Validation of an approximate approach to compute genetic correlations between longevity and linear traits

Joaquim Tarrés, Jesús Piedrafita, Vincent Ducrocq

Genet. Sel. Evol. 38 (2006) 65–83

DOI: 10.1051/gse:2005027

Original article

Validation of an approximate approach to compute genetic correlations between longevity and linear traits

Joaquim TARRÉS[a]*, Jesús PIEDRAFITA[a], Vincent DUCROCQ[b]

[a] Grup de Recerca en Remugants, Departament de ciència animal i dels aliments, Universitat autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
[b] Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas Cedex, France

(Received 17 June 2005; accepted 20 September 2005)

Abstract – The estimation of genetic correlations between a nonlinear trait such as longevity and linear traits is computationally difficult on large datasets. A two-step approach was proposed and was checked *via* simulation. First, univariate analyses were performed to get genetic variance estimates and to compute pseudo-records and their associated weights. These pseudo-records were virtual performances free of all environmental effects that can be used in a BLUP animal model, leading to the same breeding values as in the (possibly nonlinear) initial analyses. By combining these pseudo-records in a multiple trait model and fixing the genetic and residual variances to their values computed during the first step, we obtained correlation estimates by AI-REML and approximate MT-BLUP predicted breeding values that blend direct and indirect information on longevity. Mean genetic correlations and reliabilities obtained on simulated data confirmed the suitability of this approach in a wide range of situations. When nonzero residual correlations exist between traits, a sire model gave nearly unbiased estimates of genetic correlations, while the animal model estimates were biased upwards. Finally, when an incorrect genetic trend was simulated to lead to biased pseudo-records, a joint analysis including a time effect could adequately correct for this bias.

simulation / genetic correlation / reliability / longevity

1. INTRODUCTION

Functional traits refer to traits related to the ability to remain productive. Their importance increases significantly in situations where production is limited or constrained (quotas) [16]. In general, with the exception of some type traits, functional traits exhibit two problems: rather low heritabilities and insufficient information early in life. These lead to genetic evaluations with low

* Corresponding author: joaquim.tarres@vit.de

reliabilities for young sires [14]. Fortunately, more heritable traits can be used as early predictors of these functional traits. For example, in dairy cattle, early predictors of, *e.g.*, somatic cell count or functional longevity can be found in the long list of type traits recorded in each breed [2, 6, 18, 21, 22]. A technique to properly combine these pieces of information is needed.

The optimal estimation procedure to combine information from different linear traits is known to be the multiple trait BLUP evaluation [8, 20]. MT-BLUP provides an improved accuracy of the evaluation on each trait through an increase of the amount of information, an improved data structure through better connectedness and a correction of biases due to selection on correlated traits. A multiple trait evaluation automatically accounts for the fact that traits are correlated and that the relative accuracy of the evaluation for each trait may greatly vary between the animals [14].

However an MT-BLUP evaluation on functional and production traits altogether, although conceptually possible, is not routinely feasible. Traits are often described by very different models. Some of these models are not linear; others involve repeated measures and/or more than one random effect or are analysed accounting for heterogeneous variances. Above all, the amount of data to manipulate in national evaluations is tremendous. Despite huge and fast improvements in computing power, computational considerations are still a limiting factor. Furthermore, a large set of dispersion parameters must be estimated accurately before being included in such an evaluation.

Ducrocq *et al.* [14] proposed a two-step approach for multiple trait evaluation of functional and production traits. First, univariate analyses are performed for each trait to get genetic variance estimates and to compute pseudo-records and their associated weights. Pseudo-records here can be regarded as a generalization of deviations or records corrected for environmental factors to more complex situations such as repeated records and nonlinear traits. Combining these pseudo-records in a multiple trait animal model while fixing the genetic and residual variances, one can get correlation estimates and approximate MT-BLUP breeding values that blend direct and indirect information.

Among functional traits, longevity was found to be the most important in most studies on relative economic weights in dairy cattle [5, 16]. Routine genetic evaluations of bulls on the length of productive life of theirs daughters rely on the modelling of a hazard function, which describes the limiting probability for a cow alive just prior to time t of being culled at time t. This allows a conceptually natural analysis of records from animals that are still alive (censored records) together with already culled animals (uncensored records). Moreover, this non-linear model used to describe the hazard function can

include time-dependent fixed effects (*e.g.*, herd-year-season, stage of lactation), which permit to precisely account for changes in culling policies over time.

The aim of this paper was to check *via* simulation the two-step approach for multiple trait genetic evaluation of longevity and linear traits. After the analysis of a reference situation, a sensitivity analysis was performed to check the suitability of this approach in a wide range of situations.

2. MATERIAL AND METHODS

2.1. General strategy

Observations of longevity (t) and of two linear traits (y_1, y_2) with known genetic and residual correlations were simulated. For longevity traits, the current models of analysis used are so different (they have to deal with non linearity, censoring and non normal residuals) that an exact multiple trait approach is usually not feasible, at least on moderate or large size data sets. The proposed approach was aimed at summarising the data in such a way that the simplest linear animal model can be used for each trait. This first step requires the calculation of a pseudo-record $y^*_{i,m}$ for each animal m and trait i corrected for all non genetic effects and an associated weight $w_{i,m}$ indicating the amount of information for that animal. These pseudo-records are obtained from a univariate (or a simpler multivariate) analysis, after estimation of the relevant dispersion parameters. Then, all pseudo-records are analysed together using a classical MT-BLUP framework assuming an animal model:

$$y^*_{i,m} = \mu_i + a_{i,m} + e_{i,m} \quad (1)$$

where μ_i is the overall mean for trait i, $a_{i,m}$ is the additive genetic value of animal m for trait i and $e_{i,m}$ is the residual. In order to account for the variable amount of information summarised in $y^*_{i,m}$, its residual variance is assumed to be heterogeneous: $var(e_{i,m}) = \sigma^2_{e,i}/w_{i,m}$, where $\sigma^2_{e,i}$ is the residual variance for trait i.

The derivation of pseudo-records and their weights is based on the following principle: when analysed using the simplistic univariate BLUP animal model (1), these records should lead to EBV equal or as close as possible to the EBV obtained with the complete model and in the case of nonlinear traits, with the adequate methodology. Then, the application of a MT-BLUP animal model based on equation (1) is straightforward. It provides the appropriate EBV for all traits and all animals. However, MT-BLUP requires an adequate knowledge

of the correlations between traits. The variances are supposed to be the ones estimated for the simpler univariate analysis used to compute pseudo-records.

2.2. Simulation of the dataset

The records of 5000 animals m roughly resembling the length of productive life of dairy cows were simulated using the following Weibull log normal frailty model:

$$h(t_m) = \lambda\, \rho\, (\lambda\, t_m)^{\rho-1} \exp\left(\mathbf{x}_m^{'}\beta + \mathbf{z}_m^{'}\mathbf{a}\right) \tag{2}$$

where $h(t_m)$ is the hazard function at time $t_m > 0$, the Weibull parameters ρ and λ are strictly positive. A ρ value of 2.0 (increasing hazard) and a λ value such that the median time t_{med} was 1000 days were used for the simulation. The two linear traits were simulated based on the model $y_{i,m} = \mu_{i,m} + \mathbf{x}_{i,m}^{'}\beta_i + \mathbf{z}_{i,m}^{'}\mathbf{a}_i + e_{i,m}$. Two means $\mu_1 = 100$ and $\mu_2 = 200$ were also arbitrarily added without a lack of generality. For each trait i, two fixed effects $\beta_i^{'} = \left(\beta_{i,1}^{'}\ \beta_{i,2}^{'}\right)$, with respectively 10 and 100 unbalanced levels, were generated from a uniform distribution and appropriate change of scale.

The true breeding values of the 5000 animals $a_{i,m}$, that were progeny of 50 unrelated sires, were obtained by adding half the breeding value of their sire $s_{i,m}$ to a value $u_{i,m}$ covering the dam contribution and Mendelian sampling, *i.e.* representing three quarters of the total genetic variance. The values were drawn from a $MVN(0, \mathbf{G})$ distribution, where $\mathbf{G}$ is the genetic covariance matrix. In the reference situation, genetic variances for longevity and linear characters were respectively 0.20, 400 and 600. Genetic correlations between all pairs of traits were 0.4.

Residual values $e_{i,m}$ for the two linear traits were also generated from values drawn from a $BVN(0, \mathbf{R})$ distribution, where $\mathbf{R}$ is the desired residual covariance matrix. Residual variances were chosen to lead to heritabilities of 0.25 for the first linear character and of 0.10 for the second. The distribution of the residual component for the longevity measure is proportional to an extreme value distribution. The residual correlation between the two linear traits was 0.4. In the reference situation, a residual correlation of 0 was assumed between longevity and the two linear traits.

2.3. Calculation of pseudo-records and their associated weight

Longevity was analysed using a Weibull frailty model [12, 14]. First, the genetic variance was estimated using a sire model. Then, assuming that this estimated variance is the correct one, longevity pseudo-records and their weights

were obtained using two procedures. The first one was a two-step procedure where the first step involved the estimation of fixed effects and sire EBV using a sire model, and the second step consisted in calculating the progeny's EBV $\hat{a}_{i,m}$, considering fixed effects and sire EBV as known. This results in an approximation of the EBV solutions $\hat{a}_{i,m}$ from a Weibull animal model (see [9] for details). The second procedure consisted in directly estimating the progeny's EBV $\hat{a}_{i,m}$ from a Weibull animal model. The resulting pseudo-record for longevity for animal m in both procedures was (for details, see [15]).

$$y^*_{i,m} = \frac{\delta_m}{w_{i,m}} + \hat{a}_{i,m} - 1 \tag{3}$$

where $\delta_m = 0/1$ is the censoring code. The associated weight $w_{i,m} = \hat{H}_{i,m}$ is the cumulative risk $\hat{H}_{i,m}$ of animal m from time 0 to culling or censoring time. The calculations were done using a modified version (version 5.0) of the Survival Kit [12].

For linear traits, instead of a univariate analysis for each trait, we decided to use a bivariate animal model, which is simple to implement here. First, genetic and residual dispersion parameters were estimated *via* an AI-REML procedure. The pseudo-record for trait i and animal m was simply the record corrected for fixed effects:

$$y^*_{i,m} = y_{i,m} - \mathbf{x}'_{i,m}\hat{\beta}_i. \tag{4}$$

The associated weight $w_{i,m}$ for the approximate MT-BLUP evaluation was the diagonal element of the least square part of the mixed model equations (MME) after absorption of the "contemporary group" fixed effect, that is the effect with the largest number of levels. This way, a lower weight is given to observations from small contemporary groups. It may seem inconsistent to absorb this fixed effect on one side and still correct for its estimate in (4). But as pointed out by a referee, the absorption matrix $\mathbf{M}$ is idempotent ($\mathbf{MM} = \mathbf{M}$). Therefore, the use of the absorption matrix $\mathbf{M}$ to weight $\mathbf{y}^*$ leads to the correct genetic estimates:

$$\begin{aligned}\left(\mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{A}^{-1}\lambda\right)\hat{\mathbf{a}} &= \mathbf{Z}'\mathbf{M}\mathbf{y}^* \\ &= \mathbf{Z}'\mathbf{M}\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right) = \mathbf{Z}'\mathbf{M}\left(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right) \\ &= \mathbf{Z}'\mathbf{M}\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\ \mathbf{y} = \mathbf{Z}'\mathbf{M}\mathbf{M}\mathbf{y} = \mathbf{Z}'\mathbf{M}\mathbf{y}.\end{aligned}$$

Using y or y* leads to the same solution vector $\hat{\mathbf{a}}$.

2.4. Joint analysis

Genetic and residual correlations were estimated *via* an AI-REML procedure applied to pseudo-records assuming that the variances are known and

equal to the estimates obtained in the first step (the residual variance for longevity was fixed to 1, since this value plays the role of the residual variance in the variance ratio $1/\sigma_a^2$ to be multiplied to the inverse relationship matrix [9, equation (12)]). Data were analysed using either a sire or an animal model to compare the performance of both models. Several REML packages exist but none is really adapted to model equations such as (1) with heterogeneous residual variances. A simple trick avoids this limitation [14]. Let $v_{i,m} = \sqrt{w_{i,m}}$. Multiplying both sides of the model equation (1) by $v_{i,m}$, one gets:

$$y^{\#}_{i,m} = v_{i,m}\, y^{*}_{i,m} = v_{i,m}\mu_i + v_{i,m}a_{i,m} + \varepsilon_{i,m}. \tag{5}$$

Now, the residual part $\varepsilon_{i,m}$ has homogeneous variance: $Var[\varepsilon_{i,m}] = v^2_{i,m}\ Var[e_{i,m}] = \sigma^2_{e,i}$. The REML estimation of the dispersion parameters of model (5) considering $y^{\#}_{i,m}$ as the data and $v_{i,m}$ as a continuous covariate gives results identical to the analysis of model (1) [14]. A version of I. Misztal's AI-REML software was modified to impose the constraints that genetic and residual variances are fixed [7]. Indeed, this is equivalent to impose structured genetic and residual (co)variance matrices where the only unknown parameters are the correlations. The AI-REML equations were expressed as functions of the unknown parameters (correlations) and of the first derivatives of the (co)variances matrices with respect to these [7]. Finally, an MT-BLUP evaluation based on pseudo-records (and their weights) was performed using the estimated genetic and residual (co)variance matrices.

2.5. Reliabilities

Two different longevity EBV were obtained for each animal: one from direct evaluation (*via* the Weibull model) and the other taking into account indirect information from correlated traits (MT-BLUP approach). Average reliabilities for longevity of progeny and their sires were also computed in two ways. First, asymptotic mean reliability was obtained as the mean of the diagonal elements of the inverse of the Weibull Hessian matrix at convergence of the maximisation process. The second way was to compute the correlation between the simulated (true) BVs and the estimated ones (*via* a Weibull model or *via* the MT-BLUP approach). The average reliability is the square of this correlation.

2.6. Sensitivity analysis

2.6.1. Genetic parameters for longevity

Two different genetic variances for longevity (either 0.05 or 0.20 (reference)) were used, indicating low and high genetic variation. These variances gave heritabilities for longevity of 0.05 and 0.19 respectively [23].

Also different genetic correlations (0.20, 0.40 and 0.60) were simulated between all three traits (longevity and the two linear traits) as well as a situation where correlations of longevity with linear traits 1 and 2 were equal to 0.4 and –0.4, respectively, and the two linear traits were assumed uncorrelated.

2.6.2. Level of random censoring

To see how censoring affected the results, four levels of random censoring (no censoring and approximately 30, 60 and 90%) were applied to the simulated datasets. Compared with the reference situation (no censoring), random censoring was generated in the following way. All sires from the same dataset had on average the same percentage of censored records among their progeny. Censoring randomly occurred at time C_m equal to 400, 800 or 1200 d, mimicking the end of a reproductive cycle. In datasets with 90% of censored records, all censoring times C_m were in the 400 days group. In the case of 60% of censored records, 40% of the censoring times C_m were in the 400 days group, 30% in 800 days and 20% in 1200. For a 30% of censored records, the respective values were 14%, 10% and 7%. Finally, the actual longevity measure available for analysis was set to $\min(C_m, T_m)$, where T_m was the failure time generated as in the reference situation.

2.6.3. Batches of progeny with different censoring rates

The reliability of the younger sires' proofs would increase with the multiple trait approach as a function of censoring percentage. In contrast with the previous section, sires were simulated with different percentages of censoring among their progeny. The actual longevity measure for the first 1000 animals (progeny of the 10 "young" bulls) was set to $\min(400, T_m)$, *i.e.*, censored at 400 days if the actual failure time T_m was higher than 400 days. Longevity for the next 1000 animals was set to $\min(800, T_m)$ days and the records of the following 1000 animals were set to $\min(1200, T_m)$ days. Finally, the last 2000 animals were not censored (progeny of old sires).

2.6.4. Non zero residual correlation between longevity and linear traits

So far we assumed a zero residual correlation between longevity and linear traits. Assuming that longevity data follows a Weibull distribution, equation (2) is equivalent to:

$$\log t = -\frac{(\rho \log \lambda + \mathbf{x}'\beta + \mathbf{z}'\mathbf{a})}{\rho} + \frac{e}{\rho} \tag{6}$$

where e follows an extreme value distribution [17] with $Var(e) = \pi^2/6$.

In order to generate a nonzero correlation, records for linear traits were simulated as:

$$y_i = \mu_i + \mathbf{x}_i'\beta_i + \mathbf{z}_i'\mathbf{a}_i + \varepsilon_i + \omega_i \quad \text{for} \quad i = 1,\ 2 \tag{7}$$

where the complete residual values e_i were decomposed into two components: one (ω_i) correlated with longevity and the other (ε_i) correlated with the other linear trait. That is to say, for $i = 1,\ 2$:

$$\begin{aligned} \text{cov}(e, e_i) &= \text{cov}(e, \omega_i) + \text{cov}(e, \varepsilon_i) = r_{longi}\sigma_e\sigma_{e_i} + 0 = r_{longi}\sigma_e\sigma_{e_i} \\ \text{cov}(e_1, e_2) &= \text{cov}(\varepsilon_1, \varepsilon_2) = r_{12}\sigma_{e_1}\sigma_{e_2}. \end{aligned}$$

To obtain $\omega_{i,m}$, first the failure time y_m was simulated in (6). Then, the residual e_m was obtained as $e_m = \rho \log t_m + \rho \log \lambda + \mathbf{x}_m'\beta + \mathbf{z}_m'\mathbf{a}$. The component correlated with longevity $\omega_{i,m}$ was generated as $\omega_{i,m} = b_{longi}e_m$, *i.e.*, the regression of e_i on e with $b_{longi} = r_{longi}(\sigma_{e_i}/\sigma_e)$.

Finally, the values $\varepsilon_{i,m}$ were generated from a bivariate normal distribution with the adequate covariance matrix. Note that the non zero residual correlation between longevity and the linear traits was obtained using equation (6) to model $\log t$ while the final model of analysis with pseudo-records is on a different scale with heterogeneous residual variance. Also, a positive correlation r_{longi} between linear trait i and $\log t$ corresponds to a negative relationship between the linear trait and the hazard. This will have implications on the interpretation of the results.

2.6.5. Biases generated by incorrect univariate analyses

In this section, it is assumed that a genetic trend on all traits existed, *i.e.*, progeny born in years 0, 1, ...,10 do not have the same average genetic level, and it was incorrectly estimated (biased). To generate such a situation, the simulation models for longevity and linear traits were modified to:

$$h(t_m) = \lambda\ \rho\ (\lambda\ t_m)^{\rho-1} \exp\left(\mathbf{x}_m'\beta + \delta_j + \mathbf{z}_m'\mathbf{a}\right) \tag{8}$$

$$y_{i,m} = \mu_{i,m} + \mathbf{x}_{i,m}'\beta_i + \delta_{i,j} + \mathbf{z}_{i,m}'\mathbf{a}_i + e_{i,m} \quad \text{for} \quad i = 1,\ 2 \tag{9}$$

where, for an animal born in year $j = 0, 1, ..., 10$ (11 years), an effect δ_j equal to 5% of a genetic standard deviation per year was added for each trait. To create an unbalanced but connected design, each sire was assumed to have half his progeny in year j and the other half in year $j + 1$ (where $j = 0$ to 10). So, sires had their progeny in different years.

The analysis of such a situation was done ignoring the year effect in the first step (univariate analysis) leading to potentially biased estimates of variances, pseudo-records and weights. However, to see how the joint analysis of the data can cope with such biases, a year effect was included in the second step, *i.e.*, for the estimation of the genetic and residual correlations and for the MT-BLUP evaluation of the three traits together.

3. RESULTS

3.1. Reference situation

Means of the genetic and residual variances estimates over 200 replicates for linear traits were always in the confidence interval of the mean. In the case of longevity, the sire variance is slightly underestimated for the reference situation (0.0449 estimated *vs.* 0.05 simulated). However, this bias may be considered negligible [10].

The calculation of the longevity pseudo-records and their weights was obtained by two alternative procedures: an approximate two-step procedure based on a sire model and a direct procedure based on the exact animal model. The correlations between both procedures were 0.9985 for pseudo-records and 0.9910 for their weights. Therefore, the approximate two-step procedure gave results very similar to the more demanding exact one. Only results from the exact procedure are reported hereafter because the moderate size of our datasets allowed the use of the animal model.

Next, univariate BLUP analyses were performed based on pseudo-records for longevity. The resulting EBV were compared with the ones calculated from the appropriate Weibull animal model. The correlations between both were 0.9980 for sires and 0.9999 for progeny. Their standard deviations were also nearly identical. As desired, the use of the pseudo-records in the univariate BLUP evaluation based on an animal model led to sire and progeny EBV equivalent to the EBV obtained in the Weibull analysis.

The pseudo-records were useful to compute genetic and residual correlations under a multiple trait sire or an animal model. All mean genetic correlation estimates were similar to the simulated ones whatever the estimation

Table I. Estimates of genetic and residual correlations between longevity (Long) and two linear traits (L1, L2) using the AI-REML approach under a sire and an animal model for two levels of animal genetic variation for longevity: low (0.05) and high (or reference) (0.20).

Model of Analysis			Sire				Animal			
Longevity heritability			Low		High		Low		High	
Correlations[a]		True	Mean[a]	STD[a]	Mean[a]	STD[a]	Mean[a]	STD[a]	Mean[a]	STD[a]
Genetic	Long-L1	0.4	0.382	0.182	0.371	0.139	0.388	0.128	0.374	0.079
	Long-L2	0.4	0.410	0.209	0.389	0.160	0.402	0.179	0.380	0.121
	L1-L2	0.4	0.377	0.158	0.373	0.160	0.379	0.156	0.374	0.156
Residual	Long-L1	0.0	0.004	0.005	0.010	0.006	–0.000	0.006	–0.000	0.006
	Long-L2	0.0	0.003	0.005	0.007	0.006	–0.000	0.005	–0.000	0.006
	L1-L2	0.4	0.406	0.011	0.406	0.011	0.415	0.024	0.416	0.024

[a] Mean and standard deviations (STD) over 200 replicates.

models (Tab. I). However, it was necessary to impose constraints such that the genetic and residual variances were known and equal to the estimated ones. Otherwise convergence was rarely obtained, due to genetic variances quickly going to 0. The standard deviations of genetic correlations were between 0.139 and 0.160 for the sire model and were slightly smaller for the animal model (Tab. I). The average asymptotic standard errors provided by the AI-REML algorithm were lower than these standard deviations (between 0.100 and 0.113 for the sire model). This was expected because it was assumed that the variances were known without error. Nevertheless, these asymptotic standard errors provided a rough idea about the magnitude of the accuracy of the estimates.

Finally, average reliabilities for longevity computed as the squared correlations between true breeding values and EBV were 0.890 for sires and 0.538 for progeny when just direct information was used in a Weibull model (Tab. II). These reliabilities were nearly identical to the asymptotic ones computed from EBV standard errors. When the genetic correlation between longevity and linear characters was accounted for in the MT-BLUP approach, the gain in reliability was very limited for sires (0.002 in absolute terms) and slightly higher (0.017) for progeny. This small increase may be related to the initial level of reliability, which was already high for sires.

Table II. Mean and standard deviation of reliabilities (squared correlations between true and estimated breeding values) of longevity obtained for sires and progeny with a Weibull model and the MT-BLUP approach under a sire and an animal model.

			Weibull model		MT-BLUP	
Genetic variance	Model of analysis	Animals	Mean[a]	STD[a]	Mean[a]	STD[a]
Low (0.05)	Sire	Sires	0.731	0.066	0.743	0.064
	Animal	Sires	0.732	0.066	0.748	0.063
		Progeny	0.403	0.043	0.434	0.040
High (0.20)	Sire	Sires	0.890	0.030	0.892	0.030
	Animal	Sires	0.892	0.030	0.895	0.029
		Progeny	0.538	0.032	0.555	0.029

[a] Mean and standard deviations (STD) over 200 replicates.

3.2. Sensitivity analysis

3.2.1. Effect of the genetic variance of longevity

Reducing the genetic variance for longevity from 0.20 (reference) to 0.05 did not greatly affect neither the average estimates of genetic variances nor the correlations (Tab. I), although the standard deviations of the latter increased to 0.182–0.209 (sire model). The gain in reliability for the MT-BLUP approach was higher than in the reference situation, 0.012 for sires and 0.031 for progeny (Tab. II).

3.2.2. Effect of the genetic correlation

Again, the characteristics of the estimates of genetic and residual variances and genetic correlations (results not shown) were not substantially modified by the level of genetic correlation in a range from 0.2 to 0.6. This was also true when genetic correlations of 0.4 between longevity and the first linear trait and –0.4 with the second were imposed. When indirect information was included, the gain in reliability increased with the amount of the genetic correlation simulated (Fig. 1). These gains were substantially greater when genetic correlations with different signs were simulated.

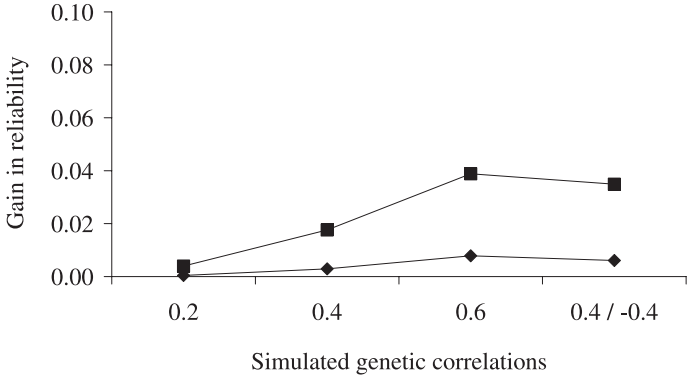

Figure 1. Increase in reliability for longevity of sires (♦) and progeny (■) obtained by adding indirect information with the MT-BLUP approach under an animal model with respect to the Weibull model. Different levels of genetic correlations were simulated: 0.2, 0.4 and 0.6 and +0.4/-0.4.

3.2.3. Effect of level of random censoring

The mean of genetic correlations estimated for datasets with different degrees of censoring were always close to the simulated values. The estimation procedure seems robust even with 90% of censoring, since the average values were only slightly underestimated. Nevertheless, the amount of censoring clearly affected the standard error of the estimates, which increased moderately until 60% of censoring (from 0.139–0.160 to 0.159–0.196 for the sire model), but the increase was substantial when censoring reached 90% (0.245–0.294).

Average reliabilities for longevity logically decreased when the level of censoring increased because the amount of direct information decreased. The gains in reliability with the inclusion of indirect information initially increased with censoring rate. In the case of 90% censoring, there was no gain in reliability for progeny. This is explained by a reduced accuracy of the genetic correlation estimates in this extreme situation as a consequence of the small number of informative records. This was checked by simulating a population of 25 000 progeny under the same conditions and with 90% censoring. Then, the standard deviation of the genetic correlation estimates was reduced to 0.08−0.11 (animal model) and the gain in reliability for progeny was about 0.06.

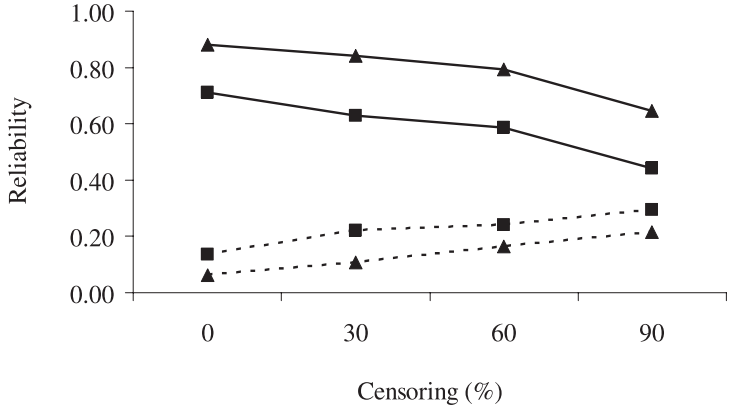

Figure 2. Mean (continuous line) and standard deviations (discontinuous line) of reliabilities for longevity of sires obtained with the Weibull model as a function of censoring percentage among progeny. Different levels of genetic variance for longevity were simulated: low (■) and high (▲).

3.2.4. Progeny batches with different censoring rates

When sires were simulated with different percentages of censoring among their progeny, the average reliabilities for longevity decreased and their standard deviations increased when censoring rate increased (Fig. 2). The average reliability of young sires with 90% of their daughters censored was 0.36 and 0.60 for low and high genetic variation, respectively. When indirect information was taken into account (Fig. 3), there was no gain in reliability for old sires (up to 30% censored). However, the gain for young sires was important when genetic variation was high (0.04–0.05), and even more so when it was low (up to 0.10). This can be attributed to the fact that the information of older sires allowed a more accurate estimation of variances and genetic correlations compared with the previous situation (all sires with the same percentage of censoring).

3.2.5. Effect of non zero residual correlations

When non zero residual correlations between longevity and linear traits exist, estimates of genetic and residual variances were again similar to the simulated values (results not shown). In the joint analysis, the sire model gave again virtually unbiased estimates of genetic correlations. However, the estimates from the animal model were clearly biased (Tab. III). The direction of

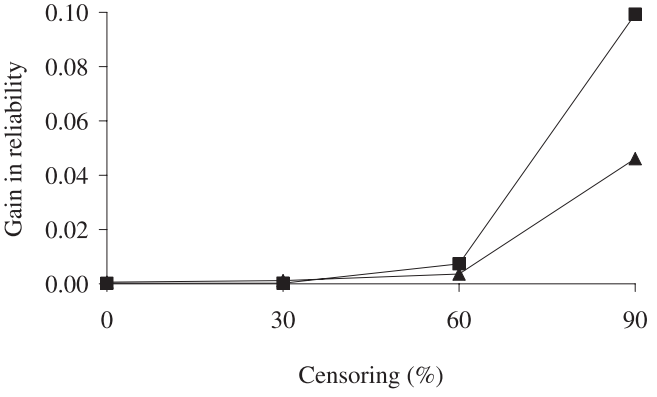

Figure 3. Increase in reliability for longevity of sires obtained by adding indirect information with the MT-BLUP approach as a function of censoring percentage. Different levels of genetic variance for longevity were simulated: low (■) and high (▲).

Table III. Estimates of genetic and residual correlations between longevity (Long) and the two linear traits (L1, L2) using the AI-REML approach under a sire and an animal model. Two situations were compared: non zero residual correlation (0.4) and incorrect univariate analysis. A high genetic variation for longevity (0.20) was assumed.

		Non zero residual			Incorrect univariate		
Correlations[a]		True	Sire	Animal	True	Sire	Animal
Genetic	Long-L1	0.4	0.355	–0.371	0.4	0.383	0.359
	Long-L2	0.4	0.360	–0.533	0.4	0.398	0.366
	L1-L2	0.4	0.382	0.554	0.4	0.400	0.394
Residual	Long-L1	0.4[b]	–0.092[c]	–0.087[c]	0	0.010	–0.001
	Long-L2	0.4[b]	–0.103[c]	–0.093[c]	0	0.007	–0.000
	L1-L2	0.4	0.388	0.369	0.4	0.406	0.414

[a] Mean over 200 replicates.
[b,c] Scale change: not comparable to the values on the same line; [b] corresponds to the correlation between the linear trait and log *t*, a negative sign for [c] is expected.

the bias depended on the sign of the residual correlation. The standard deviations were similar to those obtained with zero residual correlation. It should be noted that the estimated residual correlations on longevity trait were not comparable to the simulated ones (Tab. III). The latter ones corresponded to the modelling of log t with a residual variance of $\pi^2/6$ [11]. They were not directly transposable to the pseudo-record scale which assumes a heterogeneous residual variance and corresponds to a modelling of the hazard, not of log *t*.

Average reliabilities for sires and progeny were similar to those obtained with zero residual correlation (results not shown).

3.2.6. Effect of incorrect univariate analysis

When the univariate analyses were incorrect (for example, because of the existence of a hidden bias in the estimated genetic trend and/or an incorrect modelling of fixed effects), estimated genetic variances (0.0485 for longevity) were slightly increased compared to the reference situation (0.0449) and residual variances were slightly underestimated. The pseudo-records were also biased. A year effect was included in the joint analysis of the data to try to capture this bias. For longevity, the slope of the regression of the year effects estimated with the MT-BLUP approach on year was 0.0221 (sire model) and 0.0223 (animal model) and was very close to the simulated value (0.0224). Similar results were obtained for the year effects of the two linear traits. With the inclusion of this year effect, and in spite of the use of biased genetic and residual variances, all genetic correlations were on average similar to the simulated ones for both estimation models (Tab. III). The standard deviations of these correlations were also similar to those of the reference situation, between 0.141 and 0.171. On the contrary, the average reliabilities for sires and progeny were slightly lower (0.01 to 0.02) than the reference ones.

4. DISCUSSION

Although conceptually possible, the exact joint analysis of longevity data with early predictors or other functional and production traits is not routinely feasible on large data sets. This is due to the need for very different models for these traits (*e.g.*, accounting for nonlinearity, censoring and non normal residuals for longevity traits) and above all, to the fact that the amount of data to manipulate in national evaluations is tremendous. Despite huge and fast improvements in computing power, computational considerations are still a limiting factor. To avoid this, a less demanding two step approach was proposed by Ducrocq *et al.* [14] and was checked here *via* simulation. The results obtained under a sire and an animal model confirmed the suitability of the proposed approach in a wide range of situations.

The two-step approach starts with the estimation of dispersion parameters *via* univariate analyses (or simpler multivariate analyses of subsets of traits) and the evaluation of all recorded animals to get pseudo-records. These pseudo-records are performances free of all environmental effects that can be

used in a BLUP animal model to get the same breeding values as in the Weibull animal model. The longevity pseudo-records and their associated weights can be obtained using an animal model if the dataset is small. However, if there are computational constraints to implement it (*i.e.*, for large national applications) a two-step procedure to get approximate animal solutions based on a sire model is a less demanding alternative. The correlations between both procedures were very high for pseudo-records as well as for weights and, therefore, both of them could be used indistinctly.

Combining these pseudo-records into a multiple trait sire model and fixing genetic and residual variances to the previously estimated values, AI-REML estimates of genetic and residual correlations were virtually unbiased. This confirms the suitability of the multiple trait approach under a sire model for analysing this kind of data. A concern related to the necessary constraint of assuming that genetic and residual variances are known is that potentially biased genetic and residual variances might lead to biased estimates of correlations. In fact, it was found that the multiple trait approach under a sire model is quite robust: if the genetic trend is wrongly estimated in univariate analyses and estimates of variances, pseudo-records and weights are biased, the joint analysis of the data can correct and even estimate this bias by including a time (year) effect. Again, this leads to nearly unbiased estimates of genetic correlations.

The adequacy of the estimation of genetic correlations in the multiple trait animal model was first assessed under the assumption of a zero residual correlation between longevity and linear traits. This assumption is natural when different traits are recorded in different countries, *i.e.*, on different animals, but is no longer satisfying when the traits are observed on the same animals. Then residual correlations can differ substantially from 0 for some pairs of traits [14]. In such a situation, it was found that the genetic and residual correlations should be estimated under a sire model. The correlation between individual pseudo-record residuals is clearly deviating from the true residual correlation. A sire model somewhat averages residuals over progeny of a same sire and is more robust for variance component estimation. Then, the estimated correlations can be used in an MT-BLUP animal model.

After the MT-BLUP evaluation, a gain in true reliability for longevity is observed with respect to the situation when just direct information is used in a Weibull model. This gain is greater when the reliability in the initial Weibull model is lower. However, this gain was very limited in uncensored datasets, at least with the moderate size of our dataset. The increase in reliability was more noticeable when progeny batches with different censoring rates were

simulated. Then, the information from older sires allowed an accurate estimation of genetic variances and correlations and the multiple trait evaluation used this information to significantly improve the reliability for young sires (with 90% censored daughters).

The suitability of the two-step multiple trait approach was also assessed under situations where progeny of all sires had an extreme percentage of censoring. The estimation procedure for variances and correlations seems nearly unbiased but very imprecise in extreme cases, *e.g.*, with 90% censoring for all progeny groups and small size datasets. In these extreme situations, the gain in reliability is negligible. But the increase in true reliability due to the joint analysis can be quite important, when the dataset is large enough (*e.g.*, at least 25 000 records) to accurately estimate variances and correlations. These extreme levels of censoring usually exist only for a fraction of the bulls for dairy cattle length of productive life, but apply more generally to the whole population for example in piglet [4], beef calf [19] or laying hen [13] survival.

It was not our intention to compare this approach with other approximate strategies, which, for example, directly combine sets of estimated breeding values using selection index theory. Although computationally more demanding, the strategy proposed here has several attractive features: it accommodates nonlinear traits, it gets as close as possible to a true multiple trait BLUP which has well known theoretical characteristics and it offers a framework to the approximate estimation of genetic correlations between complex traits.

5. CONCLUSIONS

In conclusion, one can note that the two-step approach is an operational tool that can be implemented in many situations where a multiple trait approach is desirable but not applicable, either because of the huge size of the datasets analysed or the complexity and heterogeneity of the models to be implemented. Applications have been reported for total merit index constructions [14], joint analyses of longevity, discrete and linear traits [1], and joint cow and bull international evaluation [3].

ACKNOWLEDGEMENTS

Some suggestions of the two anonymous referees contributed significantly to the improvement of the manuscript. This work was started during a short stay of J. Tarrés at the Institut national de la recherche agronomique in Jouy-en-Josas, France and was supported by a grant from the "*Ministerio de Educación, Cultura y Deporte*" of Spain.

REFERENCES

[1] Besbes B., Ducrocq V., Protais M., An approximate total merit index combining linear traits, a survival trait and a categorical trait in laying hens, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France, Communication n° 20-05.

[2] Buenger A., Ducrocq V., Swalve H.H., 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 (2001) 1531–1541.

[3] Canavesi F., Boichard D., Ducrocq V., Gengler N., De Jong G., Liu Z., An alternative procedure for international evaluations: production traits european joint evaluation (PROTEJE), in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France, Communication n° 01-59.

[4] Casellas J., Noguera J.L., Varona L., Sánchez A., Arqué M., Piedrafita J., Viability of Iberian × Meishan F_2 newborn pigs. II. Survival analysis up to weaning, J. Anim. Sci. 82 (2004) 1925–1930.

[5] Colleau J.J., Regaldo D., Établissement de l'objectif de sélection dans les races bovines laitières, Renc. Rech. Ruminants 8 (2001) 329–332.

[6] Druet T., Sölkner J., Gengler N., Use of multitrait evaluation procedures to improve reliability of early prediction of survival, J. Dairy Sci. 82 (1999) 2054–2068.

[7] Druet T., Jaffrezic F., Boichard D., Ducrocq V., Modelling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows, J. Dairy Sci. 86 (2003) 2480–2490.

[8] Ducrocq V., Multiple trait prediction: principles and problems, in: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, 6–11 August 1994, Vol. 18, Guelph, Canada, pp. 452–462.

[9] Ducrocq V., A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life, Interbull Bulletin 27 (2001) 147–152.

[10] Ducrocq V., An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters, in: Proceedings of the 55th Annual Meeting of the European Association for Animal Production, 3–8 September 2004, Bled, Slovenia, Paper G6.11.

[11] Ducrocq V., Casella G., A Bayesian analysis of mixed survival models, Genet. Sel. Evol. 28 (1996) 505–529.

[12] Ducrocq V.P., Sölkner J., The Survival Kit v3.12", a FORTRAN package for large analysis of survival data, in: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Vol. 27, University of New-England, Armidale, Australia, pp. 447–450.

[13] Ducrocq V., Besbes B., Protais M., Genetic improvement of laying hens viability using survival analysis, Genet. Sel. Evol. 32 (2000) 23–40.

[14] Ducrocq V., Boichard D., Barbat A., Larroque H., Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional

traits into a total merit index, in: Proceedings of the 52nd Annual Meeting of the European Association for Animal Production, 26–29 August 2001, Budapest, Hungary, Paper GI.4.

[15] Ducrocq V., Delaunay I., Boichard D., Mattalia S., A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations, Interbull Bulletin 30 (2003) 101–111.

[16] Essl A., Length of productive life in dairy cattle breeding: a review, Livest. Prod. Sci. 57 (1998) 79–89.

[17] Kalbsfleisch J.D., Prentice R.L., The statistical analysis of failure time data, Wiley, New York, 1980.

[18] Larroque H., Ducrocq V., Relationships between type and longevity in the Holstein breed, Genet. Sel. Evol. 33 (2001) 39–59.

[19] Tarres J., Casellas J., Piedrafita J., Genetic and environmental factors influencing mortality up to weaning of Bruna dels Pirineus beef calves in mountain areas. A survival analysis, J. Anim. Sci. 83 (2005) 543–551.

[20] Van der Werf J., van Arendonk J.A.M., De Vries A.G., Improving selection of pigs using correlated characters, in: Proceedings of the 43rd Annual Meeting of the European Association for Animal Production, 14–17 September 1992, Madrid, Spain.

[21] Visscher P.M., Goddard M.E., Genetic parameters for milk yield, survival, workability and type traits for Australian dairy cattle, J. Dairy Sci. 78 (1995) 205–220.

[22] Weigel K.A., Lawlor T.J., Van Raden P.M., Wiggans G.R., Use of linear type and production data to supplement early predicted transmitting abilities for productive life, J. Dairy Sci. 81 (1998) 2040–2044.

[23] Yazdi M.H., Visscher P.M., Ducrocq V., Thompson R., Heritability, reliability of genetic evaluations and response to selection in proportional hazards models, J. Dairy Sci. 85 (2002) 1563–1577.