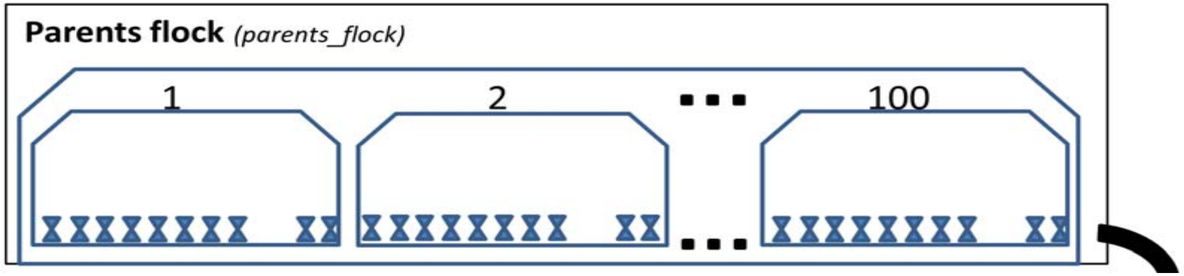
GERMAN FEDERAL INSTITUTE
FOR RISK ASSESSMENT

Possibilities and Limits of Logistic Regression in a Study of the transmission dynamics of ESBL/AmpC producing E. coli between broiler flocks

Guido Correia Carreira, Carolina Plaza Rodriguez,
Annemarie Käsbohrer

The big picture: Transmission model of ESBL/AmpC *E.coli* in the broiler production chain



Our model world:

100 flocks, each with 100 animals

Production chain with 5 stages

The big picture - Transmission model of ESBL/AmpC *E.coli* cont.

parents_flock

Prevalence in Parent FLOCKS

Prevalence in ANIMALS (of a parent flock)

Modelling impact of horizontal transmission on flock prevalence

We saw that:

Prev_0 ***(initial flock prevalence)***

Horizontal transmission

Prev_1 ***(flock prevalence after horizontal transmission)***

„Under the hood“:

Horizontal transmission:

$$Prev\_1 = Prev\_0 + \qquad \cdot (100 - Prev\_0)$$

The logistic regression model used

Logistic regression provides probability for binomial outcome. Here: probability that flock got colonised by ESBL/AmpC *E.coli* in given stage of the production chain.

$$\text{Probability that flock got colonised} = \frac{1}{1 + e^{-\left(hs \cdot \boxed{\frac{Prev_0}{nf}} + hp \cdot \boxed{col\_hist} + ho\right)}}$$

Parameterisation of the model

Question: How to chose the values for hs, hp and ho ?

Answer: One fits logistic regression model to data

Challenge: We have no appropriate data (on subsequent production cycles)

Approach: Computer experiment with theoretical scenarios which means

- making our own theoretical data,
- fitting regression model to theoretical data (using R package brglm),
- get an idea of possible values ranges

Let`s look at some scenarios

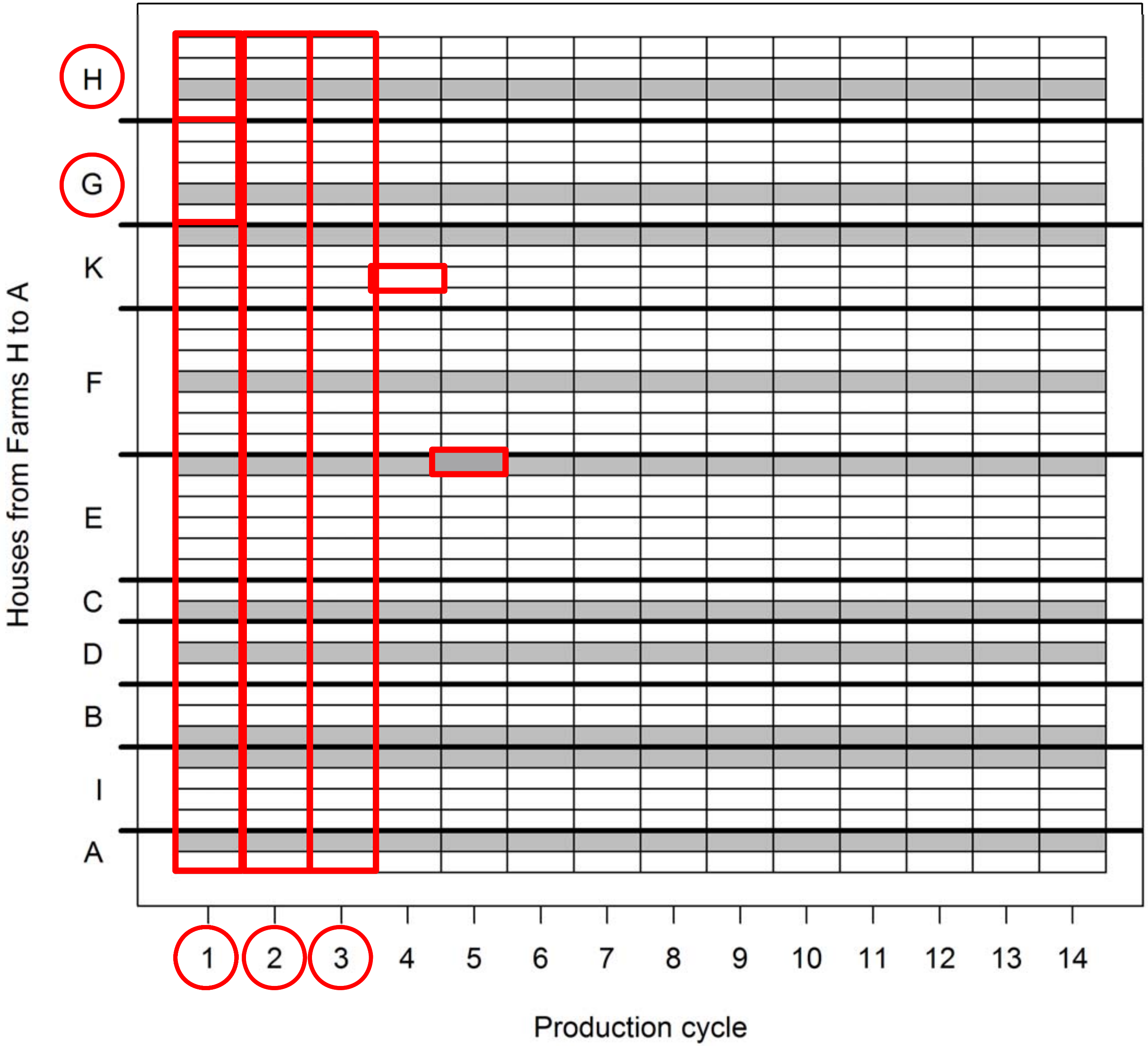
Scenario 1: Intercycle transmission dominates

Intercycle transmission means:

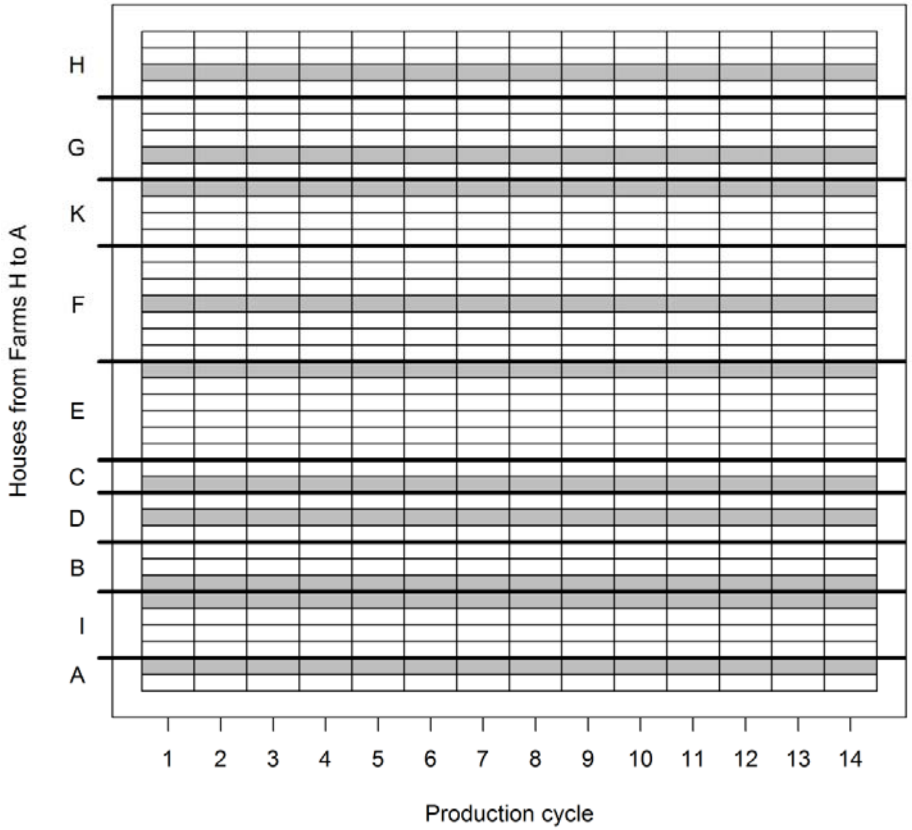
colonisation status of a house is determined by its colonisation status in previous production cycle.

and

No intrafarm transmission



Scenario 1: Intercycle transmission dominates

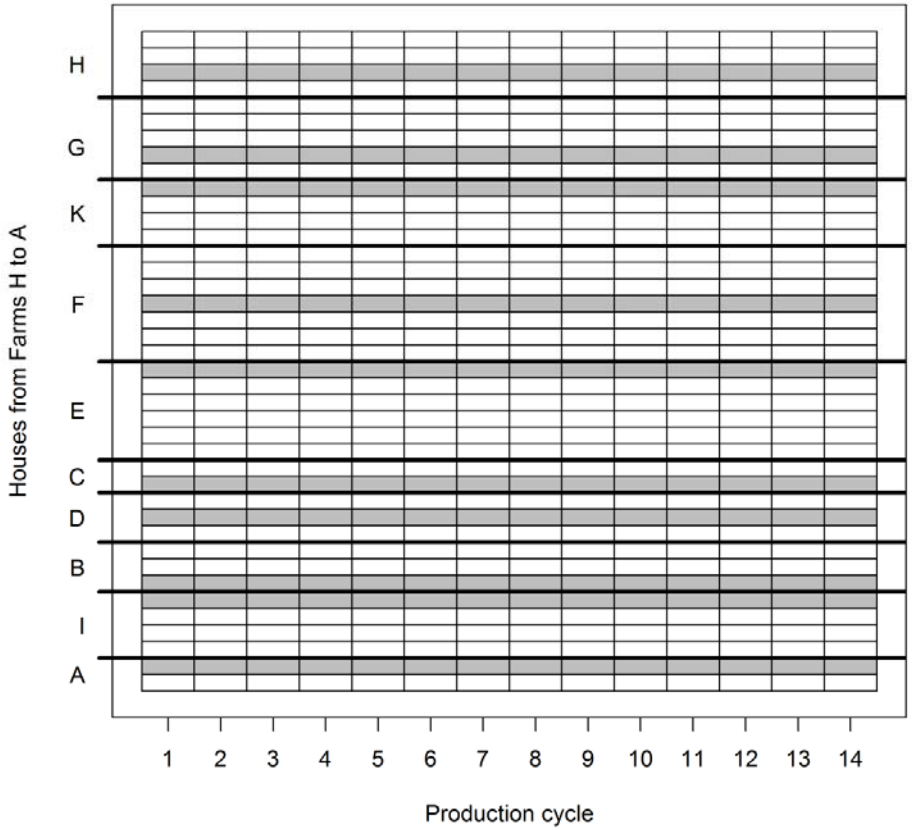


Fitting the logistic regression model:

	Value	P-Value (Wald test)
hs	8.143	0.29047
hp	13.860	< 0.001 ***
ho	-8.295	0.00357 **

$$\text{Model of scenario 1}: \frac{1}{1+e^{-\left(8.143 \cdot \frac{Prev_0}{nf} + 13.86 \cdot col\_hist - 8.295\right)}}$$

Scenario 1: Intercycle transmission dominates



Fitting the logistic regression model:

	Value	P-Value (Wald test)
hs	**8.143**	**0.29047**
hp	**13.860**	**< 0.001 *****
ho	**-8.295**	**0.00357 ****

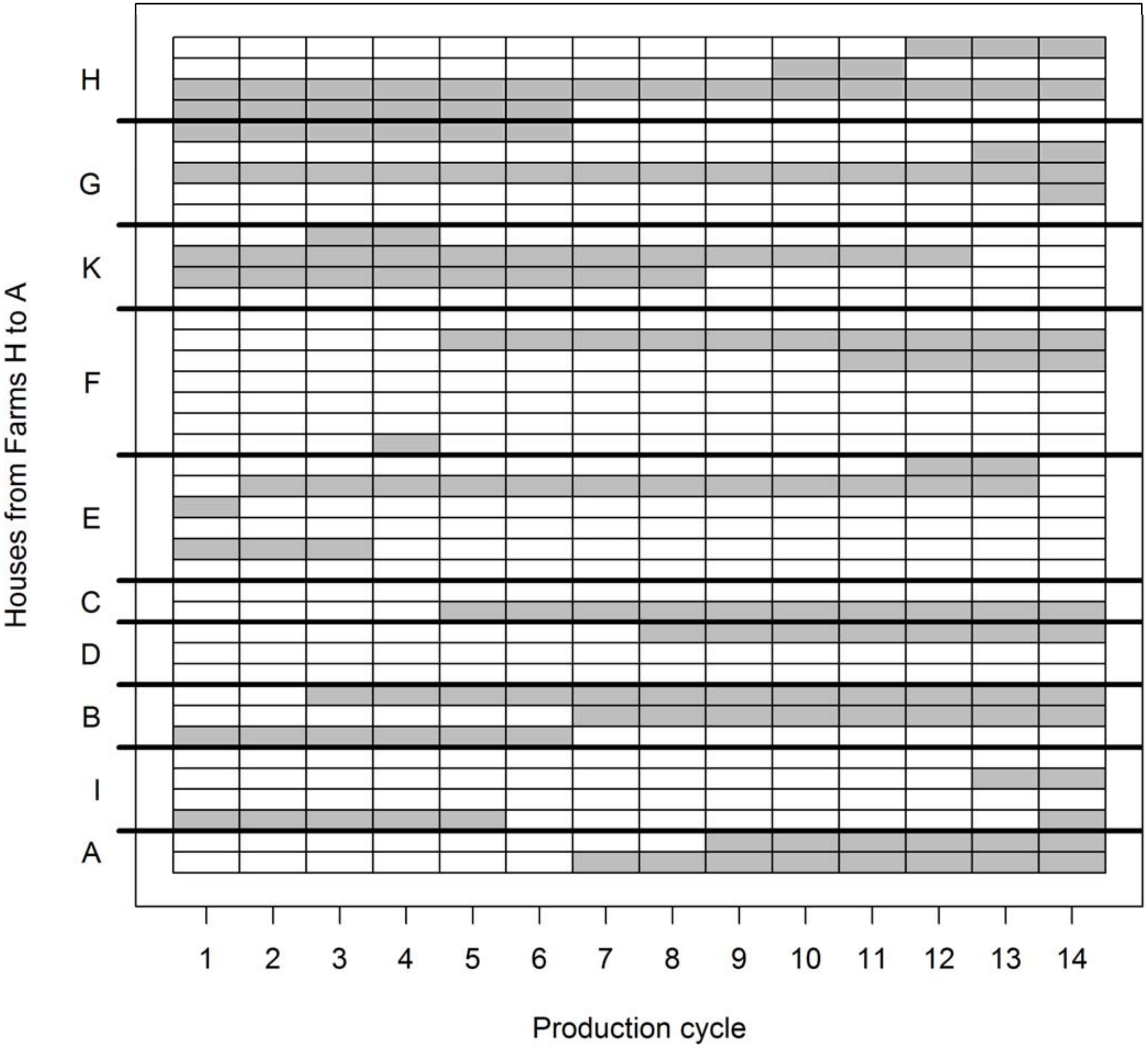
520 predictions, using a **threshold of 0.5** the model identifies 390 true negatives and 130 true positives

$$\Rightarrow \text{Accuracy} = \frac{520}{520} = 100\%$$

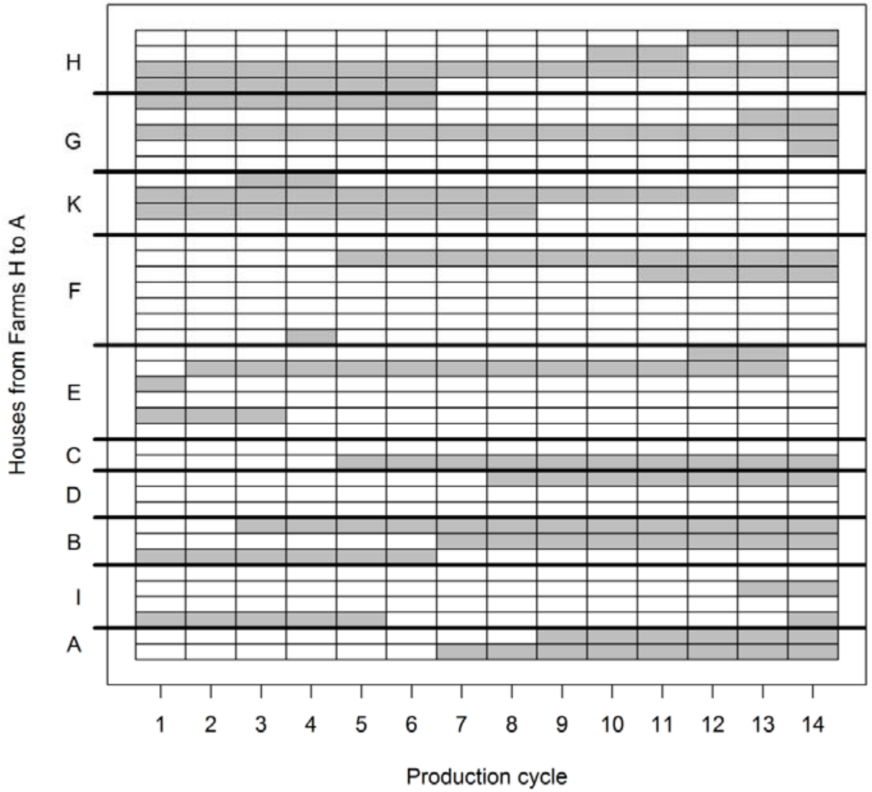
Scenario 2: Intercycle transmission dominates - probabilistic

Intercycle transmission means: colonisation status of a house correlates highly correlated with its colonisation status in the previous production cycle.

When a house was colonised in previous cycle it has 90% probability to get colonised in the current cycle



Scenario 2: Intercycle transmission dominates - probabilistic



Fitting the logistic regression model:

	Value	P-Value (Wald test)
hs	-1.0632	0.309
hp	5.2550	< 0.001 ***
ho	-2.6993	< 0.001 ***

520 predictions, using a **threshold of 0.5** the model identifies 349 true negatives and 140 true positives

⇨ Accuracy $= \frac{489}{520} \approx 94\%$

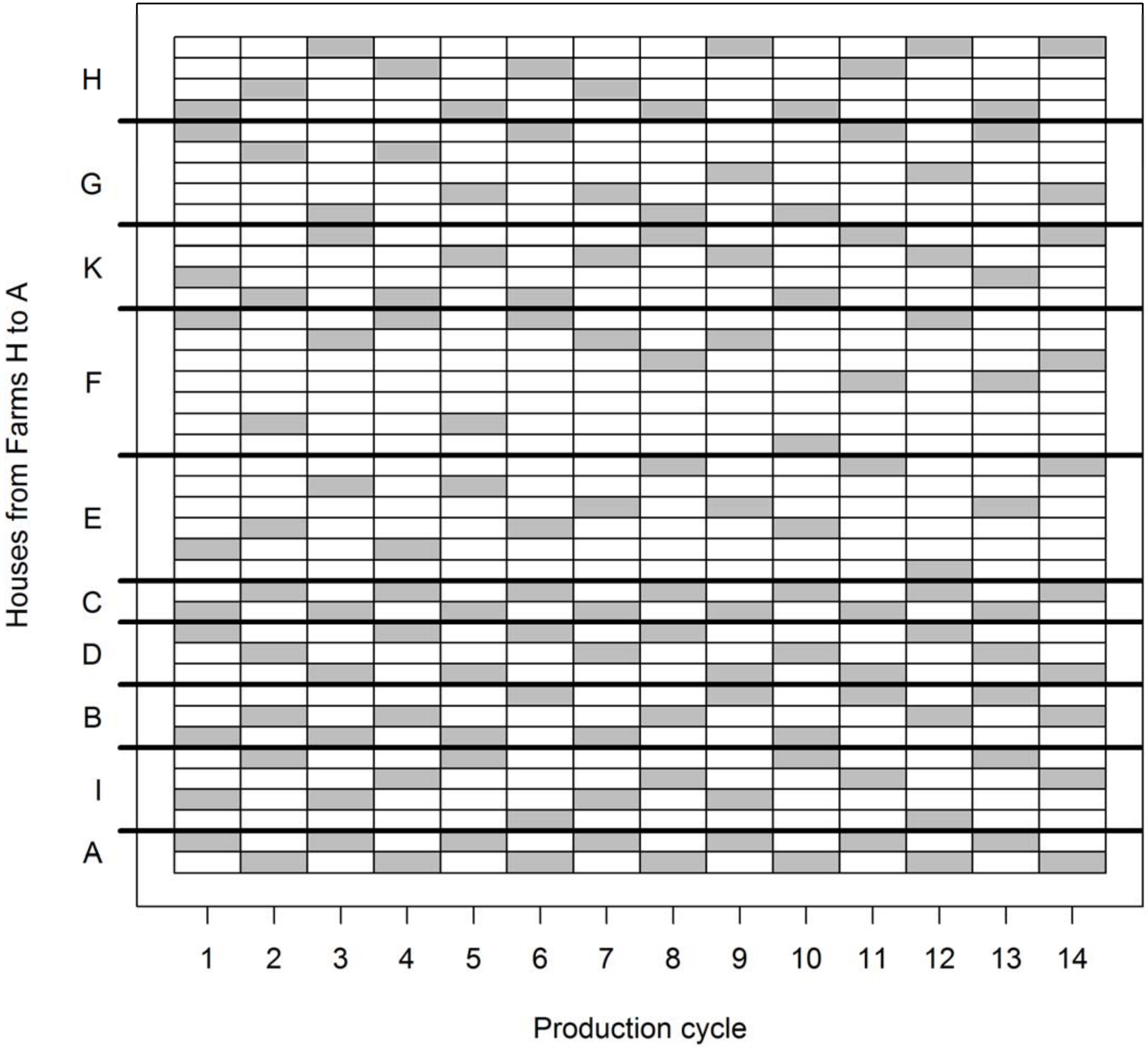
BfR

Scenario 3: No intrafarm and No intercycle transmission

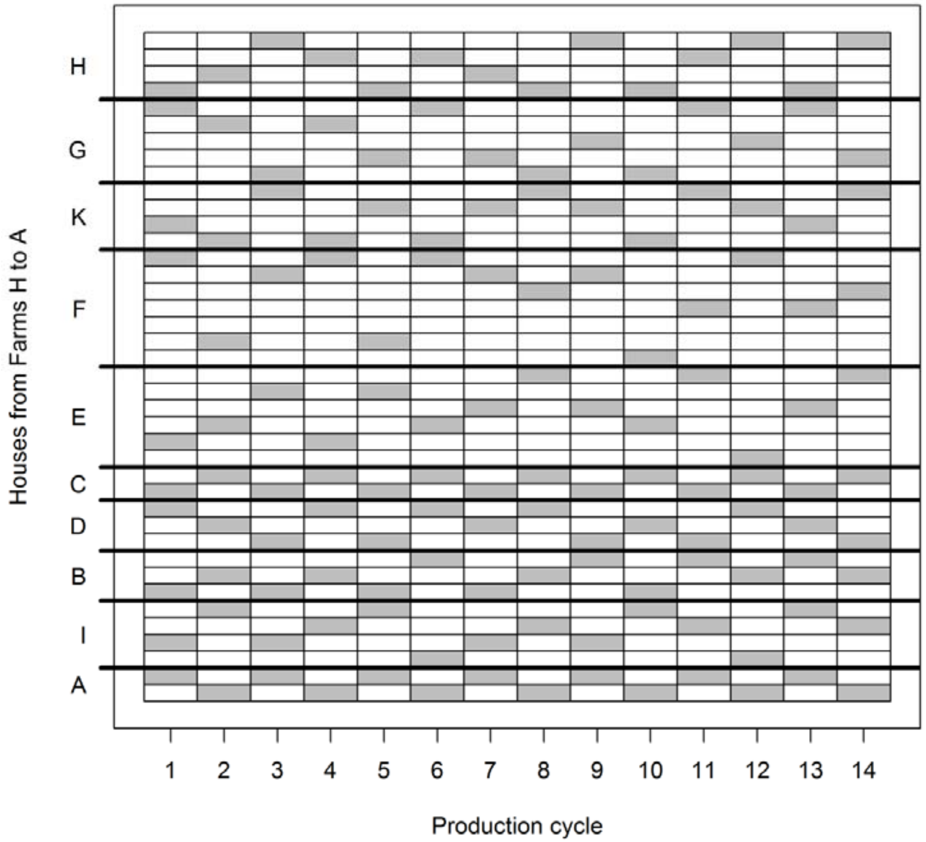
Each Farm has exactly one colonised house

And

House was never colonised in previous production cycle



Scenario 3: No intrafarm and No intercycle transmission



Fitting the logistic regression model:

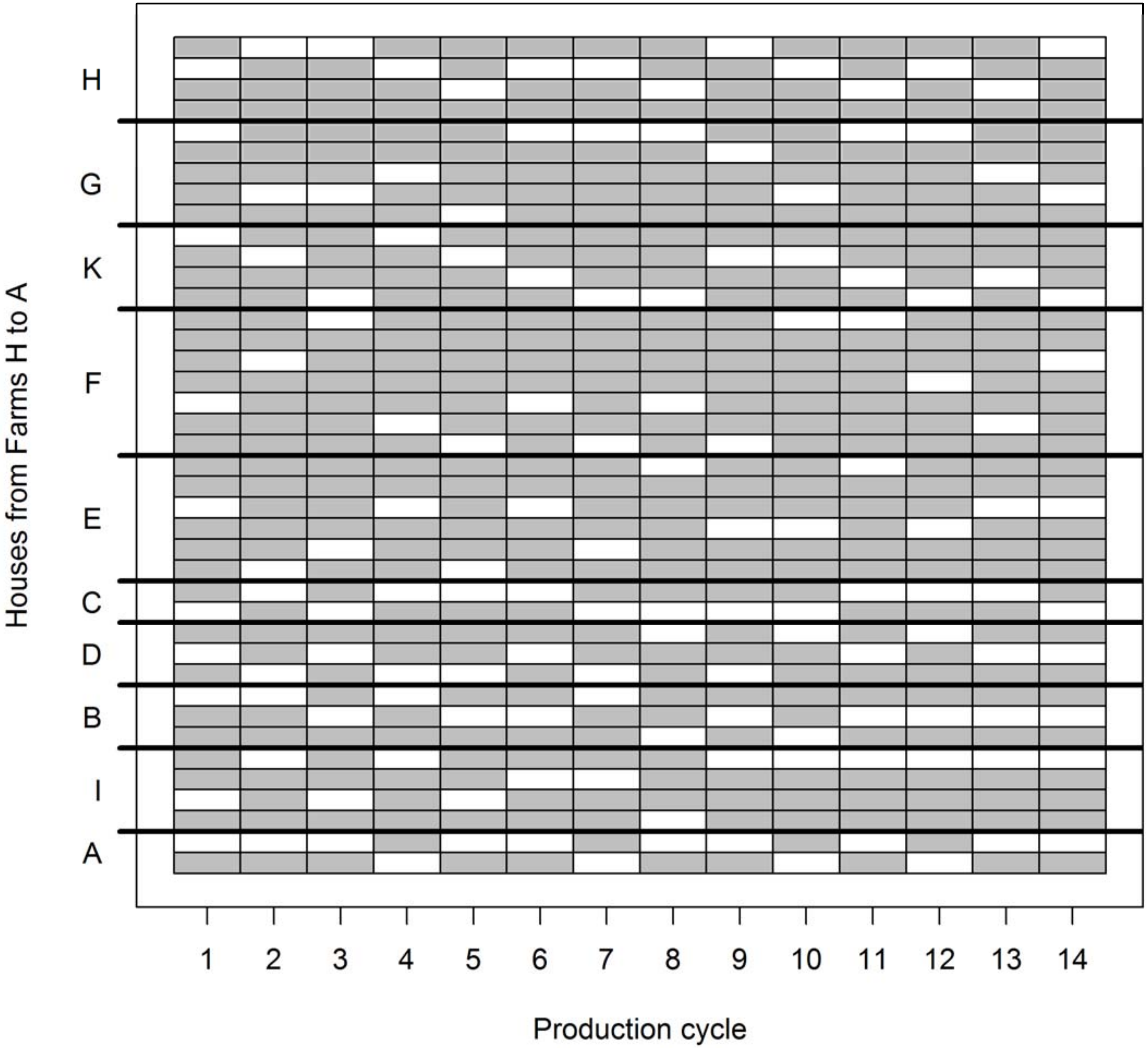
	Value	P-Value (Wald test)
hs	**-73.893**	**< 0.001 *****
hp	**1.550**	**0.448**
ho	**5.559**	**< 0.001 *****

520 predictions, using a **threshold of 0.5** the model identifies 390 true negatives and 130 true positives

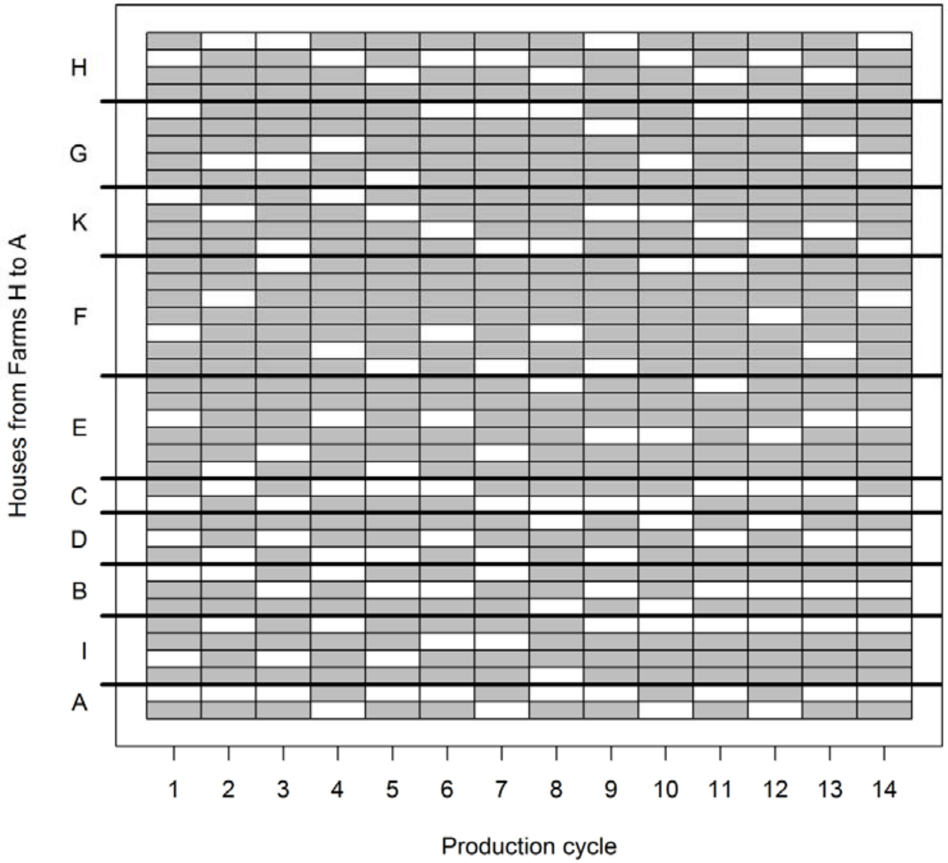
⇨ Accuracy = $\frac{520}{520} = 100\%$

Scenario 4: Intrafarm but no explicit intercylce transmission

On each farm all houses except one is colonised



Scenario 4: Intrafarm but no explicit intercylce transmission



Fitting the logistic regression model:

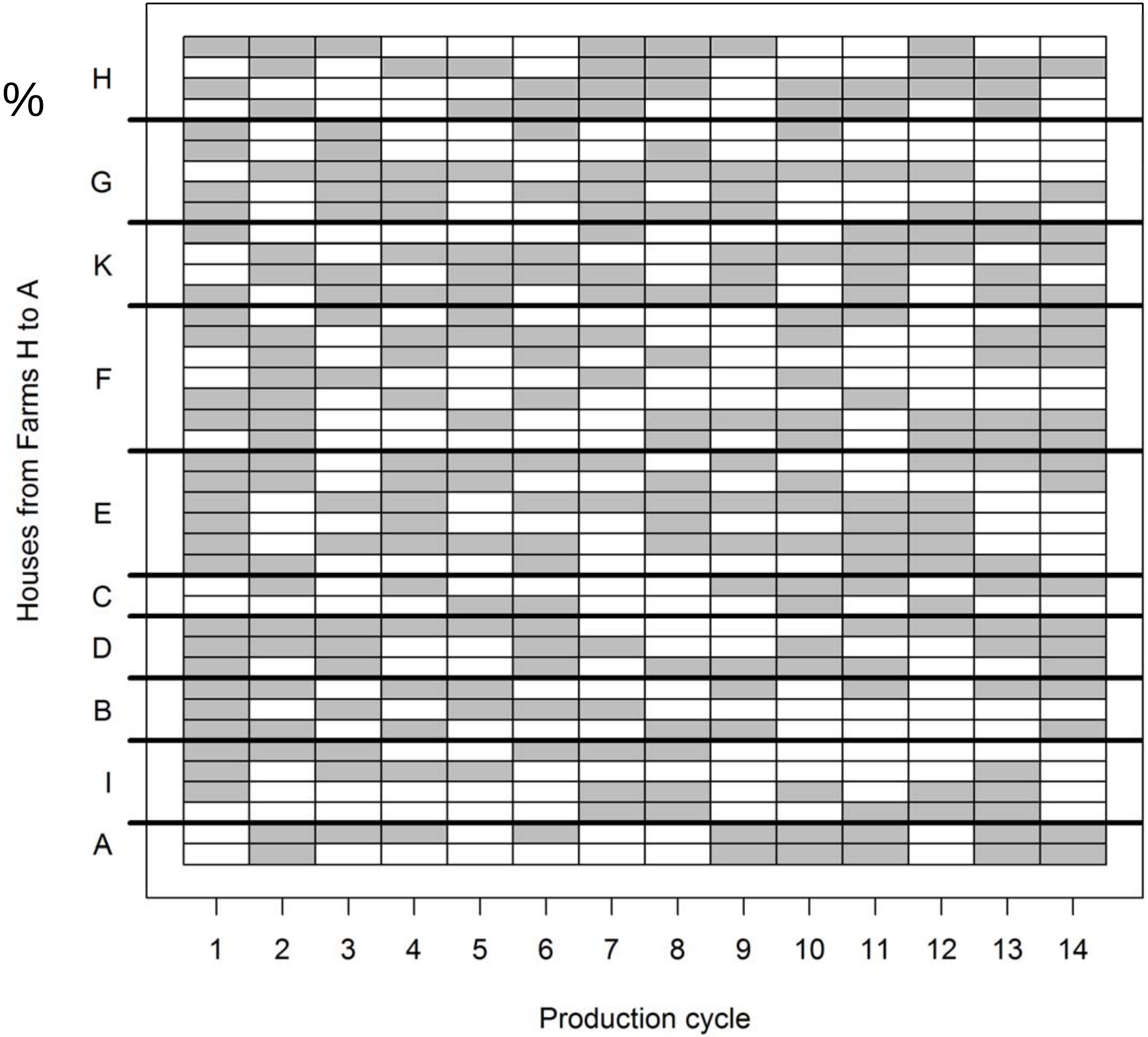
	Value	P-Value (Wald test)
hs	**-11.0713**	**< 0.001 *****
hp	**0.9272**	**0.00102 ****
ho	**7.4618**	**< 0.001 *****

520 predictions, using a **threshold of 0.6** the model identifies 86 true negatives and 367 true positives

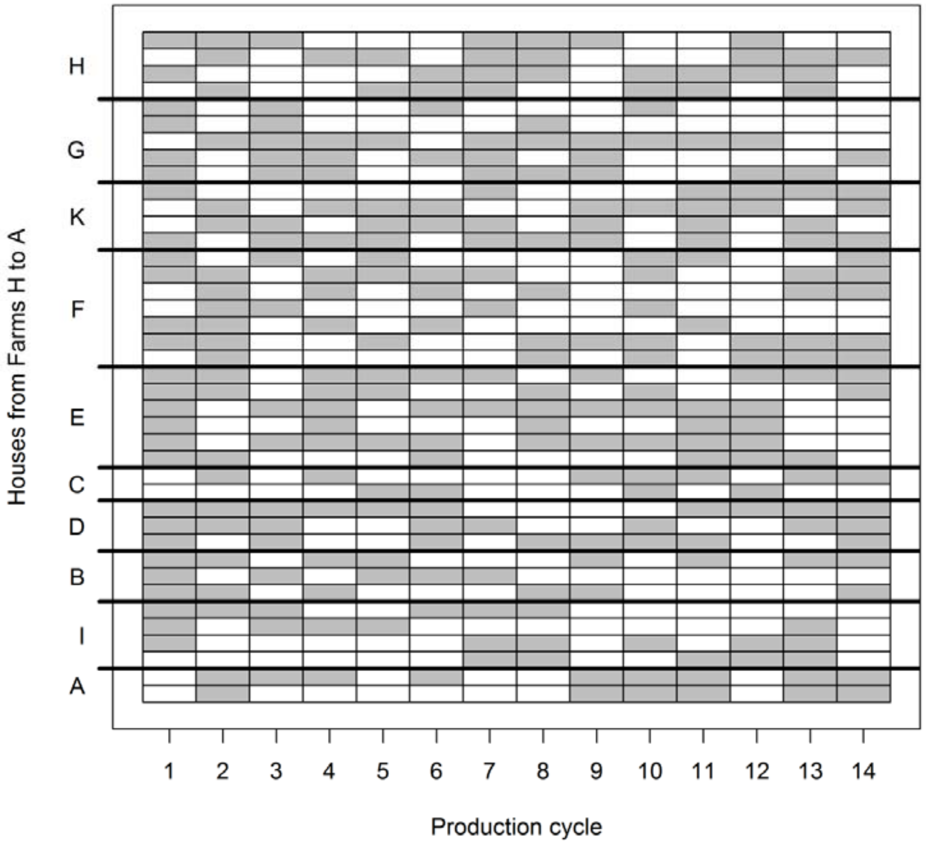
⇨ Accuracy = $\frac{453}{520} \approx 87\%$

Scenario 5: Random colonisation

Each house has a 50% chance of becoming colonised



Scenario 5: Random colonisation



Fitting the logistic regression model:

	Value	P-Value (Wald test)
hs	**0.005857**	**0.989**
hp	**0.16887**	**0.336**
ho	**-0.132886**	**0.499**

520 predictions, using a **threshold of 0.5** the model identifies 138 true negatives and 133 true positives

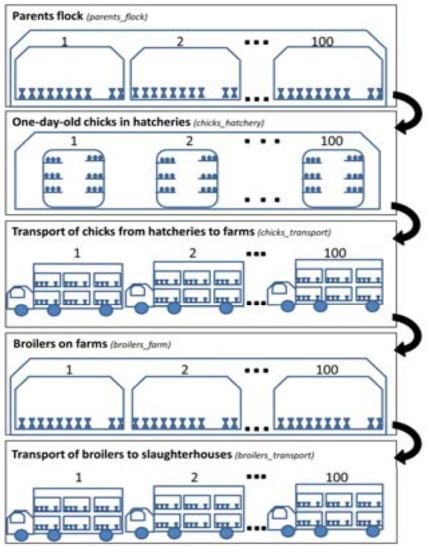
⇨ Accuracy = $\frac{271}{520} \approx 52\%$

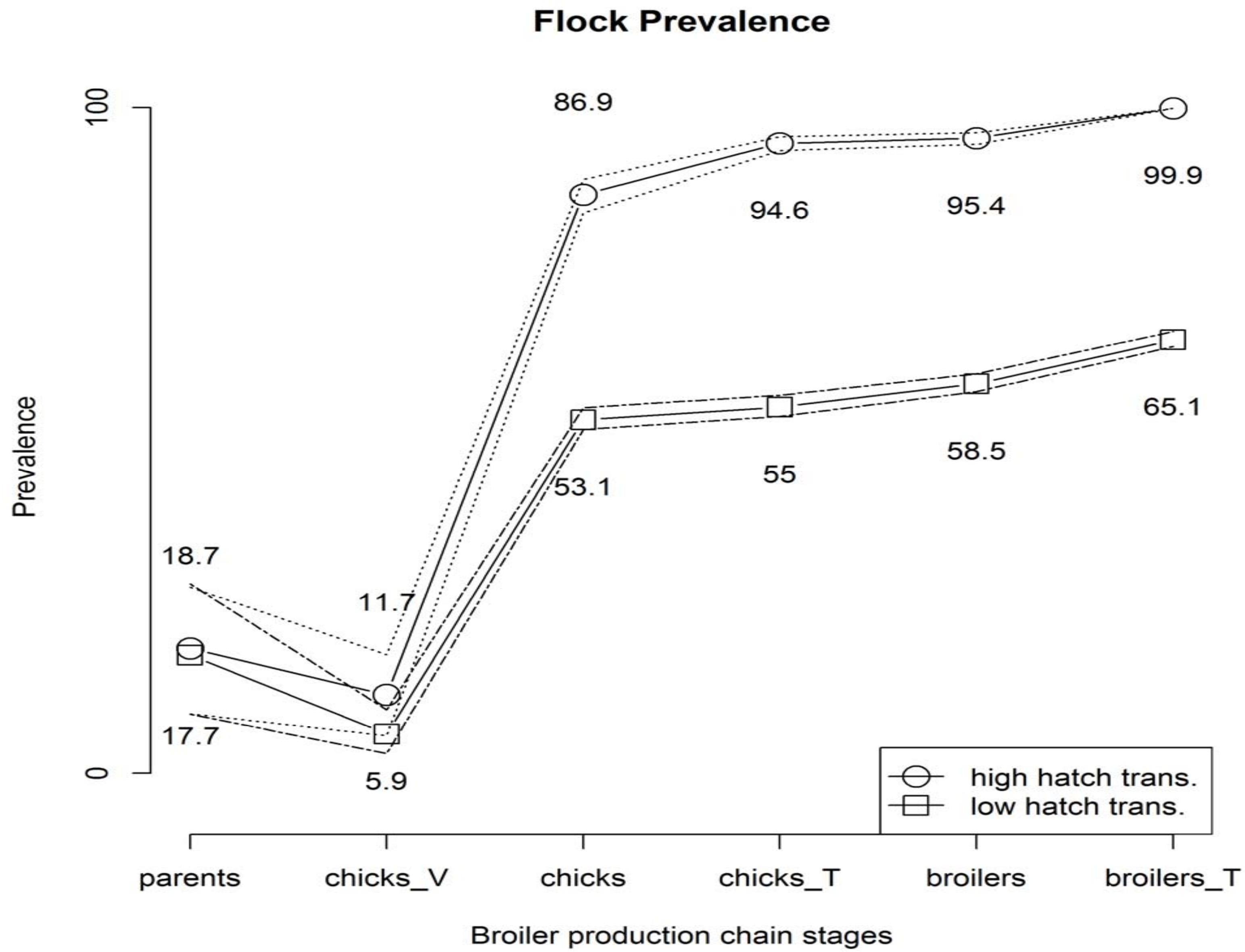
Summarising Results

Based on the five scenarios value ranges for the coeficients hs, hp and ho were found and used in the transmission model

	Minimun	Maximum
hs	-73.89	8.14
hp	0.17	13.86
ho	-8.30	7.46

Using the parameterisation in broiler production chain model





Conclusion Possibilities and limits

Possibilities:

Logistic model useful in predicting outcomes

Helps identify directions which to investigate further

Limits:

Numeric values of regression coefficients have no direct real world interpretation

At the current stage our regression model is limited in grain size of analysis and therefore limited in giving particular hints on interventions for farmers

GERMAN FEDERAL INSTITUTE
FOR RISK ASSESSMENT

15 YEARS
Science in the Service of Humanity

BfR
Bundesinstitut für Risikobewertung

RESET
Resistenzen
bei Tier und Mensch -
gemeinsame Forschung in Deutschland

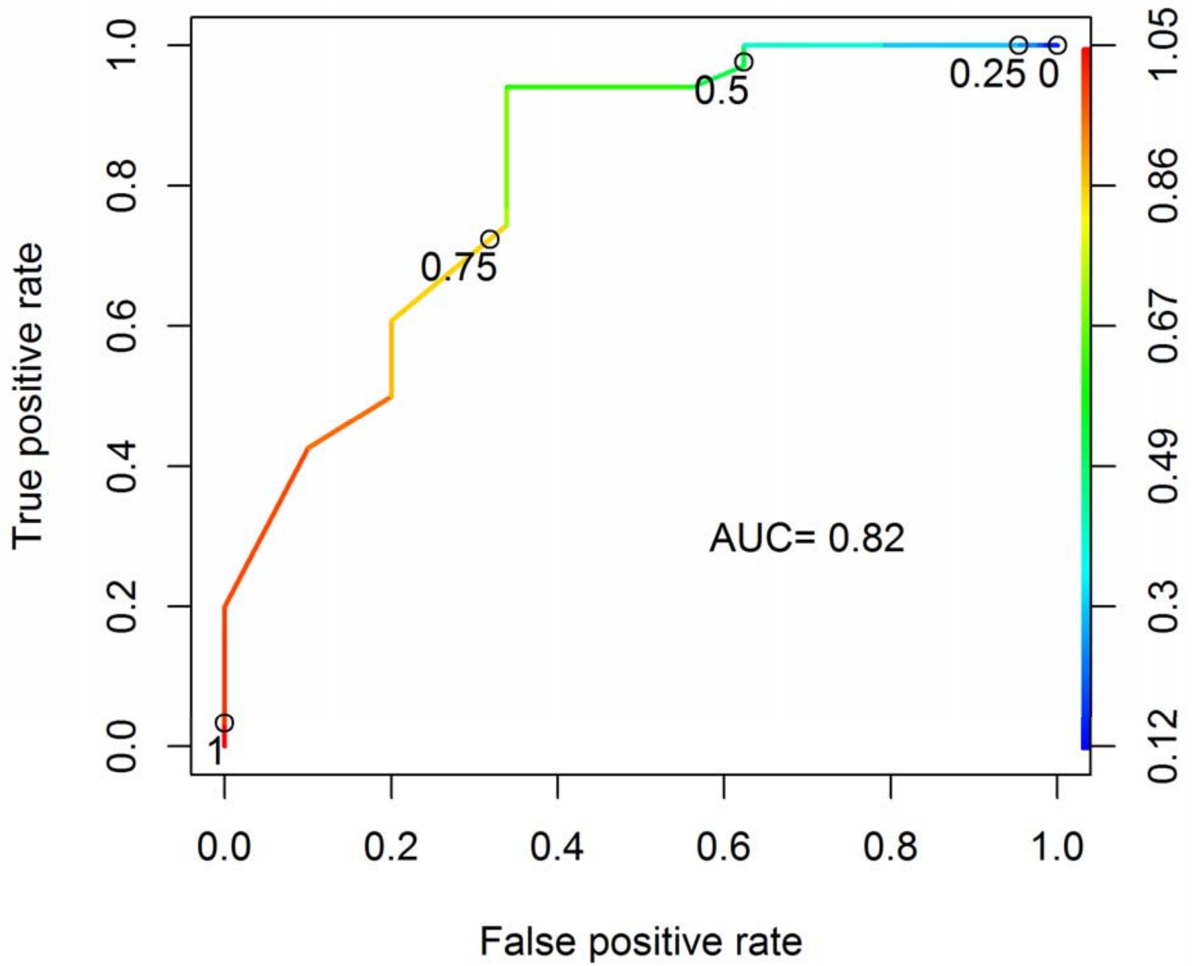
Thank you for your attention

German Federal Institute for Risk Assessment
Max-Dohrn-Str. 8-10 ● 10589 Berlin, GERMANY
Phone +49 30 - 184 12 - 0 ● Fax +49 30 - 184 12 - 47 41
bfr@bfr.bund.de ● www.bfr.bund.de/en

Backup

Choosing thresholds – ROC curve

Receiver Operating Characteristic curve = ROC curve for scenario 4



Predicting with the model

Prediction via threshold:

Example – for given independent variables and coefficients say that

$$\text{Probability that flock got col}onised = \frac{1}{1 + e^{-(2.211 \cdot 0.5 + 1 \cdot 0 - 0.7)}} \approx 0.6$$

Introduce threshold, say 0.5

Prediction: Flock is predicted to become colonized if probability > threshold

In our case: since 0.6 > 0.5 the flock is predicted to become colonized

How accurate are model predictions?

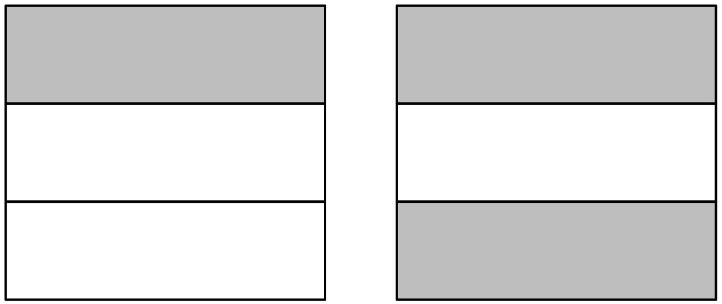
Accuracy = number of correct prediction / number of predictions

Example: Introduce threshold, say 0.5

$$\text{Probability that flock got colonised} = \frac{1}{1 + e^{-\left(8.143 \cdot \frac{Prev_0}{nf} + 13.86 \cdot col\_hist - 8.295\right)}}$$

Farm

with 3 broiler houses



Production Cycle 1

Production Cycle 2

0.9997
0.0538
0.0038

Prediction for Production Cycle 2

$$\textbf{Accuracy} = \frac{\mathbf{1}}{\mathbf{3}} \approx \mathbf{33\%}$$