Model Comparison for Genetic Analysis of Clinical Mastitis in Norwegian Red cows

Linear, Threshold,
Censored Threshold and
Zero Inflated Poisson models.

A. I. Vazquez[1], B. Heringstad[2], M. A. Perez-Cabal[3], M. Rodrigues-Motta[1],
K. A. Weigel[1], D. Gianola[1], G.J.M. Rosa[1], Y. M. Chang[1]

[1]University of Wisconsin – Madison
[2] Norwegian University of Life Sciences
[3] Universidad Polytécnica de Madrid

61st EAAP, Heraklion, Crete Island, Greece , August 23-27, 2010,
Session 12, no 4

Introduction

Many different models suggested for genetic analyses of clinical mastitis defined as single trait, multiple traits or longitudinal trait

e.g. linear, logit, threshold, multivariat threshold, longitudinal threshold, censored threshold, ZIP model, survival analyses, ……

Model comparison a challenge!

Different scales and different trait definitions

Mastitis traits

- Absence/presence during a period,

  y=0,1.

  Possible models: linear model, binary models (logit or probit threshold).

- Count of cases during a period,

  y=0, 1, 2, …

  Possible models: linear, ordinal threshold, Poisson and Zero-inflated Poisson (ZIP) models.

Objectives

Compare the performance of
Zero-Inflated Poisson (ZIP) model with
censored ordinal threshold and **linear models**
for count of mastitis cases
and
threshold and **linear models**
for absence/presence of mastitis,
for genetic analysis of clinical mastitis
in Norwegian Red cows.

Materials and Methods

Data

- 620,492 1st lactation Norwegian Red cows
- Daughters of 3,064 sires
- From 51,808 herd-5-year groups

- Mastitis data from the cows health card: veterinary treatments of clinical mastitis from -30 d before to 300 d after first calving.

Traits

- **CM:** Absence / presence of mastitis in the period from 30 days before to 300 days after first calving
- **NCM**: Number of CM treatments during 1st lactation
  - From 30 d before to 300 d after calving
  - Min. 5 days between treatments
  - 3 categories: 0, 1, and 2 or more cases of CM
  - Cows culled before 300 d were censored

No of cows with NCM	620,492
Mean NCM	0.22
% censored cows for NCM	30

Distribution of cows over NCM categories

NCM	All cows	Censored	Non-censored
0	81	77	82
1	16	19	15
≥2	3	4	3

Models – CM (binary trait)

Threshold liability model

$$\lambda = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_h\mathbf{h} + \mathbf{Z}_s\mathbf{s} + \mathbf{e}$$

λ	vector of unobserved liabilities of CM
β	vector of systematic effects
	age (21 classes) and mo-yr (288 classes) of first calving
h	vector of herd-5-year period of calving effects (51,808 levels)
s	vector of sire transmitting abilities
e	vector of residual effects
X, $\mathbf{Z}_h$, and $\mathbf{Z}_s$	incidence matrices

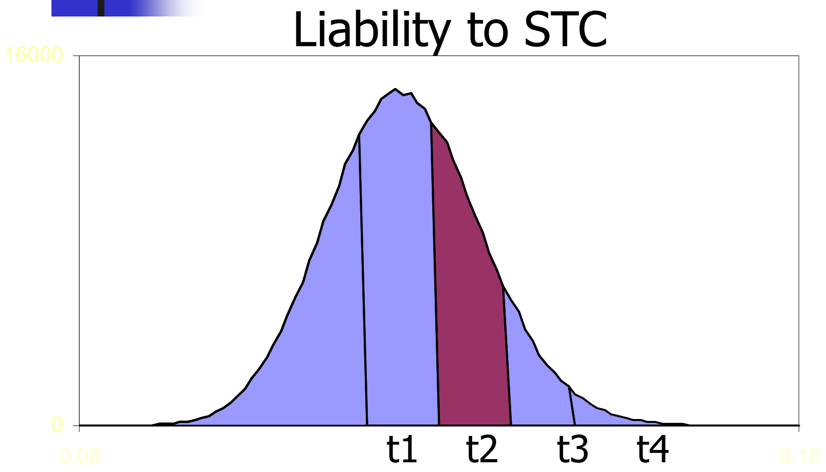
Models – NCM (ordered categorical trait)

- Linear model
- Censored threshold model

  (Chang et al. 2006; Heringstad et al. 2008)
- Zero-Inflated Poisson (ZIP)

  (Rodriguez-Motta et al., 2006)

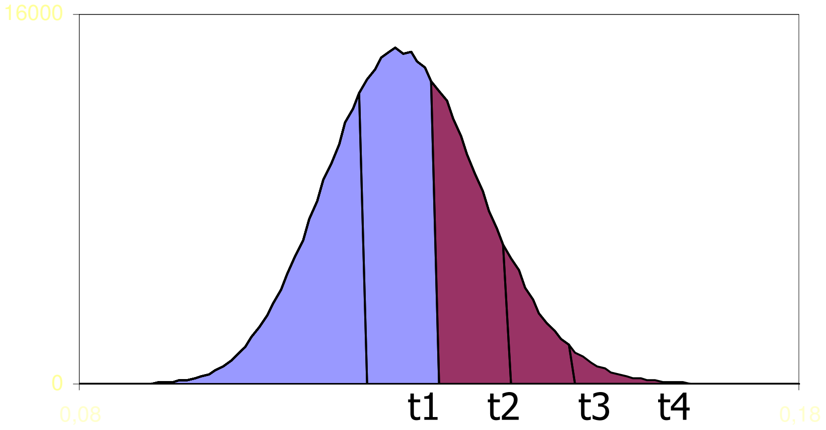
Taking censoring into account



Non-censored cow with

STC=3

t2 ≤ liability < t3



Censored cow with
STC=3

liability ≥ t2

ZIP: Zero inflated Poisson

- Account for extra zero's
- Better predictive ability than Poission model for NCM (Rodrigues-Motta et al. 2006)

S: state
$$S = \begin{cases} 0 \text{ (will never have mastitis): } p \\ 1 \text{ (might have mastitis): } 1-p \end{cases}$$

Y (mastitis):{ 0,1,2,…}
$$\begin{cases} P\{Y=0\} = p + (1-p)e^{-\mu}\dfrac{\mu^0}{0!} \\ P\{Y=y\} = (1-p)e^{-\mu}\dfrac{\mu^y}{y!} \end{cases} \qquad y = 0,1,2,\ldots$$

Model comparison

- Predictive ability
- Four-fold cross-validation
- Mean squared error of prediction (MSEP)

Results and discussion

Posterior means of sire, herd-year and residual variances, and heritability.

Variance	**Linear (binary)**	**Threshold (binary)**	**Censored threshold (count)**	**Zero-inflated Poisson (count)**	**Linear (count)**
Residual	0.144	1.000	1.000	0.034	0.220
Herd	0.009	0.132	0.139	0.266	0.014
Sire	0.001	0.019	0.023	0.043	0.002
Heritability	0.03	0.07	0.08	-	0.03

Correlation between predictions of sire effects (above the diagonal), and herd-year effects (below the diagonal).

	Linear (binary)	Threshold	CT	ZIP	Linear (counts)
Linear (binary)	-	1.00	0.96	0.97	0.98
Threshold	1.00	-	0.97	0.98	0.97
MT	0.95	0.96	-	0.97	0.97
ZIP	0.96	0.97	0.97	-	0.99
Linear (counts)	0.96	0.95	0.96	0.99	-

Top 10 sires from the linear model (for NCM) and their ranking for the other models.

Sire	**Linear (binary)**	**Threshold**	**CT**	**ZIP**	**Linear (count)**
1087	1	1	2	1	1
1312	2	2	1	2	2
2020	12	18	14	10	3
1796	6	12	10	15	4
1512	14	17	8	6	5
1239	3	3	4	4	6
1825	4	13	6	29	7
2379	5	5	5	5	8
1485	20	25	19	14	9
2083	11	10	32	17	10

Mean squared error of prediction in cross-validation, with binary response (CM).

	Linear (binary)	**Threshold**	**CT**	**ZIP**	**Linear (count)**
Total	0.160	0.161	0.180	0.157	0.163
Healthy cows	0.042	0.039	0.099	0.035	0.057
Diseased cows	0.659	0.675	0.521	0.674	0.610

Mean squared error of prediction in cross-validation, with the counts response (NCM).

	Linear (binary)	Threshold	CT	ZIP	Linear (count)
Total	0.245	0.246	0.258	0.243	0.246
Healthy cows	0.042	0.039	0.099	0.035	0.057
Cows with 1 mastitis	0.659	0.675	0.520	0.674	0.609
Cows with ≥2 mastitis	3.283	3.318	2.950	3.319	3.171

Conclusions

- Overall MSEP was smaller for ZIP
- Healthy cows, better predicted by ZIP model followed closely by threshold
- Cows with mastitis, better predicted with censored-threshold model, followed by the linear for counts model