Detecting superspreaders in wildlife reservoirs of disease

Evandro Konzen[1,2], Richard J. Delahay[1,3], Dave J. Hodgson[1], Robbie A. McDonald[1], Ellen Brooks Pollock[4], Simon E.F. Spencer[2], Trevelyan J. McKinley[1]

[1]University of Exeter, [2]University of Warwick, [3]Animal and Plant Health Agency, [4]University of Bristol

March 2024

University of Exeter

THE UNIVERSITY OF WARWICK

Natural Environment Research Council

Bovine tuberculosis (bTB) within-species spread

- Bovine tuberculosis (bTB) has high socio-economic impacts on the cattle industry worldwide;
- In the UK and Ireland, its control it complicated by infection in a wild badger (*Meles meles*) population;
- bTB transmission between cattle and badgers has been evaluated extensively;
- There remains considerable debate about
  - $\rightarrow$ the relative impacts of **within-species spread**;
  - $\rightarrow$ the efficacy of interventions to control transmission.
- Detecting **high-risk individuals** can make control measures more efficient;
- **Superspreading** is rarely characterised in wildlife populations.

Inference: key challenges and aims

- Monitoring relies on imperfect surveillance and incomplete sampling of individuals over time;
- Imperfect diagnostic test performance;
- Key epidemiological events are only partially observed;
- Hidden processes are high dimensional.

$\rightarrow$ We fit a disease **transmission model** to individual-level data from a badger population;

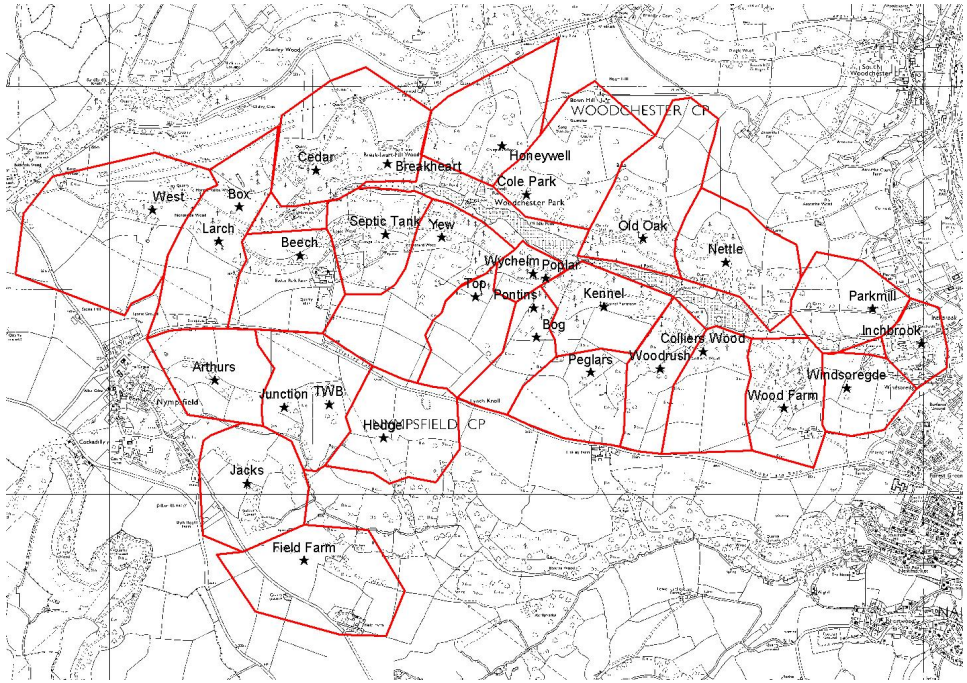
$\rightarrow$ Key **epidemiological events** are inferred alongside **model parameters**;

$\rightarrow$ Investigate the existence of **superspreading** by using a novel estimator for the individual effective reproduction number.

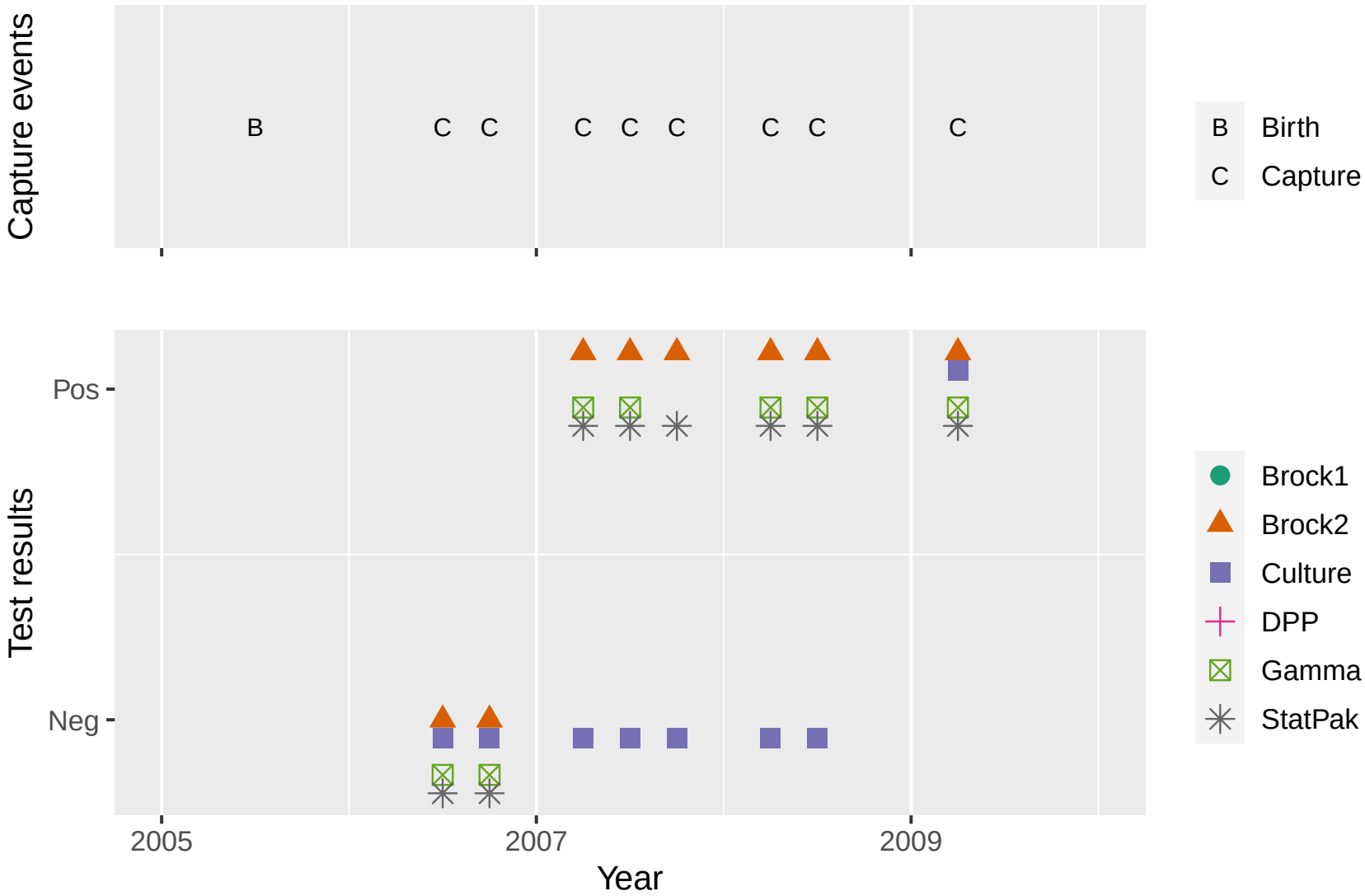
Woodchester Park data

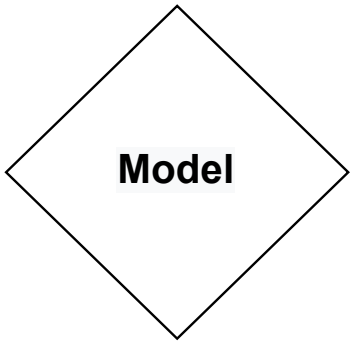
We used data from a 40-year longitudinal study (Woodchester Park, Gloucestershire, UK) (Delahay et al., 2013; Buzdugan et al., 2017):

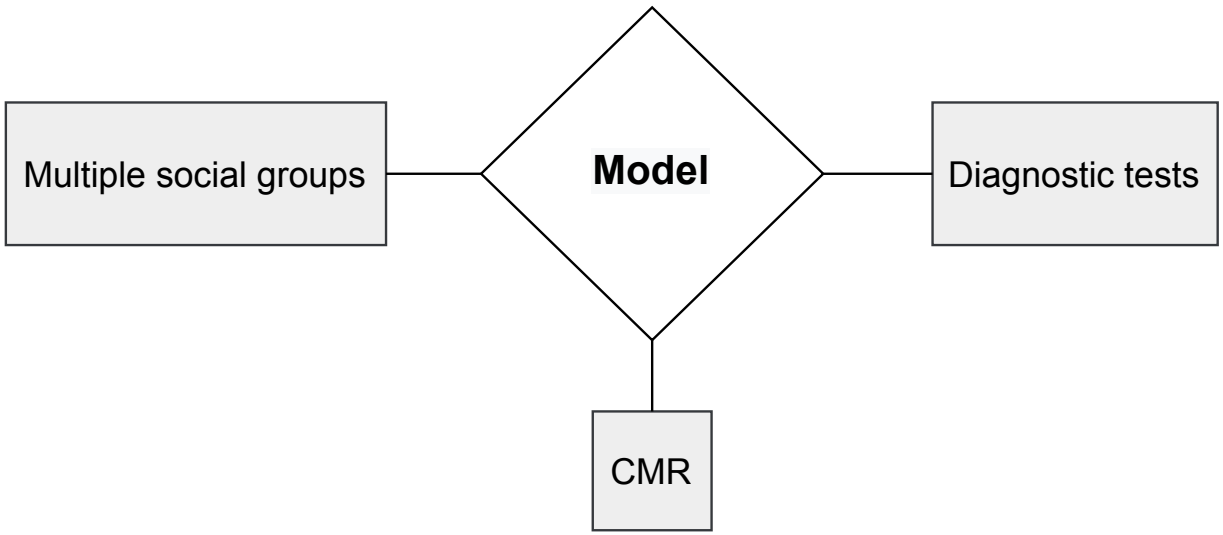
- 2,391 badgers captured in 34 social groups;
- Capture history and test results from Jan-1980 to Jan-2020;
- bTB diagnostic tests: Brock, Culture, DPP, Gamma, Stat Pak.

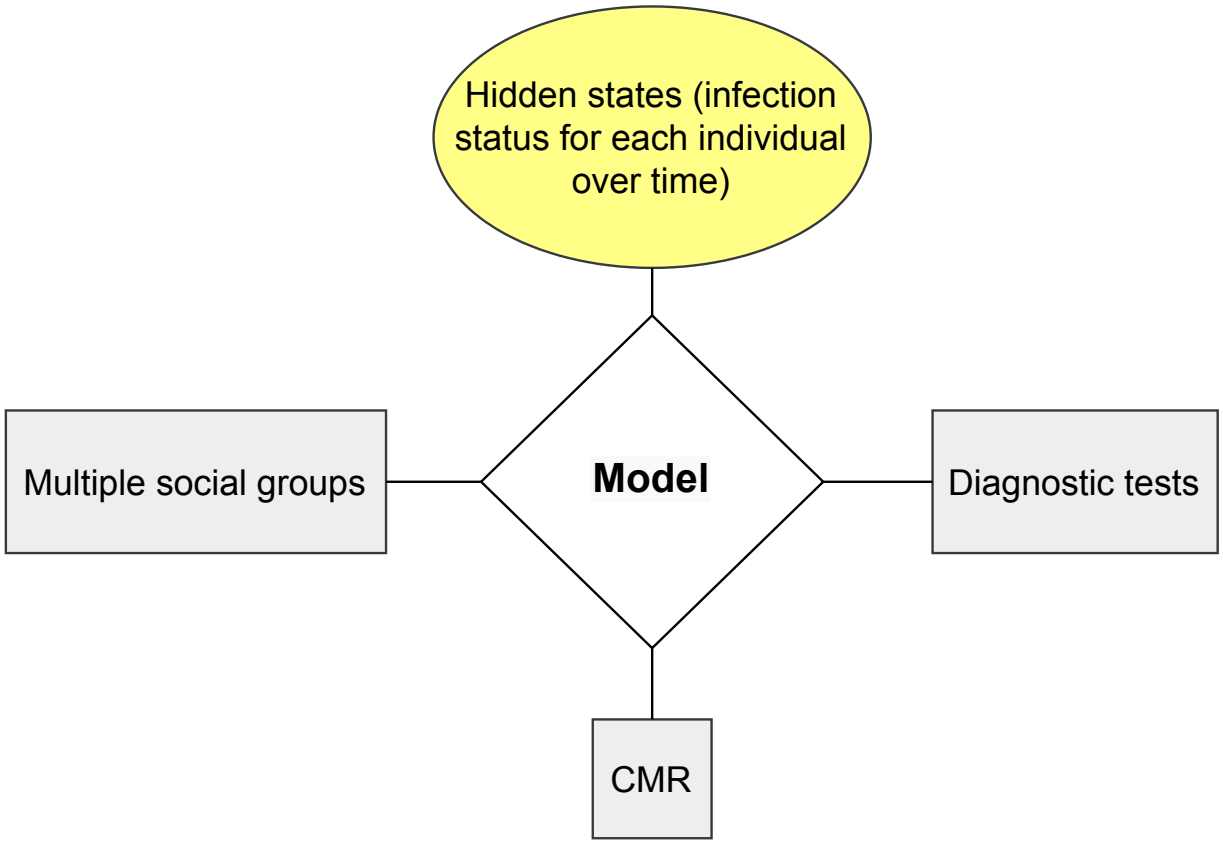


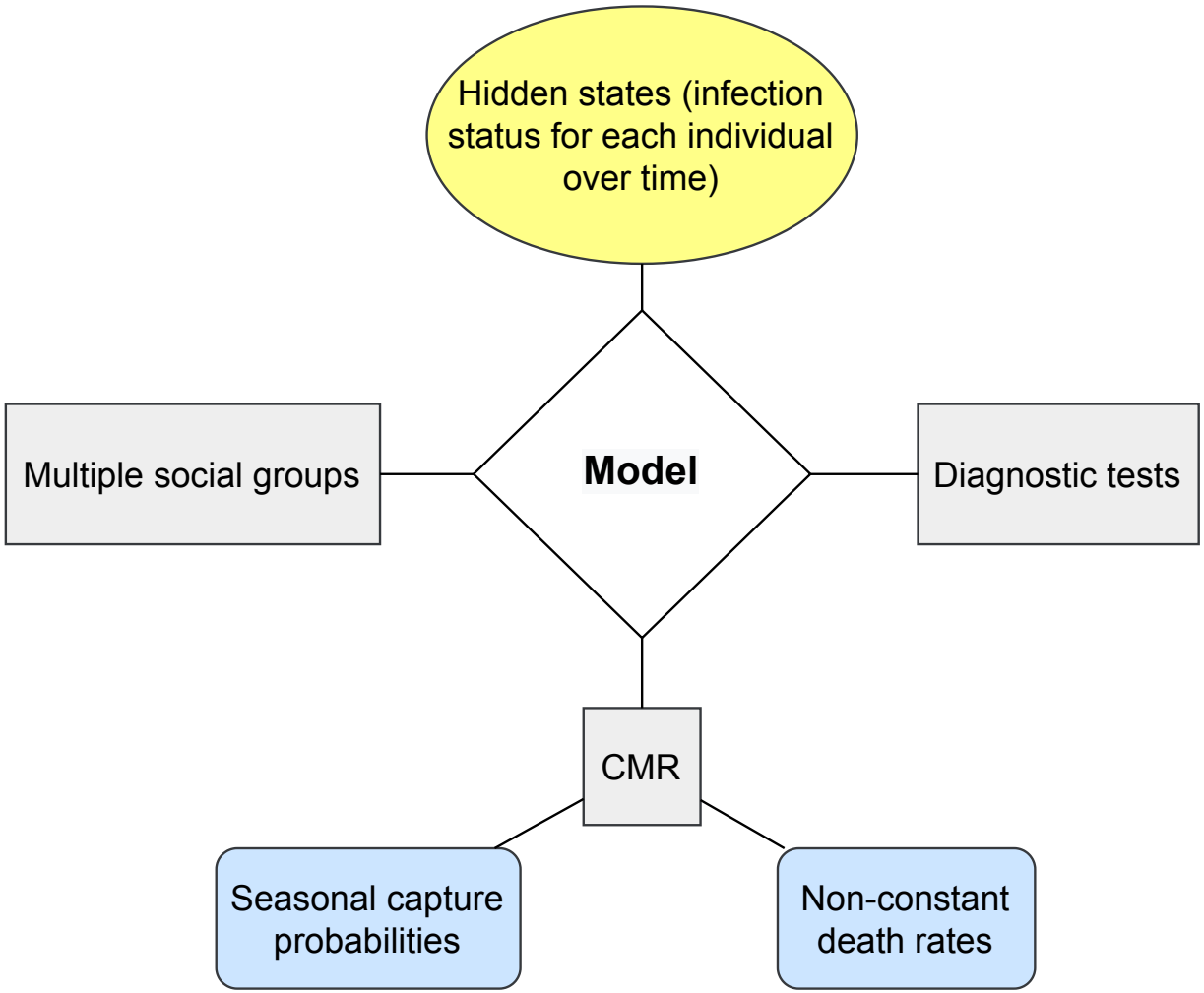
Individual-level data

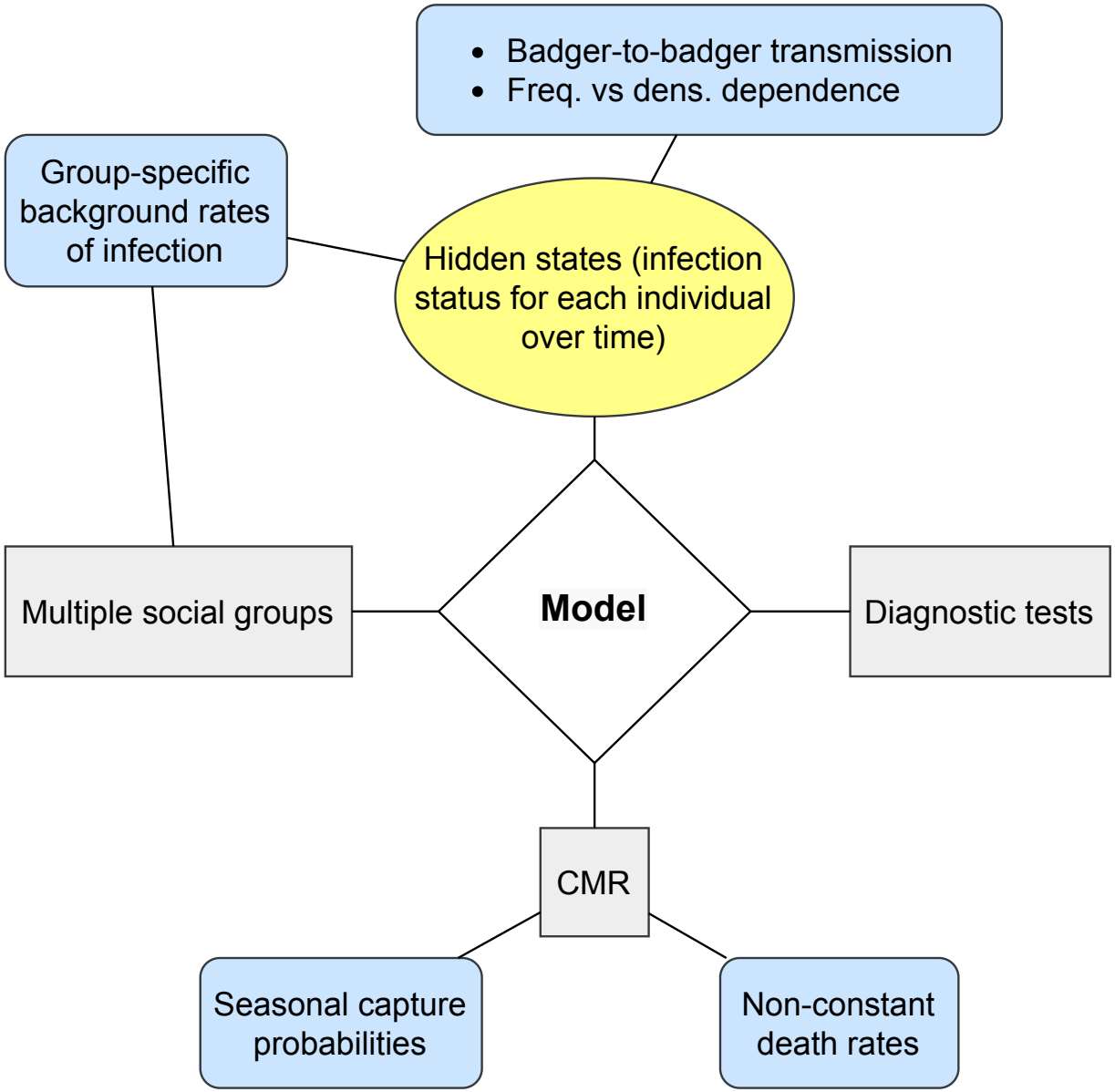


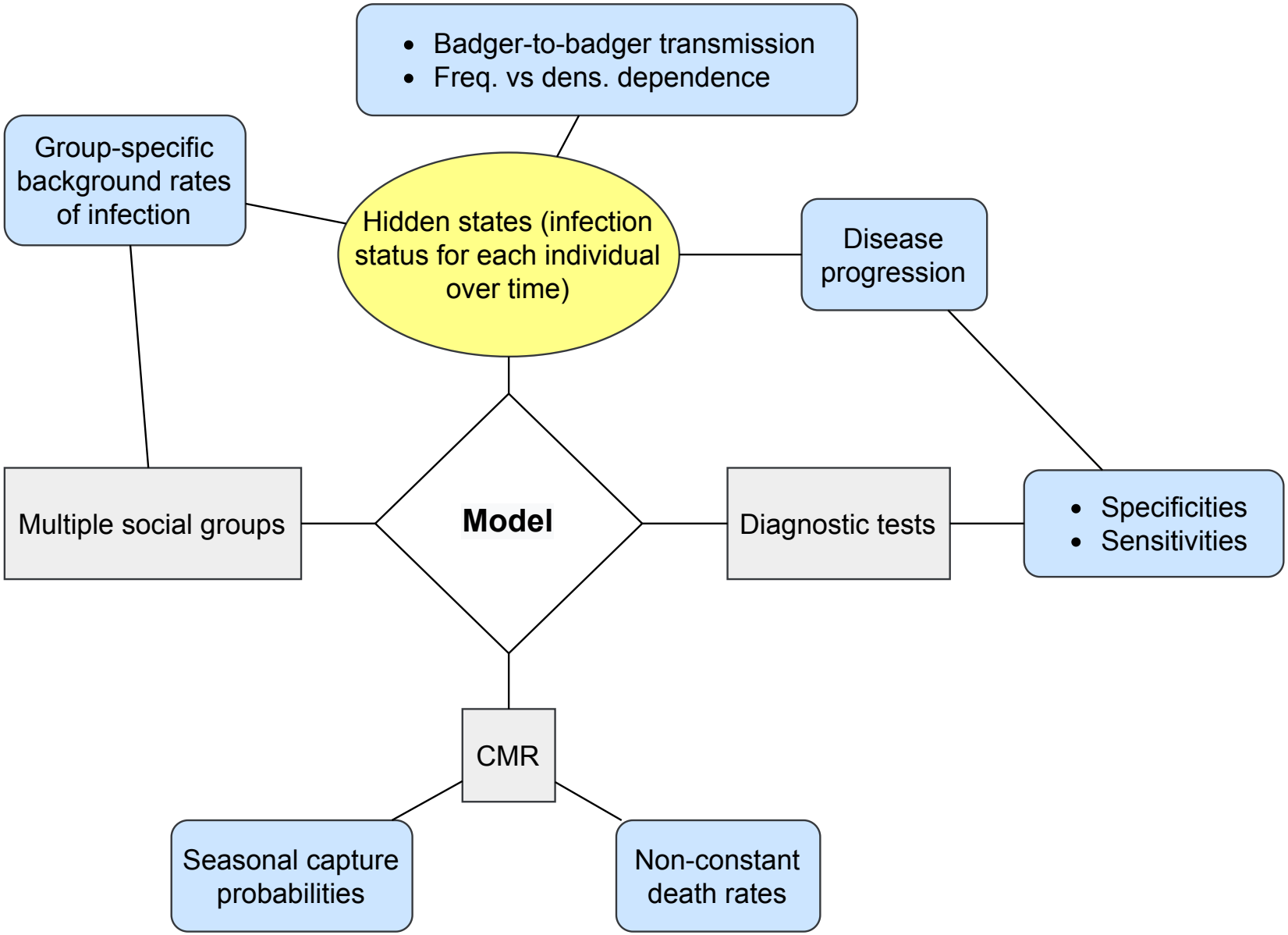

Model

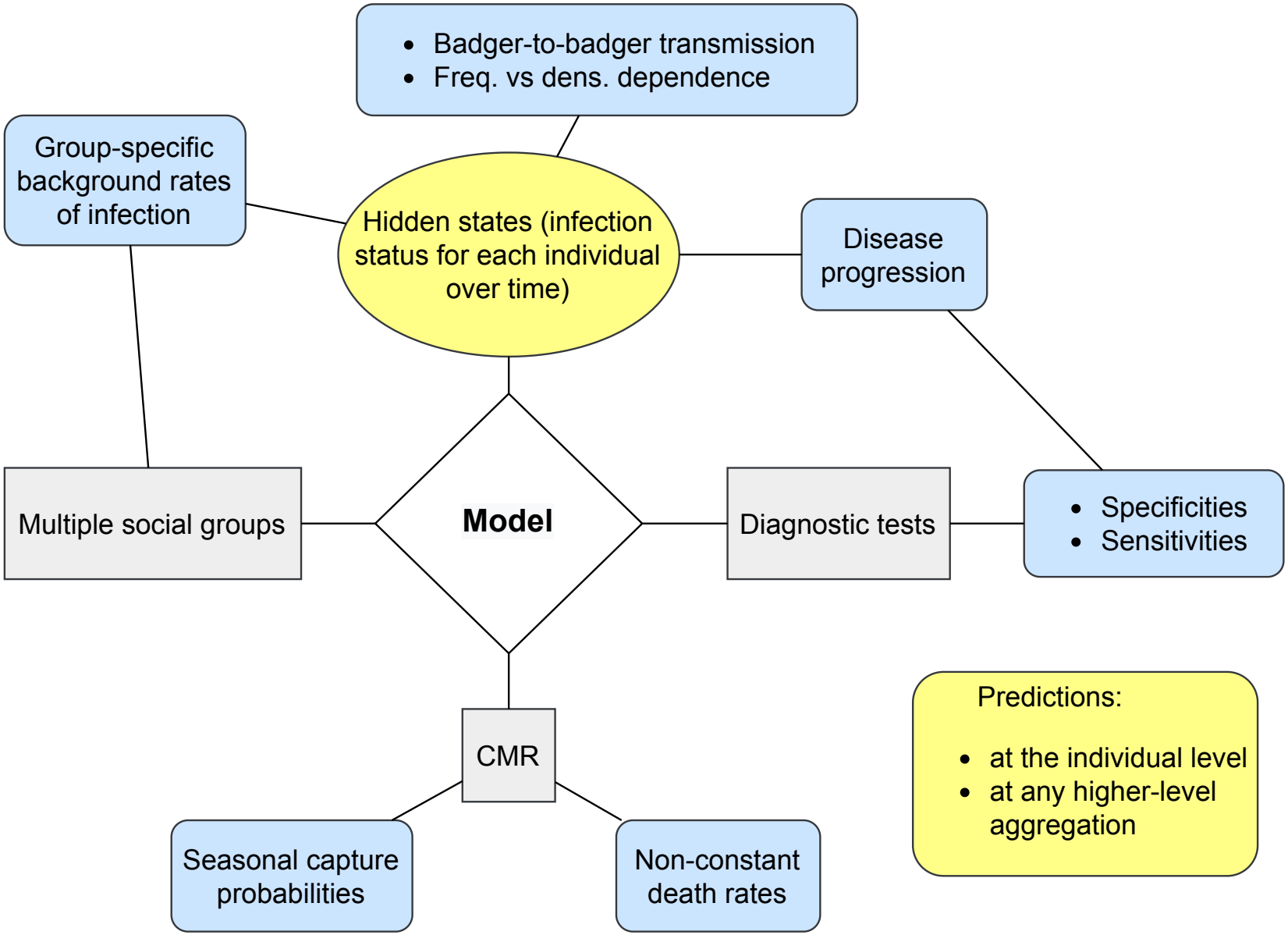

Multiple social groups
Model
Diagnostic tests
CMR


Hidden states (infection status for each individual over time)
Multiple social groups
Model
Diagnostic tests
CMR


Hidden states (infection status for each individual over time)
Multiple social groups
Model
Diagnostic tests
CMR
Seasonal capture probabilities
Non-constant death rates


• Badger-to-badger transmission
• Freq. vs dens. dependence
Group-specific background rates of infection
Hidden states (infection status for each individual over time)
Multiple social groups
Model
Diagnostic tests
CMR
Seasonal capture probabilities
Non-constant death rates


• Badger-to-badger transmission
• Freq. vs dens. dependence
Group-specific background rates of infection
Hidden states (infection status for each individual over time)
Disease progression
Multiple social groups
Model
Diagnostic tests
• Specificities
• Sensitivities
CMR
Seasonal capture probabilities
Non-constant death rates


• Badger-to-badger transmission
• Freq. vs dens. dependence
Group-specific background rates of infection
Hidden states (infection status for each individual over time)
Disease progression
Multiple social groups
Model
Diagnostic tests
• Specificities
• Sensitivities
CMR
Predictions:
• at the individual level
• at any higher-level aggregation
Seasonal capture probabilities
Non-constant death rates

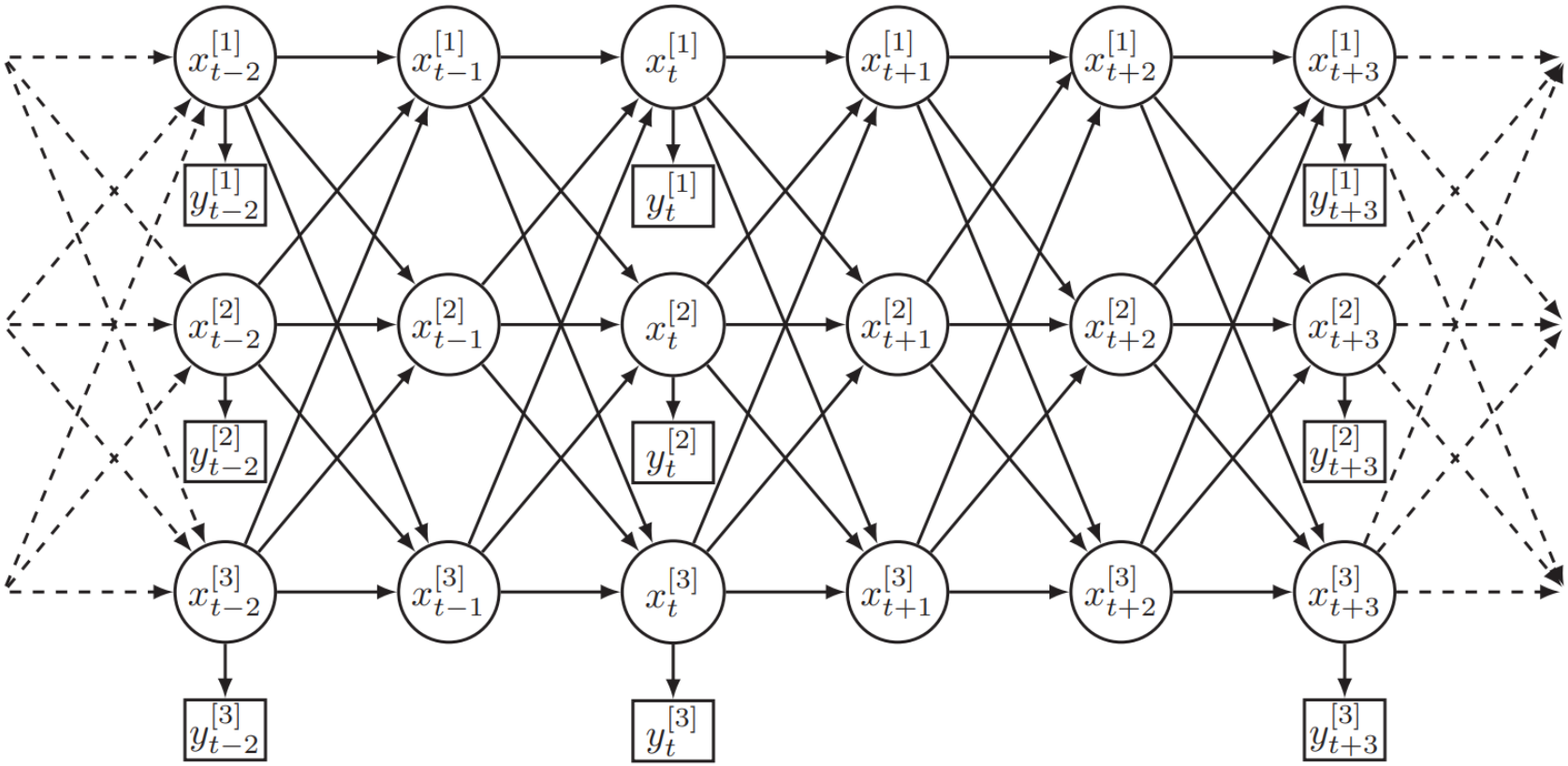
Bayesian inference of CHMM

We can view the epidemic as a coupled hidden Markov model (CHMM)

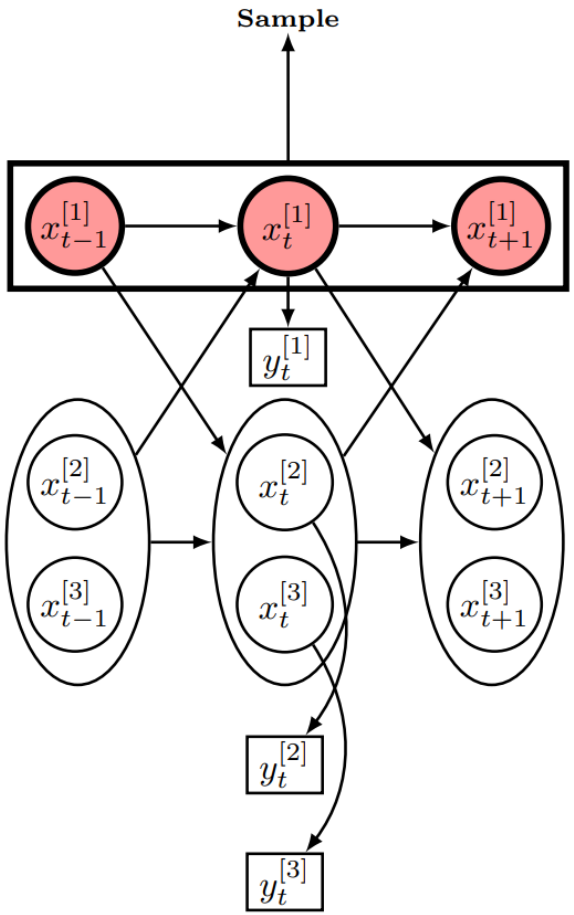
$X_t^{[i]}$: hidden state of individual $i \in \{1, \ldots, m\}$ at time $t \in \{1, \ldots, T\}$;

$Y_t^{[i]}$: diagnostic test results;

$W_t^{[i]}$: capture history.



Bayesian inference of CHMM using iFFBS[1]

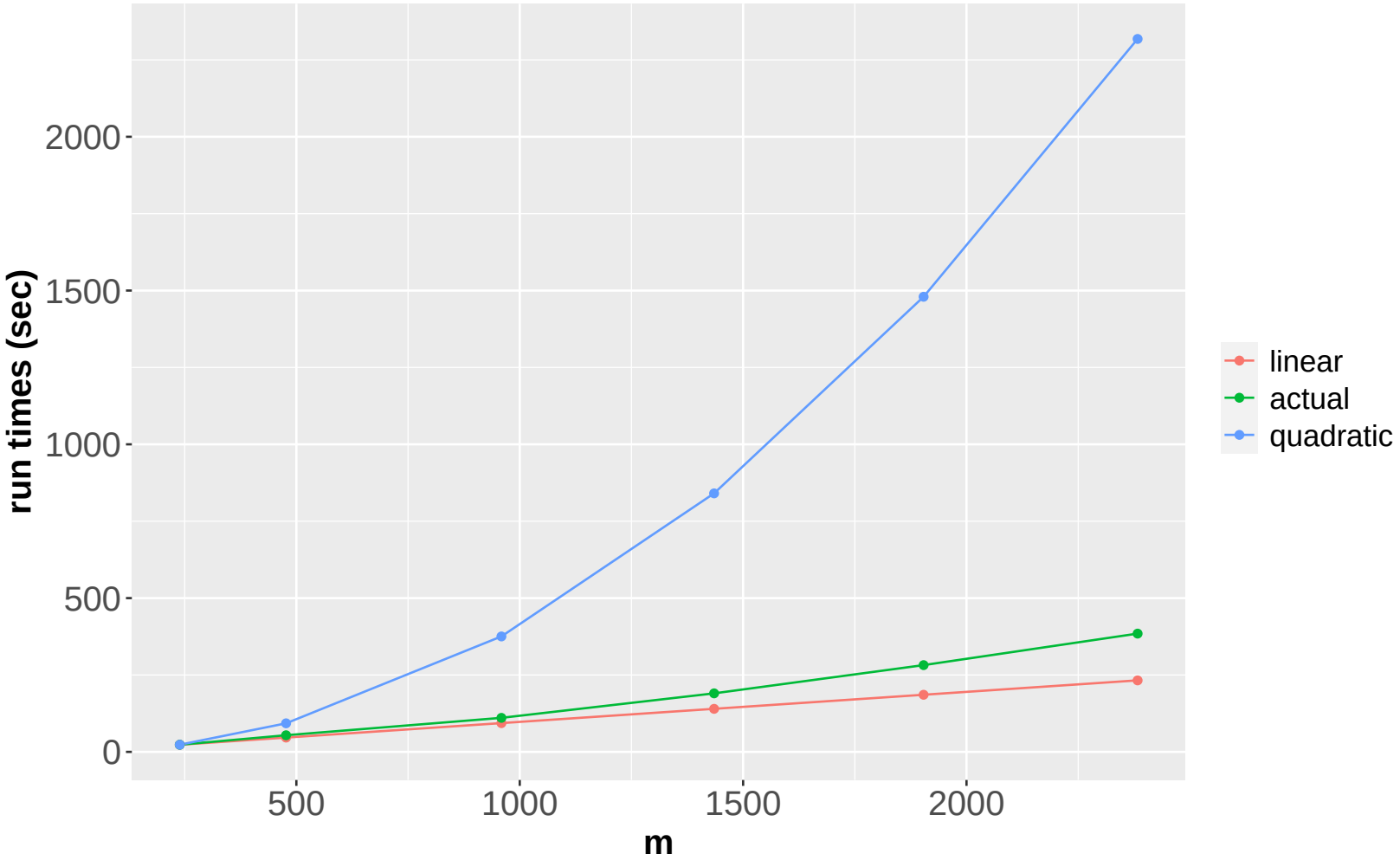


Algorithm Pseudo-algorithm of iFFBS+MCMC.

Set $r = 1$. Draw $\boldsymbol{\Theta}$ from their initial value distributions and initialise $\boldsymbol{X}_{1:T}^{[1:m]}$;
for $r = 2, \ldots, R$ **do**
  for $i = 1, \ldots, m$ **do**
    Draw $\boldsymbol{X}_{1:T}^{[i]} \sim P\left(\boldsymbol{X}_{1:T}^{[i]} \mid \boldsymbol{X}_{1:T}^{[-i]}, \boldsymbol{W}_{1:T}^{[i]}, \boldsymbol{Y}_{1:T}^{[i]}, \boldsymbol{\Theta}\right)$;
  end for
  Draw $\boldsymbol{\Theta} \sim \pi\left(\boldsymbol{\Theta} \mid \boldsymbol{X}_{1:T}^{[1:m]}, \boldsymbol{W}_{1:T}^{[1:m]}, \boldsymbol{Y}_{1:T}^{[1:m]}\right)$;
end for

[1] (Touloupou et al., 2020)

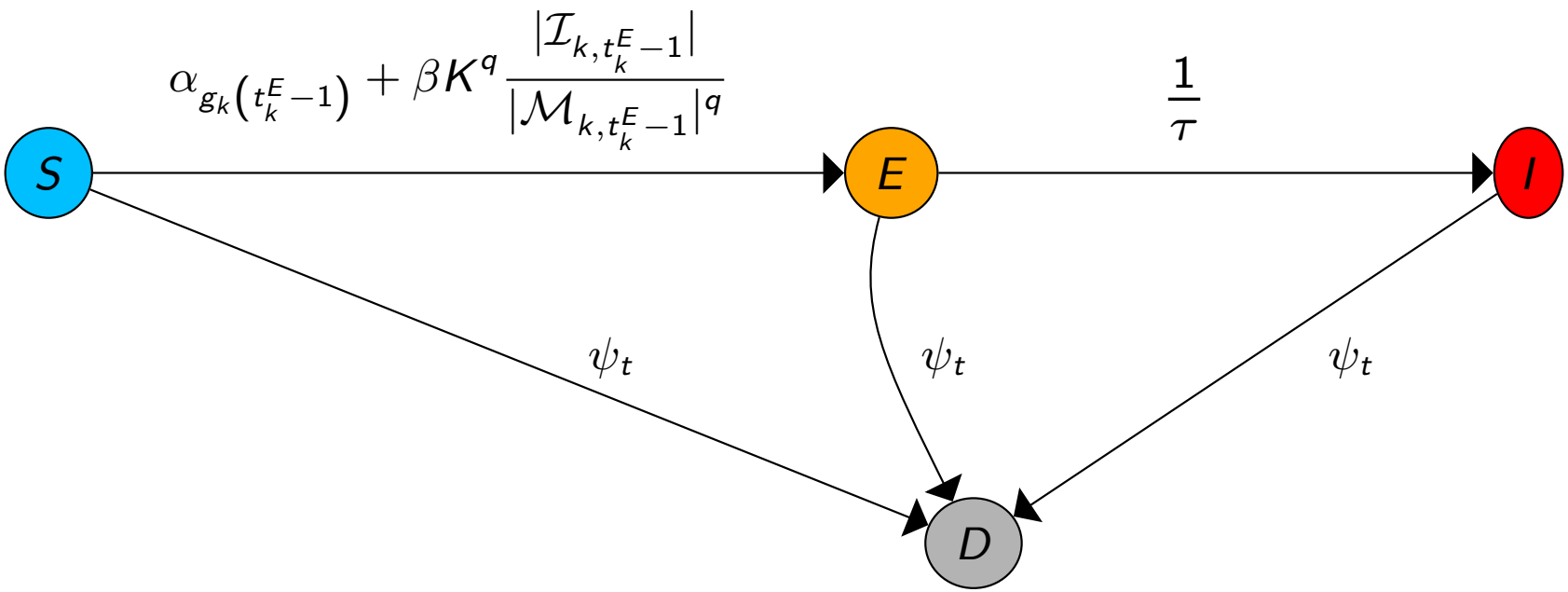
Computational complexity



iFFBS requires just:

- transition probabilities for the hidden states;
- probabilities of observing outcomes given the hidden states.

Transitions between hidden states



For modelling the death process, we used the Gompertz-Makeham hazard function

$$h(s) = c + ae^{bs}, \quad s = \text{age},$$

from which we derive the death probability $\psi_t = P\left(X_t^{[i]} = \text{dead} \;\middle|\; X_{t-1}^{[i]} = \text{alive}\right)$.

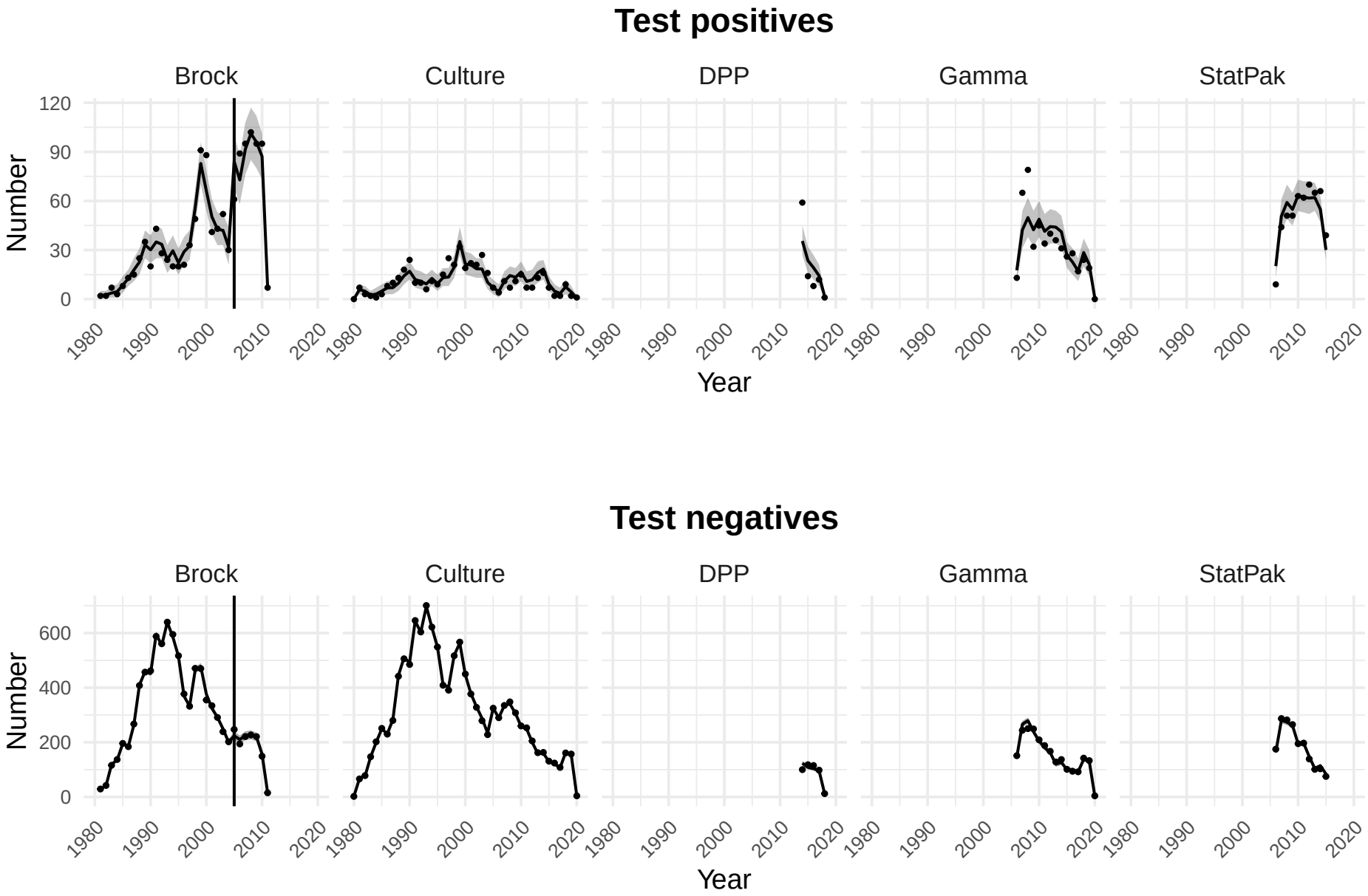
Observation processes

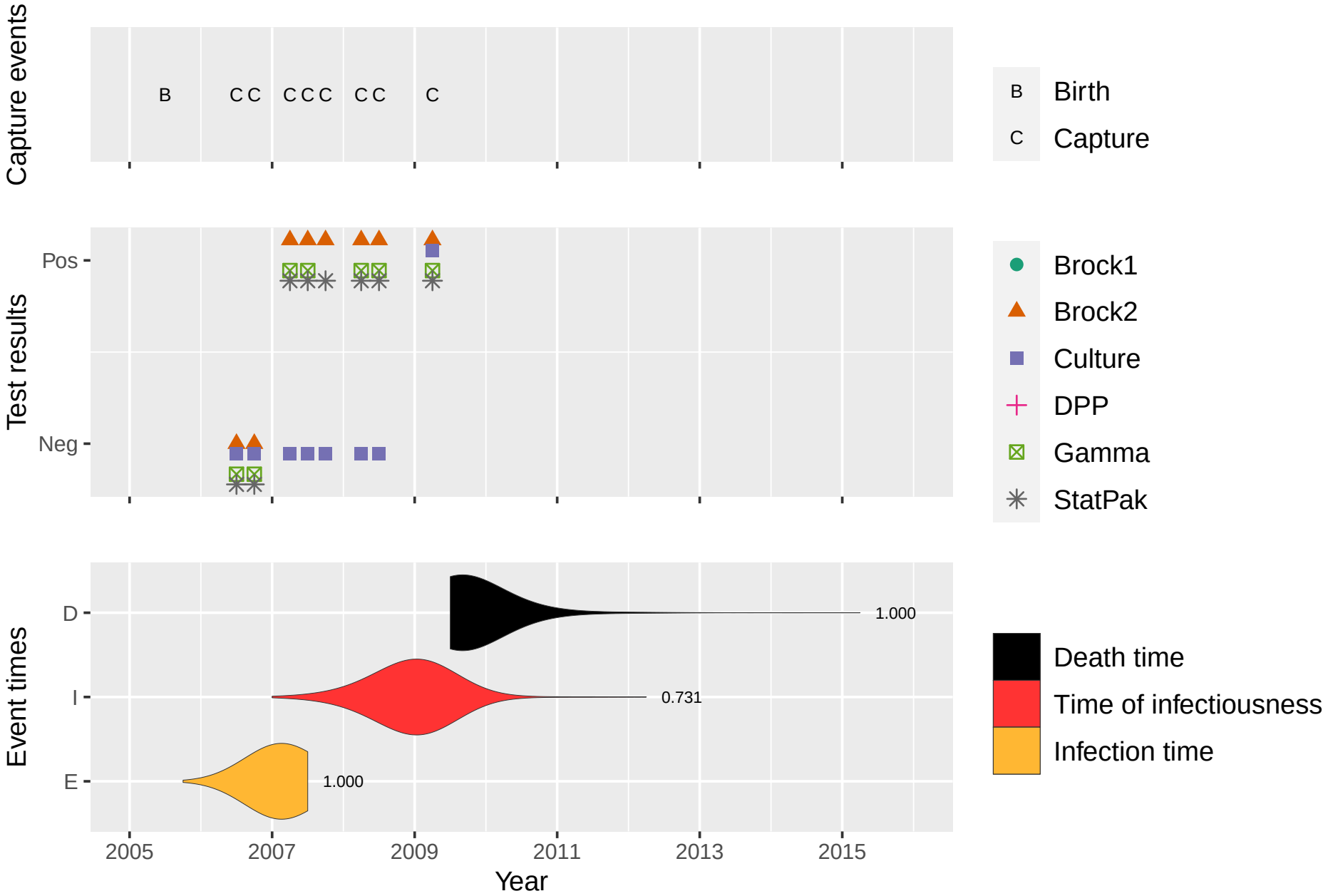
- Seasonal capture probabilities $\eta_{(t \bmod 4)}$;
- Parameters for sensitivity and specificity of each diagnostic test are inferred:

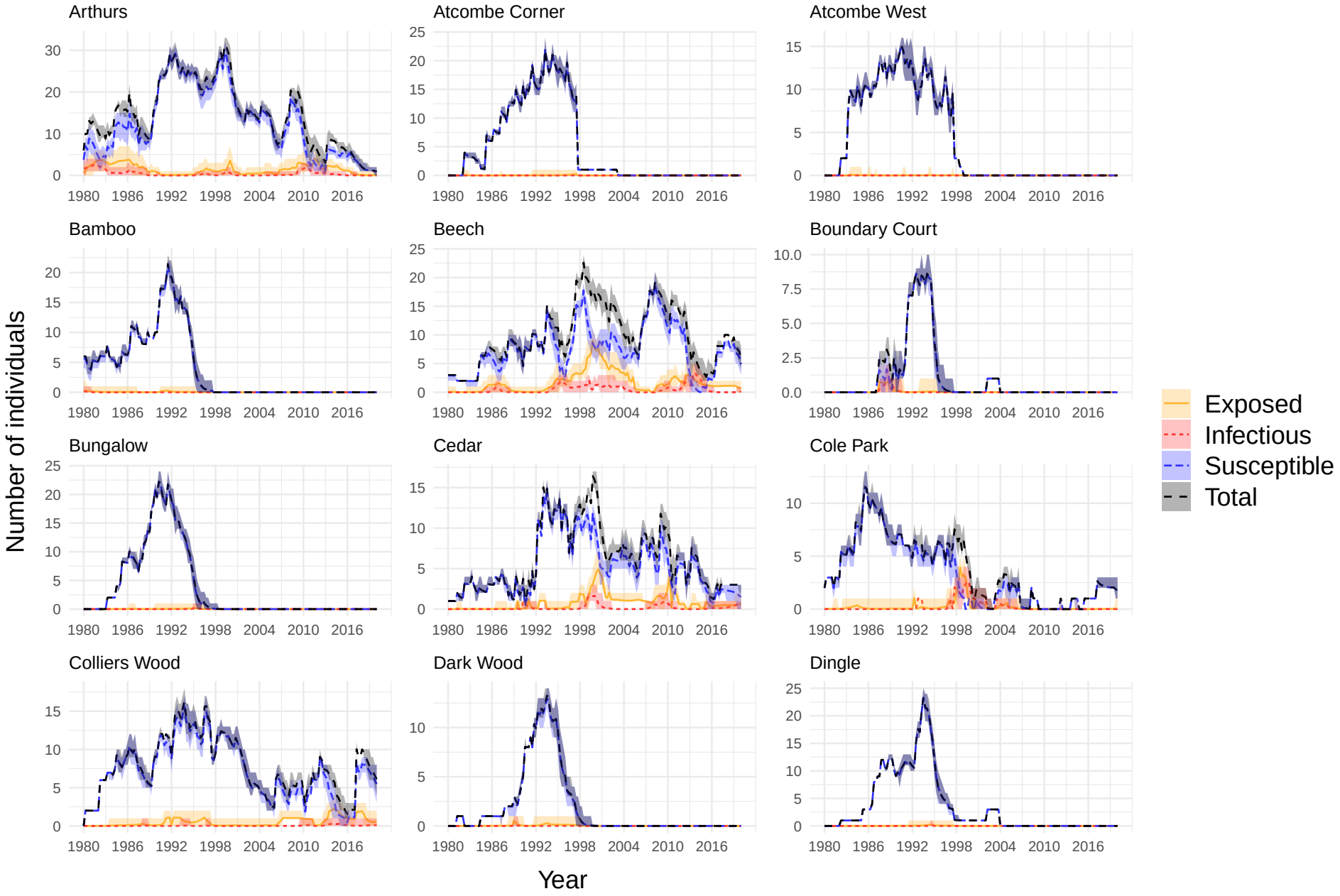
  $\rightarrow$ The sensitivity of test j for individual i at time t is given by:
  $$\theta_j = P\left(Y_{jt}^{[i]} = \text{positive} \,\middle|\, X_t^{[i]} = \textbf{infectious}\right);$$
  $$\rho_j\theta_j = P\left(Y_{jt}^{[i]} = \text{positive} \,\middle|\, X_t^{[i]} = \textbf{exposed}\right), \qquad 0 < \rho_j < 1.$$

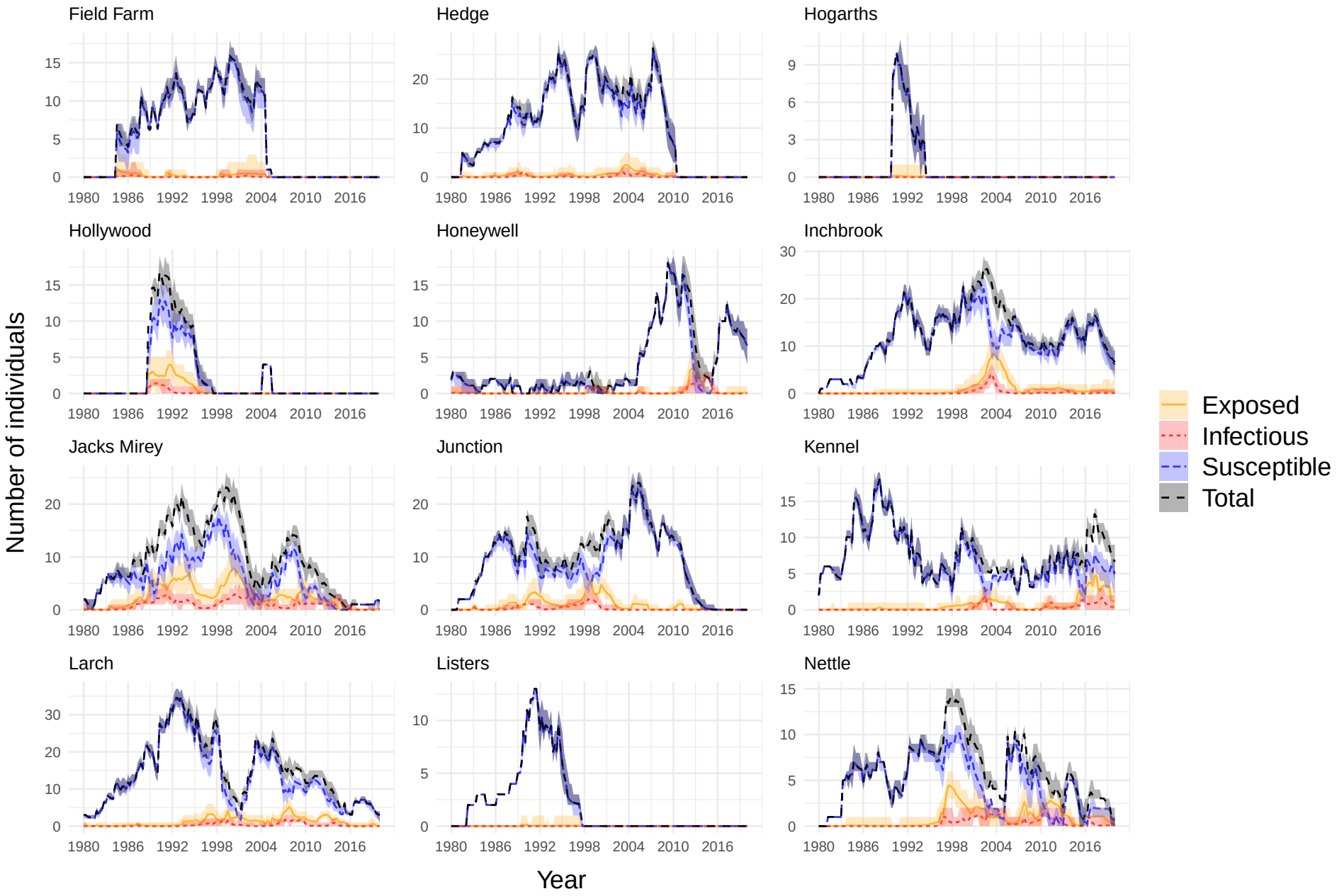
  $\rightarrow$ The specificity of test j is given by:
  $$\phi_j = P\left(Y_{jt}^{[i]} = \text{negative} \,\middle|\, X_t^{[i]} = \text{susceptible}\right).$$

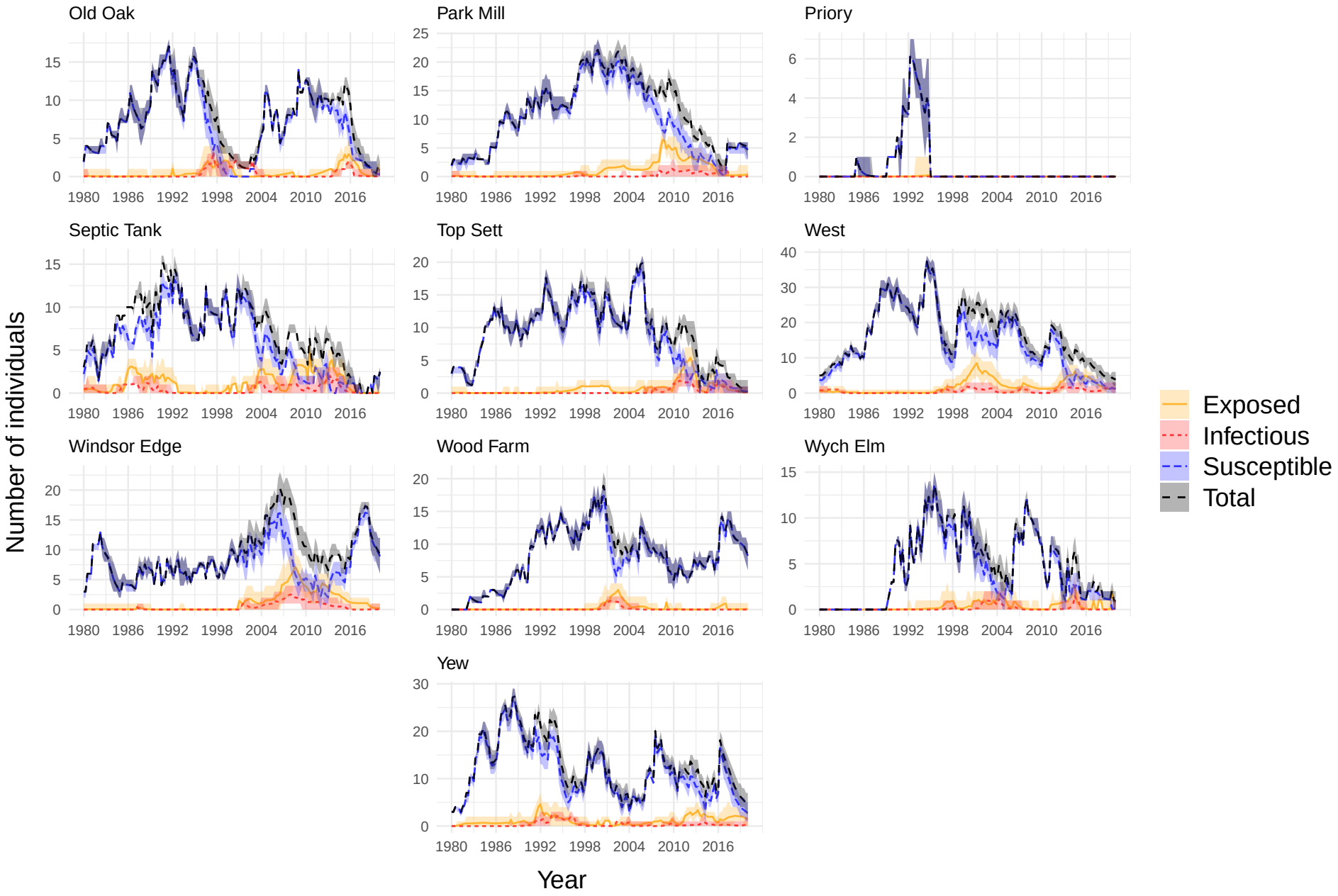
Test sensitivities are infectivity-profile dependent (Brooks-Pollock et al., 2014).

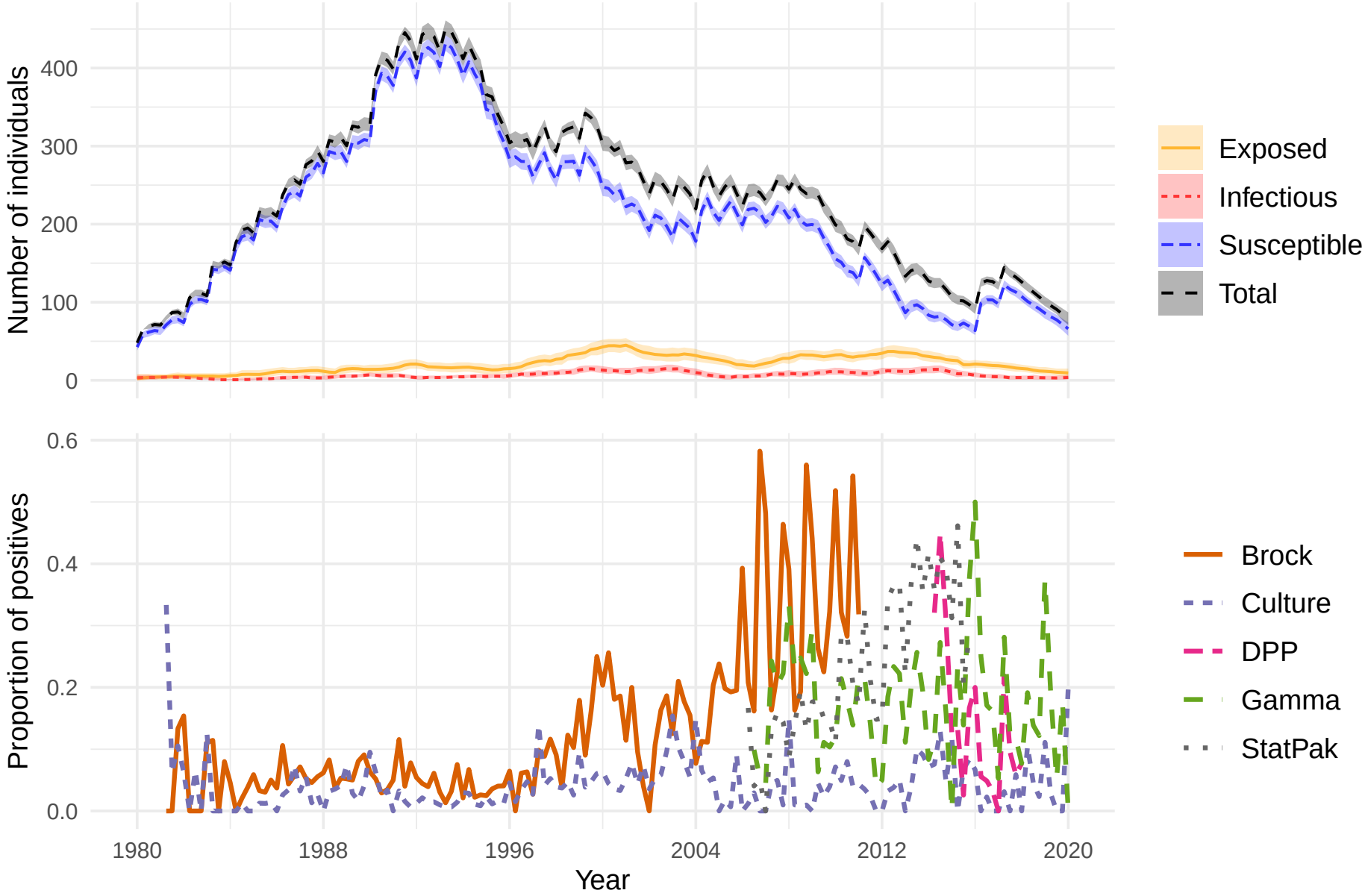

Test positives
Brock
Culture
DPP
Gamma
StatPak
Number
120
90
60
30
0
1980
1990
2000
2010
2020
Year
Test negatives
Brock
Culture
DPP
Gamma
StatPak
Number
600
400
200
0
1980
1990
2000
2010
2020
Year

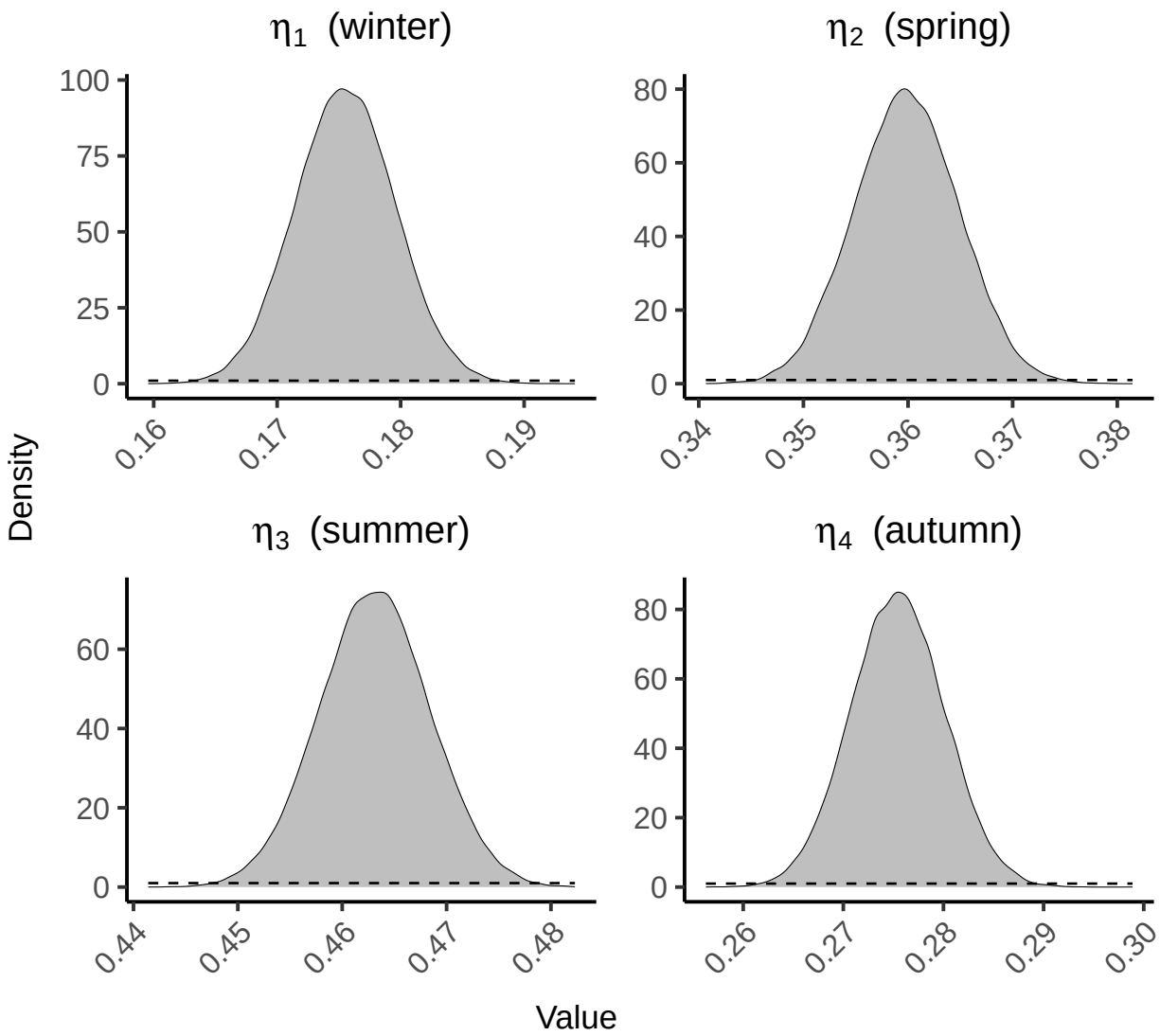

Capture events
B
C C
C C C
C C
C
Test results
Pos
Neg
Event times
D
I
E
1.000
0.731
1.000
2005
2007
2009
2011
2013
2015
Year
B Birth
C Capture
Brock1
Brock2
Culture
DPP
Gamma
StatPak
Death time
Time of infectiousness
Infection time

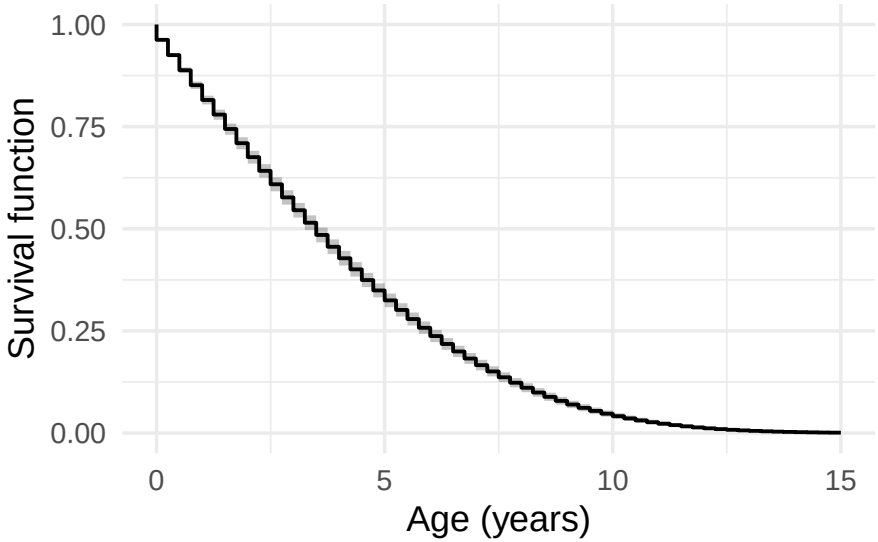

Arthurs
Atcombe Corner
Atcombe West
Bamboo
Beech
Boundary Court
Bungalow
Cedar
Cole Park
Colliers Wood
Dark Wood
Dingle
Number of individuals
Year
Exposed
Infectious
Susceptible
Total

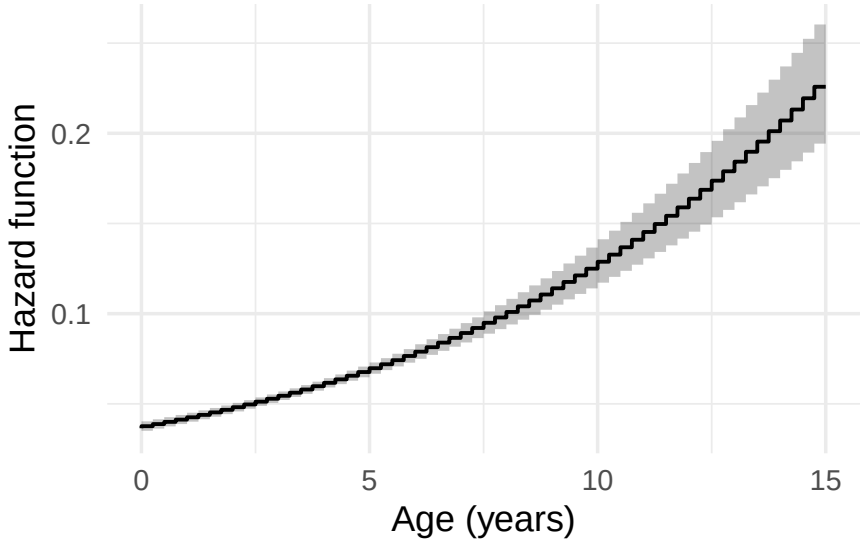

Field Farm
Hedge
Hogarths
Hollywood
Honeywell
Inchbrook
Jacks Mirey
Junction
Kennel
Larch
Listers
Nettle
Number of individuals
Year
Exposed
Infectious
Susceptible
Total

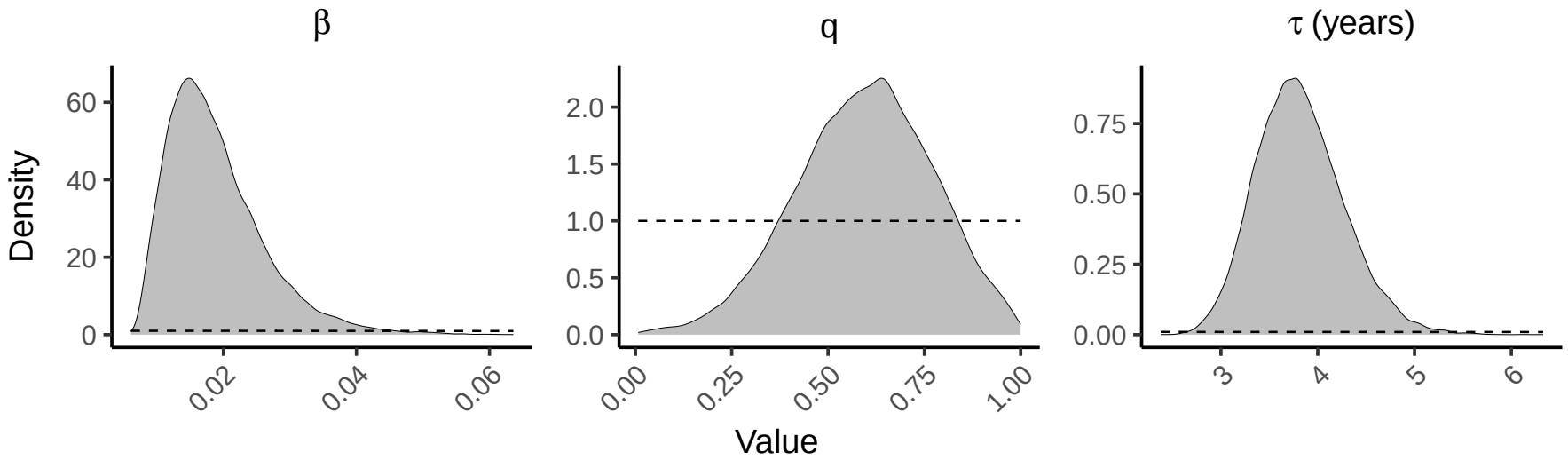

Old Oak
Park Mill
Priory
Septic Tank
Top Sett
West
Windsor Edge
Wood Farm
Wych Elm
Yew
Number of individuals
Year
Exposed
Infectious
Susceptible
Total

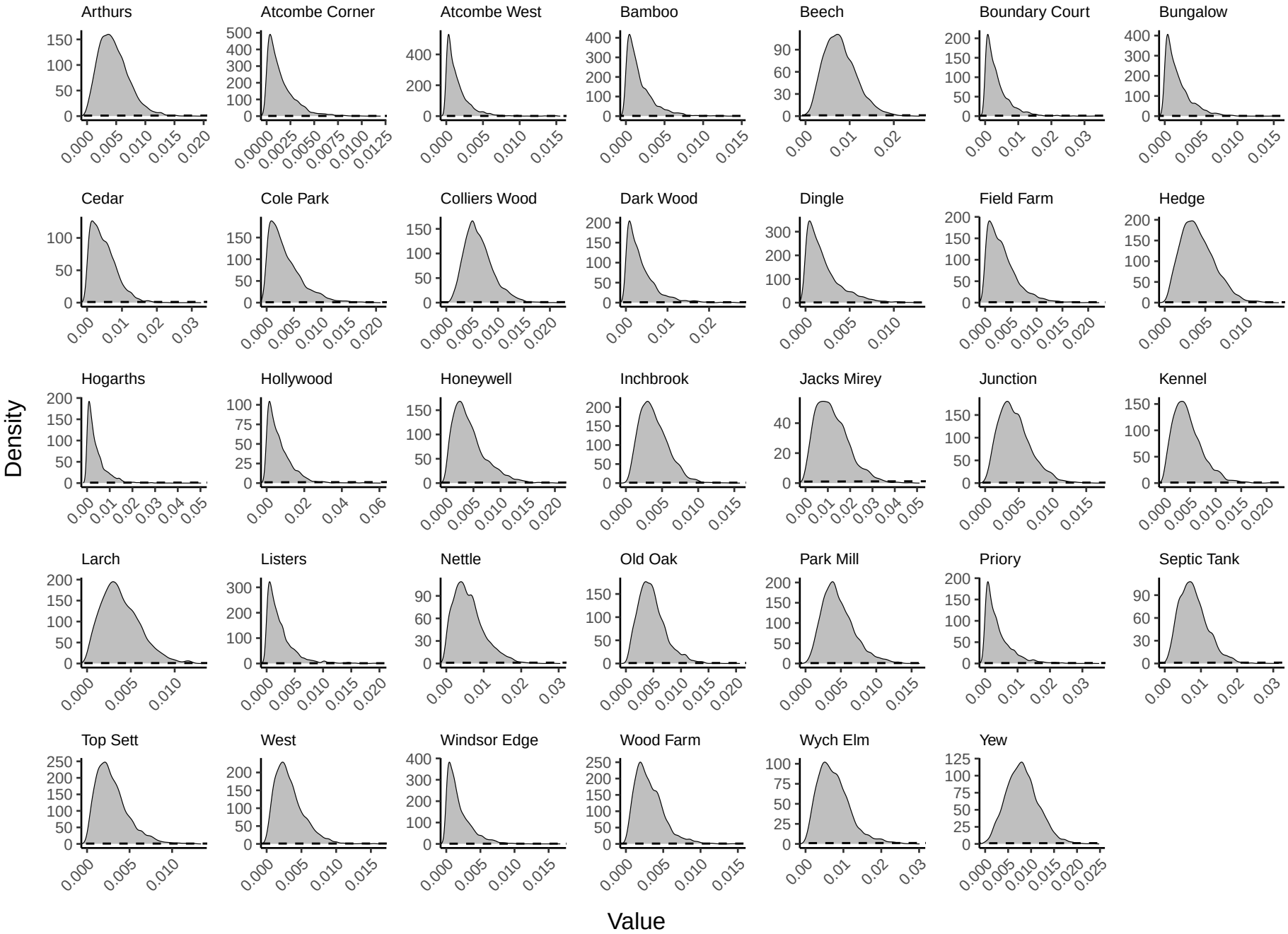

Number of individuals
400
300
200
100
0
Exposed
Infectious
Susceptible
Total
Proportion of positives
0.6
0.4
0.2
0.0
Brock
Culture
DPP
Gamma
StatPak
1980
1988
1996
2004
2012
2020
Year


η_1 (winter)
η_2 (spring)
η_3 (summer)
η_4 (autumn)
Density
Value
100
75
50
25
0
80
60
40
20
0.16
0.17
0.18
0.19
0.34
0.35
0.36
0.37
0.38
0.44
0.45
0.46
0.47
0.48
0.26
0.27
0.28
0.29
0.30


Survival function
1.00
0.75
0.50
0.25
0.00
0
5
10
15
Age (years)


Hazard function
0.2
0.1
0
5
10
15
Age (years)


β
q
τ (years)
Density
60
40
20
0
0.02
0.04
0.06
2.0
1.5
1.0
0.5
0.0
0.00
0.25
0.50
0.75
1.00
0.75
0.50
0.25
0.00
3
4
5
6
Value


Group−specific background rate of infection
Arthurs
Atcombe Corner
Atcombe West
Bamboo
Beech
Boundary Court
Bungalow
Cedar
Cole Park
Colliers Wood
Dark Wood
Dingle
Field Farm
Hedge
Hogarths
Hollywood
Honeywell
Inchbrook
Jacks Mirey
Junction
Kennel
Larch
Listers
Nettle
Old Oak
Park Mill
Priory
Septic Tank
Top Sett
West
Windsor Edge
Wood Farm
Wych Elm
Yew
Density
Value

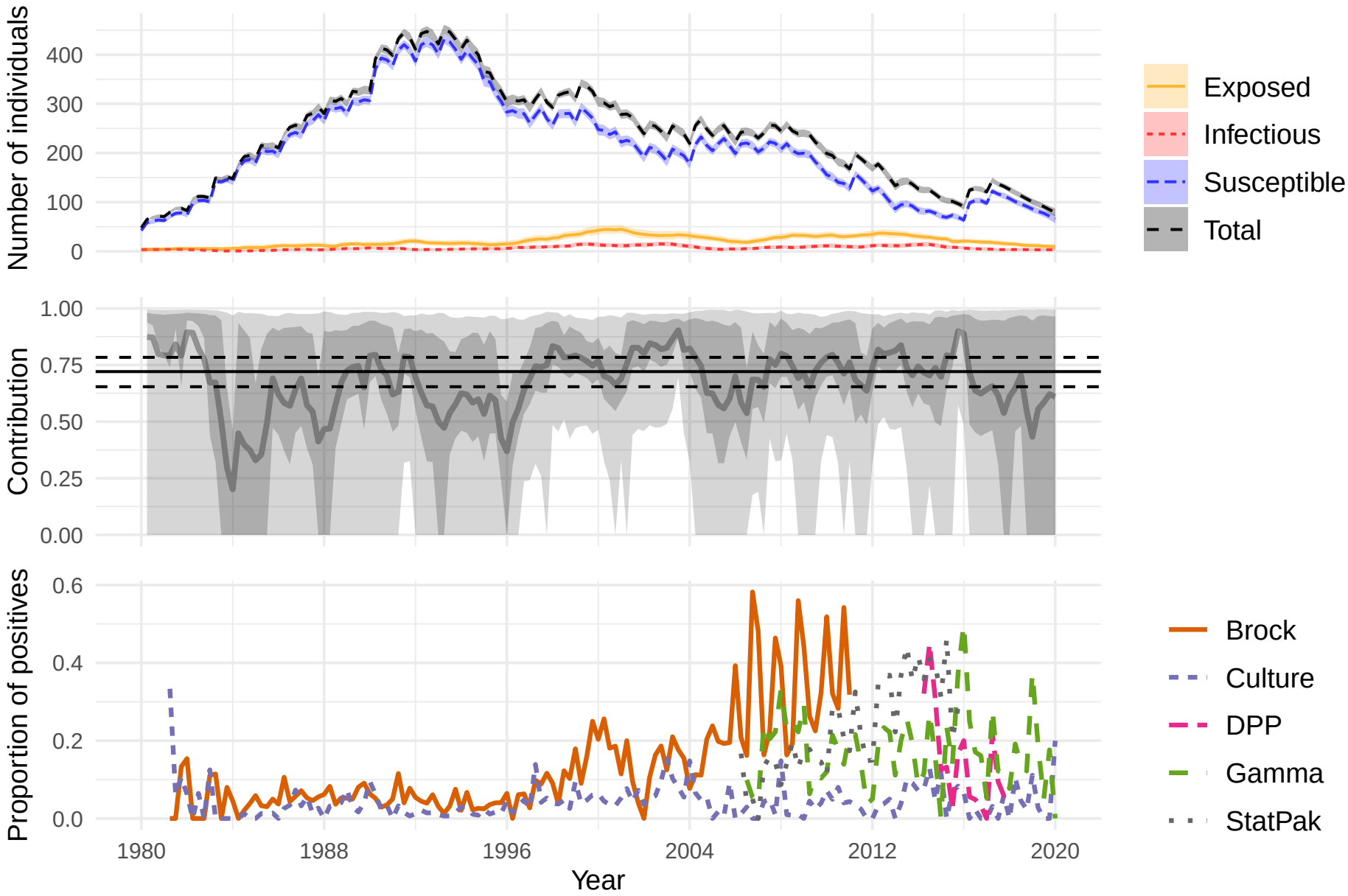
Type	Parameter	Mean	95% CI
Sensitivity (E)	Brock1	0.56	(0.49, 0.63)
	Brock2	0.73	(0.65, 0.81)
	Culture	0.072	(0.049, 0.098)
	DPP	0.41	(0.3, 0.53)
	Gamma	0.47	(0.41, 0.52)
	StatPak	0.82	(0.75, 0.87)
Sensitivity (I)	Brock1	0.91	(0.85, 0.96)
	Brock2	0.94	(0.84, 0.99)
	Culture	0.79	(0.72, 0.86)
	DPP	0.92	(0.76, 1)
	Gamma	0.85	(0.77, 0.93)
	StatPak	0.95	(0.89, 0.99)
Specificity	Brock1	0.98	(0.98, 0.99)
	Brock2	0.78	(0.76, 0.8)
	Culture	0.99	(0.99, 1)
	DPP	0.94	(0.92, 0.96)
	Gamma	0.92	(0.91, 0.93)
	StatPak	0.94	(0.93, 0.96)

Relative rates of infection

The relative contribution of direct badger-to-badger transmission on any infected badger k at its infection time t_k^E can be calculated as:

$$\frac{\beta K^q |\mathcal{I}_{k,t_k^E-1}|/|\mathcal{M}_{k,t_k^E-1}|^q}{\alpha_{g_k(t_k^E-1)} + \beta K^q |\mathcal{I}_{k,t_k^E-1}|/|\mathcal{M}_{k,t_k^E-1}|^q}.$$

These can then be aggregated at any spatio-temporal resolution.


Number of individuals
400
300
200
100
0
Exposed
Infectious
Susceptible
Total
Contribution
1.00
0.75
0.50
0.25
0.00
Proportion of positives
0.6
0.4
0.2
0.0
Brock
Culture
DPP
Gamma
StatPak
1980
1988
1996
2004
2012
2020
Year

Individual effective reproduction number (R_i)

→ Some individuals can have a disproportionately large contribution to the spread of infection.

We propose an estimator for R_i, from which a population-level R can be derived.

We define:

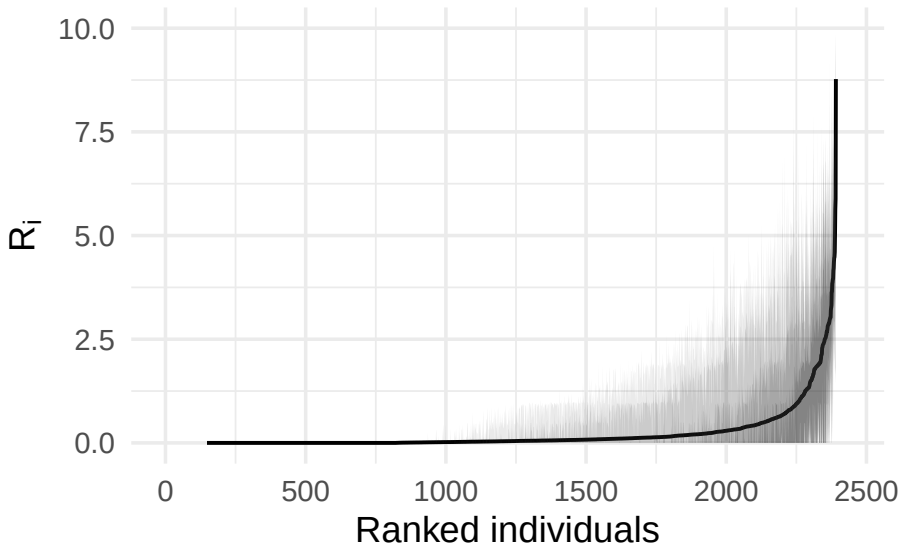
$$R_i = \sum_{k:t_k^E > t_i^E} p_{(k,i)},$$

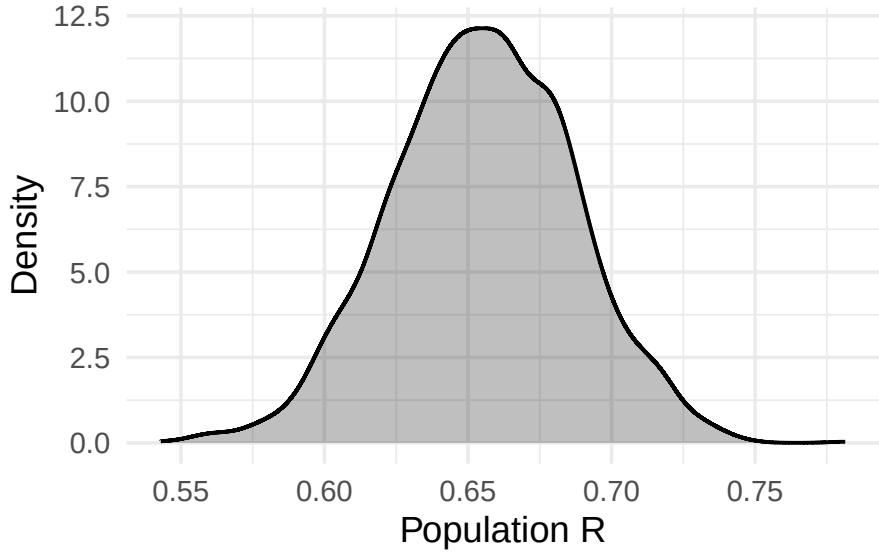
where $p_{(k,i)}$ is the relative likelihood that individual i infects individual k.[2]

Conditional on the predicted event times, $p_{(k,i)}$ is given by

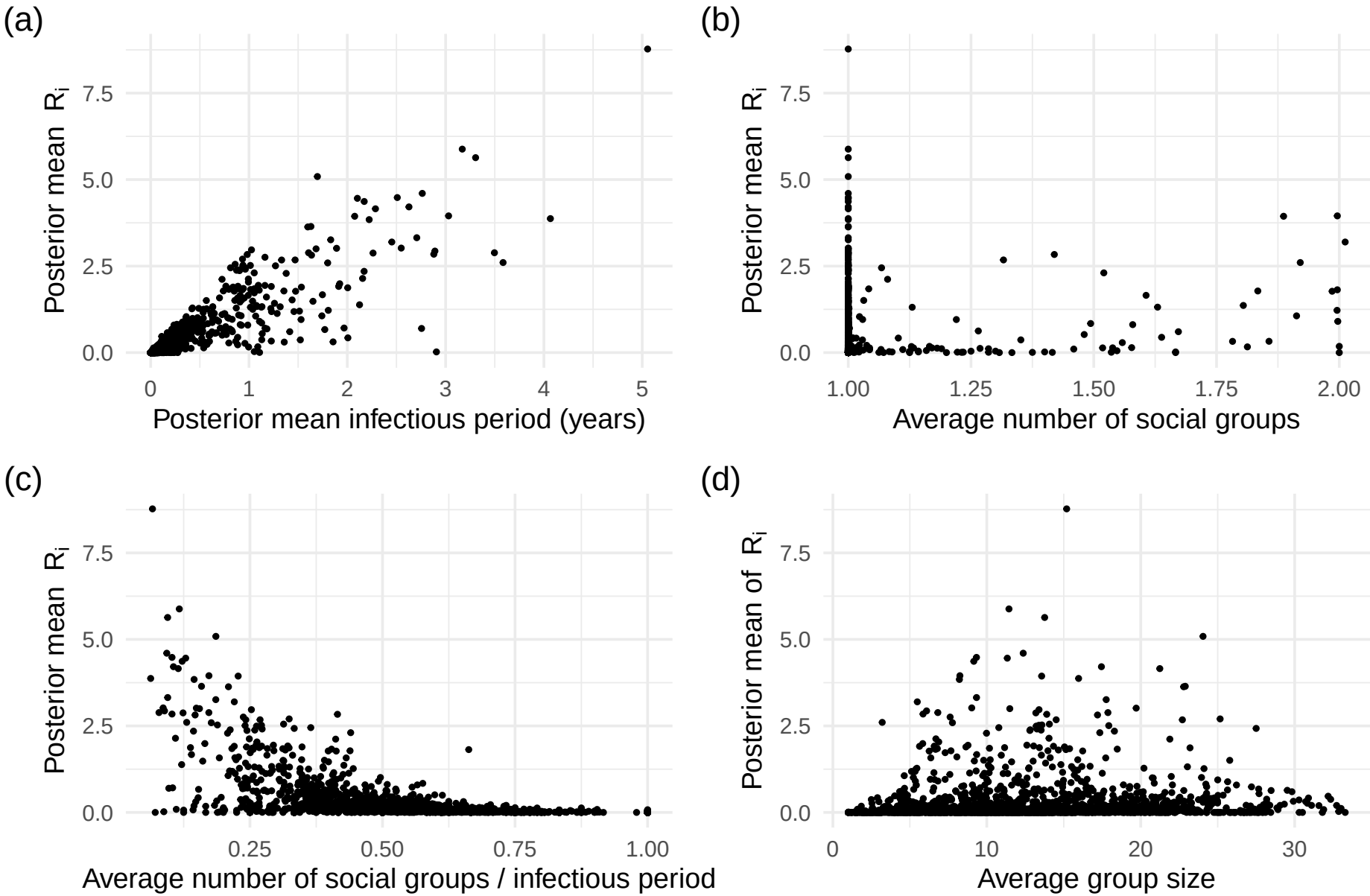
$$p_{(k,i)} = \begin{cases} \dfrac{\beta K^q / |\mathcal{M}_{k,t_k^E-1}|^q}{\alpha_{g_k}(t_k^E-1) + |\mathcal{I}_{k,t_k^E-1}|\beta K^q / |\mathcal{M}_{k,t_k^E-1}|^q} & \text{if } i \in \mathcal{I}_{k,t_k^E-1}, \\ 0 & \text{otherwise.} \end{cases}$$

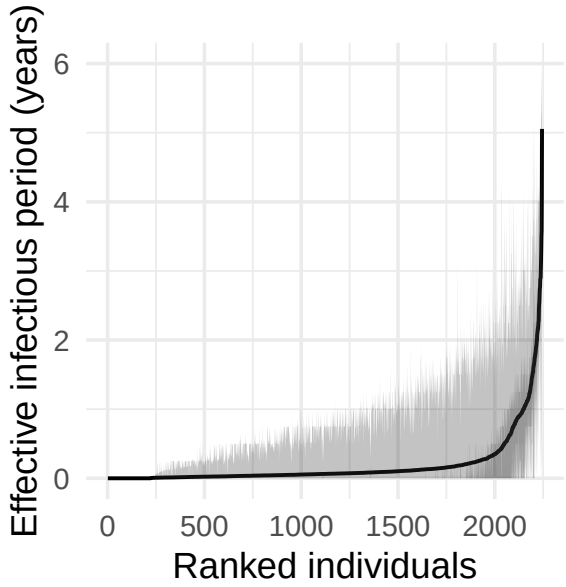
[2] Wallinga and Teunis (2004) use a generation-time interval distribution to estimate $p_{(k,i)}$.

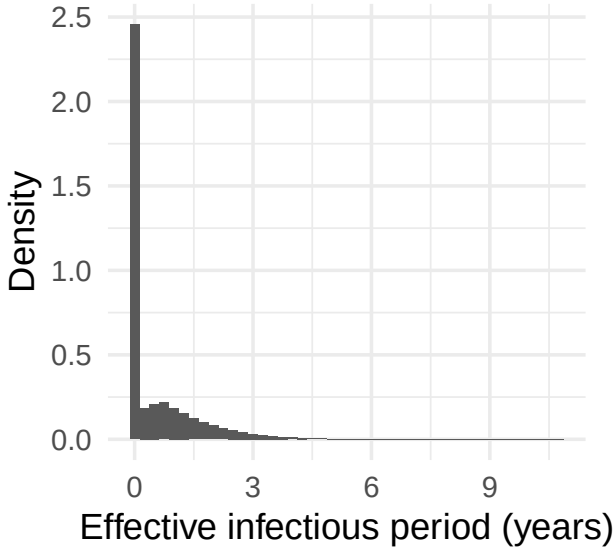


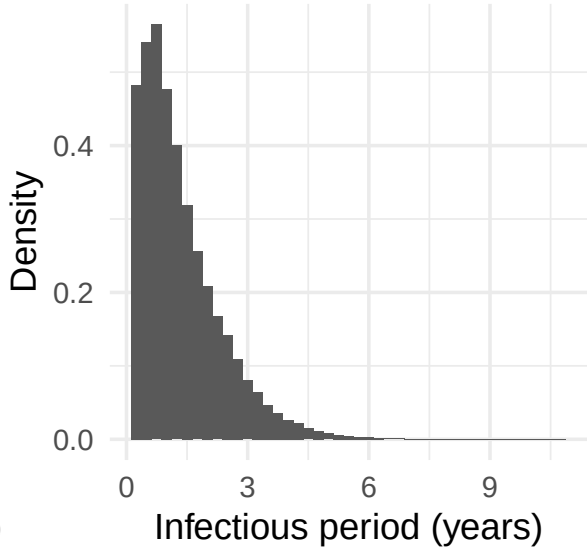


- This suggests the presence of **superspreader** badgers.
  - → This result emerges from the stochastic model **after fitting** it to the observed data.
- Population-level R is below 1 despite bTB **persistence**.
  - → Periodic **reintroductions** from external sources and/or persistent environmental contamination?


(a)
Posterior mean R_i
7.5
5.0
2.5
0.0
0
1
2
3
4
5
Posterior mean infectious period (years)
(b)
Posterior mean R_i
7.5
5.0
2.5
0.0
1.00
1.25
1.50
1.75
2.00
Average number of social groups
(c)
Posterior mean R_i
7.5
5.0
2.5
0.0
0.25
0.50
0.75
1.00
Average number of social groups / infectious period
(d)
Posterior mean of R_i
7.5
5.0
2.5
0.0
0
10
20
30
Average group size


Effective infectious period (years)
0
2
4
6
0
500
1000
1500
2000
Ranked individuals


Density
0.0
0.5
1.0
1.5
2.0
2.5
0
3
6
9
Effective infectious period (years)


Density
0.0
0.2
0.4
0
3
6
9
Infectious period (years)

Conclusions

The model:

- fits in a few hours, enabling us to conduct efficient Bayesian inference simultaneously on:
  - → all epidemiological parameters,
  - → CMR surveillance process,
  - → age-dependent mortality processes, and
  - → imperfect diagnostic testing;
- produces predictions for the hidden epidemiological states at individual-level over time;
- provides novel estimates of individual effective reproduction numbers.

Future work

The methodology could be:

- extended to allow for systematic model comparison and/or model averaging;
- applied to other disease systems.

Nevertheless, there are several areas to explore bTB in badgers further:

- more nuanced infectivity-profile dependent sensitivity and specificity processes;
- individual-level covariate dependencies;
- infection-related mortality;
- non-exponentially distributed latent period;
- sex-/age-specific differences in susceptibility and transmission;
- characteristics of superspreader badgers;
- ...

References

Brooks-Pollock, E., Roberts, G. O., and Keeling, M. J. (2014). A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature*, 511(7508):228–231.

Buzdugan, S. N., Vergne, T., Grosbois, V., Delahay, R. J., and Drewe, J. A. (2017). Inference of the infection status of individuals using longitudinal testing data from cryptic populations: Towards a probabilistic approach to diagnosis. *Scientific Reports*, 7(1):1–11.

Delahay, R. J., Walker, N., Smith, G., Wilkinson, D., Clifton-Hadley, R., Cheeseman, C., Tomlinson, A., and Chambers, M. (2013). Long-term temporal trends and estimated transmission rates for *Mycobacterium bovis* infection in an undisturbed high-density badger (*Meles meles*) population. *Epidemiology & Infection*, 141(7):1445–1456.

Touloupou, P., Finkenstädt, B., and Spencer, S. E. F. (2020). Scalable Bayesian inference for coupled hidden Markov and semi-Markov models. *Journal of Computational and Graphical Statistics*, 29(2):238–249.

Wallinga, J. and Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 160(6):509–516.

Paper is now on bioRxiv!

https://doi.org/10.1101/2024.01.26.576600

Any questions?