Genetic evaluation of longevity in dairy cattle

A new model for an old trait

Dissertation

zur Erlangung des Doktorgrades

der Fakultät für Agrarwissenschaften

der Georg-August-Universität Göttingen

vorgelegt von

Johannes Heise

geboren in Bergisch Gladbach

Göttingen, im März 2017

D 7

1. Referentin/Referent:	Prof. Dr. Henner Semianer
2. Korreferentin/Korreferent:	Prof. Dr. Roel F. Veerkamp
Tag der mündlichen Prüfung:	03.05.2017

All models are wrong, but some are useful.

George Edward Pelham Box

Table of contents

Summary

Longevity is an important trait in dairy cows, reflecting the overall functionality of a cow. The aim of this thesis was the development of a new model for the routine genetic evaluation of longevity in German Holsteins. To achieve this purpose, different studies were conducted. They are briefly summarized in the following:

Chapter 1 provides the reader with background information on base principles of survival theory and on the frame conditions of routine genetic evaluations of longevity in German Holsteins. At the end, the necessity for a new routine evaluation system is defined.

With the study in c**hapter 2**, the basis for the new model was developed. It originated from the idea that different periods in the life of a cow belong to different challenges which are related to different functional traits. Binary survival codes were defined for different periods across the first three lactations and modeled as genetically correlated traits. In order to estimate genetic parameters, an excessive estimation of variance components was conducted on data of 1,495,441 cows with two models. With the first model, 18 finely graded periods were considered. Results from this model showed a clear pattern for the genetic background of survival across the first three lactations. Periods with similar genetic background for survival were then merged in the second model, where only nine traits, three for each lactation, were considered. Afterwards, a genetic evaluation was run on data of 7,684,455 cows and estimated breeding values (**EBVs**) for sires were compared to routine EBVs for other traits. This comparison further justified the approach, showing plausible correlation patterns. In the prototype version of the new routine genetic evaluation of longevity, almost the same periods are used to define different survival traits. These are: survival from calving to 49 d, 50 d to 249 d and from 250 d to the consequent calving. These periods were defined for each of the first three lactations.

In **chapter 3** it was questioned if models for routine evaluations of longevity should include an effect of age at first calving. The idea for this study arose from the definition of age at first calving (**AFC**) which is the sum of age at first insemination (**AFI**), the interval from first to last (successful) insemination (**FLI**) and gestation length. These traits are all functional traits. In order to investigate if these traits are genetically correlated to survival, variance and covariance components were estimated between AFC, AFI, FLI and survival of different periods of the first lactation as defined in the paragraph above. Data of 721,919 German Holstein cows were analyzed. Estimated genetic correlations of AFC and FLI to survival traits late in lactation were different from zero. As a conclusion, the correction for age at first calving in models for genetic evaluations of longevity should be reconsidered, because it might remove functional genetic variance.

Chapter 4 gives a description of the prototype version of the new genetic evaluation system for longevity. The development of this prototype version was part of the project and its results are the basis for chapter 5.

The basic idea for **chapter 5** was that the differentiated genetic background of survival of different periods should also express in differentiated genome-wide associations. Therefore, this genome-wide association study (**GWAS**) was performed on deregressed EBVs of 4,849 bulls for the nine survival traits using high-density SNP-marker genotypes. Three different analyzes were performed: (1) a single-marker GWAS (2) a gene-based GWAS and (3) a gene-based mixed model, where gene regions with significant associations identified from (2) were modeled as random. Eight regions on chromosomes 5, 6, 7, 14 and 18 showed significant associations to at least one of the survival traits. Different patterns were observed for the strengths of association among the survival traits. These were in most cases plausible when compared to results from other studies. The study in chapter 5 justifies the results of chapter 2 from a genomic point of view and lays the foundation for further research on this topic. Results from this study may also be valuable when designing models for genomic evaluations of longevity in dairy cows.

In **chapter 6**, important topics that were not covered by chapters 2 to 5 were further highlighted and discussed in detail. It gives insights into different methods for the construction of an index EBV from nine survival traits. Potential for further research from observations during the study in chapter 5 is illustrated. At the end, a short prospect of the future is given for the longevity trait, the overall indicator of functionality.

CHAPTER 1:

General Introduction

General Introduction

The aim of this thesis was the development of a new model for the routine genetic evaluation of longevity in German Holstein cows. This chapter introduces basic definitions and methodological concepts. It further gives an overview of historic developments of genetic evaluations of longevity. The last section highlights problems with the current genetic routine evaluation of longevity in German Holsteins and explains the motivation to develop a new model.

Definition of longevity in dairy cows and its relevance

Longevity of dairy cows can be measured in different ways. Throughout this thesis, longevity is referred to as the time from first calving to culling, i.e. productive life. It is an important trait in dairy cows for three reasons:

1) Longevity has great impact on the **profitability** of dairy farms: During her rearing period, a cow induces costs. The longer a cow lives, the more milk she gives and the lesser are the rearing costs per kg milk which was produced by her. Further, a cow reaches her maximum milk yield per lactation in third to fifth lactation (Ray et al., 1992; vit, 2017). These two main factors are responsible for the impact of longevity on dairy profitability which various studies have proven (Allaire and Gibson, 1992; VanRaden and Wiggans, 1995; Wolfová et al., 2007) and which is reflected by relevant economic weights for longevity in total merit indexes of many major Holstein breeding countries (Miglior et al., 2005; Interbull, 2016a).

2) The same mechanism as for her rearing costs is valid also for her **environmental footprint:** during her rearing period, a cow emits environmentally detrimental substances such as, e.g., methane. The longer a cow lives, the more milk she gives and the lesser are the emissions from her rearing period per kg milk which was produced by her (as reviewed by Knapp et al., 2014).

3) Animal husbandry is in the focus of **public discussions.** A major criticism towards the dairy industry is the relatively short productive period of dairy cows (Busse, 2015). This is probably mainly due to the fact that underlying variables of longevity such as mortality rate are often considered as indicator traits for animal welfare (Winckler et al., 2003; Dechow et al., 2011; de Vries et al., 2011).

These points illustrate the necessity to improve longevity of dairy cows with all means available. This thesis illuminates important genetic aspects in this context.

The concept of time to event data and censoring

Culling is the **unique event** at the end of a cow's productive life and is used synonymously with the term disposal throughout this thesis. It can occur only once and there are no competing events which could prevent a cow from being culled one day. From this definition it is obvious that longevity in dairy cows is a time to event trait: it starts with the first calving and ends with culling.

It may occur that only partial information about a cow's longevity is known. The longevity record of this cow is then considered to be **censored** (Klein and Moeschberger, 2003)**.** Two main reasons could lead to this situation:

1) The cow was sold for further use as a dairy cow, but there is no follow-up record from the herd of destination. This occurs all over the time during the observation period, which is the time span covered by the data set.

2) The cow is still alive at the date of data cutoff for a particular genetic evaluation. This affects the group of youngest cows in a data set more than the group of older cows.

In both cases, we do not know the cow's length of productive life, but its minimum value. Both cases are called right-censored (Klein and Moeschberger, 2003). Other kinds of censoring are not relevant in our context, because they were removed from all analyses throughout this thesis. Inclusion of partial information into analyses of longevity is desired because of two reasons:

1) Genetic gain per time unit is dependent on the length of the generation interval (Falconer and Mackay, 1996). Breeders therefore are interested in selecting bulls as early as possible. If censored records were ignored, all cows would have to be given an adequate opportunity to get old in order to avoid estimation bias (see reason 2). Accurately estimated breeding values (**EBVs**) would be available too late for the decision whether to select a bull as a sire or not.

2) Excluding partial information would lead to biased estimates of longevity for first calving cohorts where a substantial proportion of animals is still alive (censored). An example of two cows, which calved for the first time at the same date, may illustrate this: the one cow was culled early and her information is included in the analyses. The other cow is still alive and her record is ignored. Her productive life is much longer than that of the first cow. This means, with a decreasing time interval from first calving to the date of data cutoff, longevity would be increasingly underestimated for first calving cohorts of animals.

Survival analysis is a field of statistics, which provides us with methods to make use of complete and censored observations at the same time.

Risk of culling and survival

Time to event can be regarded as successive survival of arbitrarily short time intervals. Following Kaplan and Meier (1958), the risk of a cow to be culled during interval t_m, conditional on surviving all intervals from t_0 to t_{m-1}, is

$$R_{t_m} = \frac{n_{culled\ during\ time\ interval\ t_m}}{n_{at\ risk\ at\ time\ interval\ t_m}}$$

where $n_{culled\ during\ time\ interval\ t_m}$ is the number of cows being culled at time interval t_m, and $n_{at\ risk\ at\ time\ interval\ t_m}$ are all cows which survived interval t_{m-1}. Cows being censored during interval t_{m-1} are not considered to be at risk at time interval t_m. In the following, the probabilities R_{t_m} and $1 - R_{t_m}$ are referred to as probabilities on the **risk-level.**

The probability of a cow at the time of her first calving (t_0) to survive until a certain time interval t_m is then

$$S_{t_m} = \prod_{j=1}^{m}(1 - R_{t_j})$$

which at the same time gives the estimate of the proportion of cows which survives from t_0 until t_m.

Life expectancy

Life expectancy is the expected longevity for a cow at the time of her first calving. Following Klein and Moeschberger (2003), the life expectancy ($L_{expected}$) is the area under the survival curve:

$$L_{expected} = \sum_{m=1}^{n} \Delta_m S_{t_m}$$

where n is the number of time intervals and Δ_m is the length of the m^{th} interval. If the survival curve is assumed to be stepwise linear, this expression becomes

$$L_{expected} = \sum_{m=1}^{n} \Delta_m \left(\frac{S_{t_m} + S_{t_{m-1}}}{2}\right)$$

with $S_{t_0} = 1$.

Measurements of longevity, effects on longevity and statistical methods applied in the context of dairy breeding

The following section gives a brief overview of different measures of longevity and of the most important statistical methods in the context of animal breeding. Deliberately, descriptions of statistical properties are only sketched roughly where thought to be necessary, because most of these methods were not applied to longevity in this thesis. The properties of linear mixed models, which were used in chapters 2, 3, 4 and 5, are briefly described in the respective chapters and can be read in detail and well presented in the book of Mrode (2014).

Measurements. The time from first calving to culling in dairy cows is commonly measured in two different units:

1) In physical units, which are usually days, months or years (e.g., Ducrocq, 2005; van Pelt et al., 2015).
2) Along the lactation cycle of a cow. This is usually presented as the number of lactations or parts of lactations, e.g., days in milk (e.g., Boettcher et al., 1999; Sasaki et al., 2015).

Although some early studies on the genetics of longevity used complete information only (Wilcox et al., 1957; Parker et al., 1960; Hargrove et al., 1969), most other studies dealt in either way with the phenomenon of censoring. The most common way is to define periods of fixed lengths, whatever unit is used, and then define a binary variable of survival observations for each period. This is the basic idea of the method of Kaplan and Meier (1958) which was described in the previous section. If multiple consequent periods are considered, modeling is undertaken on the risk-level (e.g., Boettcher et al., 1999; Sewalem et al., 2007; Holtsmark et al., 2009). This definition underlies, e.g., the national genetic routine evaluation of longevity in Canada (Sewalem et al., 2007; Interbull, 2016a). Periods could also be defined to be of different lengths, but all starting from the same time point as in the study of Sasaki et al. (2015) and implemented in the common routine genetic evaluation system for Holsteins in Denmark, Sweden and Finland (Interbull, 2016a). The survival curve is then modeled directly. Another method to deal with censored data was proposed by VanRaden and Klaaskate (1993): for censored records, remaining productive life can be estimated in a pre-processing step, using environmental effects and information on the cow's lactation status for prediction. These predicted records can then be used directly in a genetic evaluation, but with reduced weights compared to complete observations.

Effects of other traits. Two cow-related traits are frequently modeled as non-genetic effects in genetic evaluations of longevity in dairy cows:

1) Age at first calving (e.g., Buenger et al., 2001; Sewalem et al., 2007)
2) Milk, protein or fat yield, relative to the herd mean yield (e.g., Buenger et al., 2001; Ducrocq, 2005; Sewalem et al., 2007)

The question if age at first calving should be treated as a fixed effect in models for genetic evaluations of longevity is discussed in chapter 3.

The other important correction is the one for milk, protein and/or fat yield, relative to the herd mean: it was often argued that culling occurs for voluntary and involuntary reasons (e.g., Rogers et al., 1988; Weigel et al., 2003): voluntary culling is usually referred to as culling for milk yield while involuntary culling is related to functional problems. Voluntary culling is regarded to be favorable for farmers as a tool to improve the mean milk yield of their herds and involuntary culling is regarded to be the opposite. In this argumentation, only involuntary culling should be reduced to give more opportunity for voluntary culling. Further, the farmer is assumed to compare a cow's milk yield to the one of her herd mates when culling her voluntarily. In most national routine genetic evaluations, including Germany, longevity is therefore corrected for some measure of milk yield, relative to the herd mean, in order to remove the effect of voluntary culling (e.g., Ducrocq, 2005; Sewalem et al., 2007; Interbull, 2016a; vit, 2016). The resulting trait is then called *functional longevity*. For a detailed review on this topic, see also Essl (1998).

A lot of other traits, especially health traits, have a substantial impact on longevity (e.g., Rajala-Schultz and Gröhn, 1999a; b; c). These traits are discussed in more detail in chapter 2.

Summarizing considerations of this section, longevity can be regarded as an indicator trait for overall functionality.

Environmental effects. In routine genetic evaluations it is desired to correct for environmental effects. They are considered to be non-genetic but potentially confounded with genetic effects which could lead to misleading results if they were not accounted for. Most non-genetic effects are modeled as fixed effects, but some, especially herd effects, are treated as random (e.g., Ducrocq, 2005; van Pelt et al., 2015). The following effects were frequently used as covariates in genetic evaluations of longevity, assuming they have no genetic effects correlated to longevity or functional longevity, dependent on the trait definition:

1) Herd effects (e.g., Pasman and Reinhardt, 1999; Ducrocq, 2005; Sewalem et al., 2007; Sasaki et al., 2015)
2) Year effects (e.g., Pasman and Reinhardt, 1999; Sewalem et al., 2007; Sasaki et al., 2015)
3) Seasonal effects (e.g., Pasman and Reinhardt, 1999; Buenger et al., 2001)
4) Region (e.g., Ducrocq, 2005; Sasaki et al., 2015)
5) Herd size change (e.g., Pasman and Reinhardt, 1999; Sewalem et al., 2007)

Usually, herd, year and when applicable, season are considered as a herd × year × season effect (e.g., Buenger et al., 2001; Ducrocq, 2005). Region is also often considered as interaction effect region × year (e.g., Ducrocq, 2005).

Methods. Statistical models applied to genetic analyses of longevity in scientific and routine genetic evaluation context are as manifold as the measurements are. This paragraph gives a brief overview of the most important methods. They can be divided into three groups:

1) *Linear mixed models.* Since the work of Henderson (1973, 1975), linear mixed models, which treat genetic effects as random, are well established in routine genetic evaluations of all important dairy cattle traits (e.g., Interbull, 2016a). Software, even for large amounts of data, is readily available (e.g., Misztal et al., 2002; Groeneveld, 2006) and computational demands are low enough to allow for complex animal models with many correlated random effects in large-scale routine genetic evaluations (Interbull, 2016a). A special case of linear models are linear random regression models. For the analysis of longevity, their relationship to survival analysis methods (see below) was worked out in detail by Veerkamp et al. (2001). Linear mixed models were frequently used on binary survival data (e.g., Visscher and Goddard, 1995; Boettcher et al., 1999; Holtsmark et al., 2009) and are used for routine genetic evaluations of longevity in many countries (Gengler et al., 2005; VanRaden et al., 2006; Sewalem et al., 2007; Interbull, 2016a).

2) *Threshold models.* Linear mixed models assume residuals to be normally distributed (Henderson, 1973). Because this is not the case for binary response variables in linear mixed models, threshold models were suggested for the use on this kind of data in the context of animal breeding (e.g., Gianola, 1980) and were also applied to survival data, considered as survival (1/0) of consequent periods (e.g., Boettcher et al., 1999; González-Recio and Alenda, 2007).

3) *Survival analysis models.* From the above sketched methods, only linear random regression models (e.g., Veerkamp et al., 2001; van Pelt et al., 2015; Sasaki et al., 2015) can feasibly handle larger numbers of periods, but still are theoretically inadequate for use with binary data (Gianola, 1980). For the other methods, usually few and thus relatively long periods are defined (e.g., Boettcher et al., 1999; Holtsmark et al., 2009). As pointed out by Ducrocq et al. (1988) this leads, together with the discretization (1/0), to a loss of information. The authors therefore suggested a proportional hazards survival model with a Weibull parameterization of the hazard function, which is the continuous equivalent to the discrete risk-level probabilities described in the previous section. Length of productive life could then be modeled continuously. Proportional hazards survival models are currently used for routine genetic evaluations of longevity in several countries, including the Netherlands (Vollema et al., 2000), France (Ducrocq, 2005) and Germany (Pasman and Reinhardt, 1999).

The choice of the statistical method mainly relies on four basic considerations:

1) *The response variable(s)*, i.e., the representation of longevity.
2) *The explanatory variables* to be included in the model.
3) *Theoretical considerations* with regard to (1) and (2).
4) *Practical considerations.* Practical considerations can be performance considerations. For example, Boettcher et al. (1999) compared all three kinds of models for the use in genetic evaluations and stated that the threshold and survival analysis models in their study took about five to ten times the computational time compared to the linear models on mostly the same data. Because routine genetic evaluations use large amounts of data and runtimes are often in the range of days, this alone can make the difference with regard to feasibility in routine systems, which often have a severe time limitation between the data cutoff date and the mandatory publication date. Other important considerations include the availability of adequate software, the reusability of models and software which were originally developed for production traits (Veerkamp et al., 2001) and communication strategies towards dairy breeders as suggested by VanRaden and Klaaskate (1993): 'Rapid acceptance by the dairy industry might be expected if the statistical techniques currently used for yield traits work as well for longevity'.

All these points interrelate with each other. For example, if a model is desired to include genetically distinct but correlated animal effects for different periods, survival models are 'computationally impossible' for large-scale routine genetic evaluations (Ducrocq, 2005).

Frame conditions of routine genetic evaluations of longevity in German Holsteins

This section describes the frame conditions of routine genetic evaluations of longevity in German Holstein cows. This section depicts the organizational structure of involved organizations, the historic development of national genetic evaluations of longevity, its success and its current limitations.

National genetic evaluation center, breeds and sources of data

vit (IT solutions for Animal Production) is the national genetic evaluation center for the dairy breeds Holstein, Angler/Red Dairy Cattle, Jersey and Black-and-White Friesian Cattle. The organization is assigned by the German Holstein breeding organizations to conduct national routine genetic evaluations for these breeds in Germany. At the same time, vit estimates breeding values for the above mentioned breeds in Austria and Luxemburg. Data for routine evaluations are supplied by milk recording organizations of the different federal states of Germany and Luxemburg and by the Association of Austrian Cattle Breeders (ZAR). A detailed description is given in vit (2016).

Beside milk yield data, milk recording organizations collect data on newborn calves and disposed cows, including the date and reason of disposal from the herd. Reasons for disposal are coded as specified by the ADR (2006). Furthermore, milk recording organizations have the permission of most participating farmers to retrieve data about their animals from the national animal movement data base (HI-Tier, 2016), which is a reliable source of birth and disposal dates.

Historic development

In 1996, a routine genetic evaluation of functional longevity in German Holsteins was implemented, basing on a proportional hazards model (Pasman and Reinhardt, 1999). EBVs are published as **RZN** on a relative scale with mean 100 for the cow base population and a genetic standard deviation of 12 (vit, 2016). Since then, no major changes have been applied to the core evaluation system, but the weight of this trait in the total merit index (RZG) changed over time: in 2002 from 6% to 25% (Rensing et al., 2002) and to 20% in 2008 (DHV, 2008). Furthermore, type traits were analyzed for their usability as indicator traits for functional longevity (Buenger et al., 2001) and introduced in 2001 (Rensing et al., 2002) to improve reliabilities of early EBVs of sires.

The implementation of a routine genetic evaluation of functional longevity has led to considerable genetic gain for this trait as seen from Figure 1.1. Presented are mean EBVs of sires by year, weighted by the number of inseminations with their semen. After the weight of functional longevity in the RZG was increased to 25% in 2002, the genetic trend became clearly positive. Comparing periods from 2002 to 2009 and from 2010 to 2015, the slope of the genetic trend

more than doubled for the later period, which was after the introduction of the genomic evaluation system (Reinhardt et al., 2009; vit, 2016).

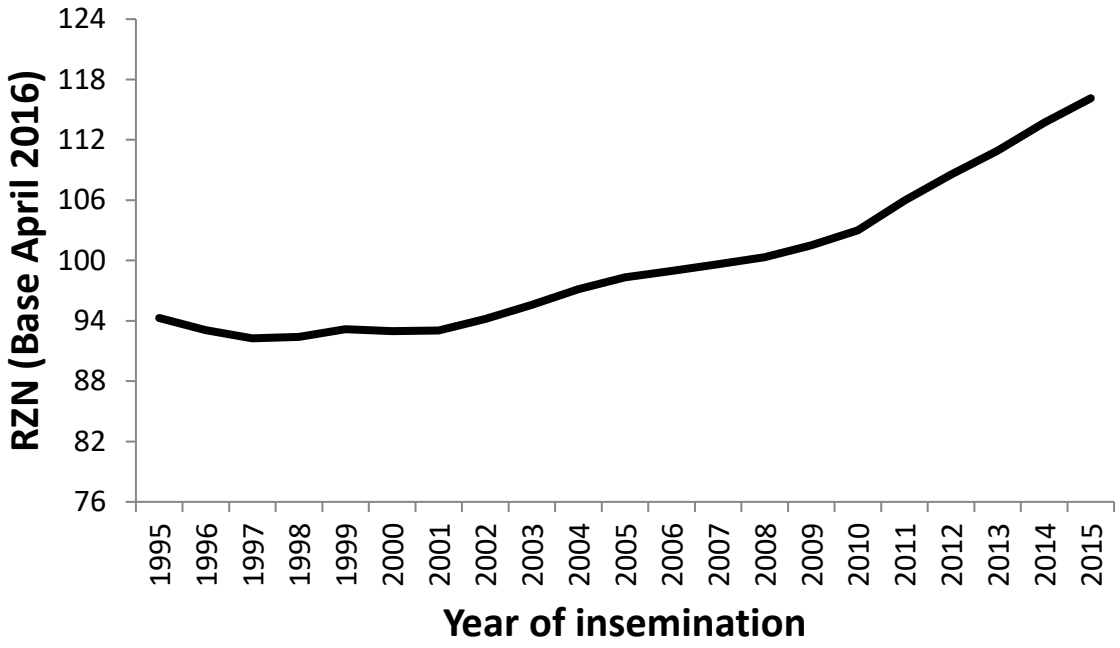


Figure 1.1: Mean EBVs of Holstein sires for functional longevity (RZN) by year, weighted by the number of inseminations with their semen. Source of data: vit, personal communication.

Participation in international routine genetic evaluations via Interbull

The Cattle Breeders' Federation (**ADR**), which is the umbrella organization of German cattle breeding organizations, is an Interbull service user (Interbull, 2017a). Interbull uses data from national routine evaluation systems for nationally published bulls to conduct international routine genetic (Multiple-trait Across Country Evaluation, MACE) and genomic (International Genomic Evaluation of Young Bulls, GMACE) evaluations. As result of this, a bull has EBVs on the scales of all participating countries and can thus be directly compared to other bulls on the different scales. This supports international trading with semen and breeding cattle and can help to improve the reliabilities of EBVs by using information from other populations (e.g., Druet et al., 1999).

Before a service user is allowed to participate in international routine evaluations with data from its national evaluation system for a specific trait, Interbull requires the participation in a test run (Interbull, 2017b). In this test run, genetic correlations to other countries are estimated as well as the sire standard deviation from the submitted sample, which represents the genetic standard deviation. The participation in a test run is required whenever there are major changes in the national evaluation system for the respective trait. Interbull further requires the validation of the national evaluation system with three different validation methods (Boichard et al., 1995; Inter-

bull, 2016b) which should all be passed if applicable. The validation procedure has to be repeated every two years and whenever the national routine evaluation system changes. Only trend validation method III could be applied to the current genetic evaluation of longevity. Results from this method are therefore the first benchmark to use for the newly developed model for this trait and this method only is described here in brief: two complete genetic evaluations are run, the one making use of the full data used for a current genetic evaluation and the other one with a data cutoff date four years earlier. EBVs of bulls from the run with full data are then modeled as a function of the EBVs from the run on the truncated data set, using weighted regression and including a term to estimate bias conditional on the additional information between the two runs. If EBVs are BLUP, there should be no trend in them, dependening on additional information between the two runs, and the estimate for the bias term should therefore be zero. Interbull accepts a maximum absolute estimate for the coefficient of this bias term of 0.02 genetic standard deviations. Detailed information on test runs and validation methods can be obtained from Interbull's Code of Practice, available on their website (Interbull, 2016b). The current routine genetic evaluation system for longevity overestimates EBVs of young bulls substantially and would probably fail this trend validation test by orders of magnitude, if it were applied.

Motivation to develop a completely new routine genetic evaluation system

Biased estimation of breeding values and probably failing the Interbull trend validation was not the only reason to consider the development of a completely new routine genetic evaluation system for longevity instead of optimizing the current one. Other reasons included the practical impossibility to use a survival analysis animal model (Ducrocq, 2005), long runtimes on already limited data and the fact that the software (Survival Kit: e.g., Mészáros et al., 2013) was no in-house development and thus practically a black box. The implementation of an animal model will be necessary in order to estimate breeding values for cows as a basis for genomic prediction from a growing cow reference population (Reents et al., 2016). From these points, requirements for a new routine genetic evaluation system for Holstein cattle in Germany can be formulated as:

1) Predictors should be best and unbiased
2) The new model should be an animal model
3) The new system must be computationally feasible
4) Software must be easy to maintain and adapt

To achieve this, a project was launched to develop a new model for the routine genetic evaluation of longevity. The following chapters show results from this project and have the following purposes:

With **chapter 2**, we studied the genetic background of survival of different periods in the life of a cow using different multiple trait sire models. This chapter includes discussion about disposal reasons, the estimation of variance components and the analysis of correlations of resulting EBVs to other traits.

In **chapter 3**, we questioned the inclusion of a fixed effect for age at first calving into models for genetic evaluations of functional longevity.

Chapter 4 provides a description of the prototype version of the new genetic evaluation system for functional longevity in German Holsteins.

In **chapter 5**, we used resulting EBVs from this prototype version and high-density genotype SNP-data to analyze associations of different genomic regions to different survival traits. We did this for three reasons: (1) validation of the results from the previous chapters, (2) as a preliminary study for a future genomic evaluation of longevity, and (3) to gain further knowledge about possible functional relationships between candidate regions and longevity.

References

ADR, (German Cattle Breeders' Federation). 2006. ADR-Empfehlung 3.1: Leistungsprüfung für funktionale Merkmale bei Bullen und Kühen (Gesundheit, Reproduktion, Nutzungsdauer, Exterieur, Melkbarkeit). Accessed March 23, 2017. http://www.adr-web.de/services/files/richtlinien-empfehlungen/20170516%20ADR-Empfehlung%203.1.pdf.

Allaire, F.R., and J.P. Gibson. 1992. Genetic Value of Herd Life Adjusted for Milk Production. J. Dairy Sci. 75:1349–1356. doi:10.3168/jds.S0022-0302(92)77886-2.

Boettcher, P.J., L.K. Jairath, and J.C.M. Dekkers. 1999. Comparison of Methods for Genetic Evaluation of Sires for Survival of Their Daughters in the First Three Lactations. J. Dairy Sci. 82:1034–1044. doi:10.3168/jds.S0022-0302(99)75324-5.

Boichard, D., B. Bonaiti, A. Barbat, and S. Mattalia. 1995. Three Methods to Validate the Estimation of Genetic Trend for Dairy Cattle. J. Dairy Sci. 78:431–437. doi:10.3168/jds.S0022-0302(95)76652-8.

Buenger, A., V. Ducrocq, and H.H. Swalve. 2001. Analysis of Survival in Dairy Cows with Supplementary Data on Type Scores and Housing Systems from a Region of Northwest Germany. J. Dairy Sci. 84:1531–1541. doi:10.3168/jds.S0022-0302(01)70187-7.

Busse, T. 2015. Die Wegwerfkuh: Wie unsere Landwirtschaft Tiere verheizt, Bauern ruiniert, Ressourcen verschwendet und was wir dagegen tun können. Karl Blessing Verlag, München.

Dechow, C., E. Smith, and R. Goodling. 2011. The effect of management system on mortality and other welfare indicators in Pennsylvania dairy herds. Anim. Welf. 20:145–158.

DHV, (German Holstein Association). 2008. Neue Gewichtung Der Zuchtwerte Bei Deutschen Holsteins. Accessed March 25, 2017. http://www.holstein-dhv.de/aktuelles/neue-gewichtung-der-zuchtwerte-bei-deutschen-holst.html.

Druet, T., J. Sölkner, A.F. Groen, and N. Gengler. 1999. Improved genetic evaluation of survival using MACE to combine direct and correlated information from yield and functional traits. Interbull Bull. 21:122-127.

Ducrocq, V. 2005. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. Anim. Sci. 80:249–256. doi:10.1079/ASC41720249.

Ducrocq, V., R.L. Quaas, E.J. Pollak, and G. Casella. 1988. Length of Productive Life of Dairy Cows. 1. Justification of a Weibull Model. J. Dairy Sci. 71:3061–3070. doi:10.3168/jds.S0022-0302(88)79906-3.

Essl, A. 1998. Longevity in dairy cattle breeding: a review. Livest. Prod. Sci. 57:79–89. doi:10.1016/S0301-6226(98)00160-2.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to Quantitative Genetics. Fourth Edition. Pearson, London.

Fischer, J., and J. Ullrich. 2016. The new system of units. Nat. Phys. 12:4–7. doi:10.1038/nphys3612.

Gengler, N., S. Vanderick, P. Mayeres, A. Gillon, and C. Croquet. 2005. Genetic evaluation of cow survival using a lactation random regression model. Interbull Bull. No 33:176–180.

Gianola, D. 1980. A Method of Sire Evaluation for Dichotomies. J. Anim. Sci. 51:1266–1271. doi:10.2527/jas1981.5161266x.

González-Recio, O., and R. Alenda. 2007. Genetic relationship of discrete-time survival with fertility and production in dairy cattle using bivariate models. Genet. Sel. Evol. 39:391. doi:10.1186/1297-9686-39-4-391.

Groeneveld, E. 2006. PEST User's Manual. Institute of Animal Science, D-31535 Neustadt, Germany.

Hargrove, G.L., J.J. Salazar, and J.E. Legates. 1969. Relationships among First-Lactation and Lifetime Measurements in a Dairy Population. J. Dairy Sci. 52:651–656. doi:10.3168/jds.S0022-0302(69)86623-3.

Henderson, C.R. 1973. SIRE EVALUATION AND GENETIC TRENDS. J. Anim. Sci. 1973:10–41. doi:10.2527/1973.1973Symposium10x.

Henderson, C.R. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics 31:423–447. doi:10.2307/2529430.

HI-Tier. 2016. Rinder-Datenbank; Allgemeine Hinweise. Accessed February 1, 2016. https://www1.hi-tier.de/info03.html.

Holtsmark, M., B. Heringstad, and J. Ødegård. 2009. Predictive abilities of different statistical models for analysis of survival data in dairy cattle. J. Dairy Sci. 92:5730–5738. doi:10.3168/jds.2009-2132.

Interbull. 2016a. National Genetic Evaluation Forms Provided by Countries. Accessed December 13, 2016. http://interbull.org/ib/geforms.

Interbull. 2016b. 6. Interbull Code of Practice - Traits and Breeds. Accessed March 21, 2017. http://www.interbull.org/ib/cop_chap6.

Interbull. 2017a. Interbull Service Users. Accessed March 24, 2017. http://www.interbull.org/ib/users_map.

Interbull. 2017b. Becoming Part of Interbull. Accessed March 25, 2017. http://www.interbull.org/ib/becomingpartofitb.

Kaplan, E.L., and P. Meier. 1958. Nonparametric Estimation from Incomplete Observations. J. Am. Stat. Assoc. 53:457–481. doi:10.1080/01621459.1958.10501452.

Klein, J.P., and M.L. Moeschberger. 2003. Survival Analysis - Techniques for Censored and Truncated Data. Second Edition. Springer, New York.

Knapp, J.R., G.L. Laur, P.A. Vadas, W.P. Weiss, and J.M. Tricarico. 2014. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. J. Dairy Sci. 97:3231–3261. doi:10.3168/jds.2013-7234.

Mészáros, G., J. Sölkner, and V. Ducrocq. 2013. The Survival Kit: Software to analyze survival data including possibly correlated random effects. Comput. Methods Programs Biomed. 110:503–510. doi:10.1016/j.cmpb.2013.01.010.

Miglior, F., B.L. Muir, and B.J. Van Doormaal. 2005. Selection Indices in Holstein Cattle of Various Countries. J. Dairy Sci. 88:1255–1263. doi:10.3168/jds.S0022-0302(05)72792-2.

Misztal, I., S. Tsurata, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. BLUPF90 and related programs (BGF90). Page in Proceedings of the 7th World Congress on Genetics Applied to Livestock Production World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Mrode, R.A. 2014. Linear Models for the Prediction of Animal Breeding Values. 3 Rev ed. Cabi Publishing, Boston, MA.

Parker, J.B., N.D. Bayley, M.H. Fohrman, and R.D. Plowman. 1960. Factors Influencing Dairy Cattle Longevity. J. Dairy Sci. 43:401–409. doi:10.3168/jds.S0022-0302(60)90175-2.

Pasman, E., and F. Reinhardt. 1999. Genetic Evaluations for Length of Productive Life of Holstein Cattle in Germany. Interbull Bull. No 21:56–59.

van Pelt, M.L., T.H.E. Meuwissen, G. de Jong, and R.F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130. doi:10.3168/jds.2014-9090.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999a. Culling of dairy cows. Part I. Effects of diseases on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:195–208. doi:10.1016/S0167-5877(99)00046-X.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999b. Culling of dairy cows. Part III. Effects of diseases, pregnancy status and milk yield on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:295–309. doi:10.1016/S0167-5877(99)00047-1.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999c. Culling of dairy cows. Part II. Effects of diseases and reproductive performance on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:279–294. doi:10.1016/S0167-5877(99)00045-8.

Ray, D.E., T.J. Halbach, and D.V. Armstrong. 1992. Season and Lactation Number Effects on Milk Production and Reproduction of Dairy Cattle in Arizona1. J. Dairy Sci. 75:2976–2983. doi:10.3168/jds.S0022-0302(92)78061-8.

Reents, R., Z. Liu, D. Segelke, K.F. Stock, E. Pasman, S. Rensing, and F. Reinhardt. 2016. Setting up a female genomic reference population for German Holsteins. Page TS1.4 in Abstracts and Presentations 40th ICAR Biennial Session, Puerto Varas, Chile.

Reinhardt, F., Z. Liu, F. Seefried, and G. Thaller. 2009. Implementation of genomic evaluation in German Holsteins. Interbull Bull. 40:219-226.

Rensing, S., E. Pasman, F. Reinhardt, and F. Feddersen. 2002. New total merit index RZG for Holsteins in Germany with more emphasis on herd life. Interbull Bull. 29:147-149.

Rogers, G.W., J.A.M.V. Arendonk, and B.T. McDaniel. 1988. Influence of Involuntary Culling on Optimum Culling Rates and Annualized Net Revenue. J. Dairy Sci. 71:3463–3469. doi:10.3168/jds.S0022-0302(88)79952-X.

Sasaki, O., M. Aihara, A. Nishiura, H. Takeda, and M. Satoh. 2015. Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. J. Dairy Sci. 98:5781–5795. doi:10.3168/jds.2014-9152.

Sewalem, A., F. Miglior, G.J. Kistemaker, P. Sullivan, G. Huapaya, and B.J. Van Doormaal. 2007. Short Communication: Modification of Genetic Evaluation of Herd Life from a Three-Trait to a Five-Trait Model in Canadian Dairy Cattle. J. Dairy Sci. 90:2025–2028. doi:10.3168/jds.2006-719.

VanRaden, P.M., C.M.B. Dematawewa, R.E. Pearson, and M.E. Tooker. 2006. Productive Life Including All Lactations and Longer Lactations with Diminishing Credits. J. Dairy Sci. 89:3213–3220. doi:10.3168/jds.S0022-0302(06)72596-6.

VanRaden, P.M., and E.J.H. Klaaskate. 1993. Genetic Evaluation of Length of Productive Life Including Predicted Longevity of Live Cows. J. Dairy Sci. 76:2758–2764. doi:10.3168/jds.S0022-0302(93)77613-4.

VanRaden, P.M., and G.R. Wiggans. 1995. Productive Life Evaluations: Calculation, Accuracy, and Economic Value. J. Dairy Sci. 78:631–638. doi:10.3168/jds.S0022-0302(95)76674-7.

Veerkamp, R.F., S. Brotherstone, B. Engel, and T.H.E. Meuwissen. 2001. Analysis of censored survival data using random regression models. Anim. Sci. 72:1–10.

Visscher, P.M., and M.E. Goddard. 1995. Genetic Parameters for Milk Yield, Survival, Workability, and Type Traits for Australian Dairy Cattle. J. Dairy Sci. 78:205–220. doi:10.3168/jds.S0022-0302(95)76630-9.

vit. 2016. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed December 13, 2016. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf.

vit. 2017. Trends Fakten Zahlen 2016. Accessed March 24, 2017. http://www.vit.de/fileadmin/user_upload/wirsindvit/jahresberichte/vit-JB2016-gesamt.pdf.

Vollema, A.R., S.V.D. Beek, A.G.F. Harbers, and G.D. Jong. 2000. Genetic Evaluation for Longevity of Dutch Dairy Bulls. J. Dairy Sci. 83:2629–2639. doi:10.3168/jds.S0022-0302(00)75156-3.

de Vries, M., E. a. M. Bokkers, T. Dijkstra, G. van Schaik, and I.J.M. de Boer. 2011. Invited review: Associations between variables of routine herd data and dairy cattle welfare indicators. J. Dairy Sci. 94:3213–3228. doi:10.3168/jds.2011-4169.

Weigel, K.A., R.W. Palmer, and D.Z. Caraviello. 2003. Investigation of Factors Affecting Voluntary and Involuntary Culling in Expanding Dairy Herds in Wisconsin using Survival Analysis. J. Dairy Sci. 86:1482–1486. doi:10.3168/jds.S0022-0302(03)73733-3.

Wilcox, C.J., K.O. Pfau, and J.W. Bartlett. 1957. An Investigation of the Inheritance of Female Reproductive Performance and Longevity, and their Interrelationships within a Holstein-Friesian Herd1,2. J. Dairy Sci. 40:942–947. doi:10.3168/jds.S0022-0302(57)94578-2.

Winckler, C., J. Capdeville, G. Gebresenbet, B. Hörning, U. Roiha, M. Tosi, and S. Waiblinger. 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. Anim. Welf. 12:619–624.

Wolfová, M., J. Wolf, J. Kvapilík, and J. Kica. 2007. Selection for Profit in Cattle: I. Economic Weights for Purebred Dairy Cattle in the Czech Republic. J. Dairy Sci. 90:2442–2455. doi:10.3168/jds.2006-614.

CHAPTER 2:

The genetic structure of longevity in dairy cows

Johannes Heise[*,†], Zengting Liu[†], Kathrin F. Stock[†], Stefan Rensing[†], Friedrich Reinhardt[†], Henner Simianer[*]

[*]Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-Universität, 37075 Göttingen, Germany

[†]Vereinigte Informationssysteme Tierhaltung w.V. (**vit**), Heideweg 1, 27283 Verden, Germany

Published in *Journal of Dairy Science* 99: 1253-1265

Abstract

Longevity of dairy cows is determined by culling. Previous studies have shown that culling of dairy cows is not an unambiguous trait but rather the result of several reasons including diseases and selection decisions. The relative importance of these reasons is not stable over time, implying that genetic background of culling may vary over life time. Data of 7.6 million German Holstein cows were used to assess the detailed genetic correlation structure among 18 survival traits defined for the first three parities. Differences of genetic factors which determine survival of different production periods were found, showing a pattern with three genetically distinct periods within each parity: early lactation (calving until day 59), mid lactation (day 60 to 299) and late lactation (day 300 until next calving). Survival in first and later parities were found to be slightly genetically different from each other. The identified patterns were in good accordance with distributions of reasons for disposal, and correlations of estimated breeding values of survival traits for different periods to production and functional traits were generally plausible compared to literature regarding effects on the risk of culling. The study shows that genetic background of survival is not only variable across but also within parities. The results of the study can help developing more accurate models for routine genetic evaluations of longevity that account for non-unity genetic correlations between survival of different periods.

Key words

Longevity, culling, dairy, genetics

Introduction

Longevity of dairy cows is an economically important trait for farmers (Allaire and Gibson, 1992) and has gained in importance as a global indicator for animal welfare (Thomsen and Houe, 2006; Pritchard et al., 2013). In the last decades numerous studies have shown that longevity is heritable, and routine genetic evaluations for longevity are conducted in all major countries of dairy breeding (Miglior et al., 2005; Interbull, 2015). Longevity results from survival at successive time periods. It is genetically often treated as the same trait over the whole life of a cow (Ducrocq, 1994; Caraviello et al., 2004; González-Recio and Alenda, 2007; Pritchard et al., 2013). However, several studies suggest that survival of different parities is genetically different (Visscher and Goddard, 1995; Boettcher et al., 1999; Veerkamp et al., 2001; Sewalem et al., 2007; Holtsmark et al., 2009). Previous studies further showed that effects of different diseases (Beaudeau et al., 1994; Gröhn et al., 1998; Rajala-Schultz and Gröhn, 1999a) and reproduction traits (Rajala-Schultz and Gröhn, 1999b; Bicalho et al., 2007) on culling are dependent on the parity and also on the stage of lactation. This implies that genetic background of survival of different periods within the same lactation may differ (Ducrocq, 1999). This hypothesis is support-

ed by distributions of disposal reasons which are reported by dairy farmers. Distribution patterns of disposal reasons depend on the parity and the stage of lactation (Seegers et al., 1998; Pinedo et al., 2010). Further, Roxström and Strandberg (2002) found culling for different reasons to be genetically different and Ducrocq (2002) found strong indications that survival late in lactation is genetically distinct to survival early in lactation regardless of lactation number. Van Pelt et al. (2015) reported the genetic background of survival to be changing over time. Their definitions of survival traits based on the overall length of productive life. Lactation based definitions of monthly survival were only recently examined by Sasaki et al. (2015) in Japanese dairy cattle using a random regression model.

The aim of our study was a systematic investigation of the genetic structure of longevity regarding different periods of first, second and third parity. Distributions of disposal reasons and correlations of estimated breeding values for the new survival traits to various production and functional traits were used to validate the genetic correlation patterns found. Because survival and threshold models are computationally highly demanding and thus not feasible for extensive multivariate genetic analyses on large data sets (Boettcher et al., 1999), a linear multiple trait model was chosen for the refined survival analyses.

Material and Methods

Data

For this study, records of Holstein dairy cows used in the German routine genetic evaluation for longevity were available. Data were restricted to years of first calving between 1998 and 2014, with cut off date February 10, 2014. Records included dates of birth and calving, the herd code and, in case the cow had left the herd, the reason for and the date of disposal. Only records with complete and valid data between first and last observed calving were considered. This means, e.g., for a cow that was culled or censored during the third lactation, records of the first and second lactation had to be present in the data set. Records of cows with unknown sires or age of first calving outside the range of 500 to 1,500 days were excluded. Herds had to have at least 15 calvings for each year in the observation period between 1998 and 2013 (data for year 2014 were not complete). After editing, 7,684,455 records remained on the data pool for the analysis of survival.

For parameter estimation, data were further restricted to years of first calving from 1998 to 2008, such that each cow in the data had the opportunity to finish at least three lactations. Because estimation of variance components would not have been computationally feasible on the full data set, ten possibly overlapping samples of 200 herds each were randomly drawn. To avoid sparse category problems, only data of five out of the 16 federal states were considered. Each sample consisted of an average of 234,498 records of daughters from 7,103 bulls. Over all samples, a

total of 1,495,441 different cow records were used for parameter estimation. Data structure by lactation is shown in Table 2.1.

Table 2.1: Distribution of records by lactation.

Lactation	Number of records	
	Parameter estimation	**Breeding value estimation**
1	1,495,441	7,684,455
2	1,137,682	5,370,587
3	790,602	3,499,842

Distribution of disposal reasons

In Germany, disposal reasons are recorded routinely when a cows exits milk recording. The farmer is requested to report his/her main reason of disposal as one of the predefined disposal reasons ‘infertility’, ‘udder diseases’, ‘claw and leg disorders’, ‘metabolic diseases’, ‘other diseases’, ‘poor milk yield’, ‘milkability’, ‘age’, ‘other reasons’ or ‘sold for dairy purposes’. Only cows being disposed for other reasons than ‘sold for dairy purposes’ were considered. Frequency distributions for disposal reasons were computed by parity and relative to calving by 10-day intervals for days in milk.

Trait definition

Traits were defined as survival of different periods of the first three parities. Genetic analyses were carried out using two different period definitions: First, for evaluating the genetic structure of survival in detail, two-month periods were defined (A). Second, adjacent periods from A with minimum genetic correlations of 0.9 were joined such that fewer periods (B) were defined to achieve a simpler model for genetic evaluations. Period definitions for A and B are specified in Table 2.2. In each case, records were coded as 1 if a cow was still alive at the end of the period and 0 if culling occurred during the period. Records of cows which were culled in a previous period or censored during a period were non-informative with regard to survival and therefore not considered. Censoring was assumed when the date of disposal was missing or when the disposal reason was ‘sold for dairy purposes’. In other words, trait n was defined as survival at the end of period n, given the cow was still alive at the end of the period n-1.

Table 2.2: Definition of periods for survival traits.

Trait definition	Days from calving	Parity		
		1	**2**	**3**
A	**0-59**	A1.1	A2.1	A3.1
	60-119	A1.2	A2.2	A3.2
	120-179	A1.3	A2.3	A3.3
	180-239	A1.4	A2.4	A3.4
	240-299	A1.5	A2.5	A3.5
	300-next calving	A1.6	A2.6	A3.6
B	**0-59**	B1.1	B2.1	B3.1
	60-299	B1.2	B2.2	B3.2
	300-next calving	B1.3	B2.3	B3.3

Model for genetic analyses

The basic model equation for all linear multiple trait models was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of survival (0/1) observations, $\mathbf{X}$ is an incidence matrix linking the observations to the fixed effects, $\mathbf{b}$ is the vector of fixed effects i.e., the effect of herd * year of calving for each period, $\mathbf{Z}$ is the incidence matrix of random sire effects, $\mathbf{s}$ is the vector of random sire effects ($\mathbf{s} \sim N(0, \mathbf{G_0} \otimes \mathbf{A})$, with the genetic covariance matrix $\mathbf{G_0}$ and the numerator relationship matrix for sires $\mathbf{A}$), and $\mathbf{e}$ is a vector of random residual effects ($\mathbf{e} \sim N(0, \mathbf{R_0} \otimes \mathbf{I})$, with the residual covariance matrix $\mathbf{R_0}$). Models using trait definitions A and B are further referred to as model A and B respectively.

Estimation of variance components

To make the parameter estimations computationally feasible, the multivariate analyses were split up such that six traits each were included simultaneously. Each six-trait combination was run on each of the sample data sets described above. This resulted in 150 runs (ten samples * 15 trait combinations) with six traits each for model A and 30 runs (ten samples * three trait combinations) with six traits each for model B. Variance components were estimated using the VCE software, version 6.0 (Groeneveld et al., 2010). Full covariance matrices were computed as raw means of all genetic parameter estimates from the different runs, ignoring results of runs where convergence was not reached (12% of all runs).

Estimation of breeding values

For the genetic evaluation using model B, the genetic variance-covariance matrix $\mathbf{G}$ was composed from the results of the runs of multivariate parameter estimations in two steps: first, a matrix $\mathbf{G_0}$ was computed, calculating approximate covariances from mean genetic variances and correlations. This matrix was decomposed $\mathbf{G_0} = \mathbf{Q\Lambda Q'}$ where $\mathbf{Q}$ is the matrix of eigenvectors of $\mathbf{G_0}$ and $\mathbf{\Lambda}$ is a diagonal matrix of corresponding eigenvalues. Next, negative eigenvalues in $\mathbf{\Lambda}$ were set to 0.001, resulting in $\mathbf{\Lambda}^*$ and a positive definite matrix $\mathbf{G}$ was then computed as $\mathbf{G} = \mathbf{Q\Lambda}^*\mathbf{Q'}$. The effect of this procedure on the correlation structure was analyzed and found to be negligible (results not shown).

Sire breeding values (EBV) were estimated from the full data set with the PEST software (Groeneveld, 2006). To validate distinct genetic correlation patterns of survival, Pearson correlation coefficients were computed between raw EBV of the particular survival traits and raw EBV of various production and functional traits from the routine national genetic evaluation for dairy cattle. Considered traits from the routine genetic evaluations were ‘functional longevity’, an index for ‘milk production’, ‘somatic cell score’, ‘stillbirth’ and ‘first to last insemination’ as described in the official documentation of the routine genetic evaluation for Holsteins in Germany (vit, 2015).

Correlations of EBV were computed for sires that were born before 2005 and had more than 50 daughters with a first calving and a minimum reliability for the above mentioned routinely EBV of 0.9. All EBV in the comparison were scaled such that higher values indicated genetic disposition for more favorable trait expressions.

Results

Survival patterns

Estimates for the risk of culling and the proportion of survived cows are shown in Figure 2.1 for trait definition A following Kaplan and Meier (1958). The highest risk of a cow to be culled given that she had survived the previous periods was found at the beginning and end of a parity (e.g., 0.076 for A2.1 and 0.122 for A2.6) while it was nearly constant for the other periods (0.035 to 0.036 for A2.2 to A2.5). For corresponding periods in parities one to three, the risk of culling increased over lactations and was highest for A3.6 (0.144). Proportions of 77.5%, 54.4% and 33.4% of all cows were still alive at the end of the last periods of the first, second and third parity, respectively.

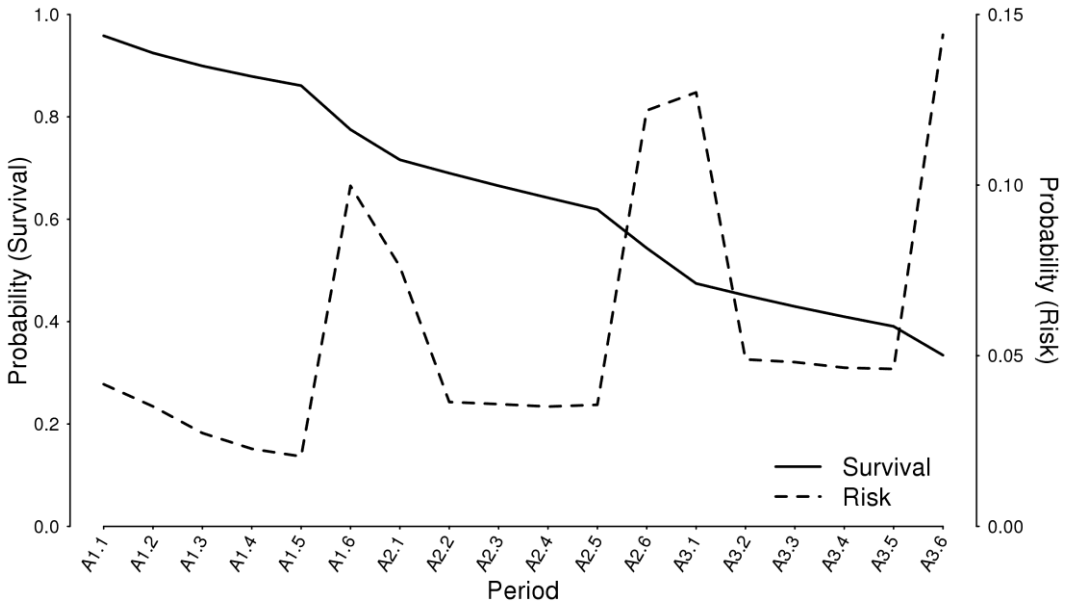


Figure 2.1: Kaplan-Meier-Estimators for survival and risk based on trait definition A (six periods per parity).

Distribution of disposal reasons

Distributions of disposal reasons by parities one to five are shown in Figure 2.2 for cows that were culled during the years 2010 to 2013. Parities four to five are shown to assess possible differences to earlier parities. Across the considered parities, main reasons for culling were ‘infertility’ (20.4%), ‘udder diseases’ (14.7%), ‘claw and leg disorders’ (12.2%) and ‘other reasons’ (30.7%). Differences in the distributions between first and later parities occurred mainly for ‘poor milk yield’, ‘milkability’, ‘udder diseases’, ‘metabolic diseases’ and ‘other diseases’, while distributions for ‘infertility’ and ‘claw and leg disorders’ were similar over parities. Frequencies for ‘metabolic diseases’, ‘other diseases’ and, for the first parity, ‘poor milk yield’ and ‘milkability’ peaked early in lactation while ‘udder diseases’ and ‘claw and leg disorders’ showed highest incidences in the middle of a lactation (about 60-180 days from calving). The frequency of ‘infertility’ as a disposal reason increased towards the end of the lactation for all parities and reached 50% for the interval 490 to 499 days from calving for the second parity.

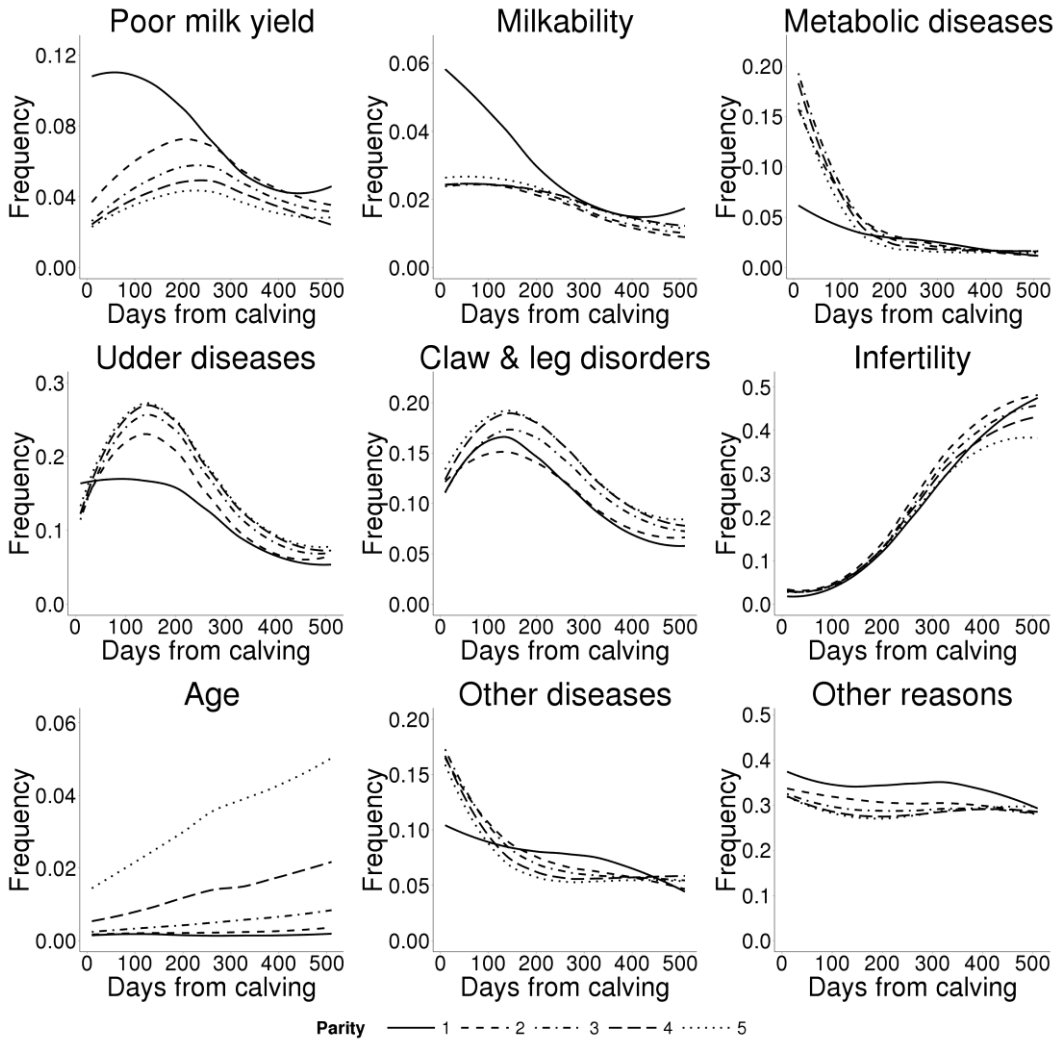


Figure 2.2: Time-dependent distributions of disposal reasons by parity. Number of disposed cows within each 10-day interval from calving are considered to be 100%. Lines were smoothed with a locally weighted regression based on first order local polynomials (Cleveland, 1979).

Genetic parameters from model A (six periods per parity)

As shown in Table 2.3, mean estimates of heritabilities on the observed scale ranged from 0.005 (A1.5) to 0.041 (A3.1) for the two-month interval trait definition in model A. First and last periods of a parity showed highest heritability estimates while those for mid-lactation periods were lower and very similar. After transformation, approximate heritabilities on the underlying scale (Dempster and Lerner, 1950) ranged between 0.038 (A1.5) and 0.105 (A3.1). Mean genetic correlations (Figure 2.3) ranged from 0.37 ($r_{g_{A1.1,A3.6}}$) to 0.96 ($r_{g_{A2.1,A3.1}}$). Standard deviations of genetic correlations over different runs ranged from 0.02 ($r_{g_{A2.1,A3.1}}$) to 0.2 ($r_{g_{A1.4,A2.6}}$). First and last periods of a parity showed lower genetic correlations to adjacent periods than the mid-lactation periods. This difference was found to be most extreme in the third lactation where

$r_{g_{A3.1,A3.2}}$ and $r_{g_{A3.5,A3.6}}$ were 0.73 and 0.61 respectively while the correlations between adjacent mid-lactation traits ranged from 0.90 to 0.93. Genetic correlations of periods one to five of the first parity to corresponding periods of the second parity were lower (0.80-0.82) than genetic correlations between the respective periods of parities two and three (0.88-0.96). Means for residual correlation estimates were close to zero. All means, standard deviations and numbers of runs with valid results that were included into the means are provided in supplementary Table 2.1.

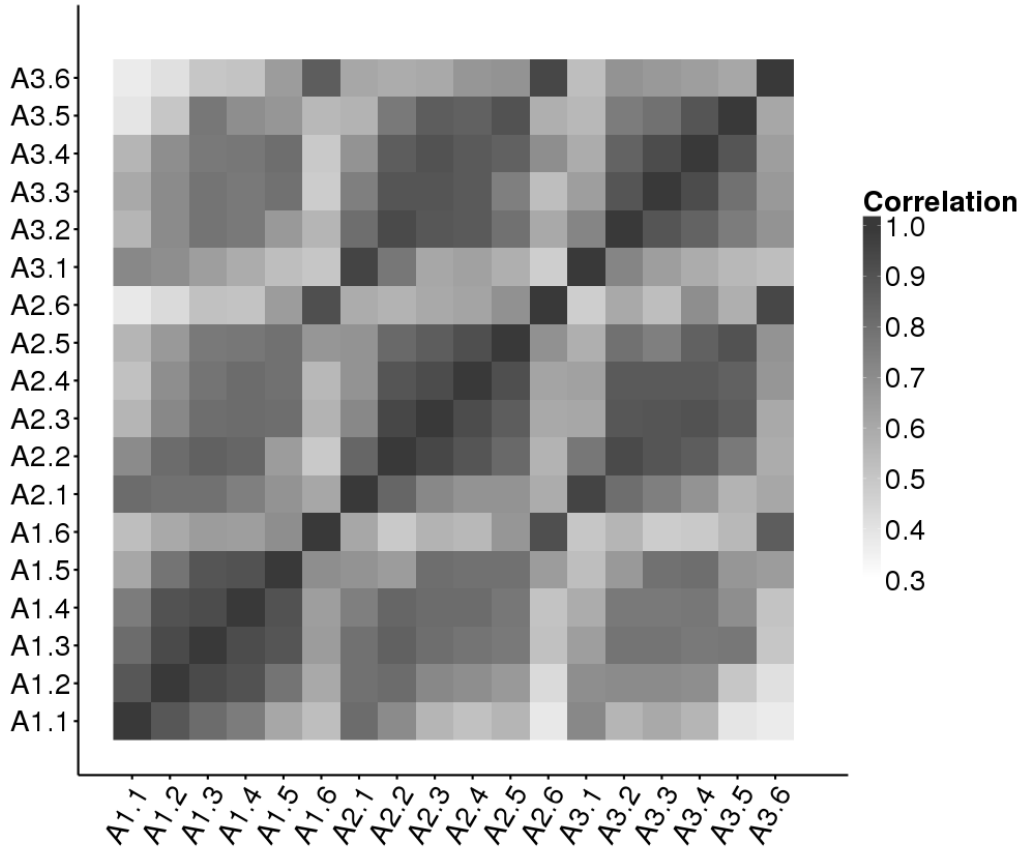


Figure 2.3: Genetic correlations from model A (six periods per partiy). Estimates are means of genetic correlations from the different runs. The values and standard deviations can be seen in detail from supplementary Table 2.1.

Table 2.3: Phenotypic frequencies, estimates of heritability from model A (six periods per parity) on the observed scale and approximated heritabilities on the underlying scale (DL: Dempster and Lerner, 1950)[1].

Trait	Phenotypic frequency	h^2		h^2 (DL)	
		Mean	SD	Mean	SD
A1.1	.96	.017	.002	.083	.009
A1.2	.96	.010	.002	.059	.009
A1.3	.97	.007	.002	.044	.010
A1.4	.98	.006	.002	.046	.014
A1.5	.98	.005	.001	.038	.011
A1.6	.90	.021	.002	.060	.005
A2.1	.92	.023	.003	.080	.010
A2.2	.96	.009	.002	.051	.012
A2.3	.96	.011	.003	.062	.014
A2.4	.96	.013	.003	.074	.018
A2.5	.96	.011	.002	.060	.013
A2.6	.88	.022	.004	.057	.010
A3.1	.87	.041	.007	.105	.018
A3.2	.95	.014	.003	.062	.015
A3.3	.95	.012	.002	.056	.009
A3.4	.95	.013	.003	.060	.014
A3.5	.95	.012	.003	.056	.014
A3.6	.85	.027	.004	.064	.009

[1]Heritability estimates are shown with means and standard deviations of estimates from the different runs.

Genetic parameters from model B (three periods per parity)

In model B, periods from model A with genetic correlations to adjacent periods of ≥0.9 were joined. As shown in Table 2.4, mean heritability estimates from model B ranged from 0.016 (B1.1) to 0.042 (B3.1). As for model A, heritabilities tended to increase over lactations. In contrast to the third parity, first periods of parities one and two showed lower or similar heritabilities than later periods of the same lactation. After transformation, approximate heritabilities on the underlying scale (Dempster and Lerner, 1950) ranged between 0.053 (B2.3) and 0.107 (A3.1). All mean genetic correlations (Figure 2.4) between periods of the same parity were below 0.9. Genetic correlations between corresponding periods of successive lactations were higher (0.82 to 0.96) than correlations within parity (0.52 to 0.81). Furthermore, corresponding periods of the

second and third parity were higher correlated (0.95 to 0.96) than corresponding periods of the first and second parity (0.82 to 0.93). The third periods showed high genetic correlations across all parities (0.93 to 0.95). Standard deviations of genetic correlations over the different runs ranged from 0.02 ($r_{g_{B2.1,B3.1}}$) to 0.12 ($r_{g_{B1.1,B3.2}}$). Means for residual correlation estimates were close to zero. All means, standard deviations and numbers of runs with valid results that contributed to the means are provided in supplementary Table 2.2.

Table 2.4: Phenotypic frequencies, estimates of heritability from model B (three periods per parity) on the observed scale and approximated heritabilities on the underlying scale (DL: Dempster and Lerner, 1950)[1].

Trait	Phenotypic frequency	h^2		h^2 (DL)	
		Mean	SD	Mean	SD
B1.1	.96	.016	.002	.080	.011
B1.2	.90	.022	.003	.065	.008
B1.3	.90	.020	.002	.058	.006
B2.1	.92	.023	.003	.078	.010
B2.2	.86	.033	.005	.081	.011
B2.3	.88	.020	.003	.053	.009
B3.1	.87	.042	.006	.107	.016
B3.2	.82	.039	.005	.084	.011
B3.3	.85	.026	.004	.061	.010

[1]Heritability estimates are shown with means and standard deviations of estimates from the different runs.

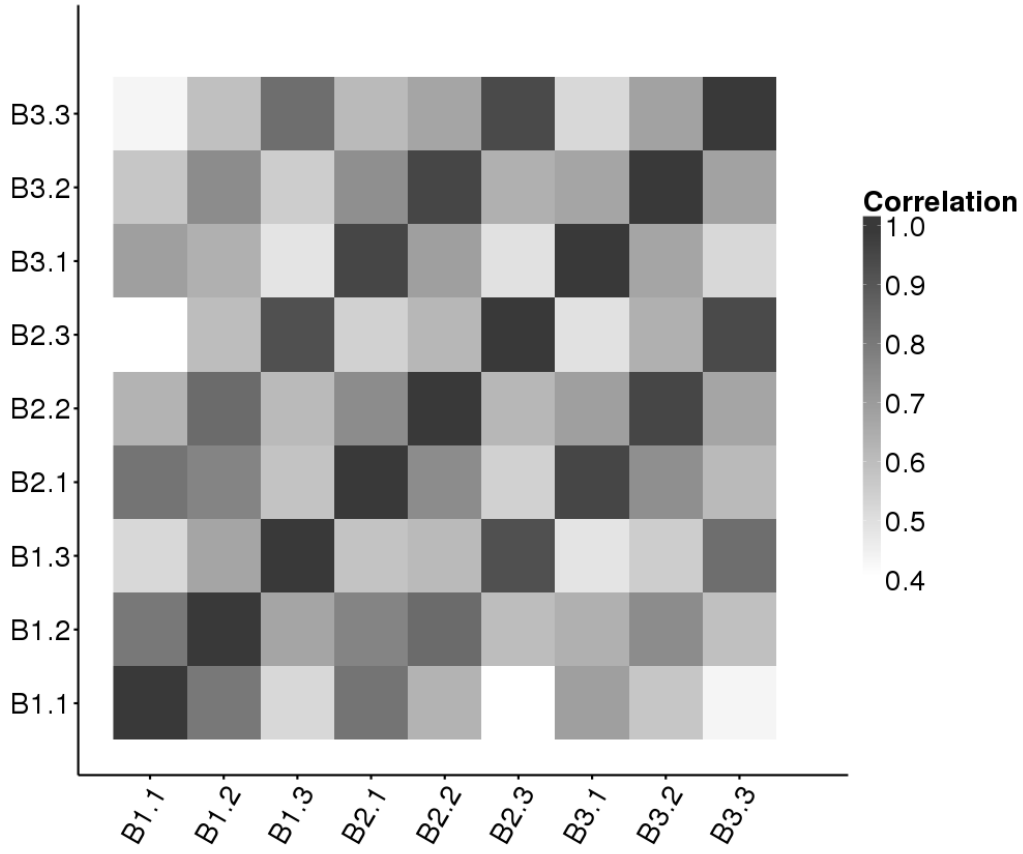


Figure 2.4: Genetic correlations from model B (three periods per parity). Estimates are means of genetic correlations from the different runs. The values and standard deviations can be seen in detail from supplementary Table 2.2.

EBV correlations to other traits

Correlations of EBV for survival traits (model B; ≥50 daughters with first calving) to EBV from the routine genetic evaluation (reliability ≥0.9) are shown in Figure 2.5. Correlations to EBV for 'functional longevity' (N=1468) ranged from 0.67 (B1.1) to 0.79 (B3.3). EBV Correlations to 'first to last insemination' were highest to the last periods of all parities (0.43 to 0.46 compared to 0.04 to 0.10 for first and second periods). Correlations to 'milk production' (N=8,743) ranged from 0.09 (B2.3) to 0.30 (B1.2) and were highest to survival traits of the first parity. All values are shown in supplementary Table 2.3.

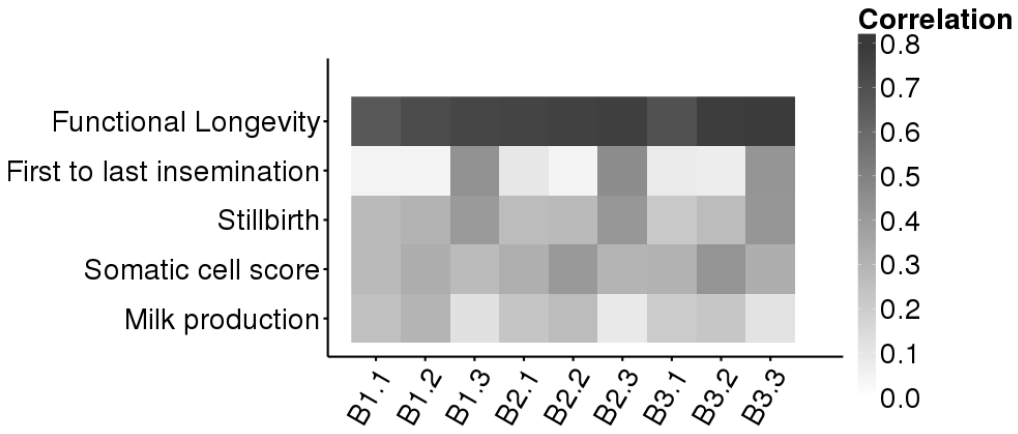


Figure 2.5: Correlations of estimated breeding values from model B (three periods per parity) to other traits for sires with more than 50 daughters on the dataset. Reliabilities of estimated breeding values were required to be ≥0.9. The values, amount of bulls contributing to the different correlation coefficients and levels of significance can be seen in detail from supplementary Table 2.3.

Discussion

Data basis

With 1,495,441 cows, the data set used for the variance component estimation was larger than in other studies on genetics of survival (Boettcher et al. (1999): 699,722 Canadian Holstein cows; Holtsmark et al. (2009): 800,331 Norwegian Red cows; van Pelt et al. (2015): 112,000 Dutch Holstein cows; Veerkamp et al. (2001): 24,741 UK Holstein cows; Visscher and Goddard (1995): 190,830 Australian Holstein cows). Although sample size is only one aspect for the quality of genetic parameter estimates, a larger information basis should imply that results are more relevant for practical applications.

Increasing risk of culling over parities was previously reported for US Holstein cows (Hadley et al., 2006; De Vries et al., 2010; Pinedo et al., 2010) and for French dairy cows (Ducrocq, 2005) which is consistent with our results. Furthermore, our results are in line with previously reported peaks of culling risk at the very beginning (Hadley et al., 2006; De Vries et al., 2010) and end (Ducrocq, 2005; Hadley et al., 2006) of a lactation. However, much higher culling rates early in first lactation were reported by Römer (2011) for herds enrolled in a program with extended recording obligations. This data might be more complete compared to data from routine milk recording systems as used in our study. Cows that calved but were disposed before the first test day of the new lactation might, dependent on the respective reporting system, not appear in such data. In our study, this effect is particularly expected for the first lactation.

Data for parameter estimation was split into different combination sets of six traits per run. For combinations, where the first trait (in this case A1.1 or B1.1) was missing, selection effects on the estimated genetic parameters were expected (Pollak and Quaas, 1981). Although the majority of combinations used in our analysis contained at least two traits of the first parity, it was not possible to include the first trait of the first parity in all runs, because the number of necessary combinations would have largely increased. By analyzing the selection effect on genetic parameters that were estimated using multiple different combinations, it was found to be minor on heritabilities and genetic correlations. E.g., the genetic correlation between A3.5 and A3.6 was estimated based on two extreme combinations amongst others: The first included the periods A1.1, A1.2, A2.5, A2.6, A3.5 and A3.6, the other one A3.1, A3.2, A3.3, A3.4, A3.5 and A3.6, i.e. only third parity traits. Mean estimates for ($r_{g_{A3.5,A3.6}}$) from the different runs were 0.61 and 0.65, respectively, mean heritability estimates for A3.5 were 0.012 and 0.013, and for A3.6 0.029 and 0.027, respectively.

Distribution of disposal reasons

Pinedo et al. (2010) and Hadley et al. (2006) referred to a slightly different set of disposal reasons in US Holstein cows (‘reproduction’, ‘low production’, ‘injury/other’, ‘disease’, ‘mastitis’, ‘udder problems’, ‘feet and legs’, ‘died’ and ‘reason not reported’). Overall frequencies for ‘reproduction’ and ‘mastitis’ were similar to disposal reasons ‘infertility’ and ‘udder diseases’, respectively from our study, but they were lower for ‘feet and legs’ than for ‘claw and leg disorders’ in our study. In French Holstein cows, higher overall frequencies were described for ‘reproduction’ (28.5%) and ‘low milk yield’ (16.6%) as disposal reasons, whereas ‘lameness and foot/leg defects’ had a much lower frequency (2.7%) (Seegers et al., 1998). Somewhat heterogeneous results of study outcomes may relate to the fact that in many cases a farmer may have more than a single reason to cull a particular cow (Fetrow et al., 2006), but he can only report one to the breeding organization, which usually will be the main reason from his point of view. The relevance of reasons as well as the composition of underlying reasons is likely to change over time and region. Further, the frequency of disposal reasons is influenced by the predefined set of disposal codes. The relatively high amount of ‘other reasons’ in our study indicated that there is room for improvement of documentation around disposals. This could be a more precisely defined code set and/or an increased motivation for the farmers to accurately report the reasons for disposal. The latter would be difficult to obtain in practice. However, given the stability of the proportion of ‘other reasons’ over the studied period, this issue did not influence the shape of distributions of other disposal reasons, justifying conclusions based on their patterns. Disposal reasons like ‘claw and leg disorders’, ‘poor milk yield’, ‘infertility’, ‘udder diseases’ and ‘other diseases’ are obviously linked to heritable traits and can therefore serve as indicators for the contribution of those traits to genetic factors affecting survival at different lactation periods.

Model for genetic evaluation

Proportional hazard models are often considered to suit censored time-to-event data best (e.g., Ducrocq, 1994; Neerhof et al., 2000). However, because computing time was much higher with threshold and survival models (Boettcher et al., 1999), a similarly extensive study regarding sample sizes and simultaneous analysis of survival traits would not have been feasible with such models within reasonable time. Meuwissen et al. (2002) used data that was simulated under a Weibull model. They analyzed the performance of a proportional hazards model, a threshold model with a logit link function and a linear model and found very similar correlations between estimated and true breeding values for all methods. Furthermore, linear multiple trait models for genetic evaluations were found to outperform threshold and survival models under practical conditions regarding the ability to predict survival of the first 365 days in milk for second crop daughters of sires (Holtsmark et al., 2009). The authors of this study assumed that this might be due to the accommodation of multiple genetic effects which could not be applied to the other models in their comparison. Additionally, Sewalem et al. (2005) reported that genetic trends from a Weibull survival model seemed to be overestimated for young sires with a high amount of censored daughter information. The authors explained this overestimation with their model which kept the parameters of the baseline hazard function constant over time although there was a systematic trend in reality. However, in the German routine genetic evaluation for functional longevity which allows the parameters of the baseline hazard function to vary over time, lactation and stage of lactation (vit, 2015), similar patterns of bias are observed. As a special case of linear models, random regression models were suggested and successfully used on survival data in the past (Veerkamp et al., 2001; Sasaki et al., 2015; van Pelt et al., 2015). Random regression models have the advantage to keep the number of parameters low and could be an interesting alternative for routine evaluations (Gengler et al., 2005).

Heritability estimates

When using linear models on binary response variables, heritability estimates on the observed scale tend to be lower than true heritabilities, with increasing underestimation for more extreme frequencies. Breeding progress and reliabilities refer to the observed scale. However, approximation to the underlying scale allows assessing the relative importance of genetic components on different traits with different phenotypic frequencies (Dempster and Lerner, 1950). Heritability estimates on the observed scale for first and last periods of each parity were higher than for other periods due to higher culling frequencies. Sasaki et al. (2015) only found increasing heritabilities towards the end of different lactations which might be related to the close-to-zero phenotypic variance at the beginning of each lactation in their study. Approximated heritabilities on the underlying scale as proposed by Dempster and Lerner (1950) for model A showed a slightly different pattern (Table 2.3), with last periods being more similar to survival traits in the middle of the lactation. First periods still tended to show higher heritabilities than the other periods. When in-

terpreting these results it must be taken into account that transformation of heritabilities may have some upward bias for low phenotypic frequencies (Stock et al., 2005).

The relatively high heritability estimates on the observed and approximated underlying scale for the first period of the third parity compared to first and second parities might be explained by an increased incidence of early lactation disorders with higher heritabilities compared to other disorders. This could, e.g, be displaced abomasum for which a relatively high heritability has been reported (Zwald et al., 2004). Its incidence is strongly increasing over parities, especially from second to third lactation (K.F. Stock, unpublished data). However, standard deviations of heritability estimates in our study were relatively high compared to the differences between periods, so results should be interpreted with caution.

Genetic correlation estimates

Contrary to heritability estimates of binary response variables from linear models, estimates of genetic correlations are the same on the observed and the underlying scale (Vinson et al., 1976; Gianola, 1982). From model A, three genetically homogenous periods per parity could be derived. Genetic background of survival seemed to be similar within each of these periods and distinct from survival of the other periods. Definitions for model B were therefore 0 to 59, 60 to 299, and 300 days from calving to the subsequent calving.

Although the estimation of genetic correlations between the particular survival traits and other traits was not covered by our study, their direction is indicated by distributions of culling reasons and EBV correlations from model B. Both assign survival of different periods to distinct trait complexes: Culling for ‘metabolic diseases’ mostly occurs at the beginning of second and later parities which might explain why survival of B2.1 and B3.1 (days 0 to 59 from calving in second and third parities) is genetically distinct to survival of other periods and also slightly distinct to survival of B1.1. This hypothesis is also supported by previous studies which found that ketosis and milk fever increase the risk of culling early in lactation while the incidence of milk fever is considerably lower in the first parity than in later parities (Beaudeau et al., 1994; Gröhn et al., 1998; Rajala-Schultz and Gröhn, 1999a). Literature further leads to the assumption that survival of the first periods of different parities might be genetically linked to displaced abomasum (Rajala-Schultz and Gröhn, 1999a) and calving traits like stillbirths (Bicalho et al., 2007).

The mid-lactation period (days 60 to 299 from calving) showed highest EBV correlations to ‘milk production’ and ‘somatic cell score’. Culling within this period is associated with a low milk yield (Rajala-Schultz and Gröhn, 1999c). An association was also found for mastitis (Rajala-Schultz and Gröhn, 1999a; Neerhof et al., 2000) which is genetically correlated to somatic cell score (Rupp and Boichard, 1999; Koeck et al., 2012). Distributions of disposal reasons are consistent with previous studies and support these results: ‘Udder diseases’ showed a peak for the mid-lactation period. From another peak for ‘claw and leg disorders’ it might also be assumed that survival of the mid-lactation periods could be genetically related to claw and leg dis-

orders. This assumption is supported by findings from Rajala-Schultz and Gröhn (1999a) and Sogstad et al. (2005) who reported claw disorders to increase the risk of culling and to have main incidences three to seven months after calving.

Cows still alive at the end of the last period of each parity (300 days from calving to subsequent calving) must have had a successful reproduction. There is probably a delay between the decision-making for culling and the time point of culling. When the cow did not conceive, she is still milked until her milk yield drops below a threshold related to economic profitability. Therefore, 'infertility' was the most frequent reason for disposal in the last period. EBV correlations of 'first to last insemination' were highest to survival of this period, confirming a genetic association. This was in accordance with results from De Vries et al. (2010), Rajala-Schultz and Gröhn (1999b) and Schneider et al. (2007) who reported increasing risk ratios towards the end of a lactation for culling of open cows compared to pregnant cows. 'Stillbirth' also showed higher EBV correlations to survival of the last parity period than to other periods. 'Stillbirth' is reported to cause lasting reproductive problems (Bicalho et al., 2007), but also to have an effect on culling early in lactation (Bicalho et al., 2007; Vergara et al., 2014). The latter could not be seen from our results. A possible explanation is that the calving is usually not reported for cows that died during or shortly after the calving of a stillborn calf. In this case, culling is assigned to the last period of the previous parity.

The clearer genetic distinction between periods of the same parity for second and third parity compared to the first parity might be explained by the different distributions of disposal reasons. While important culling reasons ('udder diseases', 'metabolic diseases', 'other diseases') show sharp peaks assigned to a single period in later parities, they are more widespread during the first parity. Further, culling reasons mainly belonging to the first parity ('poor milkyield', 'milkability') can not only be assigned to the first but also to the mid-lactation period.

Previous genetic studies using multivariate analysis for survival traits suggested that survival of the first lactation may be genetically distinct from survival of later lactations (Visscher and Goddard, 1995; Boettcher et al., 1999; Veerkamp et al., 2001; Holtsmark et al., 2009). Our study supports this assumption. From the more detailed model A, genetic correlation estimates between periods one to five of the first parity to the corresponding periods of the second parity were lower (0.80 to 0.82) than between second and third parity (0.88 to 0.96). The last period was genetically closely correlated across all considered parities indicating that previously found differences in the genetic background of survival between first and later lactations originate from all but the last periods. This pattern was also seen with regard to the frequencies for different disposal reasons. Most differences between first and later parities occurred until 300 days from calving. Distributions were very similar between second and later parities and towards the end of each parity where 'infertility' consistently became the main reason for disposal in all parities.

EBV correlations of survival traits from our study to routinely estimated EBV for functional longevity (0.67 to 0.79) were only moderate taking into account the similarity of traits. One reason

is probably the different definition. Functional longevity is defined as longevity corrected for milk yield relative to the herd mean while raw survival was used in the new model. Another explanation might be that survival of each period considered here reflected only parts of the genetic basis of total longevity.

Correlations of EBV for survival traits to 'milk production' were moderately positive (0.09 to 0.30), because survival was not corrected for any measure of voluntary culling. For routinely estimated functional longevity, the genetic correlation to 'milk production' is close to zero (vit, 2015). The EBV correlations found in our study were similar to genetic correlations estimated by Olori et al. (2002) between milk yield and survival from first to second lactation which was phenotypically corrected for milk yield. Higher genetic correlations were previously estimated between raw survival and milk yield traits (Short and Lawlor, 1992; Visscher and Goddard, 1995; Haile-Mariam et al., 2003). Conversely, Dematawewa and Berger (1998) reported negative genetic correlations of survival to milk yield traits.

Visscher et al. (1999) derived that residual covariances for multivariate genetic evaluations for longevity using binary traits and a linear animal model must be zero by construction. Using a sire model for the estimation of variance components, Olori et al. (2002) estimated the residual covariance as three times the sire covariance. In our study, however, residual covariances were estimated without further preassumptions, and resulting residual correlations were close to zero.

Consistency of results

For validation of consistency of results over different period definitions and differently split models, parameter estimates were summed up by parity and over all three parities using $\mathbf{G}^* = \mathbf{W}'\mathbf{G}\mathbf{W}$ and $\mathbf{P}^* = \mathbf{W}'\mathbf{P}\mathbf{W}$. $\mathbf{G}^*$ and $\mathbf{P}^*$ are the summarized genetic and phenotypic covariance matrices, $\mathbf{G}$ and $\mathbf{P}$ are the genetic and phenotypic covariance matrices resulting from model A or B and $\mathbf{W}$ is an incidence matrix, linking estimates to the desired measurements, i.e. to parities or whole three parity survival. For three parity survival, $\mathbf{W}'$ reduced to $\mathbf{1}'$.

The heritability estimates for three parity survival were 0.183 and 0.170 for model A and B respectively. Further, Table 2.5 shows estimates for heritabilities and genetic correlations of the complete parities for models A and B. The results show consistency over different period definitions and differently split models. Patterns for parity-wise genetic parameters were similar to results previously reported (Boettcher et al., 1999), but the estimates for genetic correlations and heritabilities were slightly higher in our study. For survival of the first and third lactation, lower genetic correlations (about 0.65) were reported by Holtsmark et al. (2009) and Veerkamp et al. (2001).

Table 2.5: Genetic correlations (off-diagonal) and heritabilities (diagonal) per parity from model A and B.

	Model A			Model B		
Parity	**1**	**2**	**3**	**1**	**2**	**3**
1	.046	.874	.816	.044	.896	.811
2		.062	.967		.056	.981
3			.079			.076

Conclusions

Our study gives evidence that genetic background of survival varies between different periods of a cow's lifetime. This variation is higher for different periods of the same parity than for corresponding periods of successive parities. Within each parity, three periods with distinct genetic background of survival were derived: 0 to 59, 60 to 299 and 300 days from calving until the consecutive calving. Most genetic correlations for survival of periods of the first parity to corresponding periods of later parities were lower than respective genetic correlations between second and third parities. The genetic structure corresponded to time-dependent distributions of disposal reasons. Correlation patterns of EBV from the linear multiple trait model to EBV of production and functional traits further confirmed the consistency of estimated genetic parameters. Although many previous studies already touched parts of the research question, our results add new aspects concerning the genetic correlations between different periods of the first three parities. They can serve as a basis for developing more accurate models for routine genetic evaluations for longevity which account for the distinct genetic correlation structure regardless of the actual type of model that will be implemented.

Acknowledgments

The German national FBF (Förderverein Bioökonomieforschung e.V.) is thanked for the financial support.

References

Allaire, F. R., and J. P. Gibson. 1992. Genetic Value of Herd Life Adjusted for Milk Production. J. Dairy Sci. 75:1349–1356. doi:10.3168/jds.S0022-0302(92)77886-2.

Beaudeau, F., K. Frankena, C. Fourichon, H. Seegers, B. Faye, and J. P. T. M. Noordhuizen. 1994. Associations between health disorders of French dairy cows and early and late culling within the lactation. Prev. Vet. Med. 19:213–231. doi:10.1016/0167-5877(94)90090-6.

Bicalho, R. C., K. N. Galvão, S. H. Cheong, R. O. Gilbert, L. D. Warnick, and C. L. Guard. 2007. Effect of Stillbirths on Dam Survival and Reproduction Performance in Holstein Dairy Cows. J. Dairy Sci. 90:2797–2803. doi:10.3168/jds.2006-504.

Boettcher, P. J., L. K. Jairath, and J. C. M. Dekkers. 1999. Comparison of Methods for Genetic Evaluation of Sires for Survival of Their Daughters in the First Three Lactations. J. Dairy Sci. 82:1034–1044. doi:10.3168/jds.S0022-0302(99)75324-5.

Caraviello, D. Z., K. A. Weigel, and D. Gianola. 2004. Comparison Between a Weibull Proportional Hazards Model and a Linear Model for Predicting the Genetic Merit of US Jersey Sires for Daughter Longevity. J. Dairy Sci. 87:1469–1476. doi:10.3168/jds.S0022-0302(04)73298-1.

Cleveland, W. S. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. J. Am. Stat. Assoc. 74:829–836. doi:10.1080/01621459.1979.10481038.

Dematawewa, C. M. B., and P. J. Berger. 1998. Genetic and Phenotypic Parameters for 305-Day Yield, Fertility, and Survival in Holsteins. J. Dairy Sci. 81:2700–2709. doi:10.3168/jds.S0022-0302(98)75827-8.

Dempster, E. R., and I. M. Lerner. 1950. Heritability of Threshold Characters. Genetics 35:212–236.

De Vries, A., J. D. Olson, and P. J. Pinedo. 2010. Reproductive risk factors for culling and productive life in large dairy herds in the eastern United States between 2001 and 2006. J. Dairy Sci. 93:613–623. doi:10.3168/jds.2009-2573.

Ducrocq, V. 1994. Statistical Analysis of Length of Productive Life for Dairy Cows of the Normande Breed. J. Dairy Sci. 77:855–866. doi:10.3168/jds.S0022-0302(94)77020-X.

Ducrocq, V. 1999. Topics that may deserve future attention in survival analysis applied to dairy cattle breeding - some suggestions. Interbull Bull. No 21:181–189.

Ducrocq, V. 2002. A piecewise Weibull mixed model for the analysis of length of productive life of dairy cows. In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Ducrocq, V. 2005. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. Anim. Sci. 80:249–256. doi:10.1079/ASC41720249.

Fetrow, J., K. V. Nordlund, and H. D. Norman. 2006. Invited Review: Culling: Nomenclature, Definitions, and Recommendations. J. Dairy Sci. 89:1896–1905. doi:10.3168/jds.S0022-0302(06)72257-3.

Gengler, N., S. Vanderick, P. Mayeres, A. Gillon, and C. Croquet. 2005. Genetic evaluation of cow survival using a lactation random regression model. Interbull Bull. No 33:176–180.

Gianola, D. 1982. Theory and Analysis of Threshold Characters. J. Anim. Sci. 54:1079–1096. doi:10.2134/jas1982.5451079x.

González-Recio, O., and R. Alenda. 2007. Genetic relationship of discrete-time survival with fertility and production in dairy cattle using bivariate models. Genet. Sel. Evol. 39:391. doi:10.1186/1297-9686-39-4-391.

Groeneveld, E. 2006. PEST User's Manual. Institute of Animal Science, D-31535 Neustadt, Germany.

Groeneveld, E., M. Kovac, and N. Mielenz. 2010. VCE User's Guide and Reference Manual Version 6.0. Institute of Farm Animal Genetics, Friedrich Loeffler Institute (FLI), D-31535 Neustadt, Germany.

Gröhn, Y. T., S. W. Eicker, V. Ducrocq, and J. A. Hertl. 1998. Effect of Diseases on the Culling of Holstein Dairy Cows in New York State. J. Dairy Sci. 81:966–978. doi:10.3168/jds.S0022-0302(98)75657-7.

Hadley, G. L., C. A. Wolf, and S. B. Harsh. 2006. Dairy Cattle Culling Patterns, Explanations, and Implications. J. Dairy Sci. 89:2286–2296. doi:10.3168/jds.S0022-0302(06)72300-1.

Haile-Mariam, M., P. J. Bowman, and M. E. Goddard. 2003. Genetic and environmental relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy cattle. Livest. Prod. Sci. 80:189–200. doi:10.1016/S0301-6226(02)00188-4.

Holtsmark, M., B. Heringstad, and J. Ødegård. 2009. Predictive abilities of different statistical models for analysis of survival data in dairy cattle. J. Dairy Sci. 92:5730–5738. doi:10.3168/jds.2009-2132.

Interbull. 2015. National genetic evaluation forms provided by countries. Accessed Jun. 26, 2015. http://interbull.org/ib/geforms.

Kaplan, E. L., and P. Meier. 1958. Nonparametric Estimation from Incomplete Observations. J. Am. Stat. Assoc. 53:457–481. doi:10.1080/01621459.1958.10501452.

Koeck, A., F. Miglior, D. F. Kelton, and F. S. Schenkel. 2012. Alternative somatic cell count traits to improve mastitis resistance in Canadian Holsteins. J. Dairy Sci. 95:432–439. doi:10.3168/jds.2011-4731.

Meuwissen, T. H. E., R. F. Veerkamp, B. Engel, and S. Brotherstone. 2002. Single and multitrait estimates of breeding values for survival using sire and animal models. Anim. Sci. 75:15–24.

Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection Indices in Holstein Cattle of Various Countries. J. Dairy Sci. 88:1255–1263. doi:10.3168/jds.S0022-0302(05)72792-2.

Neerhof, H. J., P. Madsen, V. P. Ducrocq, A. R. Vollema, J. Jensen, and I. R. Korsgaard. 2000. Relationships Between Mastitis and Functional Longevity in Danish Black and White Dairy Cattle Estimated Using Survival Analysis. J. Dairy Sci. 83:1064–1071. doi:10.3168/jds.S0022-0302(00)74970-8.

Olori, V. E., T. H. E. Meuwissen, and R. F. Veerkamp. 2002. Calving Interval and Survival Breeding Values as Measure of Cow Fertility in a Pasture-Based Production System with Seasonal Calving. J. Dairy Sci. 85:689–696. doi:10.3168/jds.S0022-0302(02)74125-8.

Pinedo, P. J., A. De Vries, and D. W. Webb. 2010. Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. J. Dairy Sci. 93:2250–2261. doi:10.3168/jds.2009-2572.

Pollak, E. J., and R. L. Quaas. 1981. Monte Carlo Study of Genetic Evaluations Using Sequentially Selected Records. J. Anim. Sci. 52:257–264. doi:10.2134/jas1981.522257x.

Pritchard, T., M. Coffey, R. Mrode, and E. Wall. 2013. Understanding the genetics of survival in dairy cows. J. Dairy Sci. 96:3296–3309. doi:10.3168/jds.2012-6219.

Rajala-Schultz, P. J., and Y. T. Gröhn. 1999a. Culling of dairy cows. Part I. Effects of diseases on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:195–208. doi:10.1016/S0167-5877(99)00046-X.

Rajala-Schultz, P. J., and Y. T. Gröhn. 1999b. Culling of dairy cows. Part II. Effects of diseases and reproductive performance on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:279–294. doi:10.1016/S0167-5877(99)00045-8.

Rajala-Schultz, P. J., and Y. T. Gröhn. 1999c. Culling of dairy cows. Part III. Effects of diseases, pregnancy status and milk yield on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:295–309. doi:10.1016/S0167-5877(99)00047-1.

Römer, A. 2011. Untersuchung zur Nutzungsdauer bei Deutschen Holstein Kühen. Züchtungskunde 83:8–20.

Roxström, A., and E. Strandberg. 2002. Genetic analysis of functional, fertility-, mastitis-, and production-determined length of productive life in Swedish dairy cattle. Livest. Prod. Sci. 74:125–135. doi:10.1016/S0301-6226(01)00300-1.

Rupp, R., and D. Boichard. 1999. Genetic Parameters for Clinical Mastitis, Somatic Cell Score, Production, Udder Type Traits, and Milking Ease in First Lactation Holsteins. J. Dairy Sci. 82:2198–2204. doi:10.3168/jds.S0022-0302(99)75465-2.

Sasaki, O., M. Aihara, A. Nishiura, H. Takeda, and M. Satoh. 2015. Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. J. Dairy Sci. 98:5781–5795. doi:10.3168/jds.2014-9152.

Schneider, M. del P., E. Strandberg, U. Emanuelson, K. Grandinson, and A. Roth. 2007. The effect of veterinary-treated clinical mastitis and pregnancy status on culling in Swedish dairy cows. Prev. Vet. Med. 80:179–192. doi:10.1016/j.prevetmed.2007.02.006.

Seegers, H., F. Beaudeau, C. Fourichon, and N. Bareille. 1998. Reasons for culling in French Holstein cows. Prev. Vet. Med. 36:257–271. doi:10.1016/S0167-5877(98)00093-2.

Sewalem, A., G. J. Kistemaker, V. Ducrocq, and B. J. Van Doormaal. 2005. Genetic Analysis of Herd Life in Canadian Dairy Cattle on a Lactation Basis Using a Weibull Proportional Hazards Model. J. Dairy Sci. 88:368–375. doi:10.3168/jds.S0022-0302(05)72696-5.

Sewalem, A., F. Miglior, G. J. Kistemaker, P. Sullivan, G. Huapaya, and B. J. Van Doormaal. 2007. Short Communication: Modification of Genetic Evaluation of Herd Life from a Three-Trait to a Five-Trait Model in Canadian Dairy Cattle. J. Dairy Sci. 90:2025–2028. doi:10.3168/jds.2006-719.

Short, T. H., and T. J. Lawlor. 1992. Genetic Parameters of Conformation Traits, Milk Yield, and Herd Life in Holsteins. J. Dairy Sci. 75:1987–1998. doi:10.3168/jds.S0022-0302(92)77958-2.

Sogstad, Å., T. Fjeldaas, and O. Østerås. 2005. Lameness and Claw Lesions of the Norwegian Red Dairy Cattle Housed in Free Stalls in Relation to Environment, Parity and Stage of Lactation. Acta Vet. Scand. 46:203–217. doi:10.1186/1751-0147-46-203.

Stock, K. f., H. Hamann, and O. Distl. 2005. Estimation of genetic parameters for the prevalence of osseous fragments in limb joints of Hanoverian Warmblood horses. J. Anim. Breed. Genet. 122:271–280. doi:10.1111/j.1439-0388.2005.00527.x.

Thomsen, P. T., and H. Houe. 2006. Dairy cow mortality. A review. Vet. Q. 28:122–129. doi:10.1080/01652176.2006.9695218.

Van Pelt, M. L., T. H. E. Meuwissen, G. de Jong, and R. F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130. doi:10.3168/jds.2014-9090.

Veerkamp, R. F., S. Brotherstone, B. Engel, and T. H. E. Meuwissen. 2001. Analysis of censored survival data using random regression models. Anim. Sci. 72:1–10.

Vergara, C. F., D. Döpfer, N. B. Cook, K. V. Nordlund, J. A. A. McArt, D. V. Nydam, and G. R. Oetzel. 2014. Risk factors for postpartum problems in dairy cows: Explanatory and predictive modeling. J. Dairy Sci. 97:4127–4140. doi:10.3168/jds.2012-6440.

Vinson, W. E., J. M. White, and R. H. Kliewer. 1976. Overall Classification as a Selection Criterion for Improving Categorically Scored Components of Type in Holsteins. J. Dairy Sci. 59:2104–2114. doi:10.3168/jds.S0022-0302(76)84494-3.

Visscher, P. M., and M. E. Goddard. 1995. Genetic Parameters for Milk Yield, Survival, Workability, and Type Traits for Australian Dairy Cattle. J. Dairy Sci. 78:205–220. doi:10.3168/jds.S0022-0302(95)76630-9.

Visscher, P., R. Thompson, H. Yazdi, G. H. Hill, and S. Brotherstone. 1999. Genetic analysis of longevity data in the UK: present practice and considerations for the future. Interbull Bull. No 21:16–22.

vit. 2015. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed Jun. 26, 2015. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf.

Zwald, N. R., K. A. Weigel, Y. M. Chang, R. D. Welper, and J. S. Clay. 2004. Genetic Selection for Health Traits Using Producer-Recorded Data. I. Incidence Rates, Heritability Estimates, and Sire Breeding Values. J. Dairy Sci. 87:4287–4294. doi:10.3168/jds.S0022-0302(04)73573-0.

Appendix

Supplementary Table 2.1: Genetic correlation estimates of model A. With means, standard deviations (sd) and numbers of contributing runs (N).

Trait	Variable	A1.2	A1.3	A1.4	A1.5	A1.6	A2.1	A2.2	A2.3	A2.4	A2.5	A2.6	A3.1	A3.2	A3.3	A3.4	A3.5	A3.6
A1.1	**mean**	0.89	0.82	0.76	0.61	0.53	0.82	0.71	0.56	0.52	0.56	0.38	0.72	0.56	0.60	0.56	0.39	0.37
	sd	0.04	0.05	0.07	0.10	0.08	0.06	0.06	0.12	0.12	0.13	0.12	0.06	0.16	0.11	0.15	0.14	0.13
	N	37	8	8	8	8	9	9	10	10	10	10	9	9	10	10	10	10
A1.2	**mean**		0.94	0.91	0.79	0.60	0.80	0.82	0.72	0.70	0.66	0.43	0.70	0.71	0.71	0.70	0.50	0.41
	sd		0.03	0.05	0.09	0.08	0.05	0.05	0.14	0.10	0.14	0.15	0.07	0.10	0.12	0.15	0.13	0.12
	N		8	8	8	8	9	9	10	10	10	10	9	9	10	10	10	10
A1.3	**mean**			0.93	0.90	0.65	0.80	0.86	0.81	0.79	0.77	0.52	0.64	0.79	0.79	0.77	0.78	0.50
	sd			0.07	0.04	0.10	0.05	0.09	0.13	0.13	0.16	0.17	0.06	0.08	0.10	0.12	0.10	0.11
	N			31	8	8	8	8	8	8	7	7	8	8	8	8	7	7
A1.4	**mean**				0.91	0.64	0.75	0.84	0.82	0.82	0.78	0.51	0.59	0.77	0.77	0.78	0.70	0.51
	sd				0.06	0.12	0.14	0.11	0.11	0.09	0.06	0.20	0.17	0.14	0.17	0.16	0.13	0.18
	N				8	8	8	8	8	8	7	7	8	8	8	8	7	7
A1.5	**mean**					0.70	0.68	0.65	0.81	0.80	0.80	0.65	0.53	0.66	0.80	0.81	0.67	0.65
	sd					0.09	0.06	0.09	0.07	0.10	0.09	0.15	0.07	0.10	0.06	0.11	0.08	0.14
	N					31	6	6	9	9	8	8	6	6	9	9	8	8
A1.6	**mean**						0.61	0.49	0.57	0.55	0.67	0.92	0.50	0.56	0.48	0.49	0.55	0.87
	sd						0.08	0.10	0.08	0.15	0.11	0.06	0.03	0.09	0.11	0.06	0.12	0.05
	N						6	6	9	9	8	8	6	6	9	9	8	8
A2.1	**mean**							0.84	0.72	0.68	0.68	0.59	0.96	0.81	0.75	0.68	0.57	0.61
	sd							0.05	0.07	0.09	0.12	0.07	0.02	0.07	0.08	0.13	0.10	0.07
	N							42	10	10	10	10	23	23	9	9	9	9
A2.2	**mean**								0.95	0.90	0.83	0.57	0.78	0.94	0.90	0.87	0.77	0.59
	sd								0.03	0.07	0.11	0.08	0.05	0.04	0.10	0.07	0.11	0.12
	N								10	10	10	10	23	23	9	9	9	9
A2.3	**mean**									0.93	0.87	0.60	0.61	0.89	0.90	0.91	0.87	0.60

Trait	Variable	A1.2	A1.3	A1.4	A1.5	A1.6	A2.1	A2.2	A2.3	A2.4	A2.5	A2.6	A3.1	A3.2	A3.3	A3.4	A3.5	A3.6
	sd									0.06	0.08	0.07	0.13	0.07	0.07	0.06	0.08	0.14
	N									47	10	10	10	10	27	27	10	10
A2.4	**mean**										0.92	0.62	0.63	0.88	0.88	0.88	0.86	0.67
	sd										0.07	0.10	0.11	0.08	0.05	0.09	0.11	0.11
	N										10	10	10	10	27	27	10	10
A2.5	**mean**											0.69	0.58	0.80	0.75	0.86	0.91	0.68
	sd											0.10	0.13	0.10	0.13	0.11	0.06	0.07
	N											45	10	10	10	10	25	25
A2.6	**mean**												0.47	0.60	0.53	0.70	0.58	0.95
	sd												0.07	0.15	0.13	0.11	0.14	0.03
	N												10	10	10	10	25	25
A3.1	**mean**													0.73	0.64	0.59	0.55	0.53
	sd													0.07	0.10	0.12	0.10	0.06
	N													53	20	20	20	20
A3.2	**mean**														0.90	0.85	0.76	0.68
	sd														0.11	0.08	0.15	0.13
	N														20	20	20	20
A3.3	**mean**															0.93	0.80	0.66
	sd															0.06	0.11	0.12
	N															56	19	19
A3.4	**mean**																0.90	0.64
	sd																0.06	0.10
	N																19	19
A3.5	**mean**																	0.61
	sd																	0.16
	N																	54

Supplementary Table 2.2: Genetic correlation estimates of model B. With means, standard deviations (sd) and numbers of contributing runs (N).

Trait	Variable	B1.2	B1.3	B2.1	B2.2	B2.3	B3.1	B3.2	B3.3
B1.1	**mean**	0.81	0.52	0.82	0.63	0.40	0.69	0.57	0.43
	sd	0.05	0.09	0.06	0.06	0.09	0.08	0.12	0.08
	N	20	20	10	10	10	10	10	10
B1.2	**mean**		0.67	0.77	0.85	0.60	0.64	0.75	0.59
	sd		0.08	0.04	0.06	0.09	0.09	0.09	0.10
	N		20	10	10	10	10	10	10
B1.3	**mean**			0.58	0.61	0.93	0.48	0.55	0.84
	sd			0.07	0.06	0.06	0.05	0.06	0.06
	N			10	10	10	10	10	10
B2.1	**mean**				0.75	0.54	0.96	0.74	0.61
	sd				0.06	0.08	0.02	0.07	0.07
	N				20	20	10	10	10
B2.2	**mean**					0.62	0.69	0.96	0.67
	sd					0.07	0.05	0.03	0.06
	N					20	10	10	10
B2.3	**mean**						0.49	0.64	0.95
	sd						0.06	0.08	0.03
	N						10	10	10
B3.1	**mean**							0.67	0.52
	sd							0.05	0.07
	N							20	20
B3.2	**mean**								0.68
	sd								0.06
	N								20

Supplementary Table 2.3: Correlations of estimated sire breeding values (EBV) for survival traits (model B) to EBV for production and functional traits. Shown are the numbers of sires (N) with >50 daughters on the data set and a reliability of EBV of ≥0.9 for routinely estimated EBV.

Trait	**N**	**B1.1**	**B1.2**	**B1.3**	**B2.1**	**B2.2**	**B2.3**	**B3.1**	**B3.2**	**B3.3**
Functional Longevity	1468	0.67	0.72	0.74	0.76	0.77	0.77	0.70	0.78	0.79
First to last insemination	484	0.04	0.04	0.44	0.10	0.04	0.46	0.08	0.07	0.43
Stillbirth	240	0.28	0.31	0.40	0.27	0.28	0.42	0.21	0.27	0.42
Somatic cell score	4231	0.28	0.32	0.27	0.32	0.41	0.30	0.31	0.43	0.32
Milk production	8743	0.25	0.30	0.12	0.23	0.27	0.09	0.21	0.23	0.11

CHAPTER 3:

Phenotypic and genetic relationships between age at first calving, its component traits, and survival of heifers up to second calving

Johannes Heise*, Kathrin F. Stock†, Friedrich Reinhardt†, Ngoc-Thuy Ha*, Henner Simianer*

*University of Goettingen, Animal Breeding and Genetics, 37075 Göttingen, Germany

†IT Solutions for Animal Production (vit), Heinrich-Schröder-Weg 1, 27283 Verden, Germany

To be published in *Journal of Dairy Science* (in press)

Abstract

The aim of this study was to answer the question whether models for genetic evaluations of longevity should include a correction for age at first calving (**AFC**) or not. For this purpose, phenotypic and genetic relationships between AFC, its component traits age at first insemination (**AFI**) and interval from first to last insemination (**FLI**) on the one hand and survival of different periods of the first lactation (**S1:** 0 to 49 d, **S2:** 50 to 249 d, **S3:** 250 d to the 2nd calving) on the other hand were investigated. Data of 721,919 German Holstein heifers, being inseminated for the first time during the years 2003 to 2012, were used for the analyses. Phenotypic correlations of AFI, FLI and AFC to S1 to S3 were negative. Mean estimated heritabilities were 0.239 (AFI), 0.007 (FLI), and 0.103 (AFC) and 0.023 (S1), 0.016 (S2), and 0.028 (S3) on the observed scale. The genetic correlation between AFI and FLI was close to zero. Genetic correlations between AFI and the survival traits were -0.08 (S1), -0.02 (S2) and -0.10 (S3), between FLI and the survival traits -0.14 (S1), -0.20 (S2) and -0.44 (S3) and between AFC and the survival traits -0.09 (S1), -0.06 (S2) and -0.20 (S3). Some of these genetic correlations were different from zero, which suggests that correcting for AFC in genetic evaluations for longevity in dairy cows might remove functional genetic variance and should be reconsidered.

Key words

Age at first calving, longevity, animal breeding, dairy cattle

Introduction

Longevity is an economically important (Allaire and Gibson, 1992) and publicly discussed trait in dairy cows. Routine genetic evaluations for this trait are conducted in all major dairy breeding countries (Miglior et al., 2005; Interbull, 2016). After years of routine genetic evaluations for longevity, some recent studies have reviewed this trait complex (van Pelt et al., 2015; Sasaki et al., 2015; Heise et al., 2016). In scientific studies (Ducrocq, 2005; Sewalem et al., 2007; van Pelt et al., 2015; Sasaki et al., 2015) as well as in routine genetic evaluations in many countries (Interbull, 2016), longevity is corrected for age at first calving, either in form of a covariate (Sewalem et al., 2007) or as a fixed class effect (Ducrocq, 2005). In the following countries, which participate in Interbull genetic evaluations, longevity is corrected for age at first calving in either form: Canada, Czech Republic, Denmark/Finland/Sweden, France, Germany, Great Britain, Hungary, Ireland, Israel, Italy, Netherlands, Poland, Republic of South Africa, Slovenia, Spain, and Switzerland. In Australia, Belgium, and New Zealand, and the United States, longevity is not corrected for age at first calving (Interbull 2016). In genetic evaluations this is only justifiable if age at first calving is predominantly reflecting environmental factors with no genetic correlation to longevity. There are reasons to reconsider this hypothesis: age at first calving can be dissected into age at first insemination, time from first to successful insemination and gestation length. All

these traits are functional traits and were shown to be heritable in previous studies (age at first insemination: e.g., Mäntysaari et al., 2002; Jamrozik et al., 2005; interval first to last insemination: e.g., Berry et al., 2003; Liu et al., 2008; gestation length: e.g., Jamrozik et al., 2005; Norman et al., 2009). Especially the interval from first to last (or successful) insemination is a widely used reproduction trait and part of the total merit index in various countries, including Germany (Interbull, 2016). Phenotypically, impaired reproductive performance increases the risk of culling (Rajala-Schultz and Gröhn, 1999), and positive genetic correlations between more favorable expressions of reproduction traits and survival were also reported (Campos et al., 1994; Haile-Mariam et al., 2003). Despite the mentioned results, literature on genetic relationships between age at first calving or age at first insemination and survival is scarce. Furthermore, most previous studies on genetic relationships between survival and other traits considered survival of relatively long periods, e.g., only survival of the complete first lactation (Visscher and Goddard, 1995; Dematawewa and Berger, 1998; Haile-Mariam et al., 2003). Recent studies (van Pelt et al., 2015; Sasaki et al., 2015; Heise et al., 2016) suggest that the genetic background for survival of different periods within the same lactation is not the same. The aim of this study was to investigate the genetic relationships between age at first calving, its underlying traits and survival of different periods of the first lactation. Knowledge about these relationships is crucial for appropriate modeling of longevity in routine genetic evaluations.

Materials and Methods

Data

The data used in this study originate from national routine genetic evaluations for Holstein cattle in Germany (vit, 2016) and were selected by applying the selection steps described in Table 3.1. The traits age at first insemination (**AFI**) and interval from first to last insemination (**FLI**) were restricted to plausible ranges: 330 d to 800 d for AFI and 0 to 210 d for FLI. Additionally, the interval from last insemination to first calving was required to be within the range from 265 to 295 d. The last step included sampling of herds within 5 different federal states such that at least 100,000 records were in each data set. This procedure resulted in 5 sample data sets consisting of in total 721,919 cows (111,388 to 174,102 each) from 10,643 sires.

Table 3.1: Data selection steps and criteria.

Selection of	Criterion
1) Insemination bulls	a) Holsteins only
	b) Semen used in at least 10 herds
2) Insemination records	a) Bulls from step 1)
	b) Conventional (non-sexed) frozen semen only
3) Cows	a) Holsteins only
	b) Both parents known
	c) Stayed in the same herd until first calving
	d) All inseminations on the cow passed step 2)
	e) Known record of first calving
	f) Record of first insemination within the years 2003 to 2012
4) Herds	a) At least 20 first inseminations (from step 3) on heifers per year from 2003 to 2015
	b) Sampling of herds within 5 of the German federal states
5) Sires	a) Per sample: more than 5 daughters per sire

Trait definitions

The individual records for AFI, FLI and age at first calving (**AFC**) were derived from the available data of the sampled animals. Following Heise et al. (2016), 3 survival traits within the first lactation were defined as follows: the first parity was divided in the periods 0 to 49 d (**S1**), 50 to 249 d (**S2**) and 250 d from calving until the second calving (**S3**). If a cow survived until the end of a period, given she had survived all previous periods, her observation for survival was coded '1', if she was culled in the period of consideration, her observation for survival was coded '0', and if the cow was sold or culled in one of the previous periods, her observation record of survival for the regarded period was considered to be missing.

Phenotypic analyses

Survival rates and risk of culling for S1 to S3 were estimated following Kaplan and Meier (1958) with the survfit() function from the survival package (Therneau, 2016) in R (R Core Team, 2016).

To evaluate phenotypic relationships between AFI, FLI and AFC and the survival traits S1 to S3, the traits AFI, FLI and AFC were coded as discrete variables (monthly steps for AFI and AFC, ranging from 12 to 21 months and from 21 to 32 months, respectively; 21 d steps for FLI, ranging from 0 to 147 d). Univariate logit threshold models were fitted as

$$\mathbf{y}^* = \mathbf{Xb}$$

where $\mathbf{y}^*$ is a latent variable which is linked to **y** via a logit-function; **y** is a vector, comprising of survival observations (1/0) for S1, S2 or S3, **X** is an incidence matrix, linking observations to classes of fixed effects and **b** is a vector of fixed effects (one of AFI, FLI or AFC). This resulted in 9 different models with survival traits as dependent and AFI, FLI and AFC as independent variables. The analysis was conducted using the function glm() from the R package stats (R Core Team, 2016).

Genetic analyses

Genetic parameters were estimated from the following linear multiple trait sire model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e}$$

where **y** is a vector of observations, **X** is an incidence matrix, linking the observations to the fixed effects, **b** is the vector of fixed effects, i.e., herd × year × season at first insemination, **Z** is the incidence matrix of the random sire effects, **s** is the vector of random sire effects ($\mathbf{s} \sim N(0, \mathbf{G_s} \otimes \mathbf{A})$, with the covariance matrix of sire effects $\mathbf{G_s}$ and the numerator relationship matrix for sires **A**), and **e** is a vector of random residual effects ($\mathbf{e} \sim N(0, \mathbf{R_s} \otimes \mathbf{I})$, with the residual covariance matrix $\mathbf{R_s}$, including the variances of dams' genetic effects and mendelian sampling). The genetic covariance matrix for animals, $\mathbf{G_0}$, was then estimated as $4\mathbf{G_s}$ and the respective residual covariance matrix $\mathbf{R_0}$ as $\mathbf{R_s} - 3\mathbf{G_s}$.

Variance components were estimated for different trait combinations using the above model and the software package VCE (Groeneveld et al., 2010). Two different trait combinations were run: 1) including S1 to S3, AFI and FLI, and 2) including S1 to S3 and AFC. Due to the dependency of AFC on AFI and FLI, models including these 3 traits simultaneously did not converge. Joint results from the different samples for genetic parameter estimates are presented as means with standard deviations.

Results

Phenotypic relationships between AFI, FLI, AFC and survival

Descriptive statistics of AFI, FLI and AFC are given in Table 3.2. Means over all samples for AFI, FLI and AFC were 16.23 months, 17.94 d and 25.96 months, respectively. Kaplan-Meier estimators for risk of culling and survival rates can be seen from Table 3.3. A proportion of 93, 86 and 77 % of cows with a reported first calving survived until the end of periods S1, S2, and S3, respectively.

Table 3.2: Descriptive statistics for the traits AFI, FLI and AFC as well as for the interval from last insemination to calving (all samples).

Description	Unit	Mean	Median	Min	Max
AFI	months	16.23	15.87	10.84	26.29
FLI	days	17.94	0	0	210
Interval last ins. to first calving	days	278.52	279	265	295
AFC	months	25.97	25.57	19.65	41.34

AFI: age at first insemination; FLI: interval from first to last insemination; AFC: age at first calving

Table 3.3: Kaplan-Meier estimators for the risk of culling, conditional on survival of the previous periods, and respective survival rates with standard errors (SE).

		Survival	
Trait	Risk of culling	Rate	SE
S1	.073	.927	.0003
S2	.078	.855	.0004
S3	.103	.766	.0005

S1 to S3: survival of the periods 0 to 49 d, 50 to 249 d and 250 d to the second calving.

Predicted survival probabilities from the logit threshold model for S1 to S3, dependent on either AFI, FLI or AFC as fixed effects, can be seen from Figure 3.1. The probabilities for S2 and S3 are conditional on survival of the preceding periods. Predicted survival probabilities for S3, affected by AFI, declined from 0.91 (12 months) to 0.86 (≥21 months), from 0.90 (0 to 20 d) to 0.85 (≥147 d) for the effect of FLI, and from 0.91 (22 months) to 0.84 (≥32 months) for AFC. Effects of AFI, FLI and AFC on S1 and S2 showed the same trend, i.e. cows with longer intervals had reduced survival probabilities.

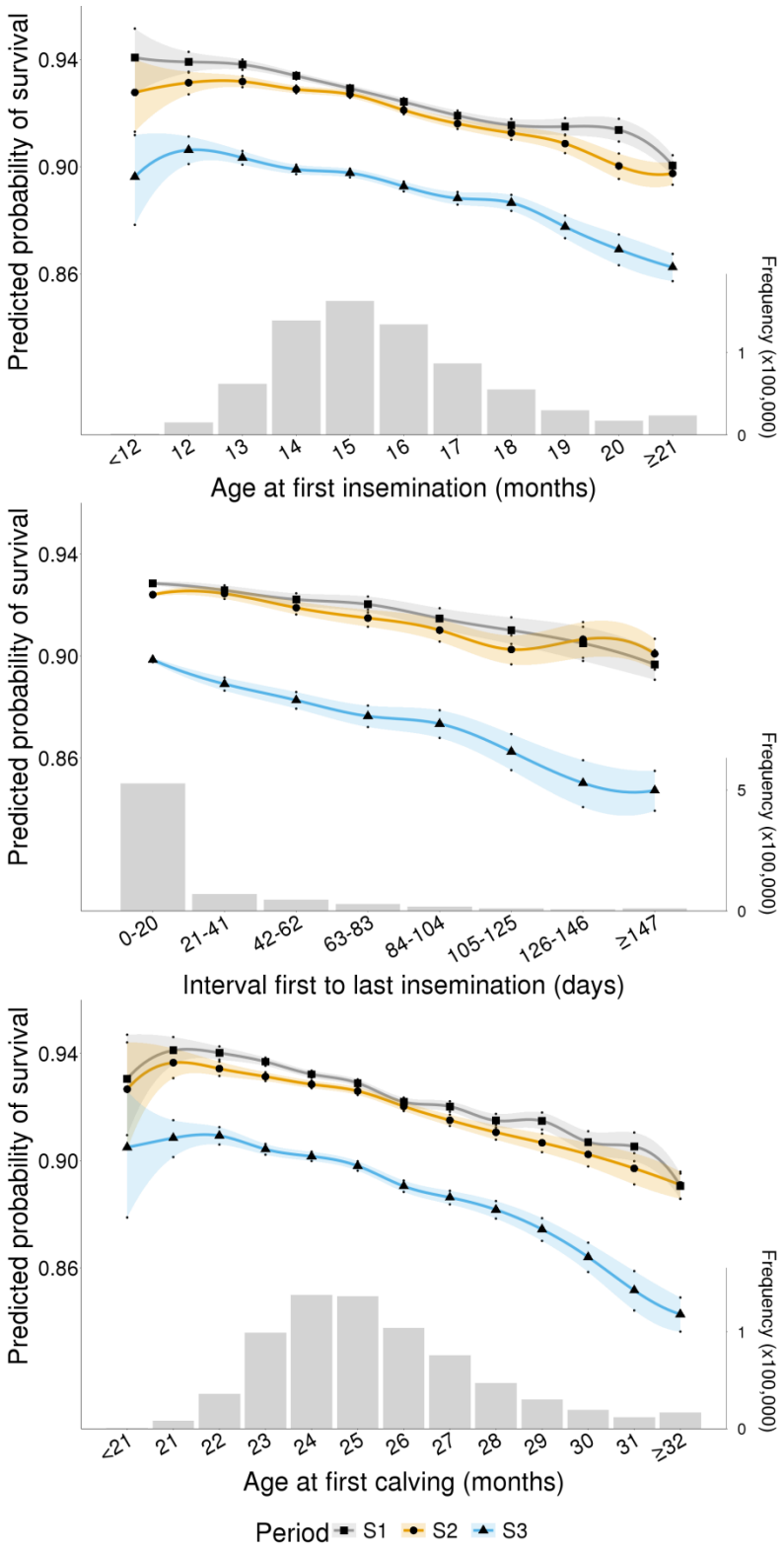


Figure 3.1: Predicted probabilities for survival (left y-axis) of different periods of the first lactation (S1 to S3) from a logit model, dependent on age at first insemination, interval first to last insemination and age at first calving as fixed effects. Ribbons indicate pointwise 95% confidence intervals. Grey bars (right y-axis) represent absolute frequencies for the respective classes.

Genetic parameters

Estimated genetic and residual relationships between AFI, FLI and AFC and the 3 survival traits are given in Table 3.4. For AFI, a mean heritability of 0.24 was estimated. Genetic correlations to the survival traits S1 and S3 were slightly negative (-0.08 and -0.10) and close to zero for S2. For FLI, a mean heritability of 0.007 was estimated. Genetic correlations between FLI and S1 (-0.14) and S2 (-0.20) were weaker than between FLI and S3 (-0.44). The mean heritability estimate for AFC was 0.10. Mean genetic correlations to survival traits were -0.09 (S1), -0.06 (S2) and -0.20 (S3).

Table 3.4: Mean estimated genetic parameters from 5 samples between survival traits S1 to S3, AFI, FLI and AFC. Heritability estimates on diagonal, genetic correlations above and residual correlations below diagonal. Standard deviations are given in parentheses.

Trait	S1	S2	S3	AFI	FLI	AFC
S1	.023	.86	.48	-.08	-.14	-.09
	(.002)	(.08)	(.06)	(.15)	(.10)	(.12)
S2	--†	.016	.58	-.02	-.20	-.06
		(.004)	(.11)	(.07)	(.06)	(.09)
S3	--†	--†	.028	-.10	-.44	-.20
			(.003)	(.09)	(.17)	(.12)
AFI	-.02	-.03	-.02	.239	.12	--*
	(.01)	(.00)	(.01)	(.069)	(.18)	
FLI	-.02	-.02	-.03	-.02	.007	--*
	(.01)	(.00)	(.00)	(.00)	(.002)	
AFC	-.03	-.03	-.04	--*	--*	0.103
	(.01)	(.00)	(.01)			(0.025)

S1 to S3: survival of the periods 0 to 49 d, 50 to 249 d and 250 d to the second calving; AFI: age at first insemination; FLI: interval from first to last insemination; AFC: age at first calving.

* Not estimated.

† Not defined, because all individuals that had observations for S2 or S3 were coded 1 for all previous survival traits.

Discussion

Data

For this study, relatively large herds were chosen compared to the German average herd size. This choice had three advantages: firstly, cows in bigger herds can be expected to be all treated in a similar way, with as few preferential treatments as possible. Secondly, the relatively large herds used in our analysis allowed for the inclusion of a fixed herd × year × season-effect. This is favorable, because a pre-analysis showed a seasonal pattern for all traits. And thirdly, herd sizes in Germany show an increasing trend (Destatis, 2013), so that our results can be considered

as a reliable basis for future decisions concerning genetic evaluations of the studied trait complex.

Model choice

Although survival traits were considered as binary traits and the other variables, AFI, FLI and AFC were non-normally distributed, a linear sire model was chosen for the estimation of variance components. This choice was mainly due to the fact that the new routine genetic evaluation for longevity of Holsteins in Germany will probably be based on a linear model, including the survival traits of the first lactation as considered here. The primary motivation of the current study was to investigate the relationship of AFC and S1 to S3 for this scenario. The choice of a linear model for the development of a new routine genetic evaluation of longevity was mainly due to the heavily increased computation time for survival and threshold models compared to linear models. Boettcher et al. (1999) reported this increase to be by a factor of 5 to 10, which would not be feasible for a multiple trait animal model in the German national genetic evaluation system. Furthermore, other studies revealed good performance of linear models for genetic evaluations of binary survival traits (Boettcher et al., 1999; Holtsmark et al., 2009). The choice of a linear model for binary variables influences the scale of h^2 (Dempster and Lerner, 1950), but it does not have a scale effect on estimates of genetic correlations (Vinson et al., 1976; Gianola, 1982). With regard to the estimation of variance components of the other non-normally distributed variables (AFI, FLI and AFC), Raheja et al. (1989) found no benefit from transforming AFI, age at last insemination or number of inseminations to approximate normal distributions.

Phenotypic relationships

The sum of AFI and FLI explained 99.6 % of the variance of AFC. This implies that gestation length only explains a minor proportion of the variance of AFC, which is in good accordance with results of Jamrozik et al. (2005).

With regard to the survival traits, Kaplan-Meier estimates of proportions of cows that survived the defined segments of the first lactation were similar to a previous study (Heise et al., 2016) in a similar data set.

The onset of puberty is known to be much stronger influenced by the body weight development than by age (as reviewed by Sejrsen and Purup, 1997), suggesting that AFI primarily reflects the growth development of heifers. Slow growing heifers were reported to have a lower probability of survival up to the second lactation than faster growing heifers (Bach, 2011), which is in accordance with our results.

The effect of FLI was somewhat stronger on S3 than on S1 and S2. This is in accordance with results from studies on disposal reasons, reporting that culling for infertility/reproduction related causes occurs mainly towards the end of the lactation (Pinedo et al., 2010; Heise et al., 2016). The effect of AFC was also somewhat stronger on S3 than on S1 and S2. Among the cows being

culled before the second calving, the probability to be culled for 'infertility' increased from 0.19 to 0.22 when the interval FLI increased from the 0 to 20 d class to the ≥147 d class. This supports that a cow, which was difficult to get pregnant as a heifer, has a slightly increased risk to be culled for reproduction problems during the first lactation. These results are in accordance with Bach (2011), who found that cows needing more than one service to conceive as a heifer had a reduced probability to finish the first lactation. For example, the probability to finish the first lactation for cows needing 2 services to conceive as a heifer was 0.74 times the probability of cows needing only 1 service.

The effect of AFC on survival traits is in general agreement with the literature (Gill and Allaire, 1976; Nilforooshan and Edriss, 2004; Ducrocq, 2005; Bach, 2011), although some studies found no effect of AFC on culling up to 310 d from calving (Ettema and Santos, 2004) and survival up to the second calving (Simerl et al., 1992).

Heritability estimates

Heritability estimates among S1 to S3 were in good agreement with results from an earlier study in a similar, partly overlapping, data set (Heise et al., 2016).

The range of heritability estimates for AFI from literature is large with 0.24 in Dutch Friesians (Jansen et al., 1987), 0.22 in Ayrshire and Friesian cows (Mäntysaari et al., 2002), 0.12 in Canadian Holsteins (Raheja et al., 1989), 0.10 to 0.64 for Holstein cows in different herds in Germany (Bergk, 2011), and 0.13 in Canadian Holstein heifers (Jamrozik et al., 2005). For beef cattle, heritability estimates for age at puberty range from 0.10 to 0.67 (as reviewed by Cammack et al., 2009). Our heritability estimate (0.24) for AFI lies well in the previously reported range, but still is relatively high compared to most other functional traits. A reason might be that farmers focus mainly on body weight when determining the optimum point for the first insemination, as proposed by, e.g., Archbold et al. (2012) since body weight is a better indicator for the onset of puberty than age (as reviewed by Sejrsen and Purup, 1997), and heritability estimates for body weight at different times during the rearing period were reported to be high, ranging from 0.41 to 0.83 (Groen and Vos, 1995; Coffey et al., 2006).

The mean heritability estimate for FLI (0.007) was somewhat lower than estimated in a previous study on German Holstein heifers (Liu et al., 2008) and at the lower end of estimates for similar traits reported in other studies (Hansen et al., 1983; Berry et al., 2003).

A wide range of heritability estimates was also reported for AFC, ranging from 0.02 (VanRaden and Klaaskate, 1993) to 0.47 (Ruiz-Sánchez et al., 2007). Our results (0.10) lie towards the lower bound of this range and well below the heritability estimate for AFI (0.24). This is also in accordance with findings of Mäntysaari et al. (2002), who reported heritability estimates of 0.22 for age at breeding, and 0.05 for age at first calving.

Genetic correlations

Mean genetic correlations of AFI with S2 were 0 and were slightly negative with S1 and S3, but estimates varied substantially between the different samples. For example, the estimates for the genetic correlation between AFI and S1 ranged from -0.29 to 0.04 across the five samples. The slightly negative correlations were in agreement with the phenotypic results from our study. This suggests that heifers with a higher genetic potential for early breeding also have a slightly higher genetic potential for survival of the first lactation. Genetic correlations between FLI and the different survival traits showed a clear pattern with smaller negative correlations to S1 and S2 and a much stronger negative correlation to S3. As for the discussed phenotypic patterns, this might result from the implied influence of reproductive performance on S3. This result must be taken cautiously due to the extremely low heritability estimate for FLI (0.007), which was close to the lower bound of the parameter space. However, this pattern was consistent over all samples and repeated in a weaker form for the genetic correlations between AFC and the respective survival traits, which might result from the fact that FLI is a component of AFC.

The close to zero estimate for the genetic correlation between AFI and FLI is in accordance with other studies on similar traits (Raheja et al., 1989; Jamrozik et al., 2005).

In our study, we concentrated on the definition of AFC as a heifer development and fertility trait and its relationship to survival traits. Previous studies have shown that AFC affects milk yield as well (e.g., Moore et al., 1991). Therefore, the interrelationship between milk yield, AFC and survival needs to be further investigated, especially when considering models for functional longevity, which means the inclusion of an effect for milk yield relative to the herd mean in order to correct for voluntary culling (e.g., Ducrocq et al., 1988).

Comparison of models with and without correction for AFC

In order to further evaluate the impact of the correction for AFC in the sample data sets used for our study, we also estimated variance components for S1, S2 and S3, including a first order regression on AFC. The estimated variance components were very similar to those without the correction, indicating that the effect of AFC explains only a minor proportion of the variability of survival. Furthermore, we compared estimated breeding values (**EBV**s) for the respective survival traits of the first lactation from the prototype version of the new genetic evaluation system for longevity between scenarios with and without correction for AFC, where AFC was considered as a class variable with levels changing weekly. Additionally, a herd × year × season effect, and a fixed effect for the region (federal states) were included. Genetic parameters were similar to those in Heise et al. (2016). Correlations between EBVs for bulls born later than 1995 and having at least 50 daughters in the data set (N=19,972) were 0.99 (S1), 0.99 (S2) and 0.98 (S3), indicating some re-ranking of bulls between the two scenarios. When comparing the top 20 bulls for S3 between both scenarios, only 11 bulls are in common, illustrating massive re-ranking of the top lists.

Implications

The aim of genetic evaluations for longevity is to assess the genetic potential of animals to resist culling, either involuntary culling only which is referred to as functional longevity (Ducrocq et al., 1988) or involuntary and voluntary culling at the same time (van Pelt et al., 2015). Both approaches include the genetic variance of longevity, contributed by functional traits. Our results indicate that AFC has low, but potentially non-zero genetic correlations to survival. Accordingly, correcting for AFC in models for genetic evaluations of survival/longevity might remove functional genetic variance from survival traits. Two other approaches could be examined to avoid this issue while still using information on AFC and its underlying traits in genetic evaluations for functional longevity: firstly, underlying traits of AFC could potentially be considered as genetically correlated traits in multivariate approaches to increase the reliability of breeding values for survival. Secondly, AFC could be corrected for its genetic component and then be used as a covariate or fixed effect in survival models, if a correction for the environmental/management part of this effect is deemed necessary.

Conclusions

Based on the parameters estimated in this study, correcting for AFC in models for genetic evaluations of survival/longevity is expected to remove functional genetic variance from survival traits and should thus be reconsidered as this might remove a part of the genetic determination of the target trait.

Acknowledgments

The German national FBF (Förderverein Bioökonomieforschung e.V., “Association for Bioeconomy Research”, Bonn, Germany) is thanked for the financial support.

References

Allaire, F.R., and J.P. Gibson. 1992. Genetic Value of Herd Life Adjusted for Milk Production. J. Dairy Sci. 75:1349–1356. doi:10.3168/jds.S0022-0302(92)77886-2.

Archbold, H., L. Shalloo, E. Kennedy, K.M. Pierce, and F. Buckley. 2012. Influence of age, body weight and body condition score before mating start date on the pubertal rate of maiden Holstein–Friesian heifers and implications for subsequent cow performance and profitability. Animal 6:1143–1151. doi:10.1017/S1751731111002692.

Bach, A. 2011. Associations between several aspects of heifer development and dairy cow survivability to second lactation. J. Dairy Sci. 94:1052–1057. doi:10.3168/jds.2010-3633.

Bergk, N. 2011. Genetisch-statistische Untersuchungen zum Wachstum weiblicher Jungrinder der Rasse Holstein auf Basis von zum Zeitpunkt der ersten Besamung erhobenen Gewichten. Martin-Luther-Universität, Halle-Wittenberg.

Berry, D.P., F. Buckley, P. Dillon, R.D. Evans, M. Rath, and R.F. Veerkamp. 2003. Genetic Relationships among Body Condition Score, Body Weight, Milk Yield, and Fertility in Dairy Cows. J. Dairy Sci. 86:2193–2204. doi:10.3168/jds.S0022-0302(03)73809-0.

Boettcher, P.J., L.K. Jairath, and J.C.M. Dekkers. 1999. Comparison of Methods for Genetic Evaluation of Sires for Survival of Their Daughters in the First Three Lactations. J. Dairy Sci. 82:1034–1044. doi:10.3168/jds.S0022-0302(99)75324-5.

Cammack, K.M., M.G. Thomas, and R.M. Enns. 2009. Review: Reproductive Traits and Their Heritabilities in Beef Cattle. Prof. Anim. Sci. 25:517–528.

Campos, M.S., C.J. Wilcox, C.M. Becerril, and A. Diz. 1994. Genetic Parameters for Yield and Reproductive Traits of Holstein and Jersey Cattle in Florida. J. Dairy Sci. 77:867–873. doi:10.3168/jds.S0022-0302(94)77021-1.

Coffey, M.P., J. Hickey, and S. Brotherstone. 2006. Genetic Aspects of Growth of Holstein-Friesian Dairy Cows from Birth to Maturity. J. Dairy Sci. 89:322–329. doi:10.3168/jds.S0022-0302(06)72097-5.

Dematawewa, C.M.B., and P.J. Berger. 1998. Genetic and Phenotypic Parameters for 305-Day Yield, Fertility, and Survival in Holsteins. J. Dairy Sci. 81:2700–2709. doi:10.3168/jds.S0022-0302(98)75827-8.

Dempster, E.R., and I.M. Lerner. 1950. Heritability of Threshold Characters. Genetics 35:212–236.

Destatis. 2013. Forst, Landwirtschaft Und Fischerei - Viehbestand. Accessed December 16, 2016. https://www.destatis.de/DE/Publikationen/Thematisch/LandForstwirtschaft/ViehbestandTierischeErzeugung/Viehbestand2030410135314.pdf?__blob=publicationFile.

Ducrocq, V. 2005. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. Anim. Sci. 80:249–256. doi:10.1079/ASC41720249.

Ducrocq, V., R.L. Quaas, E.J. Pollak, and G. Casella. 1988. Length of Productive Life of Dairy Cows. 1. Justification of a Weibull Model. J. Dairy Sci. 71:3061–3070. doi:10.3168/jds.S0022-0302(88)79906-3.

Ettema, J.F., and J.E.P. Santos. 2004. Impact of Age at Calving on Lactation, Reproduction, Health, and Income in First-Parity Holsteins on Commercial Farms. J. Dairy Sci. 87:2730–2742. doi:10.3168/jds.S0022-0302(04)73400-1.

Gianola, D. 1982. Theory and Analysis of Threshold Characters. J. Anim. Sci. 54:1079–1096. doi:10.2134/jas1982.5451079x.

Gill, G.S., and F.R. Allaire. 1976. Relationship of Age at First Calving, Days Open, Days Dry, and Herdlife to a Profit function for Dairy Cattle. J. Dairy Sci. 59:1131–1139. doi:10.3168/jds.S0022-0302(76)84333-0.

Groen, A.F., and H. Vos. 1995. Genetic parameters for body weight and growth in Dutch Black and White replacement stock. Livest. Prod. Sci. 41:201–206. doi:10.1016/0301-6226(94)00062-C.

Groeneveld, E., M. Kovac, and N. Mielenz. 2010. VCE User's Guide and Reference Manual Version 6.0. Institute of Farm Animal Genetics, Friedrich Loeffler Institute (FLI), D-31535 Neustadt, Germany.

Haile-Mariam, M., P.J. Bowman, and M.E. Goddard. 2003. Genetic and environmental relationship among calving interval, survival, persistency of milk yield and somatic cell count in dairy cattle. Livest. Prod. Sci. 80:189–200. doi:10.1016/S0301-6226(02)00188-4.

Hansen, L.B., A.E. Freeman, and P.J. Berger. 1983. Association of Heifer Fertility with Cow Fertility and Yield in Dairy Cattle. J. Dairy Sci. 66:306–314. doi:10.3168/jds.S0022-0302(83)81790-1.

Heise, J., Z. Liu, K.F. Stock, S. Rensing, F. Reinhardt, and H. Simianer. 2016. The genetic structure of longevity in dairy cows. J. Dairy Sci. 0. doi:10.3168/jds.2015-10163.

Holtsmark, M., B. Heringstad, and J. Ødegård. 2009. Predictive abilities of different statistical models for analysis of survival data in dairy cattle. J. Dairy Sci. 92:5730–5738. doi:10.3168/jds.2009-2132.

Interbull. 2016. National Genetic Evaluation Forms Provided by Countries. Accessed December 13, 2016. http://interbull.org/ib/geforms.

Jamrozik, J., J. Fatehi, G.J. Kistemaker, and L.R. Schaeffer. 2005. Estimates of Genetic Parameters for Canadian Holstein Female Reproduction Traits. J. Dairy Sci. 88:2199–2208. doi:10.3168/jds.S0022-0302(05)72895-2.

Jansen, J., J. van der Werf, and W. de Boer. 1987. Genetic relationships between fertility traits for dairy cows in different parities. Livest. Prod. Sci. 17:337–349. doi:10.1016/0301-6226(87)90081-9.

Kaplan, E.L., and P. Meier. 1958. Nonparametric Estimation from Incomplete Observations. J. Am. Stat. Assoc. 53:457–481. doi:10.1080/01621459.1958.10501452.

Liu, Z., J. Jaitner, F. Reinhardt, E. Pasman, S. Rensing, and R. Reents. 2008. Genetic Evaluation of Fertility Traits of Dairy Cattle Using a Multiple-Trait Animal Model. J. Dairy Sci. 91:4333–4343. doi:10.3168/jds.2008-1029.

Mäntysaari, P., M. Ojala, and E.A. Mäntysaari. 2002. Measures of before and after breeding daily gains of dairy replacement heifers and their relationship with first lactation milk production traits. Livest. Prod. Sci. 75:313–322. doi:10.1016/S0301-6226(01)00329-3.

Miglior, F., B.L. Muir, and B.J. Van Doormaal. 2005. Selection Indices in Holstein Cattle of Various Countries. J. Dairy Sci. 88:1255–1263. doi:10.3168/jds.S0022-0302(05)72792-2.

Moore, R.K., B.W. Kennedy, L.R. Schaeffer, and J.E. Moxley. 1991. Relationships Between Age and Body Weight at Calving and Production in First Lactation Ayrshires and Holsteins. J. Dairy Sci. 74:269-278. doi:10.3168/jds.S0022-0302(91)78170-8.

Nilforooshan, M.A., and M.A. Edriss. 2004. Effect of Age at First Calving on Some Productive and Longevity Traits in Iranian Holsteins of the Isfahan Province. J. Dairy Sci. 87:2130–2135. doi:10.3168/jds.S0022-0302(04)70032-6.

Norman, H.D., J.R. Wright, M.T. Kuhn, S.M. Hubbard, J.B. Cole, and P.M. VanRaden. 2009. Genetic and environmental factors that affect gestation length in dairy cattle. J. Dairy Sci. 92:2259–2269. doi:10.3168/jds.2007-0982.

van Pelt, M.L., T.H.E. Meuwissen, G. de Jong, and R.F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130. doi:10.3168/jds.2014-9090.

Pinedo, P.J., A. De Vries, and D.W. Webb. 2010. Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. J. Dairy Sci. 93:2250–2261. doi:10.3168/jds.2009-2572.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raheja, K.L., E.B. Burnside, and L.R. Schaeffer. 1989. Heifer Fertility and Its Relationship with Cow Fertility and Production Traits in Holstein Dairy Cattle. J. Dairy Sci. 72:2665–2669. doi:10.3168/jds.S0022-0302(89)79407-8.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999. Culling of dairy cows. Part III. Effects of diseases, pregnancy status and milk yield on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:295–309. doi:10.1016/S0167-5877(99)00047-1.

Ruiz-Sánchez, R., R.W. Blake, H.M.A. Castro-Gámez, F. Sánchez, H.H. Montaldo, and H. Castillo-Juárez. 2007. Short Communication: Changes in the Association Between Milk Yield and Age at First Calving in Holstein Cows with Herd Environment Level for Milk Yield. J. Dairy Sci. 90:4830–4834. doi:10.3168/jds.2007-0156.

Sasaki, O., M. Aihara, A. Nishiura, H. Takeda, and M. Satoh. 2015. Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. J. Dairy Sci. 98:5781–5795. doi:10.3168/jds.2014-9152.

Sejrsen, K., and S. Purup. 1997. Influence of prepubertal feeding level on milk yield potential of dairy heifers: a review. J. Anim. Sci. 75:828–835. doi:10.2527/1997.753828x.

Sewalem, A., F. Miglior, G.J. Kistemaker, P. Sullivan, G. Huapaya, and B.J. Van Doormaal. 2007. Short Communication: Modification of Genetic Evaluation of Herd Life from a Three-Trait to a Five-Trait Model in Canadian Dairy Cattle. J. Dairy Sci. 90:2025–2028. doi:10.3168/jds.2006-719.

Simerl, N.A., C.J. Wilcox, and W.W. Thatcher. 1992. Postpartum Performance of Dairy Heifers Freshening at Young Ages. J. Dairy Sci. 75:590–595. doi:10.3168/jds.S0022-0302(92)77796-0.

Therneau, E. 2016. A Package for Survival Analysis in S. Version 2.40. Accessed November 1, 2016. https://CRAN.R-project.org/package=survival.

VanRaden, P.M., and E.J.H. Klaaskate. 1993. Genetic Evaluation of Length of Productive Life Including Predicted Longevity of Live Cows. J. Dairy Sci. 76:2758–2764. doi:10.3168/jds.S0022-0302(93)77613-4.

Vinson, W.E., J.M. White, and R.H. Kliewer. 1976. Overall Classification as a Selection Criterion for Improving Categorically Scored Components of Type in Holsteins. J. Dairy Sci. 59:2104–2114. doi:10.3168/jds.S0022-0302(76)84494-3.

Visscher, P.M., and M.E. Goddard. 1995. Genetic Parameters for Milk Yield, Survival, Workability, and Type Traits for Australian Dairy Cattle. J. Dairy Sci. 78:205–220. doi:10.3168/jds.S0022-0302(95)76630-9.

vit. 2016. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed December 13, 2016. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf.

CHAPTER 4:

Prototype of the new routine genetic evaluation system for longevity in German Holsteins

Introduction

It is a question in many research projects, how the gained knowledge can be transferred into practical applications. In our case, the problem was formulated from a practical application and it was clear from the beginning that the results should be used to replace the current routine genetic evaluation of longevity for Holsteins in Germany. Results from a prototype version were used in chapter 5. This chapter gives a brief description of the key features of this prototype. The trait definition and model are described first in order to give a better understanding of the data processing steps.

Trait definition and model

Survival of nine different periods, defined in Table 4.1, is considered in the new model. Survival observations for a period are coded 1 if a cow survived this period and 0 if she was culled during this period. Survival observations in periods after culling or censoring are considered to be missing.

The following model is fitted to the data:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of survival (0/1) observations for different periods as defined in Table 4.1, $\mathbf{X}$ is an incidence matrix, linking observations to fixed effects, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{Z}$ is the incidence matrix of random animal effects, $\mathbf{a}$ is the vector of random animal effects ($\mathbf{a} \sim N(0, \mathbf{G_0} \otimes \mathbf{A})$, with the genetic covariance matrix $\mathbf{G_0}$ and the numerator relationship matrix for animals $\mathbf{A}$), and $\mathbf{e}$ is a vector of random residual effects ($\mathbf{e} \sim N(0, \mathbf{R_0} \otimes \mathbf{I})$, with the residual covariance matrix $\mathbf{R_0}$; off-diagonal elements of $\mathbf{R_0}$ are assumed to be zero). Fixed effects in the model are (1) an effect for herd × year × season of the day of entrance into each period, (2) an effect for region, and (3) an effect of milk yield, relative to the herd mean, × 5-year period. The fixed effect of relative milk yield always refers to the period prior to the survival trait under consideration. This effect is therefore dropped for L1.1.

Genetic parameters for the above model were estimated similarly to the procedure described in chapter 2.

Table 4.1: Definition of periods and opportunity windows. Opportunity windows are the minimum waiting period from calving to the date of data cutoff before an observation is included in the genetic evaluation. Days refer to days from the respective calving.

Periods (d)		Lactation			
Start	End	1	2	3	Opportunity Window (d)
0	49	L1.1	L2.1	L3.1	100
50	249	L1.2	L2.2	L3.2	300
250	Consecutive calving	L1.3	L2.3	L3.3	500

Data

IT Solutions for Animal Production (vit) is the main data supplier for German national routine genetic evaluations in Holsteins (black-and-white, red-and-white, and red-and-white dual purpose), Angler/Red Dairy Cattle, Jerseys, and the Black-and-White Friesian Cattle. Data for some federal states as well as for Austria are supplied by external milk recording organizations. The detailed organizational structure of data supply can be read from the description of the national routine genetic evaluation systems (vit, 2016).

Data processing

All programs were written by the author in Fortran 95/2003 (ISO/IEC JTC 1/SC 22, 2010) or Perl (Christiansen et al., 2012). They run on Linux machines and are controlled with bash scripts. Scripts for some graphical analyses were written in R (R Core Team, 2016).

Data transfer interface. For the new routine evaluation, a new data transfer interface was needed for two different reasons:

1) In the current genetic evaluation of longevity, data is only supplied if a cow had participated in at least one test day, due to the current functional definition of longevity, which requires a measure of milk yield. Disposals of cows which happen before their first test day are the most expensive disposals: these cows needed the full rearing costs but did not generate any income from milk. In the new routine evaluation, we wanted to make use of the information on all daughters with at least the first calving, including records of cows without any test day record.
2) The current data transfer interface is highly complex. To avoid errors due to this complex structure, a restructuration of the data transfer interface was required.

The new data transfer interface comprises of three files:

1) A file containing information about animal movements, i.e., consequent records for every cow and every herd she was sold to, comprising entry and disposal dates and disposal reasons. For cows being alive at the date of data cutoff, the disposal date and reason are missing.
2) A file containing information on all calvings for all cows, i.e., parity numbers and respective dates of calving.
3) A file containing a maximum of one record per cow with information on her last test date.

Information on a cow's milk performance is taken from the national routine genetic evaluation of milk traits, which is described in vit (2016).

Plausibility checks and data restrictions. The following plausibility checks are performed for the records of every cow:

1) Both parents must be known.
2) The date of birth must be known, valid and later than 1994.
3) The date of first calving must be known and valid; if any later calving date was invalid, the cow's longevity record is considered to be censored at her last valid date.
4) Age at first calving must be in the range from 20 to 40 months.
5) Calving intervals must be in the range from 300 to 600 days.

Coding of variables. For censored records, the *date of censoring* can be one of

1) A cow's last observed test date
2) The date of data cutoff
3) The date the cow was sold for dairy purposes
4) The date the farm finished dairy business

Survival observations for the respective periods are then coded as follows: 1 if the cow survived a period, 0 if the cow was culled in this period. Survival observations following culling or censoring are considered to be missing. Opportunity windows (Table 4.1) are applied to each period. The term opportunity window refers to a waiting period from the respective calving to the data cutoff date before the survival observation for a period is included. These opportunity windows are vital to avoid phenotypic biases of the kind explained in chapter 1. It was found from pre-studies that EBVs of young bulls are very sensitive to even slight phenotypic bias. This procedure results in up to nine survival observations per cow.

Herd × *year* × *season* effects are coded using the herd identification number, year and season (January – March, April – June, July – September, October – December) of entrance into a respective period. If classes have less than five observations, they are merged recursively within herd × year, then herd × 2-year period, then herd × 4-year period. Year refers to the entrance into a respective period.

Regions are defined as the federal states of Germany, represented by their milk recording organizations, and Austria and Luxemburg.

Relative milk yield × *5-year period* refers to the interaction effect of an overall 5-year period with the relative milk yield, which is defined as follows: yields for all test day records are pre-corrected for breed lactation curves, using estimates from the routine genetic evaluation of milk traits (vit, 2016). For the first period of either lactation, a quadratic regression on days in milk is then performed within herd × year of calving. The relative milk yield of a cow is estimated as the mean deviation of her test day milk yield(s) from this herd specific curve. This regression is performed in order to account for the variability in days in milk until lactation peak performance is reached (e.g., Macciotta et al., 2005). The milk yields for the second and third periods of either lactation are represented as accumulated lactation yield until 250 and 305 days in milk, following the guideline of the German Cattle Breeders' Federation (ADR, 2001). After estimating the deviation of a particular cow from the herd mean, this difference is categorized within herd × year of calving into five classes of 20% each. The interaction effect of relative milk yield × 5-year period is always estimated from the period prior to the survival observation under consideration in order to avoid confounding of the milk yield record with functional traits as discussed in more detail in chapter 6. Therefore, this effect is dropped for L1.1.

Solver

A new software was developed to solve the mixed model equations. The preconditioned conjugate gradients (PCG) algorithm was implemented in form of the algorithm described by Strandén and Lidauer (1999). We designed the software such that it can be used for other routine genetic evaluations as well and that only little afford should be necessary to implement multi-core functionality.

Construction of the index breeding value and approximation of its reliability

Solving the above described model results in nine estimated breeding values (**EBVs**) on the risk-level for each animal in the pedigree. At the end, only one EBV for longevity is published. To combine such EBVs from multi-trait linear models, two methods were suggested, both approximating the area under the survival curve: a non-linear method (Sewalem et al., 2007) and a linear

approximation of this non-linear method (van Pelt et al., 2015). For a detailed discussion on different methods to combine breeding values from linear multiple-trait models for survival and their relationships between each other, see chapter 6. In the prototype version of the new routine genetic evaluation of longevity, the nine different EBVs are combined to an index following an approach similar to Sewalem et al. (2007).

Table 4.2: Lengths of periods, population-wide risks and relative weights as resulting from the linear index combination method for EBVs on the risk-level (van Pelt et al., 2015).

Period	Length	Risk	Relative weight
L1.1	50	0.051	0.228
L1.2	200	0.078	0.201
L1.3	160	0.110	0.162
L2.1	50	0.042	0.128
L2.2	200	0.093	0.108
L2.3	160	0.135	0.076
L3.1	50	0.070	0.053
L3.2	200	0.118	0.035
L3.3	160	0.161	0.010

Reliabilities are approximated following Liu et al. (2004) with weights derived from the linear method, using lengths and risk estimates for different periods as seen from Table 4.2 (van Pelt et al., 2015). The following steps are applied:

Step 1: Approximation of the area under the individual survival curve. The population mean risk for culling in period t, R_t, is estimated following (Kaplan and Meier, 1958; for further details see chapter 1). To compare the genetic effects of an individual animal i to the population mean, we estimate the area under its individual survival curve, S_{ti} as follows:

The estimate for the individual survival curve is given by

$$S_{ti} = \prod_{j=1}^{t}(1 - R_j + EBV_{ji})$$

where R_j is the population mean probability for culling in period j and EBV_{ji} is the EBV of animal i for period j. The area under the survival curve for the first three lactations is then approximated with the following formula:

$$L_i = \sum_{t=1}^{9} \Delta_t \times \frac{S_{(t-1)i} + S_{ti}}{2}$$

with $S_{0i} = 1$ and Δ_t being the length of period t as seen from Table 4.2.

Step 2: Standardization. Index EBVs are then standardized with regard to the respective cow base population to a mean of 100 and a genetic standard deviation of 12 as used for all other traits in the German national routine genetic evaluation system (vit, 2016). The genetic standard deviation for the non-linear index was obtained from a simulation study which is discussed in detail in chapter 6. With this simulation study, a heritability of 0.088 was estimated for the longevity index.

References

ADR (German Cattle Breeders' Federation). 2001. ADR-Richtlinie 1.1 für das Verfahren der Durchführung der Milchleistungs- und Qualitätsprüfung (MLP) bei Rindern. Accessed March 23, 2017. http://www.adr-web.de/services/files/richtlinien-empfehlungen/ADR-Richtlinie%201.1.pdf.

Christiansen, T., brian d foy, L. Wall, and J. Orwant. 2012. Programming Perl. 4th Edition. O'Reilly Media, Sebastopol, California.

ISO/IEC JTC 1/SC 22 Programming languages, their environments and system software interfaces. 2010. ISO/IEC 1539-1:2010 - Information Technology -- Programming Languages -- Fortran -- Part 1: Base Language. International standardization subcommittee for programming languages, their environments and system software interfaces.

Kaplan, E.L., and P. Meier. 1958. Nonparametric Estimation from Incomplete Observations. J. Am. Stat. Assoc. 53:457–481. doi:10.1080/01621459.1958.10501452.

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. J. Dairy Sci. 87:1896–1907. doi:10.3168/jds.S0022-0302(04)73348-2.

Macciotta, N.P.P., D. Vicario, and A. Cappio-Borlino. 2005. Detection of Different Shapes of Lactation Curve for Milk Yield in Dairy Cattle by Empirical Mathematical Models. J. Dairy Sci. 88:1178–1191. doi:10.3168/jds.S0022-0302(05)72784-3.

van Pelt, M.L., T.H.E. Meuwissen, G. de Jong, and R.F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130. doi:10.3168/jds.2014-9090.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sewalem, A., F. Miglior, G.J. Kistemaker, P. Sullivan, G. Huapaya, and B.J. Van Doormaal. 2007. Short Communication: Modification of Genetic Evaluation of Herd Life from a Three-Trait to a Five-Trait Model in Canadian Dairy Cattle. J. Dairy Sci. 90:2025–2028. doi:10.3168/jds.2006-719.

Strandén, I., and M. Lidauer. 1999. Solving Large Mixed Linear Models Using Preconditioned Conjugate Gradient Iteration. J. Dairy Sci. 82:2779–2787. doi:10.3168/jds.S0022-0302(99)75535-9.

vit. 2016. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed December 13, 2016. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf.

CHAPTER 5:

A closer look at longevity in dairy cows: different stages – different genes

Johannes Heise[1], Sandra Kipp[2], Ngoc-Thuy Ha[1], Friedrich Reinhardt[2], Henner Simianer[1]

[1]University of Goettingen, Animal Breeding and Genetics, Albrecht-Thaer-Weg 3, 37075 Goettingen, Germany

[2]IT Solutions for Animal Production, Heinrich-Schroeder-Weg 1, 27283 Verden, Germany

Manuscript in preperation

Abstract

The aim of this genome-wide association study (**GWAS**) was the investigation of time-dependent patterns for the association between the genome and survival of different periods, defined within and across the first 3 lactations in dairy cattle. Knowledge about such patterns is important to further improve genomic and genetic evaluations of longevity, an economically important and publicly discussed trait. We used deregressed proofs for survival of 9 periods of the first 3 lactations (each 0-49 d, 50-249 d, and 250 d to the consecutive calving) and autosomal HD SNP-genotype data of 4,849 bulls in our GWAS. Associations were analyzed using (1) a single-marker GWAS, (2) a gene-based GWAS using Ensembl gene annotation data and (3) a gene-based mixed model, where gene regions with significant associations identified from (2) were modeled as random. Eight different gene regions on BTA5 (*ABCC9*), BTA6 (*GC*), BTA7 (*ARRDC3*), BTA14 (*SPATC1*) and BTA18 (a pseudogene, *SNORD112*, *CEACAM18* and *TMC4*) were found to be significantly associated to survival of at least 1 of the defined periods. We found that strengths of associations for most of these regions followed a distinct time-dependent pattern. With our results, we confirm regions found in recent GWAS and add knowledge about the differentiated genetic background of survival of different periods, defined within and across lactations.

Key words

Genome-wide association study, survival, longevity, dairy cattle, gene regions

Introduction

Longevity is an economically important trait in dairy cattle (Allaire and Gibson, 1992). Therefore, routine genetic evaluations of this trait are run in all major dairy breeding countries (Interbull, 2016). This trait can be interpreted as consecutive survival of different periods. Different periods in a cow's life refer to differing physiological challenges and different management decisions of the farmer. Recent quantitative genetic studies have shown that the genetic background of longevity does not only differ between lactations, but also within lactations (Sasaki et al., 2015; van Pelt et al., 2015; Heise et al., 2016). Accordingly, it must be assumed that different genes contribute differently to survival of different periods.

Recent genome-wide association studies (**GWAS**) for longevity in Holstein dairy cattle referred to whole productive life or herd life as a trait (Zhang et al., 2016; Nayeri et al., 2017). To our knowledge, no large scale GWAS has been reported so far that accounts for the differentiated genetic background of survival of different periods, defined within and across lactations.

Knowledge of the relationship between conventionally estimated breeding values (**EBV**, using pedigree information) and the genome, represented in our case by high-density (**HD**) SNP-marker genotypes, is important to improve future models for routine genomic and genetic evaluations of longevity. Therefore, we conducted a GWAS with the data from a prototype version of the new genetic evaluation system for longevity in German Holstein cows.

Data and Methods

Deregressed proofs of EBV

In a previous paper, we suggested to use a linear multiple trait model for genetic evaluations of longevity (Heise et al., 2016). In this model, 3 periods were considered for each of the first 3 lactations. Survival was coded as 0 (disposed) or 1 (survived) for each of the periods. A similar model was used to estimate breeding values for Holstein bulls and cows, using period definitions as seen from Table 5.1. The following model was fitted to the data:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of survival (0/1) observations, $\mathbf{X}$ is an incidence matrix, linking the observations to the fixed effects, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{Z}$ is the incidence matrix of random animal effects, $\mathbf{a}$ is the vector of random animal effects ($\mathbf{a} \sim N(0, \mathbf{G_0} \otimes \mathbf{A})$, with the genetic covariance matrix $\mathbf{G_0}$ and the numerator relationship matrix (pedigree-based) for animals $\mathbf{A}$), and $\mathbf{e}$ is a vector of random residual effects ($\mathbf{e} \sim N(0, \mathbf{R_0} \otimes \mathbf{I})$, with the residual covariance matrix $\mathbf{R_0}$; off-diagonal elements of $\mathbf{R_0}$ were assumed to be zero). Fixed effects in the model were (1) an effect for herd × year × season of the day of entrance into each period, (2) an effect for region, and (3) an effect of milk yield relative to the herd mean × 5-year period. Relative milk yield was categorized as quintiles, i.e., 5 classes of 20 % each, from lowest to highest relative milk yield. The genetic parameters assumed were similar to those in Heise et al. (2016). The mixed model equations were solved using an in-house developed multiple-purpose software package.

Table 5.1: Definition of periods for which survival was coded 0 (disposed) or 1 (survived).

Definition of periods (d from calving)		**Lactation**		
Start	**End**	**1**	**2**	**3**
0	49	L1.1	L2.1	L3.1
50	249	L1.2	L2.2	L3.2
250	Consecutive calving	L1.3	L2.3	L3.3

EBV were deregressed as

$$y_{ij}^{*} = \frac{EBV_{ij}}{r_{ij}^{2}}$$

where y_{ij}^{*} is the deregressed proof of animal i for trait j and ${r^2}_{ij}$ is the approximate reliability of EBV_{ij}. Reliabilities were approximated following Liu et al. (2004). We only considered animals with a reliability ≥ 0.7 for the EBV of L1.1.

Imputing missing genotypes on HD level

50k (Illumina BovineSNP50 Genotyping BeadChip v2, Illumina Inc., San Diego, CA) genotype data were available for 29,923 animals at 44,747 autosomal positions from the German national routine genomic evaluation (vit, 2016). Genotype information on HD level (Illumina BovineHD Bead chip, 777,692 SNP) was available for 1,366 animals. After haplotype phasing with Beagle version 4.1 r1398 (Browning and Browning, 2007), the 50k genotypes were imputed to the HD level using the software Minimac (Howie et al., 2012). Quality criteria included the call rate per individual (≥ 0.95) and the minor allele frequency (≥ 0.001). Finally, 4,849 bulls with imputed genotype information at 583,841 positions and a reliability ≥ 0.7 for the EBV of the first survival trait were available for this analysis.

Association studies

Single-marker GWAS. A single-marker GWAS was carried out using the following model:

$$\mathbf{y}^{*} = \mathbf{1}'\mu + \mathbf{xb} + \mathbf{g} + \mathbf{e}$$

where $\mathbf{y}^{*}$ is a vector of deregressed proofs, $\mathbf{1}'$ is a vector of ones, μ is a general mean term, $\mathbf{x}$ a vector containing the respective marker genotype (one of 0, 1 or 2), and $\mathbf{b}$ is the regression coefficient. The vector $\mathbf{g}$ consists of random effects for each bull with $\mathbf{g} \sim N\left(0, \mathbf{G}\sigma_g^2\right)$, where $\mathbf{G}$ is the genomic relationship matrix and σ_g^2 is the genomic variance; $\mathbf{e}$ is a vector of random residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, σ_e^2 being the residual variance. This analysis was carried out using the GCTA software (Yang et al., 2011a). For further representation, top associated SNPs were selected within 5 Mbp regions.

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

$$\frac{\hat{\sigma}_{SNP_{ij}}^2}{\hat{\sigma}_{g_j}^2} = \frac{2p_i(1-p_i)\hat{b}_{ij}^2}{\hat{\sigma}_{g_j}^2}$$

where $\sigma_{g_j}^2$ is the variance in the deregressed proofs of trait j, explained by all markers, p_i is the allele frequency of the allele coded as '1' for SNP-marker i and $\hat{b}_{ij}$ is the estimated marker effect. Approximate confidence intervals of 95 % were constructed for the proportion of variance by bootstrapping with a sample size of 500,000, drawing $\hat{b}_{ij}$ from $N(\hat{b}_{ij}, \widehat{SE}(\hat{b}_{ij})^2)$ and $\hat{\sigma}_{g_j}^2$ from

$N(\hat{\sigma}^2_{g_j}, \widehat{SE}(\hat{\sigma}^2_{g_j})^2)$. $\widehat{SE}$ refers to the estimated standard errors for $\hat{b}_{ij}$ and $\hat{\sigma}^2_{g_j}$, as calculated by the GCTA software.

Genomic inflation factors (λ_{median}) were estimated (Yang et al., 2011b).

Gene-based association analysis. Bakshi et al. (2016) proposed a fast method for a set-based association analysis, considering the structure of linkage disequilibrium (**LD**) between markers. This method uses summary statistics from a single-marker GWAS and a reference data set of individual-level genotypes to estimate the LD-structure between SNP-markers. In order to get gene-based statistics, we applied this method to the results from the single-marker GWAS using the GCTA software (Yang et al., 2011a). The Ensembl gene annotation data (Aken et al., 2016) were used to define sets of SNP-markers belonging to RNA-coding regions as well as to known pseudogenes. In the following, these regions are referred to as gene regions. Each gene region included SNP-markers between the UTRs plus 10 kbp up- and downstream flanking sequences. In total, 21,217 gene regions with on average 6.47 SNP-markers (1-205) were analyzed. The reference data set for the estimation of the LD-structure was the same as used for the single-marker analysis. A Bonferroni correction was applied to both models.

For further illustration, sets of highly associated gene regions were selected as follows: within a range of 5 Mbp up- and downstream around the mid-points of top associated gene regions, a maximum of 3 regions were selected if their p-value was below the genome-wide Bonferroni threshold on the 0.05 significance level.

Gene-based estimation of variance components. To estimate the variance, explained by the top associated gene regions, the following mixed model was fitted to the data:

$$\mathbf{y}^* = \mathbf{1}'\mu + \mathbf{g_1} + \mathbf{g_2} + \mathbf{e}$$

where $\mathbf{y}^*$ is the vector of deregressed proofs, $\mathbf{g_1}$ is a vector of random bull effects ($\mathbf{g_1} \sim N(0, \mathbf{G_1}\sigma^2_{g_1})$, with the genetic relationship matrix based on SNP-markers from the gene region, $\mathbf{G_1}$, and $\sigma^2_{g_1}$ being the corresponding variance component), $\mathbf{g_2}$ is another vector of random bull effects ($\mathbf{g_2} \sim N(0, \mathbf{G_2}\sigma^2_{g_2})$, with the genetic relationship matrix $\mathbf{G_2}$, based on all available SNP-markers outside a range of 10 Mbp up- and downstream of the considered gene region, $\sigma^2_{g_2}$ being the corresponding genomic variance); $\mathbf{e}$ is a vector of random residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$, where σ^2_e is the residual variance. The GCTA software (Yang et al., 2011a) was used for the estimation of variance components. The estimated variance ratio $\hat{\sigma}^2_{g_1}/\hat{\sigma}^2_{g_{2\,(all\,SNP)}}$ was then computed, where $\hat{\sigma}^2_{g_{2\,(all\,SNP)}}$ is the variance estimate obtained from the reduced model $\mathbf{y}^* = \mathbf{1}'\mu + \mathbf{g_{2\,(all\,SNP)}} + \mathbf{e}$ where all available SNP-markers contribute to $\mathbf{G_{2\,(all\,SNP)}}$. Approximate

standard errors for variance ratios were derived using the delta method[1] as also implemented in GCTA for heritability estimation. Additionally to this, a likelihood ratio test was performed, comparing likelihoods between the full model and the reduced model $\mathbf{y}^* = \mathbf{1}'\mu + \mathbf{g_2} + \mathbf{e}$.

Results

Single-marker GWAS

Manhattan plots for all 9 survival traits are shown in Figure 5.1. Markers from 6 different regions of 5 Mbp each on BTA5 (89 Mbp), BTA6 (89 Mbp), BTA14 (2 Mbp) and BTA18 (45, 58, and 63 Mbp) showed significant association to at least one of the survival traits. Different patterns of significance were observed: for the regions on BTA5, BTA6 and BTA18 (58 and 63 Mbp), significance was highest for the association to survival of the first and second period of a lactation, but lower for third periods. The region on BTA14 showed no significant association to survival traits of the first lactation, but to survival of L2.1 and L3.1. Contrary to these patterns, the region at 45 Mpb on BTA18 showed stronger significance for the association to survival late in lactation than to survival earlier in the same lactation.

These patterns also applied to the proportions of genomic variance, explained by the top associated SNP-markers from different regions (results not shown). Among all proportions, the top SNP on BTA6 explained the maximum proportion (4.0 %) for survival of L3.1. Close to this value, the top SNP in the region at 58 Mbp on BTA18 explained 3.8 % of the genomic variance in survival of L1.1.

Gene-based association analysis

Details about the 8 most significantly associated gene regions are shown in Table 5.2, displaying a maximum of 3 gene regions within 5 Mbp. Manhattan plots for all gene regions are shown in supplemental Figure 5.1. Highest significance was observed for the region around *CEACAM18* (Carcinoembryonic Antigen Related Cell Adhesion Molecule 18) at 58 Mbp, and a region at 44 Mbp around a pseudogene (ENSBTAG00000004994), both on BTA18. Only for the gene regions on BTA5 and 14, the top associated markers from the single-marker analyses were located within the corresponding top associated gene regions. The maximum distance was observed for the region at 44 to 45 Mbp on BTA18, where the top associated SNP-marker was almost 1.2 Mbp distant from the top associated gene region. Patterns of significance for the top associated gene regions followed the above described patterns for single SNP-markers close to these regions. Additional to the regions with significant association to longevity from the single-

[1] Off-diagonal elements of the inverse Average Information matrix were assumed to be zero, diagonal elements for $\hat{\sigma}^2_{g_1}$ and $\hat{\sigma}^2_{g_2\,(all\,SNP)}$ were taken from the output of GCTA for the full and the reduced model, respectively.

marker GWAS, gene regions on BTA7 (93 Mbp) and BTA18 (51 Mbp) were significantly associated. Both regions showed higher significance to survival of the last period of a lactation compared to earlier periods of the same lactation.

Gene-based estimation of variance components

Ratios of the gene region based variances by the genomic variances are shown in Figure 5.2. The maximum variance ratios for the different regions ranged from 0.009 ± 0.008 (ARRDC, BTA7, 93 Mbp, L2.2) to 0.116 ± 0.065 (GC, BTA6, 89 Mpb, L3.1). In general, observed patterns were similar to the patterns of significances from the gene-based GWAS. An exception was the region on BTA7, where the greatest proportion within different lactations was explained by the second period instead of the last. Likelihood ratio tests were highly significant ($p \leq 0.001$) for all regions and all survival traits except for the effect of SNORD112 (BTA18, 51 Mbp) on survival of L3.1 ($p = 0.003$).

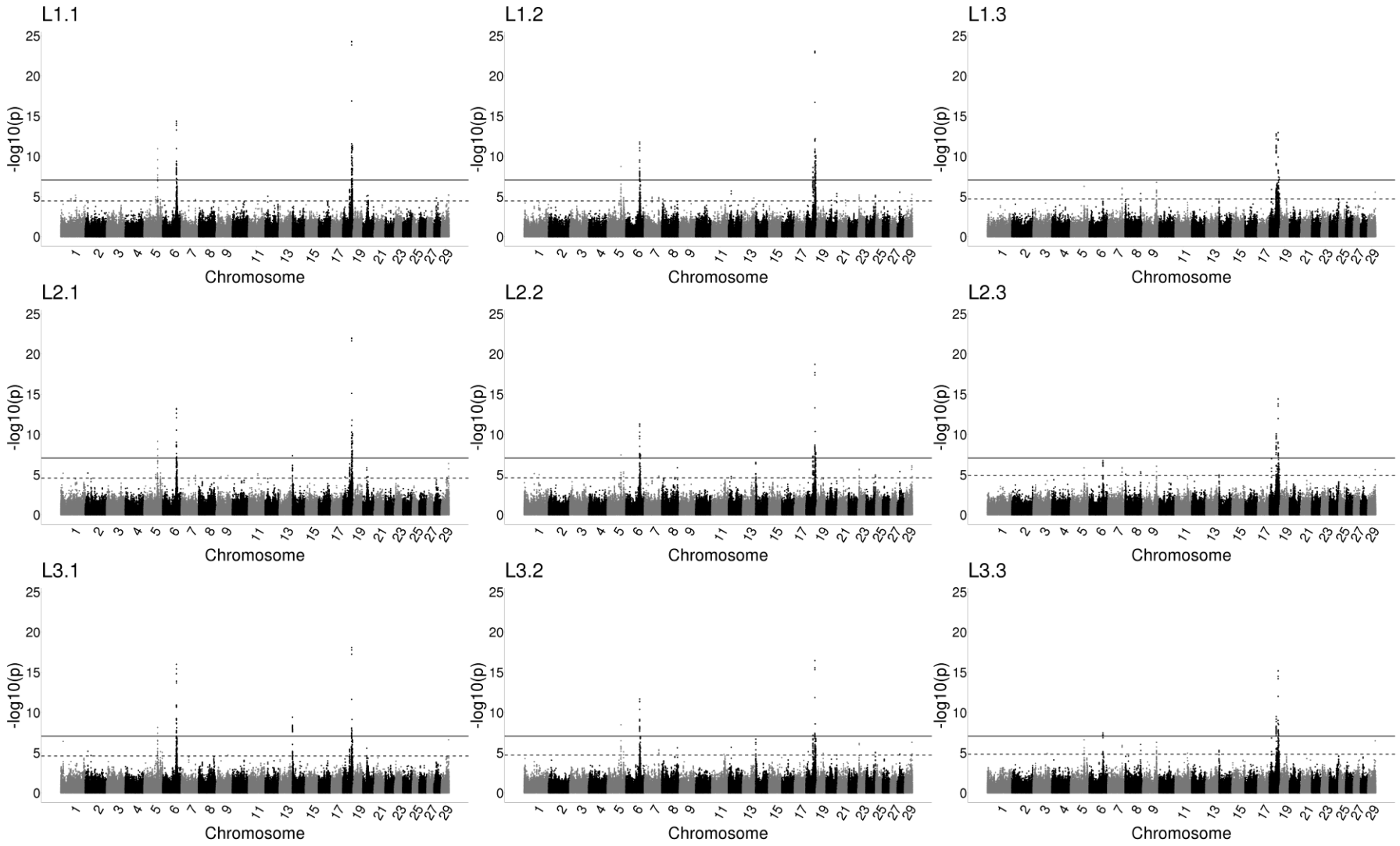


Figure 5.1: Manhattan plots for the single-marker GWAS. Rows refer to different lactations, columns to different periods within lactation. The solid line marks the genome-wide Bonferroni threshold, the dashed line the false discovery rate, each at 0.05 significance level.

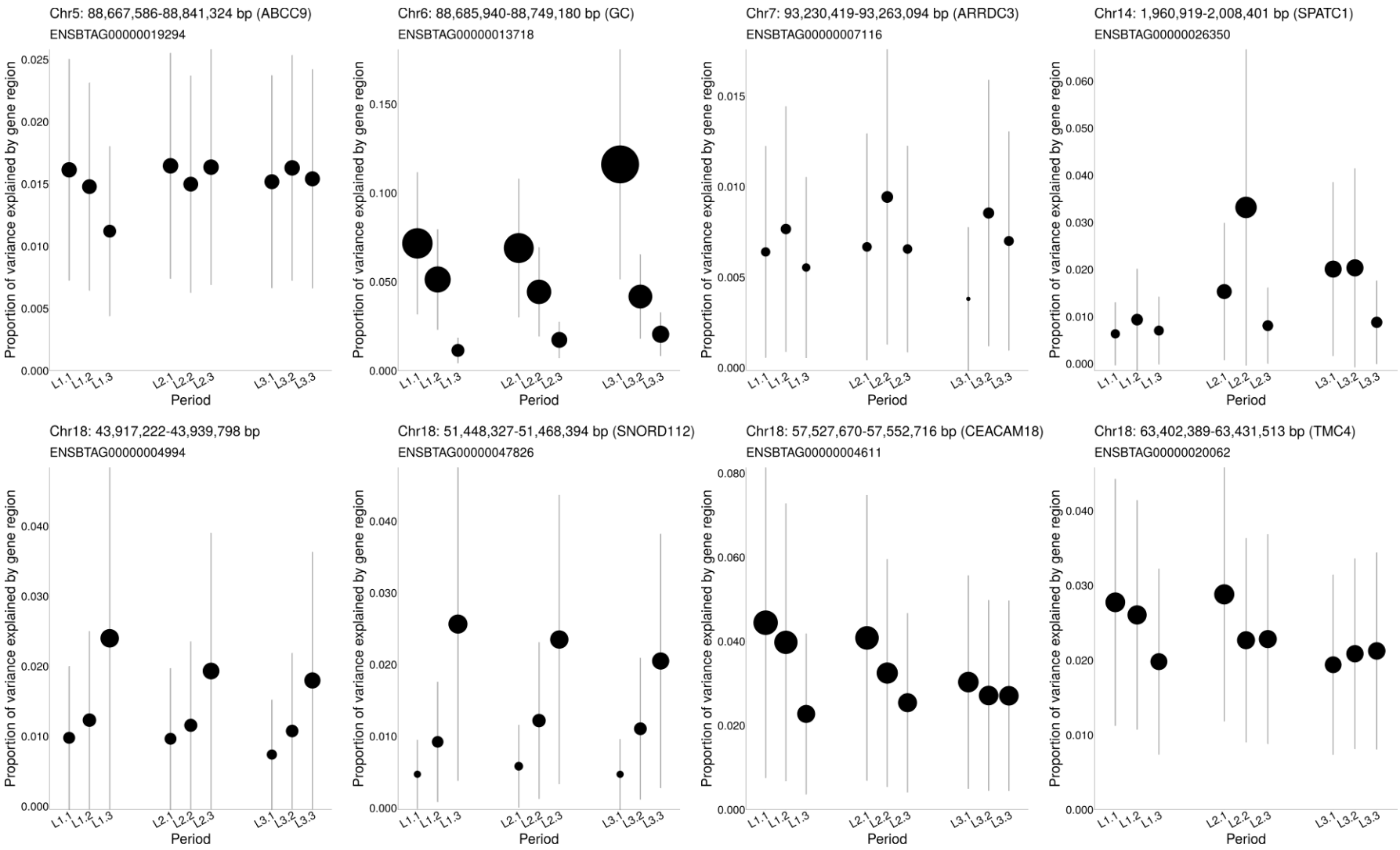


Figure 5.2: Proportion of variance, explained by top associated gene regions including 10 kbp flanking sequences. Vertical lines represent standard errors, derived from the inverse Average Information matrix using the delta method as also implemented in GCTA (Yang et al., 2011a). Point sizes correspond to the proportion of genomic variance explained.

Table 5.2: Gene regions with significant association to at least one of the survival traits. Table sorted by region and significance level within region (decreasing). A maximum of 3 gene regions within 5 Mbp-region are displayed.

Ensembl-ID	Name	Type	N SNP	Chr	Start (bp)	End (bp)	Max($-\log_{10}(p)$)	Max period	Top-SNP (bp)
ENSBTAG00000019294	*ABCC9*	protein coding	30	5	88,667,586	88,841,324	10.7	L1.1	88,795,885
ENSBTAG00000000593	*ST8STA1*	protein coding	14	5	88,277,352	88,463,869	5.7	L3.3	88,343,589
ENSBTAG00000013718	*GC*	protein coding	13	6	88,685,940	88,749,180	9.0	L3.1	88,728,581
ENSBTAG00000019716	*IL-8*	protein coding	3	6	90,549,882	90,573,647	7.3	L1.1	90,557,327
ENSBTAG00000043960	*MTHFD2L*	protein coding	20	6	90,832,934	90,995,937	5.9	L1.1	90,952,911
ENSBTAG00000007116	*ARRDC3*	protein coding	6	7	93,239,419	93,263,094	5.6	L1.3	93,254,737
ENSBTAG00000026350	*SPATC1*	protein coding	3	14	1,960,919	2,008,401	8.4	L3.1	1,967,325
ENSBTAG00000007834	*PPP1R16A*	protein coding	1	14	1,618,814	1,643,988	8.1	L3.1	1,638,045
ENSBTAG00000044263	*bta-mir-2309*	miRNA	1	14	2,061,851	2,081,925	7.7	L3.1	2,069,181
ENSBTAG00000004994		pseudogene	2	18	43,917,222	43,939,798	12.3	L1.3	43,927,101
ENSBTAG00000013175	*KIAA0355*	protein coding	4	18	44,907,050	44,979,117	8.7	L1.3	44,913,691
ENSBTAG00000003826	*SCN1B*	protein coding	7	18	45,950,622	45,980,522	8.3	L1.3	45,973,751
ENSBTAG00000047826	*SNORD112*	snoRNA	3	18	51,448,327	51,468,394	6.1	L1.3	51,460,508
ENSBTAG00000021789	*ZNF574*	protein coding	3	18	51,494,690	51,517,155	5.9	L1.3	51,510,151
ENSBTAG00000018912	*ARHGEF1*	protein coding	5	18	51,648,123	51,688,112	5.8	L1.3	51,652,913
ENSBTAG00000004611	*CEACAM18*	protein coding	3	18	57,527,670	57,552,716	21.2	L1.1	57,548,213
ENSBTAG00000039212	*CTU1*	protein coding	5	18	57,507,497	57,534,958	20.9	L1.1	57,516,245
ENSBTAG00000004608		protein coding	3	18	57,564,117	57,593,637	20.9	L1.1	57,589,121
ENSBTAG00000020062	*TMC4*	protein coding	9	18	63,402,389	63,431,513	11.0	L1.1	63,405,640
ENSBTAG00000015908	*MBOAT7*	protein coding	9	18	63,387,970	63,422,141	10.0	L1.1	63,405,640
ENSBTAG00000005841		protein coding	4	18	63,339,487	63,365,736	8.6	L1.1	63,361,108

Discussion

Genomic inflation factors

Genomic inflation factors (λ_{median}) for the single-marker analysis ranged between 0.94 and 0.98 for the different survival traits. These values are below the expectation of being substantially greater than 1, which is likely due to the LD-structure, a probably high number of contributing loci and sample size (Yang et al., 2011b, 2014). Other association studies on longevity in Holsteins reported inflation factors substantially greater than 1 (Zhang et al., 2016; Nayeri et al., 2017). The very low inflation factors might partly be due to the model, where the tested SNP is contained in the genome-wide relationship matrix. The tested SNP does therefore not only contribute to its single-marker fixed effect **b**, but also to the random correction term **g** for the correction of possible population stratification. This reduces power for the detection of associated SNP-markers (Yang et al., 2014). Therefore, significances of SNP-markers are probably estimated overly conservative in our study.

Standard errors for estimated variance ratios

In our study, approximate standard errors for the estimated ratios of variances from the gene-based variance component estimation were derived from the inverse Average Information matrix using the delta method. This method assumes that estimates of variance components are approximately normally distributed with $\hat{\sigma}_i^2 \sim N(\hat{\sigma}_i^2, \widehat{SE}(\hat{\sigma}_i^2)^2)$, where $\widehat{SE}(\hat{\sigma}_i^2)$ is the estimated standard error for the estimate $\hat{\sigma}_i^2$ of variance component i, and $\widehat{SE}(\hat{\sigma}_i^2)^2$ is the i^{th} diagonal element of the inverse Average Information matrix. Assuming a normal distribution for the estimated variance ratio, confidence intervals could then be constructed as $(\hat{r}^2 - k \cdot \widehat{SE}(\hat{r}^2), \hat{r}^2 + k \cdot \widehat{SE}(\hat{r}^2))$, where $\hat{r}^2$ is the estimated variance ratio, $\widehat{SE}(\hat{r}^2)$ its estimated standard error and $k = \Phi^{-1}(1 - \alpha/2)$ with α being the desired significance level. For further details of the entire method see Schweiger et al. (2016). The authors of the mentioned study showed that this approach yields inaccurate confidence intervals, especially when the ratio of variance components tends towards one of the boundaries of the parameter space (0 and 1), which is the case in our study. Therefore, they developed an optimized bootstrapping method to construct more accurate confidence intervals for heritability estimates and provide a software package called ALBI. This software is designed to estimate distributions of heritability estimates, obtained with one genomic relationship matrix only. Therefore, this software could not be applied to our model. However, the likelihood ratio test showed highly significant effects for all considered gene regions on all survival traits except for the effect of *SNORD112* on survival of L3.1, where p was 0.003.

Associated regions

Chromosomes 5 and 6. Regions on BTA5 and BTA6, each at 89 Mbp, were found to be significantly associated to survival. A previous study reported these regions to be associated to mastitis

resistance (Sahana et al., 2014). Following our results that the highest significance is to survival early in lactation, but udder diseases were reported to be the main disposal reason in the mid-lactation period (Heise et al., 2016), this function is not obvious. However, the region on BTA6 with the strongest association was the region around the *GC* gene (*GC*, Vitamin D Binding Protein). In humans, the corresponding protein plays a vital role in the immune system and calcium metabolism (as reviewed by White and Cooke, 2000 and Speeckaert et al., 2006). Especially the latter is in good accordance to our results: Strongest associations were found for survival early in lactation, increasing with higher lactation numbers. Hypocalcemia, also known as milk fever, shows increasing incidences over lactations (Gröhn et al., 1998; Reinhardt et al., 2011) and increases the risk of culling mainly at the beginning of a lactation (Rajala-Schultz and Gröhn, 1999). As stated by Goff (2008), hypocalcemia also reduces the effective closure of the teat sphincter muscle which could, together with the above mentioned impaired immune response, explain the association to mastitis, which was found by Sahana et al. (2014). For L3.1, we estimated the region around *GC* to explain 11.6% of the genome-wide marker variance, which is very high. The top associated SNP in this region explained about 4% and strong associations for this region were also reported in another GWAS on longevity (Zhang et al., 2016). For the region on BTA5, the *ABCC9* (ATP Binding Cassette Subfamily C Member 9) gene region showed the strongest association, followed by *ST8SIA1* (ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 1). *ABCC9* is a regulator of potassium channels and is located in a region which was found to be associated to protein production (Nayeri et al., 2016).

Chromosome 7. On BTA7, the region around the *ARRDC3* (Arrestin Domain Containing 3) gene was highest associated to survival late in lactation, which implies reproduction success. It was suggested as a candidate gene for growth and muscularity traits in beef cattle (Bolormaa et al., 2014). Obesity resistance, skin abnormalities and embryonic lethality where reported for mice in which this gene was knocked-out. Embryonic lethality was not expressed when mother were fed with high fat diets (Shea et al., 2012). If embryonic lethality or similar phenotypes were also found in cows to be caused by this gene, this would be in accordance to our results.

Chromosome 14. The region on BTA14 showed a unique pattern of significance among the associated regions: It was not significantly associated to survival of the first lactation, but to survival of the main lactation periods (up to 249 d from calving) in second and third. The highest association was found for the protein coding gene *SPATC1* (Spermatogenesis and Centriole Associated 1). This was previously reported to be associated to milk yield as well as fat and protein content (Jiang et al., 2014).

Chromosome 18. Multiple regions, spanning approximately 20 Mbp in total, were found on BTA18. The overall gene-based top association was estimated for the *CEACAM18* (Carcinoembryonic Antigen Related Cell Adhesion Molecule 18) gene at 58 Mbp, followed by *CTU1* (Cytosolic Thiouridylase Subunit 1). *CEACAM18* was estimated to explain 4.4% of the genome-wide marker variance for survival of L1.1, which was the second highest estimate for a region in our

study. This is in good accordance to other studies, which reported the region around *CEACAM18* to be associated to calving traits like calving ease, still births (Mao et al., 2014, 2016) and calving interval (Raven et al., 2016). In humans, members of the CEACAM subgroup of the CEA family are involved in many cancers (as reviewed by Hammarström, 1999) and may play a role in maternal-fetal tolerance (as reviewed by Riley, 2008). They also play an important role in immune response (as reviewed by Gray-Owen and Blumberg, 2006). The *CTU1* gene was also found as a candidate gene for calving traits (Mao et al., 2016; Raven et al., 2016) and (Purfield et al., 2015) called a mutation in this gene the most likely candidate for direct calving difficulty in the Irish Holstein-Frisian population. The genes *TMC4* (Transmembrane Channel Like 4) and *MBOAT7* (Membrane Bound O-Acyltransferase Domain Containing 7) at 63 Mbp also showed a significant association to survival. In humans of European descent, a variant in this region was reported to be associated to non-alcoholic fatty liver disease (Mancina et al., 2016). In dairy cattle, a fatty liver in the transition phase is a risk factor for ketosis shortly after calving (as reviewed by Adewuyi et al., 2005). Especially the region around 58 Mbp and its reported association with calving traits (Mao et al., 2014) is in good accordance with the pattern of significance for this region (strongest association signal on survival up to 249 d after calving). The same pattern was observed for the region at 63 Mbp.

Conversely, the regions at 45 Mbp and 51 Mbp on BTA18 showed highest significance for survival in late lactation (from day 250 to consecutive calving, a period which primarily reflects reproduction success). The region at 51 Mbp was not significant at the single-marker level. For the region at 45 Mbp, the region with the highest association was the pseudogene ENSBTAG00000004994, followed by the *KIAA0355* gene and the *SNC1B* gene. *KIAA0355* codes for an uncharacterized protein, but *SNC1B* is known to play a role in the early development of the nervous system (Brackenbury et al., 2008; Patino and Isom, 2010). It is also associated to epilepsy (Wallace et al., 1998; Heron et al., 2007). The relatively high association to the last period of a lactation could indicate that this region plays a role in the reproduction process of cows.

Beside the regions on BTA6 at 89 Mbp and on BTA18 at 45 and 58 Mbp, which were consistently found in other GWAS on longevity (Zhang et al., 2016; Nayeri et al., 2017), Nayeri et al. (2017) reported regions at 52 Mbp on BTA18 and at 93 Mbp on BTA7 to be associated with direct herd life. Both regions showed significant association to longevity at the gene-level in our study.

To investigate if any of the considered regions had effects in different directions on different survival traits, we also estimated variance components from pairwise bivariate models, where the random effects of the respective region on different survival traits were considered to be correlated. Approximately 30% of these bivariate runs did not converge, probably due to the fact that the correlation for the effects of any region on different survival traits was estimated to be close to 1 from the converged runs (results not shown).

In summary, our results can be used to further investigate candidate regions for causal relationships to longevity. The knowledge of time-dependent patterns of associations to survival could help to find functional mechanisms.

Conclusions

Most of the regions found in our study were already reported earlier. However, our study contributes novel information regarding their differentiated relevance for survival of different periods. Our study therefore underlines the potential of accounting for the differentiated genetic background of survival of different periods within and across lactations when designing models for routine genetic and genomic evaluations.

Acknowledgments

The German national FBF (Förderverein Bioökonomieforschung e.V., "Association for Bioeconomy Research", Bonn, Germany) is thanked for the financial support.

References

Adewuyi, A.A., E. Gruys, and F.J.C.M. van Eerdenburg. 2005. Non esterified fatty acids (NEFA) in dairy cattle. A review. Vet. Q. 27:117–126. doi:10.1080/01652176.2005.9695192.

Aken, B.L., S. Ayling, D. Barrell, L. Clarke, V. Curwen, S. Fairley, J. Fernandez Banet, K. Billis, C. García Girón, T. Hourlier, K. Howe, A. Kähäri, F. Kokocinski, F.J. Martin, D.N. Murphy, R. Nag, M. Ruffier, M. Schuster, Y.A. Tang, J.-H. Vogel, S. White, A. Zadissa, P. Flicek, and S.M.J. Searle. 2016. The Ensembl gene annotation system. Database 2016. doi:10.1093/database/baw093.

Allaire, F.R., and J.P. Gibson. 1992. Genetic Value of Herd Life Adjusted for Milk Production. J. Dairy Sci. 75:1349–1356. doi:10.3168/jds.S0022-0302(92)77886-2.

Bakshi, A., Z. Zhu, A.A.E. Vinkhuyzen, W.D. Hill, A.F. McRae, P.M. Visscher, and J. Yang. 2016. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. Sci. Rep. 6. doi:10.1038/srep32894.

Bolormaa, S., J.E. Pryce, A. Reverter, Y. Zhang, W. Barendse, K. Kemper, B. Tier, K. Savin, B.J. Hayes, and M.E. Goddard. 2014. A Multi-Trait, Meta-analysis for Detecting Pleiotropic Polymorphisms for Stature, Fatness and Reproduction in Beef Cattle. PLoS Genet. 10:e1004198. doi:10.1371/journal.pgen.1004198.

Brackenbury, W.J., M.B.A. Djamgoz, and L.L. Isom. 2008. An Emerging Role for Voltage-Gated Na+ Channels in Cellular Migration: Regulation of Central Nervous System Development and Potentiation of Invasive Cancers. The Neuroscientist 14:571–583. doi:10.1177/1073858408320293.

Browning, S.R., and B.L. Browning. 2007. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. Am. J. Hum. Genet. 81:1084–1097. doi:10.1086/521987.

Goff, J.P. 2008. The monitoring, prevention, and treatment of milk fever and subclinical hypocalcemia in dairy cows. Vet. J. 176:50–57. doi:10.1016/j.tvjl.2007.12.020.

Gray-Owen, S.D., and R.S. Blumberg. 2006. CEACAM1: contact-dependent control of immunity. Nat. Rev. Immunol. 6:433–446. doi:10.1038/nri1864.

Gröhn, Y.T., S.W. Eicker, V. Ducrocq, and J.A. Hertl. 1998. Effect of Diseases on the Culling of Holstein Dairy Cows in New York State. J. Dairy Sci. 81:966–978. doi:10.3168/jds.S0022-0302(98)75657-7.

Hammarström, S. 1999. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Semin. Cancer Biol. 9:67–81. doi:10.1006/scbi.1998.0119.

Heise, J., Z. Liu, K.F. Stock, S. Rensing, F. Reinhardt, and H. Simianer. 2016. The genetic structure of longevity in dairy cows. J. Dairy Sci. 99:1253–1265. doi:10.3168/jds.2015-10163.

Heron, S.E., I.E. Scheffer, S.F. Berkovic, L.M. Dibbens, and J.C. Mulley. 2007. Channelopathies in Idiopathic Epilepsy. Neurotherapeutics 4:295–304. doi:10.1016/j.nurt.2007.01.009.

Howie, B., C. Fuchsberger, M. Stephens, J. Marchini, and G.R. Abecasis. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44:955–959. doi:10.1038/ng.2354.

Interbull. 2016. National Genetic Evaluation Forms Provided by Countries. Accessed December 13, 2016. http://interbull.org/ib/geforms.

Jiang, L., X. Liu, J. Yang, H. Wang, J. Jiang, L. Liu, S. He, X. Ding, J. Liu, and Q. Zhang. 2014. Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. BMC Genomics 15:1105. doi:10.1186/1471-2164-15-1105.

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. J. Dairy Sci. 87:1896–1907. doi:10.3168/jds.S0022-0302(04)73348-2.

Mancina, R.M., P. Dongiovanni, S. Petta, P. Pingitore, M. Meroni, R. Rametta, J. Borén, T. Montalcini, A. Pujia, O. Wiklund, G. Hindy, R. Spagnuolo, B.M. Motta, R.M. Pipitone, A. Craxì, S. Fargion, V. Nobili, P. Käkelä, V. Kärjä, V. Männistö, J. Pihlajamäki, D.F. Reilly, J. Castro-Perez, J. Kozlitina, L. Valenti, and S. Romeo. 2016. The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. Gastroenterology 150:1219–1230.e6. doi:10.1053/j.gastro.2016.01.032.

Mao, X., N.K. Kadri, D.-J. De Koning, G. Sahana, and B. Guldbrandtsen. 2014. CEACAM18 as candidate for the Holstein calving QTL on BTA18. Page in Proceedings of the 10th World Congress on Genetics Applied to Livestock Production World Congress on Genetics Applied to Livestock Production, Vancouver, Canada.

Mao, X., N.K. Kadri, J.R. Thomasen, D.J. De Koning, G. Sahana, and B. Guldbrandtsen. 2016. Fine mapping of a calving QTL on Bos taurus autosome 18 in Holstein cattle. J. Anim. Breed. Genet. 133:207–218. doi:10.1111/jbg.12187.

Nayeri, S., M. Sargolzaei, M.K. Abo-Ismail, N. May, S.P. Miller, F. Schenkel, S.S. Moore, and P. Stothard. 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. BMC Genet. 17:75. doi:10.1186/s12863-016-0386-1.

Nayeri, S., M. Sargolzaei, M.K. Abo-Ismail, S. Miller, F. Schenkel, S.S. Moore, and P. Stothard. 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. J. Dairy Sci. 100:1246–1258. doi:10.3168/jds.2016-11770.

Patino, G.A., and L.L. Isom. 2010. Electrophysiology and beyond: Multiple roles of Na+ channel β subunits in development and disease. Neurosci. Lett. 486:53–59. doi:10.1016/j.neulet.2010.06.050.

van Pelt, M.L., T.H.E. Meuwissen, G. de Jong, and R.F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130. doi:10.3168/jds.2014-9090.

Purfield, D.C., D.G. Bradley, R.D. Evans, F.J. Kearney, and D.P. Berry. 2015. Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. Genet. Sel. Evol. 47:47. doi:10.1186/s12711-015-0126-4.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999. Culling of dairy cows. Part III. Effects of diseases, pregnancy status and milk yield on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:295–309. doi:10.1016/S0167-5877(99)00047-1.

Raven, L.-A., B.G. Cocks, K.E. Kemper, A.J. Chamberlain, C.J.V. Jagt, M.E. Goddard, and B.J. Hayes. 2016. Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. Mamm. Genome 27:81–97. doi:10.1007/s00335-015-9613-8.

Reinhardt, T.A., J.D. Lippolis, B.J. McCluskey, J.P. Goff, and R.L. Horst. 2011. Prevalence of subclinical hypocalcemia in dairy herds. Vet. J. 188:122–124. doi:10.1016/j.tvjl.2010.03.025.

Riley, J.K. 2008. Trophoblast Immune Receptors in Maternal-Fetal Tolerance. Immunol. Invest. 37:395–426. doi:10.1080/08820130802206066.

Sahana, G., B. Guldbrandtsen, B. Thomsen, L.-E. Holm, F. Panitz, R.F. Brøndum, C. Bendixen, and M.S. Lund. 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. J. Dairy Sci. 97:7258–7275. doi:10.3168/jds.2014-8141.

Sasaki, O., M. Aihara, A. Nishiura, H. Takeda, and M. Satoh. 2015. Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. J. Dairy Sci. 98:5781–5795. doi:10.3168/jds.2014-9152.

Schweiger, R., S. Kaufman, R. Laaksonen, M.E. Kleber, W. März, E. Eskin, S. Rosset, and E. Halperin. 2016. Fast and Accurate Construction of Confidence Intervals for Heritability. Am. J. Hum. Genet. 98:1181–1192. doi:10.1016/j.ajhg.2016.04.016.

Shea, F.F., J.L. Rowell, Y. Li, T.-H. Chang, and C.E. Alvarez. 2012. Mammalian Alpha Arrestins Link Activated Seven Transmembrane Receptors to Nedd4 Family E3 Ubiquitin Ligases and Interact with Beta Arrestins. PLoS ONE. 7:e50557. doi:10.1371/journal.pone.0050557.

Speeckaert, M., G. Huang, J.R. Delanghe, and Y.E.C. Taes. 2006. Biological and clinical aspects of the vitamin D binding protein (Gc-globulin) and its polymorphism. Clin. Chim. Acta 372:33–42. doi:10.1016/j.cca.2006.03.011.

vit. 2016. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed December 13, 2016. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf.

Wallace, R.H., D.W. Wang, R. Singh, I.E. Scheffer, A.L. George, H.A. Phillips, K. Saar, A. Reis, E.W. Johnson, G.R. Sutherland, S.F. Berkovic, and J.C. Mulley. 1998. Febrile seizures and generalized epilepsy associated with a mutation in the Na+-channel ß1 subunit gene SCN1B. Nat. Genet. 19:366–370. doi:10.1038/1252.

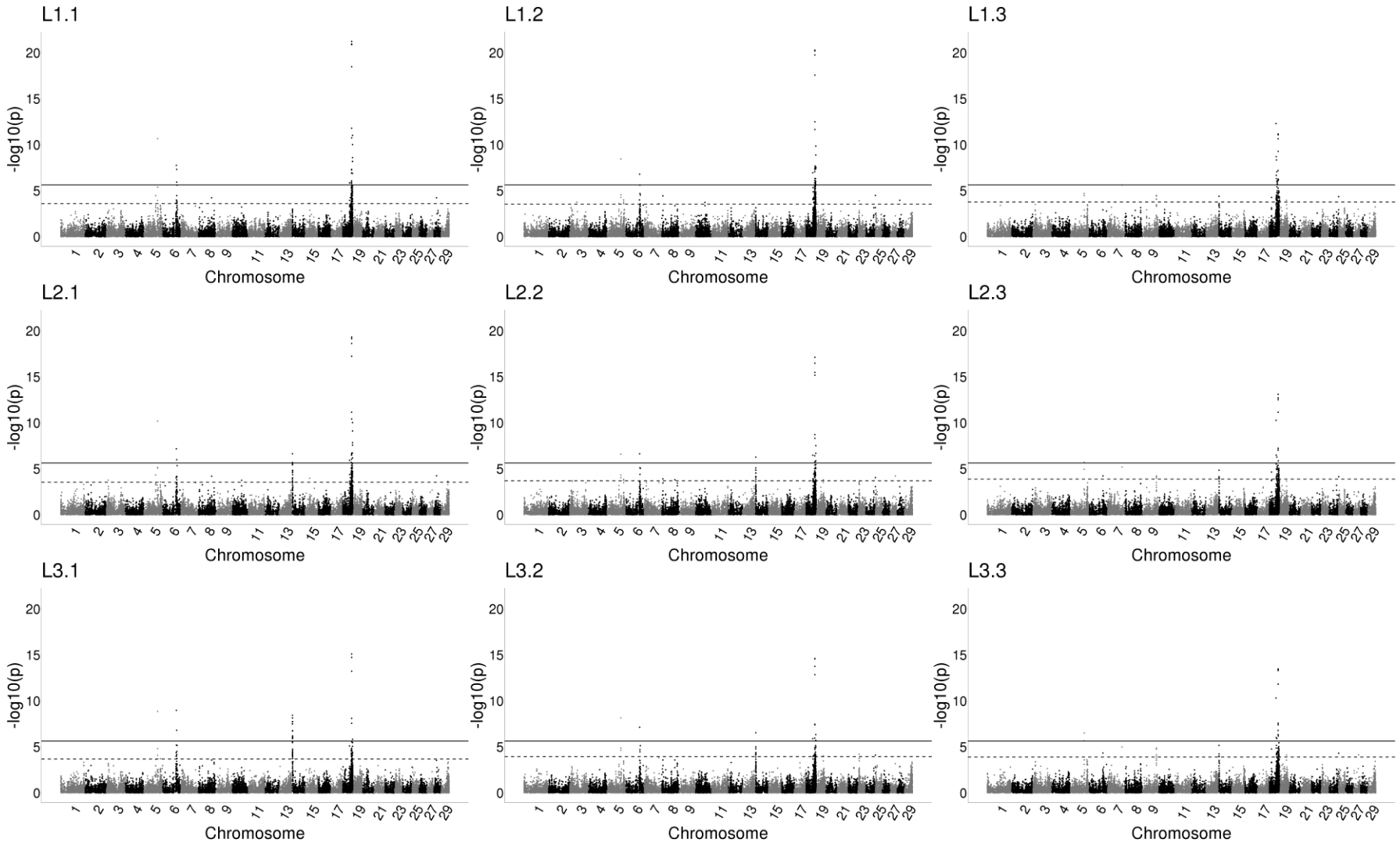
White, P., and N. Cooke. 2000. The Multifunctional Properties and Characteristics of Vitamin D-binding Protein. Trends Endocrinol. Metab. 11:320–327. doi:10.1016/S1043-2760(00)00317-9.

Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher. 2011a. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82. doi:10.1016/j.ajhg.2010.11.011.

Yang, J., M.N. Weedon, S. Purcell, G. Lettre, K. Estrada, C.J. Willer, A.V. Smith, E. Ingelsson, J.R. O'Connell, M. Mangino, R. Mägi, P.A. Madden, A.C. Heath, D.R. Nyholt, N.G. Martin, G.W. Montgomery, T.M. Frayling, J.N. Hirschhorn, M.I. McCarthy, M.E. Goddard, and P.M. Visscher. 2011b. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19:807–812. doi:10.1038/ejhg.2011.39.

Yang, J., N.A. Zaitlen, M.E. Goddard, P.M. Visscher, and A.L. Price. 2014. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 46:100–106. doi:10.1038/ng.2876.

Zhang, Q., B. Guldbrandtsen, J.R. Thomasen, M.S. Lund, and G. Sahana. 2016. Genome-wide association study for longevity with whole-genome sequencing in 3 cattle breeds. J. Dairy Sci. 99:7289–7298. doi:10.3168/jds.2015-10697.



Supplemental Figure 5.1: Manhattan plots for the gene-based GWAS. Rows refer to different lactations, columns to different periods within lactation. The solid line marks the genome-wide Bonferroni threshold, the dashed line the false discovery rate, each at 0.05 significance level.

CHAPTER 6:

General discussion

General discussion

This chapter discusses important topics of this thesis. The first sections deal with aspects that were not discussed in detail in the previous chapters, but which are vital for the success of a new genetic evaluation system for longevity. Then, potential for future research is highlighted, based on results from chapter 5. At the end, future prospects are given for the selection on functional traits.

Methods for the combination of risk-level EBVs to an index

With the prototype of the new routine genetic evaluation system for longevity in German Holsteins, nine different breeding values are estimated per animal on the risk-level, one for the conditional survival of each period. In the short and medium term, only one breeding value will be published for longevity in German Holsteins. Furthermore, Interbull requires one breeding value for direct longevity from every participating country to conduct international routine genetic evaluations (Interbull, 2016a). For these reasons, it is required to combine the nine different risk-level EBVs to one value. This value represents the new genetic evaluation of longevity towards breeding organizations and practitioners. Therefore, the methodology of combining risk-level EBVs to an index for longevity is of great relevance. This topic evoked discussions in many meetings along this longevity project, and different aspects will therefore be presented in this section.

In chapter 4, only a brief description was given on how the construction of this combined breeding value for longevity is currently performed in the prototype version of the new routine genetic evaluation system. As mentioned there, two previous studies suggested different methods to reach this goal (Sewalem et al., 2007; van Pelt et al., 2015). Our method can be easily derived from the first method. In the following, an overview of the different methods, their similarities and differences, is given. The three methods are referred to as **method I**, which is the method described by Sewalem et al. (2007) applied to our case, **method II** is our modification and **method III** is the method suggested by van Pelt et al. (2015). All three methods start from the same base assumptions but come to slightly different results. Although they are not all linear, the term ‘index’ is used for combined EBVs, regardless of the method.

Base assumptions. All methods start from the discrete survival function:

$$S_{ti} = \prod_{j=1}^{t}(1 - R_j + g_{ji})$$

where S_{ti} is the probability for animal i at its first calving to survive all periods up to period t, R_j is the estimated risk of an average animal to be culled in period j, conditional on survival of all previous periods and g_{ji} reflects the conditional genetic potential of animal i to resist culling in

period j. All methods then approximate the area under the individual survival curve, which is the life expectancy for this animal at the day of its first calving, reflecting its genetic potential (Klein and Moeschberger, 2003). Note that the derivation of the predictor for total lifespan of a sire's offspring, based on his breeding value, starts from the analog base assumption for proportional hazards survival models (Yazdi et al., 2002).

Method I. Sewalem et al. (2007) combined risk-level estimated transmitting abilities, which are half the EBVs. In the first instance, we concentrate on EBVs, because they are the basis for selection in German Holsteins (vit, 2016). Renaming variables to fit the variable nomenclature from chapter 1 and adapting the number of periods to our case, the formula proposed by the authors to combine EBVs for the conditional survival of different periods becomes:

$$DHL_i = \left(\sum_{t=1}^{10} \left(S_{(t-1)i} - S_{ti} \right) \times (N_t + D_t) \right) + K$$

where DHL_i is the estimated direct herd life for animal i in days, $S_{ti} = \prod_{j=1}^{t}(1 - R_j + EBV_{ji})$ for periods $1 \leq t \leq 9$, $S_{0i} = 1$, and $S_{10i} = 0$, N_t is the population-wide mean interval from first calving to the calving previous to the period t in days (originally in the publication of Sewalem et al. (2007): calving interval, but then, DHL_i would not approximate the complete area under the survival curve), D_t is the mean number of days of production for cows culled in period t ($D_{10} = 0$) and K is a constant, reflecting the expected number of days of production after the fourth calving. From Figure 6.1, which shows schematically the principle of this formula, it becomes clear that the approximation of the area under the survival curve is accomplished using horizontal bars.

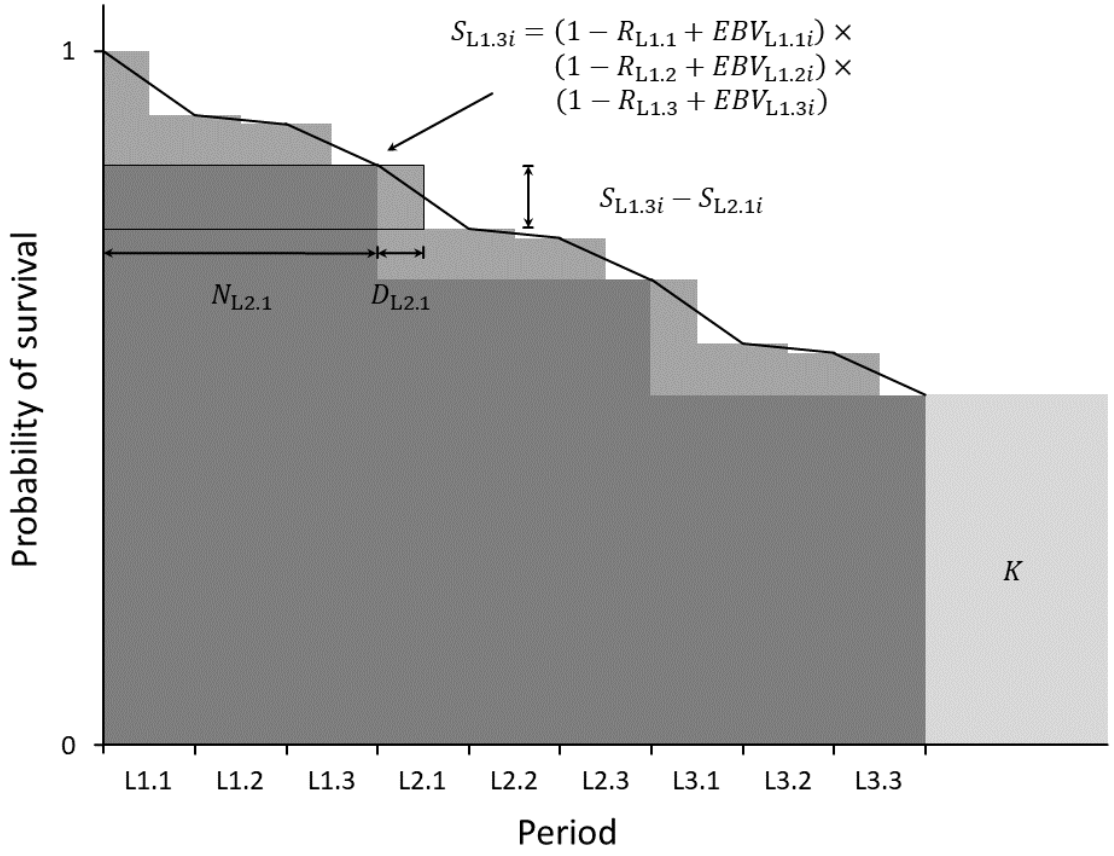


Figure 6.1: Schematic representation of the formula of Sewalem et al. (2007) to construct an index from a number of different consecutive survival traits as the area under the survival curve. Exemplarily, the formula to compute $\boldsymbol{S_{L1.3i}}$ and the decomposition of the summand for period $\boldsymbol{t}$ = L2.1 are presented, where $\boldsymbol{N_{L2.1}}$ is the calving interval for the 2nd calving and $\boldsymbol{D_{L2.1}}$ is the average number of days in milk for cows being culled in period L2.1; $\boldsymbol{K}$ is a constant, representing the average survival beyond the 4th calving. Dark grey refers to contributions by multiples of $\boldsymbol{N_t}$, medium grey to $\boldsymbol{D_t}$ and light grey to $\boldsymbol{K}$.

Method II. Our approach is slightly different, because we approximate the area under the survival curve using vertical trapezoids (see Figure 6.2):

$$L_i = \sum_{t=1}^{9} \Delta_t \times \frac{(S_{(t-1)i} + S_{ti})}{2}$$

where L_i is the index EBV for longevity of individual i, Δ_t is the mean length of period t: for the first and second period of a lactation l, exact lengths ($\Delta_{Ll.1}$ and $\Delta_{Ll.2}$) are used, and for the third period, $\Delta_{Ll.3} = CI_{l+1} - (\Delta_{Ll.1} + \Delta_{Ll.2})$, where CI_{l+1} is the mean calving interval which is terminated by the consecutive calving. S_{ti} is defined as for method I. From Figures 6.1 and 6.2, it becomes obvious that both formulas are equivalent in case the survival curve is approximated as a

stepwise linear function and K is ignored. If the survival function is not assumed to be stepwise linear, the formula of (Sewalem et al., 2007) could give slightly different results. In both cases, the index of risk-level EBVs for an individual animal is the population mean area under the survival curve, subtracted from the individual one.

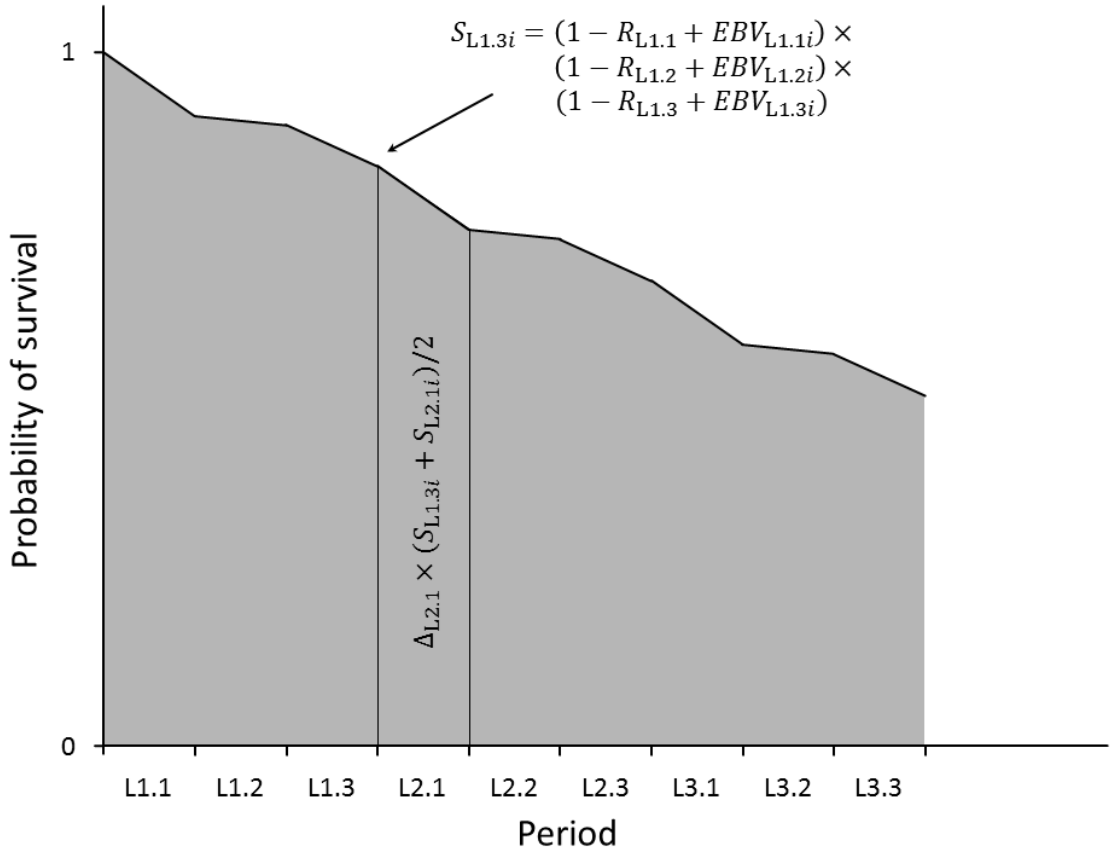


Figure 6.2: Schematic representation of the formula implemented in the prototype version of the new genetic evaluation system to construct an index from a number of different consecutive survival traits, which is the approximate area under the survival curve. Exemplarily, the formula to compute $\boldsymbol{S_{\mathrm{L1.3}}}$ and the decomposition of the summand for period $\boldsymbol{t = \mathrm{L2.1}}$ are presented, where $\boldsymbol{\Delta_{\mathrm{L2.1}}}$ is the mean length of the period **L2. 1**.

Method III. Van Pelt et al. (2015) suggested a linear approximation of the area under the survival curve. For a better understanding of their method, we consider their simplified example of only three periods with unity length each. Setting $p_t = 1 - R_t$, the area under the survival curve in the simplified example can be approximated as:

$$L_i = (p_1 + EBV_{1i}) + (p_1 + EBV_{1i})(p_2 + EBV_{2i}) + (p_1 + EBV_{1i})(p_2 + EBV_{2i})(p_3 + EBV_{3i})$$

where EBV_{ti} is the breeding value for animal i in period t. This expression matches exactly formula [2] in the paper of van Pelt et al. (2015).

By rearranging, we yield:

$$\begin{aligned}
L_i = \quad & (p_1 + p_1p_2 + p_1p_2p_3) && 1 && (0)\\
\\
+ \quad & (1 + p_2 + p_2p_3) && EBV_{1i} &&\\
+ \quad & (p_1 + p_1p_3) && EBV_{2i} && (1)\\
+ \quad & p_1p_2 && EBV_{3i} &&\\
\\
+ \quad & (1 + p_3) && EBV_{1i}EBV_{2i} &&\\
+ \quad & p_2 && EBV_{1i}EBV_{3i} && (2)\\
+ \quad & p_1 && EBV_{2i}EBV_{3i} &&\\
\\
+ \quad & 1 && EBV_{1i}EBV_{2i}EBV_{3i} && (3)
\end{aligned}$$

As van Pelt et al. (2015) state, the weights for the first order products (1) of EBVs are the first order partial derivatives of the population mean term (0) with respect to the conditional survival probability p_t of the respective period for EBV_{ti}. As easily seen from our simple example, this can be generalized for the second (2) and third (3) order products. Extension for more periods and varying lengths of periods is straightforward and becomes equivalent to method II. Van Pelt et al. (2015) suggest using only the first order product terms (1) as the index EBV of animal i. This would make the derivation of variance components for the index easy to a maximum extend, because ordinary selection index theory could then be applied: $\sigma^2_{g_L} = \mathbf{w'Gw}$ and $\sigma^2_{e_L} = \mathbf{w'Rw}$, where $\sigma^2_{g_L}$ and $\sigma^2_{e_L}$ are the genetic and residual variance of the index L, $\mathbf{w}$ is a vector with weights reflecting the lengths of periods and containing the partial derivatives of the population mean area under the survival curve as described above, and $\mathbf{G}$ and $\mathbf{R}$ are the genetic and residual covariance matrices for the different periods. If all orders of products are included in the index, the derivation of $\sigma^2_{g_L}$ and $\sigma^2_{e_L}$ is not trivial. Furthermore, it is obvious that due to the multiplicative composition of the index in methods I and II, a common expectation is no longer met: the transmitting ability, i.e., the expected realization of the genetic potential in the offspring is not exactly half the expected realization in the animal itself under the assumption that this expectation holds on the risk-level, which we approximate with the linear multiple-trait model.

Comparison of the linear and non-linear approaches

To investigate both, the genetic variance and transmitting abilities, a short simulation study and analyses on the results from the prototype version of the new genetic evaluation system were conducted to gain further understanding of the differences and similarities of the different methods and their behavior in the context of the new genetic evaluation of longevity. Because methods I and II use the same approach, only methods II and III were compared.

Simulation study. The following simulation was run ten times:

(1) Sampling nine true breeding values (**TBVs**) from $N(0, \mathbf{G})$ for 10,000 individuals, where $\mathbf{G}$ is the genetic covariance structure as estimated for the prototype version of the new genetic evaluation of longevity. The simulation of TBVs allows us to directly estimate the genetic variance of the resulting index.

(2) Construction of nine different indexes per individual, including incrementally higher orders of TBV products, beginning with the first order as suggested by van Pelt et al. (2015). Weights were derived, using the mean area under the survival curve for the German Holstein population and respecting differing lengths of periods as defined in the new genetic evaluation (for further details, see chapter 4).

(3) Comparison of the resulting nine indexes: mean correlations among all individuals over the ten runs, mean correlations among the top one percent, and estimates for $\sigma^2_{g_L}$ of the index.

(4) To gain knowledge about the expected realization of the genetic potential in the offspring, the index comprising all orders of products was built using $S_{ti} = \prod_{j=1}^{t}(1 - R_j + \frac{EBV_{ji}}{2})$ instead of $S_{ti} = \prod_{j=1}^{t}(1 - R_j + EBV_{ji})$ and was then compared to $\frac{L_i}{2}$, because this is the intuitive expectation, transferred from linear models. In the following, these terms are referred to observed and expected transmitting ability, respectively.

Real data. From the new prototype genetic evaluation system for longevity, data of 52,171 Holstein bulls with at least 20 daughters with phenotypes were analyzed. In total, these bulls had 12,107,371 daughters with survival observations.

Results and Discussion. Figure 6.3 shows scatterplots of the first three indexes from simulation run 1 against each other. This scatterplot shows that relationships are not simply linear between the first order index and higher order indexes, but almost linear between second and third. The same was observed for indexes of higher orders which are therefore not shown. The mean correlation of the ninth to the first order index was ≥ 0.999, but only 0.987 for the top one percent individuals. This means that in general, only few re-ranking due to the method is expected, but slight re-ranking for top list animals. In real data from the prototype version of the new genetic evaluation of longevity, the correlation of the first to the ninth order index was 0.994.

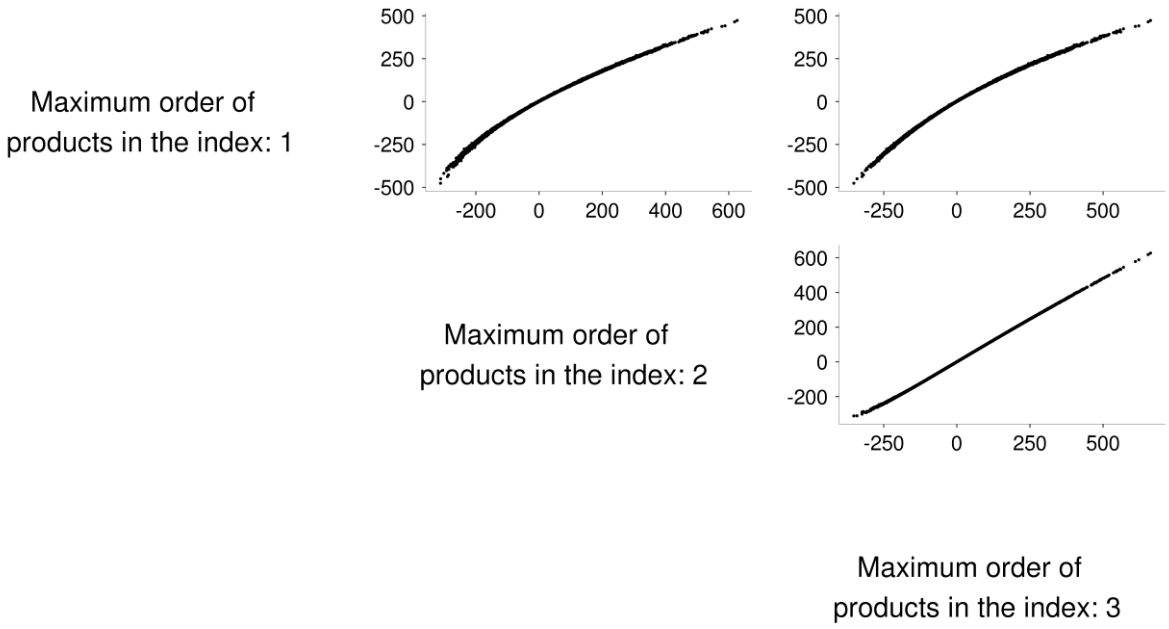


Figure 6.3: Scatter plots of simulated indexes for longevity, including first to third order of weighted EBV products.

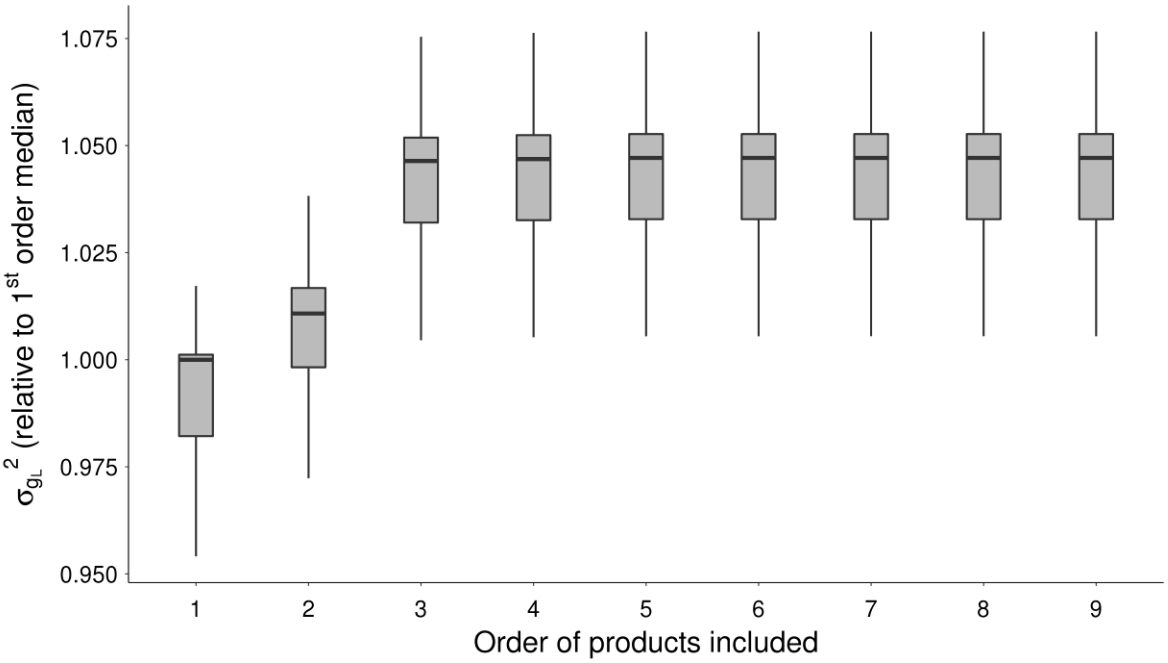


Figure 6.4: Genetic variances of indexes from ten simulation runs, incrementally including increasing orders of weighted TBV products. All variances are presented relative to the median of the first order index.

Figure 6.4 shows box plots of the genetic variance estimates for all nine indexes, relative to the median of the variances for the first order index. The estimated genetic variance increases for indexes including products up to the third order, then remains the same up to the ninth order index. This can be explained by the fact that TBVs on the risk-level have small variances and deviate around their mean: zero. High order products of such small values are practically zero, adding no further variance to the index.

Expected realizations in parent and offspring. As last point of this short simulation study and the analysis of real data from the prototype version of the new routine genetic evaluation of longevity, sire transmitting abilities were investigated for the index, comprising products of all nine TBVs (simulation) and EBVs (real data). Figure 6.5 shows expected and observed transmitting abilities for both, simulated and real data. Expected transmitting abilities are represented by half the index value, including all nine orders of EBV products. Observed transmitting abilities for the simulation are represented by the index, containing only half the EBVs of an individual on the risk-level. For real data, observed transmitting abilities for bulls were computed as averages of their daughters' EBVs, which were corrected for dam contributions ($L_{\text{daughter,corrected for dam}} = L_{\text{daughter}} - \frac{L_{\text{dam}}}{2}$). It is seen that especially for top and bottom bulls, the observed transmitting ability remains somewhat below the expected one. In real data, this is on average 0.051 σ_{g_L} for the top one percent of bulls.

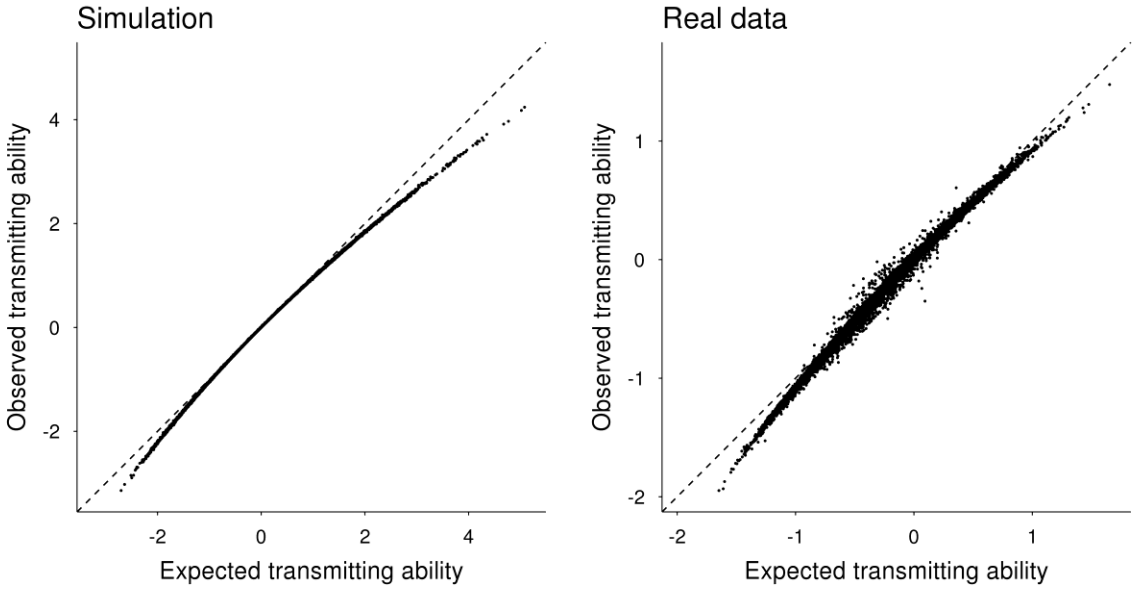


Figure 6.5: Observed versus expected transmitting abilities for the index including all nine orders of weighted TBV (EBV) products. Expected transmitting abilities are presented as half the index value; observed transmitting abilities for simulated data are the indexes built from half the parent TBVs on the risk-level, and for real data, observed transmitting abilities are represented by the means of daughter EBVs, corrected for dam, of 52,171 Holstein bulls. Scales are genetic standard deviations for both plots.

Implications. In this chapter, the mechanisms of different methods to construct an index from risk-level EBVs were shown. It was also shown that all presented methods approximate the same variable, i.e., expected longevity. Further, their similarities and differences were analyzed using simulation and real data. It can be concluded that almost no differences are expected with regard to the ranking of bulls. Furthermore, method III gives an approximation of the expected realization of longevity in the animal, which is twice the expected realization in the offspring. Both other methods give a theoretically slightly more accurate expectation for the realization of longevity in the animal itself, but slight differences exist also between the expected realization in the animal itself and twice the expected realization in its offspring. To avoid this property, method I or II could be adapted: the index for an animal could be presented as twice the index from half its EBVs on the risk-level, as originally suggested by Sewalem et al. (2007). This would reflect the expected realization of the parent's genetic potential in the offspring, but the index would be presented on the animal scale. The choice of methods should therefore depend on the purpose: if the interest is in the expected offspring realization, the adapted method I or II could be used; if the expected realization in the animal itself is of interest, e.g., when planning a certain

mating, expectations should be formulated on the risk-level and then combined to an index, using one of the methods I or II. If non-linearity shall be avoided, method III should be used.

Nowadays, a major task of pedigree-based routine genetic evaluations is to serve as a basis for genomic evaluations. Currently, only one routine genomic evaluation is run for longevity and it is not planned to conduct it directly with proofs on the risk-level. Although this might be desirable, EBVs from international routine evaluations (Multiple-trait Across Country Evaluations, **MACE**) are taken as input and only one trait for longevity is considered in MACE (Interbull, 2016a), due to the variety of definitions and methods used in the national evaluations of participants (Interbull, 2016b). If risk-level EBVs must be combined to one index value for the consequent genomic evaluation, it is desirable that this index EBV is a good expectation of the realization in the animal itself. Traits as longevity, for which accurate pedigree-based EBVs are available relatively late, benefit most from accurate genomic predictions. For these reasons, we chose method II to combine risk-level EBVs in the prototype of the new routine genetic evaluation system for longevity in German Holsteins.

Further considerations. All above mentioned methods target at an index for longevity on a time scale. Economically, culling at different time points produces different income losses and costs: on the one hand, culling risks for different periods are associated to different disorders (e.g., Rajala-Schultz and Gröhn, 1999a; b; c) and these are associated to different costs (e.g., Kossaibati and Esslemont, 1997; Gohary et al., 2016; van Soest et al., 2016; Charfeddine and Pérez-Cabal, 2017). On the other hand, the economic value of a cow is dependent on the point in the lactation cycle. For example, a cow being culled directly after her first calving induced large costs for her rearing period, but did not generate any income. Similarly, a cow, culled directly after her second calving, induced costs during her dry period, but did not generate any income in her second lactation. Compared to the same cow, but culled at the time point of her dry-off, the later culled cow produced additional costs without generating additional income. For this particular cow, earlier culling would have been economically more favorable. All three discussed methods for the construction of an index for longevity put high weight on early periods, due to the fact alone that survival in later periods is always conditional on survival of all previous periods. Economic considerations could lead to somewhat different weights, resulting in an economic longevity index which would no longer be interpretable on a time scale. Despite the more complicated interpretation of such an index, the derivation of proper weights, combining time scale with economic considerations, would be a relevant future project.

Beyond that, risk estimates as well as economic parameters are different for different farms. Farm-specific parameters could be used to derive a farm-specific genetic index for longevity or economic longevity. This functionality could be easily integrated into standard mating software tools.

Interbull test run and trend validation with Interbull's method III

Test run. With the results from the prototype version of the new routine genetic evaluation of longevity, **vit** (IT Solutions for Animal Production) participated in the Interbull MACE test run in January 2017 with data based on the national evaluation in December 2016. EBVs of 22,236 bulls, born within the years from 1986 to 2010 were used in this MACE test run. These bulls would have been published in Germany according to the German rules for publication, i.e., at least ten daughters in at least ten herds and the reliability of the EBV for longevity being ≥ 0.35. EBV correlations between the current and the new genetic evaluation were estimated to be about 0.86 in these bulls.

Across-country correlations to other participants in Interbull MACE runs were estimated and compared to correlations from the current evaluation of longevity. Estimated correlations were within the range from 0.55 (Hungary) to 0.90 (Canada, Denmark/Finland/Sweden) and did not change compared to the current genetic evaluation system.

The sire standard deviation, which reflects the genetic deviation, was estimated to be 9.86. Originally, national EBVs were scaled to a genetic standard deviation of 12, using the estimate from the above described simulation study, which might indicate that this estimate is either too high or that the sample of bulls does not represent the full genetic variance in the German Holstein population. However, because this is only a scaling factor and because the Interbull estimate for the within-sire genetic standard deviation was taken for trend validation, this has no relevance for the following statements.

Trend validation method III. The Interbull trend validation (method III) was performed for the prototype version of the new genetic evaluation system for longevity in Holsteins. Two complete genetic evaluation runs were conducted with the new prototype with data cutoff dates June 20, 2012 (truncated data set) and June 20, 2016 (full data set). Data of 3,835 bulls, born within the years 2002 to 2007, were used to perform the Interbull trend validation method III. Estimated effective daughter contributions (**EDC**, Liu et al., 2004) were used to measure additional daughter information for individual bulls over the four years between the evaluations. Three methods for the index construction were compared: (1) our original method ($h^2 = 0.088$), (2) the linear approximation, following (van Pelt et al., 2015; $h^2 = 0.097$, estimated using index theory), and our adapted method (3), where only half the EBVs were assumed on the risk-level and the index was multiplied by two afterwards ($h^2 = 0.081$, estimated using simulation). All EBVs were standardized to a mean of 100 and a genetic standard deviation of 12.

Assuming a correlation between the two evaluation runs of 0.95 and a genetic standard deviation σ_a of 9.86, the estimates for the coefficient of the bias term in the Interbull trend validation method III were 0.004 σ_a for our original method, 0.030 σ_a for the linear approximation (van Pelt et al., 2015) and 0.029 σ_a for our adapted method. This means, for the method that is currently implemented in the prototype of the new genetic evaluation system, there is no bias in the

EBVs of young bulls with regard to forward-prediction. If the index was built with one of the other methods, the Interbull trend validation with method III would have been failed, because the estimate of the bias term must not exceed 0.02 σ_a. This might indicate that EBVs on the risk-level are somewhat overestimated for youngest bulls with additional daughter information between the two runs. However, the higher estimates are based on the same assumption for σ_a (9.86) as for our original method which might be not the case. Therefore, these results should be taken with caution and further analyzed.

Nonetheless, potential causes of bias were traced as seen from Figure 6.6: phenotypic frequencies for survival observations were estimated by year × month of calving from raw phenotypes of all nine survival traits in both data sets. 1208 (solid lines) refers to the truncated data, 1608 (dashed lines) to the full data with data cutoff four years later. It can be seen that curves for all but the last period of either lactation are in perfect accordance between the two data sets. For the last periods, phenotypic frequencies are somewhat overestimated in the truncated data set. This overestimation is highest for L1.3, which has, compared to L2.3 and L3.3 a relatively high weight in the index EBV for longevity. This overestimation is likely due to the opportunity window for this period, which was assumed too short (500 days). It was chosen as the mean calving interval (approximately 410 days) plus 90 days to give cows with longer calving intervals a proper opportunity to show their consequent calving. The observed phenotypic bias could be generated by cows which did not conceive and which were milked longer than 500 days before they were culled without a consequent calving. As mentioned in chapter 1, EBVs, especially of young bulls, are very sensitive to such kind of phenotypic bias. It is therefore recommended to increase the length of the opportunity window for the last period of either lactation in order to likely decrease overestimation of EBVs. If there were still an overestimation of early EBVs after this optimization, trend validation could be performed for each trait separately in order to trace back the source of this bias.

Summarizing this section it can be said that with the currently implemented index construction method, the Interbull trend validation with method III was passed, showing no bias in the estimation of index EBVs for young bulls in the prototype version of the new routine genetic evaluation of longevity. Nonetheless, some optimizations can be applied in order to remove slight bias observed in the risk-level EBVs.

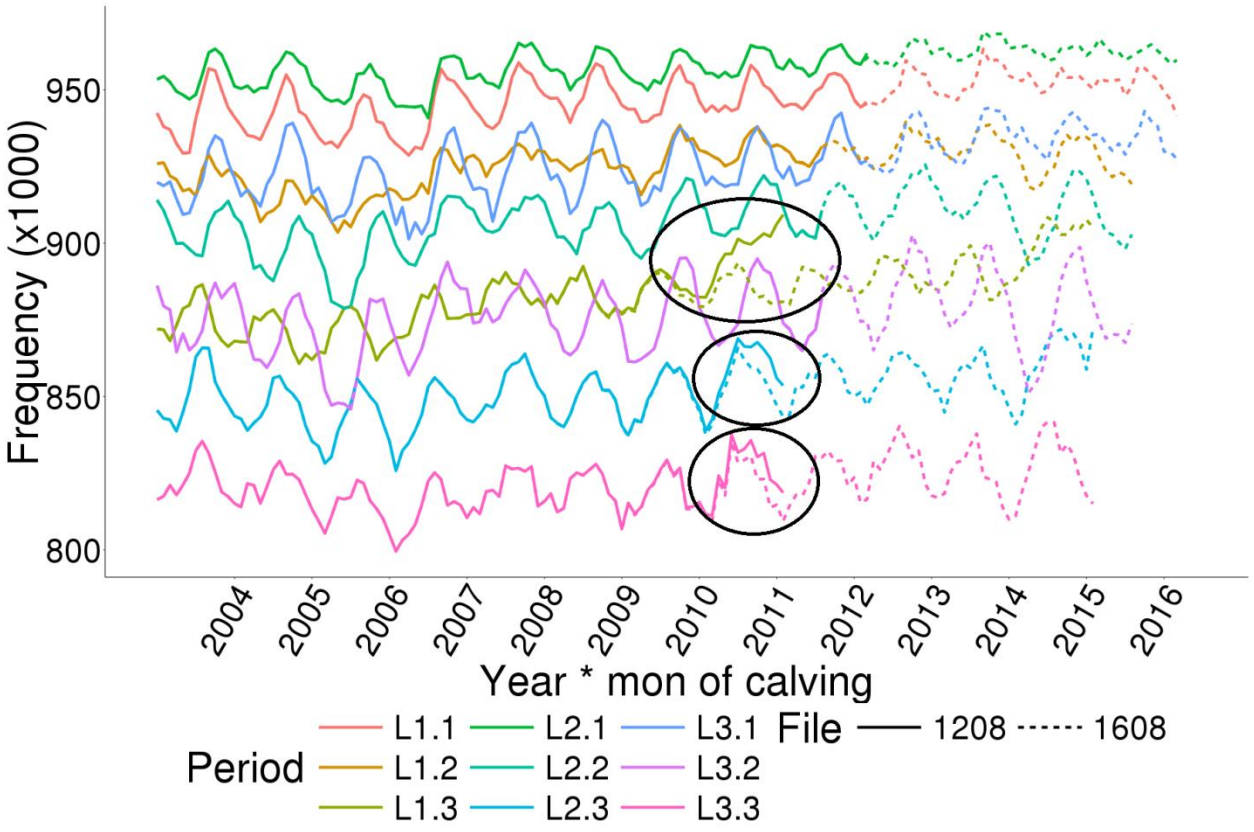


Figure 6.6: Phenotypic frequencies of the nine survival traits by year × month of calving in the truncated (1208) and the full data set (1608) used for the Interbull trend validation with method III. Periods with phenotypic overestimation in the truncated data set are marked with black ellipses.

General considerations about the choice of the model

In the following, some decisions on the choice of the model for the prototype version of the new genetic evaluation of longevity as well as some important observations from this prototype are discussed.

Linear model on binary data. A point, which might be questioned, is the use of a linear model for binary survival data. It was previously stated that such models are statistically suboptimal for this kind of data (Gianola, 1980; Ducrocq et al., 1988a). However, practical considerations play an important role when designing models for large-scale routine genetic evaluations as outlined in chapter 1. In our case, three considerations played a major role:

1) The requirement for an animal model
2) The necessity to account for multiple genetically distinct but correlated traits
3) The requirement for a computationally feasible routine evaluation system.

In routine genetic evaluations, there are internationally clocked deadlines for, e.g., the data submission to Interbull. Further, national requirements define deadlines for raw data supply. The window in between defines the available time interval for the solving process of the mixed model equations, which is usually no longer than 10 days. This could practically not be accomplished with a respective nine-trait-threshold model or a respective survival model. Furthermore, Holtsmark et al. (2009) have shown that linear multiple-trait models on binary survival observations can outperform survival and threshold models with regard to predicting survival up to 365 d under practical conditions.

The necessity for an animal multiple-trait model arises from the need for cow EBVs (to be used in a cow reference population for genomic prediction) and the results from chapters 2 and 5, which have shown that survival of different periods has different genetic determination.

Three-lactation model. The new model for longevity is a three-lactation multiple-trait model. It might be argued that dairy breeders are interested in healthy cows which survive longer than three lactations (if they have satisfying milk performance), especially due to the economic reasons outlined in chapter 1. In chapter 2, it was shown that estimated genetic correlations between survival of corresponding periods of the second and third lactation were ≥ 0.95. These periods can thus be considered to have the same genetic background. Further, distributions of culling reasons, also shown in chapter 2, suggest that this might also hold at least up to the fifth lactation. It must therefore be expected that the three-lactation model also reflects the genetic background of survival in later lactations. Another justification of the three-lactation model is easily seen from the weights applied to EBVs of different periods to combine them to an approximate linear index for longevity, following van Pelt et al. (2015) (Table 4.2 of chapter 4): most weight is given to survival of early periods, leaving only minor weights for survival of periods of the third lactation. This would be even less for the fourth, fifth and so on. For this reason, there is

presently no necessity to include survival beyond the fourth calving in a model for genetic evaluations of longevity in German Holstein cows. This statement would need to be reconsidered if the risk of culling throughout the first three lactations was substantially decreased by either, management or genetic gain.

Current versus new trait for longevity. Current and new genetic evaluation target genetically slightly different traits: the correlations between EBVs from the current system and the prototype of the new genetic evaluation were estimated to be about 0.86. This is plausible for the following reasons: in the current genetic evaluation system, longevity is measured in days from first calving. This means, a cow which did not conceive and is milked for an extended period until being culled gets a credit for this extension. In the new system, this cow is exactly treated the same as a cow being culled after 251 days in milk in the same lactation. It might be argued that a cow which has a good milk yield persistency, allowing for a prolonged lactation period is more valuable than a cow that does not. However, usually, milk yield decreases in late lactation (e.g., Wilmink, 1987; Bertilsson et al., 1997; Silvestre et al., 2009) and without a new calving this cannot be regenerated. In the current genetic evaluation of longevity, a credit is also given for prolonged calving intervals, which is not the case in the new system. Usually, long calving intervals are considered to be a sign of impaired fertility (e.g., Hare et al., 2006). Because we are interested in functional aspects of longevity, including fertility, it is consequent that prolonged calving intervals are not credited in the new routine genetic evaluation of longevity.

Age at first calving. Chapter 3 has shown that the correction for age at first calving in models for genetic evaluations of longevity should be reconsidered. In the prototype of the new routine evaluation, this effect was therefore dropped. It could be argued that age at first insemination and thus age at first calving is largely influenced by bodyweight development, which itself can be largely influenced by feeding level (as reviewed by Sejrsen and Purup, 1997), which is a management decision, and that age at first calving should therefore be corrected for in either form. For example, it could be corrected for its genetic component and then be used as a covariate in the genetic evaluation of longevity as proposed in chapter 3. However, we included a fixed effect of herd × year × season, which probably accounts for most of the management related fraction in the effect of age at first calving.

Herd size change. Another effect that was frequently corrected for in genetic evaluations of longevity is the effect of herd size change (e.g., Pasman and Reinhardt, 1999; Sewalem et al., 2007). Cows on shrinking dairy farms have a higher risk of culling compared to cows on a farm with stable size, and cows in expanding herds have a lower risk of culling (e.g., Vollema et al., 2000). Again, the fixed effect of herd × year × season is assumed to account sufficiently for this effect.

Functional longevity. Genetic evaluations of longevity are usually intended to reflect functionality of cows (as suggested by, e.g., Ducrocq et al., 1988b). An effect for milk yield, relative to the herd mean yield, was therefore adopted from the current model in order to correct for voluntary culling. However, milk yield is a fuzzy indicator of voluntary culling, because it is con-

founded with functionality: mastitis, for example, has a negative impact on milk yield which was estimated to be substantial in many studies (as reviewed by Seegers et al., 2003). If the cow was culled because of this mastitis, she would also show a lowered milk yield shortly before her culling. It is therefore desirable to exclude test day records directly prior to culling from the estimation of milk yield in order to reduce the confounding between milk yield and functionality. This is accounted for in the prototype of the new genetic evaluation of longevity by using the mean deviation from the herd mean in the period prior to the period under consideration, but the confounding cannot be removed completely: especially metabolic diseases, which are the main reason for disposal early in later lactations (Heise et al., 2016), have persistent negative effects on milk yield throughout the consequent lactation period (e.g., Rajala-Schultz et al., 1999). In these cases, it is almost impossible to distinguish between 'functional' (influenced by the disorder) and 'non-functional' effects of milk yield on longevity. Further, the potential confounding between functionality and milk yield must be expected to be greatest for early first lactation cows where only few test day records are known. Unfortunately, no better indicator exists for voluntary culling that is available on all cows considered in the routine genetic evaluation of longevity in German Holsteins.

Future research on genome-wide associations to longevity

This section arises from side results of chapter 5 and shows potential for future research. First, expectations about genomic inflation factors are briefly illustrated and then linked to observations from our genome-wide association study (**GWAS**) in chapter 5.

Genomic inflation factors (λ_{mean} and/or λ_{median} values) are often published together with GWAS (e.g., Zhang et al., 2016; Nayeri et al., 2017). The methodology behind them can be summarized following Yang et al. (2011b): the base assumption is that only few SNP-markers have a true association with the trait and most markers have not. If this were the case, the observed mean and median χ^2 values over all SNP-markers should be similar to those under the null hypothesis, where no marker is assumed to have a true association. The genomic control measures λ_{mean} and λ_{median} are the ratios of observed mean and median χ^2 values, divided by their expected counterparts under the null hypothesis. Genomic control measures can be inflated if population stratification is not eliminated properly. Yang et al. (2011b) argue that the assumption of only few SNP-markers being truly associated to the respective trait does not hold under polygenic inheritance. They formulate expectations for λ_{mean} and λ_{median}values for the case of a large number of causal variants and a quantitive trait. In the following, we concentrate on their expectation for λ_{mean}:

$$\lambda_{mean} \approx 1 + \frac{Nh^2\overline{r^2}\bar{s}}{n}$$

where N is the sample size, h^2 is the heritability, $\overline{r^2}$ is the average squared correlation coefficient between SNPs and causal variants, due to linkage disequilibrium (**LD**), $\bar{s}$ is the average number of SNPs being in LD with the causal variants and n is the total number of SNP-markers. With this formula, the expectation for λ_{mean} can be estimated straightforwardly for a specific sample and trait: N and n are initially known, h^2 can be easily estimated and was about 0.75 to 0.80 for the different survival traits in our case. For $\overline{r^2}\bar{s}$, Yang et al. (2011b) proposed to consider $\overline{r^2 s}$ instead, which could be estimated from the LD-structure in the sample. In chapter 5, we presented λ_{median} statistics, but λ_{mean} values were similar and ranged between 0.99 and 1.0 for survival of different periods.

In GWAS, potential population stratification can be taken into account by modeling a random effect for individuals, using the genetic relationship matrix as covariance matrix between the individuals (e.g., Zhang et al., 2016; Nayeri et al., 2017). We also used this method in chapter 5. For computational reasons, variance components are often only estimated once and SNP-markers are then tested against this null model. As shown by Yang et al. (2014), power for the detection of causal variants is decreased due to double counting if the genetic relationship matrix is built from all markers, including the tested SNP. They show that generating the genetic relationship matrix for the individual effects based on almost all SNPs, but excluding the SNP-marker to be tested and those under high LD with the tested SNP, yields higher power. This approach would

include the computation of the genetic relationship matrix and estimation of variance components for every tested SNP. In large samples with a high number of SNP-markers, this is seldomly computationally feasible and Yang et al. (2014) proposed to exclude the whole chromosome belonging to the marker under consideration from the computation of the genetic relationship matrix instead. Then, variance components must only be estimated for 29 null models when all autosomes in cattle are considered. Yang et al. (2014) call this method Leave-One-Chromosome-Out (**LOCO**). It is implemented in the GWAS software tool GCTA (Yang et al., 2011a).

We also analyzed our data from chapter 5 with the LOCO method, which was also used in another recent GWAS for longevity in dairy cattle (Zhang et al., 2016). Figure 6.7 shows a scatterplot of -$\log_{10}$(p-values), obtained from the run without LOCO against respective values from the run with LOCO for survival of L1.1. Figure 6.8 shows the respective Manhattan plots. It can clearly be seen that significances from the run with LOCO are by orders of magnitude higher than from the run without LOCO. Accordingly, λ_{mean} values ranged from 2.63 (L3.1) to 2.97 (L3.3) for the GWAS using the LOCO method. As mentioned above, λ_{mean} values from our GWAS without LOCO were close to 1 and thus below the expectation of being substantially larger than 1, derived from the argumentation of Yang et al. (2011b) and found in other GWAS for longevity (Zhang et al., 2016; Nayeri et al., 2017). This phenomenon should be further investigated: do the λ_{mean} values from the GWAS, performed with the LOCO method, meet our expectations basing on N, h^2 (which in our case is the proportion of genomic variance in the deregressed proofs), n, and $\overline{r^2 s}$, or are the observed χ^2 statistics inflated due population stratification effects on the chromosome which was left-out and which were therefore not accounted for? To answer this question, the genome-wide LD-structure could be analyzed in detail for our sample of 4,849 bulls. Outcomes of the formula of Yang et al. (2011b) could then be compared to observed λ_{mean} values from the different methods. Further, a GWAS could be performed where only the region around the tested SNP-marker is left-out instead of the whole chromosome. Other samples could also be taken into consideration. These results could contribute further knowledge about the genetic architecture of the trait longevity: if they confirmed that associations were estimated overly conservative in chapter 5, this would mean that more regions are significantly associated to longevity than mentioned there.

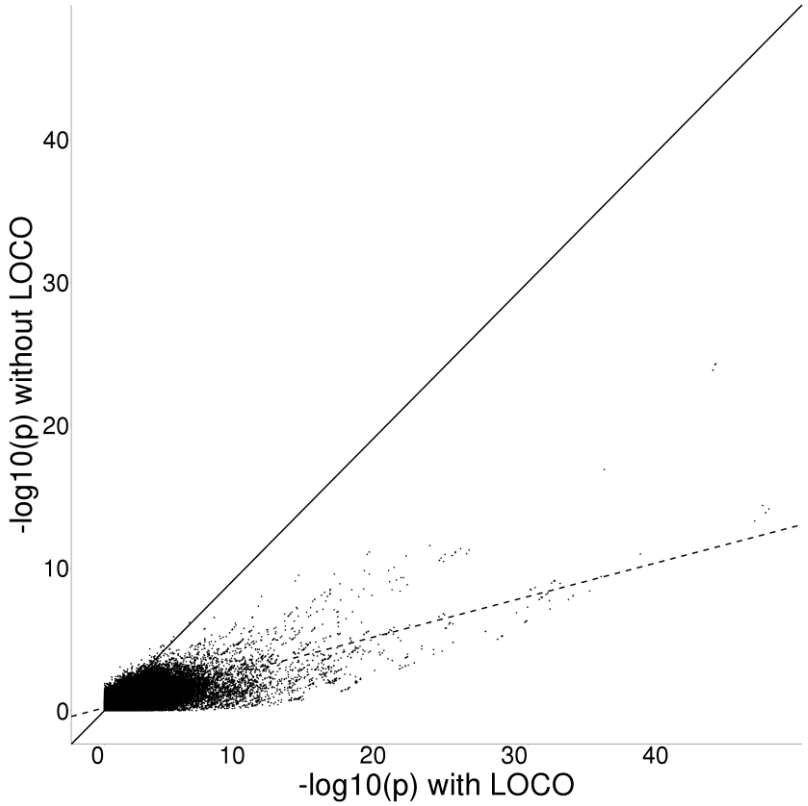


Figure 6.7: Scatterplot of -log10(p-values) from two GWAS runs on deregressed proofs for survival of L1.1 (first period of first lactation): without LOCO versus with LOCO. Bisecting line solid, regression line dashed.

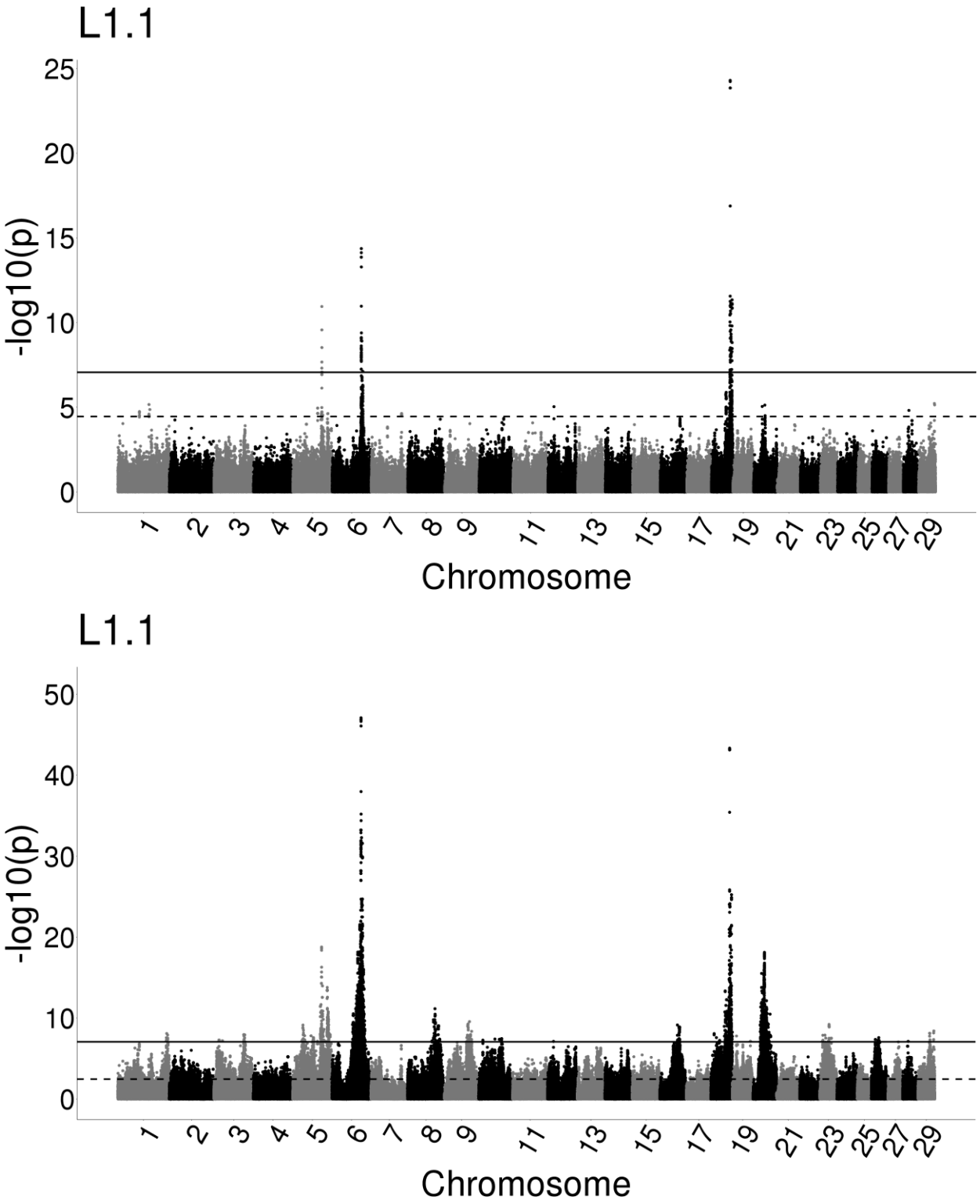


Figure 6.8: Manhattan plots from the single-marker GWAS for survival of L1.1. First plot shows results as presented in chapter 5, second plot was obtained from a GWAS with the LOCO method. The solid line marks the genome-wide Bonferroni threshold, the dashed line the false discovery rate, each at 0.05 significance level.

The future of the trait ‘longevity’ in dairy cows

Longevity, the time from first calving to culling, will remain an important trait in dairy cattle. It is an easy-to-interpret trait and can straightforwardly be used in economic considerations of dairy farmers. This trait is a natural index of many functional (and non-functional) other traits (e.g., Rajala-Schultz and Gröhn, 1999a; b; c) and is currently the only indicator for the overall functionality of a cow that is available on an almost population-wide scale. However, efforts are undertaken to measure functional traits directly on a growing number of cows (e.g., KuhVision: Reents et al., 2016). Further, a lot of research is going on in the fields of sensor techniques (as reviewed by, e.g., Rutten et al., 2013; Andriamandroso et al., 2016). With growing numbers of sensors on farms, growing quantities of data will be produced, including detailed information on different functional traits, e.g., health and fertility traits. To make maximum use of such data, it would be desirable to pool all data available, i.e., data from different sensor systems on different farms, conventional milk recording data and data from smaller phenotyping projects, and to analyze them jointly. This will be an extremely complex and challenging task. Deep learning algorithms could help with the extraction of precise (and potentially yet unknown) phenotypes from such complex data structures (as reviewed by LeCun et al., 2015). These algorithms could be trained from large but not comprehensive projects like KuhVision (Reents et al., 2016), which include precise phenotyping of health and other functional traits. The suggested procedure would increase the availability of precise phenotypes substantially with reasonable costs. These phenotypes could then serve as a basis for accurate genetic and genomic predictions for the range of low heritable functional traits for which conventional data collection (by manual documentation) is currently laborious and thus expensive (e.g., health traits). It would then be possible to select directly and effectively for functional traits instead of using longevity as an indicator trait. More precise and potentially new phenotypes could also help the farmers to make earlier and better decisions. The benefit would then be twofold.

Conclusions

Main results from this thesis:

1) It was shown that the genetic background of survival of different periods, defined within and across the first three lactations, is distinct but correlated. This was validated with a genome-wide association study which could help to further unravel the genetics of functional traits in future research projects.

2) The results of this project were implemented in a prototype version of a new routine genetic evaluation system for functional longevity in German Holstein cows. This prototype shows substantial improvement compared to the current evaluation system in terms of prediction bias. Small optimizations should be implemented and the results should be further analyzed.

References

Andriamandroso, A.L.H., J. Bindelle, B. Mercatoris, and F. Lebeau. 2016. A review on the use of sensors to monitor cattle jaw movements and behavior when grazing. BASE.

Bertilsson, J., B. Berglund, G. Ratnayake, K. Svennersten-Sjaunja, and H. Wiktorsson. 1997. Optimising lactation cycles for the high-yielding dairy cow. A European perspective. Livest. Prod. Sci. 50:5–13. doi:10.1016/S0301-6226(97)00068-7.

Charfeddine, N., and M.A. Pérez-Cabal. 2017. Effect of claw disorders on milk production, fertility, and longevity, and their economic impact in Spanish Holstein cows. J. Dairy Sci. 100:653–665. doi:10.3168/jds.2016-11434.

Ducrocq, V., R.L. Quaas, E.J. Pollak, and G. Casella. 1988a. Length of Productive Life of Dairy Cows. 1. Justification of a Weibull Model. J. Dairy Sci. 71:3061–3070. doi:10.3168/jds.S0022-0302(88)79906-3.

Ducrocq, V., R.L. Quaas, E.J. Pollak, and G. Casella. 1988b. Length of Productive Life of Dairy Cows. 2. Variance Component Estimation and Sire Evaluation. J. Dairy Sci. 71:3071–3079. doi:10.3168/jds.S0022-0302(88)79907-5.

Gianola, D. 1980. A Method of Sire Evaluation for Dichotomies. J. Anim. Sci. 51:1266–1271. doi:10.2527/jas1981.5161266x.

Gohary, K., M.W. Overton, M. Von Massow, S.J. LeBlanc, K.D. Lissemore, and T.F. Duffield. 2016. The cost of a case of subclinical ketosis in Canadian dairy herds. Can. Vet. J. 57:728–732.

Hare, E., H.D. Norman, and J.R. Wright. 2006. Trends in Calving Ages and Calving Intervals for Dairy Cattle Breeds in the United States. J. Dairy Sci. 89:365–370. doi:10.3168/jds.S0022-0302(06)72102-6.

Heise, J., Z. Liu, K.F. Stock, S. Rensing, F. Reinhardt, and H. Simianer. 2016. The genetic structure of longevity in dairy cows. J. Dairy Sci. 99:1253–1265. doi:10.3168/jds.2015-10163.

Holtsmark, M., B. Heringstad, and J. Ødegård. 2009. Predictive abilities of different statistical models for analysis of survival data in dairy cattle. J. Dairy Sci. 92:5730–5738. doi:10.3168/jds.2009-2132.

Interbull. 2016a. Code of Practice. Accessed March 28, 2017. http://www.interbull.org/ib/codeofpractice.

Interbull. 2016b. National Genetic Evaluation Forms Provided by Countries. Accessed December 13, 2016. http://interbull.org/ib/geforms.

Klein, J.P., and M.L. Moeschberger. 2003. Survival Analysis - Techniques for Censored and Truncated Data. Second Edition. Springer, New York.

Kossaibati, M.A., and R.J. Esslemont. 1997. The costs of production diseases in dairy herds in England. Vet. J. 154:41–51. doi:10.1016/S1090-0233(05)80007-3.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. Deep learning. Nature 521:436–444. doi:10.1038/nature14539.

Liu, Z., F. Reinhardt, A. Bünger, and R. Reents. 2004. Derivation and Calculation of Approximate Reliabilities and Daughter Yield-Deviations of a Random Regression Test-Day Model for Genetic Evaluation of Dairy Cattle. J. Dairy Sci. 87:1896–1907. doi:10.3168/jds.S0022-0302(04)73348-2.

Nayeri, S., M. Sargolzaei, M.K. Abo-Ismail, S. Miller, F. Schenkel, S.S. Moore, and P. Stothard. 2017. Genome-wide association study for lactation persistency, female fertility, longevity, and lifetime profit index traits in Holstein dairy cattle. J. Dairy Sci. 100:1246–1258. doi:10.3168/jds.2016-11770.

Pasman, E., and F. Reinhardt. 1999. Genetic Evaluations for Length of Productive Life of Holstein Cattle in Germany. Interbull Bull. No 21:56–59.

van Pelt, M.L., T.H.E. Meuwissen, G. de Jong, and R.F. Veerkamp. 2015. Genetic analysis of longevity in Dutch dairy cattle using random regression. J. Dairy Sci. 98:4117–4130. doi:10.3168/jds.2014-9090.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999a. Culling of dairy cows. Part I. Effects of diseases on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:195–208. doi:10.1016/S0167-5877(99)00046-X.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999b. Culling of dairy cows. Part III. Effects of diseases, pregnancy status and milk yield on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:295–309. doi:10.1016/S0167-5877(99)00047-1.

Rajala-Schultz, P.J., and Y.T. Gröhn. 1999c. Culling of dairy cows. Part II. Effects of diseases and reproductive performance on culling in Finnish Ayrshire cows. Prev. Vet. Med. 41:279–294. doi:10.1016/S0167-5877(99)00045-8.

Rajala-Schultz, P.J., Y.T. Gröhn, and C.E. McCulloch. 1999. Effects of Milk Fever, Ketosis, and Lameness on Milk Yield in Dairy Cows. J. Dairy Sci. 82:288–294. doi:10.3168/jds.S0022-0302(99)75235-5.

Reents, R., Z. Liu, D. Segelke, K.F. Stock, E. Pasman, S. Rensing, and F. Reinhardt. 2016. Setting up a female genomic reference population for German Holsteins. Page TS1.4 in Abstracts and Presentations 40th ICAR Biennial Session, Puerto Varas, Chile.

Rutten, C.J., A.G.J. Velthuis, W. Steeneveld, and H. Hogeveen. 2013. Invited review: Sensors to support health management on dairy farms. J. Dairy Sci. 96:1928–1952. doi:10.3168/jds.2012-6107.

Seegers, H., C. Fourichon, and F. Beaudeau. 2003. Production effects related to mastitis and mastitis economics in dairy cattle herds. Vet. Res. 34:475–491. doi:10.1051/vetres:2003027.

Sejrsen, K., and S. Purup. 1997. Influence of prepubertal feeding level on milk yield potential of dairy heifers: a review. J. Anim. Sci. 75:828–835. doi:/1997.753828x.

Sewalem, A., F. Miglior, G.J. Kistemaker, P. Sullivan, G. Huapaya, and B.J. Van Doormaal. 2007. Short Communication: Modification of Genetic Evaluation of Herd Life from a Three-Trait to a Five-Trait Model in Canadian Dairy Cattle. J. Dairy Sci. 90:2025–2028. doi:10.3168/jds.2006-719.

Silvestre, A.M., A.M. Martins, V.A. Santos, M.M. Ginja, and J.A. Colaço. 2009. Lactation curves for milk, fat and protein in dairy cows: A full approach. Livest. Sci. 122:308–313. doi:10.1016/j.livsci.2008.09.017.

van Soest, F.J.S., I.M.G.A. Santman-Berends, T.J.G.M. Lam, and H. Hogeveen. 2016. Failure and preventive costs of mastitis on Dutch dairy farms. J. Dairy Sci. 99:8365–8374. doi:10.3168/jds.2015-10561.

vit. 2016. Estimation of Breeding Values for Milk Production Traits, Somatic Cell Score, Conformation, Productive Life and Reproduction Traits in German Dairy Cattle. Accessed December 13, 2016. http://www.vit.de/fileadmin/user_upload/vit-fuers-rind/zuchtwertschaetzung/milchrinder-zws-online/Zws_Bes_eng.pdf.

Vollema, A.R., S.V.D. Beek, A.G.F. Harbers, and G.D. Jong. 2000. Genetic Evaluation for Longevity of Dutch Dairy Bulls. J. Dairy Sci. 83:2629–2639. doi:10.3168/jds.S0022-0302(00)75156-3.

Wilmink, J.B.M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livest. Prod. Sci. 16:335–348. doi:10.1016/0301-6226(87)90003-0.

Yang, J., S.H. Lee, M.E. Goddard, and P.M. Visscher. 2011a. GCTA: A Tool for Genome-wide Complex Trait Analysis. Am. J. Hum. Genet. 88:76–82. doi:10.1016/j.ajhg.2010.11.011.

Yang, J., M.N. Weedon, S. Purcell, G. Lettre, K. Estrada, C.J. Willer, A.V. Smith, E. Ingelsson, J.R. O'Connell, M. Mangino, R. Mägi, P.A. Madden, A.C. Heath, D.R. Nyholt, N.G. Martin, G.W. Montgomery, T.M. Frayling, J.N. Hirschhorn, M.I. McCarthy, M.E. Goddard, and P.M. Visscher. 2011b. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19:807–812. doi:10.1038/ejhg.2011.39.

Yang, J., N.A. Zaitlen, M.E. Goddard, P.M. Visscher, and A.L. Price. 2014. Advantages and pitfalls in the application of mixed-model association methods. Nat. Genet. 46:100–106. doi:10.1038/ng.2876.

Yazdi, M.H., P.M. Visscher, V. Ducrocq, and R. Thompson. 2002. Heritability, Reliability of Genetic Evaluations and Response to Selection in Proportional Hazard Models. J. Dairy Sci. 85:1563–1577. doi:10.3168/jds.S0022-0302(02)74226-4.

Zhang, Q., B. Guldbrandtsen, J.R. Thomasen, M.S. Lund, and G. Sahana. 2016. Genome-wide association study for longevity with whole-genome sequencing in 3 cattle breeds. J. Dairy Sci. 99:7289–7298. doi:10.3168/jds.2015-10697.

Acknowledgments

I would like to thank:

Prof. Dr. Henner Simianer. Thank you for supervising my PhD project and for supporting and motivating me! Your thoughts are always a valuable source of impulses.

Prof. Dr. Roel F. Veerkamp. Thank you for acting as my co-referee and for the good discussions we had at our annual meetings!

Fritz Reinhardt. Thank you for supervising my PhD project at vit! You always were like a father to me and my ideas.

Malena Erbe. Thank you for making me aware of my scientific interest and for all your good and helpful advices!

My **former colleagues** at vit. Your technical training laid the foundation of my abilities to deal with large data sets and your intense discussions gave me deep insights into the world of routine genetic evaluations. Thank you all!

My **colleagues** from the Animal Breeding and Genetics group at the University of Göttingen. Thank you for all the good discussions towards the end of my PhD project and thank you for welcoming me so kindly to Göttingen!

My **friends.** Thank you for the good times we have had!

My **parents-in-law.** Thank you for supporting Heinke and me with so many things! Thank you for all the time you spent with Noah to give us time to finish our theses!

My **parents.** Thank you for your love and patient support of all my professional and private projects!

My beloved **family.** Thank you, **Heinke**, my beloved wife, for your love and for your undeviating faith in me. Thank you for going all this way by my side! Thank you, **Noah**, my beloved son, for your love and your boisterous enthusiasm, which is so direct and pure!

Further publications

2016 **J Heise**, KF Stock, F Reinhardt & H Simianer. *Genetic relationships between age at first calving, its underlying traits and survival of Holsteins*. Page 375 in the Book of Abstracts of the 67th Annual Meeting of the European Federation of Animal Science (Belfast)

2015 **J Heise**, KF Stock, F Reinhardt & H Simianer. *Auswirkungen des Kalbeverlaufes auf Produktions- und funktionale Merkmale beim Milchrind.* DGfZ/GfT-Gemeinschaftstagung (Berlin)

2014 **J Wiebelitz**, F Reinhardt, Z Liu, KF Stock, M Erbe & H Simianer. *Genetic Evaluation of Survival Traits in German Holstein Dairy Cattle Using a Six-Trait Linear Model*. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production (Vancouver)

J Wiebelitz, F Reinhardt, H Täubert, Z Liu, M Erbe & H Simianer. *EBV trend validation for survival traits from a linear six-trait model in German dairy cattle*. Page 194 in the Book of Abstracts of the 65th Annual Meeting of the European Federation of Animal Science (Kopenhagen)

J Wiebelitz, Z Liu, F Reinhardt & H Simianer. *Möglichkeiten der Kombination von Zuchtwerten aus einem Mehr-Merkmals-Ansatz zu einem Gesamtzuchtwert für Nutzungsdauer*. DGfZ/GfT-Gemeinschaftstagung (Dummerstorf)

2013 **J Wiebelitz**, H Täubert, Z Liu, F Reinhardt, M Erbe & H Simianer. *Analyse der Abgangsursachen bei Holstein-Frisian Milchkühen*. DGfZ/GfT-Gemeinschaftstagung (Göttingen)

2012 **J Wiebelitz**, M Erbe & H Simianer. *Genauigkeit der Zuchtwertschätzung in unterteilten Populationen*. DGfZ/GfT-Gemeinschaftstagung (Halle/Saale)

Curriculum Vitae

Personal Data

Name	**Johannes Heise** né Wiebelitz
E- Mail adress	johannes.heise@uni-goettingen.de
Family status	married, one child
Date of Birth	January 17, 1987, in Bergisch Gladbach
Nationality	German

Education

Jan 2013 – current	**PhD student** Georg-August-University Göttingen, Germany
Oct 2010 – Nov 2012	**M.Sc. „Animal Sciences“** Georg-August-University Göttingen, Germany Thesis: *„Genauigkeit der genomischen Zuchtwertschätzung in unterteilten Populationen“*
Oct 2007 – Jul 2010	**B.Sc. „Animal Sciences“** Georg-August-University Göttingen, Germany Thesis: *„Inzucht beim Schleswiger Kaltblut“*

Work History

Oct 2016 – current	**Research assistant** Animal Breeding and Genetics, Department of Animal Sciences, Georg-August-University Göttingen, Germany
Jan 2013 – Sep 2016	**Research and development assistant** IT Solutions for Animal Production (vit) in Verden (Aller), Germany
Jun 2011 – Jul 2011	**Student assistant** Department of Crop Sciences, Georg-August-University Göttingen, Germany