69th Annual Meeting of the European Association for Animal Production
Dubrovnik, Croatia, 27-31 August 2018

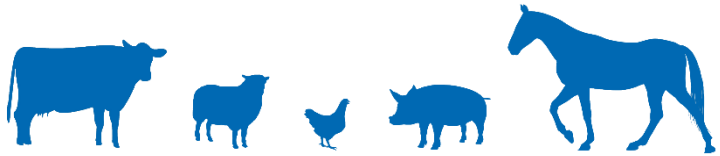
Survival analysis for health traits in German Holstein dairy cows

T. Shabalina, T. Yin, S. König

Institute of Animal Breeding and Genetics, Justus-Liebig-University Giessen
Ludwigstrasse 21 b, 35390 Giessen, Germany

30.08.2018

JLU



JUSTUS LIEBIG UNIVERSITÄT GIESSEN

Background:

- **Longevity:** Challenge in genetic evaluations because of censored data
- **Current approach:** Utilization of indicator traits (e.g., conformation and fertility traits) as early predictors for longevity in survival analysis
- **Nowadays:** German breeding organizations implemented direct health trait recording schemes
- **Alternative:** Utilization of direct health traits as early predictors in genetic evaluations for functional longevity

Aims of the study

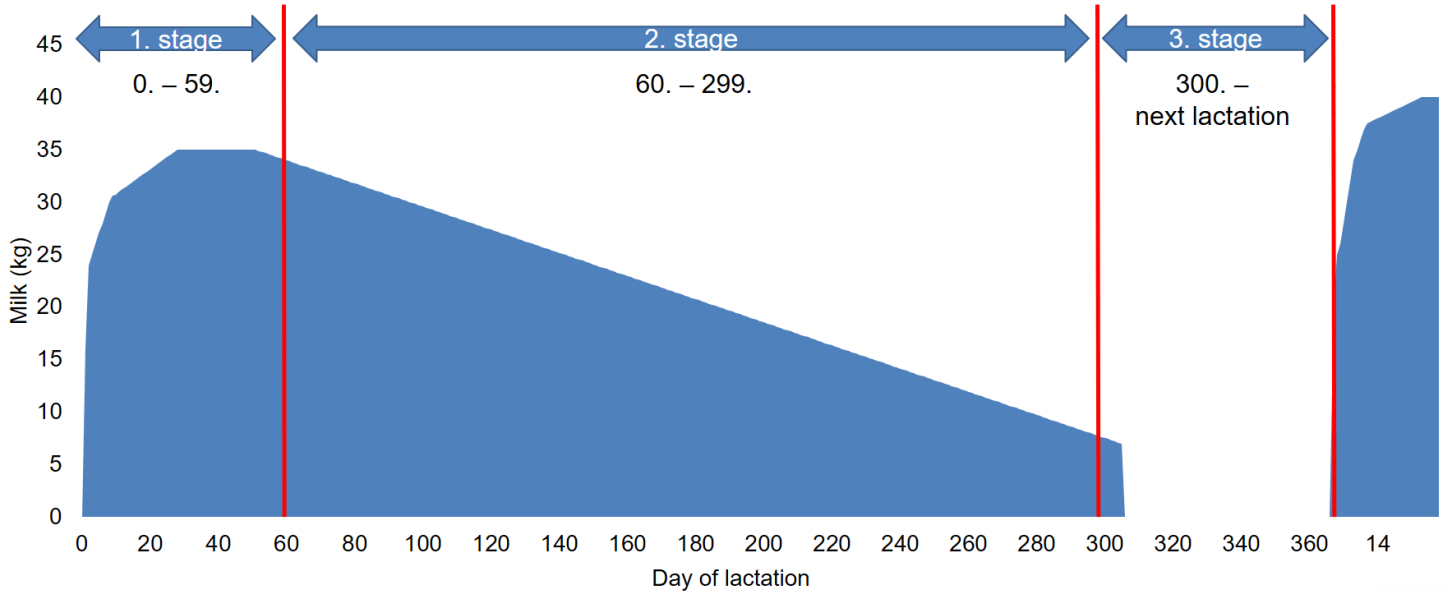
- Analysis of the effects of 13 different diseases on functional longevity (FL) using the Weibull baseline hazard function
- Analysis of genetic parameters for functional longevity
- GWAS for functional longevity and consideration of important identified SNPs in survival analysis

Datasets

	Full dataset	Range over 10 groups	Cows with SNPs
No. of cows	129,386	12,385 – 13,000	
No. of censored cows	39,171	3,084 – 3,999	
% of censored cows	30.27	29.74 – 30.76	
Average productive life (days)	977.10	970.38 – 983.09	

Illumina Bovine SNP50K

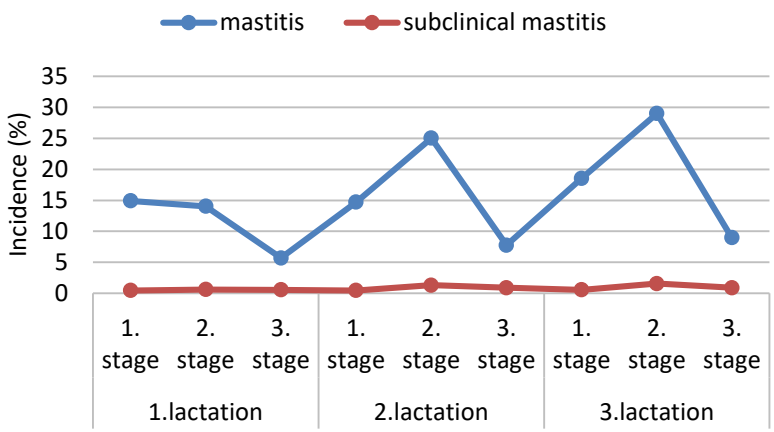
Number of SNP after quality control 40,733



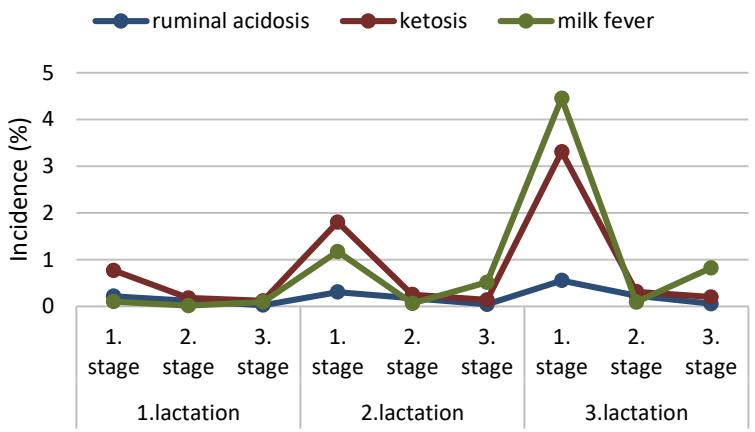
JLU

Incidences of diseases

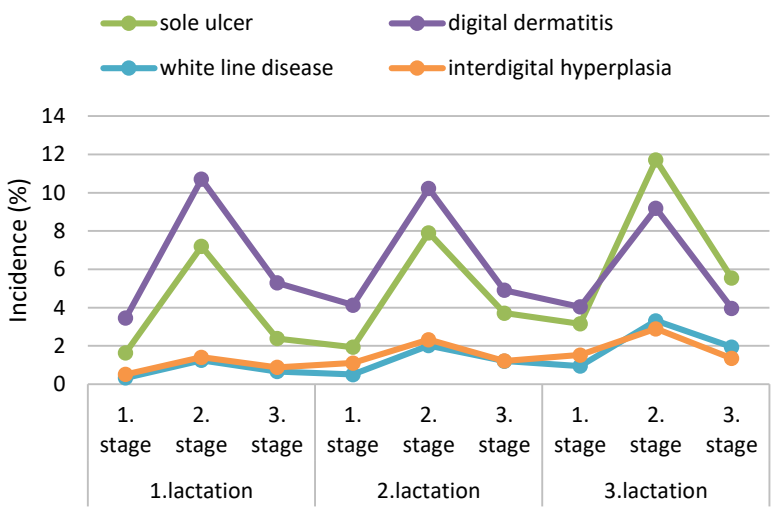
Udder diseases



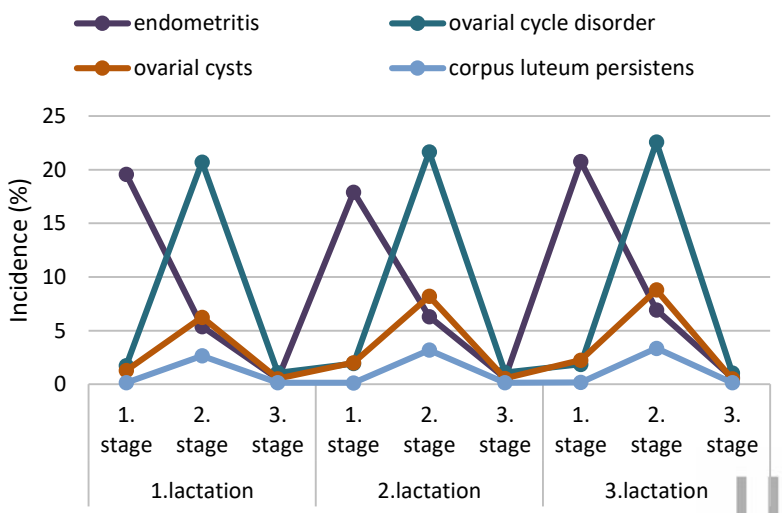
Metabolic diseases



Claw disorders



Reproductive diseases



Hazards model with effect of diseases in different stages

Survival Kit v6.12 (Mészáros et al., 2013)

$$\lambda(t) = \lambda_0(t)\exp\{H_i + FCA_j + \Delta M_k(t') + DIM_l(t') + Y_m(t') + S_n(t') + HS_o(t'') + g_p\}$$

$\lambda(t)$ = hazard function at time t of a cow getting slaughtered;

$\lambda_0(t)$ = Weibull baseline hazard function;

H_i = time-independent fixed effect of the herd (*k* = 1, …, 57);

FCA_j = time-independent fixed effect of age at first calving (*n* = 20.-23.,24.-26.,27.-29.,30.-42. mo.);

$\Delta M_k(t')$ = time-dependent covariate of milk yield from the first test-day in relation to the herd average;

$DIM_l(t')$ = time-dependent covariate of days in milk;

$Y_m(t')$ = time-dependent fixed effect of calving year (*l* = 2008, …,2017);

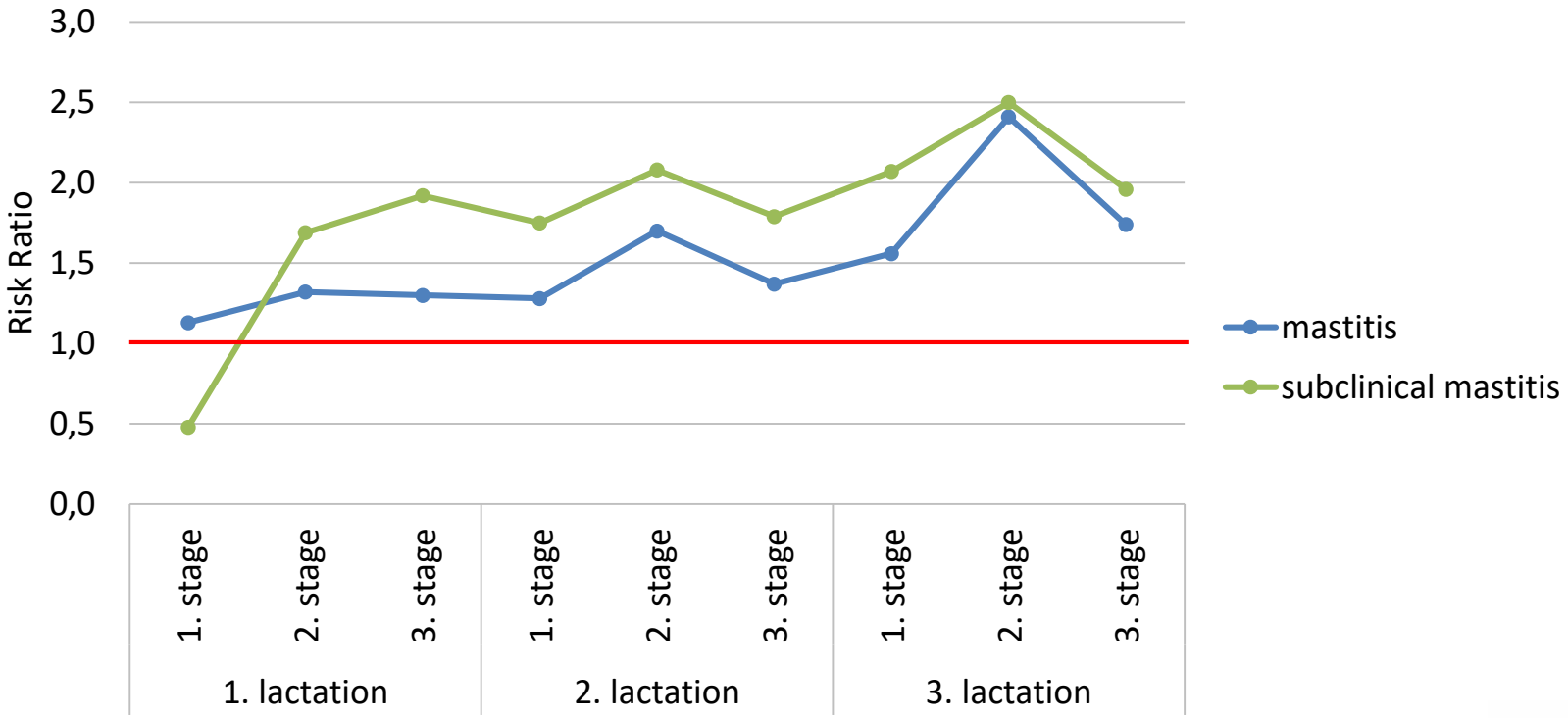
$S_n(t')$ = time-dependent fixed effect of calving season (*m* = Jan.-Apr., May-Sep., Oct.-Dec.) ;

$HS_o(t'')$ = time-dependent effect of health status at lactation stage (0 = healthy, 1 = diseased);

g_p = random genetic animal effect.

Relative risk for culling: Udder diseases

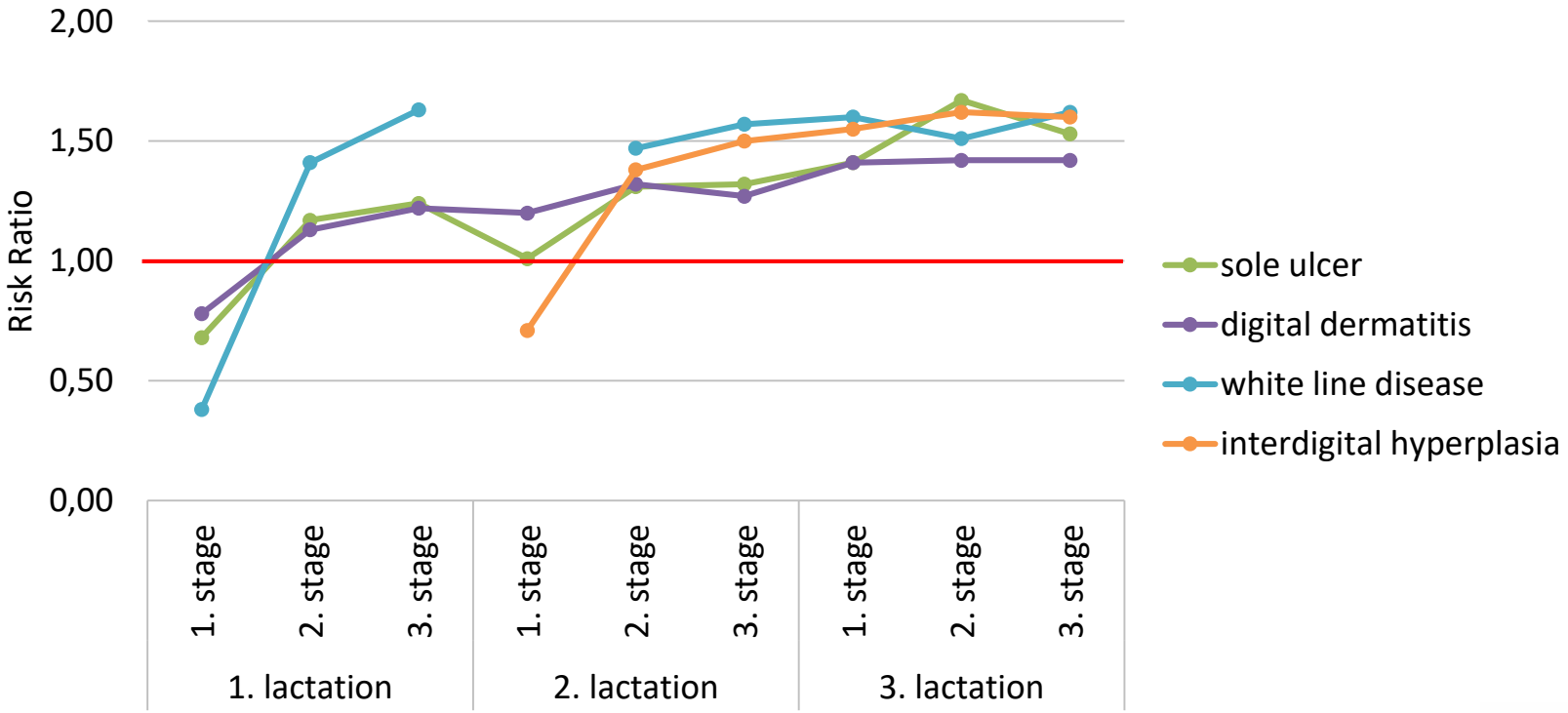
Diseases	1st lactation			2nd lactation			3rd lactation		
	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage
Mastitis	1.13	1.32	1.30	1.28	1.70	1.37	1.56	2.41	1.74
Subclinical mastitis	0.48	1.69	1.92	1.75	2.08	1.79	2.07	2.50	1.96



Relative risk for culling: Claw diseases

Diseases	1st lactation			2nd lactation			3rd lactation		
	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage
Sole ulcer	0.68	1.17	1.24	1.01	1.31	1.32	1.41	1.67	1.53
Digital dermatitis	0.78	1.13	1.22	1.20	1.32	1.27	1.41	1.42	1.42
White line disease	0.38	1.41	1.63	n.s.	1.47	1.57	1.60	1.51	1.62
Interdigital hyperplasia	n.s.	n.s.	n.s.	0.71	1.38	1.50	1.55	1.62	1.60

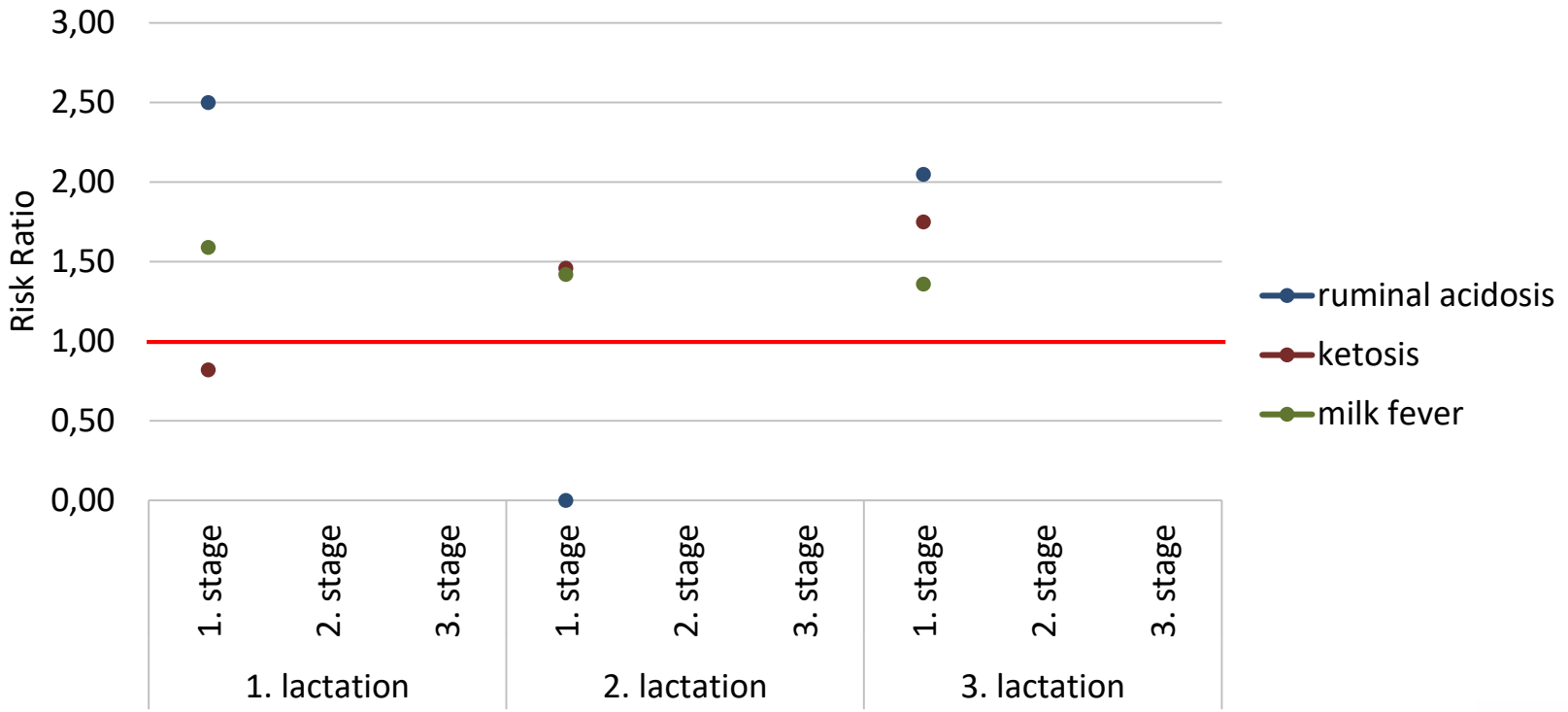
n.s. – not significant



Relative risk for culling: Metabolic diseases

Diseases	1st lactation			2nd lactation			3rd lactation		
	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage
Ruminal acidosis	2.50			n.s.			2.05		
Ketosis	0.82			1.46			1.75		
Milk fever	1.59			1.42			1.36		

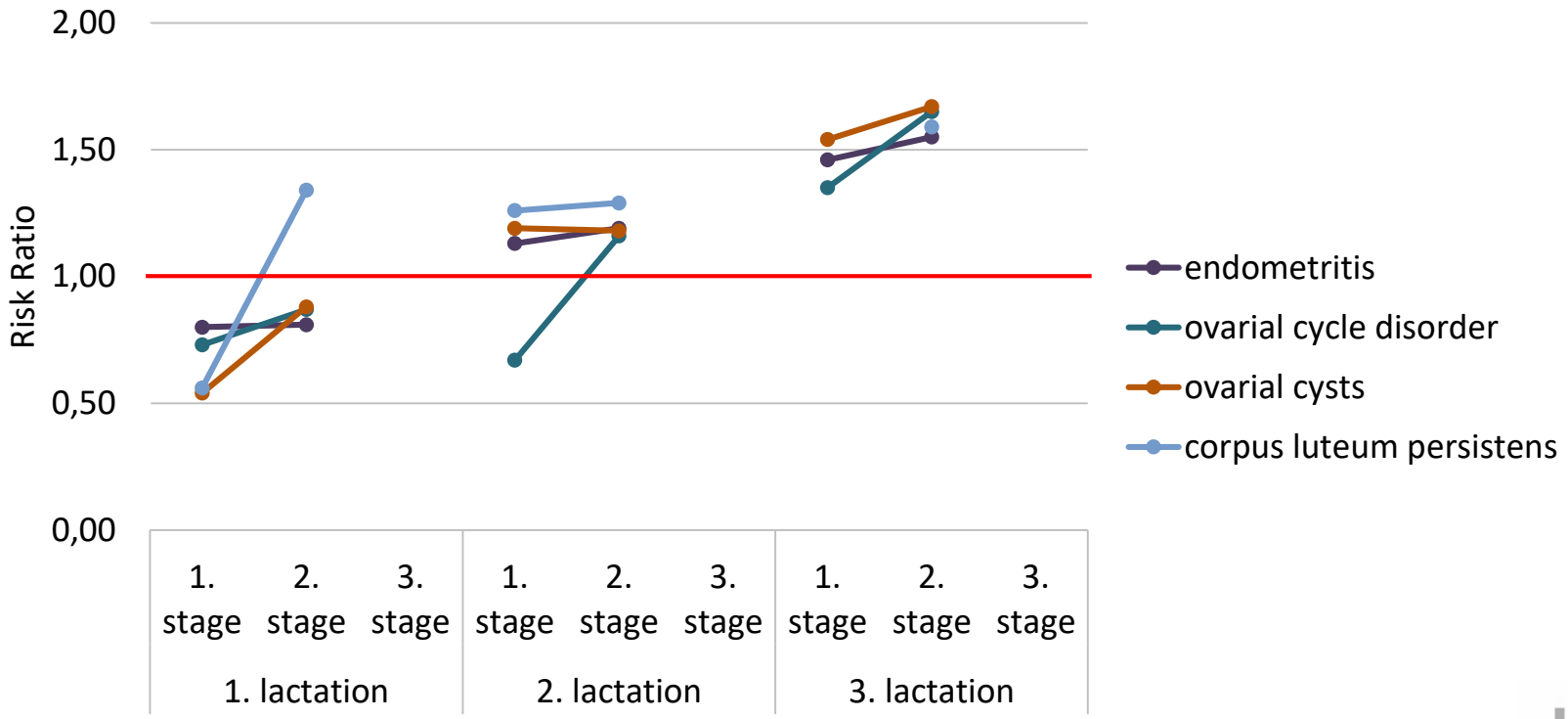
n.s. – not significant



Relative risk for culling: Reproductive diseases

Diseases	1st lactation			2nd lactation			3rd lactation		
	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage
Endometritis	0.80	0.81		1.13	1.19		1.46	1.55	
Ovarial cycle disorder	0.73	0.87		0.67	1.16		1.35	1.65	
Ovarial cysts	0.54	0.88		1.19	1.18		1.54	1.67	
Corpus luteum persistens	0.56	1.34		1.26	1.29		n.s.	1.59	

n.s. – not significant



Heritability for functional longevity

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{1}{p}}$$

σ_g^2 = additive-genetic variance

p = proportion of uncensored records

Disease in model	Genetic variance of FL (sd)	h^2 of FL
Without diseases	0.02 (0.002)	0.02
Mastitis	0.81 (0.003)	0.36
Subclinical mastitis	0.81 (0.003)	0.36
Sole ulcer	0.77 (0.004)	0.35
Digital dermatitis	0.63 (0.006)	0.30
White line disease	0.80 (0.004)	0.36
Interdigital hyperplasia	0.77 (0.004)	0.35
Ruminal acidosis	0.85 (0.003)	0.37
Ketosis	0.86 (0.003)	0.38
Milk fever	0.83 (0.003)	0.37
Endometritis	0.90 (0.002)	0.39
Ovarial cycle disorder	0.60 (0.006)	0.30
Ovarial cysts	0.84 (0.003)	0.37
Corpus luteum persistens	0.80 (0.004)	0.36

Model description

1. DMU v6 (Madsen and Jensen, 2013)

$$y_{ijklmno} = H_i + FCA_j + \Delta M_k + DIM_l + Y_m + S_n + AG_o + eijk_{lmno}$$

$y_{ijklmno}$ = length of productive life (days);

H_i = fixed effect of the herd (k = 1, …, 57);

FCA_j = fixed effect of age at first calving (n = 20.-23.,24.-26.,27.-29.,30.-42. mo.);

ΔM_k = covariate of milk yield from the first test-day in relation to the herd average;

DIM_l = covariate of days in milk;

Y_m = fixed effect of calving year (l = 2008, …,2017);

S_n = fixed effect of calving season (m = Jan.-Apr., May-Sep., Oct.-Dec.) ;

AG_o = Additive genetic animal effect;

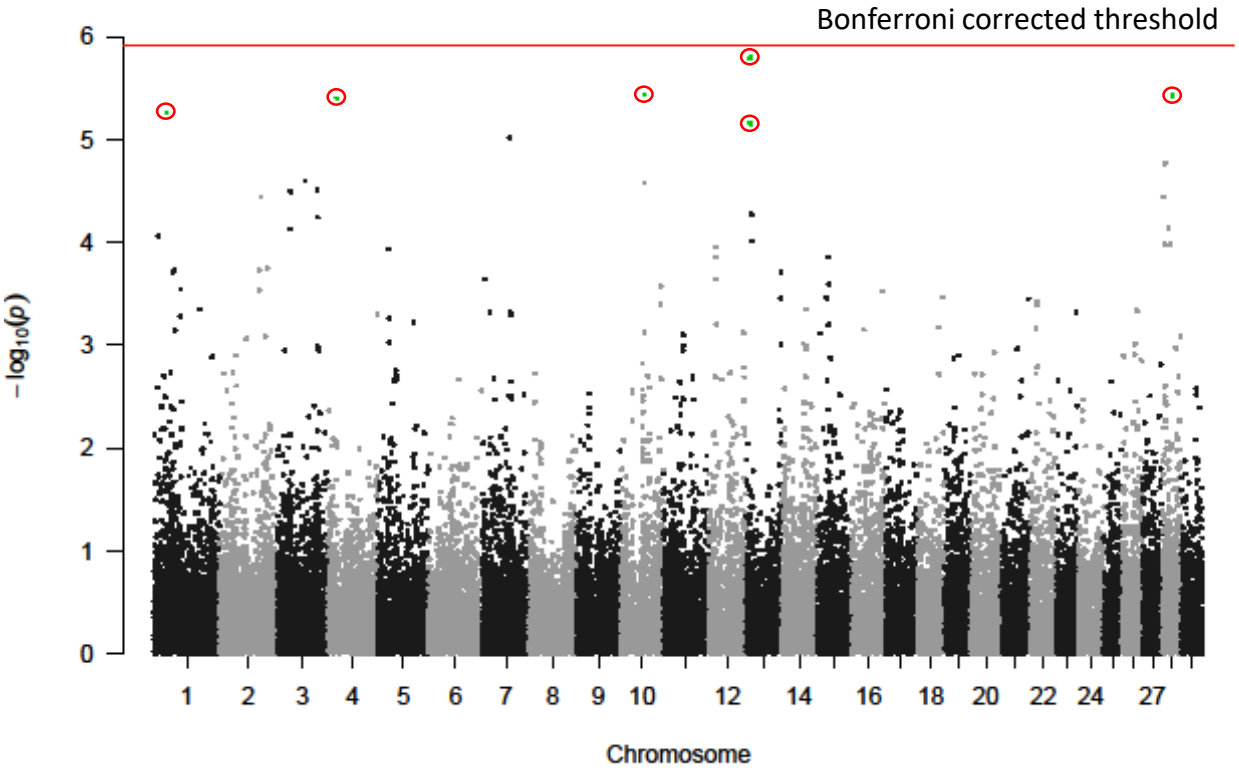
$e_{ijklmno}$ = random residual error.

2. EVB → de-regressed proofs (Garrick et al., 2009)

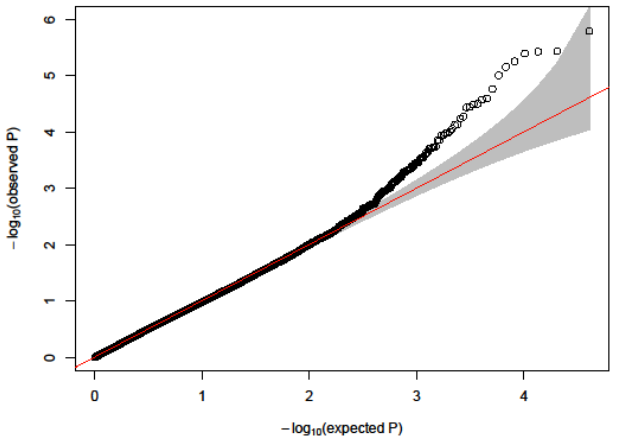
3. GWAS with GCTA v1.91.5beta (Yang et al., 2011)

Association with functional longevity

Manhattan-Plot



QQ-plot



Significant SNPs on functional longevity according to FDR

Abbreviation	Name of SNP	Bovine chromosome	Position	FDR adjusted p-value	% of variance explained by single SNP-marker
SNP1	BTA-18114-no-rs	1	30543596	0.045	5.97
SNP2	ARS-BFGL-NGS-38914	4	22227807	0.041	6.22
SNP3	ARS-BFGL-NGS-19199	10	57518113	0.041	5.97
SNP4	ARS-BFGL-NGS-25574	13	9839639	0.041	6.07
SNP5	ARS-BFGL-NGS-113280	13	9892378	0.047	5.39
SNP6	Hapmap39577-BTA-63767	28	24611840	0.041	5.54

Hazards model with SNP effects

Survival Kit v6.12 (Mészáros et al., 2013)

$$\lambda(t) = \lambda_0(t)\exp\{H_i + FCA_j + \Delta M_k(t') + DIM_l(t') + Y_m(t') + S_n(t') + SNP_o + g_p\}$$

$\lambda(t)$ = hazard function at time t of a cow getting slaughtered;

$\lambda_0(t)$ = Weibull baseline hazard function;

H_i = time-independent fixed effect of the herd (*k* = 1, …, 57);

FCA_j = time-independent fixed effect of age at first calving (*n* = 20.-23.,24.-26.,27.-29.,30.-42. mo.);

$\Delta M_k(t')$ = time-dependent covariate of milk yield from the first test-day in relation to the herd average;

$DIM_l(t')$ = time-dependent covariate of days in milk;

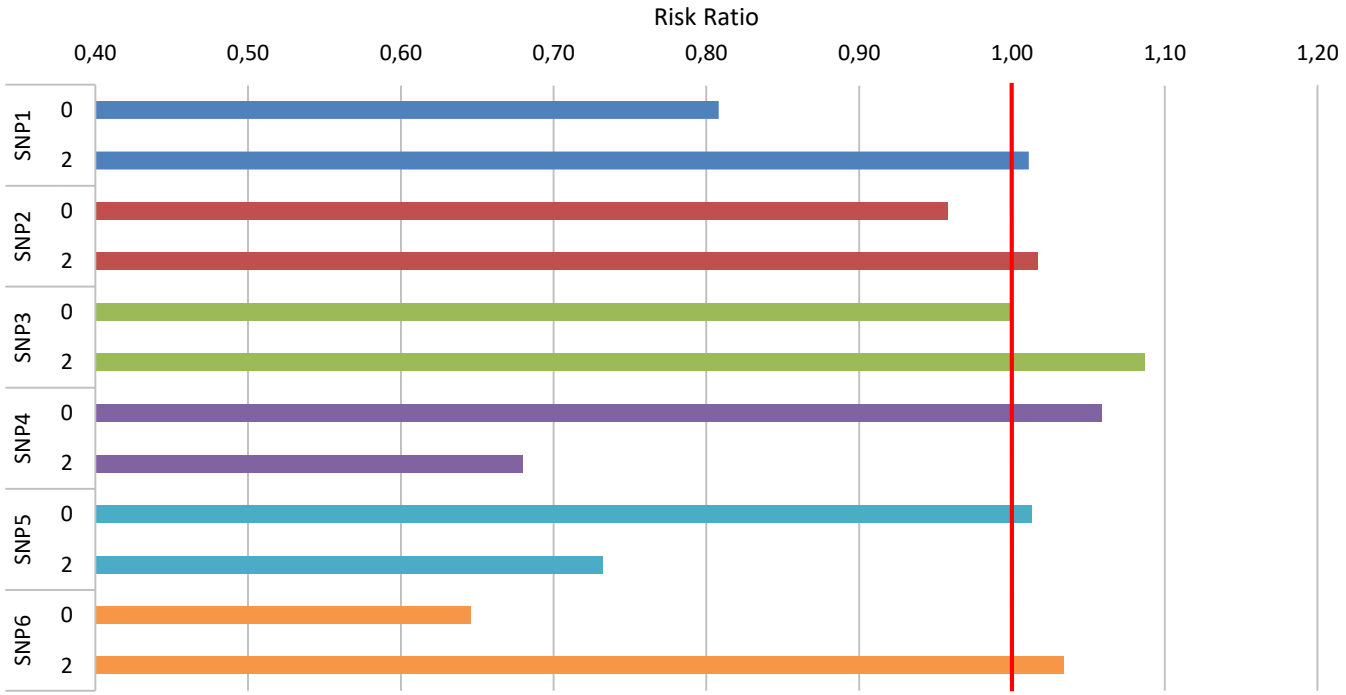
$Y_m(t')$ = time-dependent fixed effect of calving year (*l* = 2008, …,2017);

$S_n(t')$ = time-dependent fixed effect of calving season (*m* = Jan.-Apr., May-Sep., Oct.-Dec.) ;

SNP_o = time-independent fixed effect of one SNP (levels: 0, 1, 2);

g_p = random genetic animal effect.

Risk Ratio of associated SNPs on functional longevity



Abbreviation	Name of SNP	Code of SNP's levels		
		2	0	1
SNP1	BTA-18114-no-rs	AA	GG	AG
SNP2	ARS-BFGL-NGS-38914	AA	GG	AG
SNP3	ARS-BFGL-NGS-19199	AA	GG	AG
SNP4	ARS-BFGL-NGS-25574	AA	GG	AG
SNP5	ARS-BFGL-NGS-113280	AA	GG	AG
SNP6	Hapmap39577-BTA-63767	AA	CC	AC

Conclusion

Impact of health traits on FL:

- udder diseases: Increasing risk for infected cows in all lactation stages, especially in 2nd stages
- claw disorders: Importance of disease from the 2nd stage of 1st lactation
- metabolic diseases: Strong impact of diseases from the early lactation stage
- reproductive diseases: Increasing culling risk with increasing parity

Genetic parameters for FL:

- Average $h^2_{FL} = 0.35$
- Heritabilities for FL were substantially larger when considering health trait information

Genome wide association study:

- 6 SNPs significantly associated with FL (located on chromosomes 1, 4, 10, 13, 28)
- however, SNP genotype was not significant in survival analysis

The project is supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the Federal Program for Ecological Farming and Other Forms of Sustainable Agriculture.

supported by:

Federal Ministry
of Food
and Agriculture

BÖLN

Bundesprogramm Ökologischer Landbau
und andere Formen nachhaltiger
Landwirtschaft

Thank you for your attention

Appendix

Relative risk of involuntary culling for different stages

Diseases	1st lactation			2nd lactation			3rd lactation		
	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage	1. stage	2. stage	3. stage
Mastitis	1.13	1.32	1.30	1.28	1.70	1.37	1.56	2.41	1.74
Subclinical mastitis	0.48	1.69	1.92	1.75	2.08	1.79	2.07	2.50	1.96
Sole ulcer	0.68	1.17	1.24	1.01	1.31	1.32	1.41	1.67	1.53
Digital dermatitis	0.78	1.13	1.22	1.20	1.32	1.27	1.41	1.42	1.42
White line disease	0.38	1.41	1.63	n.s.	1.47	1.57	1.60	1.51	1.62
Interdigital hyperplasia	n.s.	n.s.	n.s.	0.71	1.38	1.50	1.55	1.62	1.60
Ruminal acidosis	2.50			n.s.			2.05		
Ketosis	0.82			1.46			1.75		
Milk fever	1.59			1.42			1.36		
Endometritis	0.80	0.81		1.13	1.19		1.46	1.55	
Ovarial cycle disorder	0.73	0.87		0.67	1.16		1.35	1.65	
Ovarial cysts	0.54	0.88		1.19	1.18		1.54	1.67	
Corpus luteum persistens	0.56	1.34		1.26	1.29		n.s.	1.59	

n.s. – not significant

Dataset, 10 groups

	Gruppen									
	1	2	3	4	5	6	7	8	9	10
Number of cows	13,000	13,000	13,000	13,000	13,000	13,000	13,000	13,000	13,000	12,386
Number of censored cows	3,999	3,940	3,914	3,926	3,908	3,963	3,867	3,898	3,952	3,804
Percent of censored cows	30.76	30.31	30.11	30.20	30.06	30.48	29.75	29.98	30.40	30.71
Average productive life (days)	977.42	977.60	982.48	983.09	980.90	970.39	970.68	978.75	975.01	974.58
Standard deviation	618.61	629.80	632.99	627.25	628.50	625.45	627.77	627.24	629.06	620.67

Formula

The proportion of variance explained by a single SNP-marker was estimated as

$$\frac{\hat{\mathrm{O}}^2_{SNP_{ij}}}{\hat{\mathrm{O}}^2_{g_{ij}}} = \frac{2p_i(1-p_i)\hat{b}^2_{ij}}{\hat{\mathrm{O}}^2_{g_{ij}}}$$

$\hat{\mathrm{O}}^2_{g_{ij}}$= variance in the de-regressed proofs of trait j, explained by all markers;

p_i= allele frequency of the allele, coded as '1' for SNP-marker i;

$\hat{b}^2_{ij}$= estimated marker effect.

source: Heise 2017

False discovery rate corrected p-value was estimated as

$$FDR = \frac{n}{i \times p{-}value}$$

n = Number of SNP;

i = Ranked p value.

Parameter of Weibull model

Disease in model	ρ (shape parameter of the Weibull distribution)	SE
Without diseases	1,50	0,013
Mastitis	4,09	0,036
Subclinical mastitis	4,73	0,039
Sole ulcer	4,52	0,038
Digital dermatitis	4,28	0,036
White line disease	4,72	0,039
Interdigital hyperplasia	4,70	0,039
Ruminal acidosis	4,82	0,040
Ketosis	4,79	0,040
Milk fever	4,79	0,040
Endometritis	4,16	0,034
Ovarial cycle disorder	4,19	0,036
Ovarial cysts	4,62	0,039
Corpus luteum persistens	4,71	0,039

AIC and BIC

$$AIC = -2LogLikelihood + 2k$$

$$AICc = \frac{-2LogLikelihood + 2k + 2k(k+1)}{(n-k-1)}$$

$$\text{BIC} = -2LogLikelihood + k * \ln(n)$$

k = variable in model (16);

n = number of animals (12,938).

	-2 LOG LIKELIHOOD	AIC	AICc	BIC
Mastitis	110,845	221,691	8.62	110,997
Subclinical mastitis	110,088	110,120	8.56	110,239
Sole ulcer	107,732	107,764	8.38	107,884
Digital dermatitis	97,505	97,537	7.59	97,657
White line disease	109,096	109,128	8.49	109,247
Interdigital hyperplasia	107,592	107,624	8.37	107,744
Ruminal acidosis	112,232	112,264	8.73	112,384
Ketosis	112,916	112,948	8.78	113,068
Milk fever	111,214	111,246	8.65	111,366
Endometritis	117,089	117,121	9.11	117,241
Ovarial cycle disorder	95,864	95,896	7.46	96,015
Ovarial cysts	112,245	112,277	8.73	112,396
Corpus luteum persistens	109,011	109,043	8.48	109,162