
Breeding against Infectious Disease

Estimating Gene Effects on Susceptibility and Infectivity

EAAP 2016 Belfast
Floor Biemans
Mart de Jong
Piter Bijma
floor.biemans@wur.nl



Floor Biemans



Breeding against Infectious Disease

- Overall objective: Reduce prevalence
 - Prevalence = fraction of the population infected

■ Current approach: $y = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$

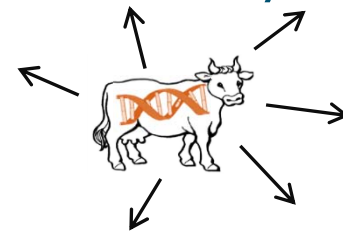


Connect disease status
to **own** breeding value

Captures genetic variation
in host susceptibility only

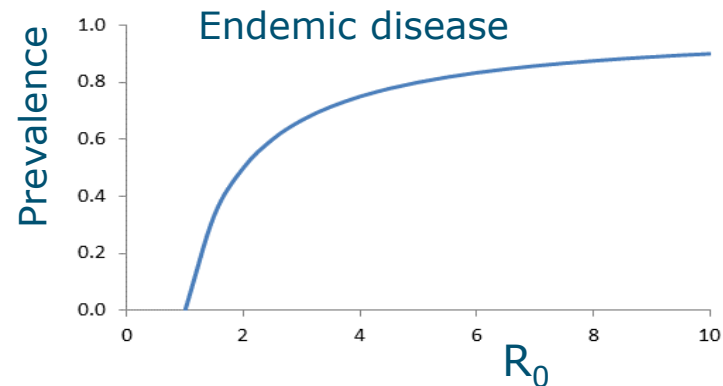
- We miss part of the genetic variation: Host Infectivity

Infectivity = propensity to infect others



Traits affecting prevalence

- R_0 determines prevalence



- R_0 : Basic reproduction ratio
 - $R_0 =$ "Number of new cases due to a case"
- Anche *et al.* 2012: $R_0 = c \times \overline{\text{susceptibility}} \times \overline{\text{infectivity}}$

Reduce prevalence \rightarrow we should also consider infectivity

Objective

- Develop methods to estimate single gene-effects on susceptibility **and infectivity**

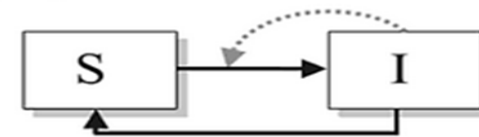
- **Challenges:**

- Infectivity:
- 0/1 trait (disease status)
 - Generalized linear models
- Time dynamics
 - Time-series data



Epidemiological & Genetic Model

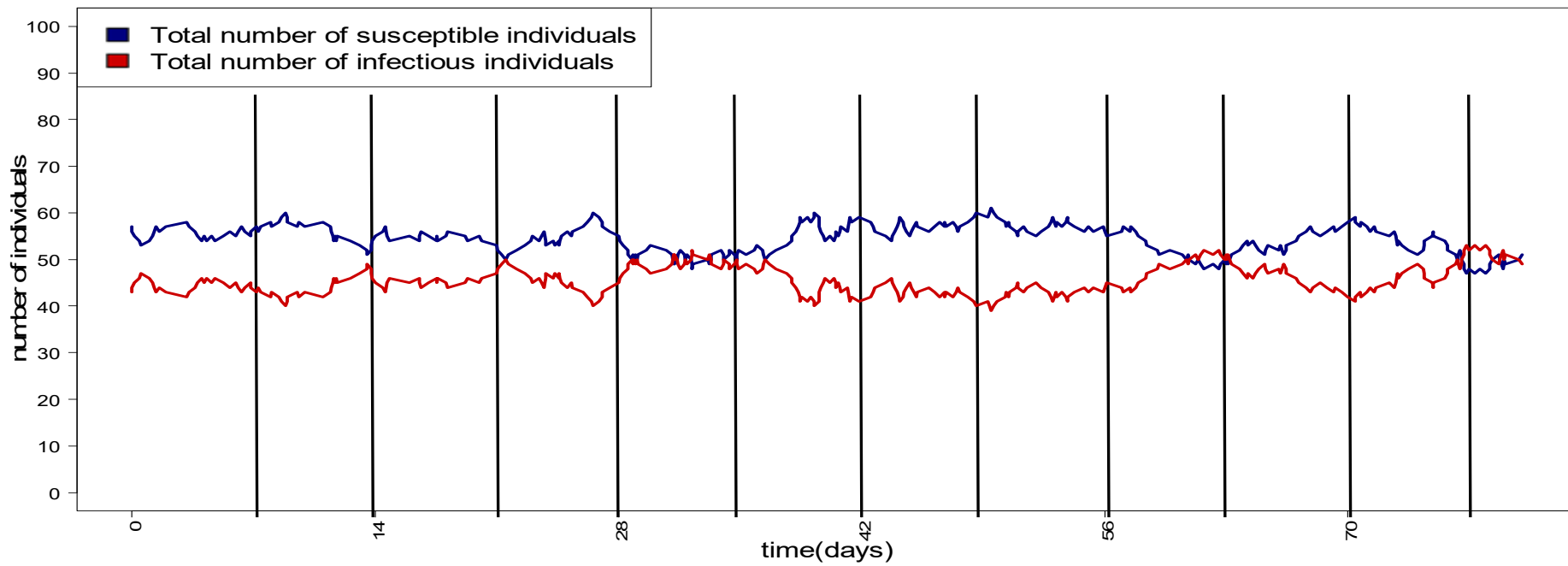
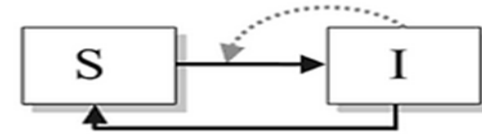
- Endemic disease
- SIS-model
 - Susceptible – Infected - Susceptible
- Genetic model
 - 2 loci, each with 2 alleles
 - Susceptibility locus; alleles **g** and **G**
 - Three genotypes: **gg**, **gG**, **GG**
 - Infectivity locus; alleles **f** and **F**
 - Three genotypes: **ff**, **fF**, **FF**



Simulated data

- Simulated data
 - 10 herds of ~ 100 individuals
 - Within-herd endemics
 - 11 observation moments per herds
 - Time series data on disease status (0/1 = S/I)

Simulated time-series data



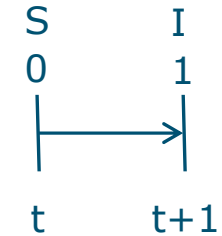
Individual disease status (0/1) recorded at each time point

Data

- Disease status of each individual at each time point (0/1 = S/I)
 - Which susceptibles have become infected
 - Which infecteds (may) have done it
- Genotyped individuals
- Length of the time interval

Methods

Model the probability of transmission in a time interval: $P(S \rightarrow I)$



- Binary data from a Poisson process
 - GLM with complementary log-log link function

- Transmission probability depends on:
 - Overall average transmission rate (c)
 - Susceptibility genotype of the (focal) individual
 - Number of infectious herd-mates at time t
 - Infectivity genotype of those herd mates

Results: Generalized Linear Model

$$\text{cloglog} \left(\varepsilon \left(\frac{c}{s} \right) \right) = c_0 + c_1 * \text{IndexG} + c_2 * \text{fractionF} + \log \left(\left(\frac{I_{tot}}{N} \right) * \Delta t \right)$$

GLM with complementary log-log link-function

Results: Generalized Linear Model

$$\text{cloglog} \left(\varepsilon \left(\frac{c}{s} \right) \right) = c_0 + c_1 * \text{IndexG} + c_2 * \text{fractionF} + \log \left(\left(\frac{I_{tot}}{N} \right) * \Delta t \right)$$

Expected number of **cases** for each susceptibility genotype, in the interval $t \rightarrow t+1$

Results: Generalized Linear Model

$$cloglog\left(\varepsilon\left(\frac{c}{s}\right)\right) = c_0 + c_1 * IndexG + c_2 * fractionF + \log\left(\left(\frac{I_{tot}}{N}\right) * \Delta t\right)$$

Susceptibility genotype of susceptibles at time t

= Allele count (0, 1, or 2)

Results: Generalized Linear Model

$$\text{cloglog} \left(\varepsilon \left(\frac{c}{s} \right) \right) = c_0 + c_1 * \text{IndexG} + c_2 * \text{fractionF} + \log \left(\left(\frac{I_{tot}}{N} \right) * \Delta t \right)$$

Infectivity genotype of infectious herd mates at time t

= Average allele count of those herd mates

Results: Generalized Linear Model

$$\text{cloglog} \left(\varepsilon \left(\frac{c}{s} \right) \right) = c_0 + c_1 * \text{IndexG} + c_2 * \text{fractionF} + \log \left(\left(\frac{I_{tot}}{N} \right) * \Delta t \right)$$

Offset

- Fraction of herd mates infected at time t
- Length of the **time** interval

Results: Generalized Linear Model

$$\text{cloglog} \left(\varepsilon \left(\frac{c}{s} \right) \right) = c_0 + c_1 * \text{IndexG} + c_2 * \text{fractionF} + \log \left(\left(\frac{I_{tot}}{N} \right) * \Delta t \right)$$

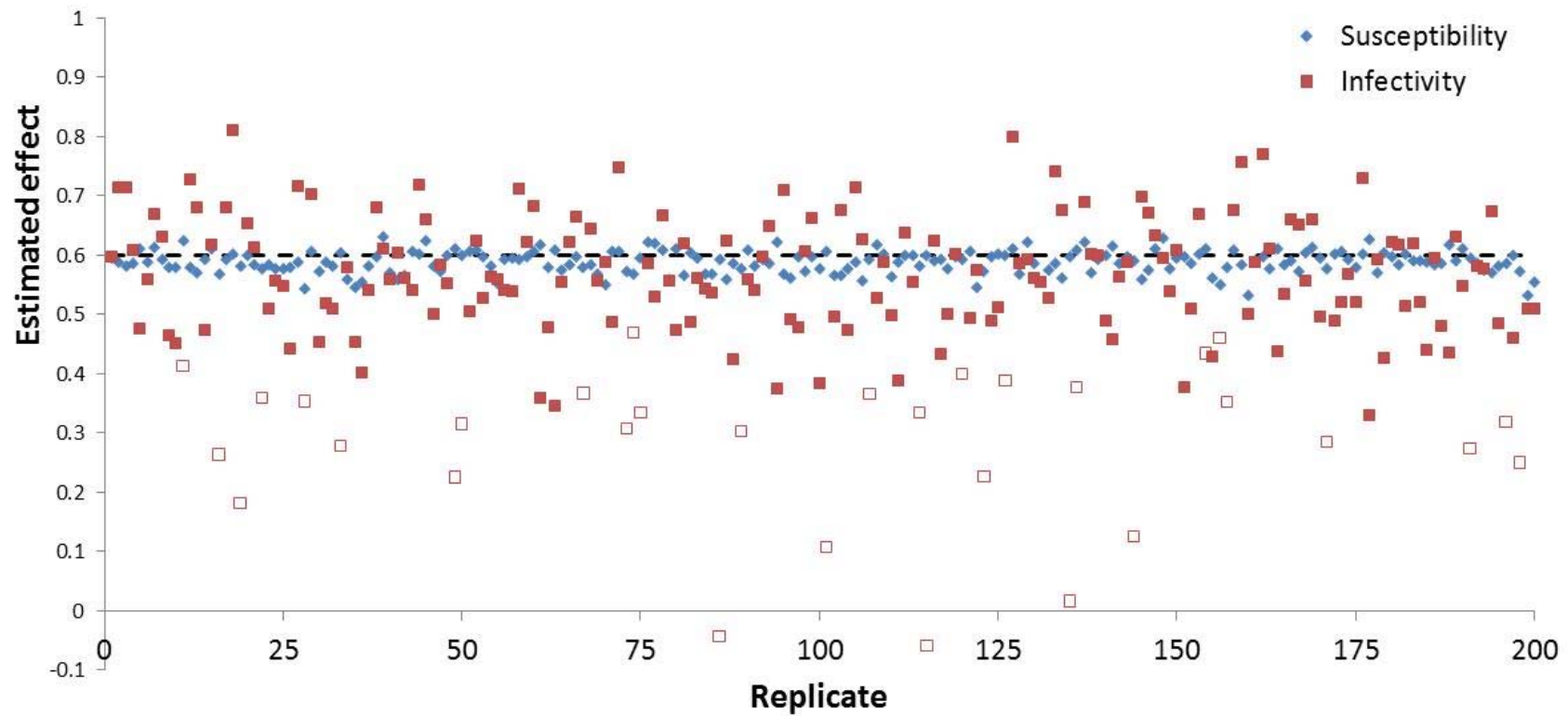
Solutions: \hat{c}_0 , \hat{c}_1 , \hat{c}_2

Estimates of interest:

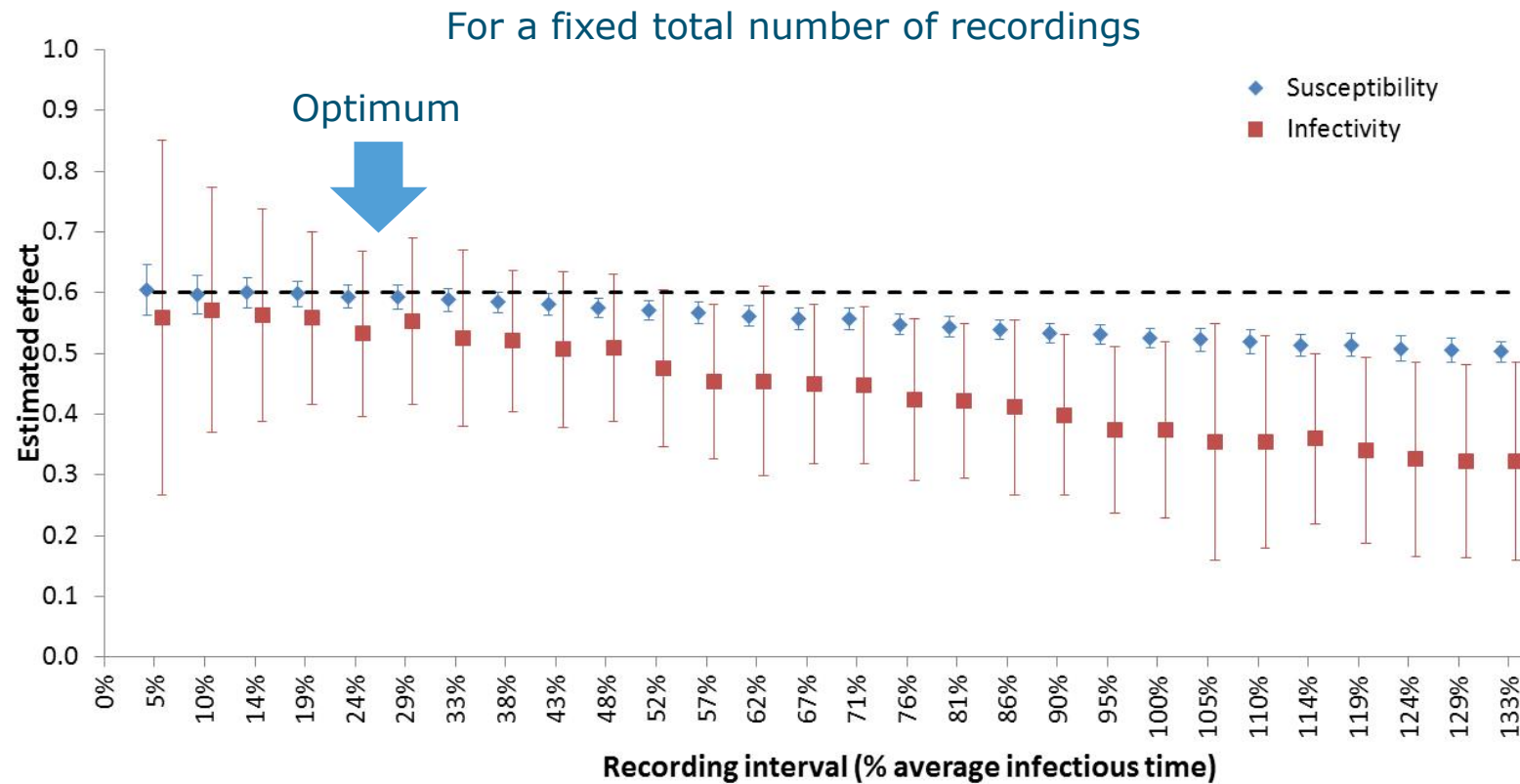
Susceptibility effect G-allele = $e^{\hat{c}_1}$

Infectivity effect F-allele = $e^{\hat{c}_2}$

Results: Estimates (at optimum recording interval)



Results: effect recording interval



Conclusions

Thanks for
your attention!

- Generalized LM for GWAS
 - Susceptibility estimates are unbiased
 - Infectivity estimates tend to be biased downwards (conservative)
 - Optimum recording interval $\sim 1/3$ of infectious period
- Ongoing
 - Application to digital dermatitis in dairy cattle (Mortellaro's disease)
- Extensions
 - Mixed models and Genomic Prediction
- Alternatives: Bayesian models (Anacleto *et al.* 2015)



Floor Biemaas