



ORIGINAL ARTICLE

Modelling genetic evaluation for dressage in Pura Raza Español horses with focus on the rider effect

M.J. Sánchez Guerrero¹, I. Cervantes¹, M. Valera² & J.P. Gutiérrez¹

¹ Department of Animal Production, Complutense University, Madrid, Spain

² Department of Agro-Forestry Sciences, ETSIA, University of Seville, Seville, Spain

Keywords

Bayesian procedure; cross-validation; genetic parameters; horse; rider–horse interaction.

Correspondence

M.J. Sánchez Guerrero, Department of Agro-Forestry Sciences, ETSIA, University of Seville, Ctra. Utrera km1, 41013 Seville, Spain.

Tel: +34 954 487748;

Fax: +34 954 486436;

E-mail: v32sagum@gmail.com

Received: 7 October 2013;

accepted: 20 February 2014

Summary

The most popular use of the Pura Raza Español horse in sport is for dressage competitions. Tests on young sport horses were first established in 2004 in Spain to collect data for the genetic evaluation of this breed's suitability for dressage. The aim of this study was to compare eight different models to find out the most appropriate way to include the rider in the genetic evaluation of dressage. A progressive removal of systematic effects from model was also analysed. A total of 8867 performance records collected between 2004 and 2011 from 1234 horses aged between 4 and 6 years old were used. The final score in the dressage test was used as the performance trait. The pedigree matrix contained 8487 individuals. A BLUP animal model was applied using a Bayesian approach with TM software. The horse's age, gender, travelling time, training level, stud of birth and event were included as systematic effects in all the models. Apart from the animal and residual effects that were present in all models, different models were compared combining random effects such as the rider, match (i.e. rider–horse interaction) and permanent environmental effects. A cross-validation approach was used to evaluate the models' prediction ability. The best model included the permanent environmental, rider and match random effects. As far as systematic effects are concerned, the event or the stud of birth was essential effects needed to fit the data.

Introduction

There is a growing demand for functional traits as selection objectives in the Pura Raza Español horse (PRE; 'Spanish Purebred') breeding programme given the increased interest in high-performance horses for sport competitions. Pura Raza Español horses with superior dressage performances have a greater economic value than the others as this is its most popular sport competition. The dressage discipline consists of a horse, guided by a rider, having to demonstrate its gaits at walk, trot and canter, and change between these gaits. All the performances are previously memorized by the rider, and a prescribed pattern of movements is followed. Apart from conformation, the breeding goal with PRE horses is therefore to improve

not only functionality in dressage, but, most importantly, gait quality, which is of great interest in dressage performance (Sánchez *et al.* 2013) especially for the selection of young horses.

Young horses (4–6 years old) were first tested in 2004 in Spain to collect data for the genetic evaluation of this breed's performance. One important environmental factor that influences horse performance is the rider (Kearsley *et al.* 2008; Bartolomé *et al.* 2013). Harmony, good communication and cooperation between horse and rider, known as 'match', are also key factors in performance outcomes in dressage competitions (Visser *et al.* 2008; Hawson *et al.* 2010; McGreevy & McLean 2010). Match has even been considered as a major influencing factor in equine welfare (McGreevy & McLean 2010). To achieve the

stated objectives in dressage, the working relationship relies heavily on how well the horses and riders cooperate. This interaction is also influenced by the level of experience of both rider and horse, together with an understanding of that particular horse's behaviour (Miller 2001; Visser *et al.* 2001, 2008), the riding skills of the rider (McGreevy & McLean 2010) and the degree of 'match' between horse and rider (Visser *et al.* 2003, 2008; McGreevy & McLean 2010; McLean & McGreevy 2010). Lewczuk (2007) also remarks that this interaction is useful when evaluating the training effect.

With the increasing professionalism of horse training and riding sport, the influence of the rider must be taken into account when interpreting gait scores (Becker *et al.* 2011). In the main exercise, known as 'dressage test', the horse is led by a rider round a track of a specific length while being evaluated by expert judges. The judges' scores are mainly based on the horse's movement, but also the degree of cooperation between horse and rider, as well as the influence of the rider's skill and experience. McLean and McGreevy (2010) suggest that equitation science can be used to provide better matches between horses and riders. The influence of the rider in top performing dressage horses is widely accepted.

The aim of this work was to assess the best model to predict dressage performance scores in PRE horses while studying the fit of the influence of the rider for use in future genetic evaluations. This study also analysed the convenience of simplifying the models by removing some of the systematic effects.

Materials and methods

Material

The data consisted of 8867 performance records from 1234 Pura Raza Español horses (1190 males and 44 females), aged between 4 and 6 years old. These horses belonged to a total of 330 studs of birth. There were an average of 3.1 different stallions by stud of birth, and each stallion performed its role in an average of 1.4 studs. Each horse had an average of 7.20 records. These were collected between 2004 and 2011 in all 179 official dressage tests (events) for young horses in Spain. In these events, the dressage discipline consists of two dressage tests, which is an exercise where different traits (canter, walk, trot, submission and overall appearance) are evaluated by 2 or 3 judges with a score from 1 to 10. The scores are then averaged and rescaled to a total score of 1–100 points. The final score of each dressage test in the

event was used as a performance trait. A description of the participants and collected data is shown in Table 1.

Pedigree information for genetic evaluation was collected from the PRE horse official studbook. The pedigree traced back all known generations for the participants totalling 8487 animals. The mean of equivalent complete generations for the participants was 10.0, which was calculated using a mean of inbreeding of 7.7% and a mean coancestry of 5.7%. These parameters were computed with ENDOG 4.8 (Gutiérrez & Goyache 2005).

Figure 1 shows the number of different riders riding one horse and the number of horses ridden by each rider. The average of different riders for one horse is 2.74, and the average of different horses ridden by one rider is 1.27. The rider–horse interaction effect combines the rider–horse pair and attempts to measure the different behaviour of a horse with specific riders, an effect referred to here as 'match'. Match had an average of 5.68 records across levels.

Genetic parameters

The genetic parameters were estimated using a Bayesian procedure applied to univariate mixed linear models. All the runs were carried out using the TM software (Legarra 2008).

The fitted models included the following systematic effects: age in years (4,5,6), gender (male, female) and travelling time (1,...,46), which referred to the combination of the following factors: 'transport to the event' (walking or by trailer/box), 'journey length to the event' (<30 min, 30 min to 2 h, 2–4 h, 4–6 h, 6–8 h and >8 h) and 'arrival time before the beginning of the event' (<6 h before, 6–12 h before, 12–24 h before and >24 h before). The horse's training (1,..., 41) was another systematic effect, defined as the com-

Table 1 Descriptive statistics of analysed data

	No. animals	No. performance records
Total	1234	8867
4 years old	779	3787
5 years old	622	3235
6 years old	320	1845
Participated at age of 4, 5 and 6	103	2199
Participated at two different ages	301	3442
Female	44	398
Male	1190	8469
Average by event	6.9	49.5
Average by stud of birth	3.7	26.9

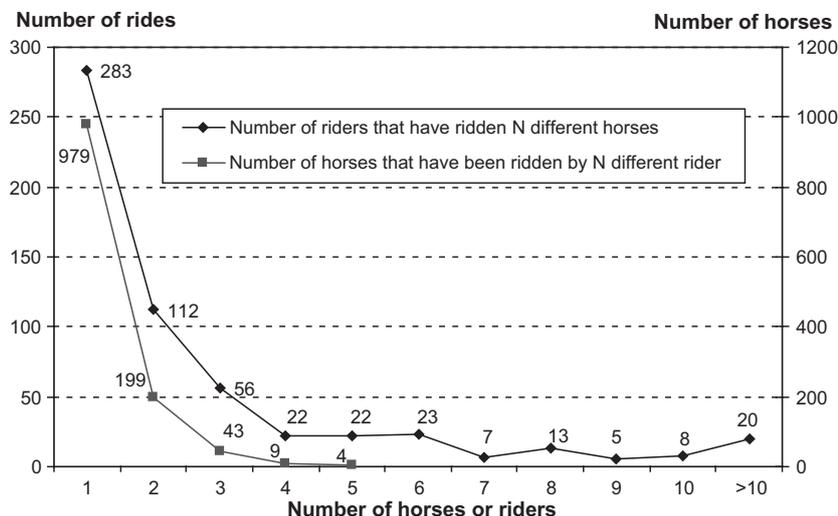


Figure 1 Horses ridden by more than one rider and vice versa.

bination of the factors ‘number of previous events in which the animal participated’ (<5 competitions, 5–10, 10–20 and >20), ‘daily hours of training’ (<3 h, 3–6 h, 6–10 h and >10 h) and ‘length of time for which the horse has been trained’ (<6 months, 6–12, 12–24 and >24 months). A new level of each effect was created for each component combination described above for these two effects. Travelling and training information was collected through a survey completed by the horse’s trainer, and in the case of travelling, veterinary travel guides were also checked. Data were checked to ensure that the distribution of records within training and travelling effect was independent. The other systematic effects were stud of birth (1,...,330) and event (1,...,179).

The equation in matrix notation for the model to be solved for a hypothetical trait considering all of the possible random effects was $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{Qr} + \mathbf{Nm} + \mathbf{e}$ and had:

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{p} \\ \mathbf{r} \\ \mathbf{m} \\ \mathbf{e} \end{pmatrix} \sim N \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{A}\sigma_u^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_p^2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_r^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_m^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2 \end{pmatrix} \right)$$

where \mathbf{y} is the vector of observations, \mathbf{X} the incidence matrix of systematic effects, \mathbf{Z} the incidence matrix of animal genetic effects, \mathbf{W} the incidence matrix of permanent environmental effects, \mathbf{Q} the incidence matrix of the rider effect, \mathbf{N} the incidence matrix of match effects, \mathbf{b} the vector of systematic effects, \mathbf{u} the vector of direct animal genetic effects, \mathbf{p} the vector of permanent environmental effects, \mathbf{r} the vector of rider

effects (1,...,571); \mathbf{m} the vector of match effects (1,..., 984), \mathbf{e} the vector of residuals, σ_u^2 the direct genetic variance, σ_p^2 the permanent environmental variance, σ_r^2 the rider variance, σ_m^2 the match variance, σ_e^2 the residual variance, \mathbf{I} an identity matrix, and \mathbf{A} the numerator relationship matrix. Eight different models were tested including different random effects. Model A included animal and residual effects, while the rest of the models included other random effects: B the rider effect, C the match effect, D the rider and the match effects, E the permanent environmental effect, F the rider and the permanent environmental effects, G the match and the permanent environmental effects and H the rider, the match and the permanent environmental effects.

Marginal posterior distributions of all parameters were estimated using the Gibbs sampling algorithm. Prior distributions for systematic effects were assigned as bounded uniform prior distribution and the variance components σ_u^2 , σ_p^2 , σ_r^2 , σ_m^2 and σ_e^2 were scaled using inverted chi-squared distributions ($\nu = 2$ and $S = 0$) to perform a flat prior distribution. Total Gibbs chain lengths of 1 000 000 samples for each analysis were defined, with a burn-in period of 100 000 and a thinning interval of 100.

Model selection criteria

The lowest deviance information criterion (DIC) value (Spiegelhalter *et al.* 2002) was used as a criterion of fitness, and a cross-validation approach was used to evaluate their prediction ability (Efron & Tibshirani 1993). Cross-validation was originally employed to

evaluate the predictive validity of linear regression equations for forecasting a performance criterion from scores on a battery of tests (Mosier 1951), and nowadays, it is usually used in quantitative genetics (Olsen *et al.* 2012; Vazquez *et al.* 2012; Andonov *et al.* 2013).

For each model, the entire data set was randomly tenfold split into a training data set containing 6650 records (75%) to estimate the parameters and solve each model, and a validation data set with 2217 records (25%) to test the predictive ability of the model using the solutions obtained with the training set. Performances of the validation data set were reconstructed using the solutions for the model effects previously obtained with the training data sets. This procedure was based on the comparisons between real and predicted performances of the validation set, calculating the Pearson correlation (r) between real and predicted performances. Ten different random replicates of this procedure were carried out, and the Pearson correlations were averaged across replicates as a measure of the predictive ability of each model.

To analyse the convenience of simplifying the models by eliminating systematic effects, the correlation between the predicted scores and the real scores was computed using the best fitting model, while the solutions for each of the systematic effects were ignored. The statistical differences between the predicting ability of the eight models were analysed with a Fisher transformation, a two-tailed test for independent samples (Fisher 1970).

Results

Genetic parameter estimates

Table 2 shows the total variance, heritabilities and the ratio of random effects variance to phenotypic variance for all the fitted models. Values are accompanied by the standard deviation of the marginal posterior distribution. It should be noted that these are not standard error of estimates as a Bayesian analysis has been performed, and the standard deviation of their marginal posterior distribution usually tends to be much higher than the standard error. Heritabilities ranged from 0.22 in the most complex model (H) to 0.59 in the simplest one (A). Repeatability of the trait, defined as the sum of the heritability and environmental permanent effect ratio assessable in models accounting for the environmental permanent effect (models E to H), was roughly equal to the heritability estimates in models A to D, which ignored the environmental permanent effect, showing an overestimation of heritability in these scenarios. The ratios for

the permanent environmental effect of the horse were not relevant in models G or H.

Models A and E, which ignored the rider effect either alone or included in 'match', also led to an overestimation of the heritability (0.58 model A and 0.38 model E), as can be observed when compared with each counterpart model, either including the environmental permanent effect (0.22–0.30 models F to H) or not (0.28–0.39 models B to D). In addition, an overestimation of heritability appeared when the match or rider effects were ignored (models B and C) in comparison with model D, which included both effects. This overestimation was also present in models F and G compared with model H.

The rider effect accounted regularly for 24% to 26% of the phenotypic variance when this effect was present in the model either after fitting the match effect (models D and H) or not (models B and F). On the other hand, the match effect was influenced by the rider effect when the latter was not included in the model accounting for 25% (model G or model C), whereas the match influence decreased to 11% (model H) or 12% (model D) when both effects involving the rider were present in the models. In fact, when models D and H were compared with models B and F, respectively, it could be seen that a good portion of the match effect was mixed with the animal effect, which clearly decreased, particularly when the environmental permanent effect was absent. The lowest heritability value of 0.22 was obtained in the model including all the effects.

Model selection criteria

Two different criteria have been used in this work to choose the best model: the DIC value, which assessed the models' goodness of fit, and a cross-validation procedure, which studied their prediction ability.

DIC values for all the models are shown in Table 2. Under this criterion, the models with a lower DIC value were understood to fit better. Model D, for instance, which only included the rider and match as a random effect besides the animal and residual effects, had the best fit, closely followed by model H, which also included the environmental permanent effect. The worst-fitting models were those which ignored the rider, the match or both: model A (only including animal and residual effects) and model E (with the environmental permanent effect added).

Table 2 shows the averaged correlations across replicates between real and predicted records under the cross-validation procedure described in the methodology section. All the correlations were found to be very

Table 2 Phenotypic variance, mean and standard deviation of the marginal posterior distributions for the heritabilities, environmental permanent ratio, rider ratio, match ratio, deviation information criteria (DIC) and correlations between predicted scores under each model (Y_i) and real scores (Y_j) for the eight univariate models

	Vp	h^2	p^2	r^2	m^2	DIC*	Correlation $Y_i Y_j$
Model A	20.84	0.59 ± 0.02	–	–	–	5.357	0.7068 ^b
Model B	22.30	0.39 ± 0.03	–	0.26 ± 0.03	–	5.294	0.7218 ^{ab}
Model C	20.09	0.36 ± 0.04	–	–	0.25 ± 0.03	5.280	0.7173 ^{ab}
Model D	21.13	0.28 ± 0.04	–	0.24 ± 0.03	0.12 ± 0.03	5.278	0.7234 ^{ab}
Model E	16.34	0.38 ± 0.09	0.19 ± 0.08	–	–	5.358	0.7043 ^b
Model F	18.98	0.24 ± 0.06	0.13 ± 0.05	0.26 ± 0.03	–	5.293	0.7297 ^{ab}
Model G	18.77	0.30 ± 0.06	0.06 ± 0.05	–	0.25 ± 0.03	5.280	0.7170 ^{ab}
Model H	19.72	0.22 ± 0.05	0.06 ± 0.04	0.24 ± 0.03	0.11 ± 0.02	5.278	0.7427 ^a

Vp, phenotypic variance; h^2 , heritability; p^2 , variance of environmental permanent effect/phenotypic variance; r^2 , variance of rider effect/phenotypic variance; m^2 , variance of match/phenotypic variance.

Different letters indicate significant differences ($p < 0.05$).

Different letters (a, b or ab) indicate significant differences ($p < 0.05$).

*DIC × 10⁻⁶.

high, as a result of a well-structured data set with several records available for each horse. Model D showed the best goodness of fit, but model H, including the complete set of random effects, showed the greatest correlation between the predicted scores within each model and the real scores (i.e. 0.7427). Model A, with only the animal as random effect besides the residual effect, was found to fit the worst, with a correlation of 0.7068. Prediction ability decreased when the effects were eliminated in the following order: both rider and match effects (A), only the match effect (C) and only the rider effect (B). For models with the environmental permanent effect, similar results were found. Significant differences were found between models H and A, and H and E. To analyse the convenience of simplifying the models by eliminating systematic effects, the correlation between the predicted scores under each model and the real scores using model H (i.e. ignoring solutions for other systematic effects) was calculated. The correlation coefficient remained similar, despite gender, age, travelling time and training being removed from the model (i.e. 0.7430, 0.7415, 0.7399 and 0.7417, respectively) and decreased significantly if the event (0.6473) or the stud of birth was omitted (0.5922). No significant differences were found between the best model with all the systematic effects and the best model without gender, age, travelling time and/or training.

Discussion

Despite difficulties in prediction, the rider effect has been included in several studies of horse performance, either as a fixed or random effect (Jaitner & Reinhardt 1993; Aldridge *et al.* 2000; Kearsley *et al.* 2008; Gómez

et al. 2010; Bartolomé *et al.* 2013). However, this was the first time to our knowledge that the rider effect has been analysed in a horse population from the view point of prediction ability. Previous studies had indicated that including the rider in the models as an additional effect significantly improved the fit of the model to the data (Kearsley *et al.* 2008; Bartolomé *et al.* 2013), and Becker *et al.* (2011) noted the importance of the rider in motion exercises, suggesting that horses will have better chances of obtaining high gait scores when guided by skilful riders than by less experienced riders.

Choosing the best way to include the rider effect, and its relationship with the animal as regards the goodness of fit, should reflect better the true state of the horse's nature. However, as far as horse competitions are concerned, it is more desirable to have a model that can predict the performance of a particular horse better when guided by a particular rider. Moreover, regarding selection decisions, prediction ability also seems to be a better criterion for choosing a model.

The final aim of BLUP methodology in the prediction of breeding values is the genetic improvement of the population (Legarra *et al.* 2005). Thus, the focus should be on the predictive ability of the models, and it was noteworthy that the models that included the rider influence in some way predicted the records better than those including neither the rider nor the match.

A previous cross-validation study in racing performance horses (Norwegian and North Swedish cold-blooded trotters) resulted in much lower values (0.26–0.27) for the correlation between predicted and real data (Olsen *et al.* 2012). The high magnitude of

the correlations obtained in our study was probably due to the robustness of the analysed data structure, with a large number of records per horse (i.e. 7.2), the equal distribution of records within the levels of the random effects related to the rider and the particularities of the PRE horse, which is a close-bred population with reliable, long-standing pedigree information.

No significant differences were found between model H and those including at least one rider-related effect (Table 2), which shows that fitting the effect as the rider on its own or as a combination of 'match' does not make much difference at least to the basic level of competition. However, all these models except the one including the whole set of effects did not significantly differ from other models, which did not fit the permanent environmental effect. Consequently, the complete model, including the whole set of effects, seems to be the most appropriate. Regarding possible simplifications of the model as regards gender, age, training or travelling time, these systematic effects could be removed from the model, as there were no significant differences in the correlations between real and predicted data using the best model. However, effects such as the event or the stud of birth are essential to fit the data. This result suggests that some of information collected was unnecessary and could be omitted from the recording data sheet. One noticeable result was the high correlations found between real and predicted records – a fact that shows how highly predictable competition results are when using an appropriate linear model.

Estimated heritabilities were higher than those reported by other authors such as Huizinga and Van der Meij (1989), Koerhuis (1992) (0.11) and Koenen *et al.* 1995 (0.17) for the Dutch Warmblood Horse, Wallin *et al.* (2003) for Swedish Warmblood riding horses (0.16–0.17), Kearsley *et al.* 2008 (0.09–0.11) and Stewart *et al.* 2010 (0.11–0.15). However, our results were similar to those obtained by Thorén *et al.* (2006) (0.35) and Philipsson *et al.* (1990) in Swedish Warmblood Horse (0.20–0.40). It must be noted that in models A to C, and E to F, heritability was overestimated, but the results from models D and H were more similar to those found by other authors. Therefore, when neither the rider effect nor the match was included, heritability was overestimated, as these three effects were confused, which showed the need in some models for splitting the animal effect used fitted in simpler models. Also, if the environmental permanent effect is not fitted, this effect could be confused within the animal effect, which also led to an overestimation of heritability. To check the influence from non-repeated records for some ani-

mals, a correlation was calculated between breeding values with and without records from animals without repetitions. The result was 0.9745, so a repeatable structure was sound. The rider effect proved regularly to be one of the most important in all the models in which it was fitted, and it was also double the match effect when both effects were fitted jointly (models D and H), which suggested that the rider effect was much more important than the understanding between rider and horse defined in the match effect. The more effects regarding riders included in the model, the more the environmental permanent influence and animal effect decreased, which indicated that when these effects were absent, the rider and match effects were hidden among the animal and permanent environmental effects. In fact, the best way to fit the rider and/or horse–rider relationship has not been analysed before using cross-validation. In view of the results obtained, the animal and rider effects and the interaction between them (match effect) should be jointly fitted when the data structure is strong enough to separate them correctly.

In the PSCJ dressage test, the horse and rider are judged as a single team by dressage judges, based on the harmonious appearance of their performance. Dressage horses are therefore selected and/or trained to be highly sensitive to the rider's cues, so that the rider can communicate with the horse by minimal, hardly perceptible, cues (Von Borstel *et al.* 2010). The results reported here show some advantage in terms both of goodness of fit and of prediction ability of models including the rider or match compared with model A which omits it, but it seems that all the models had reasonably high prediction ability. However, the satisfactory results were a consequence of the well-structured data set and underline the importance of preselection of the data. These performance tests seem to have a suitable design in which to measure the best combination of rider–horse, as the performance attained by a horse can be predicted with high reliability. Nevertheless, this ideal data structure, so useful in discriminating the importance of rider and match effects, is not usually found in field data. A single rider usually rides several horses from a particular stud in dressage competitions. In our data, for instance, 288 riders rode more than two horses. However, it is unusual for a horse to be ridden by several riders, unlike our data set, in which 251 horses had been ridden by two or more different riders.

On the other hand, just as riders' personalities vary widely, horses display a wide variety of behaviour, and hence, it is a reasonable assumption that only a small proportion of possible combinations of personal-

ities result in optimal matches between horse and rider (Hausberger *et al.* 2008).

Prediction ability has been shown to be high in the models of dressage with PRE horses due to the aiding structure mentioned above and the fact that this population has remained closed for long and is therefore fairly homogeneous. The conclusions reached in this study can only be extended if further analysis is carried out in other populations and disciplines, such as thoroughbred racing horses, where splitting rider and match effects would theoretically be easy, as most of the horses are usually guided by different riders. In fact, the culture of betting is closely tied to disciplines, and these kinds of models would help in this case to come up with more reliable predictions.

Conclusions

The results agreed with previous reports, suggesting that ignoring the rider effect would negatively affect genetic evaluations in dressage. The controversy over the inclusion of the rider and the rider–horse interaction is set to continue being of great importance in the world of horse competitions. The best model to predict the performance of a particular rider–horse pair has been shown to be that which included environmental permanent, rider and match random effects. These conclusions will help to optimize the design of performance recording in dressage of Pura Raza Español horses and might also be used for tests with other horse competitions and populations.

Acknowledgements

The authors wish to thank the National Association of Pura Raza Español Horse Breeders (ANCCE) for providing the data used in this study.

References

- Aldridge L.I., Kelleher D.L., Reilly M., Brophy P.O. (2000) Estimation of the genetic correlation between performances at different levels of show jumping competitions in Ireland. *J. Anim. Breed. Genet.*, **117**, 65–72.
- Andonov S., Ødegård J., Svendsen M., Ådnøy T., Vegara M., Klemetsdal G. (2013) Comparison of random regression and repeatability models to predict breeding values from test-day records of Norwegian goats. *J. Dairy Sci.*, **96**, 1834–1843.
- Bartolomé E., Menéndez-Buxadera A., Valera M., Cervantes I., Molina A. (2013) Genetic (co)variance components across age for Show Jumping performance as an estimation of phenotypic plasticity ability in Spanish horses. *J. Anim. Breed. Genet.*, **130**, 190–198.
- Becker A.C., Stock K.F., Distl O. (2011) Genetic correlations between free movement and movement under rider in performance tests of German Warmblood horses. *Livest. Sci.*, **142**, 245–252.
- Efron B., Tibshirani R.J. (1993) Cross-validation and other estimates of prediction error. In: B. Efron, R.J. Tibshirani (eds), *An Introduction to the Bootstrap*. Chapman & Hall, New York, pp. 237–255.
- Fisher R.A. (1970) *Statistical Methods for Research Workers*, 14th edn. Hafner Publishing, Darien, CT, pp. 362.
- Gómez M.D., Menéndez-Buxadera A., Valera M., Molina A. (2010) Estimation of genetic parameters for racing speed at different distances in young and adult Spanish Trotter horses using the random regression model. *J. Anim. Breed. Genet.*, **127**, 385–394.
- Gutiérrez J.P., Goyache F. (2005) A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.*, **122**, 172–176.
- Hausberger M., Roche H., Henry S., Visser E.K. (2008) A review of the human–horse relationship. *Appl. Anim. Behav. Sci.*, **109**, 1–24.
- Hawson L.A., McLean A.N., McGreevy P.D. (2010) Variability of scores in the 2008 Olympic dressage competition and implications for horse training and welfare. *J. Vet. Behav.*, **5**, 170–176.
- Huizinga H.A., Van der Meij G.J.W. (1989) Estimated parameters of performance in Jumping and Dressage competition of the Dutch Warmblood horse. *Livest. Sci.*, **21**, 333–345.
- Jaitner J., Reinhardt F. (1993) Estimation of breeding values for performance test traits of stallions. In: *Proceedings of the 44th Annual Meeting of the European Association for Animal Production (EAAP)*. Aarhus (Denmark).
- Kearsley C.G.S., Woolliams J.A., Coffey M.P., Brotherstone S. (2008) Use of competition data for genetic evaluations of eventing horses in Britain: analysis of the dressage, showjumping and cross country phases of eventing competition. *Livest. Sci.*, **118**, 72–81.
- Koenen E.P.C., Van Veldhuizen A.E., Brascamp E.W. (1995) Genetic parameters of linear scored conformation traits and their relation to dressage and show-jumping performance in the Dutch Warmblood riding horse population. *Livest. Prod. Sci.*, **43**, 85–94.
- Koerhuis A. (1992) Implementatie van een Multivariaat Animal Model ter Evaluatie van Sportprestaties van KWPN-Rijpasrden. Report 24. Department of Animal Breeding Wageningen Agricultural University, Wageningen, The Netherlands, pp. 48.
- Legarra A. (2008) TM Threshold Model (available at: <http://acteon.webs.upv.es/>; last accessed 17 March 2014).

- Legarra A., López-Romero P., Ugarte E. (2005) Bayesian model selection of contemporary groups for BLUP genetic evaluation in Latxa dairy sheep. *Livest. Sci.*, **93**, 205–212.
- Lewczuk D. (2007) The effect of training on linear jumping parameters in young stallions. *Equine Comp. Exerc. Physiol.*, **4**, 159–165.
- McGreevy P.D., McLean A.N. (2010) *Equitation Science*. Wiley-Blackwell, Oxon, UK, pp. 314.
- McLean A.N., McGreevy P.D. (2010) Ethical equitation: capping the price horses pay for human glory. *J. Vet. Behav.*, **5**, 203–209.
- Miller R.M. (2001) Behaviour and misbehavior of the horse. *Vet. Clin. North Am. Equine Pract.*, **17**, 379–387.
- Mosier I. (1951) The need and means of cross validation. I. Problems and designs of cross-validation. *Educ. Psychol. Measur.*, **11**, 5–11.
- Olsen H.F., Klemetsdal G., Ødegård J., Árnason T. (2012) Validation of alternative models in genetic evaluation of racing performance in North Swedish and Norwegian cold-blooded trotters. *J. Anim. Breed. Genet.*, **129**, 164–170.
- Philipsson J., Arnason T., Bergsten K. (1990) Alternative selection strategies for performance of the Swedish Warmblood Horse. *Livest. Sci.*, **24**, 273–285.
- Sánchez M.J., Gómez M.D., Peña F., Morales J.L., Molina A., Valera M. (2013) Relationship between conformation trait and gait characteristics in Spanish Purebred horses. *Arch. Tierz.*, **56**, 137–148.
- Spiegelhalter J., Best G., Bradley P., Angelika V.D.L. (2002) Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Series B Stat. Methodol.*, **64**, Part 4, 583–639.
- Stewart I.D., Woolliams J.A., Brotherstone S. (2010) Genetic evaluation of horses for performance in dressage competitions in Great Britain. *Livest. Sci.*, **128**, 36–45.
- Thorén E., Viklund Å., Koenen E.P.C., Ricard A., Bruns E., Philipsson J. (2006) Review of genetic parameters estimated at stallion and young horse performance tests and their correlations with later results in dressage and show-jumping competition. *Livest. Sci.*, **103**, 1–12.
- Vazquez A.I., Perez-Cabal M.A., Heringstad B., Rodrigues-Motta M., Rosa G.J.M., Gianola D., Weigel K.A. (2012) Predictive ability of alternative models for genetic analysis of clinical mastitis. *J. Anim. Breed. Genet.*, **129**, 120–128.
- Visser E.K., Van Reenen C.G., Hopster H., Schilder M.B.H., Knaap J.H., Barneveld A., Blokhuis H.J. (2001) Quantifying aspects of young horses' temperament: consistency of behavioural variables. *Appl. Anim. Behav. Sci.*, **74**, 241–258.
- Visser E.K., Van Reenen C.G., Engel B., Schilder M.B.H., Barneveld A., Blokhuis H.J. (2003) The association between performance in show-jumping and personality traits earlier in life. *Appl. Anim. Behav. Sci.*, **82**, 279–295.
- Visser E.K., Van Reenen C.G., Blokhuis M.Z., Morgan E.K.M., Hassmen P., Rundgren T.M.M., Blokhuis H.J. (2008) Does horse temperament influence horse-rider cooperation? *J. Appl. Anim. Welf. Sci.*, **11**, 267–284.
- Von Borstel U.U.K., Duncan I.J.H., Lundin M.C., Keeling L.J. (2010) Fear reactions in trained and untrained horses from dressage and show-jumping breeding lines. *Appl. Anim. Behav. Sci.*, **125**, 124–131.
- Wallin L., Strandberg E., Philipsson J. (2003) Genetic correlations between field test results of Swedish Warmblood riding horses as 4-year-olds and lifetime performance results in dressage and show jumping. *Livest. Sci.*, **82**, 61–71.