Hierarchical Modelling for Piglet Production

Prof. Dr. Katja Ickstadt, Faculty of Statistics, TU Dortmund University

In collaboration with

Fiona Mers and Timur Tug, Faculty of Statistics, TU Dortmund University,

Lothar Kreienbrock, Institute of Biometry, Epidemiology and
Information Processing, University of Veterinary Medicine Hanover

Overview

1. Introduction (VASIB project and data)
2. Analysis and evaluation
3. Conclusion and outlook

1. Introduction (VASIB - project)

VASIB research project: "Reducing the use of antibiotics in pig farming by integrating epidemiological information from clinical, hygiene, microbiological and pharmacological veterinary advice"

- **Collaboration:** Hanover University of Veterinary Medicine Foundation with veterinarians from the Vet-Team Reken group practice and the Institute of Farm Animal Genetics at the Friedrich Loeffler Institute and the Free University of Berlin
- **Duration**: 2016 – 2019
- **Funded**: Federal Ministry for Food and Agriculture

Aim: Use all available data from practice and additional scientific advice to optimize treatment strategies and improve general animal health through management advice.

1. Introduction (general aim of the study)

- Respiratory diseases with symptoms such as coughing are not uncommon
- Pigs are administered antibiotics and healthy animals can also be treated
  ➔ Development of antibiotic resistance
- A distinct hygiene concept is important in modern livestock farming
- **Aim**: explore influencing factors for the occurrence of coughs
  ➔ Prevention; recommendations for farmers

1. Introduction (data)

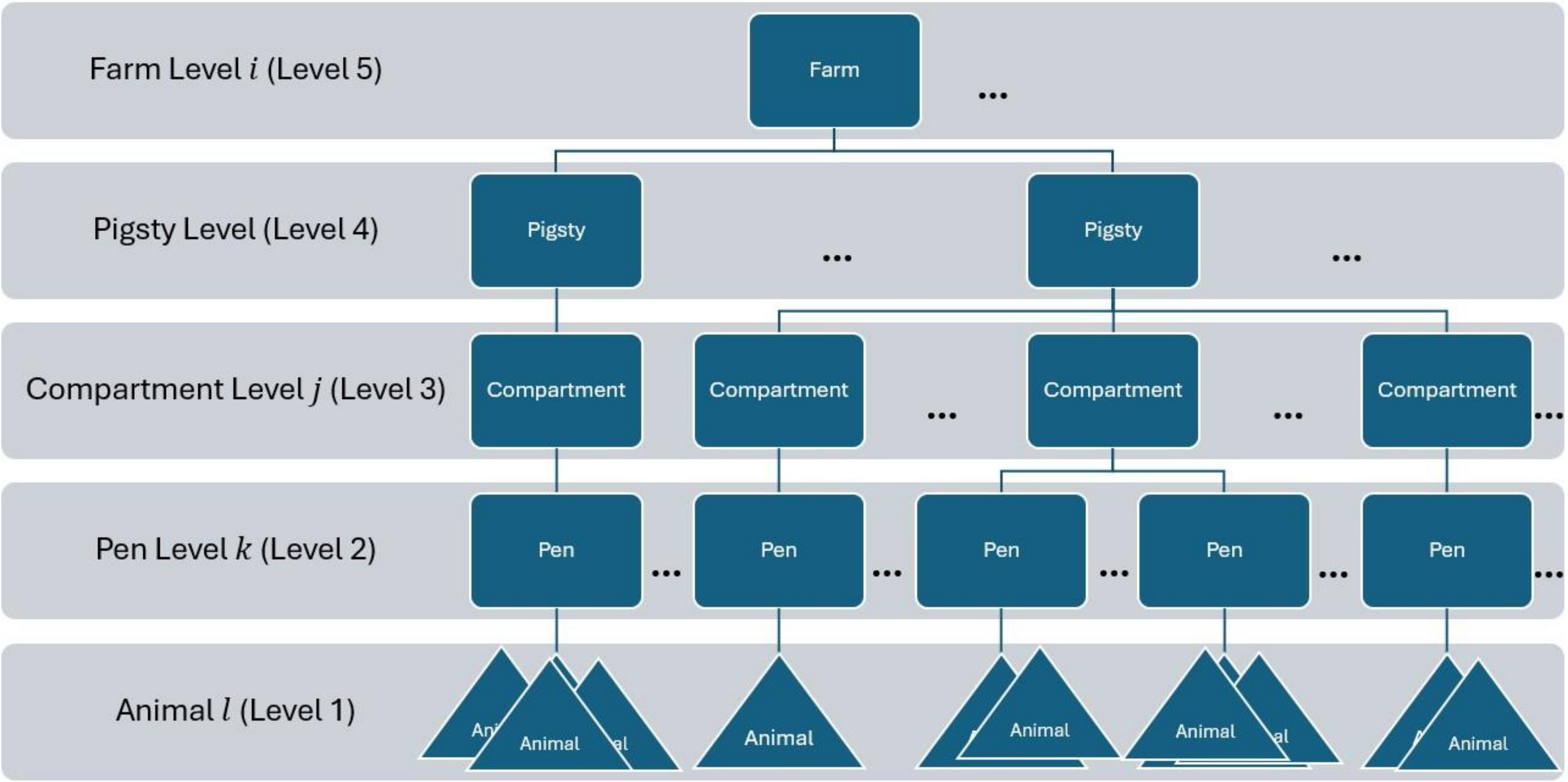
- 30 selected farms in Germany (North Rhine-Westphalia and Schleswig-Holstein) with piglet rearing and repeated problems with respiratory diseases were followed over a period of one year
- Variables were collected on two measurement dates (the first in 2016 and the final visit in 2017) with different samples of pigs
- The data structure is complex and data come from different sources

1. Introduction (preliminary statistical considerations)

- Variable selection and variable reduction
- Missing values are imputed
- Final data set for the following analysis:
  - 29 (1298) variables with 450 (1076) observations
  - 191 pigs in 63 compartments from 30 farms
- Data: nested hierarchical structure

1. Introduction (data)

Hierarchical data structure in VASIB



Observation numbers: 30 farms, 72 pigsties, 130 compartments, 300 pens and 450 animals

2. Analysis - statistical methods

Aims:

- Hierarchical models representing the hierarchical structure
- Variability within as well as between clusters
- Logistic regression: independent observations
- Simplest hierarchical model: random intercept model (random effects)

➔ Identify risk factors for coughing using frequentist and Bayesian models

$$\mathrm{logit}(\pi_{ijkl}) = \beta_0 + \sum_{p=1}^{P} \beta_p \mathrm{x}_{ijkl} + u_i + v_{ij} + w_{ijk},$$

random intercepts: $u_{\mathrm{i}} \sim N(0, \sigma_{\mathrm{u}}^2)$, $v_{\mathrm{ij}} \sim N(0, \sigma_{\mathrm{v}}^2)$ and $w_{ijk} \sim N(0, \sigma_w^2)$,

independent variables x_{ijkl}: e.g. soil condition for animal *l* in pen *k*, compartment *j* and farm *i*

2. Analysis - Model comparisons

First analysis:

- Hierarchical Bayesian model (logistic model with partial pooling)
  - Non-informative prior
  - Informative prior

Question 1: is it necessary to include the hierarchy in the model?

- Non-hierarchical Bayesian model (logistic model, Complete pooling)

Question 2: does prior knowledge (a priori distribution) have impact on results?

- Hierarchical frequentist model (logistic model)

All analyses calculated in R, Bayesian analyses conducted with brms R package.

Model comparisons

Frequentist logistic regression

- Random effects for compartments and pens
- Model convergence issues (small cluster sizes)

Bayesian models

- Non-hierarchical and hierarchical (with/without informative priors)
- Used MCMC sampling via brms and Stan in R

Key metrics for evaluation:

corrected Akaike Information Criterion (AICc), marginal and conditional R^2, and intra-class correlation coefficients (ICCc and ICCadj).

2. Analysis

- Data preparation: standardizing and transforming
- Prevalence for pig coughing in “average” pen: 38%
- Hardly any variance between pens, quite large variance between compartments
- Variables of interest:
  - Stocking density,
  - Water flow rate,
  - Floor condition,
  - Clinical and laboratory variables (all detrimental influence on chance of coughing)
- Prior specification for Bayesian models

Prior specification for different Bayesian models (BM) 1-5

BM 1 (non-hierarchical)	$\beta_0 \sim \mathcal{N}\ (0, 50)$ $\beta_1, \dots, \beta_{15} \sim \mathcal{N}\ (0, 100)$
BM 2 (hierarchical) (non-informative)	$\beta_0 \sim \mathcal{N}\ (0, 50)$ $\beta_1, \dots, \beta_{15} \sim \mathcal{N}\ (0, 100)$ $\sigma_0 \sim InvG\ (0.01, 0.01)$ $\sigma_{0i} \sim InvG\ (0.01, 0.01)$
BM 3 (hierarchical) (non-informative)	$\beta_0 \sim \mathcal{N}\ (0, 100)$ $\beta_1, \dots, \beta_{15} \sim \mathcal{N}\ (0, 1000)$ $\sigma_0 \sim InvG\ (0.01, 0.01)$ $\sigma_{0i} \sim InvG\ (0.01, 0.01)$
BM 4 (hierarchical) (highly-informative)	$\beta_0 \sim \mathcal{N}\ (0, 1)$ $\beta_1, \dots, \beta_{15} \sim \mathcal{N}\ (0, 10)$ $\sigma_0 \sim InvG\ (1, 1)$ $\sigma_{0i} \sim InvG\ (1, 1)$
BM 5 (hierarchical) (highly-informative)	$\beta_0 \sim \mathcal{N}\ (0, 1)$ $\beta_1, \dots, \beta_{15} \sim \mathcal{N}\ (0, 1)$ $\sigma_0 \sim InvG\ (0.5, 0.5)$ $\sigma_0 \sim InvG\ (0.5, 0.5)$

2. Evaluation

Estimated odds ratios for a selection of variables with associated 95% confidence and credible intervals

model	frequentistic (non-Bayesian) hierarchical	Bayesian non-hierarchical non-informative priors	Bayesian hierarchical non-informative priors	Bayesian hierarchical highly-informative priors
stocking density	1.38	**68.04**	**13.25**	**13.25**
	[0.96, 1.97]	**[51.71, 89.52]**	**[8.95, 19.60]**	**[8.95, 19.60]**
pen size	0.83	0.94	0.97	0.97
	[0.51, 1.37]	[0.70, 1.26]	0.54, 1.74]	[0.53, 1.78]
age	1.27	1.01	1.02	1.02
	[0.84, 1.92]	[0.76, 1.32]	[0.65, 1.60]	[0.64, 1.64]
soil condition	**5.31**	**4.18**	**6.49**	**6.17**
(as new vs. worn)	**[1.71, 16.43]**	**[2.36, 7.61]**	**[1.57, 30.57]**	**[1.42, 29.37]**
water flow rate	**0.57**	1.00	1.00	1.00
	[0.33, 0.96]	[0.74, 1.34]	[0.49, 2.02]	[0.49, 2.02]
air pressure	0.70	0.97	0.96	0.96
	[0.40, 1.23]	[0.73, 1.27]	[0.46, 2.02]	[0.47, 1.99]
CO_2-value	**2.13**	1.00	1.00	1.00
	[1.13, 4.03]	[0.73, 1.37]	[0.43, 2.33]	[0.43, 2.33]
NH_3-value	0.81	0.97	0.95	0.95
	[0.49, 1.33]	[0.75, 1.25]	[0.49, 1.86]	[0.49, 1.86]
H_2S-value	0.79	**0.57**	0.61	0.62
	[0.48, 1.31]	**[0.44, 0.74]**	[0.31, 1.19]	[0.32, 1.21]
temperature	3.82	**5.00**	4.01	3.78
	[0.32, 45.98]	**[1.19, 23.57]**	[0.16, 107.77]	[0.17, 81.45]

2. Evaluation

Estimated odds ratios for a selection of variables with associated 95% credible intervals

Factor	BM 1 non-hierarchical	BM 2 hierarchical non-informative	BM 3 hierarchical non-informative	BM 4 hierarchical highly- informative	BM 5 hierarchical highly- informative
stocking density (in animals/m²)	**65.07 [49.46,85.61]**	**13.55 [9.23, 19.90]**	**14.11 [9.57, 20.80]**	**13.71 [9.30, 20.21]**	**7.24 [5.07, 10.36]**
pen size (in m²)	0.94 [0.70, 1.25]	0.97 [0.53, 1.75]	0.97 [0.53, 1.74]	0.97 [0.52, 1.78]	0.96 [0.56, 1.67]
age (in days)	1.01 [0.76, 1.34]	1.02 [0.64, 1.64]	1.02 [0.64, 1.63]	1.02 [0.64, 1.63]	1.02 [0.67, 1.57]
floor condition (as new vs. worn)	**4.19 [2.34, 7.70]**	**6.30 [1.56, 28.81]**	**6.47 [1.56, 28.80]**	**6.11 [1.53, 26.06]**	3.04 [0.98, 8.94]
water flow rate (in ml/min)	1.00 [0.74, 1.34]	1.00 [0.49, 2.02]	1.00 [0.50, 2.00]	1.00 [0.51, 1.95]	1.00 [0.55, 1.83]
air pressure (in Pa)	0.70 [0.74, 1.27]	0.97 [0.46, 2.01]	0.96 [0.45, 2.04]	0.96 [0.47, 1.99]	0.98 [0.52, 1.84]
CO_2-value (in ppm)	1.00 [0.73, 1.38]	1.00 [0.43, 2.33]	1.00 [0.43, 2.34]	1.00 [0.43, 2.33]	1.00 [0.50, 2.00]
NH_3-value (in ppm)	0.97 [0.75, 1.26]	0.95 [0.49, 1.87]	0.95 [0.49, 1.85]	0.95 [0.50, 1.84]	0.97 [0.55, 1.74]
H_2S-value (in ppm)	0.58 [0.45, 0.74]	0.60 [0.31, 1.15]	0.61 [0.31, 1.20]	0.62 [0.32, 1.20]	0.75 [0.42, 1.36]
temperature (in °C)	5.02 [1.24, 24.56]	4.01 [0.18,102.86]	4.01 [0.17,110.77]	3.78 [0.15, 96.97]	1.38 [0.27, 6.91]

2. Evaluation

Model performance

- Bayesian hierarchical model with informative priors (BM5) outperformed others: best predictive accuracy (LOO-CV), highest Bayes R^2
- Non-hierarchical model is unsuitable for the problem
- Compared to the hierarchical Bayesian models, the frequentist model underestimates the variances
- Prior choice: only little influence on results of Bayesian models in this case

Important risk factors

- Independent of the choice of model:
  - ➔ Significant influence: Floor condition within a compartment (worn floor risk factor for coughing)
- Stocking density: effect varies with model structure
- Water flow rate: protective factor

3. Conclusion and outlook (methods)

- The statistical instrument is a hierarchical logistic regression model (Frequentist/Bayesian)
- Preprocessing the data is essential but costly
- The use of hierarchical Bayesian models is necessary (due to structure and convergence problems of non-Bayesian analyses)
- Different priors provide similar results
- Include interactions in the models

3. Conclusion and outlook (application)

- Objective: identify factors that influence coughing in pigs and provide farmers with information on animal health and preventive measures.
- There is little variation in coughing between farms, but a large variation within individual farms.
- Floor condition appears to be the most important variable in the dataset.
- Consider other target variables (e.g., sneezing).
- Create group variables / different groups of variables (e.g., animal hygiene, barn hygiene, etc.).

technische universität
dortmund

STIFTUNG · TIERÄRZTLICHE HOCHSCHULE HANNOVER

www.statistik.tu-dortmund.de

- Variable selection and reduction:
  - content preselection,
  - highly correlated variables (ρ=0.8),
  - variables with more than 50% missing values
  - variables with low variability (usually with only one characteristic)

- Remaining missing values imputed using
  - “predictive mean matching” (continuous variables)
  - “proportional odds model” (ordinal variables)

- Final data set: 29 (1298) variables with 450 (1076) observations 191 pigs in 63 compartments of 30 farms

Comparison overview of the four underlying models

Model	Model Characteristics				
	Flexibility	Handles Clustering	Handles Small Data	Includes Prior Knowledge	Complexity
Frequentistic Hierarchical	Moderate	Yes	No	No	Low
Bayesian Non-Hierarchical Non-Informative	Moderate	No	Yes	No	Moderate
Bayesian Hierarchical Non-Informative	High	Yes	Yes	No	High
Bayesian Hierarchical Highly Informative	Very High	Yes	Yes	Yes	Very High

Model comparison with the Bayesian models

Model	expected log pointwise predictive density (elpd_loo)	se_diff
BM5	0.0	0.0
BM3	-1.6	1.6
BM2	-2.0	1.6
BM4	-2.7	1.5
BM1	-42.5	9.1