Editorial Manager(tm) for PLoS ONE
Manuscript Draft

Manuscript Number:

Title: Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model.

Short Title: Intrinsic vs extrinsic causes of aggregation

Article Type: Research Article

Section/Category: Other

Keywords: Ixodes scapularis; Lyme disease; Negative binomial distribution; Parasite overdispersion; Peromyscus leucopus; Poisson-gamma mixture

Corresponding Author: Justin M. Calabrese

Corresponding Author's Institution: Smithsonian Conservation Biology Institute

First Author: Justin M. Calabrese

Order of Authors: Justin M. Calabrese;Justin M. Calabrese;Jesse L. Brunner;Richard S. Ostfeld

Abstract: It is well known that parasites are often highly aggregated on their hosts such that relatively few individuals host the large majority of parasites. When the parasites are vectors of infectious disease, a key consequence of this aggregation can be increased disease transmission rates. The cause of this aggregation, however, is much less clear, especially for parasites such as arthropod vectors, which generally spend only a short time on their hosts. The Poisson-gamma mixture distribution has often been used to describe aggregated parasite distributions, but lacks a clear mechanistic basis. Here, we extend this framework by linking it to a general model of parasite accumulation. Then, using blacklegged ticks (Ixodes scapularis) on mice (Peromyscus leucopus) as an empirical example, we fit the extended model to the best currently available larval tick burden datasets using a hierarchical Bayesian approach. Our results suggest that simple bad luck--inhabiting a home range with high vector density--may play a much larger role in determining parasite burdens than is currently appreciated. Along the way, we use our framework to highlight gaps in available datasets and to suggest strategies to fill them.

Suggested Reviewers: Sarah Perkins
Cardiff University
PerkinsS@cardiff.ac.uk

Howard Ginsberg
USGS Patuxent Wildlife Research Center
hginsberg@usgs.gov

Opposed Reviewers:

Dear PLoS ONE,

Attached please find, for consideration for publication in PLoS ONE, our manuscript entitled "Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model."

Aggregation of macroparasites on hosts is frequently observed but poorly understood. Despite equivocal empirical support, current dogma holds that variation in intrinsic host susceptibility drives the observed concentration of parasites on relatively few hosts. Most studies analyzing parasite burden data use phenomenological statistical models that are not capable of describing the mechanisms underlying observed patterns.

We introduce a new approach, based on a simple and general mechanistic model of parasite accumulation, that partitions the degree of parasite aggregation into components due to intrinsic variation in host susceptibility and to extrinsic variation in the density of parasites experienced by hosts. We couple our model with available data on blacklegged tick (*Ixodes scapularis*) burdens on white-footed mice (*Peromyscus leucopus*) to demonstrate that simple bad luck—living in areas with high parasite density—can play a much larger role in determining parasite burdens than is currently appreciated.

Our results are important because they have implications for efforts to control the spread of vector-borne diseases such as Lyme disease. Targeting the most heavily infested hosts can have a much larger effect on disease spread rate relative to the same control effort applied to randomly selected hosts. The success of this strategy hinges on easy and reliable identification of the hosts that will tend to accumulate the most parasites. If highly unpredictable extrinsic factors such as random variation in local parasite density exert a strong effect on parasite burdens, control efforts that assume reliable identification of key host individuals is possible may be doomed to fail and may waste limited disease management resources.

Though our empirical example focuses on blacklegged ticks and white-footed mice, the framework we establish is general and could be used to understand the factors driving parasite aggregation in other host-parasite systems. The accumulation model on which it is built can easily be tailored to accommodate system-specific details, and to exploit available data sources.

We hereby state that this work has not been published or accepted for publication, and is not under consideration for publication in another journal or book. Its submission for publication has been approved by all authors and relevant institutions, and all persons entitled to authorship have been so named.

Sincerely,

Justin Calabrese

Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model.

Justin M. Calabrese[1,2,*], Jesse L. Brunner[3,4], Richard S. Ostfeld[3]
1 Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Rd., Front Royal, VA 22630, USA
2 Helmholtz Centre for Environmental Research-UFZ, Permoserstrasse 15, 04318 Leipzig, Germany
3 Cary Institute of Ecosystem Studies, Millbrook, NY 12545, USA
4 School of Biological Sciences, Washington State University, Pullman, WA 99164, USA
∗ **E-mail: CalabreseJ@si.edu**

Abstract

It is well known that parasites are often highly aggregated on their hosts such that relatively few individuals host the large majority of parasites. When the parasites are vectors of infectious disease, a key consequence of this aggregation can be increased disease transmission rates. The cause of this aggregation, however, is much less clear, especially for parasites such as arthropod vectors, which generally spend only a short time on their hosts. The Poisson-gamma mixture distribution has often been used to describe aggregated parasite distributions, but lacks a clear mechanistic basis. Here, we extend this framework by linking it to a general model of parasite accumulation. Then, using blacklegged ticks (*Ixodes scapularis*) on mice (*Peromyscus leucopus*) as an empirical example, we fit the extended model to the best currently available larval tick burden datasets using a hierarchical Bayesian approach. Our results suggest that simple bad luck–inhabiting a home range with high vector density–may play a much larger role in determining parasite burdens than is currently appreciated. Along the way, we use our framework to highlight gaps in available datasets and to suggest strategies to fill them.

Introduction

Parasites, from nematodes and trematodes to lice and ticks, are frequently observed to be highly aggregated on their hosts with relatively few individuals hosting the large majority of parasites [1, 2]. Indeed, the distribution of parasite burdens among hosts is usually described by a negative binomial distribution (NBD) with its characteristic long right tail representing those few highly infected hosts [2]. When the parasites are vectors of infectious disease, the aggregation of vectors can inflate the potential rate of spread of an infection [3, 4]. One widely-cited example is tick-borne encephalitis (TBE), which is caused by a virus transmitted between *Ixodes ricinus* ticks when they co-feed on hosts such as yellow-necked mice, *Apodemus flavicollis* [5, 6]. Most TBE transmission occurs on the hosts with the greatest tick burdens [7]. If public health interventions could target the most infested 20% of hosts, transmission of TBE to humans could be effectively reduced by 75% [7], but similar interventions targeted at random hosts could be expected to have only negligible impact [3, 4]. Thus identifying those hosts responsible for feeding and infecting the most vectors has become a priority.

Empirical studies are often strongly influenced by available analytical tools. Currently two classes of models are applied to study parasite burdens. The first consists of regression-based approaches usually employing generalized linear models (or extensions thereof) with negative binomial error structure. Most studies using this approach focus on identifying covariates that account for variation in mean burdens among groups of hosts, and treat k, the overdisperison parameter of the NBD, as a nuisance parameter [7–9]. There is typically no link between biological processes and degrees of overdispersion (though it is possible to model k as function of other variables; [10, 11]. Still, if the overdispersion comes from lumping

together very different groups of hosts, then accounting for these differences should logically reduce the degree of dispersion in the model (i.e., k should increase) [12]. In fact there is usually substantial overdispersion left even after the key factors are included in the models [8, 9].

The second approach invokes a mixture distribution, where it is assumed that hosts randomly sample parasites from their environment, which would result in a Poisson distribution of burdens, but the sampling rate (i.e., the expected burden) varies among hosts [4, 13]. When the sampling rate of the Poisson is gamma-distributed, the marginal distribution of burdens is negative binomial [4, 14]. Thus it is variation in the sampling rate among hosts that causes overdispersion. In contrast to the regression-based approach, the Poisson-gamma mixture directly results in a NBD of burdens (indeed, it is one way to derive the NBD) and the aggregation parameter, k, can be expressed in terms of the parameters of the sampling rate distribution. The disadvantage of this approach is that the variation in sampling rates that drives aggregation is typically unobserved, and thus the causes of this variation are not identified.

Empirical studies of tick burdens on hosts have tended to use the first model category and have focused heavily on identifying intrinsic host characteristics that explain observed tick burdens. The rationale is that a priori identification of the hosts most likely to have high burdens could lead to targeted and highly effective control efforts. Unfortunately, these studies have, so far, produced equivocal results, with few consistent factors emerging across different studies, systems, sites, and years. It is often observed that males have greater burdens than females (e.g., on *A. flavicollis*, [7]), but two recent studies have shown that sex is just one of myriad host and environmental variables that each explain a small portion of the variability in burdens on several rodent species [8, 9]. Even a study that explicitly linked individual space use and activity/exploration phenotypes of Siberian chipmunks (*Tamias sibiricus*) to tick burdens also found that many variables, including their interactions, were significantly associated with burdens [15]. It is still not clear from these studies how much of the observed aggregation of tick burdens is due to variation in susceptibility among hosts, and how much is due to extrinsic factors such as variation in questing tick densities among host home ranges. If tick burdens are driven primarily by random, extrinsic factors, control efforts focusing on identifying the most susceptible hosts via host characteristics may be doomed to fail.

We have three main goals in this paper. First, we introduce a general framework for understanding the distribution of parasites on hosts. The framework consists of a simple and flexible mechanistic model of parasite accumulation and an explicit consideration of how variability among hosts enters into the parasite accumulation process. Our framework can be understood as an extension of the Poisson-gamma mixture model widely invoked to explain negative binomial parasite distributions. Second, we use a hierarchical Bayesian approach to couple our model to the best available data on black-legged ticks and white-footed mice to quantify the degree to which random variation in tick density among mouse home ranges affects the overdispersion of tick burdens on mice. Third, we use our framework to highlight existing data gaps and to suggest future studies that could address these gaps.

Methods

As mentioned above, the key weakness of using the Poisson-gamma mixture to model parasite burden distributions is that sampling rate variation is usually unobserved. Because of this, it is often not possible to identify the causes of aggregation with burden data alone. We seek here to extend the Poisson-gamma mixture framework by linking variation in the sampling rate among hosts to an underlying model of parasite accumulation. In other words, we aim to write the NBD and its aggregation parameter, k, in terms of an accumulation model and associated sources of variability among individuals. It will then be possible to use burden data in combination with other types of data that can speak directly to the accumulation process. In our empirical example, we focus on a key extrinsic factor, spatial variation in larval blacklegged tick (*I. scapularis*) density, to quantify its contribution to aggregation relative to that of intrinsic differential susceptibility among white-footed mice (*Peromyscus leucopus*) .

We first derive a simple model relating host movement and tick density to the expected burden, or sampling rate, λ, on an individual host. We assume that the realized tick burden on a host, B, is Poisson distributed with rate parameter λ. We then consider how variation in λ among hosts gives rise to an approximately negative binomial distribution of burdens.

A simple tick accumulation model

We assume that each host occupies a home range characterized by its area, A, and parasite density, D. Hosts move within their home ranges and encounter ticks as they do so. The the per day rate at which a host encounters and picks up parasites should then be proportional to the product of the distance it moves per day, M, and the density of questing larvae, D, in its home range.

We further assume that parasites are uniformly distributed within home ranges and that their density remains constant within each home range. This implies that the removal and feeding of parasites by hosts and by parasite mortality are insignificant compared to the number of questing parasites available in a home range, at least over a short period of time. This assumption is reasonable for our blacklegged tick/white-footed mouse example, given that densities of questing nymphs and larvae, as well as burdens on mice, remain high for several weeks during the peak of the season [8, 16].

Lastly, we assume that the parasites feeding on a host drop off after a successfully feeding or are removed (e.g., by grooming or host immune responses) at a constant rate, δ, independent of the density of ticks on the host (although it is possible to modify this assumption). The expected tick burden, λ, on a host at time t is therefore determined by the rates at which parasites are picked up and lost, or

$$\frac{d\lambda}{\mathrm{dt}} = \alpha MD - \delta\lambda. \tag{1}$$

Notice that the accumulation constant, α, could be broken into a number of individual constants including the width of area "sampled" by the host, the probability that given an encounter, the parasite attaches to the host, etc. These constants all enter as a product whose individual components are not separately estimable from the data at hand, and so we lump them together. If more detailed data were available, they could be kept separate.

Assuming that parasite burdens at the time of sampling have reached their equilibrium (again, for our example system, the seasonal "peaks" in tick burdens last for several weeks) [8], we focus on the stationary solution of equation (1):

$$\lambda^* = \frac{\alpha MD}{\delta}. \tag{2}$$

For notational convenience, we will refer hereafter to the equilibrium sampling rate simply as λ, dropping the $*$ superscript. Again, if data on host movement or the components of α are available, these factors could be kept separate, and the steps below could then be performed on this expanded model. Focusing on our empirical example, we assume that all factors in the model except host density are intrinsic to the individual, while variation in D is extrinsic to the individual. Defining $S = \alpha M/\delta$, collects all intrinsic components of the model into a single "susceptibility" factor.

The approximate distribution of tick burdens

Both S and D will vary among hosts and will thus be considered random variables. As both are continuous, must be non-negative, and could conceivably assume a range of different distributional shapes, we assume they are gamma distributed. We denote shape parameters of gamma distributions by η and scale parameters by ν. Thus, the gamma distribution of S is characterized by η_S and ν_S, while that of D is parameterized by η_D and ν_D.

When derived as a Poisson-gamma mixture, the probability mass function of the NBD can be written in terms of the parameters of the gamma distribution of sampling rates [17], yielding

$$\Pr\{B = b\} = \frac{\Gamma(b + \eta_\lambda)}{b!\,\Gamma(\eta_\lambda)} \left(\frac{1}{\nu_\lambda + 1}\right)^{\eta_\lambda} \left(\frac{\nu_\lambda}{\nu_\lambda + 1}\right)^{b} \tag{3}$$

where $\Gamma(\bullet)$ is the Euler gamma function, and η_λ and ν_λ are the shape and scale parameters, respectively, of the sampling rate distribution.

Equation (2) shows that λ is the product of S and D. In appendix A, we show via simulation that a gamma distribution provides a good approximation of the distribution of the product SD (Fig. A.1). We then use this fact to derive an approximation that allows us to write the parameters of the approximate gamma distribution of λ in terms of η_S, ν_S, η_D, and ν_D. The resulting approximate expressions for the shape and scale parameters of the rate distribution are (appendix A)

$$\eta_\lambda = \frac{\eta_D \eta_S}{\eta_D + \eta_S + 1} \tag{4}$$

and

$$\nu_\lambda = \nu_D \nu_S (\eta_D + \eta_S + 1). \tag{5}$$

Substituting equations (4) and (5) into equation (3), we obtain an approximation for the distribution of parasite burdens over hosts in terms of the accumulation model and associated sources of variability. The mean of the burden distribution is

$$\mu = \langle \lambda \rangle = \eta_D \eta_S \nu_D \nu_S \tag{6}$$

and the aggregation parameter of the burden distribution is $k = \eta_\lambda$. In other words, the degree of aggregation in tick burdens is determined by the shape parameter of the gamma distribution that describes how accumulation rates vary among individual hosts. The rate distribution shape parameter is, in turn, a function of η_S and η_D. Thus, this approximation links lower-level processes governing vector accumulation, which are potentially measurable, to the degree of aggregation in vector burdens on hosts.

Focusing on our blacklegged tick example, we can now develop an index that quantifies the contribution of variation in questing larval density among host home ranges, D, to the observed value of k. The limit of the expression for k as $\eta_D \to \infty$ (i.e., as the effect of variation in larval density goes away) is simply $k_S = \eta_S$. In this limit, equations (3), (4), (5), and (6) are exact (i.e., the rate parameter distribution is a gamma) and variation among individuals in sampling rate is driven entirely by differential susceptibility. The ratio k/k_S will then be < 1 as long as $\eta_D < \infty$, and the quantity $q = 1 - k/k_S$ is a measure of the degree to which the estimated value of η_D reduces the value of k conditional on the value of η_S. In other words, $q = 1$ when the aggregation of vectors on hosts is dominated by differences in vector densities among home ranges and is zero when the aggregation is due entirely to differences in individual susceptibility. Writing q in terms of η_D and η_S, we obtain

$$q = \frac{\eta_S + 1}{\eta_D + \eta_S + 1}. \tag{7}$$

Checking the limit behavior of q, we see that $q = 1$ when $\eta_D = 0$, and $q \to 0$ when $\eta_D \to \infty$, as expected.

Empirical data

We now show how the above-described extended Poisson-gamma mixture can be combined with empirical data to tease apart the contribution of extrinsic and intrinsic factors on the larval tick burdens of white-footed mice. We used two years of data (1999 and 2004) from two of the six grids (GC and TX) that

are part of the the long term records of white-footed mice from the Cary Institute of Ecosystem Studies (CIES) collected by R.S. Ostfeld and colleagues. We chose these years and grids because they offered the most observations of larval burdens and mouse home range sizes and densities of questing larvae, used to estimate D. We briefly describe their methods and data; a more detailed description can be found elsewhere [18].

We used two of six $2.25ha$ permanent small mammal trapping grids in the oak and maple dominated forests tracts of the CIES campus. Each consists of 242 Sherman traps arranged in pairs in a grid with $15m$ in between grid locations. Each grid was trapped for two consecutive nights six to eight times between April and October. Upon first capture in a trapping session the larval and nymphal *I. scapularis* were counted on each mouses' head and ears. The relationships between these counts and whole-body burdens estimated from drop off data is strong for larvae ($R^2 = 0.79$) [18].

Questing activities and larval burdens are highly seasonal [16], showing fairly distinct, but broad peaks in the late summer/early fall (mid- to late-August into early September). We therefore restricted our analyses to the data collected during these peaks, as visually identified. In addition, individual mice were often captured several times (this being a mark-recapture study). In order to avoid multiple non-independent measurements of tick burdens, we selected at random only one observation per individual mouse.

Densities of host-seeking, or "questing", larvae at a site were estimated by dragging a weighted 1m by 1m white corduroy cloth [19] along five $30m$ transects. Ticks were counted on the cloth at the end of these transects, so the grain of our tick density data is $30m^2$. Dragging was done several times during the expected peaks of larval activity, but the actual dates of dragging were inconsistent between years and trapping grids. We therefore restricted our analyses to the three or four transects that coincided with and straddled the peaks of questing larvae densities and of larval burdens.

Ideally, we would have data on larval tick densities across the home ranges of individual mice, or at least at the scale of mouse home ranges. As with every other study we are aware of, our tick density data are not paired with individual mice and estimated mouse home range areas are generally much larger than the $30m^2$ tick drags (see Appendix B). To deal with this issue, we upscaled the density data to the home range sizes using two assumptions about the spatial correlation among $30m^2$ samples, which bracket the range of possibilities. The upscaling proceeds by selecting a random home range size from the home range area distribution (appendix B). This area is then "filled" $30m^2$ at a time from the distribution of $30m^2$ larval drags (appendix B). The larval drag transects are widely spaced within the trapping grids. Filling each home range with random samples from the tick density distribution corresponds to one extreme where there is no short distance correlation in tick densities (hereafter Rnd). Thus the filled home ranges all tend towards the overall mean tick density and among home range differences are at their minimum. The other extreme, perfect short distance correlation in tick densities (hereafter Cor), can be obtained by taking a single random sample from the larval drag distribution and multiplying it by $A/30$, where A is the area of the focal home range. In this case, the large degree of heterogeneity observed among $30m^2$ drags is preserved at the scale of entire home ranges. For each grid/year combination, this procedure was repeated 10000 times for each of the Rnd and Cor assumptions. Finally, random samples of 15 areas (matching the smallest actual sample size involved, that of larval drags for each grid year combination) were drawn from the Rnd and Cor distributions for each grid/year combination. This resulted in 8 datasets: 2 grids x 2 years x 2 density upscaling assumptions.

Hierarchical Bayesian parameterization of the accumulation model

We employed a hierarchical Bayesian (HB) framework to fit our accumulation model to the larval burden and the upscaled larval density datasets (Fig. 1). The framework includes two latent variables—"true susceptibility" and "true tick density"—to account for the facts that: 1) susceptibility is not directly observed, and 2) "observations" of upscaled larval tick densities cannot be directly paired with observations of tick burdens. The overall likelihood is thus a product of the two conditionally independent likelihoods

of the data sources (burdens and upscaled densities), conditioned on the values of the latent variables (Fig. 1).

We used noninformative (uniform) priors for all four model parameters (η_S, ν_S, η_D, and ν_D) on the four Rnd datasets. For the four Cor datasets, a weakly informative half-Cauchy prior [20, 21] was used on η_S to achieve convergence (see appendix C for explanation), while uniform priors were used for the other three parameters. Though this prior introduces a slight bias in the results in favor of increasing the apparent contribution of differential susceptibility, its effect on our qualitative results is negligible: Upscaled tick densities account for most of the aggregation in burdens in the Cor datasets (table 1).

We implemented this approach via MCMC sampling in WinBugs 1.4 [22]. The WinBugs code including the priors is listed in appendix C. All analyses except for TX 1999 Cor employed a 70,000 iteration burn-in period followed by 30,000 iterations of which 5000 were kept as samples from the posterior distribution. A longer burn-in period of 150,000 iterations was used for TX 1999 Cor. For each dataset, we ran three chains started from widely spaced initial conditions. We used $\hat{R}$, the Gelman-Rubin statistic [23], to verify convergence was achieved ($\hat{R} < 1.1$ for all model parameters). Finally, we used posterior predictive simulations to check the fit of the models to the burden data and to propagate uncertainty in model parameters through to summary quantities that are functions of model parameters (μ, k and q).

Results

The parameterized HB models successfully described the observed distributions of larval blacklegged ticks on white-footed mice (Figs. 2 and 3). The expected burden distribution under the fitted model and the 95% credible regions in these figures are obtained by posterior predictive simulation. The model fits well in most cases, with some disagreement in the upper quantiles for GC 1999 Rnd, and to a lesser extent, for GC 1999 Cor (Fig. 2). Table 1 presents Bayesian posterior means and 95% credible intervals for accumulation model parameters.

Our analytical approximation of the burden distribution allows us to examine the factors that drive aggregation. Equation (4) shows that k depends only on the shape parameters of the distributions of D and S. The effect of each variable on k goes away as its distribution becomes symmetrical (shape parameter becomes large). Thus skewness in each component distribution translates into aggregation in the overall distribution of burdens.

Point estimates of q, the degree to which the aggregation in tick burdens is driven by variation in questing larval density, are higher in the Cor datasets (those in which tick distributions are spatially autocorrelated) than in the Rnd datasets (those without spatial autocorrelation) (Table 1). This indicates that strongly skewed tick density distributions (η_D small) can account for most of the aggregation observed in the burden data. Even in the Rnd datasets, where tick densities in different hosts' home ranges are more similar, point estimates for q can be as high as 0.54 (TX 1999 Rnd), and is never less than 0.1, indicating that variability in the tick density experienced by different mice can still play a substantial role in explaining observed burdens.

Table 2 compares the posterior predictive mean values of the burden distribution mean and aggregation parameter under the fitted accumulation model for both the Rnd (μ_{Rnd} and k_{Rnd}) and Cor (μ_{Cor} and k_{Cor}) datasets to their empirical counterparts (μ_{Emp} and k_{Emp}) obtained by fitting the NBD directly to the burden data via maximum likelihood. The point and interval estimates obtained by the different methods are generally similar, further demonstrating the consistency between the accumulation model and observed data. Though there is substantial overlap between the Rnd and Cor datasets, the Cor datasets produce somewhat lower values of k than those estimated directly from the burden data, suggesting that such strong spatial correlation in tick density introduces too much aggregation in burdens. This tendency suggests that, at least in extreme scenarios, variation in the tick density experienced by different individuals can be more than enough to account for the observed aggregation in parasite burdens.

Discussion

We have developed an extension of the classical Poisson-gamma mixture model of overdispersed parasite burdens by linking among-host variation in parasite sampling rate to a mechanistic model of parasite accumulation. The formulation of the accumulation model is flexible and can be tailored to different macroparasite systems as necessary. When embedded in a hierarchical Bayesian statistical framework, our model allows multiple sources of information, acting on different hierarchal levels, to be coherently integrated. The parameters of the distribution of parasite burdens can be written in terms of the components of the accumulation model and thus be linked to lower-level processes, and uncertainty in model parameters can be propagated through to quantities that are functions of model parameters (q in our empirical example). Our framework differs from other models of macroparasite burdens in that it describes the shape of the distribution of burdens as a function of a biologically relevant parameters rather than simply treating the overdispersion parameter, k, as a nuisance parameter, as in regression-based approaches, or leaving the variation in sampling rates among hosts unexplained, as in traditional applications of the Poisson-gamma mixture.

With only four free parameters, our model is able to reproduce the observed distribution of blacklegged tick burdens on white-footed mice in several places and times. Moreover, it provides a novel way to separate the contribution of intrinsic factors affecting parasite aggregation from that of extrinsic factors such as spatial variation in parasite density. The burden data provide strong information about k, while the upscaled tick density data provide information about D, both directly and indirectly through the accumulation model (Fig. 1). Overdispersion in burdens that cannot be accounted for by D is absorbed by the latent variable S and is thus attributed to differential susceptibility. Additional information, such as data on individual rates of host movement or grooming could easily be accommodated within our framework to provide more precise estimates of the parameters governing S.

We have identified the degree of patchiness in the spatial distribution of questing tick density as a key factor in explaining observed burden distributions in our empirical example. Highly patchy distributions imply strong short-distance correlation in tick density. In other words, adjacent areas will likely have similar tick densities on small spatial scales. Though we were not able to directly quantify this correlation with available data, we based our analyses on two extreme assumptions (no correlation and perfect correlation) that bracket the range of possibilities. In the no correlation extreme (Rnd), where tick densities are assumed to be fairly similar among home ranges, variation in tick densities still played an important role in some locations and years (e.g., TX 1999 Rnd), while it was less of a factor in others (e.g., TX 2004 Rnd). Importantly, the influence of this extrinsic factor did not completely disappear in any of the cases considered under the Rnd assumption (q values ranged from 0.1 to 0.54). Examining the other extreme (Cor), where the differences in tick densities among home ranges were most pronounced, we see that there is often a tendency towards slightly too much aggregation, as demonstrated by the posterior predictive mean k values in table 2. This implies that highly patchy spatial distributions of questing ticks are capable of accounting for most or all of the aggregation observed in tick burden distributions. Furthermore, questing tick density had a strong effect on observed k values at all grid/year combinations examined under the Cor assumption. As questing larvae distributions are known to be highly patchy [24, 25], we argue that the degree of correlation will tend to fall closer to the Cor extreme than to the Rnd. How the spatial scale of patchiness in tick distributions maps onto the the scale of host home range areas is an issue that merits further investigation.

Our results suggest that the often extreme differences in individual tick burdens we observe are not solely or, in many cases, even mostly caused by intrinsic differences in individual susceptibility due, for instance, to sex or life history stage. These differences in burdens can instead be explained primarily by random differences in the densities of questing ticks experienced by different hosts. As questing tick densities become more variable among home ranges, so ticks become more aggregated on a relatively small proportion of hosts. In other words, our results imply that some mice may have extremely large tick burdens simply because of bad luck; their home ranges happen to overlap with areas of high tick

density.

Spatial variation in questing larval density is very clearly a product of random processes. Gravid *I. scapularis* drop to the ground after feeding to repletion on their blood meal host, usually a deer or other larger bodied mammal, wherever that may be, and lay eggs close to where they fall [16]. The adult females and resulting larvae move no more than a meter or two while questing for a host [26]. There is no evidence, to the best of our knowledge, that gravid females choose where to drop to the ground. If there is a deterministic aspect to local questing larval densities, it is that some locations may be more favorable for larval hatching and survival [27]. Our results highlight the importance of directly quantifying questing tick density within each host home range. The availability of such data would improve our ability to pin down the mechanisms driving the aggregation of vectors on hosts. Quantifying variability in movement rates among hosts would further refine our understanding of the mechanisms governing parasite accumulation.

Our main empirical result, that variation in tick densities among home ranges can strongly affect tick burdens, is, one the one hand, not surprising. It is well known that tick spatial distributions are patchy on relatively small scales [24, 25], and it is logical to expect that this variability will affect tick burdens on hosts. On the other hand, the focus in the literature has very clearly been on trying to identify *a priori* biological characteristics that reliably predict parasite burdens [7–9]. This search carries with it an assumption that such host-related factors account for the majority of the variation in observed burdens. Our results cast substantial doubt on this assumption, and suggest that more effort should be spent on testing it and on quantifying the contribution of random, extrinsic factors. As much of the variation in tick burdens could potentially be explained by largely unpredictable, small-scale variation in the density of questing ticks, our results imply that it may be impossible to predict *a priori* the type(s) of individuals that will accumulate the largest burdens, and thus make the greatest contribution to disease transmission. Management strategies that assume such an *a priori* determination of heavily burdened individuals is possible may therefore prove ineffective and may waste limited management resources.

Acknowledgments

We thank C. Dormann, S. LaDeau, P. Leimgruber, K. Terrell, and J. Thompson for helpful comments on earlier drafts of the manuscript. We are also grateful to K. Oggenfuss and the many field assistants who collected these data at the Cary Institute with support from NSF grants DEB 0075277 and DEB 0444585, and NIH grant R01 AI40076.

References

1. Shaw DJ, Dobson AP (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. Parasitology 111: S111-S133.

2. Shaw DJ, Grenfell BT, Dobson AP (1998) Patterns of macroparasite aggregation in wildlife host populations. Parasitology 117: 597-610.

3. Woolhouse MEJ, Dye C, Etard JF, Smith T, Charlwood JD, et al. (1997) Heterogeneities in the transmission of infectious agents: implications for the design of control programs. Proceedings of the National Academy of Sciences of the United States of America 94: 338-342.

4. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438: 355-359.

5. Jones LD, Davies CR, Steele GM, Nuttall PA (1987) A novel mode of arbovirus transmission involving a nonviraemic host. Science 37: 775-777.

6. Labuda M, Jones LD, Williams T, Danielova V, Nuttall PA (1993) Efficient transmission of tick-borne encephalitis virus between co-feeding ticks. Journal of Medical Entomology 30: 295-299.

7. Perkins SE, Cattadori IM, Tagliapietra V, Rizzoli AP, Hudson PJ (2003) Empirical evidence for key hosts in persistence of a tick-borne disease. International Journal for Parasitology 33: 909.

8. Brunner JL, Ostfeld RS (2008) Multiple causes of variable tick burdens on small-mammal hosts. Ecology 89: 2259-2272.

9. Kiffner C, Vor T, Hagedorn P, Niedrig M, Ruhe F (2011) Factors affecting patterns of tick parasitism on forest rodents in tick-borne encephalitis risk areas, germany. Parasitology Research 108: 323-335.

10. Paterson S, Lello J (2003) Mixed models: getting the best use of parasitological data. Trends in Parasitology 119: 370-375.

11. Kiffner C, L
"odige C, Alings M, Vor T, R
"uhe F (2011) Body-mass or sex-biased tick parasitism in roe deer (*Capreolus capreolus*)? a gamlss approach. Medical and Veterinary Entomology 25: 39–45.

12. Wilson K, Bjornstad ON, Dobson AP, Merler S, Poglayen G, et al. (2002) Heterogeneities in macroparasite infections: patterns and processes. In: Hudson PJ, Rizzoli A, Grenfell BT, Heesterbeek H, Dobson AP, editors, The Ecology of Wildlife Diseases, New York, New York, USA: Oxford University Press. p. 197.

13. Hubbard A, Liang S, Maszle D, Qiu D, Gu X, et al. (2002) Estimating the distribution of worm burden and egg excretion of *Schistosoma japonicum* by risk group in sichuan province, china. Parasitology 125: 221–231.

14. Boswell MT, Patil GP (1970) Chance mechanisms generating negative binomial distributions. In: Patil GP, editor, Random Counts in Scientific Work, Vol. 1, Pennsylvania State University Press. pp. 3-22.

15. Boyer N, Reale D, Marmet J, Pisanu B, Chapuis JL (2010) Personality, space use and tick load in an introduced population of siberian chipmunks *Tamias sibiricus*. Journal of Animal Ecology 79: 538-547.

16. Fish D (1993) Population ecology of *Ixodes dammini*. In: Ginsberg H, editor, Ecology and Environmental Management of Lyme Disease, New Brunswick, New Jersey, USA: Rutgers University Press. pp. 25-42.

17. Hilborn R, Mangel M (1997) The ecological detective: Confronting models with data. Princeton Univ. Press.

18. Schmidt KA, Ostfeld RS, Schauber EM (1999) Infestation of *Peromyscus leucopus* and *Tamias striatus* by *Ixodes scapularis* (acari: Ixodidae) in relation to the abundance of hosts and parasites. Journal of Medical Entomology 36: 749-757.

19. Falco RC, Fish D (1992) A comparison of methods for sampling the deer tick, *Ixodes dammini*, in a lyme disease endemic area. Experimental and Applied Acarology 14: 165-173.

20. Gelman A (2006) Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1: 515–533.

21. Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press New York, New York, USA.

22. Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) Winbugs: a bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10: 325-337.

23. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Statistical Science 7: 457–472.

24. Petney TN, Van Ark H, Spickett AM (1990) On sampling tick populations: the problem of overdispersion. The Onderstepoort Journal of Veterinary Research 57: 123–127.

25. Markowski D, Hyland KE, Ginsberg HS, Hu R (1997) Spatial distribution of larval ixodes scapularis (acari: Ixodidae) on peromyscus leucopus and microtus pennsylvanicus at two island sites. The Journal of Parasitology 83: 207–211.

26. Falco RC, Fish D (1991) Horizontal movement of adult *Ixodes dammini* (acari: Ixodidae) attracted to co2-baited traps. Journal of Medical Entomology 28: 726-729.

27. Lindsay LR, Barker IK, Surgeoner GA, McEwen SA, Gillespie TJ, et al. (1998) Survival and development of the different life stages of *Ixodes scapularis* (acari: Ixodidae) held within four habitats on long point, ontario, canada. Journal of Medical Entomology 35: 189-199.

Figure Legends

Figure 1. Hierarchical Bayesian model structure of the tick accumulation model. Gray boxes identify the levels in the hierarchy, white boxes represent data, and white ovals represent low-level model elements. Arrows show the relationships among model elements.

Figure 2. Bayesian fits of the model to the four GC grid datasets, as visualized with quantile-quantile plots. The "expected" distribution (solid lines) under the fitted accumulation model, as well as the 95% credible regions (dashed lines) around the predicted line, were generated via posterior predictive simulations.

Figure 3. As in figure 2, but for the four TX grid datasets.

Supporting Information

Appendix A. Approximation of the sampling rate distribution.

Appendix B. Home range sizes and upscaled larval density.

Appendix C. WinBUGS code and information about priors.

Tables

Table 1. Grid-specific Bayesian mean estimates for the accumulation model parameters. 95% credible intervals are in parentheses below the point estimates. The sample sizes for the burden datasets are 132, 165, 96, and 91 for GC 1999, GC 2004, TX 1999, and TX 2004, respectively. A sample size of 15 was used for all upscaled density datasets.

	η_s	ν_s	η_d	ν_d	q
GC 1999 Rnd	1.89	11.71	9.50	0.17	0.27
	(1.31, 2.89)	(6.82, 17.72)	(3.47, 22.24)	(0.06, 0.38)	(0.10, 0.52)
GC 1999 Cor	5.24	4.89	1.99	0.75	0.73
	(2.27, 16.14)	(1.08, 9.90)	(1.38, 2.86)	(0.45, 1.28)	(0.56, 0.93)
GC 2004 Rnd	1.26	8.15	9.03	0.22	0.23
	(0.93, 1.82)	(5.00, 11.53)	(3.04, 18.81)	(0.08, 0.54)	(0.10, 0.50)
GC 2004 Cor	2.91	4.16	2.10	0.93	0.62
	(1.28, 9.32)	(0.92, 8.28)	(1.18, 3.63)	(0.44, 1.78)	(0.41, 0.88)
TX 1999 Rnd	3.66	11.95	3.64	0.18	0.54
	(1.47, 13.35)	(2.31, 24.07)	(1.66, 6.86)	(0.08, 0.37)	(0.29, 0.90)
TX 1999 Cor	19.52	2.02	1.21	0.70	0.92
	(4.06, 68.40)	(0.33, 6.02)	(0.92, 1.56)	(0.40, 1.16)	(0.79, 0.99)
TX 2004 Rnd	1.42	14.65	25.92	0.03	0.10
	(1.01, 1.97)	(9.54, 21.67)	(9.45, 49.40)	(0.02, 0.08)	(0.04, 0.23)
TX 2004 Cor	7.34	4.03	1.73	0.52	0.77
	(1.98, 32.07)	(0.48, 9.93)	(1.15, 2.60)	(0.29, 0.90)	(0.55, 0.96)

Table 2. The upper section contains grid-specific "empirical" maximum likelihood estimates of the mean (μ_{Emp}) and aggregation parameter (k_{Emp}) obtained by directly fitting the classical NBD to the larval burden data via maximum likelihood. Wald-type 95% confidence intervals are in parentheses below the MLEs. The bottom section presents the corresponding Bayesian posterior predictive means and 95% posterior predictive intervals for the Rnd (μ_{Rnd} and k_{Rnd}) and Cor (μ_{Cor} and k_{Cor}) datasets.

	GC 1999	GC 2004	TX 1999	TX 2004
μ_{Emp}	28.16	15.87	17.41	14.59
	(23.97, 32.36)	(13.43, 18.31)	(14.45, 20.38)	(11.86, 17.32)
k_{Emp}	1.38	1.05	1.50	1.31
	(1.06, 1.69)	(0.82, 1.28)	(1.06, 1.93)	(0.92, 1.71)
μ_{Rnd}	27.61	15.91	17.76	14.71
	(23.86, 32.64)	(13.55, 18.69)	(14.76, 21.74)	(12.14, 17.80)
k_{Rnd}	1.34	0.96	1.29	1.27
	(1.01, 1.70)	(0.74, 1.20)	(0.91, 1.77)	(0.935, 1.71)
μ_{Cor}	27.83	16.07	18.23	14.83
	(23.69, 32.82)	(13.77, 18.89)	(14.66, 22.53)	(11.99, 18.37)
k_{Cor}	1.16	0.87	1.02	1.06
	(0.87, 1.50)	(0.66, 1.15)	(0.72, 1.33)	(0.73, 1.48)

Figure1

Click here to download Figure: Figure1.eps

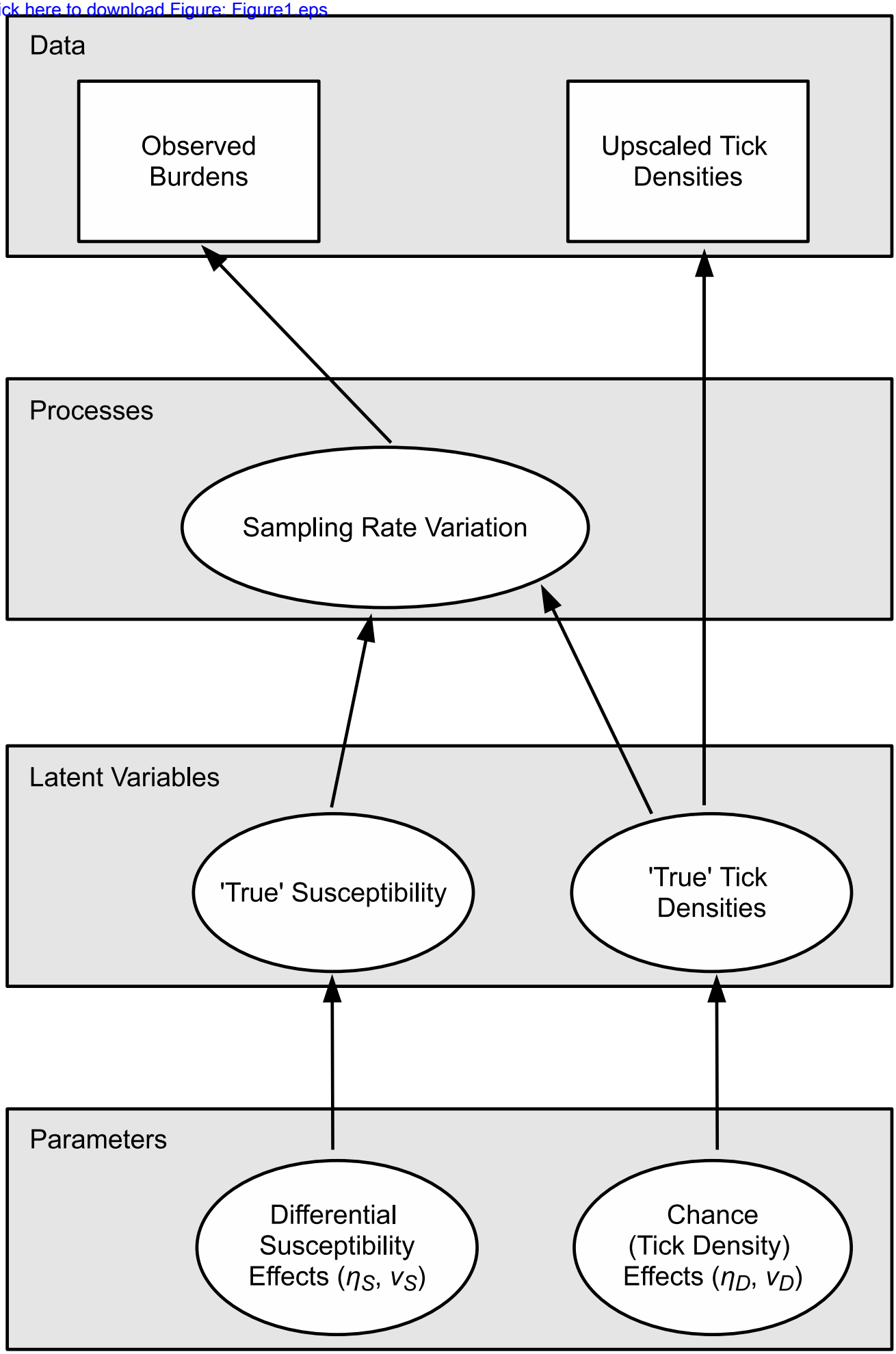


Figure2
Click here to download Figure: Figure2.eps

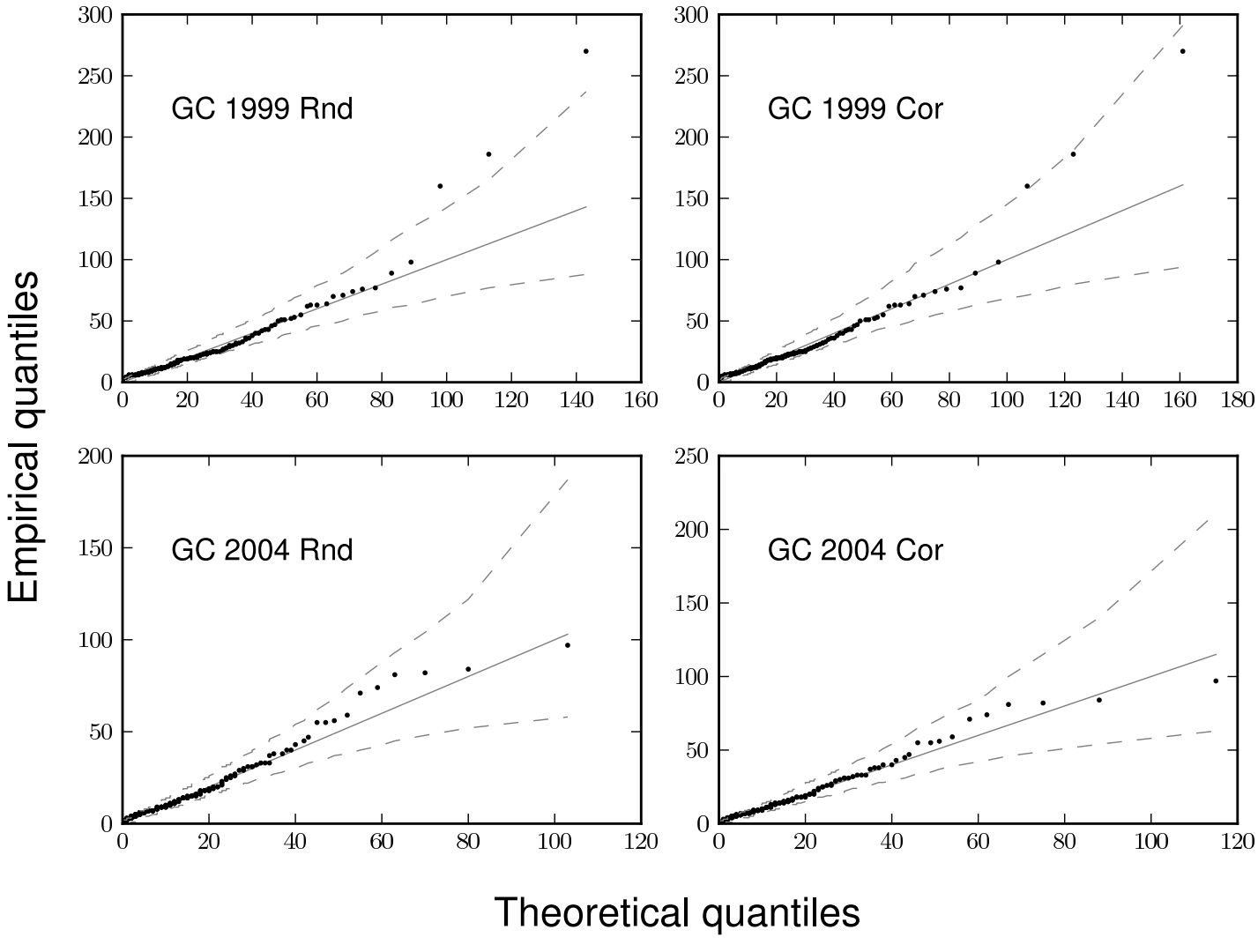
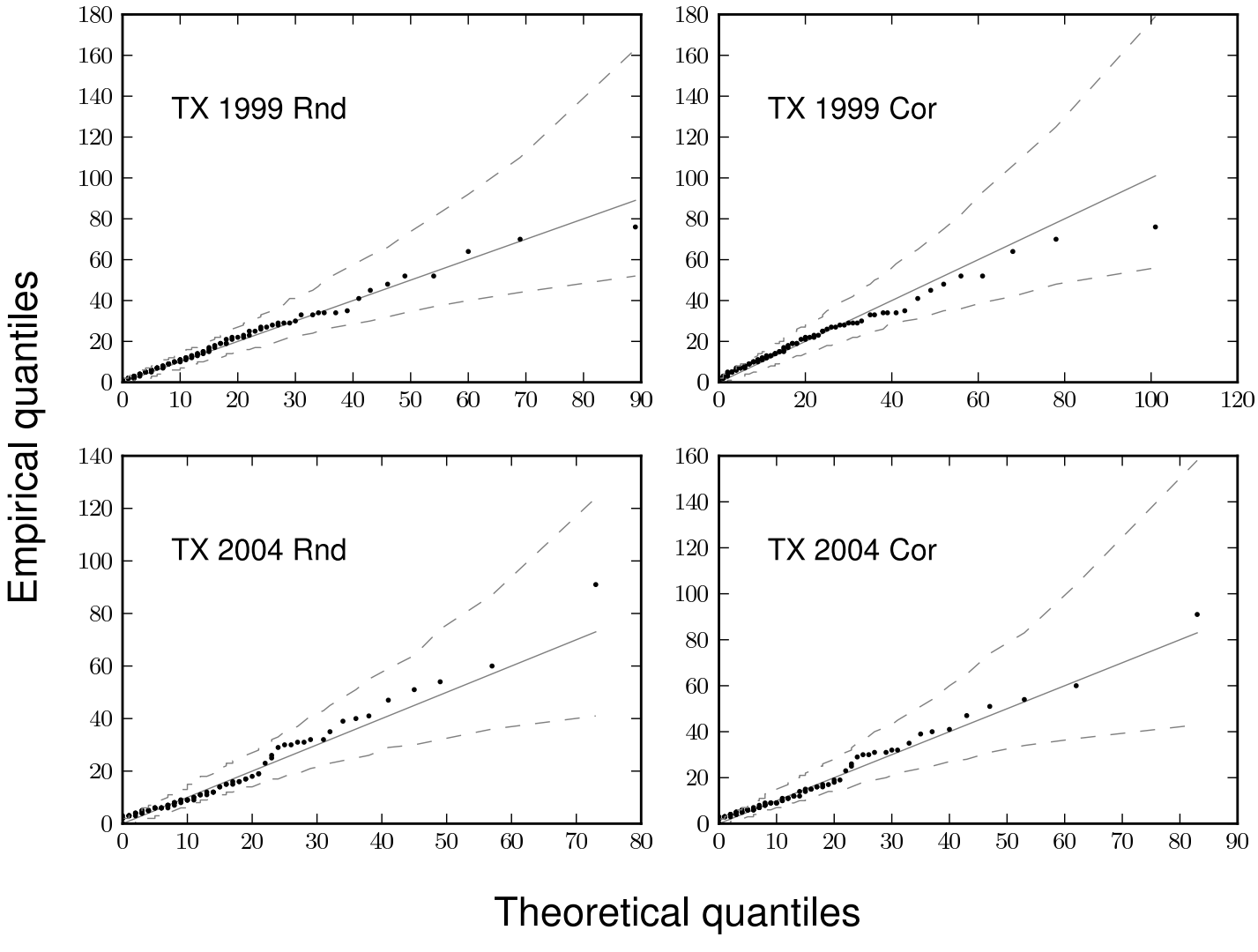


Figure3

Click here to download Figure: Figure3.eps



Supporting Information

Click here to download Supporting Information: CalabreseEtAl.Ticks.Appendices.Final.pdf