



HAL
open science

Comparison between sire-maternal grandsire and animal models for genetic evaluation of longevity in dairy cattle populations with small herds

J. Jenko, G. Gorjanc, M. Kovac, Vincent Ducrocq

► **To cite this version:**

J. Jenko, G. Gorjanc, M. Kovac, Vincent Ducrocq. Comparison between sire-maternal grandsire and animal models for genetic evaluation of longevity in dairy cattle populations with small herds. *Journal of Dairy Science*, 2013, 96 (December), pp.8002-8013. 10.3168/jds.2013-6830 . hal-01189854

HAL Id: hal-01189854

<https://hal.science/hal-01189854v1>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Comparison between sire-maternal grandsire and animal models for genetic evaluation of longevity in a dairy cattle population with small herds

J. Jenko,*¹ G. Gorjanc,† M. Kovač,† and V. Ducrocq‡

*Agricultural Institute of Slovenia, Hacquetova ulica 17, 1000 Ljubljana, Slovenia

†Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Groblje 3, 1230 Domžale, Slovenia

‡INRA, UMR 1313 Génétique Animale et Biologie Intégrative, 78350 Jouy-en-Josas, France

ABSTRACT

Survival analysis techniques for sire-maternal grandsire (MGS) and animal models were used to test the genetic evaluation of longevity in a Slovenian Brown cattle population characterized by small herds. Three genetic models were compared: a sire-MGS model for bulls and an approximate animal model based on estimated breeding values (EBV) from the sire-MGS model for cows, an animal model, and an animal model based on the estimated variance components from the sire-MGS model. In addition, modeling the contemporary group effect was defined as either a herd or a herd-year (HY) effect. With various restrictions on the minimum HY group size (from 1 to 10 cows per HY), changes in estimates of variance components, and consequently also in EBV, were observed for the sire-MGS and animal models. Variance of contemporary group effects decreased when an HY effect was fitted instead of a herd effect. In the case of a sire-MGS model, estimates of additive genetic variance were mostly robust to changes in minimum HY group size or fitting herd or HY effect, whereas they increased in the animal model when HY instead of herd effects was fitted, possibly revealing some confounding between cow EBV and contemporary group effect. Estimated heritabilities from sire-MGS models were between 0.091 and 0.119 and were mainly influenced by the restriction on the HY group size. Estimated heritabilities from animal models were higher: between 0.125 and 0.160 when herd effect was fitted and between 0.171 and 0.210 when HY effect was fitted. Rank correlations between the animal model and the approximate animal model based on EBV from the sire-MGS model were high: 0.94 for cows and 0.93 for sires when a herd effect was fitted and 0.90 for cows and 0.93 for sires when an HY effect was fitted. Validation performed on the independent validation data set revealed that the correlation between sire EBV and daughter survival were slightly higher with

the approximate animal model based on EBV from the sire-MGS model compared with the animal model. The correlations between the sire EBV and daughter survival were higher when the model included an HY effect instead of a herd effect. To avoid confounding and reduce computational requirements, it is suggested that the approximate animal model based on EBV from the sire-MGS model and HY as a contemporary group effect is an interesting compromise for practical applications of genetic evaluation of longevity in cattle populations.

Key words: longevity, survival analysis, genetic evaluation, dairy cattle

INTRODUCTION

Longevity in dairy cattle is most commonly defined as the length of productive life (Ducrocq, 2005; Terawaki et al., 2006; Chirinos et al., 2007). Common approaches for the analysis of such data are survival analysis (Forabosco et al., 2009), due to a finer time scale (in days rather than in months, years, or lactations), the ability to treat living cows as censored, and the effective handling of time-dependent explanatory variables (Smith and Quaas, 1984; Ducrocq, 1987). Genetic evaluation of longevity based on survival analysis is commonly performed with sire or sire-maternal grandsire (MGS) models (Forabosco et al., 2009) as implemented in the Survival Kit software (Mészáros et al., 2013). This software implements a Bayesian approach to estimate parameters of frailty (i.e., mixed survival) models, assuming a conjugate log-gamma or nonconjugate normal prior distribution for the frailty (random) terms. Inferences on random hyperparameters are drawn from their marginal posterior density using Laplacian approximation (Ducrocq and Casella, 1996; Mészáros et al., 2013).

The main reason for using sire and sire-MGS versus animal models is computational. The joint maximization of a complex nonlinear function with many parameters such as the one used in survival mixed models is very demanding. Efficient optimization of such functions requires repeated computation and storage

Received March 19, 2013.

Accepted August 20, 2013.

¹Corresponding author: janez.jenko@kis.si

of a Hessian matrix, which is usually denser than in standard linear mixed models. The use of AI in dairy cattle leads to a much smaller number of sires than the number of recorded animals, which makes sire and sire-MGS models the only computationally tractable ones in very large populations, in contrast to animal models.

To obtain EBV also for cows in the case of large populations, a 2-step approximate procedure was developed using results obtained from sire-MGS models (Ducrocq, 2001). This approximate animal model procedure is incorporated in the Survival Kit software (Mészáros et al., 2013). In the first step, a sire-MGS survival model is applied to obtain sire EBV. Then, in the second step, cows are evaluated assuming that other effects are known (Ducrocq, 2001).

Even if computationally possible, the application of an animal model for genetic evaluation of longevity in populations with small contemporary groups is questionable because of the lack of information for decisive separation of genetic and environmental effects. Namely, the amount of information per individual cow is limited in the presence of a single (potentially censored) record per cow for a lowly heritable trait. Therefore, modeling the contemporary group effect and its interplay with genetic effects is of considerable importance. Some authors distinguished variation in culling policy only between herds (Potočnik et al., 2011). However, within-herd culling policy also varies between years and between seasons within year, leading to the standard use of time-dependent herd-year (HY; e.g., Dürr et al., 1999; Terawaki et al., 2006) or herd-year-season (e.g., Ducrocq, 2005; Sewalem et al., 2005; Chirinos et al., 2007) interaction effects in survival analysis models. Modeling contemporary groups in a time-dependent manner within herd can better account for the conditions around the time of culling, but no clear guidelines exist to determine which approach is favorable in the case of limited amount of information per contemporary group.

In Slovenia, cattle populations are small with rather small herds, representing an example where sire-MGS and animal models can be tested and the interaction with contemporary group definition can be explored. The objective of this research was therefore to compare the application of different genetic models and different definitions of contemporary groups for genetic evaluation of longevity of Slovenian Brown cattle.

MATERIALS AND METHODS

Data

Data on longevity defined as time from the first calving until culling or the sixth calving of Brown cows

between January 1998 and December 2008 in Slovenia were used. In the data preparation procedure, records from cows <20 mo and >48 mo at the time of first calving or with the missing date of the first calving were discarded. Lengths of productive life beyond lactation 5 were censored to concentrate the study on the risk of being culled early in life, because culling of older cows may involve reasons different from those in younger cows and early cullings are the most expensive. Censoring was also applied to cows changing herds during their lifespan, with censoring occurring at the date of last milk recording in their first herd. The final data set included 42,120 cows with phenotype data. A relatively small herd size (from 6.7 on average in 1999 to 8.7 Brown cows per herd in 2008) is an intrinsic feature of herds of the Brown cattle breed in Slovenia. Therefore, a series of 5 data sets were prepared for the comparison of estimated variance components under different restrictions on minimum number of cows per HY group (1, 2, 3, 5, or 10). A pedigree file with 5 generations was created for each data set separately. To test the predictive ability of the models, the data were divided into a training set for the estimation of variance components and the prediction of breeding values and a validation data set for testing the predictive ability of survival in cows. The training data set comprised 90% of all the cows with records (37,908 cows), whereas the remaining 10% (4,212 cows) were used for validation. Cows in the validation data set came from larger herds (≥ 10 cows per HY class) and were second-crop progeny of 63 sires.

Base Model

The analysis was based on a proportional hazards model with Weibull distributions:

$$t_i \sim \text{Weibull}(\lambda_i, \rho_i), \quad [1]$$

where t_i is the longevity record of the i th cow, and λ_i and ρ_i are scale and shape parameters, respectively, of the Weibull distributions that were modeled. Proportional hazards provide a flexible framework to describe hazard functions $h(t_i)$ separately for the general aging process—the so-called baseline hazard function, $h(t_i)_0$ —and the multiplicative effect of a function of genetic and nongenetic factors $f(t_i)$:

$$h(t_i) = h(t_i)_0 \times f(t_i). \quad [2]$$

The baseline hazard function was modeled in a piecewise fashion, assuming a different Weibull hazard function for each combination of lactation number and stage of lactation, as in Terawaki et al. (2006). The

cutpoints defining stages of lactation were chosen according to the observed changes in the estimated hazard distribution derived from the Kaplan-Meier estimate of the raw survivor function (Kaplan and Meier, 1958). Cutpoints were set at d 0, 60, 270, and 380 d of lactation, and the day of drying off. Each piecewise baseline hazard function $h(t_i)_{0,p}$ was of the Weibull form [1]:

$$h(t_i)_{0,p} = \lambda_p \rho_p (\lambda_p t_i'')^{\rho_p - 1}, \quad [3]$$

with specific values of λ_p and ρ_p for each of the 25 (5 lactations \times 5 stages) periods generated by the cutpoints. In [3], t_i'' is the number of days since the previous calving (e.g., since the second calving, for any period covering the second lactation).

To focus on functional rather than true longevity, the effect of (relative) milk production was included to correct for culling due to low production, which is the major source of voluntary culling (e.g., Vukasinovic et al., 1995; Dürr et al., 1999). This effect was modeled through the definition of time-dependent groups $m_j(t_i)$ with j (1 to 10) classes on peak milk yield m (defined as the highest value out of the first 2 milk recordings within the first 120 d of lactation). Peak milk yields were compared within calving year on a whole-population basis. To test the influence of herd size on risk of culling due to low production, the comparison was done for 5 groups of herd size separately for the first and later lactations. Group sizes for primiparous cows were: 1 to 3, 4 to 6, 7 to 9, and >9 , and for multiparous they were: 1 to 5, 6 to 10, 11 to 15, and >15 cows in a herd. According to the mean and standard deviation of peak milk yield, limits of 10 deciles (groups) were calculated. The change in milk yield group during the cow's herdlife was assumed to occur on the day of peak yield. Primiparous cows were included in a special "dummy" group until the peak yield of the first lactation was obtained.

Six groups k (1 to 6) of change in herd size s , $s_k(t_i')$, were defined according to the difference in maximum number of recorded cows during the first 3 recording days of each year. Cows from herds decreasing in size by more than 30% were in group 1, cows from herds decreasing from 10 to 30% were in group 2, cows in stable herd size group (changes from -10 to 10%) were in group 3, cows from herds increasing from 10 to 30% were in group 4, cows from herds increasing by more than 30% were in group 5, and finally, a "dummy" group 6 was created for small herds with fewer than 5 cows. The change between groups of herd size variation occurred on January 1 of each year.

For the effect of region, 3 territories were defined that were distinguished mainly on the basis of the agricultural and climatic conditions. The change in region r by year-season group yp , $rypl_l(t_i')$, with l (1 to 132) classes occurred on January 1, April 1, July 1, and October 1 of each year.

The time-dependent contemporary group effect $c_m(t_i')$ or time-independent contemporary group effect c_m and additive genetic effect a_n were modeled according to different models as described in the next section. In all cases, these effects were modeled assuming a log-gamma distribution (for computational convenience, e.g., Ducrocq and Casella, 1996) for the contemporary group effect:

$$c_m \sim \text{Log-Gamma}(\gamma, \gamma), \quad [4]$$

and a normal distribution for additive genetic effect:

$$\mathbf{a} \sim N(0, k\mathbf{A}\sigma_a^2), \quad [5]$$

where the 2 hyperparameters of the log-gamma distribution are equal to γ (i.e., the mean is equal to 1 and the variance is equal to $\sigma_c^2 = \psi^{(1)}(\gamma)$, where $\psi^{(1)}$ is the trigamma function), \mathbf{A} is the numerator relationship matrix between additive genetic values \mathbf{a} , and k is the proportion of additive genetic variance σ_a^2 , with different proportions according to the models described in the next section. The final model equation was

$$f(t_i'') = \exp \left\{ m_j(t_i) + s_k(t_i') + ryp_l(t_i') + c_m(t_i') + a_n \right\}. \quad [6]$$

One complex aspect of the model is that 3 time scales are simultaneously considered: the number of days since the first calving (t_i), calendar time (t_i') "controlling" the definition of year or season, and the number of days since the previous calving (t_i'') to model a baseline hazard that follows the cyclic pattern across lactations.

Alternative Models

Alternative definitions of contemporary group and additive genetic effect were tested. Two alternatives were tested for the contemporary group effect defined as either the time-independent effect of herd or the time-dependent effect of HY, with changes occurring on January 1 of each year. In both cases, restrictions on minimum number of records were applied on the HY group. If the HY group size was below a given

threshold (1, 2, 3, 5, or 10 cows) in a certain year, all length of productive life records of cows in that herd were considered censored from that year onward.

Two models were tested to estimate the additive genetic effect of bulls and cows. The first model was a sire-MGS model, with the additive genetic effect of bulls represented as

$$a_n = a_o + \frac{1}{2} a_p, \quad [7]$$

where a_o is the additive genetic effect of the o th sire and a_p is additive genetic effect of the p th MGS of the n th cow. The relationship matrix was built for all sires based on the available male pedigree. The heritability for this model can be calculated as

$$h^2 = \frac{4\sigma_s^2}{\frac{5}{4}\sigma_s^2 + \sigma_c^2 + 1}, \quad [8]$$

where σ_s^2 is the additive sire genetic variance; the additional 1/4 added to the denominator of [8] corresponds to the maternal grandsire genetic variance; σ_c^2 is the variance of contemporary group effects, and 1 is the standard environmental variance. Expression [8] is the theoretical heritability that would be obtained in the absence of any censoring. It is higher than when censoring is accounted for, when the standard environmental variance is $1/p$, with p the proportion n of uncensored records (Yazdi et al., 2002; Mészáros et al., 2010).

With the sire-MGS model, cow EBV can be approximated by her (male) pedigree index derived from the sire-MGS model. An alternative is to derive the cow EBV using an approximation of an animal model as proposed in Ducrocq (2001).

The second model was an animal model with the inclusion of the additive genetic effect of the n th cow (a_n) directly. With the relationship matrix built for all animals in the study, recorded cows and ancestors in pedigree, the heritability for this model corresponded to

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + 1}, \quad [9]$$

where σ_a^2 is the additive genetic variance, σ_c^2 is the variance of contemporary group effects, and 1 is the standard environmental variance. Two variants of the animal model were tested, using for σ_a^2 and σ_c^2 the value estimated from either the animal model or the sire-MGS model. The different estimates of cow breeding value were compared, considering only cows that had both their sire and MGS known.

Computation

The Survival Kit v6 package (Mészáros et al., 2013) was used to estimate parameters in all models following the theoretical development by Ducrocq and Casella (1996). Approximate animal model solutions for cows were obtained as a combination of EBV from the sire-MGS model for sires and an approximation of remaining part of additive genetic effect for cows, assuming that all other effects are known and equal to their sire-MGS model estimates (Ducrocq, 2001). Estimates of variance components and correlations between evaluated breeding values were compared for all models and data sets. Predictive accuracies of the models were tested for sire-MGS and animal models separately and for bulls only. Two approaches were used to show the predictive ability of the models. First, survival of the second-crop daughters was correlated with the EBV of sires and, second, sires were grouped into 4 equal-sized groups based on their EBV in the training data set. The first group consisted of the 25% sires that had the highest EBV and so on, to the last group where the 25% sires with the lowest EBV were included. Finally, the differences in survival were visualized and compared between groups with the Kaplan-Meier survival curve.

RESULTS

Data Structure

The number of animals in each data file and the associated pedigree file are presented in Table 1, where the percentages of left-truncated (i.e., with first calvings before January 1, 1998) and right-censored data are also shown. Setting restrictions on the minimum number of cows in each HY inevitably led to a decrease in the number of cows in the analyses. In the case of a minimum of 5 cows per HY, only 13.2% (32,907 vs. 37,908 with no restrictions) of cow records were discarded from the data. The proportion of discarded records reached 50.0% (18,954 vs. 37,908 with no restrictions) when a minimum of 10 cows per HY was required. Setting restrictions on the minimum number of cow records per HY led to a larger decrease in number of herds (34.6% with 5 and 72.8% with 10 records at minimum per HY) and HY (43.2% with 5 and 77.6% with 10 records at minimum per HY). For the sire-MGS model, the decrease in the number of sires in the pedigree was 2.5% and 9.0% when setting a minimum of 5 or 10 records per HY, respectively. For the animal model, the decrease in number of animals in pedigree closely followed the reduction in number of cows: 11.0 or 42.2% with a restriction at 5 or 10 records at minimum per HY.

Table 1. Data characteristics under different restrictions on the minimum number of records in herd-year group

HY size ¹	Cow records	Left truncated (%)	Right censored (%)	Herds	Herd-years	Animals in pedigree	
						Sire-MGS model ²	Animal model
1	37,908	21.2	49.7	2,401	18,689	886	57,005
2	37,683	21.8	50.0	2,262	16,936	885	56,738
3	37,084	22.8	50.8	2,149	14,979	885	55,961
5	32,907	23.8	51.9	1,571	10,607	864	50,741
10	18,954	25.0	53.1	654	4,183	806	31,796

¹Minimum number of records per herd-year group.

²Sire-maternal grandsire model.

Computing Requirements

Computing time for the evaluation with the sire-MGS model was around 10 min for each of the models. With the animal model, computing time ranged from 40 min for the model with the highest restriction on a minimum number in HY class to approximately 2 h with the model with no restriction on a minimum HY class size. Sire-MGS models used approximately 0.16 GB, whereas animal models used around 25 GB of computer memory. This illustrates the unfeasibility of animal model evaluations with the Survival Kit for large populations.

Variance Components

With sire-MGS models, variance of contemporary group effects decreased with greater restrictions on minimum HY group size (Table 2). The decrease was highest between no restriction (0.296 for herd effect and 0.211 for HY effect) and at a minimum of 2 records per HY (0.193 for herd effect and 0.162 for HY effect). Surprisingly, the herd variance was always higher than the HY variance.

Changes in additive genetic variance with respect to restrictions on HY group size were small with sire-MGS models. There was a slight decrease (from 0.129 to

0.119 with herd effect and from 0.123 to 0.110 with HY effect) with the first restriction on herd size (minimum size of 2). Then, estimates were relatively stable up to the restriction of a minimum of 10 cows per HY, when slightly higher additive genetic variance was obtained (0.145 with herd effect and 0.129 with HY effect). Heritabilities were relatively stable to changes in variance of contemporary group effects and additive genetic variance. Estimates for heritabilities ranged from 9.1 to 11.9% (Table 2).

With the animal models, the contemporary group variance decreased with higher restriction on minimum HY group size in similar way as with sire-MGS models (Table 3). The herd variance was between 0.159 and 0.286 and comparable with estimates using the sire-MGS model. The HY variance was much smaller and ranged from 0.089 to 0.113.

Additive genetic variance estimates for the animal models were all higher compared with sire-MGS models. This was especially pronounced for the animal model with the HY effect. As a result, heritabilities in models with the herd effect were between 0.125 and 0.160, whereas in models with an HY effect, heritabilities varied between 0.171 and 0.210.

Variance component estimates between sire-MGS and animal models differed depending on the way contemporary group effect was defined. Estimates of herd

Table 2. Estimates of variance components obtained using a sire-maternal grandsire model with contemporary group defined as herd or herd-year combination¹

HY size ²	Herd			Herd-year		
	σ_c^2	σ_a^2	h^2	σ_c^2	σ_a^2	h^2
1	0.296	0.129 ± 0.020	0.096	0.211	0.123 ± 0.020	0.098
2	0.193	0.119 ± 0.019	0.097	0.162	0.110 ± 0.018	0.092
3	0.186	0.117 ± 0.019	0.096	0.159	0.108 ± 0.018	0.091
5	0.164	0.119 ± 0.020	0.099	0.146	0.108 ± 0.019	0.092
10	0.172	0.145 ± 0.028	0.119	0.134	0.129 ± 0.026	0.110

¹Where σ_c^2 = contemporary group variance; σ_a^2 = additive genetic variance and its standard error of estimation, $\sigma_a^2 = 4\sigma_s^2$, where σ_s^2 is sire additive genetic variance; h^2 = heritability.

²Minimum number of records per herd-year group.

Table 3. Estimates of variance components obtained using an animal model with contemporary group defined as herd or herd-year combination¹

HY size ²	Herd			Herd-year		
	σ_c^2	σ_a^2	h^2	σ_c^2	σ_a^2	h^2
1	0.286	0.199 ± 0.024	0.134	0.113	0.295 ± 0.031	0.210
2	0.182	0.168 ± 0.022	0.125	0.095	0.229 ± 0.029	0.173
3	0.177	0.169 ± 0.022	0.125	0.095	0.227 ± 0.028	0.172
5	0.159	0.182 ± 0.024	0.136	0.090	0.225 ± 0.029	0.171
10	0.175	0.223 ± 0.033	0.160	0.089	0.232 ± 0.036	0.176

¹Where σ_c^2 = contemporary group variance; σ_a^2 = additive genetic variance and its standard error of estimation; h^2 = heritability.

²Minimum number of records per herd-year group.

variance were stable for sire-MGS and animal models, with a maximum change of 6% (from 0.193 to 0.182) with a minimum of 2 cows per HY class. On the other hand, HY variance was from 1.5 to 1.9 times higher with the sire-MGS model than with the animal model. Additive genetic variance estimates were higher with the animal model: 1.4 to 1.5 times higher for models with a herd effect and 1.8 to 2.5 times higher for models with an HY effect. As a result, heritability estimates were also higher with the animal model: 1.3 to 1.4 times higher for the model with herd effect and 1.6 to 2.1 times higher for the model with HY effect.

EBV

The average rank correlations between the EBV derived from the pedigree index based on the sire-MGS model and animal model EBV were 0.85 when a herd effect was fitted and 0.77 when an HY effect was fitted (Table A1 in the Appendix), denoting large reranking of cows. Correlations were the highest for restrictions to a minimum of 2 and 3 cows on HY group size when a herd effect was included and to a minimum of 10 cows when an HY effect was considered.

Rank correlations between the approximate animal model based on EBV from sire-MGS model with the animal model EBV were larger (Table 4) than with the pedigree index (Table A1 in the Appendix). In general, they were larger for sires (from 0.92 to 0.95) than for cows (from 0.88 to 0.94). This can be explained by the greater reliability of sire EBV. Similarity between cow EBV from an animal model and its approximation based on sire-MGS model was lower when an HY effect was considered instead of a herd effect. The reduction in rank correlation due to this change in the model was, on average, 0.04 for cows. This could be related to the large difference in variance component estimates between the 2 models. Whatever the type of contemporary group effects, higher restrictions on the number of cows per HY increased the correlation between EBV

for cows when HY effect was considered. The opposite was observed for sires.

To assess the effect of using quite different variances, the same rank correlations were computed when the animal model was implemented, assuming that the correct variance component estimates are those derived from the sire-MGS model (Table A2 in the Appendix). This time, correlations of 0.95 or greater were found in nearly all cases, underlying the similar EBV between the 2 models. A slight reduction was found only for sires in the case of higher restrictions on HY group size.

Animal model EBV when variance components estimates were obtained either from the animal or the sire-MGS model were compared. Rank correlations were almost unity for both cows and sires when a herd effect was fitted (on average 0.99 for cows and sires, Table A3 in the Appendix). As expected, given the change in variance estimates, the correlations were lower when an HY effect was fitted, but were still relatively high (on average, 0.97 for cows and 0.98 for sires). When restriction on HY group size was high, the difference in correlations was even smaller. These results illustrate again the consequences of the likely underestimation of HY variance and overestimation of the additive genetic variance with animal model due to confounding.

Table 4. Rank correlations for cow and sire EBV from an animal model and approximate animal model based on EBV from the sire-maternal grandsire model when contemporary group is defined as herd or herd-year combination

HY size ¹	Cows		Sires	
	Herd	Herd-year	Herd	Herd-year
1	0.94	0.88	0.94	0.95
2	0.94	0.90	0.94	0.93
3	0.94	0.90	0.94	0.93
5	0.94	0.91	0.93	0.92
10	0.94	0.92	0.92	0.92
Average	0.94	0.90	0.93	0.93

¹Minimum number of records per herd-year group.

Table 5. Rank correlations for 806 sire EBV from sire-maternal grandsire model and for 31,796¹ cow EBV obtained from an animal model with different minimum requirements on herd-year group size and contemporary group defined as herd or herd-year

Item	HY size X ²				
	1	2	3	5	10
Herd X vs. herd-year X					
Sires	0.94	0.96	0.96	0.96	0.97
Cows	0.91	0.94	0.94	0.95	0.96
Herd X vs. herd 10					
Sires	0.83	0.84	0.83	0.88	—
Cows	0.86	0.86	0.86	0.89	—
Herd-year X vs. herd-year 10					
Sires	0.76	0.81	0.80	0.86	—
Cows	0.83	0.86	0.86	0.88	—

¹Animals (cows and sires) must be present in the model with a minimum of 10 cows in herd-year group.

²Minimum number of records per herd-year group.

Rank correlations between EBV obtained from models fitting either a herd or an HY effect under various restrictions on HY group size are shown in Table 5 for sires (EBV from a sire-MGS model) and for cows (EBV from an animal model). Only animals included in the model with the restriction of a minimum of 10 cows in HY class were included. Within the same group size, rank correlations were always high (at least 0.94) for sires and increased when the minimum HY group size increased. Corresponding values were somewhat lower for cows but still reasonably high (at least 0.91). When EBV for the same model were compared under different HY group size restrictions, substantial reranking occurred (Table 5). Correlation between EBV under increasing minimum HY group sizes (from 1 to 5) and EBV under the highest restriction (at least 10 cows per HY) were always ≤ 0.89 . These correlations were lower for sires than for cows, which can be related to the loss of information from the excluded daughters of a particular sire. A lower correlation between EBV was found for the model including an HY effect than with a herd effect. This indicates sensitivity of genetic evaluation to minimum HY group size, especially when an HY effect is fitted.

Model validation based on daughters from 63 sires (Figure 1) showed that the correlations between sire EBV and daughter survival were systematically better for models with an HY effect than for models with a herd effect. With higher restriction in herd size, the predictive ability was decreasing for an early survival (183 d; i.e., half a year after first calving). For later survival (1, 2, or 3 yr after first calving), an increase in correlation was observed for the highest restriction in herd size. On the other hand, a decrease was found for the highest restriction in case of survival to 4 yr after first calving. When comparing different model types, on average a slightly higher (+0.01) correlation was ob-

served for sire-MGS models. On average, correction for an HY random effect increased the correlation between the sire-predicted breeding value and observed survival by 0.04 compared with models where only a herd effect was accounted for.

The raw survivor curves for the second-crop daughters from the validation data set (Figure 2) showed an absolute difference of 0.14 in survival between the extreme groups 5 yr after first calving. When a herd effect was applied, survival of daughters from bulls in the second quartile was the highest for animal models. However, survival of daughters from the sires from first and second quartiles was similar for animal models with HY effect and sire-MGS model, whatever the contemporary group effect.

DISCUSSION

An essential objective of genetic evaluations is to produce EBV that allow the ranking of males and females as correctly as possible. Some of our results using theoretically more-appropriate models were unexpected, and in some instances, counterintuitive. For example, the fact that estimated herd variance was always higher than HY variance was not anticipated, because HY groups are intuitively associated with a finer description of the within-herd heterogeneity in culling policy over years. Nevertheless, measures of predictive ability clearly showed that the models with time-dependent HY effects were better. Although the variance explained by smaller contemporary group effects was lower, the correlation between sire EBV and daughter survival was systematically higher and the raw survival curves from second-crop daughters of groups of bulls based on their EBV ranking was more appropriate. This shows that the proportion of the variance explained with varying contemporary group

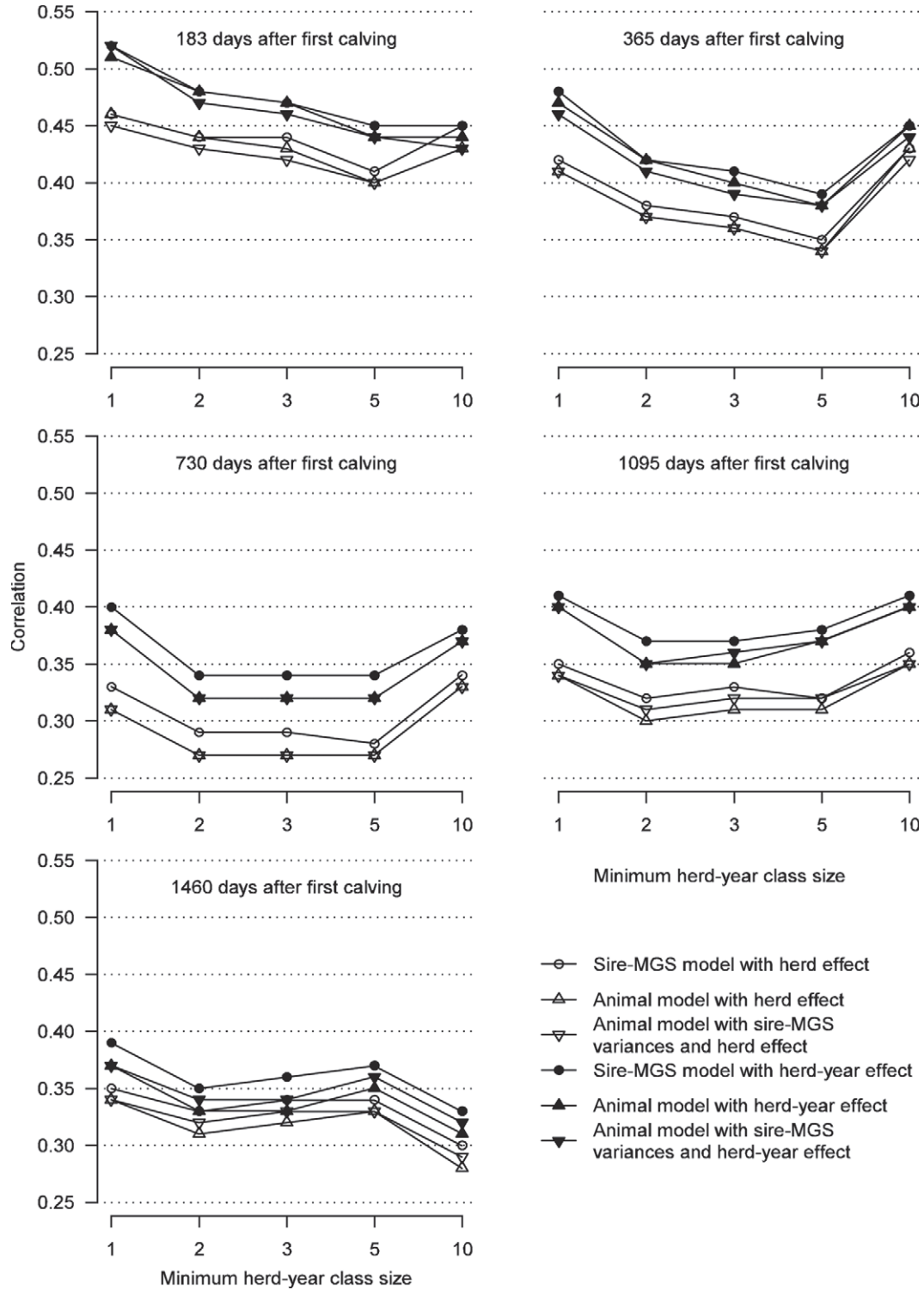


Figure 1. Correlation coefficients between sire EBV and observed survival of second-crop daughters from the validation data set when EBV were obtained from either a sire-maternal grandsire (MGS) model, an animal model, or an animal model with genetic variance derived from the sire-MGS variances. The contemporary group effect is either a herd or a herd-year effect.

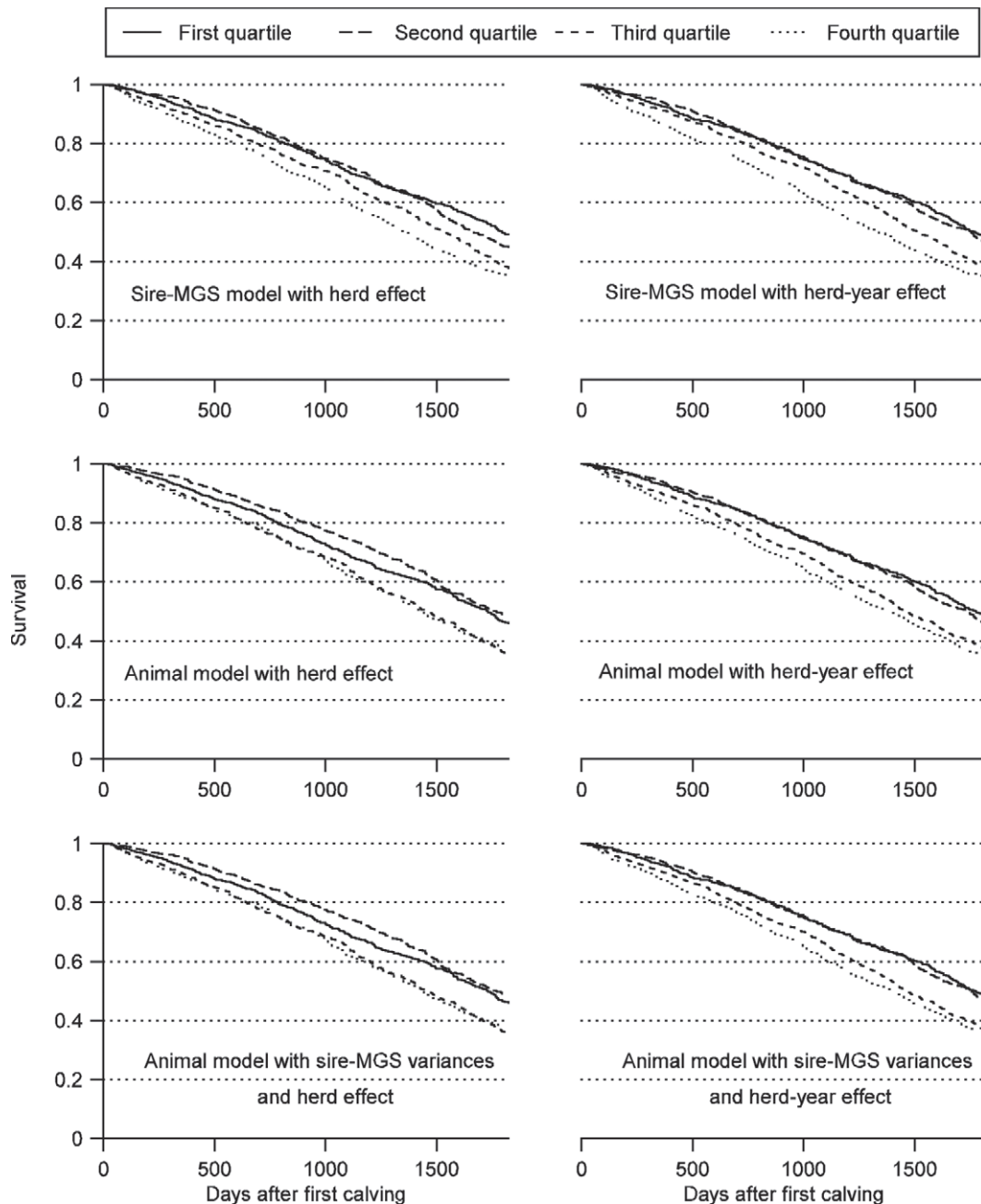


Figure 2. Raw survivor curves of the second-crop daughters from the validation data set, grouped into quartiles based on the EBV for length of productive life of their sire, computed under the following conditions: no restriction on herd-year group size, with either a herd or a herd-year contemporary effect and with a genetic part described by either a sire-maternal grandsire (MGS) effect, an animal effect, or an animal effect using a variance derived from a sire-MGS model.

effects does not reveal the real improvement in survival prediction when confounding with other components (genetic effects or residual) is likely. Compared with the results from Holtsmark et al. (2009), the obtained correlations between second-crop survival and sire EBV were lower, which may be the result of much smaller data set used in this study.

Modeling the additive genetic component of a trait is traditionally considered to be better with animal models: animal information from both the paternal and maternal sides of the pedigree as well as the individual (animal) Mendelian sampling term are taken into account. In contrast, in sire-MGS models, maternal information is only through the MGS, and the cow's Men-

delian sampling term is ignored. Consequently, only a part of the additive genetic variance ($0.3125 = 1/4 + 1/16$) is captured, with the remainder (0.6875) being included in the residual term of the model that follows a censored extreme value distribution. Damgaard et al. (2003) considered that sire or sire-MGS survival models are inconsistent with the additive genetic infinitesimal model by construction. However, some authors (e.g., Meuwissen et al., 2002) found very little effect due to this inconsistency, which was the source of a controversy (see Ducrocq, 2006). In dairy cattle, AI bulls have many daughters, leading to precise estimates of their breeding values, even for lowly heritable traits, whereas cows have only single observation and very few relatives. Given the usually low heritability of longevity, the contribution of (potentially censored) daughter performances to the breeding value of their dam is very limited. In other words, the longevity breeding value of a cow is mainly estimated using information from her male ancestors and her own (possibly censored) length of productive life.

Korsgaard et al. (1998) suggested the use of an alternative model where an extra residual term is explicitly included to account for the part of additive genetic variance ignored in sire-MGS model. However, estimation of this extra term and of its characteristics (e.g., variance) is very imprecise and heavily dependent on the a priori distribution assumed (Ducrocq, 2006), to such a point that its use is not recommended. Our situation makes the identifiability problem even worse because of the small HY groups. Results of variance component estimation suggest a large confounding effect between the additive genetic effect, the contemporary group effect, and the residual. Indeed, the high heritability estimates for longevity obtained with an animal model (up to 21%) are rather unusual. Assuming a too-large heritability would lead to overestimated EBV reliabilities. Most heritability estimates obtained from sire-MGS model were reported to be below 0.10 in longevity studies (Vollema, 1998; Chirinos et al., 2007; Mészáros et al., 2008), although some authors reported (theoretical) heritability up to 0.18 (Roxström and Strandberg, 2002; Potočnik et al., 2011). In our case, increased genetic variance with an animal model was compensated for by a large decrease in HY variance. The data structure may not be adequate to properly disentangle these 2 effects when genetic and HY variances are estimated. On the other hand, variance components appeared more stable with the sire-MGS model.

Another difficulty encountered with the animal survival model is its computational limitation. The repeated computation and inversion of a nearly dense matrix of a size larger than the total number of fixed

and random effects (Ducrocq and Casella, 1996) hinders the use of animal models for routine genetic evaluation and even more for the estimation of dispersion parameters. Approximations are therefore needed. Correlations between the animal model EBV and their approximation were much higher than those obtained with the pedigree estimate (Table A1 in the Appendix). However, whatever the model, these correlations were substantially lower than the correlations (almost unity) obtained by Ducrocq (2001) with the same approach but on a simulated data set and an idealized situation, where nearly no changes between sire and dam EBV were detected. The reasons for this might be the relatively large number of records in each HY-season group and of progeny per sire in Ducrocq (2001), but also the use of the same additive genetic variance in the sire-MGS and animal models. Here, the 2 additive genetic variances used were very different and, as mentioned before, the one obtained from the animal model appears overestimated to such a degree that the estimated additive genetic variance computed from the sire-MGS appears more appropriate.

For a given minimum size of HY class, the 2 options to model the contemporary group effect led to similar rankings. As larger contemporary group size increases the accuracy of sire evaluations (Meyer et al., 1989), it is not surprising that the higher the restriction, the more similar the ranking except that here, size restrictions led to fewer recorded daughters per sire. An alternative that needs to be studied is the definition of herd or HY class groups for small herds instead of excluding them from the analysis, as done in the current study. These classes could be formed according to similarity on milk production level. When dealing with HY groups, the clustering of similar HY groups within herd could be performed to increase the number of animals within a particular HY group (Van Bebber et al., 1997).

CONCLUSIONS

Genetic evaluation for the length of productive life of populations characterized by small herd size is possible but should be implemented with caution. Whatever the model used, when no restriction at all was imposed on the minimum HY group size, herd or HY variance was substantially inflated. Ranking of EBV was quite sensitive to these restrictions. When a herd effect was fitted, estimates of herd variance were similar whether a sire-MGS model or an animal model was used. In contrast, when an HY effect was used, estimates of HY variance were smaller, especially with animal model. This had an effect on the prediction of second-crop daughters' survival regarding the sires' EBV, where the inclusion of HY instead of herd effect was more

appropriate when an animal model was applied. Also, the correlation between the second-crop survival and sires' EBV was improved for both animal and sire-MGS models when an HY effect was used. Conclusions for the estimates of additive genetic variance were different. When a sire-MGS model was used, estimates were very similar whether a herd or an HY effect was fitted, whereas estimates were much larger with the animal model, especially when an HY effect was fitted. These findings suggest that in the case of small herds, the confounding between animal genetic effect, contemporary group effect and residual may be large. Variance component estimates derived from the sire-MGS model did not lead to more robust estimation of EBV when used in animal model. Regarding the model predictive ability, the most appropriate is the application of a sire-MGS model with HY contemporary group effect. Therefore, we recommend basing genetic evaluations of bulls and cows for longevity on the approximate animal model based on EBV from the sire-MGS model.

ACKNOWLEDGMENTS

The work was done during the stay of the first author at l'Unité Mixte de Recherche 1313 (Génétique Animale et Biologie Intégrative) of the Institut National de la Recherche Agronomique in Jouy-en-Josas, France. Erasmus student mobility and placement program is acknowledged for financial support.

REFERENCES

- Chirinos, Z., M. J. Carabaño, and D. Hernández. 2007. Genetic evaluation of length of productive life in the Spanish Holstein-Friesian population. Model validation and genetic parameters estimation. *Livest. Sci.* 106:120–131.
- Damgaard, L. H., I. R. Korsgaard, J. Simonsen, O. Dalsgaard, and A. H. Andersen. 2003. Genetic versus non-genetic Weibull log-normal sire frailty models: A simulation study. *Interbull Bull.* 30:35–42.
- Ducrocq, V. 1987. An analysis of length of productive life in dairy cattle. PhD Thesis. Cornell Univ., Ithaca, NY.
- Ducrocq, V. 2001. A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life. *Interbull Bull.* 27:147–152.
- 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.
- Ducrocq, V. 2006. Current status and prospects of survival analysis in animal breeding. CD-ROM Commun. no. 26–07 in Proc. 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil. WCGALP, Belo Horizonte, Brazil.
- Ducrocq, V., and G. Casella. 1996. A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28:505–529.
- Dürr, J. W., E. G. Monardes, and R. I. Cue. 1999. Genetic analysis of herd life in Quebec Holsteins using Weibull models. *J. Dairy Sci.* 82:2503–2513.
- Forabosco, F., J. H. Jakobsen, and W. F. Fikse. 2009. International genetic evaluation for direct longevity in dairy bulls. *J. Dairy Sci.* 92:2338–2347.
- Holtsmark, M., B. Heringstad, and J. Ødegard. 2009. Predictive abilities of different statistical models for analysis of survival data in dairy cattle. *J. Dairy Sci.* 92:5730–5738.
- Kaplan, E. L., and P. Meier. 1958. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53:457–481.
- Korsgaard, I. R., P. Madsen, and J. Jensen. 1998. Bayesian inference in the semiparametric log normal frailty model using Gibbs sampling. *Genet. Sel. Evol.* 30:241–256.
- Mészáros, G., C. Fuerst, B. Fuerst-Waltl, O. Kadlečík, R. Kasarda, and J. Sölkner. 2008. Genetic evaluation for length of productive life in Slovak Pinzgau cattle. *Arch. Tierzucht.* 51:438–448.
- Mészáros, G., J. Pálos, V. Ducrocq, and J. Sölkner. 2010. Heritability of longevity in Large White and Landrace sows using continuous time and grouped data models. *Genet. Sel. Evol.* 42:13.
- 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.
- Meuwissen, T. H. E., R. F. Veerkamp, B. Engel, and S. Brotherstone. 2002. Single and multitrait estimates of breeding value for survival using sire and animal models. *Anim. Sci.* 75:15–24.
- Meyer, K., H.-U. Graser, and K. Hammond. 1989. Estimates of genetic parameters for first lactation test day production of Australian black and white cows. *Livest. Prod. Sci.* 21:177–199.
- Potočnik, K., V. Gantner, J. Krsnik, M. Štepec, B. Logar, and G. Gorjanc. 2011. Analysis of longevity in Slovenian Holstein cattle. *Acta Agric. Slov.* 98:93–100.
- 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.
- 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.
- Smith, S. P., and R. L. Quaas. 1984. Productive life span of bull progeny groups: Failure time analysis. *J. Dairy Sci.* 67:2999–3007.
- Terawaki, Y., T. Katsumi, and V. Ducrocq. 2006. Development of a survival model with piecewise Weibull baselines for the analysis of length of productive life of Holstein cows in Japan. *J. Dairy Sci.* 89:4058–4065.
- Van Bebber, J., N. Reinsch, W. Junge, and E. Kalm. 1997. Accounting for herd, year and season effects in genetic evaluations of dairy cattle: A review. *Livest. Prod. Sci.* 51:191–203.
- Vollema, A. R. 1998. Selection for longevity in dairy cattle. PhD Thesis. Wageningen Agricultural Univ., Wageningen, the Netherlands.
- Vukasinovic, N., J. Moll, and N. Kunzi. 1995. Genetic relationships among longevity, milk production, and type traits in Swiss Brown cattle. *Livest. Prod. Sci.* 41:11–18.
- 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.

APPENDIX

Table A1. Rank correlations for cow¹ EBV from an animal model and their pedigree index approximation from sire-maternal grand sire model (sire plus half the maternal grandsire EBV) with contemporary group defined as herd or herd-year

HY size ²			
	n	Herd	Herd-year
1	36,039	0.84	0.71
2	35,934	0.86	0.77
3	35,542	0.86	0.77
5	32,561	0.85	0.79
10	21,787	0.83	0.80
Average		0.85	0.77

¹For cows with known sire and maternal grandsire.

²Minimum number of records per herd-year group.

Table A3. Rank correlations between cow and sire EBV from an animal model using variance components estimated from a sire-maternal grandsire model or an animal model with contemporary group defined as herd or herd-year combination

HY size ¹	Cows		Sires	
	Herd	Herd-year	Herd	Herd-year
1	0.99	0.95	0.99	0.98
2	0.99	0.97	0.99	0.98
3	0.99	0.97	0.99	0.98
5	0.99	0.97	0.99	0.98
10	0.99	0.98	0.99	0.98
Average	0.99	0.97	0.99	0.98

¹Minimum number of records per herd-year group.

Table A2. Rank correlations for cow and sire EBV from an animal model [with variance components from a sire-maternal grandsire (sire-MGS) model] and approximate animal model based on EBV from the sire-MGS model, when contemporary group is defined as herd or herd-year combination

HY size ¹	Cows		Sires	
	Herd	Herd-year	Herd	Herd-year
1	0.96	0.95	0.94	0.98
2	0.96	0.95	0.95	0.95
3	0.96	0.95	0.95	0.95
5	0.96	0.95	0.94	0.94
10	0.95	0.95	0.93	0.95
Average	0.96	0.95	0.94	0.95

¹Minimum number of records per herd-year group.