prevalence (right axis) would increase to about 21% over the next 25 years. (b) Models that reduced survival by 10% produced only a modest reduction in predicted prevalence, but forecast dramatic declines in mean population size (left axis) and total infected animals (left axis). (c) Models with transmission coefficients reduced by 10% also produced a modest reduction in predicted prevalence without affecting population numbers. (d) Models that reduced both survival and transmission by 10% diminished both prevalence and total numbers of infected animals. Gray bars are total numbers of deer predicted, black bars are numbers of infected deer predicted, and the line represents predicted prevalence; vertical lines span the entire range of prevalence predictions, and the wider portions of those lines are ± 1 standard deviation of prevalence estimates from 1,000 sample paths. Diamonds are independent prevalence estimates derived from field data (Schmitt et al., 1997; MDNR, 1998); capped bars bound the estimates' 95% confidence intervals.

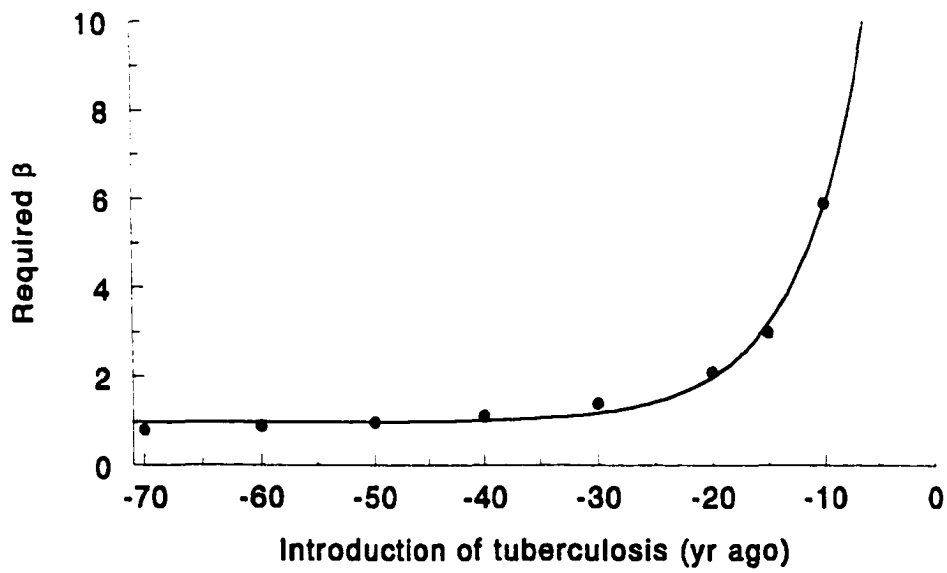


Figure 1.4. Our deterministic model also was used retrospectively to examine the likely historic duration of the tuberculosis epidemic in Michigan. Assumptions about epidemic duration influenced the values of transmission coefficients (β) necessary for the model to approximate observed prevalence data (Schmitt et al., 1997). We assumed tuberculosis was introduced 40-years-ago; a more recent introduction (e.g., 5- to 10-years-ago) would have required considerably higher coefficients to elevate prevalence to levels observed in 1995.

APPENDIX (*chapter 1*)

1.A.1. Quantitative description of the model

We modeled the epizootic and population dynamics of tuberculosis in the affected white-tailed deer herd as a Markov process on a 12-dimensional random vector (Table 1.A.1). Each dimension marked the number of individuals in a single sex/age/health state at the end of the cervid year. The index set was 1 cervid year. We defined the cervid year as the following sequence of 5 events: 1 day before fawning, 1 day after fawn recruitment, 1 day after vertical (maternal) disease transmission, 1 day after horizontal disease transmission, 1 day after all adult mortality. Transitions between consecutive vectors were made using the following rules:

Rule 1: The number of susceptible adult does/bucks 1 day before fawning in year *i* equals the number of susceptible adult does/bucks 1 day after all adult mortality in year *i*-1 plus the number of susceptible fawn does/bucks 1 day after all adult mortality in year *i*-1.

Rule 2: The number of 2nd year incubating adult does/bucks 1 day before fawning in year *i* equals the number of first year incubating adult/fawn does/bucks 1 day after all adult mortality in year *i*-1 plus the number of 1st year incubating fawn does/bucks 1 day after all adult mortality in year *i*-1.

Rule 3: The number of 1st year incubating adult/fawn does/bucks 1 day before fawning in year *i* equals 0.

Rule 4: The number of infectious adult does/bucks 1 day before fawning in year *i* equals the number of infectious adult does/bucks 1 day after all adult mortality in year

i-1 plus the number of 2nd year incubating adult does/bucks 1 day all adult mortality in year i-1.

Rule 5: The number of susceptible fawn does/bucks 1 day after fawn recruitment in year i is distributed as a binomial (number of adult does end of year i-1, $p_{(\text{adult doe recruiting})}/2$)

Rule 6: The adult susceptible/incubating/infectious population 1 day after fawn recruitment in year i equals the adult susceptible/incubating/infectious population 1 day before fawning in year i.

Rule 7: The number of infected fawn does/bucks 1 day after vertical disease transmission in year i is distributed as a binomial (number infectious adult does 1 day before fawning in year i, $p_{(\text{vertical transmission})}{}^{*}p_{(\text{adult doe recruiting})}$).

Rule 8: The number of susceptible fawn does/bucks 1 day after vertical disease transmission in year i equals the number of susceptible fawn does/bucks 1 day after fawn recruitment in year i minus the number of infected fawn does/bucks 1 day after vertical transmission in year i.

Rule 9: The adult susceptible/incubating/infectious population 1day after vertical transmission in year i equals the adult susceptible/incubating/infectious population 1 day after fawn recruitment in year i.

Rule 10: The number of infected fawn does/bucks 1 day after horizontal transmission in year i is distributed as a binomial(number of susceptible doe/buck fawns 1 day after vertical transmission in year i, ξ_1) + number of infected fawn does/bucks 1 day after horizontal transmission in year i: $\xi_1 = 1 - (1-1/T)^{M(i)\beta(1)}$, T = the total population size 1 day after fawn recruitment in year i, N_i = the total number of infectious individuals

1 day after fawn recruitment in year i, and β_1 = the average number of infectious contacts between a single infectious individual and does or fawns per cervid year.

Rule 11: The number of infected does 1 day after horizontal disease transmission in year i is distributed as a binomial (number of susceptible does 1 day after vertical disease transmission in year i, ξ_1).

Rule 12: The number of infected bucks 1 day after horizontal disease transmission in year i is distributed binomial(number of susceptible bucks 1 day after vertical disease transmission in year i, ξ_2): $\xi_2 = 1 - (1-1/T)^{N(i)\beta(2)}$, T = the total population size 1 day after fawn recruitment in year i, N_i = the total number of infectious individuals 1 day after fawn recruitment in year i and β_2 = the average number of infectious contacts between a single infectious individual and adult bucks per cervid year.

Rule 13: The number of adult susceptible/incubating/infectious does/bucks 1 day after all adult mortality in year i is distributed as a binomial (number of susceptible/incubating/infectious 1 day after horizontal disease transmission in year i, $p_{(\text{survival for adult does/bucks})}$).

We used Monte Carlo simulation to investigate the range of possible sample paths for this process (Table 1.A.2) and to derive ranges and other variance estimates on various relevant functionals of the these vectors (e.g., yearly prevalence and probability of transition from susceptible to infected) for each of the seven management scenarios mentioned in the previous section. Thus, the deterministic model used in the retrospective assessment projects as the expected sample path.

1.A.2. Derivation of equation 1.4

Rearranging equation 1.3 to give the probability of escaping infection and thus not receiving 1 or more infectious contacts gives

$$1-\left[1-\left(1-\frac{1}{N}\right)^{I\beta}\right]=\left(1-\frac{1}{N}\right)^{I\beta}.$$

Now consider a 2nd *constant* source of risk of infection that is independent of the risk of infection from the $I\beta$ infectious contacts that occur during the time period in question. Denote α as the probability of not becoming infected by the "point source" during the time period. Given the independent and non-additive relationship between the point source and the $I\beta$ infectious contacts, the probability of any given individual escaping all of the infectious contacts *and* infection from the point source during the time period is given by

$$\alpha\left(1-\frac{1}{N}\right)^{I\beta}.$$

Consequently, the probability of any given individual becoming infected during the time period is

$$1-\alpha\left(1-\frac{1}{N}\right)^{I\beta} \qquad \text{(Q.E.D.)}.$$

1.A.3. Equation 1.4 in continuous time

Equation 1.4 can be used as an alternative to the traditional "mass action assumption" of most continuous time epidemic models:

$$\frac{dS}{dt}=-\frac{\beta}{N}SI.$$

A clear derivation of this equation comes from Heesterbeek and Roberts (1995): "The I infectives make on average βI potentially infective contacts per unit of time. Only a fraction S/N of those contacts is with a susceptible. So, $\beta IS/N$ is the number of new cases arising per unit of time." Replacing this mechanism with the mechanism of equation 1.4 gives

$$\frac{dS}{dt} = -S\left[1-\alpha\left(1-\frac{1}{N}\right)^{I\beta}\right]. \tag{1.5}$$

For example, consider the classical epidemic model

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta}{N}SI \\
\frac{dI}{dt} &= \frac{\beta}{N}SI - \gamma I \\
\frac{dR}{dt} &= \gamma I,
\end{aligned}$$

where R is the number of immune or recovered individuals and γ is the recovery rate. Note that in this simple example, the population size and recovery rate are constant, and infected individuals are instantly infectious (i.e., no incubation period). Replacing the mass action assumption with equation 1.5 gives

$$\begin{aligned}
\frac{dS}{dt} &= -S\left[1-\left(1-\frac{1}{N}\right)^{I\beta}\right] \\
\frac{dI}{dt} &= S\left[1-\left(1-\frac{1}{N}\right)^{I\beta}\right] - \gamma I \\
\frac{dR}{dt} &= \gamma I.
\end{aligned}$$

The dynamics of this model are similar to the dynamics of the classical model (Fig. 1.A.1).

REFERENCES (*appendix, chapter 1*)

Heesterbeek, J. A. P. and Roberts, M. G. (1995). Mathematical models for microparasites of wildlife. In *Ecology of infectious disease in natural populations*, B. T. Grenfell and A. P. Dobson (eds), 90-122. Cambridge: Cambridge University Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. 1990. *Numerical recipes (FORTRAN)*. Cambridge: Cambridge University Press.

(*appendix*) **Table 1.A.1**

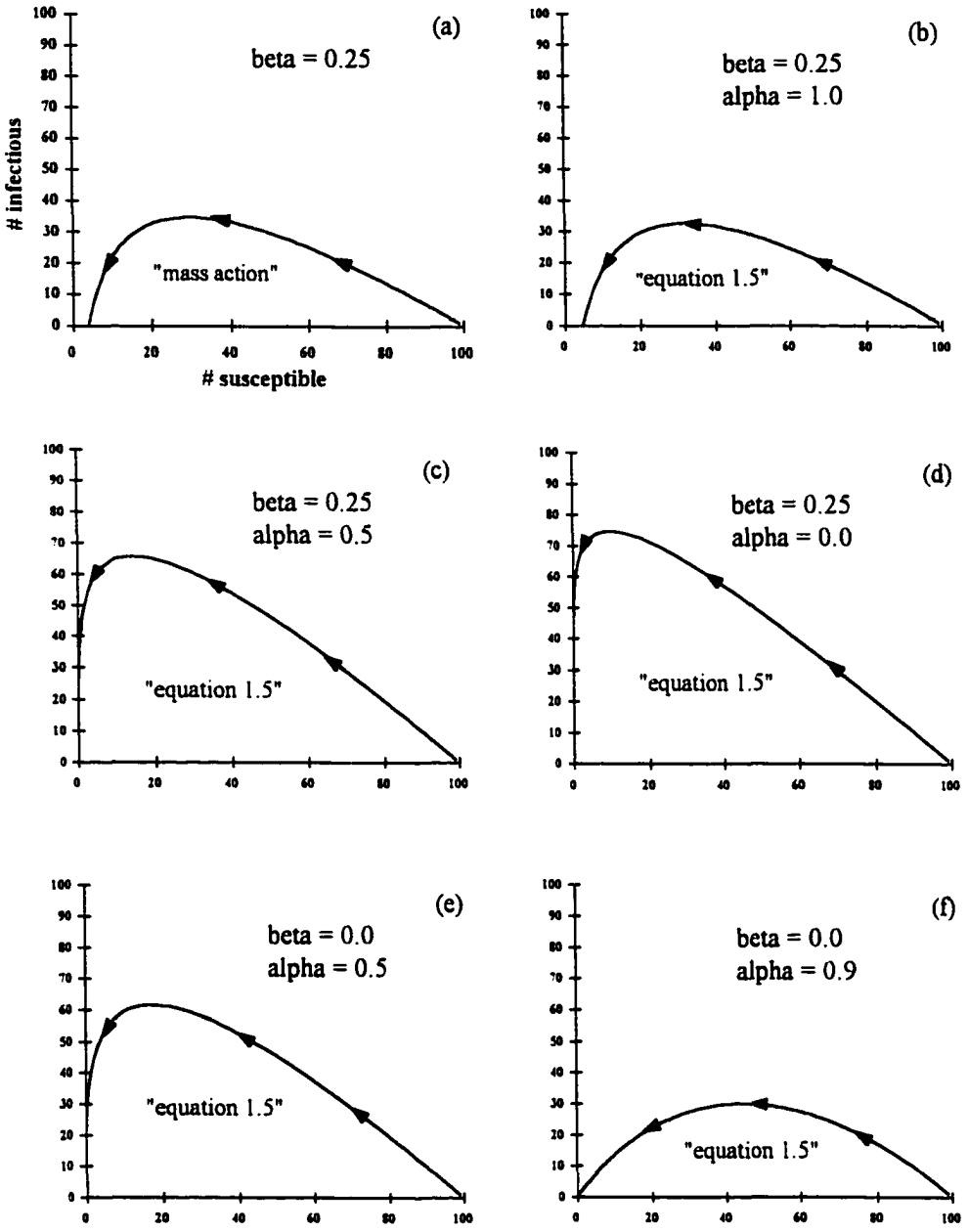
Random vector (X_1, X_2, ..., X_{12}) used to model a bovine tuberculosis (Mycobacterium bovis) epizootic in a free-ranging white-tailed deer.

Vector position	Interpretation of vector position
X_1	# of susceptible fawn does
X_2	# of susceptible adult does
X_3	# of 1st year incubating fawn does
X_4	# of 1st year incubating adult does
X_5	# of 2nd year incubating adult does
X_6	# of infectious adult does
X_7	# of susceptible fawn bucks
X_8	# of susceptible adult bucks
X_9	# of 1st year incubating fawn bucks
X_{10}	# of 1st year incubating adult bucks
X_{11}	# of 2nd year incubating adult bucks
X_{12}	# of infectious adult bucks

(*appendix*) **Table 1.A.2**

Simulated management scenarios and predicted probabilities for susceptible white-tailed deer becoming infected with bovine tuberculosis within a given year. See text for description of parameters and their estimation.

Scenario	Coefficients (β)		Adult survival		Transition probabilities		
	female	male	female	male	5 yr	10 yr	25 yr
Status quo	0.50	8.10	0.75	0.45	0.012	0.019	0.047
Reduce transmission (lower β by 10%)	0.45	7.29	0.75	0.45	0.009	0.010	0.031
Reduce transmission (lower β by 25%)	0.38	6.08	0.75	0.45	0.006	0.008	0.013
Reduce transmission (lower β by 50%)	0.25	4.05	0.75	0.45	0.003	0.003	0.002
Reduce adult survival (by 10%)	0.5	8.1	0.68	0.41	0.011	0.017	0.037
Reduce transmission and adult survival	0.38	6.08	0.68	0.41	0.006	0.007	0.010
Reduce transmission and adult survival	0.25	4.05	0.68	0.41	0.002	0.001	0.001



(*appendix*) Figure 1.A.1. Comparison of the dynamics of the "mass action assumption" (a) with the dynamics of "equation 1.5" (b-f). Arrows indicate direction of epidemic in time. Initial conditions (a-f) are 99 susceptible, 1 infectious, and 0 recovered. Assumptions (a-f) are constant population size, constant recovery rate (γ=0.15), and instantaneous infectiousness once infected. Note when α<1 (i.e., c-f) the final size of the epidemic is 0 remaining susceptible individuals. Trajectories were determined numerically (Press et al. 1990).

CHAPTER 2

A Comparison of Several Theories of Epidemic Dynamics Using AIC

SUMMARY

We defined and compared 4 classes of epidemic models (41 models in total) using Akaike's information criteria (AIC)-based model selection and 2 data sets (measles, "common cold") from the literature. The 4 classes included multinomial, structured multinomial, nested logistic regression, and biological-mechanistic models. In all, 52 possible model-data set combinations were fit and AICc values calculated. Model selection results were not consistent between data sets and no universal model emerged; only six models were highly plausible (ΔAICc $\leq$ 4) candidates for at least one of the data sets. Although we detected some evidence of underlying structure, neither data set suggests that the number of infectious or susceptible individuals, population size, or time strongly affected epidemic dynamics. Our findings underscore the need for replicated data and experimental evidence to support selection of any particular epidemic model. The results of our analyses raise questions about invoking popular epidemic models in attempts to explain or predict specific host/parasite dynamics.

Key words: Greenwood model; Reed-Frost model; mass action; binomial chain; epidemic modeling; AIC; model selection; statistical inference; measles; common cold.

2.1. Introduction to the problem

Models of specific epidemics are abundant in the literature (Dietz and Schenzle, 1985; Barlow, 1995). Several general theories of epidemic dynamics exist and have allowed authors a variety of approaches to specific problems (Bailey, 1975; Anderson and May, 1991; Barlow, 1995). Unfortunately, the assumptions of these theories have not been seriously scrutinized. Only a very few authors (e.g., Bailey, 1975; Becker, 1979) have attempted to describe existing data and compare alternative models from within a rigorous statistical framework; of these, none have applied the modern paradigm of information theory based model fitting and selection (Burnham and Anderson, 1998). Here we define and compare 4 classes of epidemic models (41 models in total) using 2 data sets from the literature and Akaike's information criteria (AIC)-based model selection.

We divide models of epidemic dynamics into two general categories: "statistical models" which attempt only to describe the pattern of the data, and "mechanistic models" which attempt to mimic the processes assumed to have generated the data. Mechanistic models are often more appealing to scientists because they offer potential insights that might lead to further testable hypotheses; they differ from statistical models in their ability to address the effects of changes in the system not previously observed. This may explain the dominance of mechanistic epidemic models in the literature (Dietz and Schenzle, 1985; Barlow, 1995).

The pivotal issue of mechanistic models of epidemic dynamics is how to mathematically describe the transition from the susceptible to the infected state throughout the course of the epidemic. Authors have rationalized a variety of approaches, modeling this transition as dependent on various combinations of constants, random variates, and functions of the number infectious (I), the number susceptible (S), and the total population size (N).

Although often neglected, many other phenomena are potential sources of significant effects. First, does infectiousness or susceptibility vary by individual? For example, is susceptibility a function of measurable covariates such as age, gender, occupation, and medical history, or of unknown and hence unmeasurable covariates? Second, does the mechanism of transmission vary by situation or pathogen? Third, what is the consequence of the population being open to births/deaths and emigration/immigration during the epidemic? Fourth, does infectiousness and/or susceptibility change through time independently of the population's composition in terms of I and S. Finally, does the spatial arrangement and movements of members of the population affect transmission?

2.2. Data

A literature search revealed two data sets useful in addressing these questions. Both are frequency tabulations of epidemic "binomial chains" in which the progress of successive viral infections was tracked over time within human family households. For example, one class of chains from households of 5 members with 1 initial infection is: initial infection, followed by 2 new infections and recovery of the first, 1 new infection and recovery of the second and third, and no new infections and recovery of the fourth; the notation for this chain is 1-2-1-0. The first infection is assumed to be independent of the last infection because this first infectious individual had ceased to be infectious after infecting the 2 in some way; the Markov property is assumed. Chains apparently were not initiated on the same calendar date. Collecting data in this format was presumably accomplished by visiting each household separately at time intervals beginning on the date when the first infection was recognized and continuing with re-visits at intervals that equaled the sum of the infectious and incubation

periods of the disease in question. To avoid misclassifying chains, a requirement of this scenario is that infectious and incubation periods be known, somewhat constant, and that the infectious period is short relative to the incubation period.

Both data sets were analyzed previously by different authors using traditional methodologies of hypothesis testing (Table 2.1). The first data set, "Heasman and Reid", is a tabulation of binomial chains from 664 households of size 5 with 1 initial infection of the common cold in each household (Heasman and Reid, 1961). This data set was the subject of extensive analyses by Becker (1980, 1981, 1989). The second data set, "Providence," is a tabulation of binomial chains from 434 households of sizes 3 and 4 with 1 initial infection of measles in each household (Wilson et al., 1939; Bailey, 1975). This data set was analyzed as two separate sets, conditional on household size, by Bailey (1975).

While far from ideal, together these 2 data sets display variation in several of the components of the epidemic problem. Each household of each data set represents a replicate of a specific epidemic. Both data sets exhibit variation in I and S during the course of the epidemic. The Providence data, while not open to births/deaths or immigration/emigration, shows variation in household size, thereby simulating this effect via comparison between households of different sizes within the same data set. In both sets, the spatial component to the epidemic process is largely constant because, when compared to entire communities, houses are approximately the same size and are likely to share similar daily activity dynamics. Further, households are most likely to meet the "homogeneity of mixing" assumption (Bailey, 1975) of most mechanistic models of epidemic dynamics and necessary for many of the models compared here. Unfortunately, no additional information was collected that might have further helped our analyses; several, potentially useful covariates (e.g., age and gender

composition of each household) are unknown. Finally, both illnesses are viral respiratory tract infections of comparable epidemiology. In both cases, the pathogen is thought to transmit via intimate contact between individuals, inhalation or ingestion of exhaled infectious aerosolized droplets or secretions, or by contact with surfaces infected by those droplets or secretions (Gershon, 1995; Gwaltney, 1995). Entry into the new host is believed to occur through the respiratory tract or mucous membranes.

2.3. Models

2.3.1 *Rationale*

For convenience we divide our set of 41 candidate models into 4 classes (multinomial, structured multinomial, nested logistic, biological-mechanistic) based on our interpretation of each individual model as "mechanistic" or "statistical" and whether or not individual models are nested within the global model of each class. For example, model class 2 contains 14 models nested under one global model and is considered to be more mechanistic than model class 1, but less so than model class 3.

2.3.2 *Model class 1, multinomial (1 member)*

The multinomial model, **MULT**, recognizes no generating process beyond the simple observation of different categories of chains, each with different cell probabilities. For example, under this model the likelihood of the Providence data is

$$\frac{334!}{34!25!36!239!}p_1^{34}p_2^{25}p_3^{36}\left(1-\sum_{i=1}^{3}p_i\right)^{239}\cdot\frac{100!}{4!3!1!4!3!8!10!67!}p_4^4p_5^3p_6^1p_7^4p_8^3p_9^8p_{10}^{10}\left(1-\sum_{i=4}^{10}p_i\right)^{67}.$$

2.3.3 *Model class 2, structured multinomial (15 members)*

Global model **NTSI** recognizes the first level of structure generating each multinomial cell of each data set; each chain classification is the result of a disease transmission process. Under this model, the probability of transition from the susceptible to the infected state is conditional on the population size (N), the time step (T), the number of available susceptible individuals (S), and the number of infectious individuals (I). The number recovered (R), is not included since $R=N-S-I$. For example, the expected number of 1-2-0 chains (household size 4) under model **NTSI** in the Providence data is

$$100(3p^2_{N=4,T=1,S=3,I=1}q_{N=4,T=1,S=3,I=1}q_{N=4,T=2,S=1,I=2})$$

which we interpret as the expected number of instances of 2 of 3 susceptibilities (S=3) converting ($p^2_{N=4,T=1,S=3,I=1}$) to the infectious state in the presence of 1 infectious (I=1) by the end of the first time step (T=1), followed by 1 susceptible (S=1) not converting ($q_{N=4,T=2,S=1,I=2}$ = 1 - $p_{N=4,T=2,S=1,I=2}$), in the presence of 2 infectious (I=2) by the end of the second time step (T=2). The coefficient "3" comes from the fact that there are 3 ways to observe this sequence of events among 3 individuals (+ 1 initial infectious). Model class 2 contains 14 nested sub-models (**N, T, S, I, NT, NS, NI, TI, TS, SI, NTS, NTI, NSI, TSI**), some of which are redundant due to issues of parameter identifiability when applied to these 2 data sets because N = 3 or 4 is small. Model **I** has been previously investigated by Becker (1981) using the Heasman and Reid data.

2.3.4 *Model class 3, nested logistic (16 members)*

Nested logistic regression models, global model **logitNTSI**, consider transition from the susceptible to the infected state as a logit transform of some linear combination of the current population state,

$$p_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 N+\beta_2 T+\beta_3 S+\beta_4 I)]}.$$

These models offer nearly the same structure as models in class 2, but their forced additivity results in greater parsimony. This approach was briefly explored by Becker (1989) using the Heasman and Reid data but not in a framework of model selection.

Included in this class is our interpretation of the "mass action assumption" (Hamer, 1906; Muench, 1959; Anderson and May, 1979). Unfortunately, this differential equation has no unique interpretation as a stochastic model or in discrete time (Bailey 1975):

$$\frac{dS}{dt} = -\frac{\beta}{N} SI.$$

The clearest derivation comes from Heesterbeek and Roberts (1995): "The I infectives make on average βI potentially infective contacts per unit of time. Only a fraction S/N of those contacts is with a susceptible. So, $\beta IS/N$ is the number of new cases arising per unit of time." This mechanism, regardless of any specific interpretation as a stochastic model or in discrete time, requires a positive monotonic I effect. Consequently, we represent the "mass action assumption" as a logit transform of I/N, model **logitI/N**:

$$p_{N,T,S,I} = \frac{1}{1+\exp\left[-\left(\beta_0+\beta_1\frac{I}{N}\right)\right]}.$$

2.3.5 *Model class 4, biological-mechanistic (10 members)*

For heuristic purposes, all members of this class are considered as modifications of the McCarty-Miller model, **MM** (McCarty and Miller, 1998), which recognizes two possible biological mechanisms of disease transmission: a random apportionment of Ib (the product of I infectious individuals and b infectious contacts per infectious individual) total infectious contacts among the N individuals of the entire population at each time step, and a possible constant level of background risk (1- a):

$$p_{N,T,S,I} = 1 - a\left(1 - \frac{1}{N}\right)^{Ib}.$$

Model **MM** simplifies to the Reed-Frost model, **RF** (Bailey 1975), if N is held constant and a=1 (i.e., no background risk) and the Greenwood model, **G** (Bailey, 1975; Greenwood, 1931), if b=0 and N is held constant. Thus, the Greenwood model has a mechanistic interpretation as a single point source of infection, making the probability of transition from the susceptible to the infected state a constant regardless of population composition.

Model **MM** also simplifies to the James-Rossiter model (1989), **JR**, if a=1 and N=S; corresponding to a random apportionment of the Ib infectious contacts across only the susceptible portion of the population. Model **MM** can be modified to account for variation of the degree of risk to infection from a point source specific to a population by replacing a with a beta random variate. For example, the expected number of 1-2-0 chains (household size 4) in the Providence data under model **MM** with b=0 and a~beta (α, β) is given by

$$\mathrm{E}[100(3(1-a)^2a^2] = 300\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 (1-a)^2a^2(1-a)^{\alpha-1}a^{\beta-1}da,$$

which simplifies to

$$300\frac{(\alpha+1)\alpha(\beta+1)\beta}{(\alpha+\beta+3)(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)}.$$

This modification is equivalent to the beta-Greenwood model, **BG**, previously investigated by Bailey (1975) using the Providence data; allowing separate beta random variates for each population size gives model **BGN**. Model **MM** can be further modified to give the beta-Reed-Frost model, **BRF** (Becker, 1981), if a=1, and the term $(1 - 1/N)^b$ is replaced by a beta random variate; conditioning on N gives Model **BRFN** (see *appendix* for additional clarification).

Finally, included in this class is the propagating point source model, **PP**, presented here for the first time, which considers each new infectious individual as a new point source of risk to infection:

$$p_{N,T,S,I} = 1-(1-p)^I.$$

While mechanistically distinct, model **PP** is functionally equivalent to model **MM** when a=b=1 and the term $1/N$ is replaced by the parameter p; only with the Providence data set is this model distinct.

2.4. Model fitting and selection

2.4.1 *AIC and Akaike weights*

Akaike's information criteria-based model selection requires 4 steps (Buckland et al., 1997; Burnham and Anderson, 1998): selection of candidate models for each data set, maximizing the resulting likelihoods, calculation of an AIC value for each model-data set combination, and comparison of the resulting AIC values with selection leaning towards models with the lowest AIC values. Akaike's information criteria, corrected for small sample biases (AICc),

is given by:

$$\mathrm{AICc} = -2\ln \mathcal{L}(\hat{\underline{\theta}} | \mathrm{data}, \mathrm{model}) + 2K + \frac{2K(K+1)}{n-K-1},$$

where K is the number of model parameters ($\underline{\theta}$) estimated from the data (as maximum likelihood estimates) and n is the sample size. Theoretically, this procedure selects the model achieving the most efficient, or "best", compromise between bias and precision (Burnham and Anderson, 1998; McQuarrie and Tsai, 1998). Models differing in AICc value from the best model by more than 4-7 units are generally poor candidates (Burnham and Anderson, 1998). Alternatively, Akaike weights provide a useful scaled metric (0-1) of model selection uncertainty for model i among the R candidate models:

$$w_i = \frac{\exp\left(-\frac{1}{2}\Delta \mathrm{AIC}_i\right)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2}\Delta \mathrm{AIC}_r\right)},$$

where $\Delta\mathrm{AIC}_i$ = AIC_i - minimum AIC; models with high weights are good candidates (Burnham and Anderson, 1998).

2.4.2 *Likelihoods*

We include only one loglikelihood (less multinomial coefficient, and not simplified to assist interpretation) as an example, the Heasman and Reid data under model **NTSI**, where $(\ldots)_{N,T,S,I}$ indexes every term within the parentheses:

$423 \cdot \ln[(q^4)_{5,1,4,1}] + 131 \cdot \ln[4(pq^3)_{5,1,4,1}(q^3)_{5,2,3,1}] + 36 \cdot \ln[12(pq^3)_{5,1,4,1}(pq^2)_{5,2,3,1}(q^2)_{5,3,2,1}] +$

$14 \cdot \ln[24(pq^3)_{5,1,4,1}(pq^2)_{5,2,3,1}(pq)_{5,3,2,1}(q)_{5,4,1,1}] + 4 \cdot \ln[24(pq^3)_{5,1,4,1}(pq^2)_{5,2,3,1}(pq)_{5,3,2,1}(p)_{5,4,1,1}] +$

$2 \cdot \ln[12(pq^3)_{5,1,4,1}(pq^2)_{5,2,3,1}(p^2)_{5,3,2,1}] + 8 \cdot \ln[12(pq^3)_{5,1,4,1}(p^2q)_{5,2,3,1}(q)_{5,3,1,2}] +$

$2 \cdot \ln[12(pq^3)_{5,1,4,1}(p^2q)_{5,2,3,1}(q)_{5,3,1,2}] + 2 \cdot \ln[4(pq^3)_{5,1,4,1}(p^3)_{5,2,3,1}] + 24 \cdot \ln[6(p^2q^2)_{5,1,4,1}(q^2)_{5,2,2,2}]$

$+ 11 \cdot \ln[12(p^2q^2)_{5,1,4,1}(pq)_{5,2,2,2}(q)_{5,3,1,1}] + 3 \cdot \ln[12(p^2q^2)_{5,1,4,1}(pq)_{5,2,2,2}(p)_{5,3,1,1}] +$

$1 \cdot \ln[6(p^2q^2)_{5,1,4,1}(p^2)_{5,2,2,2}] + 3 \cdot \ln[4(p^3q)_{5,1,4,1}(q)_{5,2,1,3}] +$

$0 \cdot \ln[4(p^3q)_{5,1,4,1}(p)_{,5,2,1,3}] + 0 \cdot \ln[(p^4)_{5,1,4,1}]$ (see *appendix* for additional clarification).

Standard approaches (Becker, 1981) give simple closed-form solutions for maximum likelihood estimates of the parameters for all models in classes 1 and 2 by equating the first derivative of each loglikelihood to 0 and solving for each parameter. For example, the maximum likelihood estimator of parameter $p_{N=4,T=2}$ of model **NT** is:

$$\hat{p}_{N=4,T=2} = \frac{y_{N=4,T=2}}{x_{N=4,T=2}},$$

where $y_{N=4,T=2}$ is the total number of individuals converting from the susceptible to the infected state in households of size 4 at the second time step of each household's epidemic, and $x_{N=4,T=2}$ is the total number of such trials (e.g., for the Providence data, $y_{N=4,T=2} = 21$ and $x_{N=4,T=2} = 40$, so $p_{N=4,T=2} = 0.525$). We used numerical optimization (Powell, 1964) in cases where no simple closed-form estimator exists.

2.5. Results and discussion

2.5.1 *Complications*

We fit and calculated AICc values for 52 possible model-data set combinations and assessed goodness-of-fit using the χ^2 statistic (Tables 2.2, 2.3). Model selection results are not consistent between data sets and no universal model emerges; only six models are especially plausible ($\Delta AIC \leq 2$) candidates for at least one of the data sets. Although we detected some evidence of underlying structure, neither data set suggests that the number of infectious or susceptible individuals, population size, or time strongly affected epidemic dynamics. While our analyses raise several questions about the validity and utility of current theories of epidemic dynamics, two issues must be considered as we attempt to address them. Were the data collected accurately, and were sample sizes sufficient to detect weak or complicated effects?

Misclassification of chains could have occurred in numerous ways. For example, some instances of the chain 1-2-1 could have been incorrectly classified as 1-1-1-1. Both measles and the common cold have an infectious prodrome of 1-4 days prior to the appearance of symptoms and subsequent diagnosis (Gershon, 1995; Gwaltney, Jr., 1995). If one of the 2 secondary infections occurred early, and one occurred late, both individuals may have been infectious but only one may have been symptomatic and detected. Further, we have no guarantee that new infections were not introduced into some households from outside sources during the period of data collection.

Heterogeneity among households is also an issue. Perhaps, households of size 4 contained more young children than households of size 3 and children are more or less susceptible or infectious than adolescents and adults. Also, were resistant individuals present

in any households prior to the start of the study? This seems more likely in epidemics of measles than the common cold because, unlike the common cold, prior infections of measles confer long-term immunity (Gershon, 1995; Gwaltney, 1995). Without quantitative information about such covariates, the presence of undesirable variance components in these data sets seems likely.

Finally, AICc selects the "best" approximating model and not necessarily the most faithful representation of the generating process; thus, sample size issues are a concern (Burnham and Anderson, 1998). For example, if model **MM** were the true generating process, model **MM** could still avoid selection in favor of model **G** if the level of background risk (parameter a) is sufficiently higher than the risk due to direct transmission from infectious individuals to susceptibilities (the product Ib). Further, regardless of the number of instances of each type of chain, data from households of size 5 have only 15 initial degrees of freedom with respect to our candidate models because there are only 16 possible chains. Similarly data from households of sizes 3-4 combined have only 10 initial degrees of freedom.

2.5.2 *Evidence for structure*

The low Akaike weights of model **MULT** in both analyses suggest that the structure represented by a binomial chain interpretation of each cell of the multinomial is plausible; that each chain is the result of a specific disease transmission process and it is useful to model this process as generating the multinomial. As a caveat, we point out that the fully parameterized version of **MULT**, in which every chain has a unique probability of occurrence and a perfect fit is achieved, would always be selected over the generating process if it were present as a candidate model and of higher dimension than **MULT**. Thus, while selection against model

MULT is encouraging, its selection as the best approximating model for a particular data set would not rule out the chain binomial interpretation of this type of data.

2.5.3 *Evidence for heterogeneity*

Selection of model **BG** among class 4 models in both data sets and among all models in the Providence data suggests that heterogeneity of one or more conditions among households is plausible in the Providence data and at least possible in the Heasman and Reid data. Given the lack of covariate information, the smaller Akaike weight of model **BG** in the Heasman and Reid data is perhaps more surprising than its large weight in the Providence data; in part, the former may possibly be explained by the multiple etiologies that may cause "colds" (Gwaltney, 1995). Misclassification of chains seems more likely in the Heasman and Reid data than in the Providence data given the shorter incubation and longer infectious period of the common cold relative to measles (Gershon, 1995; Gwaltney, 1995). However this may contradict the potential for hidden resistant individuals associated with the Providence measles epidemic.

2.5.4 *Evidence for I, S, T, and N effects*

Four observations illustrate a lack of conformity between current theories of epidemic dynamics and the data sets we examined. Neither data set suggests a strong I effect; that risk to infection increases or decreases when the number infectious exceeds 1. All model-data set combinations incorporating an I effect gave Akaike weights of <0.12, although it seems possible that the highly infectious nature of these pathogens and relatively large proportion of the "population" initially infected may have overwhelmed such effects. Further, while we

recognize that the Providence data set shows variation in I for only households of size 4, separate analyses by household size result in no remarkable differences in the model selection results; model **BG** is still strongly favored even when the data are partitioned by household size and no evidence for an I effect emerges. Second, the Providence data only hints at a possible N effect as model **BGN** ranked second but with a low weight of 0.165. The 30% increase in population size is apparently not large enough to detect an N effect with this sample size (i.e., $N_{\text{size 3}} = 334$, $N_{\text{size 4}} = 100$), if it is present in these data. Third, while a T effect seems biologically reasonable, only the Heasman and Reid data offer any support for this notion with the relatively high weights of models **logitT** and **logitTS**, 0.166 and 0.144 respectively. It is difficult to argue that even a weak T effect could have been masked by heterogeneity in the Heasman and Reid data, given the relatively low rank of model **BG** and the extremely good fit of virtually all candidate models. Fourth, reasonable evidence is suggested for a monotonic S effect in the Heasman and Reid data as model **logitS** produced a weight of 0.312; however, in the Providence data all models incorporating this effect gave weighs <0.001. We can give no biological explanation for this outcome. Perhaps this result is an artifact of Freedman's paradox (Freedman, 1983; Burnham and Anderson 1998); or, the Heasman and Reid data set is an unfortunate sample which deviates greatly from the expected sample of its generating process.

2.5.5. *Evidence for simple generating mechanisms*

Failure to detect an I effect eliminates several of the proposed mechanisms from further consideration here. Model **JR** did not generate these data because several of the reported chains are an impossible result of its suggested mechanism. For example, chain 1-3-0 of the

Heasman and Reid data could not occur since the probability of the last susceptible escaping infection is not defined (e.g., since $(1-1/S)^{3b}$ is not reasonable when $S = 0$).

Selection of the beta-Greenwood model for the Providence data has a pleasing mechanistic interpretation as constant background risk varying in intensity among households. Explanatory hypotheses are plentiful. Perhaps infections were spread to susceptibilities through contact with common infected objects within each home such as food, towels, sinks, or eating/drinking utensils; or perhaps the infectious agent was uniformly distributed in the air throughout each home. These conditions could produce constant risk; the addition of new infectious individuals may have little or no further effect other than replenishing and maintaining risk through time. Because this model was not selected by both data sets, however, further conjecture is problematic.

2.6.6 *Conclusion*

"A discipline is a science only if the theories it entertains are universal and falsifiable" (Popper 1972; Dolby 1982). Thus, although we recognize the difficulty of falsifying theories expressed as probability statements (Dolby, 1982), we are concerned with the proliferation of models of specific epidemics in the absence of a well-supported underlying theory of epidemic dynamics. Even a cursory literature search reveals hundreds of such models published since 1930. Further, although it may be unrealistic to expect that a single unifying model of epidemic dynamics can adequately represent the myriad of host/parasite relationships encountered in nature, we are troubled by the lack of significant supporting data sets, replication, and experimental evidence supporting any particular epidemic model. While the data analyzed here are admittedly imperfect, the results of our analyses raise questions about

the propriety of blindly invoking popular epidemic models in attempts to explain or predict specific host/parasite dynamics. We conclude that until these issues are sorted out and a well supported theory of epidemic dynamics emerges, models attempting forecasts of the outcomes of specific epidemics will remain little more than "just so stories" (Kipling, 1902), whose only real utility may lie in exercising our own imaginations.

ACKNOWLEDGMENTS (*chapter 2*)

Support from the Colorado Division of Wildlife, W-153-R is gratefully acknowledged. Richard Bartman, Sara Oyler-McCance and Mary Conner provided useful reviews of earlier drafts.

REFERENCES (*chapter 2*)

Anderson, R. M. and May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature* **280**, 361-367.

Anderson, R. M. and May, R. M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*, second edition. London: Charles Griffin and Company.

Barlow, N. D. (1995). Critical evaluation of wildlife disease models. In *Ecology of infectious disease in natural populations*, B. T. Grenfell and A. P. Dobson (eds), 230-259. Cambridge: Cambridge University Press.

Becker, N. (1989). *Analysis of infectious disease data*. New York: Chapman and Hall.

Becker, N. (1980). An epidemic chain model. *Biometrics* **6**, 249-254.

Becker, N. (1981). A general chain binomial model for infectious diseases. *Biometerics* **37**, 251-258.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1977). Model selection: an integral part of inference. *Biometrics* **53**, 603-618.

Burnham, K. P. and Anderson, D. A. (1998). *Model selection and inference: A practical information theoretic approach*. New York: Springer-Verlag.

Dietz, K. and Schenzle, D. (1985). Mathematical models for infectious disease statistics. In *A celebration of statistics*, A. C. Atkinson and S. E. Fienberg (eds), 167-204. New York: Springer Verlag.

Dolby, G. R. (1982). The role of statistics in the methodology of the life sciences. *Biometerics* **38**, 1069-1083.

Freedman, D. A. (1983). A note on screening regression equations. *The American Statistician* **37**, 152-155.

Gershon, A. A. (1995). Measles virus (rubeola). *Mandell, Douglas and Bennett's principles and practice of infectious diseases - 4th edition.* New York: Churchill Livingstone.

Greenwood, M. (1931). On the statistical measure of infectiousness. *Journal of Hygiene Cambridge* **31**, 336-351.

Gwaltney, J. M. Jr. (1995). The common cold. *Mandell, Douglas and Bennett's principles and practice of infectious diseases - 4th edition.* New York: Churchill Livingstone.

Hamer, W. H. (1906). Epidemic disease in England-the case evidence of variability and persistence of type. *Lancet* **1**, 733-9.

Heasman, M. A. and Reid, D. D. (1961). Theory and observation in family epidemics of the common cold. *British Journal of Preventive Social Medicine* **15,** 12-16.

Heesterbeek, J. A. P. and Roberts, M. G. (1995). Mathematical models for microparasites of wildlife. In *Ecology of Infectious Disease in Natural Populations*, B. T. Grenfell and A. P. Dobson (eds), 90-122. Cambridge: Cambridge University Press.

James, A. D. and Rossiter, P. B. (1989). An epidemiological model of rinderpest. I. Description of the model. *Tropical Animal Health and Production* **21**, 59-68.

Kipling, R. (1902). *Just so stories*. London: MacMillan.

McCarty, C. W. and Miller, M. W. (1998). A versatile model of disease transmission applied to forecasting bovine tuberculosis dynamics in white-tailed deer. *Journal of Wildlife Diseases* **34**, 722-730.

McQuarrie, A. D. R. and Tsai, C. L. (1998). *Regression and Time Series Model Selection*. Singapore: World Scientific.

Muench, H. (1959). *Catalytic Models in Epidemiology*: Cambridge: Harvard University Press.

Popper, K. R. (1972). *The Logic of Scientific Discovery*. London: Huchinson.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal* 7, 155-162.

Wilson, E. B., Bennett, C. Allen, M., and Worchester, J. (1939). Measles and scarlet fever in Providence, R. I., 1929-34 with respect to age and size of family. *Procedings of the American Philosophical Society* **80**, 357-476.

Table 2.1

Data sets used in a comparison of theories of epidemic dynamics.

Heasman and Reid (common cold data; Heasman and Reid, 1961)		Providence (measles data; Bailey, 1975)	
chain	frequency	chain	frequency
1-0	423	size 3	
1-1-0	131	1-0	34
1-1-1-0	36	1-1-0	25
1-1-1-1-0	14	1-1-1	36
1-1-1-1-1	4	1-2	239
1-1-1-2	2		
1-1-2-0	8	size 4	
1-1-2-1	2	1-0	4
1-1-3	2	1-1-0	3
1-2-0	24	1-1-1-0	1
1-2-1-0	11	1-1-1-1	4
1-2-1-1	3	1-1-2	3
1-2-2	1	1-2-0	8
1-3-0	3	1-2-1	10
1-3-1	0	1-3	67
1-4	0		
		total: 434	
total: 664			

Table 2.2

Results of a comparison of 21 models of epidemic dynamics using binomial chain data from households of size 5 from an epidemic of the common cold (Heasman and Reid, 1961).

Model	Class	*K*	χ^2 (df)	ΔAICc	AICc weight
logitS	3	2	2.34 (9)	0[1]	0.312
logitT	3	2	4.11 (9)	1.26	0.166
logitTS	3	3	2.42 (9)	2.00	0.114
logitSI	3	3	2.32 (9)	2.02	0.114
logitTI	3	3	2.47 (9)	2.07	0.111
S	2	4	1.39 (8)	3.31	0.060
logitTSI	3	4	2.32 (8)	4.01	0.042
BG	4	2	7.84 (10)	4.58	0.032
T	2	4	6.43 (8)	5.19	0.023
SI	2	7	1.36 (5)	7.99	0.006
TS	2	7	1.36 (5)	8.00	0.006
TI	2	7	1.38 (5)	8.01	0.006
BRF	4	2	14.2 (10)	8.24	0.005
TSI	2	8	1.35 (7)	10.04	0.002
RF	4	1	17.8 (11)	11.22	0.001
MM	4	2	15.7 (7)	12.78	0.001
I	2	3	14.0 (8)	12.90	0.000
logitI	3	2	18.6 (7)	13.18	0.000
G	4	1	22.2 (12)	14.98	0.000
MULT	1	15	----- (0)	22.78	0.000
JR	4	1	NA	∞	0.000

1. AICc: 1643.28

Table 2.3

Results of a comparison of 31 models of epidemic dynamics using binomial chain data from household sizes 3-4 from an epidemic of measles, Providence Rhode Island (Bailey, 1975).

Model	Class	K	χ^2 (df)	ΔAICc	AICc weight
BG	4	2	6.13 (8)	0[1]	0.824
BGN	4	4	5.30 (6)	3.22	0.165
BRF (BPP[2])	4	2	22.5 (8)	9.81	0.006
MULT	1	10	----- (0)	10.71	0.004
RFN	4	4	22.3 (6)	13.64	0.001
T	2	3	83.0 (6)	68.91	0.000
TI	2	4	81.5 (4)	70.93	0.000
logitT	3	2	89.8 (6)	71.78	0.000
TS	2	5	81.6 (5)	72.48	0.000
NT	2	5	78.8 (3)	72.53	0.000
logitTS	3	3	90.1 (5)	73.22	0.000
logitNT	3	3	89.9 (5)	73.46	0.000
NS	2	5	76.8 (5)	73.45	0.000
logitTI	3	3	89.8 (5)	73.57	0.000
NSI (others[3])	2	6	77.9 (2)	73.97	0.000
logitNS	3	3	75.0 (5)	74.10	0.000
logitNSI (others[4])	3	4	77.5 (4)	75.09	0.000
S	2	3	88.7 (4)	78.89	0.000
SI	2	4	88.8 (3)	80.78	0.000
logitS	3	2	93.8 (5)	82.52	0.000
logitSI	3	3	92.2 (4)	83.17	0.000
G	4	1	101. (7)	85.93	0.000
logitI	3	2	99.0 (6)	90.90	0.000
I	2	2	99.0 (5)	90.93	0.000
NI (logitNI)	2,3	3	101. (5)	92.59	0.000
logitI/N	3	2	95.0 (6)	92.68	0.000
N (GN[5], logitN)	2,3,4	2	99.7 (6)	95.92	0.000
MM	4	2	99.4 (6)	95.93	0.000
RF (PP)	4	1	125. (7)	114.18	0.000
RFN	4	2	119. (6)	116.20	0.000
JR	4	1	NA	∞	0.000

1. AICc = 858.08
2. Model **PP** with $p \sim \text{beta}(\alpha,\beta)$
3. In this data set, **NSI** is equivalent to **NTSI, TSI, NTS, and NTI**
4. In this data set, **logitNSI** is equivalent to **logitNTSI, logitTSI, logitNTS, and logitNTI**
5. Model **G**, with p conditional on N

APPENDIX (*chapter 2*)

Miscellaneous notes and associated references

2.A.1. Heasman and Reid data: model structures, parameter estimates, miscellaneous interval estimates, and miscellaneous estimated variance-covariance matrices

2.A.1.1 *Model* **logitS**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 S)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -819.630 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 2.34 (df = 9; Sakamoto, 1991); AICc = 1643.28; ΔAICc = 0; AIC weight = 0.312 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 0.88206; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) for β_0 is (0.42872, 1.3539).
(3) MLE of β_1 = 0.30847; 95% PLI is (0.17990, 0.43318).
(4) Estimated variance-covariance matrix from 8000 bootstrap replicates (Efron and Tibshirani, 1993), $\mathrm{vc}(\bar{\beta}_0, \bar{\beta}_1)$, is

$$\begin{bmatrix} 0.0631 & -0.0171 \\ & 0.00484 \end{bmatrix}.$$

Corresponding estimated standard error-correlation matrix is

$$\begin{bmatrix} 0.251 & -0.979 \\ & 0.0696 \end{bmatrix}.$$

continued ...

2.A.1.2 *Model* **logitT**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 T)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -820.261 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 4.11 (df = 9; Sakamoto, 1991); AICc = 1644.54; ΔAICc = 1.26; AIC weight = 0.166 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 2.4571.
(3) MLE of β_1 = -0.34123.

2.A.1.3 *Model* **logitTS**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 T + \beta_2 S)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -819.627 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 2.42 (df = 9; Sakamoto, 1991); AICc = 1645.28; ΔAICc = 2.00; AIC weight = 0.114 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 0.96852.
(3) MLE of β_1 = -0.019829.
(4) MLE of β_2 = 0.29193.

2.A.1.4 *Model* **logitSI**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 S + \beta_2 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -819.627 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 2.32 (df = 9; Sakamoto, 1991); AICc = 1645.30; ΔAICc = 2.02; AIC weight = 0.114 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 0.88653.
(3) MLE of β_1 = 0.30825.
(3) MLE of β_2 = 0.0034895.

2.A.1.5 *Model* **logitTI**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 T+\beta_2 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -819.653 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 = 2.47 (df = 9; Sakamoto, 1991); AICc = 1645.35; ΔAICc = 2.07; AIC weight = 0.111 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 2.7372.
(3) MLE of β_1 = -0.31852.
(4) MLE of β_2 = -0.30119.

2.A.1.6 *Model* **S**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_S.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -819.265 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 1.39 (df = 8; Sakamoto, 1991); AICc = 1646.59; ΔAICc = 3.31; AIC weight = 0.060 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimates (MLE's; Lehmann, 1983) of $p_{5,4,1,1}$, $p_{5,3,1,2}$, $p_{5,3,1,1}$, and $p_{5,2,1,3}$ = 0.20021; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) for $p_{5,2,1,3}$ is (0.10159, 0.33162).
(3) MLE's of $p_{5,3,2,1}$ and $p_{5,2,2,2}$ is = 0.20000; 95% PLI is (0.14736, 0.26067).
(4) MLE of $p_{5,2,3,1}$ = 0.13747; 95% PLI is (0.1132, 0.16649).
(5) MLE of $p_{5,1,4,1}$ = 0.10775; 95% PLI is (0.096273, 0.11984).
(6) Estimated variance-covariance matrix from 8000 bootstrap replicates (Efron and Tibshirani, 1993), $\mathrm{vc}\left(\hat{p}_{5,1,4,1}, \hat{p}_{5,2,3,1}, \hat{p}_{5,2,2,2}\hat{p}_{5,2,1,3}\right)$, is

$$\begin{bmatrix} 1.10E-4 & 6.81E-6 & -2.16E-5 & -1.28E-4 \\ & 2.87E-4 & -1.05E-5 & 3.66E-6 \\ & & 7.67E-4 & 1.35E-5 \\ & & & 3.70E-3 \end{bmatrix}.$$

Corresponding estimated standard error-correlation matrix is

$$\begin{bmatrix} 0.0105 & 0.0384 & -0.0743 & -0.201 \\ & 0.0169 & 0.0224 & 0.00356 \\ & & 0.0277 & 0.00802 \\ & & & 0.0608 \end{bmatrix}.$$

2.A.1.7 *Model* **logitTSI**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 T + \beta_2 S + \beta_3 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -819.614 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 2.32 (df = 8; Sakamoto, 1991); AICc = 1647.29; ΔAICc = 4.01; AIC weight = 0.042 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 1.5917.
(3) MLE of β_1 = -0.12267.
(4) MLE of β_2 = 0.19077.
(5) MLE of β_3 = -0.11523.

continued ...

2.A.1.8 *Model* **BG**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} \sim \text{beta}(\alpha, \beta),$$

$$\text{i.e. } f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}.$$

(1) Model/data structure is

chain	general structure	structure under model **BG**
1-0	$(1-p)^4$	$(\beta+3)(\beta+2)(\beta+1)\beta\ [(\alpha+\beta+3)(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
1-1-0	$4p(1-p)^6$	$4\alpha(\beta+5)(\beta+4)...\beta\ [(\alpha+\beta+6)(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-1-1-0	$12p^2(1-p)^7$	$12(\alpha+1)\alpha(\beta+6)...\beta\ [(\alpha+\beta+8)...(\alpha+\beta)]^{-1}$
1-1-1-1-0	$24p^3(1-p)^7$	$24(\alpha+2)(\alpha+1)\alpha(\beta+6)...\beta\ [(\alpha+\beta+9)...(\alpha+\beta)]^{-1}$
1-1-1-1-1	$24p^4(1-p)^6$	$24(\alpha+3)...\alpha(\beta+5)...\beta\ [(\alpha+\beta+9)...(\alpha+\beta)]^{-1}$
1-1-1-2	$12p^4(1-p)^5$	$12(\alpha+3)...\alpha(\beta+4)...\beta\ [(\alpha+\beta+8)...(\alpha+\beta)]^{-1}$
1-1-2-0	$12p^3(1-p)^5$	$12(\alpha+2)...\alpha(\beta+4)...\beta\ [(\alpha+\beta+7)...(\alpha+\beta)]^{-1}$
1-1-2-1	$12p^4(1-p)^4$	$12(\alpha+3)...\alpha(\beta+3)...\beta\ [(\alpha+\beta+7)...(\alpha+\beta)]^{-1}$
1-1-3	$4p^4(1-p)^3$	$4(\alpha+3)...\alpha(\beta+2)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}$
1-2-0	$6p^2(1-p)^4$	$6(\alpha+1)...\alpha(\beta+3)...\beta\ [(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-2-1-0	$12p^3(1-p)^4$	$12(\alpha+2)...\alpha(\beta+3)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}$
1-2-1-1	$12p^4(1-p)^3$	$12(\alpha+3)...\alpha(\beta+2)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}$
1-2-2	$6p^4(1-p)^2$	$6(\alpha+3)...\alpha(\beta+1)\beta\ [(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-3-0	$4p^3(1-p)^2$	$4(\alpha+2)...\alpha(\beta+1)\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}$
1-3-1	$4p^4(1-p)$	$4(\alpha+3)...\alpha\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}$
1-4	p^4	$(\alpha+3)...\alpha\ [(\alpha+\beta+3)...(\alpha+\beta)]^{-1}$

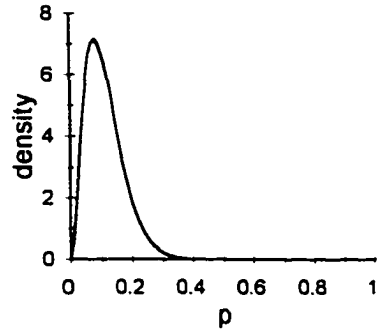
(2) Maximized log-likelihood ($\ln\mathcal{L}$) = -821.920 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 7.84 (df = 10; Sakamoto, 1991) AICc = 1647.86; ΔAICc = 4.48; AIC weight = 0.032 (Burnham and Anderson, 1998).

(3) Maximum likelihood estimate (MLE; Lehmann, 1983) of α = 2.8710.

(4) MLE of β = 22.737.

(5) *continued ...*

(5) *continued*



$\hat{E}(p) = 0.112;$

$\hat{V}ar(p) = 0.00373;$

$\hat{S}D(p) = 0.0612.$

2.A.1.9 *Model* T, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_T.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -820.203 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 = goodness-of-fit 6.43 (df = 8; Sakamoto, 1991); AICc = 1647.86; ΔAICc = 5.19; AIC weight = 0.023 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{5,1,4,1}$ = 0.10775; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) for $p_{5,1,4,1}$ is (0.096273, 0.11984).
(3) MLE's of $p_{5,2,3,1}$, $p_{5,2,2,2}$, and $p_{5,2,1,3}$ = 0.14459; 95% PLI is (0.11944, 0.17233).
(4) MLE's of $p_{5,3,2,1}$, $p_{5,3,1,1}$, and $p_{5,3,1,2}$ = 0.19833; 95% PLI is (0.13748, 0.27083).
(5) MLE of $p_{5,4,1,1}$ = 0.22241; 95% PLE is (0.074943, 0.44333).
(6) Estimated variance-covariance matrix from 8000 bootstrap replicates (Efron and Tibshirani, 1993), $\mathrm{vc}\left(\hat{p}_{5,1,4,1}, \hat{p}_{5,2,1,3}, \hat{p}_{5,3,1,2}, \hat{p}_{5,4,1,1}\right)$, is

$$\begin{bmatrix} 1.08E-4 & 5.72E-6 & -2.46E-5 & 1.63E-4 \\ & 1.94E-4 & 1.20E-5 & -2.39E-5 \\ & & 1.01E-3 & -5.00E-5 \\ & & & 1.01E-2 \end{bmatrix}.$$

Corresponding estimated standard error-correlation matrix is

$$\begin{bmatrix} 0.0104 & 0.0396 & -0.0744 & 0.118 \\ & 0.0139 & 0.0271 & -0.0129 \\ & & 0.0318 & -0.0118 \\ & & & 0.133 \end{bmatrix}.$$

2.A.1.10 *Model* **SI**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_{S,I}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -818.550 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 7; χ^2 goodness-of-fit = 1.36 (df = 5; Sakamoto, 1991); AICc = 1651.27; ΔAICc = 7.99; AIC weight = 0.006 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimates (MLE's; Anzzalini, 1996) of $p_{5,4,1,1}$ and $p_{5,3,1,1}$ = 0.21932.
(3) MLE of $p_{5,3,2,1}$ = 0.19644.
(4) MLE of $p_{5,2,3,1}$ = 0.13733.
(5) MLE of $p_{5,1,4,1}$ = 0.10767.
(6) MLE of $p_{5,3,1,2}$ = 0.20069.
(7) MLE of $p_{5,2,2,2}$ = 0.20507.
(8) MLE of $p_{5,2,1,3}$ = 6.1429E-7.

continued...

2.A.1.11 *Model* **TS**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_{T,S}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -818.553 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 7; χ^2 goodness-of-fit = 1.36 (df = 5; Sakamoto, 1991); AICc = 1651.28; ΔAICc = 8.00; AIC weight = 0.006 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{5,1,4,1}$ = 0.10768.
(3) MLE of $p_{5,2,1,3}$ = 5.2495E-7.
(4) MLE of $p_{5,2,2,2}$ = 0.20504.
(5) MLE of $p_{5,2,3,1}$ = 0.13732.
(6) MLE's of $p_{5,3,1,2}$ and $p_{5,3,1,1}$ = 0.20863.
(7) MLE of $p_{5,3,2,1}$ = 0.19633.
(8) MLE of $p_{5,4,1,1}$ = 0.22211.

2.A.1.12 *Model* **TI**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_{T,I}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -818.561 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 7; χ^2 goodness-of-fit = 1.38 (df = 5; Sakamoto, 1991); AICc = 1651.29; ΔAICc = 8.01; AIC weight = 0.006 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{5,1,4,1}$ = 0.10774; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) for $p_{5,1,4,1}$ is (0.096274, 0.11984).
(3) MLE of $p_{5,2,1,3}$ = 8.0508E-7; 95% PLI is (0.00000, 0.47263).
(4) MLE of $p_{5,2,2,2}$ = 0.20509; 95% PLI is (0.12590, 0.30340).
(5) MLE of $p_{5,2,3,1}$ = 0.13733; 95% PLI is (0.11132, 0.16649).
(6) MLE of $p_{5,3,1,2}$ = 0.20077; 95% PLI is (0.036388, 0.49934).
(7) MLE of $p_{5,3,2,1}$ and $p_{5,3,1,1}$ = 0.19846; 95% PLI is (0.13526, 0.27370).
(8) MLE of $p_{5,4,1,1}$ = 0.22282; 95% PLI is (0.074955, 0.44330).

continued...

2.A.1.13 *Model* **BRF**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = r^I \text{ and } r \sim \text{beta}(\alpha, \beta),$$

$$\text{i.e. } f(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1} .$$

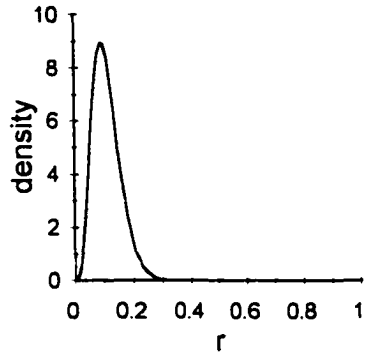
(1) Model/data structure is

chain	general structure	structure under model **BRF**
1-0	$(1-p)^4$	$(\beta+3)(\beta+2)(\beta+1)\beta\ [(\alpha+\beta+3)(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
1-1-0	$4p(1-p)^6$	$4\alpha(\beta+5)(\beta+4)...\beta\ [(\alpha+\beta+6)(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-1-1-0	$12p^2(1-p)^7$	$12(\alpha+1)\alpha(\beta+6)...\ \beta\ [(\alpha+\beta+8)...(\alpha+\beta)]^{-1}$
1-1-1-1-0	$24p^3(1-p)^7$	$24(\alpha+2)(\alpha+1)\alpha(\beta+6)...\beta\ [(\alpha+\beta+9)...(\alpha+\beta)]^{-1}$
1-1-1-1-1	$24p^4(1-p)^6$	$24(\alpha+3)...\alpha(\beta+5)...\beta\ [(\alpha+\beta+9)...(\alpha+\beta)]^{-1}$
1-1-1-2	$12p^4(1-p)^5$	$12(\alpha+3)...\alpha(\beta+4)...\beta\ [(\alpha+\beta+8)...(\alpha+\beta)]^{-1}$
1-1-2-0	$*12p^3(1-p)^6$	$12(\alpha+2)...\alpha(\beta+5)...\beta\ [(\alpha+\beta+8)...(\alpha+\beta)]^{-1}$
1-1-2-1	$*12p^3q^4(1-q^2)$, where q = 1-p	$12[\ (\alpha+2)...\alpha(\beta+3)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}$ - $(\alpha+2)...\alpha(\beta+5)...\beta\ [(\alpha+\beta+8)...(\alpha+\beta)]^{-1}\ ]$
1-1-3	$4p^4(1-p)^3$	$4(\alpha+3)...\alpha(\beta+2)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}$
1-2-0	$*6p^2(1-p)^6$	$6(\alpha+1)\alpha(\beta+5)...\beta\ [(\alpha+\beta+7)...(\alpha+\beta)]^{-1}$
1-2-1-0	$*12p^2q^5(1-q^2)$, where q = 1-p	$12(\alpha+1)\alpha(\beta+4)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1} \bullet$ $[\ 1 - (\beta+6)(\beta+5)\ [(\alpha+\beta+8)(\alpha+\beta+7)]^{-1}\ ]$
1-2-1-1	$*12p^3q^4(1-q^2)$, where q = 1-p	$12(\alpha+2)...\alpha(\beta+3)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1} \bullet$ $[\ 1 - (\beta+5)(\beta+4)\ [(\alpha+\beta+8)(\alpha+\beta+7)]^{-1}\ ]$
1-2-2	$*6p^2q^2(1-q^2)^2$, where q = 1-p	$6(\alpha+1)\alpha(\beta+1)\beta[(\alpha+\beta+3)...(\alpha+\beta)]^{-1} \bullet$ $[\ 1 - 2[\ (\beta+3)(\beta+2)\ [(\alpha+\beta+5)(\alpha+\beta+4)]^{-1}\ ] +$ $(\beta+5)...\beta\ [(\alpha+\beta+7)...(\alpha+\beta+4)]^{-1}\ ]$
1-3-0	$*4p^3(1-p)^4$	$4(\alpha+2)...\alpha(\beta+3)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}$
1-3-1	$4p^3q(1-q^3)$, where q = 1-p	$4[\ (\alpha+2)...\alpha\beta\ [(\alpha+\beta+3)...(\alpha+\beta)]^{-1}$ - $\alpha(\beta+5)...\beta\ [(\alpha+\beta+6)...(\alpha+\beta)]^{-1}\ ]$
1-4	p^4	$(\alpha+3)...\alpha\ [(\alpha+\beta+3)...(\alpha+\beta)]^{-1}.$

Stared (*) values are differences between model **BG** and model **BRF**.

(2) Maximized log-likelihood ($\ln\mathcal{L}$) = -823.752 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 14.2 (df = 10; Sakamoto, 1991); AICc = 1651.52; ΔAICc = 8.24; AIC weight = 0.005 (Burnham and Anderson, 1998).

(3) Maximum likelihood estimate (MLE; Lehmann, 1983) of $\alpha = 4.9713$.
(4) MLE of $\beta = 39.694$.
(5)



$\hat{E}(r) = 0.1113;$
$\hat{V}ar(r) = 0.002166;$
$\hat{S}D(r) = 0.0465.$

2.A.1.14 *Model* **TSI**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_{T,S,I}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$ = -818.549 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 8; χ^2 goodness-of-fit= 1.35 (df = 7; Sakamoto, 1991); AICc = 1653.32; ΔAICc = 10.04; AIC weight = 0.002 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{5,1,4,1} = 0.10768$; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) for $p_{5,1,4,1}$ is (0.096274, 0.11984).
(3) MLE of $p_{5,2,1,3} = 5.0002E\text{-}7$; 95% PLI is (0.00000, 0.47264).
(4) MLE of $p_{5,2,2,2} = 0.20512$; 95% PLI is (0.12590, 0.30341).
(5) MLE of $p_{5,2,3,1} = 0.13735$; 95% PLI is (0.11132, 0.16649).
(6) MLE of $p_{5,3,1,1} = 0.21428$; 95% PLI is (0.058138, 0.46616).
(7) MLE of $p_{5,3,1,2} = 0.20000$; 95% PLI is (0.036387, 0.49935).
(8) MLE of $p_{5,3,2,1} = 0.19643$; 95% PLI is (0.13019, 0.27639).
(9) MLE of $p_{5,4,1,1} = 0.22222$; 95% PLI is (0.074948, 0.44332).

continued...

2.A.1.15 *Model* **RF**, *Heasman and Reid data*

Model is

2.A.

$$q_{N,T,S,I} = r^I .$$

Model is

$$q_{N,T,S,I} = r^I .$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -826.489 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 1; χ^2 goodness-of-fit = 17.8 (df = 11; Sakamoto, 1991); AICc = 1654.50; ΔAICc = 11.22; AIC weight = 0.001 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (Lehmann, 1983) of r = 0.88383; 95% profile likelihood interval (Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) of r is (0.87305, 0.89406); 95% bootstrap confidence interval (Efron and Tibshirani, 1993) from 10,000 replicates is (0.874, 0.895).

2.A.1.16 *Model* **MM**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = 1 - a\left(1 - \frac{1}{N}\right)^{bI} .$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -826.021 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 15.7 (df = 7; Sakamoto, 1991); AICc = 1656.06; ΔAICc = 12.78; AIC weight = 0.001 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of a = 0.96437; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) for a is (0.95237, 0.97580); 95% bootstrap confidence interval (95% BCI; Efron and Tibshirani, 1993) from 10,000 replicates is (0.893, 1.00).
(3) MLE of b = 0.30646; 95% PLI is (0.26631, 0.34892); 95% BCI from 10,000 replicates is (0.270, 0.452).

continued ...

2.A.1.17 *Model* **I**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = p_I.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -825.071 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 14.0 (df = 8; Sakamoto, 1991); AICc = 1656.18; ΔAICc = 12.90; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimates (MLE's; Lehmann, 1983) of $p_{5,1,4,1}$, $p_{5,2,3,1}$, $p_{5,3,2,1}$, $p_{5,4,1,1}$, and $p_{5,3,1,1}$ = 0.11687.
(3) MLE's of $p_{5,3,1,2}$ and $p_{5,2,2,2}$ = 0.20455.
(4) MLE of $p_{5,2,1,3}$ = 5.0000E-7.

2.A.1.18 *Model* **logitI**, *Heasman and Reid data*

Model is

$$p_{N,T,S,I} = \frac{1}{1+\exp\left[-(\beta_0+\beta_1 I)\right]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -826.165 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 18.6 (df = 7; Sakamoto, 1991); AICc = 1656.46; ΔAICc = 13.18; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = 2.5600.
(3) MLE of β_1 = -0.53984.

2.A.1.19 *Model* **G**, *Heasman and Reid data*

Model is

$$q_{N,T,S,I} = r.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -828.125 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 1; χ^2 goodness-of-fit = 22.2 (df = 12; Sakamoto, 1991); AICc = 1658.26; ΔAICc = 14.98; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of r = 0.88102; 95% profile likelihood interval (Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) is (0.87000, 0.89148).

2.A.1.20 *Model* **MULT**, *Heasman and Reid data*

Model is each *class* of chain has a unique probability of occurrence.

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -816.660 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 15; AICc = 1666.06; ΔAICc = 22.78; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of P(observe chain 1-0 in one trial) = 423/664 = 0.63705; estimated standard error (SE; sqrt($\widehat{pq}/n$)) of the MLE of P(observe chain 1-0 in one trial) = 0.018660.
(3) MLE of P(... 1-1-0 ...) = 131/664 = 0.19729; SE = 0.015444.
(4) MLE of P(... 1-1-1-0 ...) = 36/664 = 0.054217; SE = 0.0087878.
(5) MLE of P(... 1-1-1-1-0 ...) = 14/664 = 0.021084; SE = 0.0056612.
(6) MLE of P(... 1-1-1-1-1 ...) = 4/664 = 0.0060241; SE = 0.0030030.
(7) MLE of P(... 1-1-1-2 ...) = 2/664 = 0.0030120; SE = 0.0021266.
(8) MLE of P(... 1-1-2-0 ...) = 8/664 = 0.012048; SE = 0.0042339.
(9) MLE of P(... 1-1-2-1 ...) = 2/664 = 0.0030120; SE = 0.0021266.
(10) MLE of P(... 1-1-3-0 ...) = 2/664 = 0.0030120; SE = 0.0021266.
(11) MLE of P(... 1-2-0 ...) = 24/664 = 0.036145; SE = 0.0073551.
(12) MLE of P(... 1-2-1-0 ...) = 11/664 = 0.016567; SE = 0.0049535.
(13) MLE of P(... 1-2-1-1 ...) = 3/664 = 0.0045181; SE = 0.0026026.
(14) MLE of P(... 1-2-2 ...) = 1/664 = 0.0015060; SE = 0.0015049.
(15) MLE of P(... 1-3-0 ...) = 3/664 = 0.0045181; SE = 0.0026026.
(16) MLE of P(... 1-3-1 ...) = 0/664 = 0; 95% profile likelihood interval (95% PLI; Kalbfleisch and Sprott, 1970; Venazon and Moolgavkar, 1988) is (0, 0.003).
(17) MLE of P(... 1-4 ...) = 0/664 = 0; 95% PLI is (0, 0.003).

continued ...

2.A.2. Providence data: model structures, parameter estimates, miscellaneous interval estimates, and miscellaneous estimated variance-covariance matrices

2.A.2.1 *Model* **BG**, *Providence data*

Model is

$$p_{N,T,S,I} = p, \text{ with } p \sim \text{beta}(\alpha, \beta),$$
$$\text{i.e. } f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}.$$

(1) Model/data structure is

(size) chain	general structure	structure under model **BG**
(3) 1-0	$(1-p)^2$	$(\beta+1)\beta\ [(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
1-1-0	$2p(1-p)^2$	$2\alpha(\beta+1)\beta\ [(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
1-1-1	$2p^2(1-p)$	$2(\alpha+1)\alpha\beta\ [(\alpha+\beta+2)...(\alpha+\beta)]^{-1}$
1-2	p^2	$(\alpha+1)\alpha\ [(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
(4) 1-0	$(1-p)^3$	$(\beta+2)(\beta+1)\beta\ [(\alpha+\beta+2)...(\alpha+\beta)]^{-1}$
1-1-0	$3p(1-p)^4$	$3\alpha(\beta+3)...\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}$
1-1-1-0	$6p^2(1-p)^4$	$6(\alpha+1)\alpha(\beta+3)...\beta\ [(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-2	$3p^2(1-p)^2$	$3(\alpha+1)\alpha(\beta+1)\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]$
1-1-1-1	$6p^3(1-p)^3$	$6(\alpha+2)...\alpha(\beta+2)...\beta\ [(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-1-2	$3p^3(1-p)^2$	$3(\alpha+2)...\alpha(\beta+1)\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}$
1-2-1	$3p^3(1-p)$	$3(\alpha+2)(\alpha+1)\alpha\beta\ [(\alpha+\beta+3)...(\alpha+\beta)]^{-1}$
1-3	p^3	$(\alpha+2)...\alpha\ [(\alpha+\beta+2)...(\alpha+\beta)]^{-1}$.

(2) Maximized log-likelihood ($\ln\mathcal{L}$) = -427.025 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 6.13 (df = 8; Sakamoto, 1991); AICc = 858.08; ΔAICc = 0; AIC weight = 0.824 (Burnham and Anderson, 1998).

(3) Maximum likelihood estimate (MLE; Lehmann, 1983) of α = 1.2298.

(4) MLE of β = 0.28963.

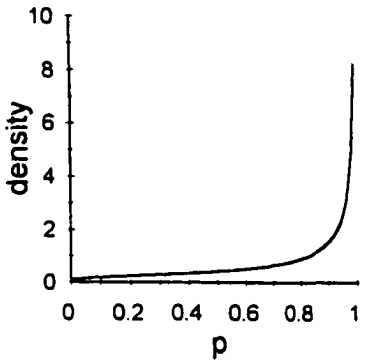
(5) Estimated variance-covariance matrix from 8000 bootstrap replicates (Efron and Tibshirani, 1993), $\text{vc}(\hat{\alpha}, \hat{\beta})$, is

$$\begin{bmatrix} 0.0633 & 0.0133 \\ & 0.00360 \end{bmatrix}.$$

Corresponding estimated standard error-correlation matrix is

$$\begin{bmatrix} 0.252 & 0.880 \\ & 0.0600 \end{bmatrix}.$$

(6)



$\hat{E}(p) = 0.8094;$

$\hat{V}ar(p) = 0.06124;$

$\hat{S}D(p) = 0.2475.$

continued ...

2.A.2.2 *Model* **BGN**, *Providence data*

Model is

$$p_{N,T,S,I} = p_N \text{ with } p_3 \sim \text{beta}(\alpha_3,\beta_3), \text{ and } p_4 \sim \text{beta}(\alpha_4,\beta_4),$$

$$\text{e.g. } f(p_3) = \frac{\Gamma(\alpha_3+\beta_3)}{\Gamma(\alpha_3)\Gamma(\beta_3)} p_3^{\alpha_3-1}(1-p_3)^{\beta_3-1}.$$

(1) Model/data structure is

(size) chain	general structure	structure under model **BGN**
(3) 1-0	$(1-p_3)^2$	$(\beta_3+1)\beta_3\ [(\alpha_3+\beta_3+1)(\alpha_3+\beta_3)]^{-1}$
1-1-0	$2p_3(1-p_3)^2$	$2\alpha_3(\beta_3+1)\beta_3\ [(\alpha_3+\beta_3+2)(\alpha_3+\beta_3+1)(\alpha_3+\beta_3)]^{-1}$
1-1-1	$2p_3^2(1-p_3)$	$2(\alpha_3+1)\alpha_3\beta_3\ [(\alpha_3+\beta_3+2)...(\alpha_3+\beta_3)]^{-1}$
1-2	p_3^2	$(\alpha_3+1)\alpha_3\ [(\alpha_3+\beta_3+1)(\alpha_3+\beta_3)]^{-1}$
(4) 1-0	$(1-p_4)^3$	$(\beta_4+2)(\beta_4+1)\beta_4\ [(\alpha_4+\beta_4+2)...(\alpha_4+\beta_4)]^{-1}$
1-1-0	$3p_4(1-p_4)^4$	$3\ \alpha_4(\beta_4+3)...\beta_4\ [(\alpha_4+\beta_4+4)(\alpha_4+\beta_4+3)...(\alpha_4+\beta_4)]^{-1}$
1-1-1-0	$6p_4^2(1-p_4)^4$	$6(\alpha_4+1)\alpha_4(\beta_4+3)...\beta_4\ [(\alpha_4+\beta_4+5)...(\alpha_4+\beta_4)]^{-1}$
1-2	$3p_4^2(1-p_4)^2$	$3(\alpha_4+1)\alpha_4(\beta_4+1)\beta_4\ [(\alpha_4+\beta_4+4)...(\alpha_4+\beta_4)]^{-1}$
1-1-1-1	$6p_4^3(1-p_4)^3$	$6(\alpha_4+2)...\alpha_4(\beta_4+2)...\beta_4\ [(\alpha_4+\beta_4+5)...(\alpha_4+\beta_4)]^{-1}$
1-1-2	$3p_4^3(1-p_4)^2$	$3(\alpha_4+2)...\alpha_4(\beta_4+1)\beta_4\ [(\alpha_4+\beta_4+4)...(\alpha_4+\beta_4)]^{-1}$
1-2-1	$3p_4^3(1-p_4)$	$3(\alpha_4+2)...\alpha_4\beta_4\ [(\alpha_4+\beta_4+3)...(\alpha_4+\beta_4)]^{-1}$
1-3	p_4^3	$(\alpha_4+2)...\alpha_4\ [(\alpha_4+\beta_4+2)...(\alpha_4+\beta_4)]^{-1}$.

(2) Maximized log-likelihood ($\ln\mathcal{L}$) = -426.198 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 5.30 (df = 6; Sakamoto, 1991); AICc = 861.30; ΔAICc = 3.22; AIC weight = 0.165 (Burnham and Anderson, 1998).

(3) Maximum likelihood estimate (MLE; Lehmann, 1983) of α_3 = 1.0907; 95% bootstrap confidence interval (95% BCI; Efron and Tibshirani, 1993) from 8,000 replicates for α_3 is (0.695, 1.79).

(4) MLE of β_3 = 0.26396; 95% BCI is (0.165, 0.435).

(5) MLE of α_4 = 1.5794; 95% BCI is (0.880, 3.23).

(6) MLE of β_4 = 0.34081; 95% BCI is (0.176, 0.701).

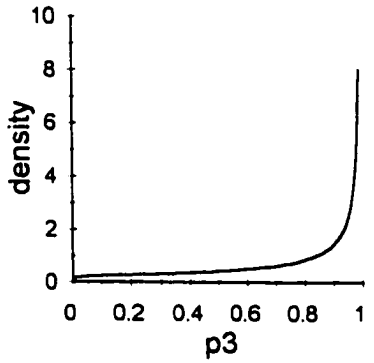
(7) Estimated variance-covariance matrix from 8000 bootstrap replicates (Efron and Tibshirani, 1993), $\text{vc}\left(\hat{\alpha}_3,\hat{\beta}_3,\hat{\alpha}_4,\hat{\beta}_4\right)$, is

$$\begin{bmatrix} 0.0809 & 0.0174 & -0.00271 & -0.000215 \\ & 0.00470 & -0.000391 & -4.97E-5 \\ & & 0.367 & 0.0702 \\ & & & 0.0190 \end{bmatrix}.$$

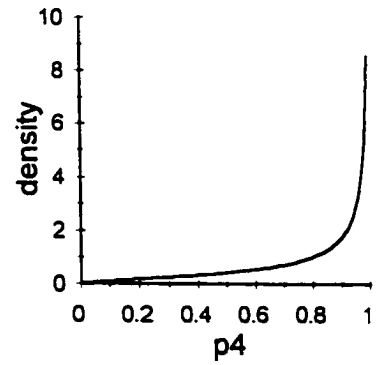
Corresponding estimated standard error-correlation matrix is

$$\begin{bmatrix} 0.284 & 0.893 & -0.0497 & -0.00549 \\ & 0.0686 & -0.0297 & -0.00525 \\ & & 0.604 & 0.842 \\ & & & 0.138 \end{bmatrix}.$$

(8)



$\hat{E}(p_3) = 0.80515;$
$\hat{V}ar(p_3) = 0.066624;$
$\hat{S}D(p_3) = 0.25812.$



$\hat{E}(p_4) = 0.82251;$
$\hat{V}ar(p_4) = 0.049990;$
$\hat{S}D(p_4) = 0.22359.$

continued ...

2.A.2.3 *Model* **BRF** (*or* **BPP**), *Providence data*

Model is

$$q_{N,T,S,I} = r^I \text{ with } r \sim \text{beta}(\alpha, \beta),$$

$$\text{i.e. } f(r) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} r^{\alpha-1}(1-r)^{\beta-1} .$$

(1) Model/data structure is

(size) chain	general structure	structure under model **BRF**
(3) 1-0	$(1-p)^2$	$(\beta+1)\beta\ [(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
1-1-0	$2p(1-p)^2$	$2\alpha(\beta+1)\beta\ [(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
1-1-1	$2p^2(1-p)$	$2(\alpha+1)\alpha\beta\ [(\alpha+\beta+2)...(\alpha+\beta)]^{-1}$
1-2	p^2	$(\alpha+1)\alpha\ [(\alpha+\beta+1)(\alpha+\beta)]^{-1}$
(4) 1-0	$(1-p)^3$	$(\beta+2)(\beta+1)\beta\ [(\alpha+\beta+2)...(\alpha+\beta)]^{-1}$
1-1-0	$3p(1-p)^4$	$3\alpha(\beta+3)...\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}$
1-1-1-0	$6p^2(1-p)^4$	$6(\alpha+1)\alpha(\beta+3)...\beta\ [(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-2	$*3p^2(1-p)^3$	$3(\alpha+1)\alpha(\beta+2)...\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]$
1-1-1-1	$6p^3(1-p)^3$	$6(\alpha+2)...\alpha(\beta+2)...\beta\ [(\alpha+\beta+5)...(\alpha+\beta)]^{-1}$
1-1-2	$3p^3(1-p)^2$	$3(\alpha+2)...\alpha(\beta+1)\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}$
1-2-1	$*3p^3q(1+q)$, where $q = 1-p$	$3[\ (\alpha+2)(\alpha+1)\alpha\beta\ [(\alpha+\beta+3)...(\alpha+\beta)]^{-1} + (\alpha+2)(\alpha+1)\alpha(\beta+1)\beta\ [(\alpha+\beta+4)...(\alpha+\beta)]^{-1}\ ]$
1-3	p^3	$(\alpha+2)...\alpha\ [(\alpha+\beta+2)...(\alpha+\beta)]^{-1}$.

Stared (*) values are differences between model **BRF** and model **BG**.

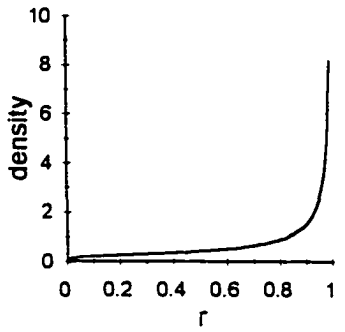
(2) Maximized log-likelihood ($\ln\mathcal{L}$) = -431.928 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 22.5 (df = 8; Sakamoto, 1991); AICc = 867.89; ΔAICc = 9.81; AIC weight = 0.006 (Burnham and Anderson, 1998).

(3) Maximum likelihood estimate (MLE; Lehmann, 1983) of α = 1.1727.

(4) MLE of β = 0.27985.

(5) *continued*...

(5) *continued*



$\hat{E}(r) = 0.80734;$

$\hat{V}ar(r) = 0.06342;$

$\hat{S}D(r) = 0.25183.$

2.A.2.4 *Model* **MULT**, *Providence data*

Model is that each *class* of chain has a unique probability of occurrence.

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -424.133 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 10; AICc = 868.79; ΔAICc = 10.71; AIC weight = 0.004 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of P(observe chain 1-0 in one trial of household size 3) = 34/334 = 0.1018; estimated standard error (SE; $\text{sqrt}(\hat{p}\hat{q}/n)$) of the MLE of P(observe chain 1-0 in one trial of household size 3) = 0.030239.
(3) MLE of P(... 1-1-0 ... size 3) = 25/334 = 0.07485; SE = 0.014399.
(4) MLE of P(... 1-1-1 ... size 3) = 36/334 = 0.1078; SE = 0.016969.
(5) MLE of P(... 1-2 ... size 3) = 239/334 = 0.7156; SE = 0.024685.
(6) MLE of P(... 1-0 ... size 4) = 4/100 = 0.04; SE = 0.010722.
(7) MLE of P(... 1-1-0 ... size 4) = 3/100 = 0.03 SE = 0.017059.
(8) MLE of P(... 1-1-1-0 ... size 4) = 1/100 = 0.01; SE = 0.0099499.
(9) MLE of P(... 1- 1-1-1 ... size 4) = 4/100 = 0.04; SE = 0.010722.
(10) MLE of P(...1-2-0 ... size 4) = 8/100 = 0.08; SE = 0.085790.
(11) MLE of P(... 1-1-2 ...size 4) = 3/100 = 0.03; SE = 0.017059.
(12) MLE of P(... 1-2-1 ...size 4) = 1/100 = 0.01; SE = 0.0099499.
(13) MLE of P(... 1-3 ... size 4) = 67/100 = 0.67; SE = 0.047021.

continued...

2.A.2.5 *Model* **BRFN**, *Providence data*

Model is

$$q_{N,T,S,I} = r_N^I \text{ and } r_N \sim \text{beta}(\alpha_N, \beta_N),$$

$$\text{i.e. } f(r_N) = \frac{\Gamma(\alpha_N + \beta_N)}{\Gamma(\alpha_N)\Gamma(\beta_N)} r_N^{\alpha_N - 1}(1 - r_N)^{\beta_N - 1} .$$

(1) Model/data structure is

(size) chain	general structure	structure under model **BRF**
(3) 1-0	$(1-p_3)^2$	$(\beta_3+1)\beta_3\ [(\alpha_3+\beta_3+1)(\alpha_3+\beta_3)]^{-1}$
1-1-0	$2\,p_3(1-p_3)^2$	$2\,\alpha_3(\beta_3+1)\beta_3\ [(\alpha_3+\beta_3+2)(\alpha_3+\beta_3+1)(\alpha_3+\beta_3)]^{-1}$
1-1-1	$2\,p_3^2(1-p_3)$	$2(\alpha_3+1)\alpha_3\beta_3\ [(\alpha_3+\beta_3+2)...(\alpha_3+\beta_3)]^{-1}$
1-2	p_3^2	$(\alpha_3+1)\alpha_3\ [(\alpha_3+\beta_3+1)(\alpha_3+\beta_3)]^{-1}$
(4) 1-0	$(1-p_4)^3$	$(\beta_4+2)(\beta_4+1)\beta_4\ [(\alpha_4+\beta_4+2)...(\alpha_4+\beta_4)]^{-1}$
1-1-0	$3p_4(1-p_4)^4$	$3\alpha_4(\beta_4+3)...\beta_4\ [(\alpha_4+\beta_4+4)...(\alpha_4+\beta_4)]^{-1}$
1-1-1-0	$6p_4^2(1-p_4)^4$	$6(\alpha_4+1)\alpha_4(\beta_4+3)...\beta_4\ [(\alpha_4+\beta_4+5)...(\alpha_4+\beta_4)]^{-1}$
1-2	$*3p_4^2(1-p_4)^3$	$3(\alpha_4+1)\alpha_4(\beta_4+2)...\beta_4\ [(\alpha_4+\beta_4+4)...(\alpha_4+\beta_4)]$
1-1-1-1	$6p_4^3(1-p_4)^3$	$6(\alpha_4+2)...\alpha_4(\beta_4+2)...\beta_4\ [(\alpha_4+\beta_4+5)...(\alpha_4+\beta_4)]^{-1}$
1-1-2	$3p_4^3(1-p_4)^2$	$3(\alpha_4+2)...\alpha_4(\beta_4+1)\beta_4\ [(\alpha_4+\beta_4+4)...(\alpha_4+\beta_4)]^{-1}$
1-2-1	$*3p_4{}^3q_4(1+q_4)$, where $q_4 = 1-p_4$	$3[\ (\alpha_4+2)(\alpha_4+1)\alpha_4\beta_4\ [(\alpha_4+\beta_4+3)...(\alpha_4+\beta_4)]^{-1} + (\alpha_4+2)(\alpha_4+1)\alpha_4(\beta_4+1)\beta_4\ [(\alpha_4+\beta_4+4)...(\alpha_4+\beta_4)]^{-1}\ ]$
1-3	p_4^3	$(\alpha_4+2)...\alpha_4\ [(\alpha_4+\beta_4+2)...(\alpha_4+\beta_4)]^{-1}$.

Stared (*) values are differences between model **BRFN** and model **BGN**.

(2) Maximized log-likelihood ($\ln\mathcal{L}$) = -431.809 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 22.3 (df = 6; Sakamoto, 1991); AICc = 871.72; ΔAICc = 13.64; AIC weight = 0.001 (Burnham and Anderson, 1998).

(3) Maximum likelihood estimate (MLE; Lehmann, 1983) of α_3 = 1.0908.

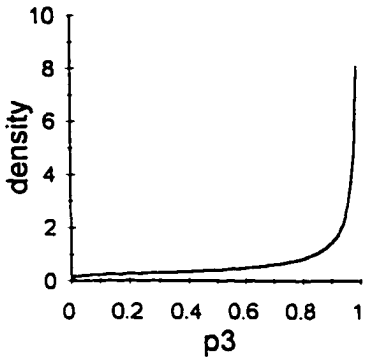
(4) MLE of β_3 = 0.26397.

(5) MLE of α_4 = 1.3196.

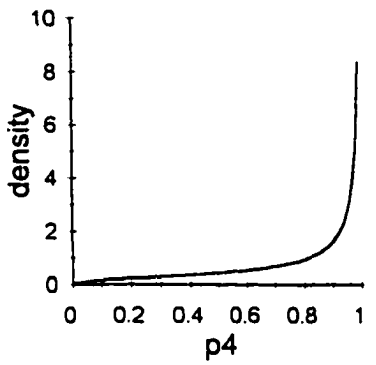
(6) MLE of β_4 = 0.30235.

(7) *continued...*

(7) *continued*



$\hat{E}(p_3) = 0.81359;$
$\hat{V}ar(p_3) = 0.057843;$
$\hat{S}D(p_3) = 0.24051.$



$\hat{E}(p_4) = 0.80516;$
$\hat{V}ar(p_4) = 0.066622;$
$\hat{S}D(p_4) = 0.25811.$

2.A.2.6 *Model* T, *Providence data*

Model is

$$p_{N,T,S,I} = p_T.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -460.465 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 83.0 (df = 6; Sakamoto, 1991); AICc = 926.99; ΔAICc = 68.91; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{4,1,2,1}$ and $p_{3,1,2,1}$ = 0.81296.
(3) MLE of $p_{4,2,1,2}$, $p_{4,2,2,1}$, and $p_{3,2,1,1}$ = 0.56446.
(4) MLE of $p_{4,3,1,1}$ = 0.79933.

continued...

2.A.2.7 *Model* **TI**, *Providence data*

Model is

$$p_{N,T,S,I} = p_{T,I}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -460.461 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 81.5 (df = 4; Sakamoto, 1991); AICc = 929.01; ΔAICc = 70.93; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{4,1,3,1}$ and $p_{3,1,2,1}$ = 0.81302.
(3) MLE of $p_{4,2,2,1}$ and $p_{3,2,1,1}$ = 0.56627.
(4) MLE of $p_{4,2,1,2}$ = 0.55556.
(5) MLE of $p_{4,3,1,1}$ = 0.80000.

2.A.2.8 *Model* **logitT**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 T)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -462.914 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 89.8 (df = 6; Sakamoto, 1991); AICc = 930.06; ΔAICc = 71.98; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -2.4709.
(3) MLE of β_1 = 1.0161.

continued...

2.A.2.9 *Model* **TS**, *Providence data*

Model is

$$p_{N,T,S,I} = p_{T,S}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -460.209 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 5; χ^2 goodness-of-fit = 81.6 (df = 5; Sakamoto, 1991); AICc = 930.56; ΔAICc = 72.48; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,1,2,1} = 0.81140$.
(3) MLE of $p_{4,1,3,1} = 0.81690$.
(4) MLE of $p_{3,2,1,1}$ and $p_{4,2,1,2} = 0.58228$.
(5) MLE of $p_{4,2,2,1} = 0.50000$.
(6) MLE of $p_{4,3,1,1} = 0.80000$.

2.A.2.10 *Model* **NT**, *Providence data*

Model is

$$p_{N,T,S,I} = p_{N,T}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -460.236 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 5; χ^2 goodness-of-fit = 78.8 (df = 3; Sakamoto, 1991); AICc = 930.61; ΔAICc = 72.53; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,1,2,1} = 0.81140$.
(3) MLE of $p_{3,2,1,1} = 0.59016$.
(4) MLE of $p_{4,1,3,1} = 0.81678$.
(5) MLE of $p_{4,2,2,1}$ and $p_{4,2,1,2} = 0.52513$.
(6) MLE of $p_{4,3,1,1} = 0.79923$.

continued...

2.A.2.11 *Model* **logitTS**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 T + \beta_2 S)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -462.622 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit= 90.1 (df = 5; Sakamoto, 1991); AICc = 931.30; ΔAICc = 73.22; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -2.0488.
(3) MLE of β_1 = 0.88710
(4) MLE of β_2 = -0.12766.

2.A.2.12 *Model* **logitNT**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 N + \beta_2 T)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -462.736 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); K = 3; χ^2 goodness-of-fit = 89.9 (df = 5; Sakamoto, 1991); AICc = 931.54; ΔAICc = 73.46; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -2.0488.
(3) MLE of β_1 = -0.097919.
(4) MLE of β_2 = 1.0293.

continued...

2.A.2.13 *Model* **NS**, *Providence data*

Model is

$$p_{N,T,S,I} = p_{N,S}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -460.702 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 5; χ^2 goodness-of-fit = 76.8 (df = 5; Sakamoto, 1991); AICc = 931.53; ΔAICc = 73.45; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,2,1,1}$ = 0.59016.
(3) MLE of $p_{3,1,2,1}$ = 0.81140.
(4) MLE of $p_{4,2,1,2}$ and $p_{4,3,1,1}$ = 0.60870.
(5) MLE of $p_{4,2,2,1}$ = 0.50000.
(6) MLE of $p_{4,1,3,1}$ = 0.81690.

2.A.2.14 *Model* **logitTI**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 T + \beta_2 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -462.794 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 89.8 (df = 5; Sakamoto, 1991); AICc = 931.65; ΔAICc = 73.57; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -2.6877.
(3) MLE of β_1 = 0.97794.
(4) MLE of β_2 = 0.25429.

continued...

2.A.2.15 *Model* **NTSI** (*or* **NSI, TSI, NTS, NTI**), *Providence data*

Model is

$$p_{N,T,S,I} = p_{N,S,T,I}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -459.926 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 6; χ^2 goodness-of-fit = 77.9 (df = 2; Sakamoto, 1991); AICc = ; ΔAICc = 73.97; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,1,2,1}$ = 0.80689.
(3) MLE of $p_{3,2,1,1}$ = 0.59016.
(4) MLE of $p_{4,1,3,1}$ = 0.82667.
(5) MLE of $p_{4,2,1,2}$ = 0.55556.
(6) MLE of $p_{4,2,2,1}$ = 0.50000.
(7) MLE of $p_{4,3,1,1}$ = 0.80000.

2.A.2.16 *Model* **logitNS**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 N + \beta_2 S)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -463.060 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 75.0 (df = 5; Sakamoto, 1991); AICc = 931.18; ΔAICc = 74.10; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -1.8013.
(3) MLE of β_1 = 0.69507.
(4) MLE of β_2 = -0.84351.

continued...

2.A.2.17 *Model* **logitNTSI** (*or* **logitNSI, logitTSI, logitNTS, logitNTI**), *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0 + \beta_1 N + \beta_2 S + \beta_3 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -462.537 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 77.5 (df = 4; Sakamoto, 1991); AICc = 933.17; ΔAICc = 75.09; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -1.3932.
(3) MLE of β_1 = 0.86255.
(4) MLE of β_2 = -0.98273.
(5) MLE of β_3 = 0.64868.

2.A.2.18 *Model* **S**, *Providence data*

Model is

$$p_{N,T,S,I} = p_S.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -465.456 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 88.7 (df = 4; Sakamoto, 1991); AICc = 936.97; ΔAICc = 78.89; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,2,1,1}$, $p_{4,2,1,2}$, and $p_{4,3,1,1}$ = 0.59546.
(3) MLE of $p_{3,1,2,1}$ and $p_{4,2,2,1}$ = 0.79706.
(4) MLE of $p_{4,1,3,1}$ = 0.82657.

continued ...

2.A.2.19 *Model* **SI**, *Providence data*

Model is

$$p_{N,T,S,I} = p_{S,I}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -465.381 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 4; χ^2 goodness-of-fit = 88.8 (df = 3; Sakamoto, 1991); AICc = 938.86; ΔAICc = 80.78; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,2,1,1}$ and $p_{4,3,1,1}$ = 0.60608.
(3) MLE of $p_{4,2,1,2}$ = 0.55607.
(4) MLE of $p_{3,1,2,1}$ and $p_{4,2,2,1}$ = 0.79711.
(5) MLE of $p_{4,1,3,1}$ = 0.82670.

2.A.2.20 *Model* **logitS**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 S)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -468.284 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 93.8 (df = 5; Sakamoto, 1991); AICc = 940.60; ΔAICc = 82.52; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -0.26118.
(3) MLE of β_1 = -0.49194.

continued...

2.A.2.21 *Model* **logitSI**, *Providence data*

Model is:

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 S+\beta_2 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -467.597 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 92.2 (df = 4; Sakamoto, 1991); AICc = 941.25; ΔAICc = 83.17; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -0.98747.
(3) MLE of β_1 = -0.44029.
(4) MLE of β_2 = 0.60231.

2.A.2.22 *Model* **G**, *Providence data*

Model is

$$q_{N,T,S,I} = r.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -470.999 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 1; χ^2 goodness-of-fit = 101. (df = 7; Sakamoto, 1991); AICc = 944.01; ΔAICc = 85.93; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (Lehmann, 1983) of r = 0.20764.

2.A.2.23 *Model* **logitI**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 I)]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -472.474 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 99.0. (df = 6; Sakamoto, 1991); AICc = 948.98; ΔAICc = 90.90; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -2.4706.
(3) MLE of β_1 = 1.1239.

2.A.2.24 *Model* **I**, *Providence data*

Model is

$$p_{N,T,S,I} = p_I \cdot$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -472.474 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 99.0. (df = 5; Sakamoto, 1991); AICc = 949.01; ΔAICc = 90.93; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,2,1,1}$, $p_{3,1,2,1}$, $p_{4,3,1,1}$, $p_{4,2,2,1}$, and $p_{4,1,3,1}$ = 0.79352.
(3) MLE of $p_{4,2,1,2}$ = 0.55552.

2.A.2.25 *Model* **NI** (*or* **logitNI**), *Providence data*

Model is

$$p_{N,T,S,I} = p_{N,I} \cdot$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -472.307 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 3; χ^2 goodness-of-fit = 101. (df = 5; Sakamoto, 1991); AICc = 950.67; ΔAICc = 92.59; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,1,2,1}$, and $p_{3,2,1,1}$ = 0.78875.
(3) MLE of $p_{4,3,1,1}$, $p_{4,2,2,1}$, and $p_{4,1,3,1}$ = 0.80424.
(4) MLE of $p_{4,2,1,2}$ = 0.55552.

continued ...

2.A.2.26 *Model* **logit(I/N)**, *Providence data*

Model is

$$q_{N,T,S,I} = \frac{1}{1+\exp\left[-\left(\beta_0 + \beta_1 \frac{I}{N}\right)\right]}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -473.365 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 95.6. (df = 6; Sakamoto, 1991); AICc = 950.76; ΔAICc = 92.68; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of β_0 = -2.2353.
(3) MLE of β_1 = 2.9184.

2.A.2.27 *Model* **N** (*or* **GN**, **logitN**)

Model is

$$p_{N,T,S,I} = p_N.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -474.986 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 = 99.7 (df = 6; Sakamoto, 1991); AICc = 954.00; ΔAICc = 95.92; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of $p_{3,1,2,1}$ and $p_{3,2,1,1}$ = 0.78876.
(3) MLE of $p_{4,3,1,1}$, $p_{4,2,2,1}$, $p_{4,1,3,1}$, and $p_{4,2,1,2}$ = 0.79131.

2.A.2.28 *Model* **MM**

Model is

$$p_{N,T,S,I} = 1 - a\left(1 - \frac{1}{N}\right)^{bI}.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -474.991 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 99.4 (df = 6; Sakamoto, 1991); AICc = 954.01; ΔAICc = 95.93; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE; Lehmann, 1983) of a = 0.21044.
(3) MLE of b = 0.23150E -8.

2.A.2.29 *Model* **RF** (*or* **PP**)

Model is

$$q_{N,T,S,I} = r^I.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -485.386 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 1; χ^2 goodness-of-fit = 125. (df = 7; Sakamoto, 1991); AICc = 972.26; ΔAICc = 114.18; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (Lehmann, 1983) of r = 0.78244.

2.A.2.30 *Model* **RFN**

Model is

$$q_{N,T,S,I} = r_N^I.$$

(1) Maximized log-likelihood ($\ln\mathcal{L}$) = -485.250 (Edwards, 1976; Anzzalini, 1996; Royall, 1997); number of parameters (K) = 2; χ^2 goodness-of-fit = 119. (df = 6; Sakamoto, 1991); AICc = 974.28; ΔAICc = 116.20; AIC weight = 0.000 (Burnham and Anderson, 1998).
(2) Maximum likelihood estimate (MLE) of r_3 = 0.78873.
(3) MLE of r_4 = 0.76934.

2.A.3. Miscellaneous methods and derivations

2.A.3.1 *Comments on FORTRAN code and numerical methods*

We used Powell's method written in FORTRAN 77 (Powell, 1964; Press et al., 1992) to perform all numerical optimizations. We executed these computations on an 166+ Cyrix 686 micro-computer running the linux 1.2.13 operating system (Peterson, 1998; http://metlab.unc.edu/LDP). We used the "f2c" compiler, which translates FORTRAN 77 to C (Feldman and Weinberger, 1990) followed by "gcc" to make each executable. We verified hardware and compiler reliability by compiling and executing numerous examples of the FORTRAN code with "xlf" on an IBM RISC System/6000 (lamar.colostate.edu)

mini-computer running the AIX operating system (Hoskins and Davies, 1999). We also verified the reliability of Powell's optimization method using "proc NLP" of SAS (Cody and Smith, 1997; http://www.sas.com) on this system; We compared results obtained using Powell's method with the results of several optimization methods available with "proc NLP"; we found no discrepancies.

2.A.3.2 *Example calculation of chain probabilities under model* **BG**.

For example, $3pq^4$ under model BG implies

$$\frac{3\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 pq^4 p^{\alpha-1}q^{\beta-1}dp$$

$$= \frac{3\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta+4)}{\Gamma(\alpha+\beta+5)}\int_0^1 \frac{\Gamma(\alpha+\beta+5)}{\Gamma(\alpha+1)\Gamma(\beta+4)}p^{(\alpha+1)-1}q^{(\beta+4)-1}dp.$$

Now, since

$$\int_0^1 \frac{\Gamma(\alpha+\beta+5)}{\Gamma(\alpha+1)\Gamma(\beta+4)}p^{(\alpha+1)-1}q^{(\beta+4)-1}dp = 1,$$

we get

$$\frac{3\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta+4)}{\Gamma(\alpha+\beta+5)}.$$

Recalling that $\Gamma(x+n+1) = (x+n)\Gamma(x+n)$,

where $x \in \Re$ and $n \in$ Integers, gives

$$\frac{3\alpha(\beta+3)(\beta+2)(\beta+1)\beta}{(\alpha+\beta+4)(\alpha+\beta+3)(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)}.$$

Other chain probabilities, under any of **BG**, **BGN**, **BRF**, or **BRFN**, are derived similarly.

REFERENCES (*appendix, chapter 2*)

Anzzalini, A. (1996). *Statistical inference - based on the likelihood.* London: Chapman and Hall.

Cody, R. P. and Smith, J. K. (1997). *Applied statistics and the SAS programing language*. 4th edition. Prentice Hall College Division.

Edwards, A. W. F. (1976). *Likelihood: an account of the statistical concept of likelihood and its application to scientific inference*. London: Cambridge University Press.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London: Chapman and Hall.

Feldman, S. I. and Weinberger, P. J. (1990). A portable FORTRAN 77 compiler. In *UNIX Time sharing systems programer's manual* - 10th edition, volume 2. AT&T Bell Laboratories.

Hoskins, J. and Davies, D. (1999). *Exploring IBM RS/6000 computers: the instant insider's guid to IBM's unix workstations and servers*. Maximum Press.

Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society, Series B* **32**, 175-208.

Royall, R. M. (1997). *Statistical evidence: a likelihood paradigm*. London: Chapman and Hall.

Peterson, R. (1998). Linux: the complete reference. Osborne McGraw-Hill.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal* 7, 155-162.

Press, W. H., Teukolsky, S. A., Vetterling, and Flannery, B. P. (1992). *Numerical recipes in FORTRAN 77*. Cambridge University Press.

Sakamoto, Y. (1991). *Categorical data analysis by AIC*. Tokyo: KTK Scientific Publishers.

Venzon, D. J. and Moolgavkar. (1988). A method for computing profile-likelihood-based confidence limits. Applied Statistics **37**, 87-94.

CHAPTER 3

Evaluating the Ability of Information-Based Model Selection to Detect Mechanisms of Disease Transmission in Binomial Chain Epidemic Data

SUMMARY

In this chapter we use Monte Carlo methods to evaluate the ability of Akaike's information criteria (AIC) and it's Baysian complement, BIC, to detect the McCarty-Miller model (McCarty and Miller, 1998) as the generating process of epidemic binomial chain data of the same sample sizes and similar rates of infection as found in the 2 data sets analyzed in Chapter 2 (common cold, Heasman and Reid, 1961; measles,Wilson et al., 1939). As predicted, BIC out-performed AIC in all scenarios in which the generating process was in the set of candidate models. Our results do not support the notion suggested in Chapter 2 that the selection of a biologically unreasonable model by AIC as the best approximating model for the Heasman and Reid data is an unfortunate artifact of sample size.

Key words: Greenwood; Reed-Frost; binomial chain; epidemic modeling; model selection; statistical inference; AIC; Monte Carlo; measles; common cold.

3.1. Introduction to the problem

3.1.1 *Review*

Models of specific epidemics are numerous in the literature (Dietz and Schenzle, 1985; Barlow, 1995). Several general theories of epidemic dynamics exist and have allowed authors a variety of approaches to specific problems (Bailey, 1975; Anderson and May 1984; Barlow, 1995). However, a recent application of model selection using Akaike's information criteria (AIC; Akaike 1973), corrected for possible small sample bias, AICc (Sugiura 1978; Hurvich and Tsai 1989; Burnham and Anderson, 1998) to 2 classic data sets ("Heasman and Reid", and "Providence"; Table 3.1), has raised doubts about the validity and utility of these theories (Chapter 2). In this chapter we use Monte Carlo methods (Morgan, 1984; Efron and Tibshirani, 1993; Hjorth, 1994) to evaluate the ability of AICc and it's Baysian complement, BIC (Akaike 1978; Schwarz, 1978; Burnham and Anderson, 1998), to detect the McCarty-Miller model, **MM** (McCarty and Miller, 1998) as the generating process of epidemic binomial chain data of the same sample sizes and similar rates of infection as found in these 2 data sets. Specifically, we address three questions: (1) What parameter values of generating process **MM** are conducive to selection of **MM** as the "best" model by either AICc or BIC?, (2) If selection of **MM** fails, what models are favored by AICc and BIC?, (3) Finally, could the selection of a biologically unreasonable model (**logitS**) by AICc as the best approximating model for the Heasman and Reid data set be somehow explained even if **MM** were the generating process?

3.1.2 *Model selection under efficiency and consistency*

Information theory based model selection requires 4 steps (Buckland et al., 1997; Burnham and Anderson, 1998; McQuarrie and Tsai, 1998): (1st) selection of candidate models for the data, (2nd) maximizing the resulting log-likelihoods, (3rd) calculation of an information criteria for each model-data set combination, and (4th) selection of the "best" model based on a comparison of the corresponding criteria values. Two general classes of information criteria have resulted from the Frequentist and Baysian philosophies (Burnham and Anderson, 1998; McQuarrie and Tsai, 1998): efficient criteria (e.g., TIC, AIC, AICc, QAIC, QAICc, TIC, FPE, Cp, PRESS), and consistent criteria (e.g., BIC, SIC, HQ). Diversity within these 2 classes is primarily the result of differing interpretations of what constitutes a "best" model.

A selection criteria that asymptotically (i.e., sample size $\to \infty$) chooses the model from the set of candidates with the minimum mean squared error distribution is called efficient (Shibata, 1980; McQuarrie and Tsai, 1998). This approach assumes that the true generating process is of infinite dimension and therefore is not a candidate model; selecting the "best approximating" model is therefore the goal. The theoretical measure of discrepancy between the approximating model (M_A) and the generating process (M_T) targeted by AIC and AICc is *Kullback-Leibler discrepancy*, K-L (Kullback and Leibler, 1951):

$$\Delta_{\text{K-L}}(M_T, M_A) = E_{F_T}\left[\log\left(\frac{f_T(x)}{f_A(x)}\right)\right].$$

Kulback-Leibler discrepancy is not used directly as a model selection criteria since (1) the true generating process (M_T) is usually not known, (2) additional uncertainty must be accounted for since the data only provide an estimate of $f_A(x)$, and (3) even if $f_A(x)$ is know, computation

of the required expectation may be impractical (Burnham and Anderson, 1998). Akaike's information criteria, corrected for small sample biases (AICc), is given by

$$\text{AICc} = -2\ln\mathcal{L}(\hat{\underline{\theta}}|\text{data}, \text{model}) + 2K + \frac{2K(K+1)}{n-K-1},$$

where n is the sample size and K is the number of model parameters ($\underline{\theta}$) estimated from the data (as maximum likelihood estimates) and n is the sample size. Removal of the last term gives AIC. Akaike's information criteria is a biased estimator of the expected K-L discrepancy.

A selection criteria that asymptotically (i.e., sample size $\to \infty$) chooses the correct model (i.e., the generating process) from the set of candidates with probability one is called consistent (Bozdogan, 1987; McQuarrie and Tsai, 1998). This approach assumes that the generating process is represented in the set of candidate models. BIC is a consistent criteria:

$$\text{BIC} = -2\ln\mathcal{L}(\hat{\underline{\theta}}|\text{data}, \text{model}) + \ln(n)\cdot K,$$

where n is the sample size and K is the number of model parameters ($\underline{\theta}$) estimated from the data (as maximum likelihood estimates) and n is the sample size. BIC is not efficient and it does not estimate expected K-L discrepancy (Burnham and Anderson, 1998; McQuarrie and Tsai, 1998). The derivation of BIC assumes equal priors on all candidate models and weak priors on the estimates of the parameters of each model (Burnham and Anderson, 1998). In many traditional settings (e.g., generalized additive models, time series), BIC consistently out-performs AIC when the generating process is present as a model in the candidate set and sample size is large (MQuarrie and Tsai, 1998).

While information criteria may be used to select the "best" model, current theory suggests that making inferences from only the selected model may be less than optimal (Burnham and Anderson, 1998). Three methods of assessing model selection uncertainty have been recommended. As a general rule, models differing in AICc value by 2-4 units from the best model should not be ruled out as good candidates and models differing by >4-7 units are very weak contenders. Similar recommendations for BIC have not yet appeared. Alternatively, Akaike weights provide a normalized metric (0-1) of model selection uncertainty for model i among the R candidate models:

$$w_i = \frac{\exp\left[-\frac{1}{2}\Delta \mathrm{AIC}_i\right]}{\sum_{r=1}^{R} \exp\left[-\frac{1}{2}\Delta \mathrm{AIC}_r\right]},$$

where $\Delta \mathrm{AIC}_i$ = AIC_i - minimum AIC; models with high weights are good candidates (Burnham and Anderson, 1998). In a Baysian context, Akaike weights using $\Delta \mathrm{BIC}_i$ can be taken to give the posterior probability of each candidate model being the generating process. Finally, bootstrapping provides a robust method of assessing model selection uncertainty and is applicable in virtually all model selection problems regardless of the selection criteria (Burnham and Anderson, 1998).

3.2. Candidate models and the results of Chapter 2

3.2.1 *General classification of models*

Candidate models for the Heasman and Reid and Providence data sets may be conveniently divided into 4 classes based on an interpretation of each individual model as being

"mechanistic" or "statistical" and whether or not individual models are nested within the global model of each class. For example, model class 2 contains 14 models nested under one global model and is considered to be more mechanistic than model class 1, but less so than model class 3.

3.2.2 *Model class 1, multinomial (1 member)*

The multinomial model, **MULT**, recognizes no generating process beyond the observation of different categories of chains, each with different cell probabilities. For example, under this model the likelihood of the Hesiman and Reid data (Table 3.1):

$$\frac{664!}{423!131!36!14!4!2!8!2!2!24!11!3!1!3!0!0!} \times$$

$$p_1^{423} p_2^{131} p_3^{36} p_4^{14} p_5^{4} p_6^{2} p_7^{8} p_8^{2} p_9^{2} p_{10}^{24} p_{11}^{11} p_{12}^{3} p_{13}^{1} p_{14}^{3} p_{15}^{0} \left(1 - \sum_{i=1}^{15} p_i \right)^{0}.$$

Model **MULT** is not selected by AICc for either the Heasman and Reid or Providence data sets; ΔAICc > 10.0 and Akaike weights <0.005 in both cases.

3.2.3 *Model class 2, structured multinomial (15 members)*

Global model **NTSI** recognizes the first level of structure generating each multinomial cell of each data set; that each chain classification is the result of a disease transmission process. Under this model, the probability of transition from the susceptible to the infected state is conditional on the population size (N); the time step (T); the number of available susceptibilities (S); and the number infectious (I). The number recovered (R), is not included

since $R=N-S-I$. For example, the expected number of chains 1-2-0 of the Heasman and Reid data set is

$$664\left(6p^2_{N=5,T=1,S=4,I=1}q^2_{N=5,T=1,S=4,I=1}q^2_{N=5,T=2,S=2,I=2}\right),$$

which represents the expected number of instances of 2 of 4 susceptibilities (S=4) converting ($p^2_{N=5,T=1,S=4,I=1}$) to the infectious state in the presence of 1 infectious (I=1) at the end of the first time step (T=1) and 2 not converting ($q_{N=5,T=1,S=4,I=1} = 1 - p_{N=5,T=1,S=4,I=1}$), followed by 2 susceptibles (S=2) not converting ($q_{N=5,T=2,S=2,I=2}$), in the presence of 2 infectious (I=2) at the end of the second time step (T=2). The coefficient "6" comes from the fact that there are 6 ways to observe this sequence of events among 3 individuals (+ 1 initial infectious). Model class 2 contains 14 nested sub-models (**N**, **T**, **S**, **I**, **NT**, **NS**, **NI**, **TI**, **TS**, **SI**, **NTS**, **NTI**, **NSI**, **TSI**). Because N = 3 or 4 is small, models **NTS**, **NTI**, **NSI**, and **TSI** are all equivalent to global model **NTSI** when applied to the Providence data set. Models in class 2 are not selected by AICc for either the Heasman and Reid or Providence data sets; ΔAICc > 3.0 and Akaike weights $\leq$0.06 in all cases.

3.2.4 *Model class 3, nested logistic (15 members)*

Nested logistic regression models, global model **logitNTSI**, consider transition from the susceptible to the infected state as a logit transform on some linear combination of the current population state,

$$p_{N,T,S,I} = \frac{1}{1+\exp[-(\beta_0+\beta_1 N+\beta_2 T+\beta_3 S+\beta_4 I)]}.$$

These models offer nearly the same structure as models in class 2, but force additivity resulting in greater parsimony. Because $N = 3$ or 4 is small, **logitNTS**, **logitNTI**, **logitNSI**, **logitTSI** are all equivalent to global model **logitNTSI** when applied to the Providence data set. Included in this class is our interpretation of the "mass action assumption" (Chapters 1,2), which we represent as a logit transform of I/N, model **logitI/N**:

$$p_{N,T,S,I} = \frac{1}{1 + \exp\left[-\left(\beta_0 + \beta_1 \frac{I}{N}\right)\right]}.$$

All model in this class, except **logitI** and **logitI/N**, are good candidates for the Heasman and Reid data set:

logitS, ΔAICc = 0, weight = 0.312;

logitT, ΔAICc = 1.26, weight = 0.166;

logitTS, ΔAICc = 2.00, weight = 0.114;

logitSI, ΔAICc = 2.02, weight = 0.114;

logitTI, ΔAICc = 2.07, weight = 0.111;

logitTSI, ΔAICc = 4.01, weight = 0.042;

logitI, ΔAICc = 13.18, weight < 0.001.

Curiously, models in this class are very poor candidates for the Providence data; ΔAICc > 71.0 and Akaike weights <0.001 in all cases.

3.2.3 *Model class 4, biological mechanistic (10 members)*

For heuristic purposes, all members of this class can be considered as modifications of the McCarty-Miller model, **MM** (Chapter 2; McCarty and Miller, 1998), which recognizes two

possible biological mechanisms of disease transmission: a random apportionment of Ib (# infectious · average number of infectious contacts per infectious individual) total infectious contacts among the N individuals of the entire population at each time step, and a possible constant level of background risk ($1-a$):

$$p_{N,T,S,I} = 1 - a\left(1 - \frac{1}{N}\right)^{IB}.$$

Model **MM** simplifies to the Reed-Frost model, **RF** (Bailey 1975), if N is held constant and a=1 (i.e., no background risk) and the Greenwood model, **G** (Bailey, 1975; Greenwood, 1931), if B=0 and N is held constant. Model **MM** also simplifies to the James-Rossiter model (1989), **JR**, if a=1 and N=S; corresponding to a random apportionment of the IB infectious contacts across only the susceptible portion of the population. Model **MM** can be modified to account for variation of the degree of risk to infection from a point source specific to a population by replacing a with a beta random variate. For example, the expected number of 1-1-1-2 chains in the Heasman and Reid data set under model **MM** with B=0 and $(1-a) \sim \text{beta}(\alpha,\beta)$ is given by:

$$\mathrm{E}[664(12(1-a)^4a^5] = 7968\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 (1-a)^4a^5(1-a)^{\alpha-1}a^{\beta-1}da,$$

which may be written as,

$$7968\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+4)\Gamma(\beta+5)}{\Gamma(\alpha+\beta+9)}\int_0^1 \frac{\Gamma(\alpha+\beta+9)}{\Gamma(\alpha+4)\Gamma(\beta+5)}(1-a)^{(\alpha+4)-1}a^{(\beta+5)-1}da,$$

which since,

$$\frac{\Gamma(\alpha+\beta+9)}{\Gamma(\alpha+4)\Gamma(\beta+5)}\int_0^1 (1-a)^{(\alpha+4)-1}a^{(\beta+5)-1} = 1,$$

simplifies to,

$$7968\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+4)\Gamma(\beta+4)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+9)},$$

and since $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, finally giving,

$$7968\frac{(\alpha+3)(\alpha+2)(\alpha+1)\alpha(\beta+4)(\beta+3)\ldots\beta}{(\alpha+\beta+8)(\alpha+\beta+7)\ldots(\alpha+\beta)}.$$

This modification is equivalent to the beta-Greenwood model, **BG**, previously investigated by Bailey (1975) using the Providence data set; allowing separate beta random variates for each population size gives model **BGN**. Model **MM** can be further modified to give the Kermack-McKendrick model, **BRF** (Becker, 1981), if a=1, and the term $(1 - 1/N)^B$ is replaced by a beta random variate; allowing separate beta random variates for each population size gives model **BRFN**.

Finally, included in this class is the propagating point source model, **PP**, which considers each new infectious individual as a new point source of risk to infection:

$$p_{N,T,S,I} = 1-(1-p)^I.$$

While mechanistically distinct, model **PP** is functionally equivalent to model **MM** when a=B=1 and the term $1/N$ is replaced by the parameter p; only with the Providence data set is this model distinct.

Models in class 4 are very poor candidates for the Heisman and Reid data set; ΔAICc > 4.50 and Akaike weights <0.04 in all cases. In sharp contrast, models **BG** and **BGN** are the only plausible candidates for the Providence data set: **BG,** ΔAICc = 0, weight = 0.824; **BGN** , ΔAICc = 3.22, weight = 0.165 .

3.4.2. Likelihoods

We include only one log-likelihood (less multinomial coefficient, and not simplified to assist interpretation) as an example, the Providence data under model **NTSI**, where $(\ldots)_{N,T,S,I}$ indexes every term within the parentheses:

$$34\cdot\ln[(q^2)_{3,1,2,1}] + 25\cdot\ln[2(pq)_{3,1,2,1}(q)_{3,2,1,1}] + 36\cdot\ln[2(pq)_{3,1,2,1}(p)_{3,2,1,1}] + 239\cdot\ln[(p^2)_{3,1,2,1}] + 4\cdot\ln[(q^3)_{4,1,3,1}] + 3\cdot\ln[3(pq^2)_{4,1,3,1}(q^2)_{4,2,2,1}] + 1\cdot\ln[6(pq^2)_{4,1,3,1}(pq)_{4,2,2,1}(q)_{4,3,1,1}] + 4\cdot\ln[6(pq^2)_{4,1,3,1}(pq)_{4,2,2,1}(p)_{4,3,1,1}] + 3\cdot\ln[3(pq^2)_{4,1,3,1}(p^2)_{4,2,2,1}] + 8\cdot\ln[3(p^2q)_{4,1,3,1}(q^2)_{4,2,1,2}] + 10\cdot\ln[3(p^2q)_{4,1,3,1}(p)_{4,2,1,2}] + 67\cdot\ln[(p^3)_{4,1,3,1}]$$ (see *appendix* of *Chapter 2* for additional clarification).

Standard approaches (Becker, 1981) give simple closed-form solutions for maximum likelihood estimates of the parameters for all models in classes 1 and 2 by equating the first derivative of each log-likelihood to 0 and solving for each parameter. For example, the maximum likelihood estimator of parameter $p_{N=4,T=1}$ of model **NI** is:

$$\hat{p}_{N=4,T=2} = \frac{y_{N=4,I=1}}{x_{N=4,I=1}},$$

where $y_{N=4,I=1}$ is the total number of individuals converting from the susceptible to the infected state in households of size 4 at the second time step of each household's epidemic, and $x_{N=4,I=1}$ is the total number of such trials. We used numerical optimization (Powell, 1964) in cases where no simple estimator exists.

3.4. Monte Carlo study: methods, results and discussion

3.4.1 *Monte Carlo methods*

We investigated 18 scenarios of data generated under model **MM** (Figures 3.1-3.18); 6 mimicking the Heasman and Reid data, and 12 mimicking the Providence data. We chose combinations of the parameters of model **MM** that produce similar rates of infection as found in these 2 data sets and generated 10,000 fictitious data sets for each scenario. We used a simple linear congruential generator (Knuth, 1981), modeling the probability of conversion of each susceptible individual to the infected state at each step of the chain as an independent Bernoulli trial, to generate each chain of each household; 664 instances for scenarios simulating the Heasman and Reid data, and 434 instances for scenarios simulating the Providence data (*appendix*). We confirmed that 10,000 replicates were sufficient to give estimates of model selection frequency reliable to 2 decimal places by comparing the results of 10,000 replicates of scenario 3 (Figure 3.3) with 50,000 replicates of scenario 3 (e.g., Burnham and Anderson, 1998).

Our candidate models were chosen based on the results of Chapter 2. We chose 21 models to challenge data from the 6 Heasman and Reid scenarios (Figures 3.1-3.6): **MULT**, **T**, **S**, **I**, **TI**, **SI**, **TS**, **TSI**, **logitT**, **logitS**, **logitI**, **logitTI**, **logitSI**, **logitTS**, **logitTSI**, **G**, **BG**, **RF**, **BRF**, **MM**, and **JR**. This is the same set used in Chapter 2 for the AICc based analysis of the Heasman and Reid data. We chose 11 models to challenge data from the 12 Providence scenarios: **N**, **S**, **I**, **NI**, **logitS**, **G**, **BG**, **RF**, **RFN**, **BRF**, and **MM**. This set is a subset of the set used in Chapter 2 for the AICc based analysis of the Providence data. We chose a reduced set for 4 reasons. First, the reduced set shortens computation time and minimizes problems associated with the numerical optimization routine (Powell, 1964) used

to maximize many of the log-likelihoods. Second, because the Providence data set is small, it is unlikely that models having >3 parameters will be selected by AICc or BIC; the results of Chapter 2 support this belief. Third, many of the models in the original set of 31 models used in Chapter 2 incorporate similar effects. Finally, not all 31 models were necessary to address our stated goals.

We used the same FORTRAN source code used in Chapter 2 (see *appendix*) to calculate the maximized log-likelihoods for every simulated data set-model combination. FORTRAN code was compiled using f2c (Feldman and Weinberger, 1990; http://netlib.bell-labs.com/f2c, 1999) and run on a Cyrix 686 166+ under the Linux operating system (Slackware version 1.2.13; Barkakati, 1996). Computations times were practical; 46 hours for scenarios 1-6 (Figures 3.1-3.6). We verified the reliability of the f2c translator and the Cyrix 686 processor by comparing the results of scenarios computed using this system with results obtained by compiling ("xlf"; http://lamar.colostate.edu) and running the same scenarios on an IBM RS6000 under the UNIX operating system (AIX version 4; IBM copyright 1982, 1996; http://lamar.colostate.edu).

3.4.2 *Detecting the generating process*

Two observations deserve notice: (1) The generating process was more likely to be correctly identified in the simulated Heasman and Reid data than in the simulated Providence data, and (2) BIC out-performed AICc in all but one scenario. AICc correctly identified the generating process >55% of the time in only 4 of the 6 Heasman and Reid scenarios (Figures 3.1, 3.2, 3.5, 3.6) and only once (Figure 3.16) in the 12 Providence scenarios. Performing slightly better, BIC correctly identified the generating process >55% of the time in 5 of the 6

Heasman and Reid scenarios (Figures 3.1, 3.2, 3.4, 3.5, 3.6) but again only once (Figure 3.16) in the 12 Providence scenarios. Apparently the combinations of values for the parameters of model **MM** necessary to produce data similar to the Providence data are not sufficient to provoke consistent selection of **MM** over more parsimonious models such as **RF** and **G** or more flexible models such as **RFN**.

3.4.3 **RFN** *vs.* **MM**

Model **MM** and model **RFN** each have 2 parameters when applied to the Providence data set. However model **RFN** will always give an equal or better fit than model **MM** because it does not constrain the N effect as **MM** does. Consequently, it is not surprising that model **RFN** was selected more frequently than **MM** in 7 of the 12 Providence scenarios (Figures 3.7-3.12, 3.15) by both AICc and BIC. Because the number of parameters in model **MM** is independent of the variation in N, both AICc and BIC would be more likely to select **MM** over **RFN** given greater variation in N than the 2 values, $N = 3$ or 4, present in the Providence data set.

3.4.4 *Accidental evidence for heterogeneity*

The empirical probability of selecting a model incorporating individual heterogeneity, models **BG** and **BRF**, by either AICc or BIC is very low in all scenarios considered. The probability of selection exceeds standard alpha levels in only 1 scenario; model **BRF** is selected by AICc with probability 0.14 and BIC with probability 0.07 in scenario 6 (Figure 3.6); the empirical probability of the selection of either **BG** or **BRF** is <0.05 in all other cases. Consequently, the selection of model **BG** by AICc with a high Akaike weight of 0.824 as the best candidate

for the Providence data set is probably not the result of this data deviating even strongly from the expected sample if the true generating process was **MM** or another similar process.

Not surprisingly, model selection analyses of data generated under model **RF** (Figures 3.4-3.6) are more likely to result in selection of **BRF** than **BG**. Similarly, model selection analyses of data generated under model **BG** (Figures 3.4-3.6) are more likely to result in selection of **BG** than **BRF**. Further, it is not surprising that scenario 6 (Figure 3.6) resulted in the highest selection probability for a model incorporating heterogeneity, since it was generated under the most extreme example of model **RF** investigated (i.e., model **MM** with α=1.0, B=0.75 versus, for example, scenario 5, **MM** with α=1.0, B=0.5).

3.4.5 *Is* **logitS** *unreasonable under* **MM***?*

The empirical probability of selecting model **logitS** by either AICc or BIC is low in all scenarios considered; $p<0.05$ for all Heasman and Read scenarios (Figures 3.1-3.6); $p<0.2$ for all Providence scenarios (Figures 3.7-3.18). Consequently, the selection of model **logitS** by AICc (weight 0.312) as the best candidate for the Heasman and Read data set is probably not the result of this data deviating even strongly from the expected sample if the true generating process was **MM** or another similar process.

3.5. Conclusions

It is unrealistic to expect that the generating process is present as a model in any set of candidate models (Burnham and Anderson, 1998). However, it is not unrealistic to suppose that a key central generating mechanism exists and that individual observations may obscure this mechanism from detection with "noise"caused by individual heterogeneity or errors in data collection. This is especially arguable in the case of infectious epidemics in which it is

known that the disease always physically transmits from host to host but that individual hosts and environments vary greatly; versatility is probably the key to successful pathogen evolution. Studies such as these are useful to reveal what can be detected in the absence of such noise and can lead to many useful specific and general observations.

First, studies collecting binomial chain data should record sufficient covariate information to model individual heterogeneity; minimally, gender, age, and degree of crowding. Unfortunately, this information was either not collected, collected haphazardly, or not adequately reported (e.g. "crowding" information presented by Heasman and Reid) to model these effects in either the Providence or Heasman and Reid data sets.

Second, for data of these sample sizes and rates of infection, model **MM** will not be selected by either AICc or BIC over model **G** unless threshold values for the B parameter of this model are exceeded. Roughly, if model **MM** were the underlying generating process of the Heisman and Reid data and a candidate model, it would not be selected by AICc or BIC unless $B \geq 0.25$ (i.e., 0.25 infectious contacts per infectious individual) and the effects of individual heterogeneity were somehow controlled. A threshold value is more difficult to determine for the Providence data from our study because of the similarity of model **MM** to model **RFN**.

Third, studies questioning the existence of an N effect as modeled by **MM** using binomial chain data should collect data from populations of more than 2 different sizes. Model **RFN** will always give an equal or better fit than model **MM** because it does not constrain the N effect as **MM** does. However, the number of parameters in model **MM** is independent of the number of populations sizes; unlike model **RFN**, in which the number of parameters increases by one unit for every populations size. Consequently, both AICc and

BIC will penalize model **RFN** but not **MM** as the number of population sizes increases.

Finally, in the absence of heterogeneity, selection of model **logitS** by either AICc or BIC is highly unlikely if model **MM** is the generating process for data of the same sample sizes and similar rates of infection as found in the Heasman and Reid and Providence data sets. Consequently, the selection of model **logitS** by AICc as the best approximating model for the Heasman and Reid data remains biologically unexplained.

REFERENCES (*chapter 3*)

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds), 267-281. Budapest: Akademia Kiado.

Akaike, H. (1978). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics* **30**, Part A, 9-14.

Barkakati, N. (1996). *Linux Secrets*. Forester City, California: IDG Books Worldwide, Inc.

Barlow, N. D. (1995). Critical evaluation of wildlife disease models. In *Ecology of infectious disease in natural populations*, B. T. Grenfell and A. P. Dobson (eds), 230-259. Cambridge: Cambridge University Press.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* **52**, 345-370.

Buckland, S. T., Burnham, K. P., and Augustin N. H. (1977). Model selection: an integral part of inference. *Biometrics* **53**, 603-618.

Burnham, K. P. and D. A. Anderson. (1998). *Model selection and inference: A practical information theoretic approach*. New York: Springer-Verlag.

Dietz, K. and Schenzle, D. (1985). Mathematical models for infectious disease statistics. In *A celebration of statistics*, A. C. Atkinson and S. E. Fienberg (eds), 167-204. New York: Springer Verlag.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman Hall.

Feldman, S. A. and Weinberger, P. J. (1990). A portable fortran 77 compiler. In *UNIX Time Sharing System Programer's Manual*, 10th edition, volume 2, AT&T Bell Laboratories. New York: Holt, Rinehart and Winston.

Greenwood, M. 1931. On the statistical measure of infectiousness. *Journal of Hygiene Cambridge* **31**, 336-351.

Heasman, M. A. and Reid, D. D. (1961). Theory and observation in family epidemics of the common cold. *British Journal of Preventive Social Medicine* **15**, 12-16.

Heesterbeek, J. A. P. and Roberts, M. G. (1995). Mathematical models for microparasites of wildlife. In *Ecology of Infectious Disease in Natural Populations*, B. T. Grenfell and A. P. Dobson (eds), 90-122. Cambridge: Cambridge University Press.

Hjorth, J. S. U. (1994). Computer intensive statistical methods: validation, model selection and bootstrap. London: Chapman Hall.

Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Knuth, D. E. (1981). *Seminumerical Algorithms*, 2nd edition, volume 2 of *The Art of Computer Programming*. Reading, Massachusetts: Addison-Wesley.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79-86.

McCarty, C. W. and Miller, M. W. (1998). A versatile model of disease transmission applied to forecasting bovine tuberculosis dynamics in white-tailed deer. *Journal of Wildlife Diseases* **34**, 722-730.

McQuarrie A. D. R. and C. L. Tsai. *Regression and Time Series Model Selection.* World Scientific, Singapore.

Morgan, B. J. T. (1984). *Elements of Simulation*. London: Chapman Hall.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal* 7, 155-162.

Schwarz, G. (7978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.

Shibata, R. (1980). Asymptotic efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* **8**, 147-164.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods* 7, 275-277.

Wilson, E. B., Bennett, C. Allen, M., and Worchester, J. (1939). Measles and scarlet fever in Providence, R. I., 1929-34 with respect to age and size of family. *Procedings of the American Philosophical Society* **80**, 357-476.

Table 3.1

Data sets used in a challenge of theories of epidemic dynamics. Both are frequency tabulations of epidemic "binomial chains" in which the progress of successive viral infections was tracked within human family households. For example, one class of chains from households of 5 members with 1 initial infection is initial infection, followed by 2 new infections and recovery of the first, 1 infection and recovery of the second and third, and no new infections and recovery of the fourth; the notation for this chain is 1210.

Heasman and Reid (common cold data; Heasman and Reid, 1961)		Providence (measels data; Bailey 1975)	
chain	frequency	chain	frequency
10	423	size 3:	
110	131	10	34
1110	36	110	25
11110	14	111	36
11111	4	12	239
1112	2		
1120	8	size 4:	
1121	2	10	4
113	2	110	3
120	24	1110	1
1210	11	1111	4
1211	3	112	3
122	1	120	8
130	3	121	10
131	0	13	67
14	0		
		total:	434
total:	664		

Expected distribution under model MM (actual dist. of data)	
chain	frequency
10	272.0 (423)
110	139.3 (131)
1110	66.8 (36)
11110	26.7 (14)
11111	6.7 (4)
1112	4.2 (2)
1120	20.9 (8)
1121	5.2 (2)
113	2.2 (2)
120	65.3 (24)
1210	26.1 (11)
1211	6.5 (3)
122	4.1 (1)
130	13.6 (3)
131	5.1 (0)
14	1.1 (0)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
MULT	15	18 (0.44)	0
T	4	131 (0.36)	0
S	4	988 (0.47)	0
I	2	1156 (0.46)	19 (0.70)
TI	7	19 (0.29)	0
SI	7	57 (0.25)	0
TS	7	0	0
TSI	8	56 (0.43)	0
logitT	2	0	0
logitS	2	0	0
logitI	2	149 (0.27)	95 (0.57)
logitTI	3	57 (0.25)	0
logitSI	3	19 (0.20)	0
logitTS	3	75 (0.30)	18 (0.43)
logitTSI	4	0	0
G	1	7239 (0.42)	9812 (0.91)
BG	2	36 (0.53)	0
RF	1	0	56 (0.69)
BRF	2	0	0
MM	2	0	0
JR	1	0	0

Figure 3.1. (*left*) Expected distribution of 664 binomial chains under model G (model MM with a=0.8, B=0.0) and actual distribution of the Heisman and Read data (Heisman and Read, 1964). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model G (model MM with a=0.8, B=0.0) as the generating process for 664 binomial chains. Note that these conditions produce a more infectious epidemic than observed in the Heisman and Read data.

Expected distribution under model **MM** (actual dist. of data)	
chain	frequency
10	435.7 (423)
110	141.2 (131)
1110	38.1 (36)
11110	7.6 (14)
11111	0.8 (4)
1112	0.5 (2)
1120	4.7 (8)
1121	0.5 (2)
113	0.2 (2)
120	26.1 (24)
1210	5.2 (11)
1211	0.6 (3)
122	0.3 (1)
130	2.2 (3)
131	0.4 (0)
14	0.1 (0)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
MULT	15	0	0
T	4	479 (0.33)	0
S	4	574 (0.29)	0
I	2	574 (0.34)	0
TI	7	0	0
SI	7	95 (0.33)	0
TS	7	0	0
TSI	8	0	0
logitT	2	0	0
logitS	2	431 (0.23)	47 (0.72)
logitI	2	720 (0.23)	48 (0.41)
logitTI	3	48 (0.18)	0
logitSI	3	288 (0.26)	0
logitTS	3	47 (0.40)	0
logitTSI	4	144 (0.37)	0
G	1	5500 (0.29)	8423 (0.74)
BG	2	144 (0.21)	0
RF	1	956 (0.32)	1482 (0.62)
BRF	2	0	0
MM	2	0	0
JR	1	0	0

Figure 3.2. (*left*) Expected distribution of 664 binomial chains under model **G** (model **MM** with a=0.9, B=0.0) and actual distribution of Heisman and Read data (Heisman and Read, 1964). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **G** (model **MM** with a=0.9, B=0.0) as the generating process for 664 binomial chains. Note that these conditions produce a slightly less infectious epidemic than observed in the Heisman and Read data.

Expected distribution under model MM (actual dist. of data)		
chain	frequency	
10	348.5	(423)
110	150.3	(131)
1110	57.1	(36)
11110	17.0	(14)
11111	3.0	(4)
1112	1.7	(2)
1120	11.1	(8)
1121	2.9	(2)
113	0.8	(2)
120	41.4	(24)
1210	17.1	(11)
1211	3.0	(3)
122	2.4	(1)
130	5.7	(3)
131	2.7	(0)
14	0.3	(0)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
MULT	15	21 (0.41)	0
T	4	266 (0.34)	0
S	4	956 (0.43)	0
I	2	1492 (0.39)	0
TI	7	83 (0.30)	0
SI	7	40 (0.25)	0
TS	7	48 (0.26)	0
TSI	8	40 (0.29)	0
logitT	2	0	0
logitS	2	117 (0.32)	23 (0.43)
logitI	2	4 (0.18)	0
logitTI	3	81 (0.27)	7 (0.30)
logitSI	3	121 (0.27)	0
logitTS	3	7 (0.33)	0
logitTSI	4	24 (0.41)	0
G	1	4073 (0.35)	6262 (0.82)
BG	2	156 (0.44)	41 (0.53)
RF	1	2327 (0.42)	3629 (0.80)
BRF	2	144 (0.65)	38 (0.65)
MM	2	0	0
JR	1	0	0

Figure 3.3. (*left*) Expected distribution of 664 binomial chains under model **MM** (a=0.9, B=0.25) and actual distribution of Heisman and Read data (Heisman and Read, 1964). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.9, B=0.25) as the generating process for 664 binomial chains. Note that these conditions produce a more infectious epidemic than observed in the Heisman and Read data.

Expected distribution under model **MM** (actual dist. of data)	
chain	frequency
10	531.2 (423)
110	103.1 (131)
1110	15.9 (36)
11110	1.7 (14)
11111	0.1 (4)
1112	0.1 (2)
1120	0.9 (8)
1121	0.1 (2)
113	0.0 (2)
120	8.4 (24)
1210	1.9 (11)
1211	0.1 (3)
122	0.1 (1)
130	0.3 (3)
131	0.1 (0)
14	0.0 (0)

		# selected | 10,000 simulations (mean weight | selection)			
model	K	AICc		BIC	
MULT	15	0		0	
T	4	199	(0.31)	0	
S	4	386	(0.36)	0	
I	2	254	(0.34)	0	
TI	7	30	(0.19)	0	
SI	7	5	(0.19)	0	
TS	7	0		0	
TSI	8	0		0	
logitT	2	175	(0.19)	30	(0.33)
logitS	2	331	(0.24)	77	(0.42)
logitI	2	118	(0.27)	19	(0.55)
logitTI	3	25	(0.26)	0	
logitSI	3	481	(0.24)	2	(0.33)
logitTS	3	0		0	
logitTSI	4	36	(0.23)	0	
G	1	3046	(0.25)	3761	(0.56)
BG	2	105	(0.34)	31	(0.47)
RF	1	4323	(0.31)	5881	(0.64)
BRF	2	471	(0.39)	150	(0.51)
MM	2	0		0	
JR	1	15	(0.50)	49	(0.95)

Figure 3.4. (*left*) Expected distribution of 664 binomial chains under model **RF** (model **MM** with a=1.0, B=0.25) and actual distribution of the Heisman and Read data (Heisman and Read, 1964). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **RF** (model **MM** with a=0.1, B=0.25) as the generating process for 664 binomial chains. Note that these conditions produce a less infectious epidemic than observed in the Heisman and Read data.

Expected distribution under model **MM** (actual dist. of data)	
chain	frequency
10	425.0 (423)
110	143.6 (131)
1110	40.7 (36)
11110	8.6 (14)
11111	1.0 (4)
1112	0.6 (2)
1120	4.8 (8)
1121	1.2 (2)
113	0.2 (2)
120	22.7 (24)
1210	10.2 (11)
1211	1.2 (3)
122	1.4 (1)
130	2.0 (3)
131	1.2 (0)
14	0.1 (0)

		# selected | 10,000 simulations (mean weight | selection)			
model	K	AICc		BIC	
MULT	15	26	(0.40)	0	
T	4	167	(0.32)	0	
S	4	301	(0.38)	0	
I	2	561	(0.42)	0	
TI	7	65	(0.29)	0	
SI	7	75	(0.24)	0	
TS	7	102	(0.29)	0	
TSI	8	0		0	
logitT	2	0		0	
logitS	2	226	(0.27)	22	(0.42)
logitI	2	16	(0.24)	11	(0.51)
logitTI	3	148	(0.29)	0	
logitSI	3	251	(0.27)	0	
logitTS	3	17	(0.43)	0	
logitTSI	4	27	(0.32)	0	
G	1	658	(0.29)	943	(0.64)
BG	2	36	(0.25)	0	
RF	1	6631	(0.40)	8959	(0.83)
BRF	2	693	(0.44)	65	(0.66)
MM	2	0		0	
JR	1	0		0	

Figure 3.5. (*left*) Expected distribution of 664 binomial chains under model **RF** (model **MM** with a=1.0, B=0.5) and actual distribution of the Heisman and Read data (Heisman and Read, 1964). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **RF** (model **MM** with a=0.1, B=0.5) as the generating process for 664 binomial chains. Note that these conditions produce an epidemic approximately equivalent to the Heisman and Read data.

Expected distribution under model **MM** (actual dist. of data)	
chain	frequency
10	340.0 (423)
110	149.9 (131)
1110	58.6 (36)
11110	18.1 (14)
11111	3.3 (4)
1112	1.9 (2)
1120	10.7 (8)
1121	4.2 (2)
113	0.9 (2)
120	34.7 (24)
1210	23.3 (11)
1211	4.2 (3)
122	5.5 (1)
130	5.0 (3)
131	4.9 (0)
14	0.4 (0)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
MULT	15	22 (0.63)	0
T	4	25 (0.26)	0
S	4	199 (0.45)	6 (0.49)
I	2	928 (0.49)	0
TI	7	122 (0.38)	0
SI	7	53 (0.33)	0
TS	7	57 (0.30)	0
TSI	8	42 (0.33)	0
logitT	2	0	0
logitS	2	26 (0.35)	10 (0.59)
logitI	2	0	0
logitTI	3	62 (0.25)	0
logitSI	3	43 (0.28)	0
logitTS	3	15 (0.49)	0
logitTSI	4	0	0
G	1	38 (0.30)	139 (0.64)
BG	2	36 (0.36)	0
RF	1	6920 (0.57)	9587 (0.92)
BRF	2	1412 (0.62)	258 (0.68)
MM	2	0	0
JR	1	0	0

Figure 3.6. (*left*) Expected distribution of 664 binomial chains under model **RF** (model **MM** with a=1.0, B=0.75) and actual distribution of the Heisman and Read data (Heisman and Read. 1964). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **RF** (model **MM** with a=0.1, B=0.75) as the generating process for 664 binomial chains. Note that these conditions produce a more infectious epidemic than observed in the Heisman and Read data.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	13.0	(34)
110	20.9	(25)
111	85.0	(36)
12	215.1	(239)
size 4:		
10	3.2	(4)
110	2.1	(3)
1110	2.8	(1)
120	4.4	(8)
1111	6.1	(4)
112	9.6	(3)
121	39.9	(10)
13	31.9	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	129 (0.40)	215 (0.51)
S	3	222 (0.49)	21 (0.46)
I	3	0	0
NI	3	1131 (0.52)	94 (0.68)
logitS	2	543 (0.44)	597 (0.58)
G	1	0	51 (0.37)
BG	2	0	0
RF	1	16 (0.31)	270 (0.52)
RFN	2	6071 (0.43)	6831 (0.53)
BRF	2	6 (0.63)	5 (0.72)
MM	2	1882 (0.39)	1916 (0.46)

Figure 3.7. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=1.0, B=4.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=1.0, B=4.0) as the generating process for 434 binomial chains.

Expected distribution under model MM (dist. of actual data)		
chain	frequency	
size 3:		
10	5.8	(34)
110	10.1	(25)
111	66.3	(36)
12	251.8	(239)
size 4:		
10	1.3	(4)
110	0.7	(3)
1110	1.1	(1)
120	2.3	(8)
1111	3.6	(4)
112	7.5	(3)
121	39.1	(10)
13	44.4	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	176 (0.43)	258 (0.55)
S	3	387 (0.48)	37 (0.64)
I	3	1 (0.25)	0
NI	3	1128 (0.55)	62 (0.42)
logitS	2	942 (0.44)	1108 (0.56)
G	1	1 (0.21)	110 (0.41)
BG	2	0	0
RF	1	27 (0.29)	259 (0.51)
RFN	2	5736 (0.42)	6579 (0.50)
BRF	2	13 (0.51)	13 (0.47)
MM	2	1589 (0.37)	1574 (0.43)

Figure 3.8. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=1.0, B=5.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=1.0, B=5.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	2.6	(34)
110	4.7	(25)
111	48.8	(36)
12	277.9	(239)
size 4:		
10	0.6	(4)
110	0.2	(3)
1110	0.4	(1)
120	1.1	(8)
1111	1.9	(4)
112	5.8	(3)
121	34.9	(10)
13	55.6	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	213 (0.37)	284 (0.50)
S	3	462 (0.43)	20 (0.57)
I	3	0	0
NI	3	1552 (0.41)	0
logitS	2	1656 (0.40)	1806 (0.51)
G	1	10 (0.21)	156 (0.44)
BG	2	0	0
RF	1	40 (0.32)	377 (0.48)
RFN	2	4807 (0.39)	6138 (0.48)
BRF	2	5 (0.70)	5 (0.56)
MM	2	1255 (0.32)	1214 (0.39)

Figure 3.9. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=1.0, B=6.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=1.0, B=6.0) as the generating process for 434 binomial chains.

Expected distribution under model MM (dist. of actual data)		
chain	frequency	
size 3:		
10	18.8	(34)
110	28.6	(25)
111	92.2	(36)
12	194.4	(239)
size 4:		
10	3.8	(4)
110	2.6	(3)
1110	3.4	(1)
120	6.3	(8)
1111	6.7	(4)
112	9.9	(3)
121	38.1	(10)
13	29.1	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	288 (0.40)	361 (0.48)
S	3	352 (0.45)	32 (0.43)
I	3	23 (0.20)	0
NI	3	516 (0.45)	28 (0.57)
logitS	2	722 (0.42)	696 (0.54)
G	1	33 (0.23)	418 (0.46)
BG	2	15 (0.45)	8 (0.44)
RF	1	49 (0.33)	796 (0.53)
RFN	2	4609 (0.42)	4414 (0.50)
BRF	2	22 (0.42)	7 (0.65)
MM	2	3371 (0.38)	3240 (0.46)

Figure 3.10. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.8, B=3.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.8, B=3.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	8.3	(34)
110	14.1	(25)
111	74.8	(36)
12	236.8	(239)
size 4:		
10	1.6	(4)
110	0.9	(3)
1110	1.4	(1)
120	3.4	(8)
1111	4.1	(4)
112	8.0	(3)
121	39.0	(10)
13	41.7	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	201 (0.39)	274 (0.53)
S	3	389 (0.44)	24 (0.41)
I	3	2 (0.26)	0
NI	3	713 (0.51)	63 (0.51)
logitS	2	1113 (0.41)	1141 (0.54)
G	1	32 (0.22)	360 (0.43)
BG	2	8 (0.30)	8 (0.33)
RF	1	70 (0.31)	709 (0.55)
RFN	2	4713 (0.41)	4751 (0.49)
BRF	2	22 (0.32)	8 (0.36)
MM	2	2719 (0.36)	2662 (0.43)

Figure 3.11. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.8, B=4.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.8, B=4.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	3.7	(34)
110	6.6	(25)
111	56.3	(36)
12	267.3	(239)
size 4:		
10	0.7	(4)
110	0.3	(3)
1110	0.5	(1)
120	1.7	(8)
1111	2.2	(4)
112	5.7	(3)
121	35.7	(10)
13	53.2	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	387 (0.36)	418 (0.49)
S	3	495 (0.45)	31 (0.53)
I	3	0	0
NI	3	1267 (0.42)	0
logitS	2	1564 (0.40)	1639 (0.51)
G	1	37 (0.19)	339 (0.45)
BG	2	8 (0.26)	0
RF	1	89 (0.27)	767 (0.51)
RFN	2	4083 (0.38)	4790 (0.46)
BRF	2	68 (0.36)	29 (0.44)
MM	2	2002 (0.34)	1987 (0.40)

Figure 3.12. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.8, B=5.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.8, B=5.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	16.49	(34)
110	25.66	(25)
111	89.80	(36)
12	202.0	(239)
size 4:		
10	2.225	(4)
110	1.349	(3)
1110	1.939	(1)
120	6.896	(8)
1111	4.956	(4)
112	8.811	(3)
121	36.69	(10)
13	37.13	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	931 (0.39)	612 (0.54)
S	3	505 (0.46)	32 (0.61)
I	3	214 (0.22)	0
NI	3	116 (0.38)	1 (0.48)
logitS	2	1225 (0.34)	739 (0.48)
G	1	813 (0.25)	3522 (0.54)
BG	2	193 (0.41)	49 (0.55)
RF	1	457 (0.32)	1760 (0.54)
RFN	2	1376 (0.37)	739 (0.43)
BRF	2	227 (0.39)	81 (0.43)
MM	2	3943 (0.34)	2465 (0.42)

Figure 3.13. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.5, B=2.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.5, B=2.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	7.331	(34)
110	12.49	(25)
111	71.81	(36)
12	242.4	(239)
size 4:		
10	0.9386	(4)
110	0.4686	(3)
1110	0.7396	(1)
120	3.506	(8)
1111	2.767	(4)
112	6.558	(3)
121	35.89	(10)
13	49.13	(67)

		# selected | 10,000 simulations (mean weight | selection)			
model	K	AICc		BIC	
N	2	608	(0.37)	501	(0.48)
S	3	556	(0.47)	46	(0.69)
I	3	75	(0.21)	0	
NI	3	277	(0.45)	0	
logitS	2	1914	(0.36)	1363	(0.49)
G	1	350	(0.25)	2032	(0.52)
BG	2	63	(0.38)	32	(0.42)
RF	1	380	(0.27)	1932	(0.54)
RFN	2	2451	(0.38)	1596	(0.45)
BRF	2	64	(0.49)	16	(0.74)
MM	2	3247	(0.34)	2482	(0.40)

Figure 3.14. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.5, B=3.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.5, B=3.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	3.258	(34)
110	5.873	(25)
111	53.59	(36)
12	271.3	(239)
size 4:		
10	0.3960	(4)
110	0.1582	(3)
1110	0.2663	(1)
120	1.683	(4)
1111	1.417	(3)
112	4.479	(8)
121	31.95	(10)
13	59.65	(67)

		# selected | 10,000 simulations (mean weight | selection)			
model	K	AICc		BIC	
N	2	727	(0.37)	609	(0.50)
S	3	508	(0.42)	16	(0.42)
I	3	0		0	
NI	3	359	(0.37)	0	
logitS	2	1979	(0.35)	1525	(0.46)
G	1	315	(0.25)	1695	(0.49)
BG	2	35	(0.34)	4	(0.49)
RF	1	563	(0.29)	2055	(0.53)
RFN	2	3001	(0.34)	2262	(0.42)
BRF	2	81	(0.35)	15	(0.36)
MM	2	2432	(0.31)	1819	(0.37)

Figure 3.15. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.5, B=4.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.5, B=4.0) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	13.36	(34)
110	21.38	(25)
111	85.50	(36)
12	213.8	(239)
size 4:		
10	0.8000	(4)
110	0.3840	(3)
1110	0.6143	(1)
120	7.679	(8)
1111	2.458	(4)
112	6.144	(3)
121	30.72	(10)
13	51.20	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	832 (0.29)	56 (0.51)
S	3	522 (0.40)	16 (0.56)
I	3	760 (0.30)	50 (0.53)
NI	3	73 (0.36)	0
logitS	2	847 (0.29)	114 (0.51)
G	1	5664 (0.27)	9087 (0.70)
BG	2	477 (0.36)	35 (0.65)
RF	1	345 (0.28)	585 (0.59)
RFN	2	44 (0.29)	0
BRF	2	82 (0.34)	12 (0.52)
MM	2	354 (0.26)	45 (0.34)

Figure 3.16. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.2, B=0.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.2, B=0.0) as the generating process for 434 binomial chains. Note that under these conditions (i.e., a=0.2, B=0.0), data generated under **MM** are equivalent to data generated under model **G**.

Expected distribution under model MM (dist. of actual data)		
chain	frequency	
size 3:		
10	8.907	(34)
110	14.90	(25)
111	76.37	(36)
12	233.8	(239)
size 4:		
10	0.5196	(4)
110	0.2232	(3)
1110	0.3691	(1)
120	5.328	(8)
1111	1.762	(4)
112	5.0867	(3)
121	30.19	(10)
13	56.52	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	862 (0.29)	90 (0.46)
S	3	548 (0.38)	3 (0.49)
I	3	400 (0.27)	11 (0.63)
NI	3	21 (0.27)	0
logitS	2	896 (0.29)	152 (0.45)
G	1	5304 (0.27)	8475 (0.68)
BG	2	414 (0.39)	72 (0.55)
RF	1	594 (0.29)	1048 (0.59)
RFN	2	147 (0.29)	24 (0.39)
BRF	2	205 (0.34)	26 (0.50)
MM	2	609 (0.25)	99 (0.37)

Figure 3.17. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.2, B=0.5) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.2, B=0.5) as the generating process for 434 binomial chains.

Expected distribution under model **MM** (dist. of actual data)		
chain	frequency	
size 3:		
10	5.938	(34)
110	10.29	(25)
111	66.90	(36)
12	250.9	(239)
size 4:		
10	0.3375	(4)
110	0.1291	(3)
1110	0.2195	(1)
120	3.658	(8)
1111	1.244	(4)
112	4.145	(3)
121	28.86	(10)
13	61.41	(67)

		# selected | 10,000 simulations (mean weight | selection)	
model	K	AICc	BIC
N	2	752 (0.31)	93 (0.51)
S	3	551 (0.36)	0
I	3	93 (0.21)	0
NI	3	30 (0.23)	0
logitS	2	1274 (0.30)	322 (0.45)
G	1	4695 (0.26)	7827 (0.65)
BG	2	352 (0.36)	30 (0.40)
RF	1	733 (0.28)	1454 (0.58)
RFN	2	323 (0.29)	31 (0.43)
BRF	2	167 (0.27)	0
MM	2	1030 (0.25)	243 (0.34)

Figure 3.18. (*left*) Expected distribution of 434 binomial chains under model **MM** (a=0.2, B=1.0) and actual distribution of the Providence data (Bailey, 1975). (*right*) Results of 10,000 Monte Carlo replicates of model selection analysis with model **MM** (a=0.2, B=1.0) as the generating process for 434 binomial chains.

APPENDIX (*chapter 3*)

FORTRAN code used to generate a single data set using model **MM** ($a = 0.9$, B=0.0) as the generating process of 664 binomial chains each starting with one initial infectious individual in a population of size 5; thus mimicking the Heasman and Reid data set (Heasman and Reid, 1961). Brackets, "[...]", indicate comments, line continuations, and line numbers.

```
[C] program  data5.f
IMPLICIT REAL (X)
REAL TOTAL5, B, A
INTEGER CHAIN, SEED1, S, I, COUNT

TOTAL5 = 664.0
A = 0.9
B = 0.0

[C] data values are initialized to 1E-4 to avoid problems associated with 0 valued cells
DATA X1, X11, X111, X1111, X11111
[+] / 1E-4, 1E-4, 1E-4, 1E-4, 1E-4/
DATA X1112, X112, X1121, X113, X12
[+] / 1E-4, 1E-4, 1E-4, 1E-4, 1E-4/
DATA X121, X1211, X122, X13, X131, X14
[+] / 1E-4, 1E-4, 1E-4, 1E-4, 1E-4, 1E-4/

OPEN (FILE = "DATA5", UNIT = 1)
OPEN (FILE = "SEEDSTORE", UNIT = 2)

READ (2,*) SEED1

DO [100], K = 1, INT(TOTAL5)
S = 4
I = 1
COUNT = 0
CHAIN = 10000
   DO [150] J = 1, 4
      IF (I .EQ. 0) THEN
         S = 0
      ENDIF

      IF (S .GT. 0) THEN
         COUNT = 0
            DO [125] II = 1, S
               IF (RAN4(SEED1,0) .LT. PSI5(ALPHA,BETA,I)) THEN
                  COUNT = COUNT + 1
               ENDIF
            [125] CONTINUE
         S = S - COUNT
         I = COUNT
      ELSE
         I = 0
      ENDIF
      CHAIN = CHAIN + I * 10**(5-J-1)
   [150] CONTINUE
```

```
IF (CHAIN .EQ. 10000) THEN
   X1 = X1 + 1
ELSEIF (CHAIN .EQ. 11000) THEN
   X11 = X11 + 1

         .
         .
         .
ELSEIF (CHAIN .EQ. 13100) THEN
   X131 = X131 + 1
ELSEIF (CHAIN .EQ. 14000) THEN
   X14 = X14 + 1
ENDIF

[100]  CONTINUE

WRITE (1,*) X1

         .
         .
         .
WRITE (1,*) X14
WRITE (1,*) "THIS IS A PLACEHOLDER"

CLOSE(1)
XDUMMY = RAN4(SEED1,1)
CLOSE(2)

END

REAL FUNCTION PSI5(A,B,I)
REAL B, A
INTEGER I
PSI5 = 1.0 - A*(4.0/5.0)**(B*REAL(I))
END
FUNCTION RAN4(SEED, TOGGLE)

[C]   simple linear congruential generator (Knuth, 1981)
INTEGER SEED, TOGGLE
PARAMETER (M=714025, IA=1366, IC=150889)
SEED = MOD(SEED*IA+IC,M)
   IF (TOGGLE .EQ. 1) THEN
      REWIND (2)
      WRITE (2,*) SEED
   ENDIF
RAN4 = FLOAT(SEED)/FLOAT(M)
RETURN
END
```

CONCLUSION

Epidemic Dynamics: Truth, Models, and Data

SUMMARY

I close this dissertation by discussing various complications of quantifying risk in epidemic models and reviewing current literature (emphasis on 1/1/98 - 4/30/99). Concepts addressed include spatial and temporal factors, individual and population heterogeneity, transmission, environmental effects, simultaneous epidemics, and multiple avenues of risk. I provide many examples of each factor. I conclude that despite the volume of literature on epidemic modeling, many fundamental questions remain open to debate and many potentially fruitful avenues of expansion are open. In response, I make several suggestions to direct future research in this area: (1) whenever possible, research should emphasize experimental challenges of current and new theories of epidemic dynamics under the paradigm of multiple working hypotheses; (2) when true experiments are not possible, quasi-experiments and correlative studies should adopt information theory based analyses and entertain several *a priori* models; (3) epidemic modeling and the concept of "epidemic" should be expanded to include the propagation of other phenomenon through populations such as ideas, fads, religions, computer viruses, product specific accidents, cancer, poisoning, substance abuse, mental illness, eating disorders, child abuse, and domestic violence; and, (4) several principles of ecology deserve more attention as factors in epidemic dynamics: convergent and divergent

evolution, interspecific competition, fragmentation, metapopulations, source and sink populations, and island biogeography.

4.1. Introduction

4.1.1 *Preliminaries*

It is a fact of life that at all times and in all places we are at risk to infection. Quantifying this risk is not simple and persists as an active topic of both academic and practical interest (e.g., Grenfell and Dobson, 1995; Mollison, 1995; Isham and Medley, 1995; Donnelly and Ferguson, 1999). The central question is "How do we describe risk in the most useful and efficient manner?" In this discussion I address this question and its related issues by integrating concepts from the philosophy of science, ecology, and epidemic modeling using examples from the current literature on risk, epidemics and epidemic dynamics with emphasis on literature published 1/1/98 - 4/30/99.

Because questions and answers must have a context, I begin by defining essential vocabulary. In the broadest sense, "risk" is the chance of change in an individual or population. "Risk to infection," my focus here, is the chance of change in an individual or population that can be attributed to some external factor which propagates through (in most cases) the biological world in time. The study of "risk" is the main topic of "epidemiology", which I consider to be "a science." I adopt Popper's definition of science, "A discipline is a science if the theories it entertains are universal and falsifiable," and that theories are falsifiable only through controlled experiments (Platt, 1964; Popper, 1972; Dolby, 1982). A distinction between statistical and mechanistic models is also required: a statistical model attempts to mimic the pattern of the data while a mechanistic model attempts to mimic the

process that created the data (Chapter 2). Also necessary are the concepts of projection, forecasting, and model fitting and selection: a projection is the logical consequences of a model; a forecast is a prediction of the future or the hidden past or present using any means (Chapter 1; Caswell, 1989); model fitting and selection is the science of selecting the "best" model for a given data set (Chapters 2-3; Burnham and Anderson, 1998).

4.1.2 *Efficient descriptions of risk*

What is an efficient description of risk? The answer at least satisfies Ocham's razor (Burnham and Anderson, 1998): the principle in philosophy that assumptions introduced to explain a thing must not be multiplied beyond necessity. In terms of model fitting and selection the answer is parsimony: a model is parsimonious if it achieves a good compromise between goodness-of-fit to the data and the precision of each parameter's estimate (Chapter 2; Burnham and Anderson, 1998). In terms of forecasting (e.g., Chapter 1), the definition is not as clear and involves many additional factors depending on the specific situation and goal of the problem- examples are measurability and economics.

4.1.3 *Useful descriptions of risk*

What is a useful description of risk? Two interpretations are advantageous since some models can be useful for their precision and some for their uncertainty. Models of risk capable of making precise forecasts would adequately describe the essential statistics and dynamics of the cause (e.g. threshold dynamics, final size; Heesterbeek and Roberts, 1995) and assist in the design of management strategies (e.g. vaccination protocols; Anderson and May, 1991). Less capable models are still useful, but in a different sense because they expose important

testable hypotheses (e.g., Chapter 1).

4.1.4 *Epidemics*

Epidemics, literally "upon the population," occur when there is a noticeable appreciation in a risk (or risks) within a population or among populations. Epidemics occur at 3 spatial scales: (1) epidemics occur within individuals as a condition spreads throughout the body- an unusual but technically correct interpretation (e.g., Afenya and Bentil, 1998; Wu and Ding, 1999); (2) between individuals- the most common interpretation; and, (3) among populations of individuals. If "risk" itself is a function of the number or presence of infected individuals in (or associated with) the population, the epidemic is said to be an "infectious epidemic." The term "infectious disease epidemic" is reserved for situations in which harm comes to individuals of the population who are infected.

4.2. Concepts and components of epidemics

4.2.1 *Truth is complex*

Epidemic dynamics and individual risk are certainly complex. The most general model of risk recognizes only 2 interacting concepts, the susceptibility of the individual and the hostility of the environment: risk to individual i at time t is a function of the state of individual i at time t and the state of environment at time t. Many useful divisions of this global model have been explored in the literature. The earliest treatments (e.g., Hamer, 1906; Kermack and McKendrick, 1927; Greenwood, 1931) focus on the individual as a member of one of 4 sets at any point in time with the total population remaining closed (Bailey, 1975; Anderson and May, 1991): susceptible; infected but not infectious and therefore incubating; infectious; and

immune. Subsequent work builds on this foundation, recognizing additional components of the epidemic process while attempting to produce useful and efficient models of more specific problems (Chapters 1-2): variation in modes of transmission; individual heterogeneity with respect to infectiousness, susceptibility and incubation; host and vector population dynamics; spatial spread; and, environmental factors.

4.2.2 *Susceptibility*

The simplest interpretation of "susceptibility" has two components: (1) susceptibility is a binomial condition, either the individual is susceptible or immune; and, (2) every member of the population in question is identically susceptible. A more realistic interpretation of "susceptibility" admits 4 additional concepts. First, susceptibility varies among individuals at any given time as the result of some combination of two factors: (1) the probability of successful infection given equivalent contact with infectious material varies among individuals; and, (2) the probability of contact with infectious material varies among individuals. Second, susceptibility of the individual varies in time. This can occur through some combination of at least 3 mechanisms: (1) behavioral changes in the susceptible individual; (2) changes in the health of the susceptible individual which lower the dose necessary for infection (e.g., malnutrition, immunity from previous exposure or vaccination may be transient); and, (3) accumulation of contacts with infectious sources resulting in the achievement of a threshold necessary for true infection. Third, susceptibility varies by population. The mean susceptibility of individuals may be different from population to population. Fourth, the concept of "susceptibility" applies among populations. Populations may be differentially susceptible to initial infection.

Current literature reveals that differential susceptibility is the rule rather than the exception (Table 4.1). Because differential susceptibility can have major effects on projection and forecasting exercises, interest in this issue among epidemic modelers has recently surged (e.g. Halloran, 1996; Castillo-Chavez et al., 1997; Dushoff, 1997; Dwyer, 1997; D'Amico et al, 1998; Gilbert et al., 1998; Karlaftis and Tarko 1998; Lui, 1998; Moller et al., 1998; Siegmund et al., 1998; Yang and Silveira 1998; Ferguson et al., 1999). However, absent from the literature are comprehensive assessments of the relative importance of the factors predicting differential susceptibility and their distributions in populations of interest for specific diseases (HIV is an exception: Table 4.1).

4.2.3 *Infectiousness*

Like "susceptibility", the simplest interpretation of "infectiousness" has two components: (1) infectiousness is a binomial condition, either the source is infectious or not; and, (2) infectious sources are identically infectious. A more realistic interpretation of infectiousness admits 4 additional concepts. First, infectiousness varies among sources at any given time as the result of some combination of 3 factors: (1) some sources are more likely to confer infection given equivalent contact with a susceptible individual, perhaps providing a larger dose per contact or a more effective delivery per contact; (2) some infectious sources may make more infectious contacts per unit time; and, (3) the infectiousness of any given source may vary in time and different sources may be at different stages in their own infectiousness. Third, infectiousness varies by population. Mean infectiousness may vary by population, in spite of an equivalent number of infectious sources. Fourth, the concept of "infectiousness" applies among populations. Populations may be differentially infectious to other populations in spite

of an equivalent number of infectious sources in each.

The concept of "differential infectiousness" has received less attention among epidemic modelers than the concept of "differential susceptibility" (e.g., Blythe and Anderson, 1988; Cairns, 1990; Iannelli et al., 1992; Stigum, 1994; Baker and Stevens, 1995; Garnett and Anderson, 1996; Halloran, 1996; Castillo-Chavez et al., 1997; Taylor et al., 1998; Hyman et al., 1999). This may be because clear examples of differential infectiousness are relatively rare (Table 4.2).

4.2.4 *Transmission and transmission models*

Adding to the complexity of differential susceptibility and infectiousness is the concept of "transmission" (Fig. 4.1). Transmission is the act of passing a dose of infectious material from a source to another entity: vertical transmission occurs when a dose is transmitted from a parent to an offspring (e.g., "maternal transmission" in the model of Chapter 1); horizontal transmission occurs when a dose is passed between peers (e.g., Chapter 2: models **RF**, **MM**, and the "mass action assumption)"; and, infection from a "point source" (e.g., Chapter 2: models **G**, **BG**, and **PP**) occurs when a dose is received from a reservoir of infectious material whose infectiousness is largely independent of the number of infectious individuals in the populations (e.g., a contaminated water supply). Both Inter- and intra-specific transmission occur through 4 general mechanisms (numerous examples of each: Cecil et al, 1996; Olendorf et al., 1999): (1) by direct contact (e.g., intercourse, kissing, across the placenta, breast feeding) or by a proximate inanimate mechanical vector (e.g., droplet spray from coughing, sharing a drinking glass); (2) through an inanimate mechanical vector (e.g., contaminated water or food); (3) through an animate mechanical vector, one not actually infected (e.g.,

filariasis: larvae are transmitted from the bloodstream of one person to another by mosquitos); and, (4) through a biological vector, one actually involved with the life cycle of the parasite (e.g., malaria: mosquito vector is an intermediate host). Further complicating these classifications is the concept of an intermediate vector, one that disseminates the infectious material to the actual vector. For example, flies and cockroaches can transport infectious material from feces to food (e.g., toxoplasmosis).

Modeling the transition from the susceptible to the infected or incubating state is the central problem of mechanistic epidemic modeling (Chapter 2). Like risk, the most global model is simple: risk of infection to individual i at time t is a function of the state of individual i at time t and the state of environment at time t. Of course, this "simple" global model is deceptive; the goal of creating useful epidemic models is to determine which variables will produce an acceptable approximation to truth (Chapters 1-3).

4.2.5 *Incubation*

Clinically, "incubation" refers to the time period between infection, or the initiation of the illness, and the first appearance of symptoms; however, for our purposes, incubation is the period between infection and infectiousness. Again, heterogeneity is the rule; virtually all diseases show a 0.5 to 4 fold range in incubation periods among cases (human examples: Cecil et al., 1996; Olendof et al., 1999). With 3 exceptions, investigators have given little attention to modeling this heterogeneity. First, the incubation period of some infectious diseases varies inversely with the size of the initial infectious dose (e.g., Typhoid fever: Cecil et al, 1996; Olendorf et al., 1999). Consequently, especially dose-potent infectious sources, those that transmit a larger than average dose per contact but do not necessarily have more

infectious contacts per unit time, may affect epidemic dynamics by shortening the incubation periods of some individuals, perhaps increasing the speed of the epidemic through the population. Second, the incubation period of some infectious plant diseases varies with temperature (e.g., Godoy et al., 1999; Xu, 1999). This may be a general rule among poikilotherms and hibernating homeotherms, but documentation is lacking. Third, the mean incubation period of HIV resulting from blood transfusions, appears to vary by age (1.97 years for children (0-4 yrs old at infection), 8.23 years for adults (5-59 yrs) , and 5.50 years for elderly patients (≥60 yrs); Medley et al., 1987). These issues deserve more attention.

4.2.6 *Spatial spread*

Examples of the spatial spread of pathogens, both within and among populations, are abundant in the current literature (Table 4.3). Spatial spread from population to population occurs in 2 ways: the infectious agent migrates by a vector, or infectious individuals migrate. Boats and planes are important intermediate vectors of human pathogens from population to population; and, as an interesting caveat, the confined conditions in planes and boats might result in a small epidemic among passengers while in transit, thus increasing the probability of transmission to the new population when the destination is reached (e.g., Klontz et al., 1989). Epidemic models addressing spatial spread are abundant (e.g., Thieme, 1980; Barlow, 1991; Bolker et al., 1995; Sattenspiel and Dietz, 1995; Clancy, 1996; Rhodes and Anderson, 1996; Bonabeau et al., 1998; Filipe and Gibson, 1998; Murray and Salomon, 1998; Swinton, 1998; Ball, 1999).

4.2.7 *Climate, weather and pollution effects*

Climate is the most significant predictor of the geographic distribution of a species; seasonal patterns often predict the behavior and grouping of populations within geographic ranges (e.g., migration to breeding areas); and, day-to-day weather usually affects the movements and behavior of individual members of populations (e.g., schools close on "snow days"). Consequently, climate and weather are factors affecting differential susceptibility and infectiousness. Curiously, this issue is relatively undeveloped in the literature (Table 4.4).

4.3. Synthesis and recommendations

4.3.1 *Characteristics of the current literature on epidemic modeling*

Examples of epidemic models are numerous in the literature. For this discussion I reviewed over 200 papers and books on epidemic dynamics and related topics. To quantify the current trends of this literature, I queried *Biological Abstracts*, and *Medline* and censussed *Biometrics*, *Ecology* and *The American Naturalist*, for instances of epidemic models published 1/1/98-4/30/99; I found 42 examples. I classified 33 (77%) as adopting a mechanistic approach, 8 (19%) as adopting a statistical approach, and only 1 incorporating both approaches (e.g., Ellner et al., 1998). I further classified 33 (77%) as exercises in model projection and forecasting and only 10 (23%) as exercises in model fitting and selection. Based on this sample, the proceeding discussion, and Chapters 1-3, I have 5 recommendations.

4.3.2 *First recommendation*

The most striking feature of the literature on epidemic modeling is the dearth of experimental evidence for virtually any theory proposed. For example, none of the 43 recent papers (1/1/98-4/30/99) I reviewed reported on the results of an experiment. In fact, among the 200+ papers reviewed, <10 reported any sort of experimental challenge (i.e., random allocation of treatments and controls, replication, etc..) to proposed theories (notable exceptions: Dwyer, 1991,1992; D'Amico et al., 1996; Dwyer et al., 1997). Experimental manipulations, preferably under the paradigm of "multiple working hypotheses" (Chamberlain, 1964), are the best way to obtain reliable knowledge (Platt, 1964; Popper, 1972). More experimental challenges of the underlying theories of epidemic dynamics are needed.

4.3.3 *Second recommendation*

The Paradigm of information theory-based model selection using Akaiki's Information Criteria (AIC) offers numerous advantages over other approaches (Chapters 2-3; Burnham and Anderson, 1998). Because many models can be fit to a data set, it is important to entertain other possible explanations of a phenomenon and use the available data in the most efficient manner (i.e., bias variance tradeoff: Chapters 2-3; Burnham and Anderson 1998). Although this paradigm is now common in many fields (e.g., time series analysis, capture-recapture: Brunham and Anderson, 1998), in my review I found only 2 examples in the epidemic modeling literature in which AIC was used to pick a best approximating model from a candidate set (i.e., Dwyer et al., 1997; Wu and Ding, 1999). This approach should be integrated whenever data from quasi-experiments are encountered- the rule, rather than the exception, of most epidemic data.

4.3.4 *Third recommendation*

The concept of "epidemic" is broader than usually assumed in the literature on epidemic modeling. Any phenomenon whose incidence in a population increases with time is an epidemic (Table 4.5). There are few examples of epidemic models of other phenomenon besides the host-parasite system in the literature. These analogies deserves more attention.

4.3.5 *Fourth recommendation*

Ecology is "the study of the natural environment and of the relations of organisms to each other and to their surroundings" (Ricklefs and Gary, 1999). Although the integration of epidemiology and ecology has received considerable recent attention (e.g, Grenfell and Dobson, 1995) many of the classic concepts and problems of ecology deserve more attention within the context of infectious disease.

Ecology defines "fitness" as the "genetic contribution by an individual's descendants to future generations of the population" (Ricklefs and Gary, 1999). Fitness is an important component of evolution: the selection of a trait occurs if, and only if, the trait is heritable and some factor in the environment confers a fitness advantage to those individuals who posses the trait. Four questions are immediate. First, what makes a pathogen "fit?" Second, what is the fitness advantage, if any, of "virulence?" Third, by what diversity of mechanisms do pathogens and their hosts coevolve? Fourth, can examples of convergent and divergent evolution among pathogens be identified in the current literature? Although the evolution of pathogens is well documented, especially in response to antibiotics and vaccines (e.g., Aggarwal, 1999; Beltran et al., 1999; Villa et al., 1999) and many pathogens show remarkably similar epidemiology (e.g., Herpes and AIDS, Dobbins et al., 1999), these questions have

received relatively little attention when compared to the volume of literature on epidemic models (notable exceptions: Anderson and May, 1982; Massad, 1987; Herre, 1993; Ewald, 1994; Andreasen and Christiansen, 1995; Read et al., 1995; Goff et al., 1997; Ebert, 1998; Hastings, 1998; Bergstrom et al., 1999).

Competition is the "use or defense of a resource by one individual that reduces the availability of that resource to other individuals, whether of the same species or other species" (Ricklefs and Gray, 1996). Under what conditions do pathogens compete with one another? Can pathogens affect the competitiveness of their hosts (Yan et al., 1998)?

A meta-population is a population of populations; (Hanski and Gilpin, 1997). The member populations are often classified into 2 categories: "sources" provide excess members to colonize new areas and maintain "sink" populations, those that would not persist without constant immigration. Are there source and sink populations of infectious disease? For example, are developed countries sink populations for pathogens which are readily detected and cured with adequate medical care in developed countries. Documentation of possible meta-populations in the current literature are rare (e.g., Kern et al., 1999). These hypotheses are testable.

Fragmentation, the disruption of habitat into discontinuous parts (Bissonette, 1997), has at least 3 implications for host-parasite dynamics. First, fragmentation can affect the distribution of animate and biological vectors, possibly limiting the geographic range of a pathogen. Second, fragmentation which separates populations of hosts, should reduce the rate of transfer of pathogens among populations. Third, the de-fragmentation of host habitats should favor pathogens. Has the increased density and connectedness of the global human population "de-fragmented" the habitat (i.e. humans) of some human parasites? These

hypotheses are testable.

The theory of island biogeography makes two predictions about the colonization of islands from mainlands (MacArthur, and Wilson, 1967; Whittaker, 1998): (1) islands closer to mainlands should have greater species diversity than islands further from mainlands; and, (2) larger islands should have greater species diversity than smaller islands. This concept applies to endemics and epidemics at two scales: the individual as an island and the population as an island. By analogy, the theory predicts that individuals living in isolation should harbor a less diverse community of parasites and be less likely to suffer new colonization. At the population level, isolated populations should have less parasite diversity. Because epidemics represent new colonization, by analogy, the theory predicts that large cosmopolitan populations should experience a greater diversity of epidemics per unit time than smaller isolated populations. These hypotheses are testable.

Acquavella, J., Olsen, G., Cole, P., Ireland, B., Kaneene, J., Schuman, S., and Holden, L. (1998). Cancer among farmers: a meta-analysis. *Annals of Epidemiology* **8**, 64-74.

Afenya, E. K. and Bentil, D. E. (1998). Some perspectives on modeling leukemia. *Mathematical Biosciences* **150**, 113-130.

Aggarwal, R., McCasuteland, K. A., Dilawari, J. B., Sinha, S. D., and Robertson, B. H. (1999). Genetic variability of hepatitis E virus within and between three epidemics in India. *Virus Research* **59**, 35-48.

Anderson, R. M. and May, R. M. (1982). Coevolution of hosts and parasites. *Parisitology* **85**, 411-426.

Anderson, R. M. and May, R. M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford: Oxford University Press.

Andreasen, V. (1989). Disease regulation of age-structured host populations. *Theoretical Population Biology* **36**, 214-239.

Andreasen, V. and Christiansen, F. B. (1995). Slow coevolution of a viral pathogen and its diploid host. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* **348**, 341-354.

Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*, 2nd edition. London: Charles Griffin and Company.

Baker, R. D. and Stevens, R. H. (1995). A random-effects model for analysis of infectious disease final-state data. *Biometrics* **51**, 956-968.

Ball, F. (1999). Stochastic and deterministic models for SIS epidemics among a population partitioned into households. *Mathematical Biosciences* **156**, 41-67.

Barlow, N. D. (1991). A spacially aggregated disease/host model for bovine Tb in New Zealand possum populations. *Journal of Applied Ecology* **28**, 777-793.

Barlow, N. D. (1995). Critical evaluation of wildlife disease models. In *Ecology of infectious disease in natural populations*, B. T. Grenfell and A. P. Dobson (eds), 230-259. Cambridge: Cambridge University Press.

Barlow, N. D. (1996). The ecology of wildlife disease control: simple models revisited. *Journal of Applied Ecology* **33**, 303-314.

Beltran, P., Delgado, G., Navarro, A., Trujillo, F., Selander, R. K., and Cravioto, A. (1999). Genetic diversity and population structure of *Vibrio cholerae*. *Journal of Clinical Microbiology* **37**, 581-590.

Bergstrom, C. T., McElhany, P., and Real, L. A. (1999). Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5095-5100.

Bissonette, J. A. (1997). *Wildlife and Landscape Ecology: Effects of Pattern and Scale*. New York: Springer.

Blythe, S. P. and Anderson, R. M. (1988). Variable infectiousness in HIV transmission models. *IMA Journal of Mathematics Applied in Medicine & Biology* **5**, 181-200.

Boiko, A. V. and Pogorelova, N. P. (1998). Influence of technogenic pollutions on bacterial associations of a water reservoir. *Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii* **0**, 23-25.

Bolen, J. R., Kresnow, M., and Sacks, J. J. (1998). Reported bicycle helmet use among adults in the United States. *Archives of Family Medicine* 7, 72-77.

Bolker, B. M., Altmann, M., Aubert, M., Ball, F., Barlow, N. D., Bowers, R. G., Dobson, A. P., Elkington, J. S., Garnett, G. P., Gilligan, C. A., Hassel, M. P., Isham, V., Jacquez, J. A., Kleczkowski, A., Levin, S. A., May, R. M., Metz, J. A. J., Mollison, D., Morris, M., Real, L. A., Sattenspiel, L., Swington, J., White, P., and Williams, B. G. (1995). Group report: Spatial dynamics of infectious diseases in natural populations. In *Ecology of Infectious Disease in Natural Populations*, B. T. Grenfell and A. P. Dobson (eds), 399-420. Cambridge: Cambridge University Press.

Bonabeau, E., Toubiana, L., and Flahault, A. (1998). The geographical spread of influenza. *Proceedings of the Royal Society of London - Series B: Biological Sciences* **265**, 2421-2425.

Boschini, A., Di Perri, G., Legnani, D., Fabbri, P., Ballarini, P., Zucconi, R., Boros, S., and Rezza, G. (1999). Consecutive epidemics of Q fever in a residential facility for drug abusers: Impact on persons with human immunodeficiency virus infection. *Clinical Infectious Diseases* **28**, 866-872.

Braden, C. R. (1995). Infectiousness of a university student with laryngeal and cavitary tuberculosis. *Clinical Infectious Diseases* **21**, 565-570.

Briggs, C. J. and Godfray, H. C. J. (1995). The dynamics of insect-pathogen interactions in stage structured populations. *American Naturalist* **145**, 855-887.

Brooks, A. J., Stuewig, J., and LecRoy, C. W. (1998). A family based model of Hispanic adolescent substance use. *Journal of Drug Education* **28**, 65-86.

Burnham, K. P. and Anderson, D. A. (1998). *Model selection and inference: A practical information theoretic approach*. New York: Springer-Verlag.

Caces, F. and Harford, T. (1998). Time series analysis of alcohol consumption and suicide mortality in the United States, 1934-1987. *Journal of Studies on Alcohol* **59**, 455-461.

Cairns, A. J. (1990). Epidemics in heterogeneous populations: II. Nonexponential incubation periods and variable infectiousness. *IMA Journal of Mathematics Applied in Medicine & Biology* 7, 219-230.

Castaneda, E., Penuela, I., Vela, M. C., Group Colombian Pneumococcal Study, and Tomasz, A. (1998). Penicillin-resistant Streptococcus pneumoniae in Colombia: Presence of international epidemic clones. *Microbial Drug Resistance-Mechanisms Epidemiology & Disease* **4**, 233-239.

Castillo-Chavez, C. (ed.) (1989). *Mathematical and Statistical Approaches to AIDS Epidemiology*. New York: Springer-Verlag.

Castillo-Chavez, C., Huang, W., and Li, J. (1997). The effects of females' susceptibility on the coexistence of multiple pathogen strains of sexually transmitted diseases. *Journal of Mathematical Biology* **35**, 503-522.

Caswell, H. (1989). *Matrix population models, construction, analysis, and interpretation*. Sunderland, Massachusetts: Sinauer Associates.

Cates, W., Jr., Chesney, M. A., and Cohen, M. S. (1997). Primary HIV infection- a public health opportunity. *American Journal of Public Health* **87**, 1928-1930.

Cecil, R. L., Plum, F., and Bennett, J. C. (eds) (1996). *Cecil Textbook of Medicine*. New York: W. B. Saunders.

Celentano, D. D., Hodge, M. J., Razak, M. H., Beyrer, C., Kawichai, S., Cegielski, J. P., Nelson, K. E., and Jittiwutikarn, J. (1999). HIV-1 incidence among opiate users in northern Thailand. *American Journal of Epidemiology* **149**, 558-564.

Chamberlain, T. C. (1964). The method of multiple working hypotheses. *Science* **148**, 754-759.

Cisse, B., Aaby, P., Simondon, F., Samb, B., Soumare, M., and Whittle, H. (1999). Role of schools in the transmission of measles in rural Senegal: Implication for measles control in developing countries. *American Journal of Epidemiology* **149**, 295-301.

Clancy, D. (1996). Carrier-borne epidemic models incorporating population mobility. *Mathematical Biosciences* **132**, 185-204.

Cook-Mozaffari, P., Newton, R., Beral, V., and Burkitt, D. P. (1998). The geographical distribution of Kaposi's sarcoma and of lymphomas in Africa before the AIDS epidemic. *British Journal of Cancer* **78**, 1521-1528.

Cox, B. and Skegg, K. (1993). Contagious suicide in prisons and police cells. *Journal of Epidemiology & Community Health* **47**, 69-72.

D'Amico, V., Elkinton, J. S., Dweyer, G., Burand, J. P., and Bounaccorsi, J. P. (1996). Virus transmission in gypsy moths is not a simple mass action process. *Ecology* 77, 201-206.

Datta, S., Halloran, M. E., and Longini, I. M., Jr. (1998). Augmented HIV vaccine trial design for estimating reduction in infectiousness and protective efficacy. *Statistics in Medicine* **17**, 185-200.

Day, J. F. and Curtis, G. A. (1999). Blood feeding and oviposition by *Culex nigripalpus* (Diptera: *Culicidae*) before, during and after a widespread St. Louis encephalitis virus epidemic in Florida. *Journal of Medical Entomology* **36**, 176-181.

Dearwater, S. R., Coben, J. H., Campbell, J. C., Nah, G., Glass, N., McLoughlin, E., and Bekemeier, B. (1998). Prevalence of intimate partner abuse in women treated at community hospital emergency departments. *Journal of the American Medical Association* **5**, 433-438.

Deparis, X., Roche, C., Murgue, B., and Chungue, E. (1998). Possible dengue sequential infection: Dengue spread in a neighborhood during the 1996/97 dengue-2 epidemic in French Polynesia. *Tropical Medicine & International Health* **3**, 866-871.

De Roo, A., Ado, B., Rose, B., Guimard, Y., Fonck, K., and Colebunders, R. (1998). Survey among survivors of the 1995 Ebola epidemic in Kikwit, Democratic Republic of Congo: Their feelings and experiences. *Tropical Medicine & International Health* **3**, 883-885.

Diehl, D., Niu, H. S., and Rupprecht, C. E. (1998). Efficacy of an oral vaccinia-rabies glycoprotein recombinant vaccine in controlling epidemic raccoon rabies in New Jersey. *Journal of Wildlife Diseases* **34**, 752-763.

Dietz, K. (1967). Epidemics and rumors: A survey. *Royal Statistical Society* **130**, 505-508.

Dobbins, J. G., Mastro, T. D., Nopkesorn, T., Sangkharomya, S., Limpakarnjanarat, K., Weniger, B. G., and Schmid, D. S. (1999). Herpes in the time of AIDS: A comparison of the epidemiology of HIV-1 and HSV-2 in young men in northern Thailand. *Sexually Transmitted Diseases* **26**, 67-74.

Dolby, G. R. (1982). The role of statistics in the methodology of the life sciences. *Biometerics* **38,** 1069-1083.

Donisi, A., Tomasoni, D., Ripamonti, D., Milini, P., Palvarini, L., Cattane, A., Paraninfo, G., Casari, S., Cadeo, G. P., and Carosi, G. (1998). Changing patterns of HIV transmission and better targeting for intervention strategies. *International Journal of STD & AIDS* **9**, 740-743

Donnelly, C. A. and Ferguson, N. M. (1999). *Statistical Aspects of BSE; Models for an Epidemic*. New York: Chapman & Hall.

Duncan, S. C., Duncan, T. E., Biglan, A., and Ary, D. (1998a). Contributions of the social context to the development of adolescent substance use: a multivariate latent growth modeling approach. *Drug & Alcohol Dependence* **50**, 57-71.

Duncan, T. E., Duncan, S. C., and Hops, H. (1998b). Latent variable modeling of longitudinal and multilevel alcohol use data. *Journal of Studies on Alcohol* **59**, 399-408.

Dushoff, J. G. (1997). *Modeling the effects of host heterogeneity on the spread of human diseases*. PhD dissertation. Princeton: Princeton University.

Dwyer, G. (1991). The role of density, stage, and patchiness in the transmission of an insect virus. *Ecology* **72**, 559-574.

Dwyer, G. (1992). On the spatial spread of insect pathogens: theory and experiment. *Ecology* **73**, 479-494.

Dwyer, G., Elkinton, J. S., and Buonaccorsi, J. P. (1997). Host heterogeneity in susceptibility and disease dynamics: Tests of a mathematical model. *American Naturalist* **150**, 685-707.

Ebert, D. (1998). Experimental evolution of parasites. *Science* **282**, 1432-1435

Egan, D. J. and Robinson, D. O. (1979). Models of a heroin epidemic. *American Journal of Psychiatry* **136**, 1162-1167.

Elliott, A. M., Hayes, R. J., Halwiindi, B., Luo, N., Tembo, G., Pobee, J. O., Nunn, P. P., and McAdam, K. P. (1993). The impact of HIV on infectiousness of pulmonary tuberculosis: a community study in Zambia. *AIDS* **7**, 981-987.

Ellner, S. P., Bailey, B. A., Bobashev, G. V., Gallant, A. R., Grenfell, B. T., and Nychka, D. W. (1998). Noise and nonlinearity in measles epidemics: Combining mechanistic and statistical approaches to population modeling. *American Naturalist* **151**, 425-440.

Etard, J. F. and Borel, E. (1992). Man-water contacts and urinary schistosomiasis in a Mauritanian village. *Revue d Epidemiologie et de Sante Publique* **40**, 268-275.

Ewald, P. W. (1994). *Evolution of infectious disease*. New York: Oxford University

Ferguson, N., Anderson, R., and Gupta, S. (1999). The effect of antibody-dependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 790-794.

Ferson, M. J., Young, L. C., and Stokes, M. (1998). Changing epidemiology of hepatitis A in the 1990's in Sydney, Australia. *Epidemiology & Infection* **121**, 631-636.

Filipe, J. A. N. and Gibson, G. J. (1998). Study and approximating spatio-temporal models for epidemic spread and control. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **353**, 2153-2162.

Fleming, D. T. and Wasserheit, J. N. (1999). From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. *Sexually Transmitted Infections* **75**, 3-17.

Garnett, G. P. and Anderson, R. M. (1996). Antiviral therapy and the transmission dynamics of HIV-1. *Journal of Antimicrobial Chemotherapy* **37**, 135-150.

Georges, A., Leroy, E. M., Renaut, A. A., Benissan, C. T., Nabias, R. J., Ngoc, M. T., Obiang, P. I., Lepage, J. P. M., Bertherat, E. J., Benoni, D. D., Wickings, E. J., Amblard, J. P., Lansoud-Soukate, J. M., Milleliri, J. M., Baize, S., and Georges-Courbot, M. (1999). Ebola hemorrhagic fever outbreaks in Gabon, 1994-1997: Epidemiologic and health control issues. *Journal of Infectious Diseases* **179**, S65-S75.

Gibney, L., Choudhury, P., Khawaja, Z., Sarker, M., and Vermund, S. H., (1999). Behavioural risk factors for HIV/AIDS in a low-HIV prevalence Muslim nation: Bangladesh. *International Journal of STD & AIDS* **10**, 186-194.

Gilbert, P. B., Self, S. G., and Ashby, M. A. (1998). Statistical methods for assessing differential vaccine protection against human immunodeficiency virus types. *Biometrics* **54**, 799-814.

Godoy, C. V., Amorim, L., and Filho, A. B. (1999). Influence of leaf wetness duration and temperature on the development of maize rust caused by *Puccinia polysora*. *Fitopatologia Brasileira* **24**, 160-165.

Goff, L. J., Ashen, J., and Moon, D. (1997). The evolution of parasites from their hosts: A case study in the parasitic red algae. *Evolution* **51**, 1068-1078.

Goffman, W. and Newill, V. A. (1964). Generalization of epidemic theory: An application to the transmission of ideas. *Nature* **204**, 225-228.

Goran, M. I., Shewchuk, R., Gower, B. A., Nagy, T. R., Carpenter, W. H., and Johnson, R. K. (1998). Longitudinal changes in fatness in white children: no effect of childhood energy expenditure. *American Journal of Clinical Nutrition* **67**, 309-316.

Goyer, R. A. (1996). Results of lead research: prenatal exposure and neurological consequences. *Environmental Health Perspectives* **104**, 1050-1054.

Grassi, E., Petrillo, M., Guidi, E., Angelini, L., Raspadori, F., and Gregorio, P. (1998). The control of cholera epidemics: Continuity or discordance in the health interventions. *Igiene Moderna* **109**, 541-556.

Greenwood, M. (1931). On the statistical measure of infectiousness. *Journal of Hygiene, Cambridge* **31**, 336-351.

Gupta, P., Kumar, P., and Aggarwal, O. P. (1998). Knowledge, attitude and practices related to dengue in rural and slum areas of Delhi after the dengue epidemic of 1996. *Journal of Communicable Diseases* **30**, 107-112.

Halloran, M. E. (1996). Epidemiologic effects of varicella vaccination. *Infectious Disease Clinics of North America* **10**, 631-655.

Halloran, M. E., Anderson, R. M., Azevedo-Neto, R. S., Bellini, W. J., Branch, O., Burke, M. A., Compans, R., Day, K., Gooding, L., Gupta, S., Katz, J., Kew, O., Keyserling, H., Krause, R., Lal, A. A., Massad, E., McLean, A. R., Rosa, P., Rota, P., Wiener, P., Wynn, S. G., and Zanetta, D. M. (1998). Population biology, evolution, and immunology of vaccination and vaccination programs. *American Journal of the Medical Sciences* **315**, 76-86.

Hamer, W. H. (1906). Epidemic disease in England- the case evidence of variability and persistence of type. *Lancet* **1**, 733-739.

Han, L. L., Popovici, F., Alexander, J. P., Jr., Laurentia, V., Tengelsen, L. A., Cernescu, C., Gary, H. E., Jr., Ion-Nedelcu, N., Campbell, G. L., and Tsai, T. F. (1999). Risk factors for West Nile virus infection and meningoencephalitis, Romania 1996. *Journal of Infectious Diseases* **179**, 230-233.

Hanski, I. and Gilpin, M. E. (eds) (1997). Metapopulation biology: ecology, genetics, and evolution. San Diego, California: Academic Press.

Hastings, I. M. (1998). A model for the origins and spread of drug-resistant malaria. *Parasitology* **115**, 133-141.

Haverkos, H. W. (1998). HIV/AIDS and drug abuse: Epidemiology and prevention. *Journal of Addictive Diseases* **17**, 91-103.

Hazell, P. (1993). Adolescent suicide clusters: evidence, mechanisms and prevention. *Australian & New Zealand Journal of Psychiatry* **27**, 653-665.

Heesterbeek, J. A. P. and Roberts, M. G. (1995). Mathematical models for microparasites of wildlife. In *Ecology of Infectious Disease in Natural Populations*, B. T. Grenfell and A. P. Dobson (eds), 90-122. Cambridge: Cambridge University Press.

Herre, E. A. (1993). Population structure and the evolution of virulence in a nematode parasite of fig wasps. *Science* **259**, 1442-1445.

Hoppensteadt, F. C. and Murray, J. D. (1981). Threshold analysis of a drug use epidemic model. *Mathematical Bioscience* **53**, 79-87.

Hotta, S. (1999). Dengue vector mosquitoes in Japan: The role of *Aedes albopictus* and *Aedes aegypti* in the 1942-1944 dengue epidemics of Japanese Main Islands. *Medical Entomology & Zoology* **49**, 267-274.

Hyman, J. M., Li, J., and Stanley, E. A. (1999). The differential infectivity and staged progression models for the transmission of HIV. *Mathematical Biosciences* **155**, 77-109.

Iannelli, M., Loro, R., Milner, F., Pugliese, A., and Rabbiolo, G. (1992). An AIDS model with distributed incubation and variable infectiousness: applications to i.v. drug users in Latium, Italy. *European Journal of Epidemiology* **8**, 585-593.

Ishman, V. and Medley, G. (eds) (1995). *Models for Infectious Human Diseases: Their Structure and Relation to Data*. Cambridge: Cambridge University Press.

Jackson, C., Henriksen, L., Dickinson, D., Messer, L., and Robertson, S. B. (1998). A longitudinal study predicting patterns of cigarette smoking in late childhood. *Health Education & Behavior* **25**, 436-447.

James, J., Gosho, C., Wohl, R. W. (1979). The relationship between female criminality and drug use. *International Journal of the Addictions* **14**, 215-229.

Judson, F. N. (1992). Interactions between human papillomavirus and human immunodeficiency virus infections. *IARC Scientific Publications (Lyon)* **119**, 199-207.

Kamimura, K. (1998). Studies on the population dynamics of the principal vector mosquito of Japanese encephalitis. *Medical Entomology & Zoology* **49**, 181-185.

Karlaftis, M. G. and Tarko, A. P. (1998). Heterogeneity considerations in accident modeling. *Accident Analysis & Prevention* **30**, 425-433.

Kermack, W. O. and McKendrick, A. G. (1927). Contributions to the mathematical theory of epidemics, part 1. *Proceedings of the Royal Society of Edinburgh* **115**. 700-721.

Kern, B., Depner, K. R., Letz, W., Rott, M., Thalheim, S., Nitschke, B., Plagemann, R., and Liess, B. (1999). Incidence of classical swine fever (CSF) in wild boar in a densely populated area indicating CSF virus persistence as a mechanism for virus perpetuation. *Journal of Veterinary Medicine Series B* **46**, 63-67.

Kilian, A. H. D., Gregson, S., Ndyanabangi, B., Walusaga, K., Kipp, W., Sahlmuller, G., Garnett, G. P., Asiimwe-Okiror, G., Kabagambe, G., Weis, P., and Von Sonnenburg, F. (1999). Reductions in risk behaviour provide the most consistent explanation for declining HIV-1 prevalence in Uganda. *AIDS (Hagerstown)* **13**, 391-398.

King, N. J., Eleonora, G., and Ollendick, T. H. (1998). Etiology of childhood phobias: current status of Rachman's three pathways theory. *Behaviour Research & Therapy* **36**, 297-309.

Klontz, K. C., Hynes, N. A., Gunn, R. A., Wilder, M. H., Harmon, M. W., and Kendal, A. P. (1989). An outbreak of Influenza A/Taiwan/1/86(H1N1) infections at a naval base and its association with airplane travel. *American Journal of Epidemiology* **129**, 341-348.

Koike, S. (1999). Low-level lead exposure and children's intelligence from recent epidemiological studies in the U.S.A. and other countries to progress in reducing lead exposure and screening in the U.S.A. *Japanese Journal of Hygiene* **52**, 552-561.

Krieger, J. N., Nirapathpongporn, A., Chaiyaporn, M., Peterson, G., Nikolaeva, I., Akridge, R., Ross, S. O., and Coombs, R. W. (1998). Vasectomy and human immunodeficiency virus type 1 in semen. *Journal of Urology* **159**, 820-826.

Lazzarin, A., Saracco, A., Musicco, M., and Nicolosi, A. (1991). Man-to-woman sexual transmission of the human immunodeficiency virus. *Archives of Internal Medicine* **151**, 2411-2416. Erratum appears in *Archives of Internal Medicine* **152**, 876.

Le Grange, D., Telch, C. F., and Tibbs, J. (1998). Eating attitudes and behaviors in 1,435 South African Caucasian and non-Caucasian college students. *American Journal of Psychiatry* **155**, 250-254. Comment in *American Journal of Psychiatry* **156**, 981-982.

Lindeman, A. K. (1999). Quest for ideal weight: costs and consequences. *Medicine & Science in Sports & Exercise* **31**, 1135-1140.

Lipskaya, G. Y., Cherkasova, E. A., Ivanova, O. E., and Drozdov, S. G. (1998). Geotyping of wild poliovirus strains on the territory of Russia and CIS countries in 1987-1995. *Zhurnal Mikrobiologii, Epidemiologii i Immunobiologii* **0**, 70-74.

Longini, I. M., Jr., Datta, S., and Halloran, M. E. (1996). Measuring vaccine efficacy for both susceptibility to infection and reduction in infectiousness for prophylactic HIV-1 vaccines. *Journal of Acquired Immune Deficiency Syndromes* **13**, 440-447.

Lui, K. J. (1998). Interval estimation of the risk ratio between a secondary infection, given a primary infection and the primary infection. *Biometrics* **54**, 706-711.

MacArthur, R. H. and Wilson, E. O. (1967). *The Theory of Island Biogeography.* Princeton: Princeton University Press.

Maris, R. W. (1997). Social and familial risk factors in suicidal behavior. *Psychiatric Clinics of North America* **20**, 519-550.

Martens, W. J. (1998). Health impacts of climate change and ozone depletion: an ecoepidemiologic modeling approach. *Environmental Health Perspectives* **106**, 241-251.

Massad, E. (1987). Transmission rates and the evolution of pathogenicity. *Evolution* **41**, 1127-1130.

Medley, G. F., Anderson, R. M., Cox, D. R., and Billard, L. (1987). Incubation period of AIDS in patients infected via blood transfusion. *Nature* **328**, 719-721.

Michelson, K. N., Thomas, J. C., Boyd, C., and Janssens, A. H. (1999). Chlamydia trachomatis infection in a rural population: The importance of screening men. *International Journal of STD & AIDS* **10**, 32-37.

Misra, S. N., Sengupta, D., and Satpathy, S. K. (1998). AIDS in India: Recent trends in opportunistic infections. *Southeast Asian Journal of Tropical Medicine & Public Health* **29**, 373-376.

Moller, A. P., Sorci, G., and Erritzoe, J. (1998). Sexual dimorphism in immune defense. *American Naturalist* **152**, 605-619.

Mollison, D. (ed.) (1995). *Epidemic models: Their Structure and Relation to Data*. Cambridge: Cambridge University Press.

Moran, P. (1999). The epidemiology of antisocial personality disorder. *Social Psychiatry & Psychiatric Epidemiology* **34**, 231-242.

Mulder, R. T., Beautrais, A. L., Joyce, P. R., and Fergusson, D. M. (1998). Relationship between dissociation, childhood sexual abuse, childhood physical abuse, and mental illness in a general population sample. *American Journal of Psychiatry* **155**, 806-811.

Murray, C. J. and Salomon, J. A. (1998). Modeling the impact of global tuberculosis control strategies. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1381-1386.

Murray, J. F. (1998). Tuberculosis and HIV infection: A global perspective. *Respiration* **65**, 335-342.

Nakasone, T. (1999). Tuberculosis among health care workers in Okinawa Prefecture. *Kekkaku* **74**, 389-395.

Nunn, P., Mungai, M., Nyamwaya, J., Gicheha, C., Brindle, R. J., Dunn, D. T., Githui, W., Were, J. O., and McAdam, K. P. (1994). The effect of human immunodeficiency virus type-1 on the infectiousness of tuberculosis. *Tubercle & Lung Disease* **75**, 25-32.

Olendorf, D., Jeryan, C., and Boyden, K. (eds). (1999). *Gale encyclopedia of medicine*. Michigan: Gale Research Company.

Operskalski, E. A., Stram, D. O., Busch, M. P., Huang, W., Harris, M., Dietrich, S. L., Schiff, E. R., Donegan, E., and Mosley, J. W. (1997). Role of viral load in heterosexual transmission of human immunodeficiency virus type 1 by blood transfusion recipients. Transfusion Safety Study Group. *American Journal of Epidemiology* **146**, 655-661.

Petersen, L. R., Satten, G. A., Dodd, R., Busch, M., Kleinman, S., Grindon, A., and Lenes, B. (1994). Duration of time from onset of human immunodeficiency virus type 1 infectiousness to development of detectable antibody. *Transfusion* **34**, 283-289.

Pierre, J. A. and Fournier, A. M. (1999). Human immunodeficiency virus infection in Haiti. *Journal of the National Medical Association* **91**, 165-170.

Platt, J. R. (1964). Strong inference. *Science* **146**, 347-353.

Popper, K. R. (1972). *The Logic of Scientific Discovery*. London: Huchinson.

Pumpuni, C. B., Mendis, C., and Beier, J. C. (1997). *Plasmodium yoelii* sporozoite infectivity varies as a function of sporozoite loads in *Anopheles stephensi* mosquitoes. *Journal of Parasitology* **83**, 652-655.

Rae-Grant, N., McConville, B. J., Fleck, S., Kennedy, J. S., Vaughan, W. T., Jr., Steiner, H., Bernard, V., Deci, P., MacMillan, H., Makini, G., Jr., Offord, D. R., Silverman, M. M., and Lynelle, T. (1999). Violent behavior in children and youth: Preventive intervention from a psychiatric perspective. *Journal of the American Academy of Child & Adolescent Psychiatry* **38**, 235-241.

Read, A. F., Albon, S. D., Antonovics, J., Apanius, V., Dwyer, G., Holt, R. D., Judson, O., Lively, C. M., Martin-Lof, A., McLean, A. R., Metz, J. A. J., Schmid-Hempel, P., Thrall, P. H., Via, S., and Wilson, K. (1995). Group report: Genetics and evolution of infectious disease in natural populations. In *Ecology of Infectious Disease in Natural Populations*, B. T. Grenfell and A. P. Dobson (eds), 450-477. Cambridge: Cambridge University Press.

Reifman, A., Barnes, G. M., Dintcheff, B. A., Farrell, M. P., and Uhteg, L. (1998). Parental and peer influences on the onset of heavier drinking among adolescents. *Journal of Studies on Alcohol* **59**, 311-317.

Rhodes, C. J. and Anderson, R. M. (1996). Persistence and dynamics in lattice models of epidemic spread. *Journal of Theoretical Biology* **180**, 125-133.

Ricklefs, R. E. and Gary, L. M. (1999). *Ecology*, 4th edition. New York: W. H. Freeman & Co.

Rida, W. N. (1996). Assessing the effect of HIV vaccination on infectiousness. *Statistics in Medicine* **15**, 2393-2412.

Rida, W., Fast, P., Hoff, R., and Fleming, T. (1997). Intermediate-size trials for the evaluation of HIV vaccine candidates: a workshop summary. *Journal of Acquired Immune Deficiency Syndromes* **16**, 195-203.

Ridzon, R., Kent, J. H., Valway, S., Weismuller, P., Maxwell, R., Elcock, M., Meador, J., Royce, S., Shefer, A., Smith, P., Woodley, C., and Onorato, I. (1997). Outbreak of drug-resistant tuberculosis with second-generation transmission in a high school in California. *Journal of Pediatrics* **131**, 863-868.

Rieder, H. L. (1995). Epidemiology of tuberculosis in Europe. *European Respiratory Journal - Supplement* **20**, S620-S632.

Roberts, R. B., Tennenberg, A. M., Eisner, W., Hargrave, J., Drusin, L. M., Yurt, R., and Krieswirth, B. N. (1998). Outbreak in a N.Y. city teaching hospital burn center caused by the Iberian epidemic clone of MRSA. *Microbial Drug Resistance-Mechanisms Epidemiology & Disease* **4**, 175-183.

Roscoe, D. E., Holste, W. C., Sorhage, F. E., Campbell, C., Niezgoda, M., Buchannan, R., Diehl, D., Niu, H. S., and Rupprecht, C. E. (1998). Efficacy of an oral vaccinia-rabies glycoprotein recombinant vaccine in controlling epidemic raccoon rabies in New Jersey. *Journal of Wildlife Diseases* **34**, 752-763.

Rowe, D. C. and Rodgers, J. L. (1994). A social contagion model of adolescent sexual behavior: explaining race differences. *Social Biology* **41**, 1-18

Sattenspiel, L. and Dietz, K. (1995). A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences* **128**, 71-91.

Sehgal, S. (1998). HIV epidemic in Punjab, India: Time trends over a decade. *Bulletin of the World Health Organization* **76**, 509-513.

Semiotrochev, V. L., Khakimov, M. M., and Shoumarov, S. B. (1998). General regularities of epidemic manifestation of cholera in Central Asia Republics. *Uzbekiston Tibbiet Zhurnali* **0**, 74-75.

Shaw, V. N., Hser, Y., Anglin, M. D., and Boyle, K. (1999). Sequences of powder cocaine and crack use among arrestees in Los Angeles county. *American Journal of Drug & Alcohol Abuse* **25**, 47-66.

Shiboski, S. C. and Padian, N. S. (1998). Epidemiologic evidence for time variation in HIV infectivity. *Journal of Acquired Immune Deficiency Syndromes* **19**, 527-535.

Siegmund, K. D., Province, M. A., Higgins, M., Williams, R. R., Keller, J., and Todorov, A. A. (1998). Modeling disease incidence rates in families. *Epidemiology* **9**, 557-562.

Smith, D. R., Fedorka-Cray, P. J., Mohan, R., Brock, K. V., Wittum, T. E., Morley, P. S., Hoblet, K. H., and Saif, L. J. (1998). Epidemiologic herd-level assessment of causative agents and risk factors for winter dysentery in dairy cattle. *American Journal of Veterinary Research* **59**, 994-1001.

Stigum, H., Falck, W., and Magnus, P. (1994). The core group revisited: the effect of partner mixing and migration on the spread of gonorrhea, Chlamydia, and HIV. *Mathematical Biosciences* **120**, 1-23.

Stoolmiller, M. (1998). Comment on "Social contagion, adolescent sexual behavior, and pregnancy: a nonlinear dynamic EMOSA model." *Developmental Psychology* **34**, 1114-1118.

Swinton, J. (1998). Extinction times and phase transitions for spatially structured closed epidemics. *Bulletin of Mathematical Biology* **60**, 215-230.

Taylor, D. R., Jarosz, A. M., Lenski, R. E., and Fulbright, D. W. (1998). The acquisition of hypovirulence in host-pathogen systems with three trophic levels. *American Naturalist* **151**, 343-355.

Thieme, H. R. (1980). Some mathematical considerations of how to stop the spatial spread of a rabies epidemic. *Lecture Notes in Biomathematics* **38**, 310-319.

Tiedemann, A. and Firsching, K. H. (1998). Combined whole-season effects of elevated ozone and carbon dioxide concentrations on a simulated wheat leaf rust (*Puccinia recondita* f. sp. *tritici*) epidemic. *Zeitschrift Fuer Pflanzenkrankheiten und Pflanzenschutz* **105**, 555-566.

Topley, J. M., Bamber, S., Coovadia, H. M., and Corr, P. D. (1998). Tuberculous meningitis and co-infection with HIV. *Annals of Tropical Paediatrics* **18**, 261-266.

Toyota, M. (1994). Validity of "the risk index" to predict the infectiousness of tuberculosis patients. *Kekkaku* **69**, 375-377.

Vernazza, P. L., Eron, J. J., Fiscus, S. A., and Cohen, M. S. (1999). Sexual transmission of HIV: infectiousness and prevention. *AIDS* **13**, 155-166.

Veugelers, P. J., Cornelisse, P. G., Craib, K. J., Marion, S. A., Hogg, R. S., Strathdee, S. A., Montaner, J. S., O'Shaughnessy, M. V., and Schechter, M. T. (1998). Models of survival in HIV infection and their use in the quantification of treatment benefits. *American Journal of Epidemiology* **148**, 487-496.

Vila, J., Ruiz, J., Navia, M., Becerril, B., Garcia, I., Perea, S., Lopez-Hernandez, I., Alamo, I., Ballester, F., Planes, A. M., Martinez-Beltran, J., and Jimenez, D. A. T. (1999). Spread of amikacin resistance in *Acinetobacter baumannii* strains isolated in Spain due to an epidemic strain. *Journal of Clinical Microbiology* **37**, 758-761.

Villalbi, J. R., Galdos-Tanguis, H., Cayla, J. A., Casanas, P., Ferrer, A., and Nebot, M. (1999). Tuberculosis infection and disease among schoolchildren: The influence of the HIV epidemic and of other factors. *Journal of Epidemiology & Community Health* **53**, 112-117.

Vitek, C. R., Brisgalov, S. P., Bragina, V. Y., Zhilyakov, A. M., Bisgard, K. M., Brennan, M., Kravtsova, O. N., Lushniak, B. D., Lyerla, R., Markina, S. S., and Strebel, P. M. (1999). Epidemiology of epidemic diphtheria in three regions, Russia, 1994-1996. *European Journal of Epidemiology* **15**, 75-83.

Whittaker, R. J. (1998). *Island Biogeography: Ecology, Evolution, and Conservation.* Oxford: Oxford University Press.

Whittle, H. C., Aaby, P., Samb, B., Jensen, H., Bennett, J., and Simondon, F. (1999). Effect of subclinical infection on maintaining immunity against measles in vaccinated children in West Africa. *Lancet (North American Edition)* **353**, 98-102.

Wu, H. and Ding, A. A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410-418.

Xu, X. (1999). Effects of temperature on the length of the incubation period of rose powdery mildew (*Sphaerotheca pannosa* var. *rosae*). *European Journal of Plant Pathology* **105**, 13-21.

Yan, G., Stevens, L., Goodnight, C. J., and Schall, J. J. (1998). Effects of a tapeworm parasite on the competition of Tribolium beetles. *Ecology* **79**, 1093-1103.

Yang, H. M. and Silveira, A. S. (1998). The loss of immunity in directly transmitted infections modeling: effects on the epidemiological parameters. *Bulletin of Mathematical Biology* **60**, 355-372.

Zhang, Y., Ma, J., Zhao, S., et al. (1998). Studies on immunization strategy against hepatitis A in a highly prevalent area. *Zhonghua Yufang Yixue Zazhi* **32**, 295-296.

Zule, W. A. and Desmond, D. P. (1999). An Ethnographic comparison of HIV risk behaviors among heroin and methamphetamine injectors. *American Journal of Drug & Alcohol Abuse* **25**, 1-23.

Table 4.1

Ten examples from the current literature (1/1/98 - 4/30/99) of differential susceptibility to infectious disease at both the individual and population level (human hosts, unless otherwise indicated).

Disease	Comments and references
Chlamydia trachomatis	Susceptibility varies by gender: men are more susceptible (Michelson et al., 1999).
Dengue	(1) Susceptibility varies by age and previous exposure to different strains: previous exposure can *increase* susceptibility (Deparis et al., 1998; Ferguson et al., 1999). (2) Individuals with more knowledgeable about dengue may be less susceptible (Gupta et al., 1998).
Diphtheria	Susceptibility varies by age and gender: young, middle aged, and women are more susceptible (Vitek et al., 1999).
Ebola	Susceptibility varies by cultural and socio-economic factors (De Roo et al., 1998; Georges et al., 1999): willingness to care for ill family members increases susceptibility (De Roo et al., 1998).
Hepatitis A	(1) Susceptibility may vary by and within age classes: children of immune mothers show varying degrees of immunity for first 13 months (Zhang et al., 1998). (2) Susceptibility varies by age, gender, drug use, socioeconomic status, and sexual preference: gay men, IV drug users, disadvantaged youth, and males are more susceptible and very young children are less susceptible (Ferson et al., 1998).
HIV	(1) IV drug users are more susceptible (Donisi et al, 1998; Haverkos, 1998; Celentano et al., 1999): heroin users more susceptible than methamphetamine users (Zule and Desmond, 1999); susceptibility changing as drug use patterns change (Gibney et al., 1999). (2) Susceptibility varies by age, ethnicity, marital status, occupation, and education (Celentano et al., 1999): older individuals are more susceptible to horizontal transmission (Donisi et al., 1998); female prostitutes are more susceptible (Rae-Grant et al., 1999); single illiterate woman are more susceptible (Kilian et al., 1999). (3) Strong evidence that risk factors are changing: heterosexuals becoming more susceptible (Donisi et al., 1998).
Q fever	Susceptibility increases given current infection with HIV and previous exposure (Boschini et al., 1999).
Tuberculosis	(1) Susceptibility increases given current infection with HIV (Misra et al., 1998; Murray, 1998; Topley et al., 1998; Villalbi et al., 1999). (2) Susceptibility varies by occupation: gold miners are more susceptible to tuberculosis and other pulmonary diseases (Corbet et al., 1999); nurses are more susceptible (relative risk, 2.3; Nakasone, 1999).
West Nile fever	Susceptibility varies by living conditions: mosquitoes in the home and having a flooded basement increases susceptibility (Han et al., 1999).
Winter dysentery (in cattle)	Susceptibility of a population to outbreak varies among populations according to husbandry practices (Smith et al., 1998).

Table 4.2

Six examples from the recent literature (1/1/90 - 4/30/99) of differential infectiousness at both the individual and population level (human hosts, unless otherwise indicated).

Pathogen	Comments and reference
HIV	(1) Infectiousness varies according to behavior: "...a small subset of infected people may be responsible for a disproportionate number of infections." (Hyman et al., 1999); anal sex is more infectious (Lazzrin et al., 1991). (2) Infectiousness varies by case: "Although this series was small, the prospective observations suggest that viral load was the only characteristic in the recipient that contributed to heterosexual infectiousness." (Operskalski et al., 1997); infectiousness varies over time by case (Shiboski and Padian, 1998; Hyman and Stanley, 1999). (3) Evolving educational programs are reducing infectiousness (Cates et al., 1997; Kilian et al, 1999). (4) Infectiousness varies by gender: men are more infectious and vasectomy may reduce infectiousness (Kreiger et al., 1998; Vernazza et al., 1999). (5) Evolving treatment methods may be reducing infectiousness (Veugelers et al, 1998). (6) Undiagnosed cases are more infectious (Cates et al., 1997): window period between initial infectiousness and detectability (i.e. patient/blood donor becomes seropostive) averages 45 days, but some remain infectious and seronegative for 6 months (Peterson et al., 1994). (7) Vaccination breakthrough might lead to reduced infectiousness relative to normal course of infection (Longini et al., 1996; Rida, 1996, Rida et al., 1997; Datta et al., 1998). (8) "Available data leave little doubt that other STDs facilitate HIV transmission through direct, biological mechanisms..." (Fleming and Wasserheit, 1999).
Human papillomavirus	Infectiousness affected by secondary conditions: HIV-induced immunosuppression increases the severity, duration and infectiousness of anogenital warts (Judson, 1992).
Measles	Infectiousness may vary according to previous exposure: "Vaccinated index cases may have been less infectious than unvaccinated index cases, since they produced fewer clinical cases among exposed children (relative risk = 0.55, 95% CI 0.29-1.04)." (Cisse et al., 1999).
Plasmodium yoelii (in mice)	"When *Plasmodium yoelii*-infected *Anopheles stephensi* mosquitoes were each allowed to feed on a single mouse, we noted that sporozoites from mosquitoes with higher sporozoite loads were more infectious in 13 of 30 (43%) mice." (Pumpuni, et al., 1997).
Schistosomiasis	Some infectious water sources are more infectious than others (Etard and Borel, 1992).
Tuberculosis	(1) Infectiousness varies from case to case (Toyota, 1994; Rieder, 1995; Ridzon et al., 1997). (2) Infectiousness varies as cases progress: "Infectiousness increased significantly in the last of three semesters during which the source case was symptomatic." (Braden, 1995). (3) Infectiousness may vary due to secondary infections: HIV may reduce infectiousness (Elliot et al., 1993; Nunn et al., 1994).

Table 4.3

Ten examples from the current literature (1/1/98 - 4/30/99) of the spatial spread, both within and among populations, of infectious disease (human hosts, unless otherwise indicated).

Pathogen	Comments and reference
Cholera	(1) "Since 1991, a large recrudescence of cholera epidemics occurred, both in the continents where cholera is endemic, and in traditionally cholera-free areas." (Grassi et al., 1998). (2) Migrants from endemic countries caused out-breaks in Tadjikistan, Uzbekistan and Kazakhstan in 1993-94 (Semiotrochev et al., 1994).
Dengue	(1) Dengue has spread among Japanese main islands (Hotta, 1999). (2) Intra-familial transmission is different from inter-familial transmission: "Analysis of dengue spread in Teroma (French Polynesia) confirmed that dengue transmission occurs primarily in the house, thus vector control campaigns should incorporate focal insecticide spraying and systematic daily use of insecticide in houses." (Deparis et al., 1998).
Green mold (in commercial mushrooms)	Within populations: "Spatial analysis revealed that green mold foci were more likely to occur in neighboring sections along the beds rather than above, below, or across from each other." (Rosye et al., 1999).
HIV	Among populations: spread from urban to rural areas in Haiti (Pierre and Fournier, 1999); "Although Bombay (Mumbai) appears to be the main focus for acquired immunodeficiency syndrome (AIDS) in India, rapid spread has occurred through other major cities as well." (Sehgal, 1998; also, Misra et al., 1998).
Influenza	"This analysis indicates that diffusion over long distances, possibly due to global transportation systems, is so quick that homogeneous global mixing occurs before the epidemic builds up within infected patches." (Bonabeau et al., 1998).
Measles	Intra-familial transmission is different from inter-familial transmission: "The frequency of measles and subclinical measles - defined as a four-fold or greater rise in antibody titre without clinical signs or symptoms - was related to intensity of exposure according to whether the index case was in the same hut, household, or compound." (Whittle et al., 1999).
Methicillin-resistant *Staphylococcus aureus*	"This outbreak is the first known report of the Iberian MRSA clone in the United States." (Roberts et al., 1998).
Penicillin-resistant *Streptococcus pneumoniae*	"The global spread of multidrug-resistant Streptococcus pneumoniae clones is well documented in the literature." (Castaneda et al, 1998).
Poliovirus	33 strains of wild type poliovirus have spread across Russia and the CIS (Lipskaya et al., 1998).
Rabies (in racoons)	"The maximum linear movement (12.9 km) among five ear-tagged rabid raccoons in the study area was significantly greater than that of 19 normal radio-collared raccoons (2.58 km) in the area." (Diehl et al., 1998).

Table 4.4

Four examples from the current literature (1/1/98-4/30/99) of the effects of climate, weather, land use changes, and pollution on epidemic dynamics (human hosts, unless otherwise indicated).

Pathogen	Comments and reference (models: Martens, 1998)
Japanese encephalitis	"The population of the vector mosquito, *Culex tritaeniorhynchus*, has decreased in rice fields where new methods for rice cultivation have been applied. Nowadays, rice fields have small amounts of water due to technology which controls the management of intermittent water supply, new varieties of rice have been developed to shorten the period of cultivation, and pesticides are extensively used., ..., Moreover, the global warming, the changes in environment and emergence of insecticide-resistant mosquitoes increase the risk of epidemic in Japan and other Asian countries at any time." (Kamimura, 1998).
Saint Louis encephalitis (SLE)	"Epizootic and epidemic transmission of SLE to sentinel chickens and humans in Indian River County (*Florida*) was greatest immediately following heavy rainfalls that synchronized *Culex nigripalpus* oviposition (*the vector*) and blood feeding." (Day and Curtis, 1999).
Water borne pathogenic bacteria	"The industrial pollution of a water reservoir (the Volga), ... The total amount of bacteria increases simultaneously with the decrease of their specific diversity, ..., which may lead to the growth of the epidemic danger of the polluted areas of the water reservoir." (Boiko and Pogorelova, 1998).
Wheat leaf rust (in wheat plants)	Speaking of experimental results: "Ozone significantly reduced disease severity, uredospore production and increased the latent period of leaf rust on young plants, consequently inhibiting the epidemic spread on upper leaves of mature plants." (Tiedemann and Firsching, 1998).

Table 4.5

Fifteen examples of phenomenon that propagate like infectious diseases.

Phenomenon	Typical route of infection and comments on pathogenisis.
Alcoholism	(1) Three routes: by example from parent to child (vertical transmission); by example among peers (horizontal transmission); and, through advertising (point sources of infection). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., Duncan et al., 1998b; Reifman et al., 1998).
Cancer	(1) Not communicable but source of causative carcinogens is analogous to a point source of infection. (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., HIV victims more susceptible to some cancers: Cook-Mozaffari et al., 1998; farmers are more susceptible to some cancers: Acquavella et al., 1998).
Computer viruses	(1) Computer files are infectious; inanimate vectors are computer disks and email. (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply- users who use anti-viral programs are less susceptible.
Crime	(1) Compulsion to commit crimes may be infectious among peers. (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (general: Moran, 1999; shoplifting: James et al., 1979).
Domestic violence, child abuse	Possible route: by example from parent to child (vertical transmission). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., Dearwater et al., 1998; Mulder et al., 1998).
Drug abuse	(1) Two routes: by example from parent to child (vertical transmission), and by example among peers (horizontal transmission). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., Brooks et al., 1998; Shaw et al., 1999). (3) At least 2 examples of the application of epidemic models to the propagation of drug abuse are in the literature (Egan and Robinson, 1979; Hoppensteadt and Murray, 1981).
Eating disorders (bulimia, anarexia)	(1) Two routes: by example from parent to child (vertical transmission), and by example among peers (horizontal transmission). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (by race: le Gange et al., 1998; by gender: Lindeman, 1999).
Mental illness	(1) Some personality disorders may transmit vertically through mechanisms of behavior and genetics (e.g., phobias: King et al., 1998). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (by upbringing: Mulder et al., 1998).

table continued...

Table 4.5 *(continued)*
Examples of phenomenon that propagate like infectious diseases.

Phenomenon	Typical route of infection and comments on pathogenisis.
Obesity	(1) In some cases: by example from parent to child. (2) Concepts of differential susceptibility apply (e.g., Goran et al., 1998).
Poisoning	Not communicable but cases contracted from point sources. (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., children are more susceptible to lead poisoning: Goyer, 1996; Koike, 1999).
Product specific accidents	(1) Products themselves transmit to consumers (hosts) via advertising (point sources) and word-of-mouth (person-person); with the product comes the chance of injury (an additional point source risk). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., bicycle riders who wear helmets - use varies by age - are less likely to suffer injury: Bolen et al., 1998).
Religions, philosophies, values, ideas, fads, jokes, rumors	(1) Three routes: by example from parent to child (vertical transmission); by example among peers (horizontal transmission); and, through advertising (point sources). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply. (2) At least 2 examples of the application of epidemic models to the transmission of ideas are in the literature: ideas (Goffman and Newill, 1964), and rumors (Dietz, 1967).
Smoking	(1) Three routes: by example from parent to child (vertical transmission); by example among peers (horizontal transmission); and, through advertising (point sources of infection). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., Duncan et al., 1998a; Jackson et al., 1998).
Suicide	(1) In some cases compulsion to commit suicide may be infectious: suicide clusters (Cox and Skegg, 1993; Hazell, 1993). (2) Concepts of heterogeneity, susceptibility, incubation, and infectiousness apply (e.g., Marsi, 1997; Caces and Harford, 1998).
Teen pregnancy; adolescent sexual behavior	(1) Age of first intercourse may be affected by peer pressure. (2) Concepts of heterogeneity, susceptibility, incubation apply (e.g., Rowe and Rodgers, 1994). (3) At least 2 epidemic models exist in the literature (Rowe and Rodgers, 1994; Stoolmiller, 1998).

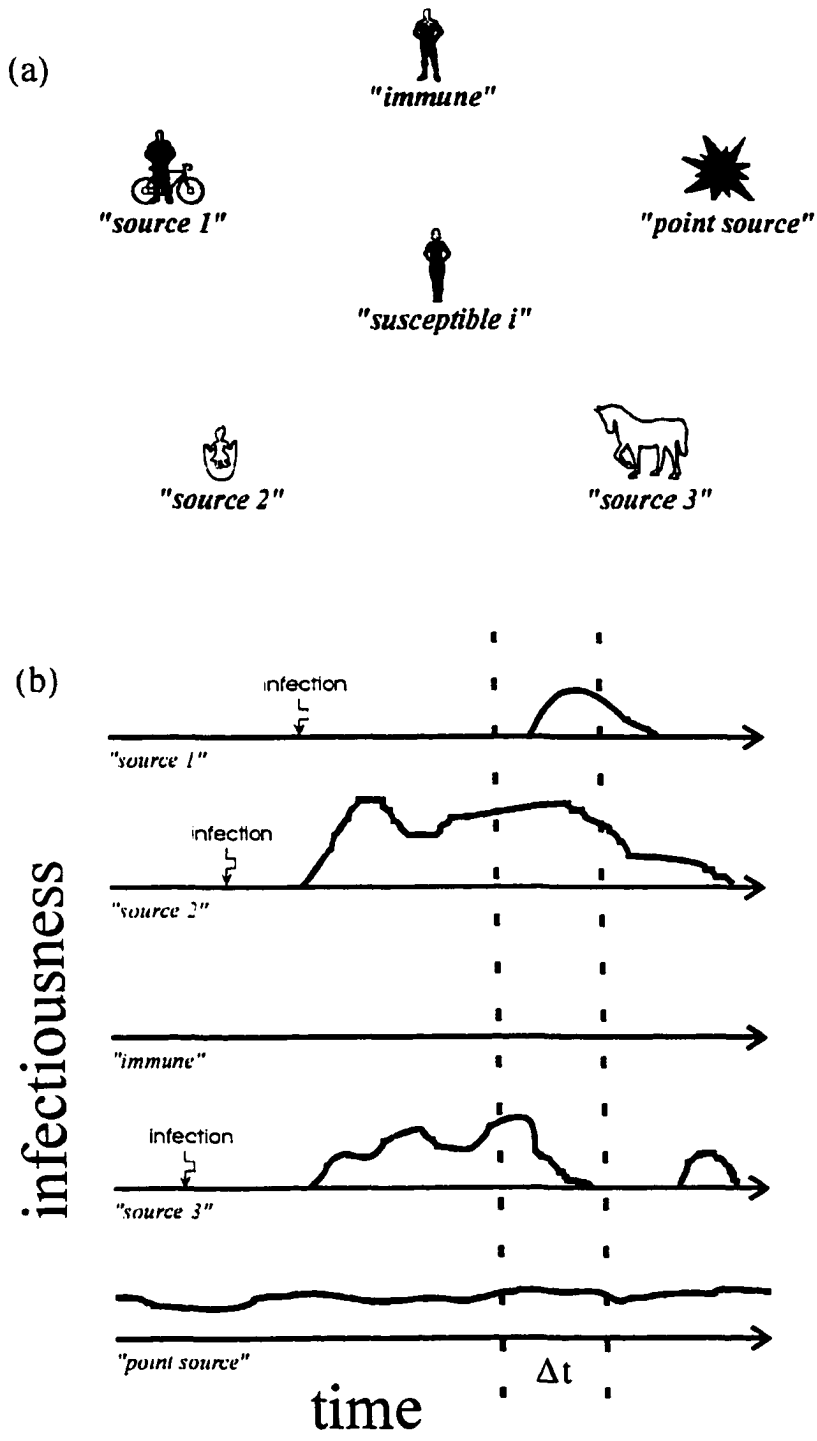


Figure 4.1. Graphic depiction of the complex interaction of heterogeneity, susceptibility, incubation, infectiousness, and transmission. (a) The risk of infection experienced by "*susceptible i*" during Δt is a complex function of her biological susceptibility during Δt, her interaction with each member of her environment during Δt, and (b) the infectiousness of each source in her environment during Δt.