Borrelia burgdorferi Sensu Lato Spirochetes in Wild Birds in

Northwestern California: Associations with Ecological Factors,

Bird Behavior and Tick Infestation

Tianyi Zheng

The University of British Columbia

Introduction

Newman et al. investigated the birds' roles as bridging vectors for subadult *Ixodes pacificus* ticks and as potential host for the Lyme disease spirochete *B.burgdorferi sensu stricto* in north-western California. [1] Lyme disease is an infectious disease caused by *Borrelia* bacteria, which is spread by ticks. [2] In North America, Lyme disease is caused by *B.burgdorferi sensu stricto*, which are mainly from the nymph stages of *Ixodes scapularis* and *Ixodes pacificus*.[3] *Ixodes pacificus* is known to infested more than 100 species of lizard, birds and mammals in California, and *Ixodes scapularis* has similar impact on the East Coast. [4] While most researches have focused on the role of mammals as spirochete reservoir hosts, birds are potentially important hosts because of their species-rich and ecologically diverse. [1] Birds' dispersal distances are generally much larger than for mammals and they migrate annually. Thus, they may have an impact on the geographic spread of ticks and *borrelia* bacteria.

There are 3 main goals of this paper: First, the author is trying to identify birds' species that are potentially important to the enzootic maintenance of Lyme disease spirochetes in northwestern California. Second, the author is trying to evaluate the individual nesting and feeding behaviors that contribute to the transmission of spirochetes. Third, they are trying to investigate whether avian taxonomic relationships are predictive of *Ixodes pacificus* larval and nymph loads, and of *B.burgdorferi* infection.[1]

The experimental site for this study is located at the foothills of the Mayacmas Mountains in northwestern California. [1] Birds were captured during the breeding seasons of 2003 and 2004 by

using mist netting. [1] Birds were identified to species and examined for presence of ticks. Ticks were removed with fine-tipped forceps, mostly around the eyes and beak where most *Ixodes* ticks attach. [1] Genomic DNA from ticks and bird blood samples was extracted using DNeasy Blood and Tissue Kit. [1] Each sequence was directly compared with sequences of the same loci from various spirochete genospecies available in the GenBank database. [1]

Explanatory variable	Levels (or data type)	Level names
log(average body weight) (LOG.AVEBWT)	(continuous)	(continuous)
Breeding status (BRD)	breeding	BREED
	non-breeding	NONBRD
Feeding substrate/behavior (FDSUB)	air (aerial, flycatching, hovering)	AIR
	bark	BARK
	foliage gleaning	FOLIAGE
	ground	GROUND
	stalking	STALK
Main food (MNFD)	insects	INSECT
	omnivore	OMNIVORE
	mammals	MAMMAL
	seeds	SEEDS
Main habitat (MNHAB)	chaparral	CHAP
	grass	GRASS
	oak woodland-grass	OAKW
	dense oak woodland	XW
Nest substrate (NEST)	bark	BARK
	cavity	CAVITY
	ground	GROUND
	opportunistic	OPP
	shrub	SHRUB
	tree	TREE
	non-breeding	NON
Resident status (RESSTAT)	resident	RES
	non-resident	NONRES
Year (YEAR)	2 levels by data collection date	2003
		2004
Number of nymphs removed from bird (N_NYM)	(count data)	(count data)
Species (SPECIES)	52 instances by 4-letter species codes	[52 levels (excludes REVI)]
Genus (GENUS)	44 instances by genus name	[44 levels] (see Table 1)
Family (FAMILY)	24 instances by family name	[24 levels] (see Table 1)
Order (ORDER)	4 instances by order name	[4 levels] (see Table 1)

doi:10.1371/journal.pone.0118146.t002

Table 1. Explanatory variables and their levels included in the analysis. (From Table 2 of the original paper)

Methods

Modeling goal	Model family	Data	Unit of analysis	Outcome
Generalized linear models				
Bird infection (presence/absence)	binomial	orders: Passeriformes and Piciformes	individual bird	Best supported model: Bird infection ~ N_NYM + YEAR + ORDER
Bird infection (presence/absence)	binomial	order: Passeriformes only	individual bird	Best supported model: Bird infection ~ N_NYM +YEAR + LOG_BWT
Larval infestation (presence/absence)	binomial	orders: Passeriformes and Piciformes	individual bird	Best supported model: Larval presence ~ YEAR + ORDER + FDSUB
Larval infestation (presence/absence)	binomial	order: Passeriformes only	individual bird	Best supported model: Larval presence ~ YEAR + NEST + FDSUB
Nymphal infestation (presence/absence)	binomial	orders: Passeriformes and Piciformes	individual bird	Best supported model: Nymphal presence ~ YEAR + FAMILY
Nymphal infestation (presence/absence)	binomial	order: Passeriformes only	individual bird	Best supported model: Nymphal presence ~ YEAR + FAMILY + MNHAB
Zero-inflated negative binomial models				
Larval infestation (count data)	(count model: negative binomial with log link) | (zero-inflation model: binomial model with logit link)	orders: Passeriformes and Piciformes	individual bird	Best supported model: Number of larvae ~ (MNHAB) | (LOG.AVEBWT + YEAR)
Nymphal infestation (count data)	(count model: negative binomial with log link) | (zero-inflation model: binomial model with logit link)	orders: Passeriformes and Piciformes	individual bird	Best supported model: Number of nymphs ~ (MNHAB + FDSUB) | (YEAR + FDSUB)
Body weight Null Models				
BNM: Bird infection	single-variable linear regression	all observations	species	not significant
BNM: Bird infection	single-variable linear regression	positive counts only	species	not significant
BNM: Larval infestation	single-variable linear regression	all observations	species	not significant
BNM: Larval infestation	single-variable linear regression	positive counts only	species	not significant
BNM: Nymphal infestation	single-variable linear regression	all observations	species	not significant
BNM: Nymphal infestation	single-variable linear regression	positive counts only	species	not significant

doi:10.1371/journal.pone.0118146.t003

Table 2. Bird infection and tick infestation modeling goals and best supported models, based on AIC values. (Table 3 in paper)

Original Analysis:

There are 3 parts in the main analysis of the paper: Body weight null models, generalized linear models and zero-inflated negative binomial models.

1. Body weight Null Models:

The author believes that the birds with higher weight are more likely to be exposed to BBSL infection because of the larger surface area of skin. [1] Thus, body weight should serve as the adequate explanatory variable for the null model of infestation and infection prevalence. [1] They

fit single-variable linear regressions for Bird infection (for each individual bird, the bird infection status is a binary data, which has 0=negative/uninfected and 1=positive/infected, but since the unit of analysis is species here, the bird infection data is a number between 0 and 1, which can be interpreted as the infected ratio within the species), Larval infestation and Nymphal infestation (average number of *Ixodes pacificus* larvae/nymphs removed for each species). The author fit the models with all observations and with positive counts only (only include species that has at least one bird infected/ one larvae/nymphs removed) groups to investigate whether there is any differences between the groups. The unit of analysis in the Null model is species.

2. Generalized linear models:

The author states that based on preliminary analyses, data are not normally distributed, which violates the assumptions of homoscedasticity of error. [1] Moreover, birds vary across species, which leads to "unbalanced" data structure for each analysis. [1] Thus, the author considers generalized linear model as the tool to handle both zero-inflated data and unbalanced data. They fit a binomial GLM on Bird infection, Larval infestation and Nymphal infestation. (All binary data as presence/absence) The author chose a candidate set for each model from the combinations of covariates in Table 1 that make sense biologically. The best supported model was selected from the candidate sets based on Akaike Information Criterion(AIC) values and second order AIC(AICc) values. The author did not specify why the analysis is based on order Passeriformes and Passeriformes only. It might be these two orders are most common in North-California.

3. Zero-inflated negative binomial models:

As the reason stated in the GLM section, the author fit zero-inflated negative binomial models to investigate the tick load (number of larvae and number of nymphs, both are count data). The zero-inflated negative binomial model consists of a negative binomial count model with a log link, and a zero-inflation binomial model with a logit link. The author chose a candidate set of models from the combinations of variables in Table 1 that biologically make sense. The best supported model was selected from the candidate sets based on Akaike Information Criterion(AIC) values and second order AIC(AICc) values.

Alternative analysis:

1. Body weight Null Models:

The author did not specify why the average body weights are log-transformed, so I compared the log-transformation with untransformed and square root transformed average body weights. I checked the normality assumptions for each null model and tried to replace it with a GLM model. The choice of the family for GLM model was based on the distribution fitting result for the response variable in the null model.

2. Generalized linear models:

The author did not specify why they choose logit as the link function instead of probit, so I fitted all best supported models with probit link function and compared them with the logit results.

3. Zero-inflated negative binomial models:

Instead of the negative binomial count model, I tried to fit a zero-inflated poisson model and a quasipoisson GLM for Larval and Nymphal count data.

Results

1. Body weight Null Models:

   Reproduction for the linear regression fit plots:

   Solid line: with all observations   Dashed line: with positive counts only

   (a) Bird infection ~ log (average body weights) (b) Larval infestation ~ log (average body weights) (c) Nymphal infestation ~ log (average body weights)
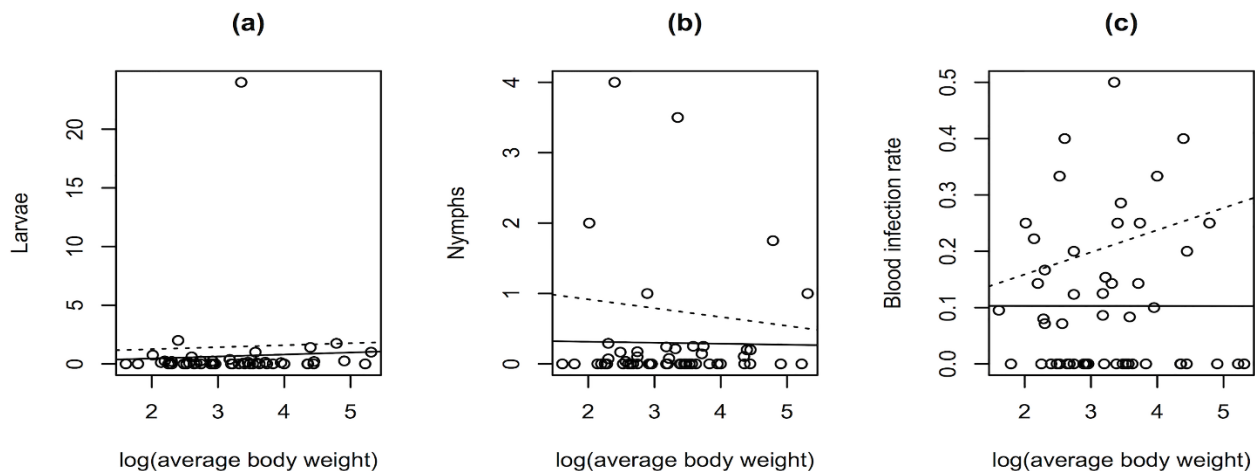

Fig 1. Body weight as a null model for presence or absence of tick infestation and *B. burgdorferi s.l.* infection in birds. (From Fig 2 of the original paper)
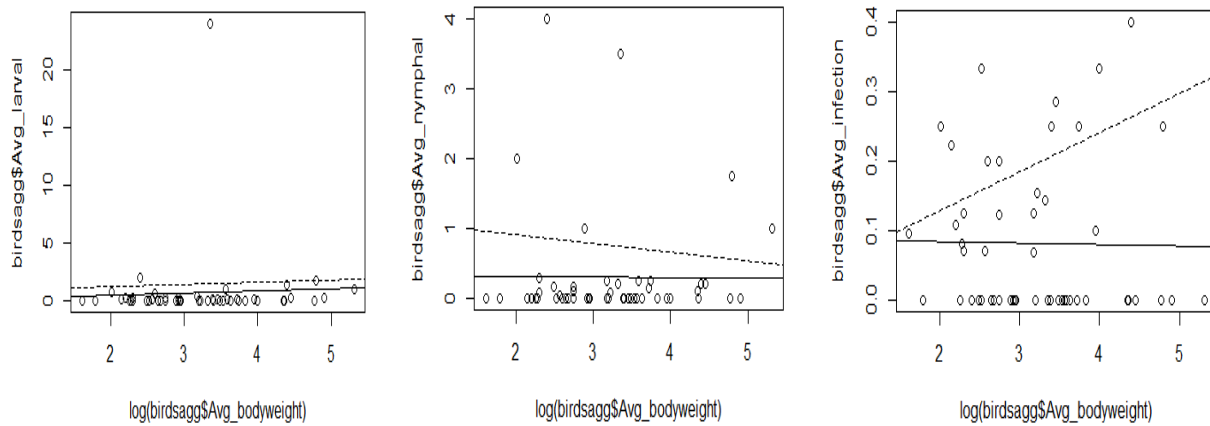
Fig 2. Reproduction of Fig1.

Reproduced P-values associated with each model, compared with the original P-values.

For all observations:

Model	a	b	c
Original P-value	0.9956	0.7797	0.9118
Reproduction P-value	0.9162	0.6989	0.9899

For positive counts only:

Model	a	b	c
Original P-value	0.1566	0.8791	0.6635
Reproduction P-value	0.2135	0.8791	0.6635

Histograms for different transformations on average of body weights and the fit for normal distribution on top.

Histogram of birdsagg$Avg_bodyweight

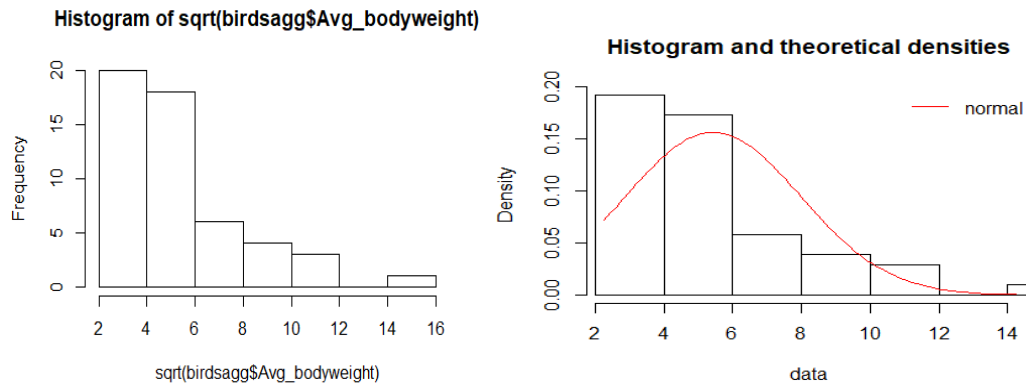**Histogram and theoretical densities**

Fig 2. Histogram on the original average body weights.

Histogram of sqrt(birdsagg$Avg_bodyweight)

Histogram and theoretical densities

Fig 3. Histogram on the square root transformed average body weights.

Histogram of log(birdsagg$Avg_bodyweight)

Histogram and theoretical densities
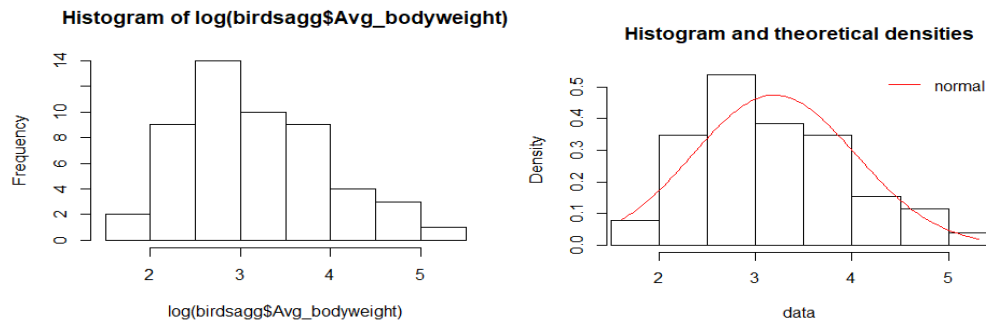
Fig 4. Histogram on the log-transformed average body weights.

Histogram for the average infection rate for different species and the fit for normal, gamma, Weibull distributions on top.
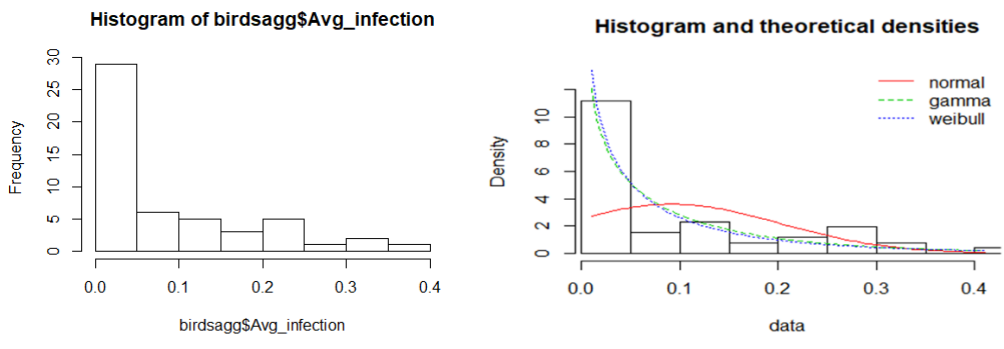

Fig 5. Histogram on the original average body weights, with distribution fitting results.

The residual plot and normal Q-Q plot for single-variable linear regression:

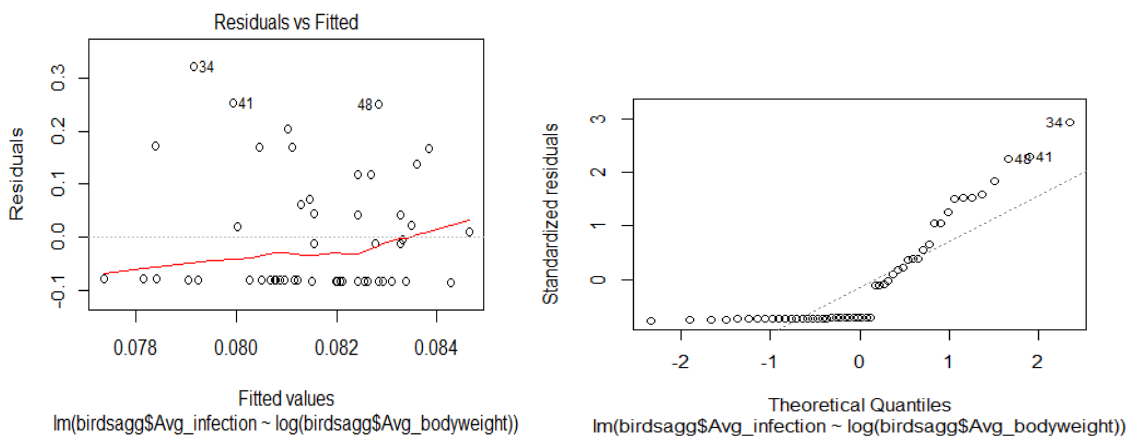Average infection rate ~ log (average body weight)

Fig 6. Residual plot and normal Q-Q plot for single-variable linear regression

The residual plot and normal Q-Q plot for Gamma GLM:
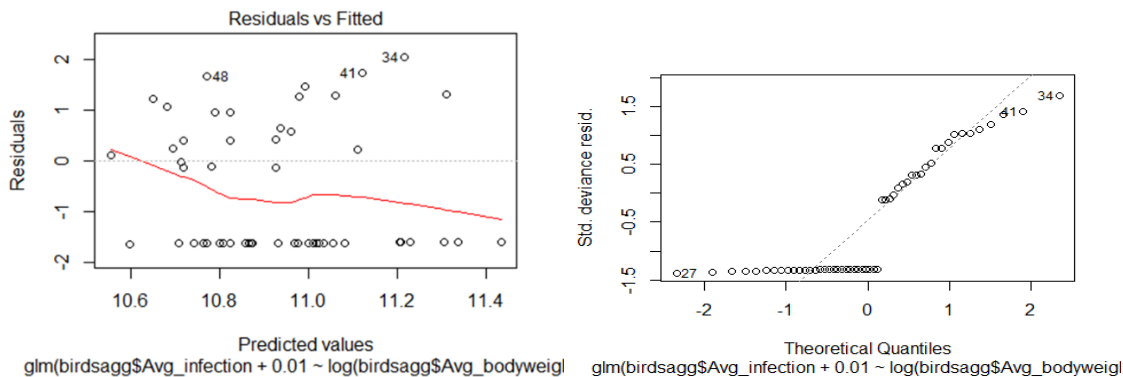
Average infection rate ~ log (average body weight)

Fig 7. Residual plot and normal Q-Q plot for Gamma GLM

2. Generalized linear models:

Since all the model selections for best supported models are the same with the original paper, I will not paste them here, they are in the R markdown output document.

Refit the best supported models with probit link function:

Best supported model type	AICc for logit	AICc for probit
Bird infection (2 orders)	397.46	397.21
Bird infection (1 order)	397	397.16
Larval infestation (2 orders)	352.51	352.42
Larval infestation (1 order)	332.11	331.29
Nymphal infestation (2 orders)	319.81	320.62
Nymphal infestation (1 order)	300.61	301.09

3. Zero-inflated negative binomial models:

Since all the model selections for best supported models are the same with the original paper, I will not paste them here, they are in the R markdown output document.

AICc Value	Zero-inflated negative binomial models	Zero-inflated poisson models
Larval infestation	526.0233	590.0875
Nymphal infestation	466.898	472.1667

For the negative binomial count model and poisson count model, the sets of significant covariates are the same for both Larval infestation and Nymphal infestation. But for quasipoisson GLM, the sets are different.

Discussions

I noticed that the dataset for the analysis in the original paper is not the same as which published on Dryad. Throughout the paper, the author states that the number of species n=53, however, as in the Table 2 from the original paper and in the published dataset, n=52.

1. Body weight Null Models:

   From the transformation plots, I noticed that the square root transformation has a reasonable performance in handling overdispersion and outliers. However, the log transformation performed better since it moved towards normality and I have notice that in the plot of the square root transformation, there is still a gap between the main data and the outlier.

   On the other hand, if the dataset is highly zero-inflated, when applying the log transformation, we need to add a small constant since log (0) is undefined, which can potentially introduce bias to our estimation. However, square root transformation do not have this issue since sqrt(0)=0.

   From the Gamma GLM fitting results, despite that GLM extends linear regression by allowing us to specify the distribution that the data are assumed to have come from, it does not indicate that the GLM is superior than linear regression and it will always fit better. In our scenario, since the gamma function is not a good approximation for the average infection rate, the gamma GLM fitting result is not better than the single-variable linear regression model, as indicated in the residual and normal Q-Q plots. Despite the strict assumptions of parametric statistics, the simple linear regression is easy to implement and interpret.

2. Generalized linear models:

   The results from logit and probit models are close enough in terms of AICc value. Many authors in the application fileds (e.g. Paap, R. from applied econometrics [6]) claim that the fitting and performance from logit and probit functions are very similar to each other.

3. Zero-inflated negative binomial models:

   Quasi-Poisson and negative binomial distributions are effective tools for handling overdispersion. [7] However, we should not compare the quasi-poisson model AIC with the negative binomial model AIC since quasi models are based on quasi-AIC (QAIC). We can compare the fitting results from poisson and negative binomial models since they are based on AIC and from the same family.

References

(1) Newman EA, Eisen L, Eisen RJ, Fedorova N, Hasty JM, et al. (2015) *Borrelia burgdorferi* Sensu Lato Spirochetes in Wild Birds in Northwestern California: Associations with Ecological Factors, Bird Behavior and Tick Infestation. PLOS ONE 10(2): e0118146. https://doi.org/10.1371/journal.pone.0118146

(2) "Clinical practice. Lyme disease" N. Engl. J. Med. 370 (18): 1724–31. doi:10.1056/NEJMcp1314325. PMC 4487875 Freely accessible. PMID 24785207. Archived from the original (PDF) on 19 October 2016.

(3) Burgdorfer W, Lane RS, Barbour AG, Gresbrink RA, and Anderson JR (1985) The western black-legged tick, Ixodes pacificus: a vector of Borrelia burgdorferi. Am J Trop Med Hyg 34: 925–930. pmid:3898886

(4) Castro MB, Wright SA (2007) Vertebrate hosts of Ixodes pacificus (Acari: Ixodidae) in California. J Vector Ecol 32: 140–149. pmid:17633435

(5) Weisbrod AR, Johnson RC (1989) Lyme disease and migrating birds in the Saint Croix River Valley. Appl Environ Microbiol 55: 1921–1924. pmid:2782872

(6) Paap, R. and Franses, P. H. (2000), A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables. J. Appl. Econ., 15: 717–744. doi:10.1002/jae.580

(7) O'Hara, R. B. and Kotze, D. J. (2010), Do not log-transform count data. Methods in Ecology and Evolution, 1: 118–122. doi:10.1111/j.2041-210X.2010.00021.x