Model Fitting and Comparison

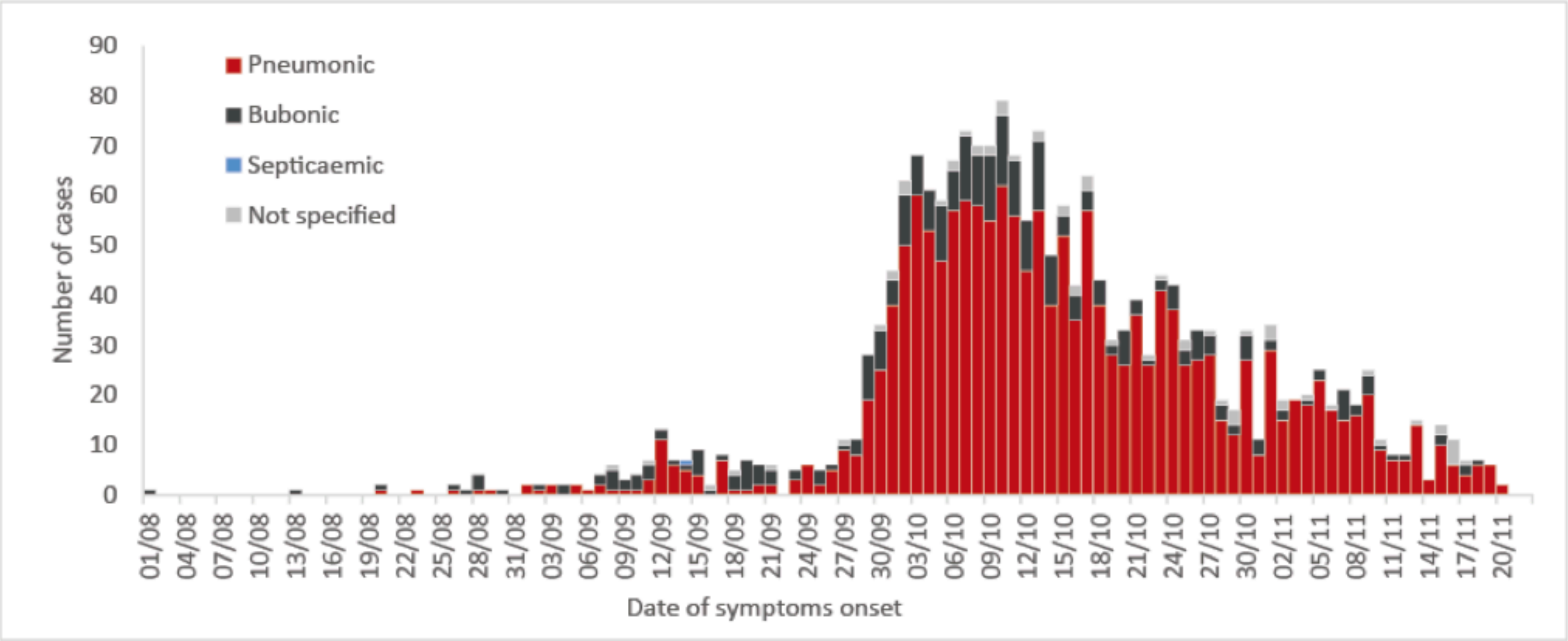
Cara Brook

University of California, Berkeley

E2M2 2020

Ranomafana, Madagascar

Which model is best?



http://www.who.int/csr/don/27-november-2017-plague-madagascar/en/

There are many statistical methods used to ‘fit’ models to data.

There are many statistical methods used to ‘fit’ models to data.

The method best suited for your work will depend on your model and your data.

There are many statistical methods used to ‘fit’ models to data.

The method best suited for your work will depend on your model and your data.

What are some measures of model fit used in E2M2 so far?

There are many statistical methods used to ‘fit’ models to data.

The method best suited for your work will depend on your model and your data.

What are some measures of model fit used in E2M2 so far?

Least squares

Log likelihood

AIC

(uses least squares or log-likelihood but penalizes by number of fitted parameters)

There are many statistical methods used to ‘fit’ models to data.

The method best suited for your work will depend on your model and your data.

What are some measures of model fit used in E2M2 so far?

Least squares

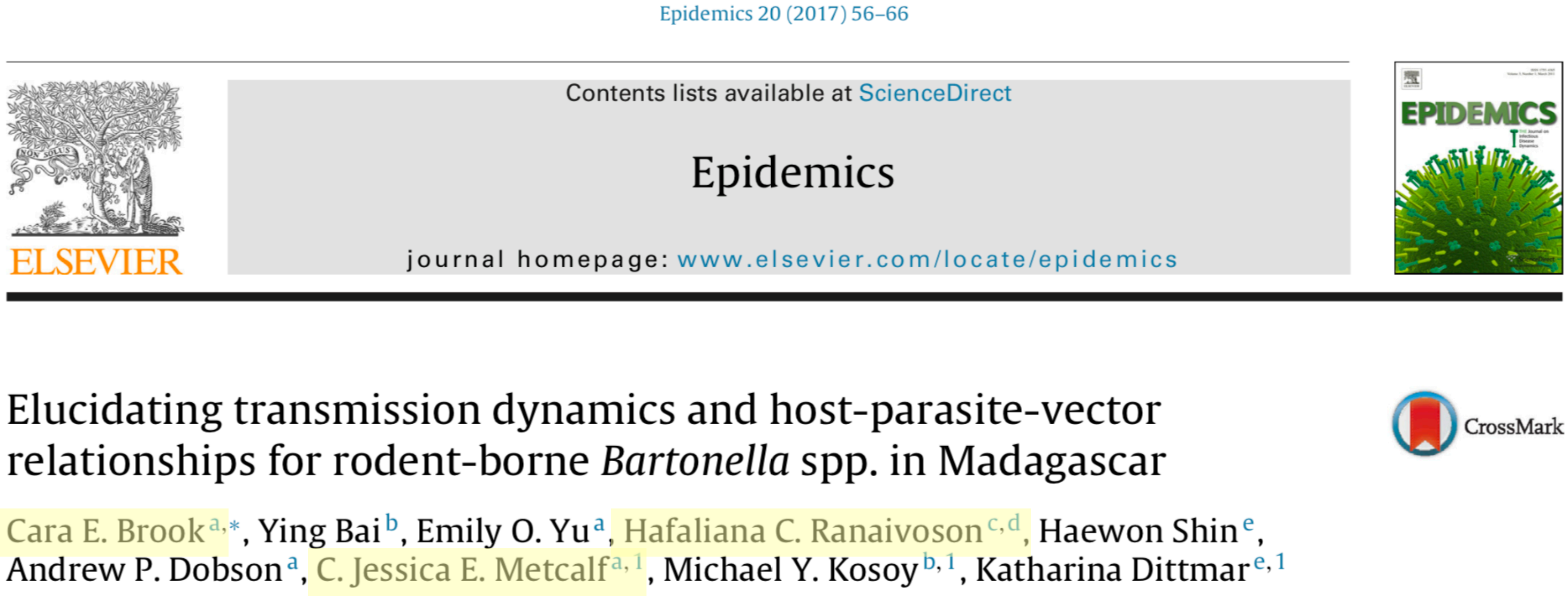
Log likelihood

AIC

(uses least squares or log-likelihood but penalizes by number of fitted parameters)

The R function ‘optim’ can be used to minimize these measures of model difference from the data.

An example of model selection: *Bartonella spp.* in Madagascar rats



Epidemics 20 (2017) 56–66

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

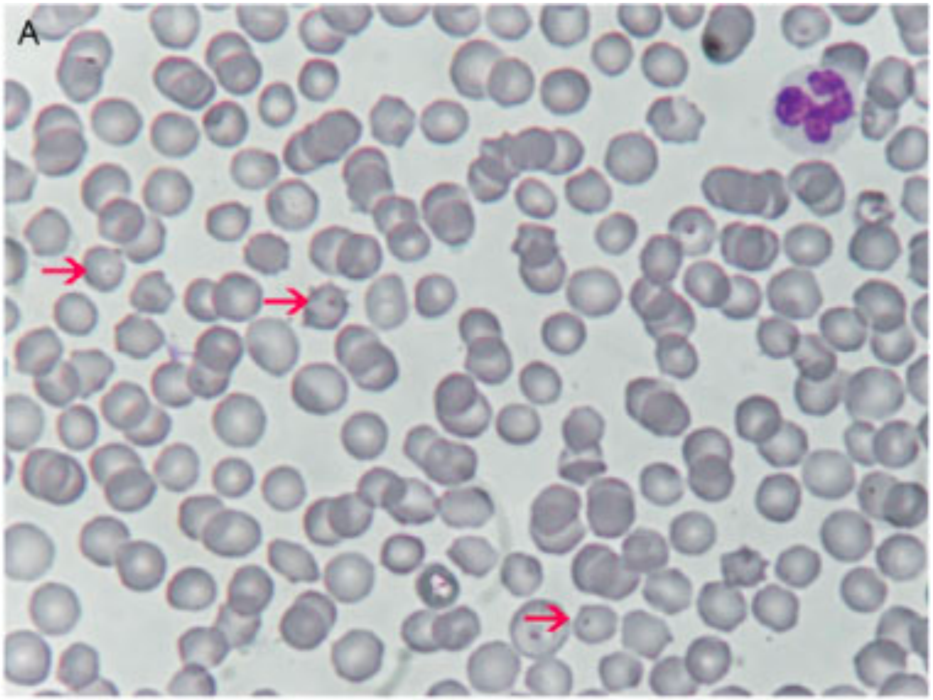
ELSEVIER

Elucidating transmission dynamics and host-parasite-vector relationships for rodent-borne *Bartonella* spp. in Madagascar

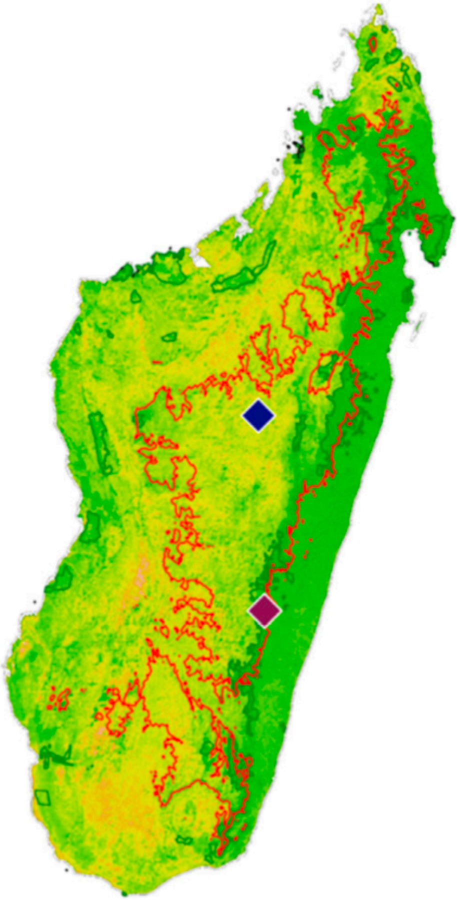
Cara E. Brook[a,*], Ying Bai[b], Emily O. Yu[a], Hafaliana C. Ranaivoson[c,d], Haewon Shin[e], Andrew P. Dobson[a], C. Jessica E. Metcalf[a,1], Michael Y. Kosoy[b,1], Katharina Dittmar[e,1]

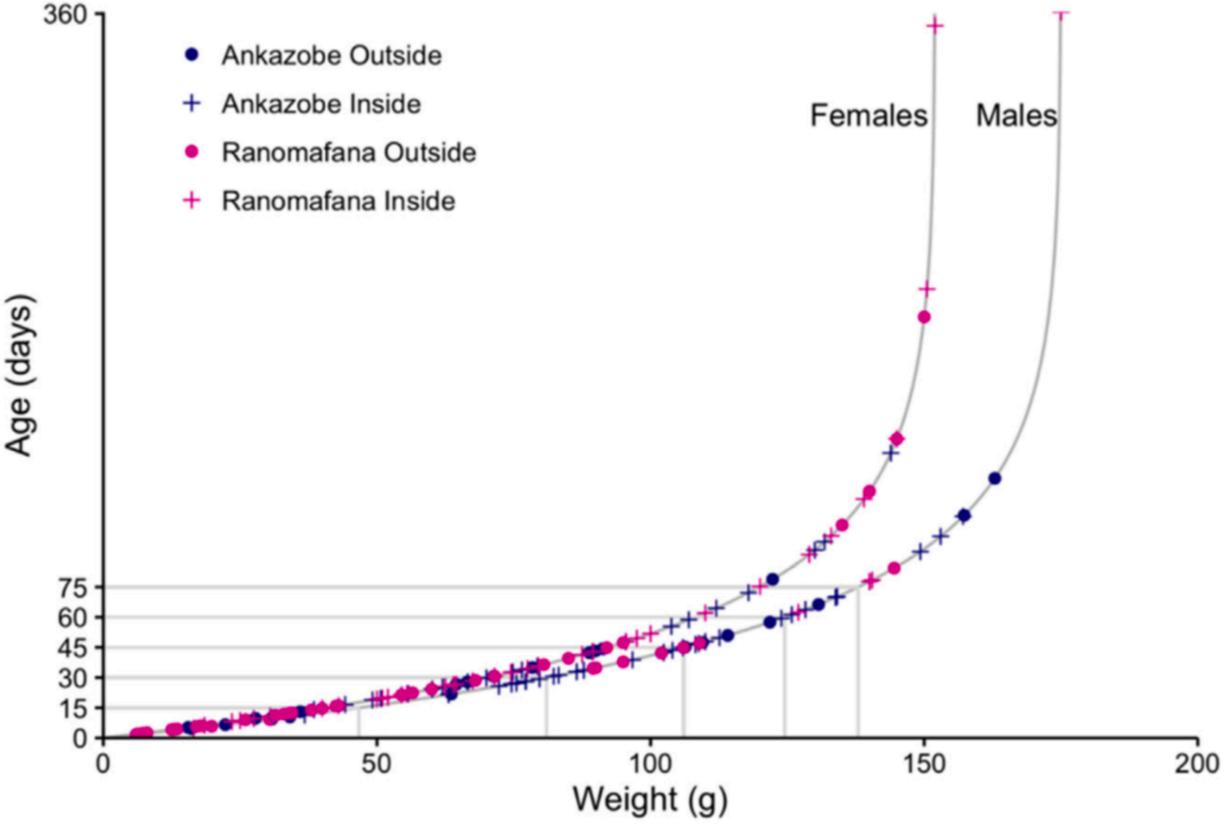
Bartonella spp.

- persistent erythrocyctic bacteria that are sometimes zoonotic
- vectored by ticks, fleas, sand flies, mosquitoes
- at least 8 human-infecting species
  - *Bartonella bacilliformis* = Carrion’s disease
  - *Bartonella henselae* = cat scratch fever
  - *Bartonella quintana* = trench fever

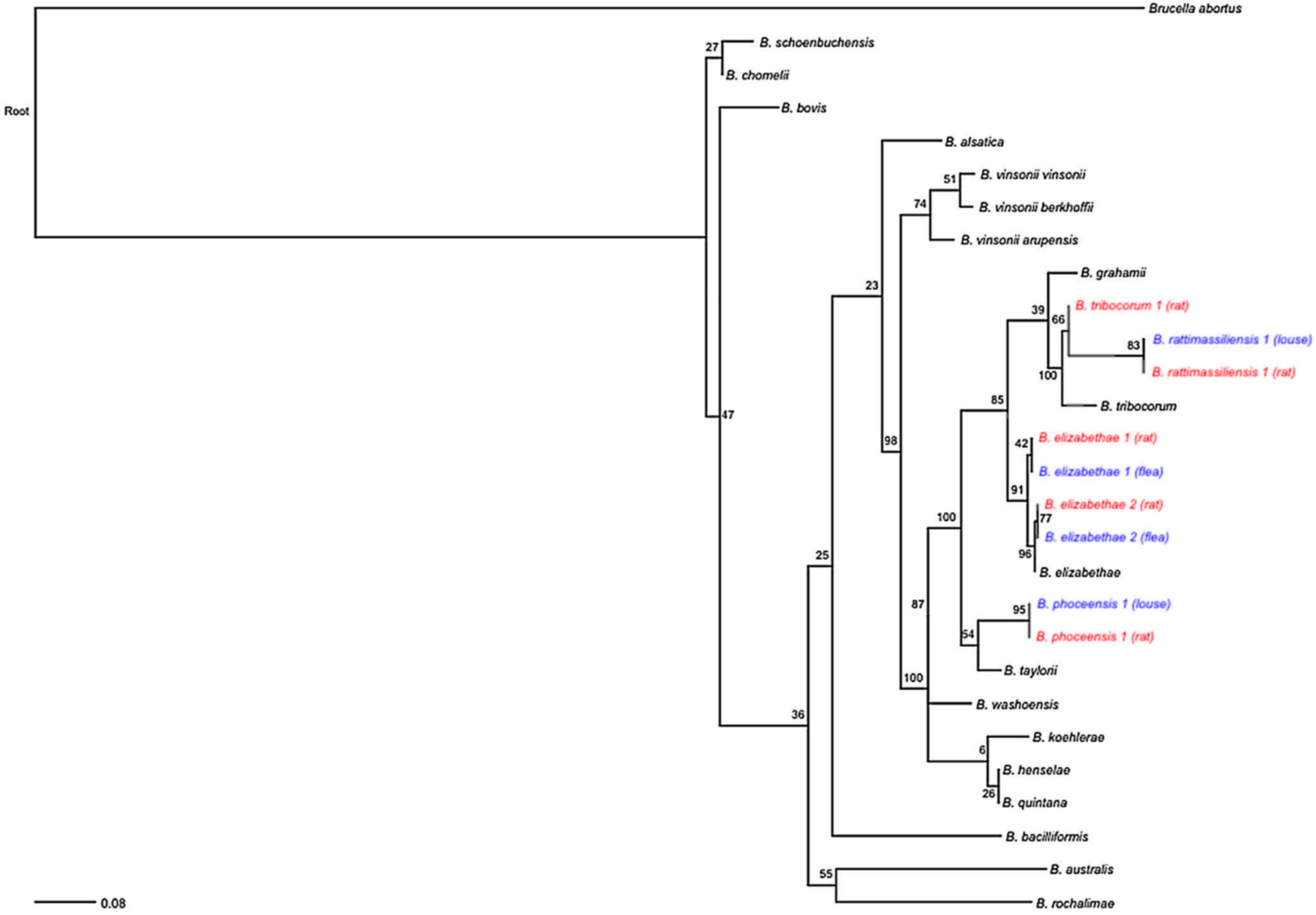


We first collected samples from rats in two sites Madagascar.

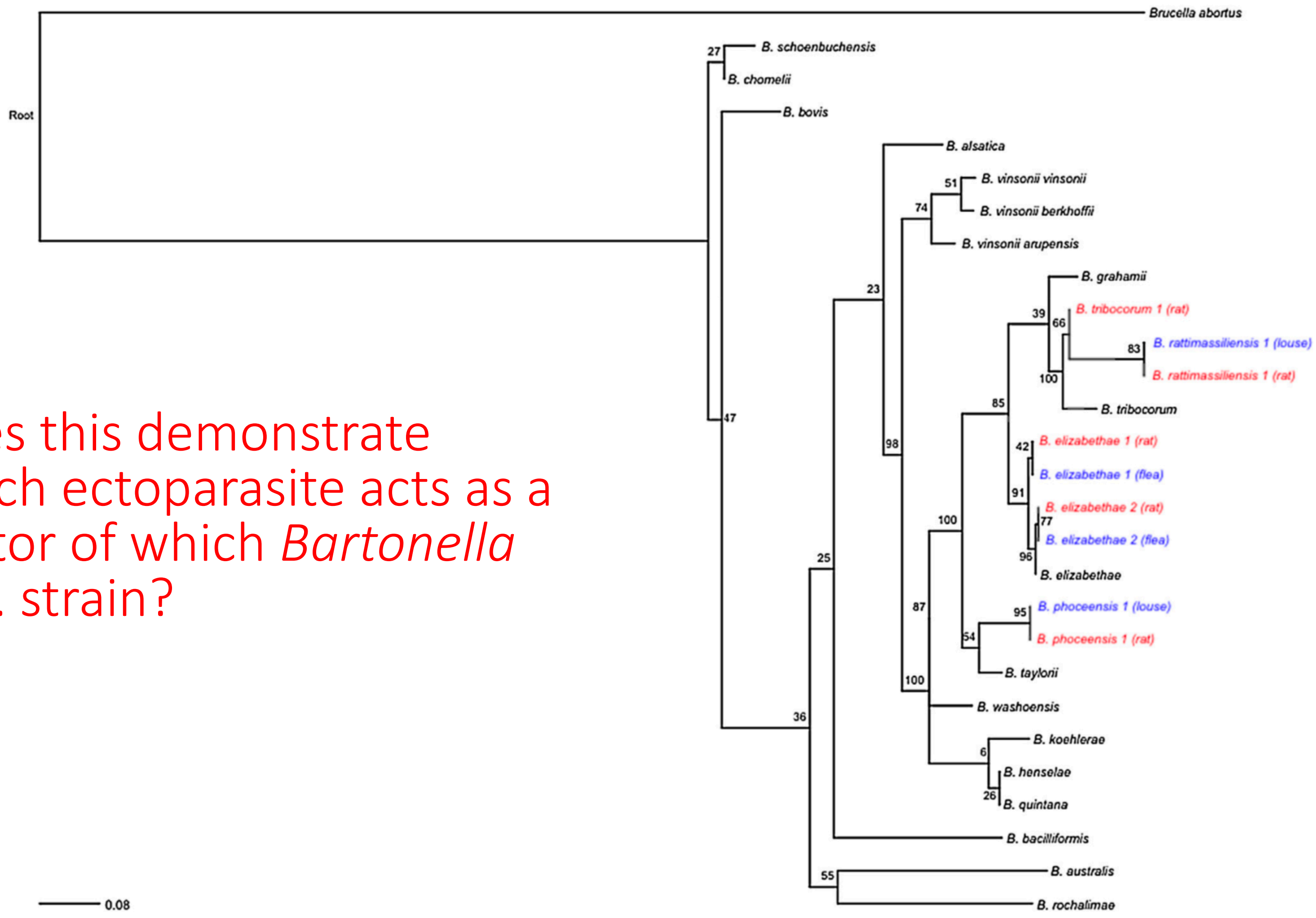




Statistically, we demonstrated an association between genotypes of *Bartonella* spp. found in rats and their ectoparasites.



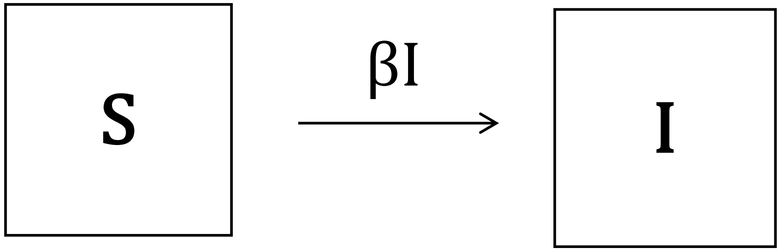
Statistically, we demonstrated an association between genotypes of *Bartonella* spp. found in rats and their ectoparasites.



Does this demonstrate which ectoparasite acts as a vector of which *Bartonella* spp. strain?

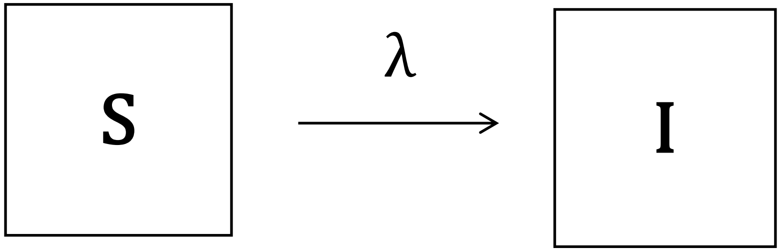
Then, we asked: *How does the rate of becoming infected vary with age?*

Age-prevalence data allows for powerful inference into the dynamics of pathogen transmission.



for a persistent, non-immunizing infection

Age-prevalence data allows for powerful inference into the dynamics of pathogen transmission.

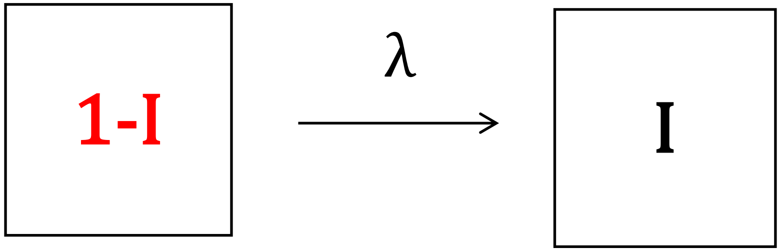


$$S \xrightarrow{\lambda} I$$

where λ, the force of infection, is the per capita rate at which susceptible hosts become infected

Age-prevalence data allows for powerful inference into the dynamics of pathogen transmission.

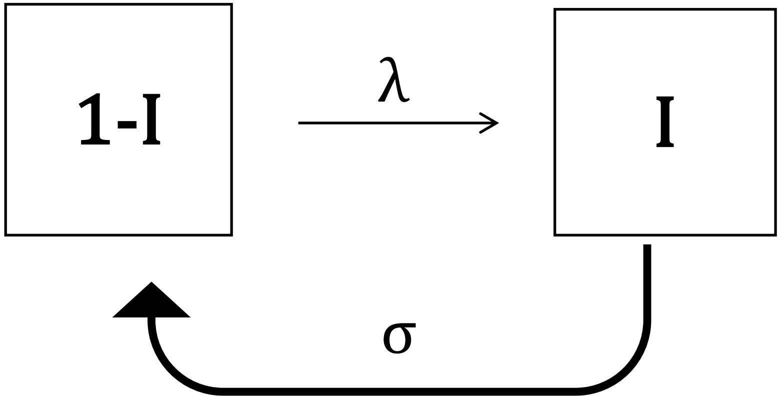
with a persistent infection, we can assume that, if not infected, you must be susceptible….



$$1-I \xrightarrow{\lambda} I$$

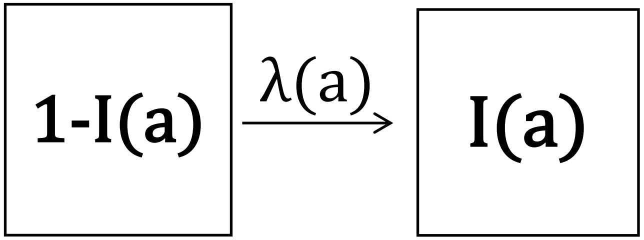
where λ, the force of infection, is the per capita rate at which susceptible hosts become infected

Age-prevalence data allows for powerful inference into the dynamics of pathogen transmission.

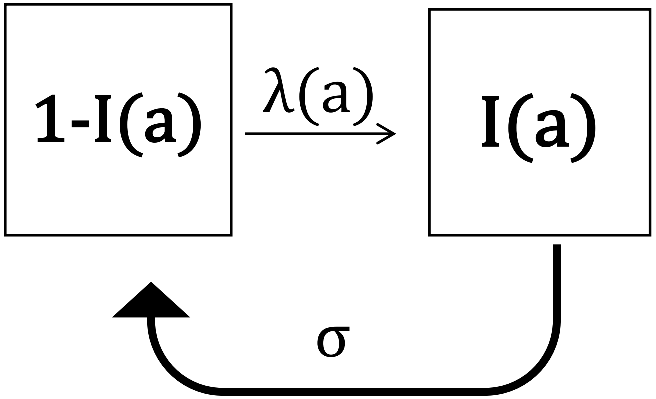


and σ is the rate of recovery from infection

Age-prevalence data allows for powerful inference into the dynamics of pathogen transmission.

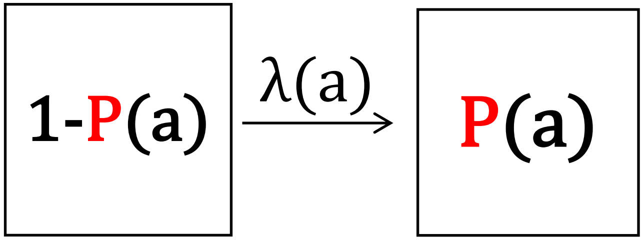


$$\frac{dI(a)}{da} = \lambda(a)\left(1 - I(a)\right)$$



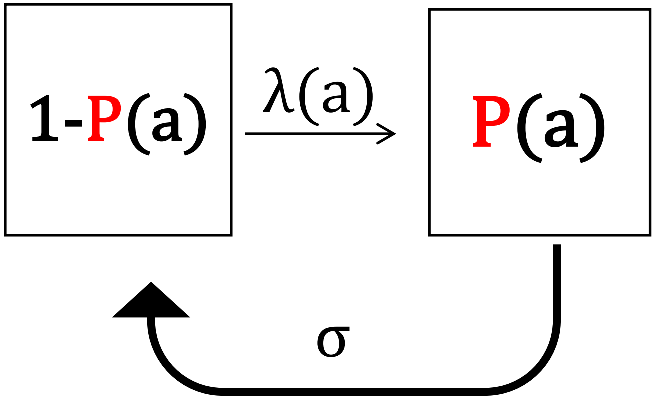
$$\frac{dI(a)}{da} = \lambda(a)\left(1 - I(a)\right) - \sigma I(a)$$

Age-prevalence data allows for powerful inference into the dynamics of pathogen transmission.



$1-P(a)$ $\xrightarrow{\lambda(a)}$ $P(a)$

$$\frac{d\,P(a)}{da} = \lambda(a)\,(1 - P(a))$$



$1-P(a)$ $\xrightarrow{\lambda(a)}$ $P(a)$

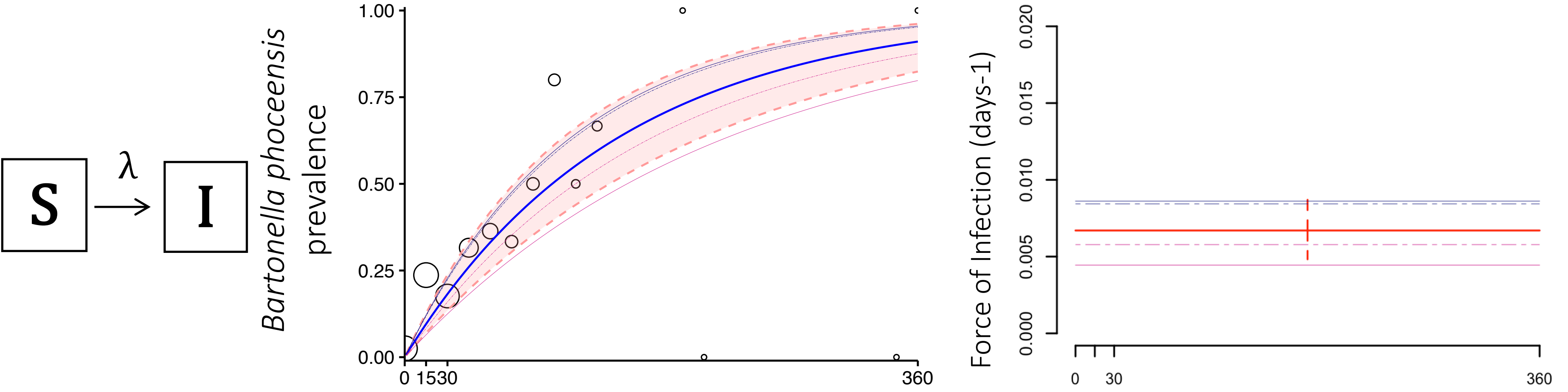
σ

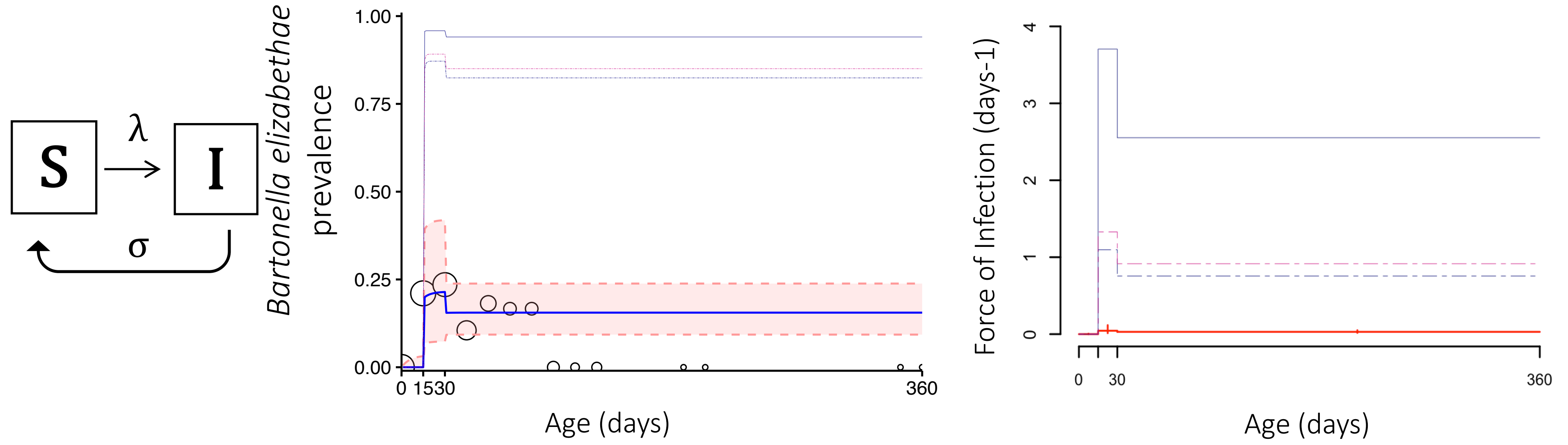
$$\frac{d\,P(a)}{da} = \lambda(a)\,(1 - P(a)) - \sigma(a)P(a))$$

similar techniques can also be applied to age-seroprevalence data for immunizing infections

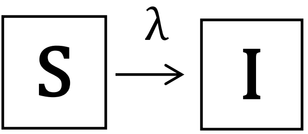
Let’s see which model works best for your data!

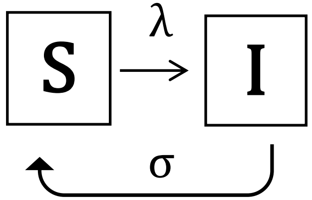
We found that an SI model offered the best fit to *B. phoceensis* data while the SIS model offered the best fit to the *B. elizabethae* data.

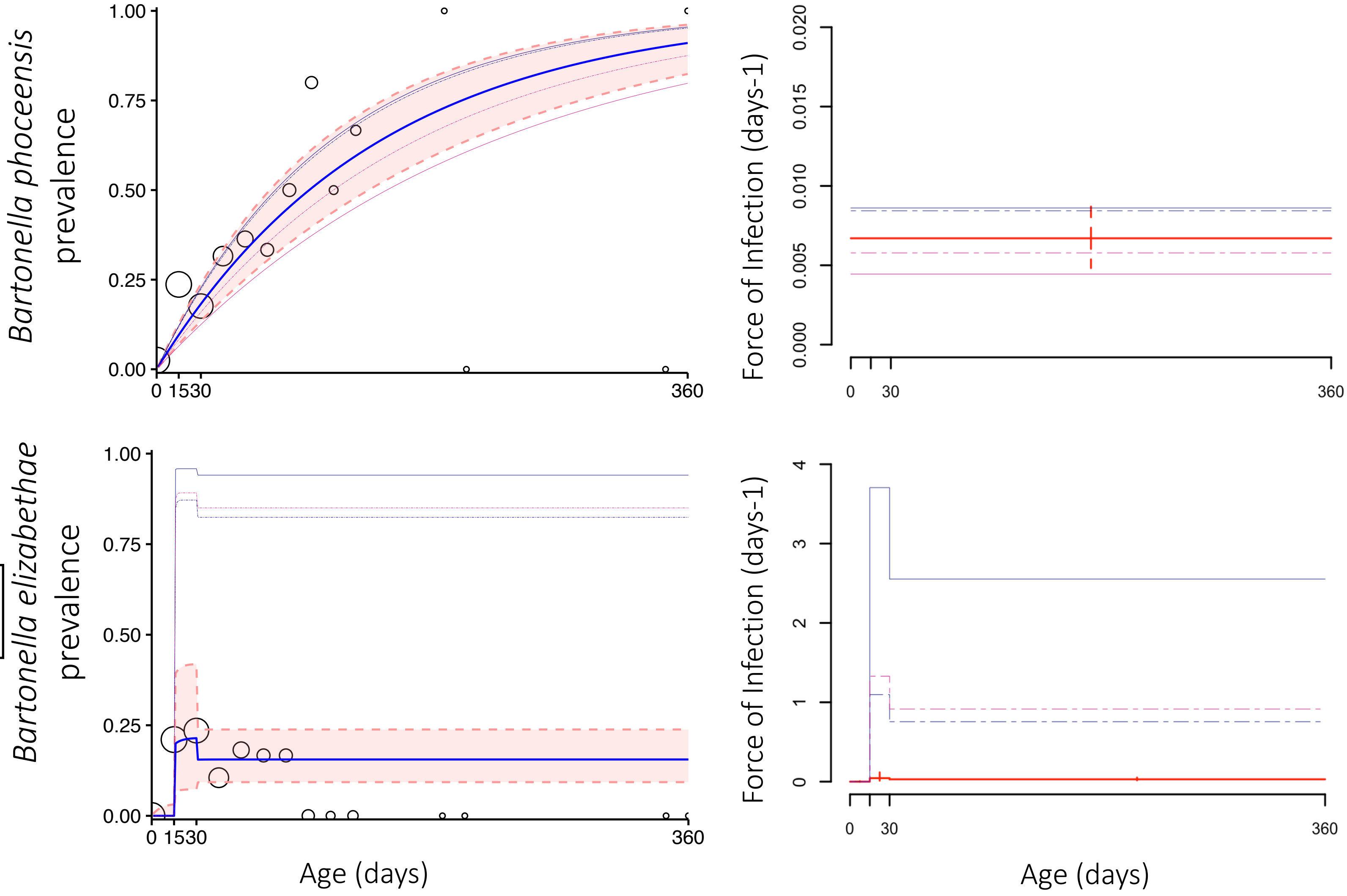


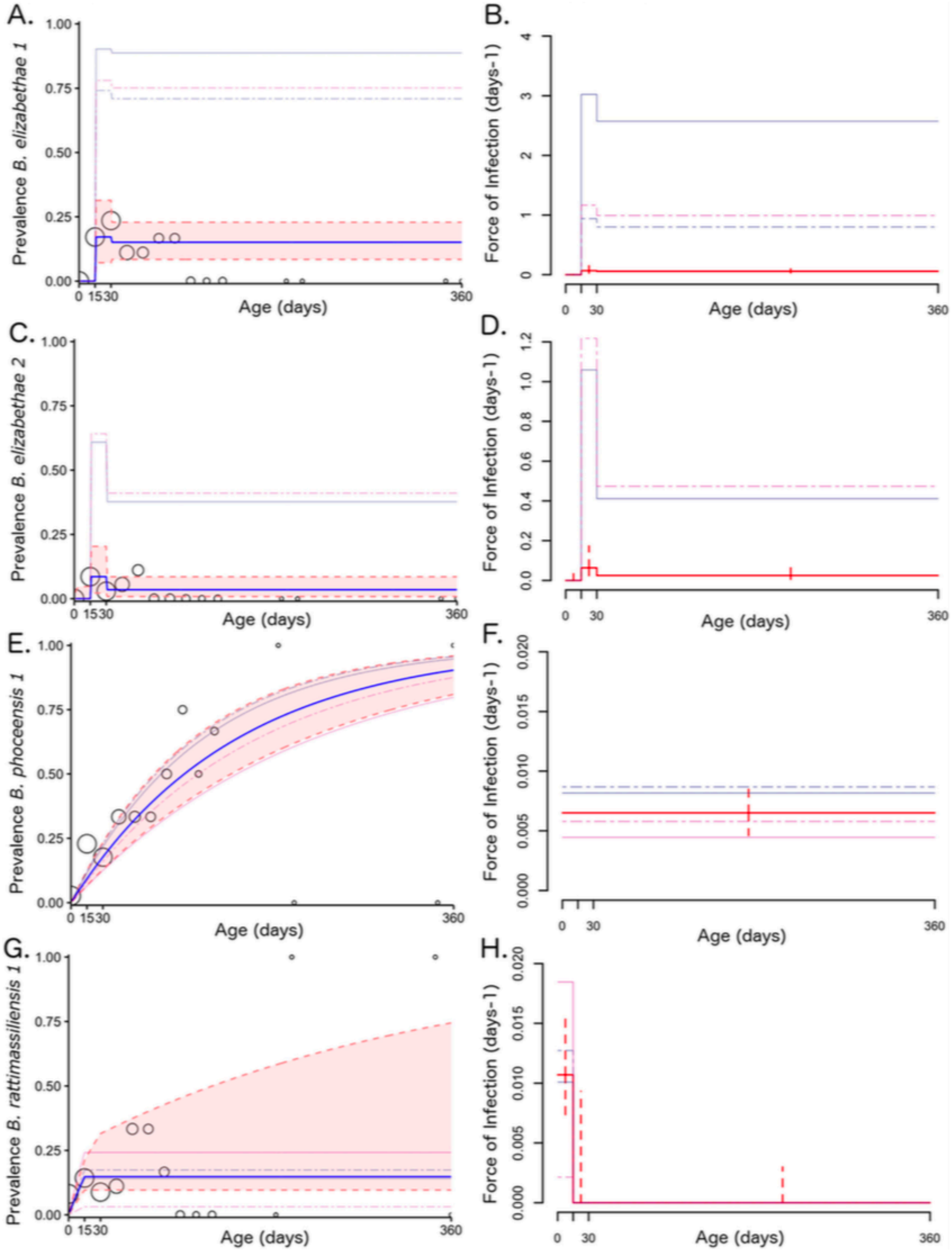


The age-structured FOI identifies age cohorts most influential in an epidemic. Juveniles showed the highest FOI.








A.
Prevalence B. elizabethae 1
1.00
0.75
0.50
0.25
0.00
0 1530
360
Age (days)
B.
Force of Infection (days-1)
4
3
2
1
0
0 30 360
Age (days)
C.
Prevalence B. elizabethae 2
1.00
0.75
0.50
0.25
0.00
0 1530
360
Age (days)
D.
Force of Infection (days-1)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0 30 360
Age (days)
E.
Prevalence B. phoceensis 1
1.00
0.75
0.50
0.25
0.00
0 1530
360
Age (days)
F.
Force of Infection (days-1)
0.020
0.015
0.010
0.005
0.000
0 30 360
Age (days)
G.
Prevalence B. rattimassiliensis 1
1.00
0.75
0.50
0.25
0.00
0 1530
360
Age (days)
H.
Force of Infection (days-1)
0.020
0.015
0.010
0.005
0.000
0 30 360
Age (days)