

# Cross-validation analysis for genetic evaluation models for ranking in endurance horses

S. García-Ballesteros<sup>1</sup>, L. Varona<sup>2</sup>, M. Valera<sup>3</sup>, J. P. Gutiérrez<sup>1</sup> and I. Cervantes<sup>1†</sup>

<sup>1</sup>Departamento de Producción Animal, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, E-28040 Madrid, Spain; <sup>2</sup>Unidad de Genética Cuantitativa y Mejora Animal, Universidad de Zaragoza, E-50013 Zaragoza, Spain; <sup>3</sup>Departamento de Ciencias Agro-Forestales, Universidad de Sevilla, Ctra. Utrera km 1, 41013 Sevilla, Spain

(Received 19 December 2016; Accepted 11 May 2017; First published online 21 June 2017)

---

*Ranking trait was used as a selection criterion for competition horses to estimate racing performance. In the literature the most common approaches to estimate breeding values are the linear or threshold statistical models. However, recent studies have shown that a Thurstonian approach was able to fix the race effect (competitive level of the horses that participate in the same race), thus suggesting a better prediction accuracy of breeding values for ranking trait. The aim of this study was to compare the predictability of linear, threshold and Thurstonian approaches for genetic evaluation of ranking in endurance horses. For this purpose, eight genetic models were used for each approach with different combinations of random effects: rider, rider–horse interaction and environmental permanent effect. All genetic models included gender, age and race as systematic effects. The database that was used contained 4065 ranking records from 966 horses and that for the pedigree contained 8733 animals (47% Arabian horses), with an estimated heritability around 0.10 for the ranking trait. The prediction ability of the models for racing performance was evaluated using a cross-validation approach. The average correlation between real and predicted performances across genetic models was around 0.25 for threshold, 0.58 for linear and 0.60 for Thurstonian approaches. Although no significant differences were found between models within approaches, the best genetic model included: the rider and rider–horse random effects for threshold, only rider and environmental permanent effects for linear approach and all random effects for Thurstonian approach. The absolute correlations of predicted breeding values among models were higher between threshold and Thurstonian: 0.90, 0.91 and 0.88 for all animals, top 20% and top 5% best animals. For rank correlations these figures were 0.85, 0.84 and 0.86. The lower values were those between linear and threshold approaches (0.65, 0.62 and 0.51). In conclusion, the Thurstonian approach is recommended for the routine genetic evaluations for ranking in endurance horses.*

---

**Keywords:** racing performance, breeding value, Thurstonian approach, rider, equine

## Implications

Among the most important decisions that have to be taken in breeding programs is the prediction of a reliable breeding value for a specific trait. The race level is an important factor in the genetic evaluations and Thurstonian models have been demonstrated that they can deal better with this effect. Models that improve the accuracy of prediction will give a better response to selection as it is proportional to the accuracy. Correlations between breeding values across models were high, and then genetic progress seems to be warranted irrespective of the most realistic model.

## Introduction

The ranking in the race is one of the most common traits used as a measure of performance in the genetic evaluation of

competition horses (Tavernier, 1991; Ricard and Chanu, 2001; Ricard and Legarra, 2010). The discontinuous nature of rankings in competitions has promoted the use of different transformations of this trait using linear approaches. The most frequent criteria used in Europe to evaluate the breeding value in sport horses are transformed ranks (Ruhlmann *et al.*, 2009; Ricard and Legarra, 2010): the squared root of ranks (Lühns-Behnke *et al.*, 2006), the Snell-transformation score (Snell, 1964; Gómez *et al.*, 2006) or normalized scores (Aldridge *et al.*, 2000). An alternative is to apply statistical tools that can work with discrete categories, such as the threshold model (Gianola, 1982), that can use full rank without any transformation. However, sometimes it is difficult to define the number of thresholds. The way this categorical trait is scored underestimates the horse variance and the genetic effect and, as a consequence, the heritability of the underlying performance is underestimated (Ricard and Legarra, 2010).

† E-mail: icervantes@vet.ucm.es

On the other hand, the rank of horses is from 1 to the number of participants in each event regardless of the average level of the participants. For example, the first in a short race will not probably be the first in a long race. In fact, endurance competitions are long distance races that vary from 40 to 160 km/day. This type of competition is carried out cross country, where not only speed is important, but also endurance and the adapted physical condition of the horses are essential. Moreover, horses can be excluded after veterinary controls when they exceed a determined heart rate or certain metabolic parameters or reasons related to limbs during the endurance exercise. Rider and rider–horse interaction also plays a crucial role in the competition performance and in welfare aspects (Hausberger *et al.*, 2008; Visser *et al.*, 2008; McLean and McGreevy, 2010). All these factors make it very difficult to define models for genetic evaluation. Ricard and Touvais (2007) studied endurance performance and compared several models to fix the race effect and they concluded that the level of competition is the true measurement of difficulty of the race. However, they established that the rivalry between horses is the most important factor in the race effect. In fact, the model of choice should be able to adjust the results for the competitive level of each event. The so called Thurstonian approach is a competitive model because it describes the competitive ability of animals as a continuous unobserved value to fix the race effect (Ricard and Legarra, 2010). This model for quantitative genetic effects could be applied to data consisting of ranks using the truncated Gaussian distribution, and the variance parameters are estimated through a full Bayesian approach (Gianola and Simianer, 2006). The competitive models were previously used in other sport performance such as in trotter horses (Gómez *et al.*, 2011), where horses with the highest ranking in a low race level of competition had similar genetic value as the horses with a lower rank in a high race level, but their use is scarce.

The Spanish Arabian Horse Breeder Association (AECCA) is currently developing a breeding program aimed at improving endurance performance. This selection objective is common in other crossbreeds such as Anglo-Arabian, Hispano-Arabian and Spanish Sport Horse. The refinement in the genetic model used for breeding value predictions will improve the genetic response in endurance performance.

Therefore, the aim of this study was to compare the prediction abilities of linear, threshold and Thurstonian models for the genetic evaluation of ranking trait in endurance horses as well as to deepen the abilities of the less used models.

## Material and methods

### Data

The total data set consisted of 4065 ranking records from 966 horses (415 males, 386 females and 165 geldings) aged between 5 and 24 years. Most of them were Arabian horses (69.2%) and the rest were Anglo-Arabian (19.6%), Spanish Sport Horse (7.1%) and other breeds such as Spanish

Purebred, Spanish-Arabian and Thoroughbred horses (4.1%). The Arabian horse is the most used breed in this performance, but many breeders are interested how other sport horses perform regardless of the breed. The records were collected during 581 endurance races held between 2000 and 2014. The average records per event were seven with a minimum of three and a maximum of 42. 13% of events had more than 11 participants. The number of records per horse ranged between 1 and 25 with an average of 4.2 (Table 1).

Regarding rider information, the number of different riders in the data set was 777. The average of different riders riding one horse were 1.7 and the percentage of horses that were ridden by different riders was 39.8%. Also, the average number of riders who rode different horses was 2.1 and the percentage of riders who rode at least two different horses was 44.1%.

The rider–horse interaction known as ‘match’ effect involves the relationship, communication and co-operation between rider and horse. This is influenced by the level of experience and behavior of both rider and horse. This is arrived at by combining the rider–horse pair and measure the different behavior of a horse with specific rider. This effect can be disentangled because one rider rides different horses and vice versa. Match had an average of 2.5 records across 1651 levels.

Pedigree information for genetic evaluation totaled 8733 animals: 47% Arabian horses, 21% Thoroughbred,

**Table 1** Structure and descriptive statistic of endurance races held between 2000 and 2014

	Number of horses	Number of records
Total	966	4065
Female	386	1527
Male	415	1724
Gelding	165	814
Anglo-Arabian horses	189	719
Spanish sport horses	69	258
Arabian horses	668	2938
Other breeds	40	150
Average records by event	–	7.0
Average records by horses	4.2	–
≤6 years old	157	238
7 years old	302	547
8 years old	379	734
9 years old	359	706
10 years old	279	559
11 years old	242	494
12 years old	167	309
13 years old	103	203
14 years old	68	146
>14 years old	61	129
Participated at two different ages	205	–
Participated at three different ages	144	–
Participated at four different ages	80	–
Participated at least five different ages	85	–

15% Spanish Purebred, 10% Anglo-Arabian, 2% Spanish Sport Horse and 5% other breeds. The mean of equivalent complete generations was 7.5 for Arabian, 3.5 Thoroughbred, 6.1 Anglo-Arabian and 3.9 Spanish Sport Horse participants. This parameter was computed with ENDOG 4.8 (Gutiérrez and Goyache, 2005).

## Methods

### Models

The genetic model included the following systematic effects: gender (three classes: male, female and gelding), age (10 classes: first are horses of 5 to 6 years, next classes from 2 to 9 are horses from 7 to 14 years old, respectively, and tenth are horses >14 years) and race (581 classes). The race includes, among other things, the level of difficulty of the event. The random effects were: rider, match, permanent environmental, animal genetic and residual. The model to be solved for a hypothetical trait considering all of the possible random effects was:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{Qr} + \mathbf{Nm} + \mathbf{e}$$

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

where  $\mathbf{y}$  is the vector of observations containing the trait rank at finish,  $\mathbf{X}$  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 effects,  $\mathbf{N}$  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,  $\mathbf{m}$  the vector of match effects,  $\mathbf{e}$  the vector of residuals.

Eight different models were tested that included different random effects. Model A did not include any random effects besides the animal and residual effects, model R additionally included the rider effect, M the match effect, P the permanent environmental effect, RP the rider and permanent environmental effects, MP the match effect and the permanent environmental effect, RM the rider and match effects, RMP the rider, match and permanent environmental effects.

The genetic models were performed using three approaches:

1. A Thurstonian model suggested by Gianola and Simianer (2006): this model assumed an underlying variable that is transformed into the ranking in the competition and makes it possible to include and evaluate the competitive level of horses in races. The model for this underlying

variable ( $\mathbf{l}$ ) was:

$$\mathbf{l} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{Wp} + \mathbf{Qr} + \mathbf{Nm} + \mathbf{e}$$

and the joint posterior density was:

$$\begin{aligned} & \rho(\mathbf{l}, \mathbf{b}, \mathbf{u}, \mathbf{p}, \mathbf{r}, \mathbf{m}, \sigma_u^2, \sigma_p^2, \sigma_r^2, \sigma_m^2 | \mathbf{T}, \delta) \\ & \propto \prod_{k=1}^K \left\{ \prod_{j=1}^J \left[ N(l_{jk} | \mu_{jk}, 1) \right]^{\delta_{jk}} T_k \right\} \\ & N(\mathbf{u} | \mathbf{0}, A\sigma_u^2) N(\mathbf{p} | \mathbf{0}, I\sigma_p^2) N(\mathbf{r} | \mathbf{0}, I\sigma_r^2) \\ & N(\mathbf{m} | \mathbf{0}, I\sigma_m^2) \rho(\sigma_u^2) \rho(\sigma_p^2) \rho(\sigma_r^2) \rho(\sigma_m^2) \end{aligned}$$

where  $\mathbf{T}$  is the matrix of indicator sets ( $T_k$ ) that denotes the order of liabilities, or ranking, observed in competition  $k$  and  $\delta = \{\delta_{jk}\}$  vector of indicator variables that takes 1 or 0 depending on the presence of the  $j$  observation on the  $k$  competition. Further,  $K$  is the number of matches and  $J$  the number of observations. The earlier distributions for the variance components were assumed uniform between appropriate bounds.

The analysis was performed using a Gibbs sampler with a data augmentation approach for the liabilities. Given the liabilities, the conditional sampling of each level of  $\mathbf{u}$ ,  $\mathbf{r}$ ,  $\mathbf{m}$  and  $\mathbf{p}$  was univariate Gaussian and the conditional distribution of  $\sigma_u^2$ ,  $\sigma_r^2$ ,  $\sigma_m^2$  and  $\sigma_p^2$  was univariate inverted  $\chi^2$ . Thus, the key step for the implementation was the conditional sampling of liabilities from the predictive distributions. The conditional predictive distributions for the liabilities of competition  $k$  are proportional to:

$$\rho(l_k | \text{ELSE}) \propto \prod_{j=1}^J \left[ N(l_{jk} | \mu_{jk}, 1) \right]^{\delta_{jk}} T_k$$

where ELSE denotes the unknown liabilities at all other competition events and systematic, additive genetic, RMP horse effects; thus, within event  $k$ , the liabilities are not independent, because the sample space is restricted by the knowledge of the ranking, represented by  $T_k$ . The conditional sampling of liabilities is performed sequentially. The liability of the rank at finish position one was set to 0. The next one was calculated as:

$$l_{nk} \sim TN_{(l_{n-1(k)}, l_{n+1(k)})}(\mu_{nk}, 1)$$

With  $nk$  the competing horse  $n$  in competition  $k$ ;  $n-1$  is the competing horse in the previous position of ranking than  $n$  horse and  $n+1$  the horse with the next position of ranking. The performance of the excluded horses is assumed to be worse than the last classified horse.

2. Threshold model (Gianola, 1982; Gianola and Foulley, 1983; Sorensen and Gianola, 2002): under this model, it is assumed that an underlying non-observable variable exists, defining the different categories of the categorical trait if this underlying variable exceeds a particular threshold value. Eight categories were defined, with the excluded horses in the eighth category, seven or higher

ranking the seventh category and the other (1 to 6) categories corresponding to the rank value.

- Linear model: assuming that the analyzed trait was a continuous variable with normal distribution and using the raw ranks obtained per horse in each race. The excluded horses in the race were included after the last position at +1 additional position within races.

In linear and threshold models, the analyses were carried out with TM software (Legarra *et al.*, 2008). 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. Thurstonian models were solved using software developed by the authors (Gómez *et al.*, 2011).

#### Cross-validation

A cross-validation approach was used to evaluate the prediction ability (Efron and Tibshirani, 1993) of different models. For each genetic model, the entire data set was randomly split into a training data set containing 3060 records (75%) to estimate the parameters and solve each model, and a validation data set with 1005 records (25%). Performances of the validation data set were reconstructed using the solutions for the model effects previously obtained with the training data sets. To compare real and predicted performances of the validation set, the Pearson's correlations ( $\rho$ ) between real and predicted ranking performances were calculated. In total, 20 different random replicates of this procedure were carried out, and the Pearson's correlations were averaged across replicates as a measure of the predictive ability of each model.

#### Breeding value analysis

The breeding values of horses for ranking trait were predicted with the best model of each approach regarding

cross-validation procedure above. To evaluate the possible changes in the genetic ranking of individuals according to their breeding values, the Pearson's and Spearman's correlations between breeding values, predicted with the best models inside approaches, were calculated for all animals in the pedigree, with top 20% and top 5% of animals according to genetic ranking.

## Results

Table 2 shows the heritabilities, proportion of total phenotypic variance attributed to rider, match and environmental permanent effects (ratios), and the averages of correlations between real and predicted performances for each model and approach. The heritability of the trait had from low to medium magnitude across models and approaches, as expected for a categorical trait. The range of heritability values were similar for linear and threshold models, between 0.06 to 0.13 and 0.06 to 0.14, respectively, whereas Thurstonian model values were higher and ranged between 0.08 and 0.21. The highest heritability across models was obtained using the model without random effects (beside animal and residual), followed by models with one additional random effect in all approaches. The RMP model that included all the identified random effects had the lowest heritability.

Regarding the importance of random effects, the permanent environmental effect ratio ranged between 0.03 and 0.05 in linear and threshold models and this range was higher for Thurstonian models (0.03 to 0.11). The rider effect had the similar trend, obtaining values of 0.04 in linear models, the range was higher in threshold models (0.04 to 0.06), and the largest values for this effect were obtained in

**Table 2** Heritabilities ( $h^2$ ), rider effect ratio ( $r^2$ ), match effect ratio ( $m^2$ ), environmental permanent effect ratio ( $p^2$ ), residual ratio ( $e^2$ ) and average of correlations between real and predicted performances ( $\rho$ ) and their standard deviation for different approaches (linear, threshold and Thurstonian) and models (A to RMP)

Model...	A	R	M	P	RP	MP	RM	RMP
<b>Linear</b>								
$h^2$	0.13	0.11	0.10	0.08	0.06	0.07	0.09	0.06
$r^2$	–	0.04	–	–	0.04	–	0.04	0.04
$m^2$	–	–	0.06	–	–	0.05	0.03	0.03
$p^2$	–	–	–	0.05	0.05	0.04	–	0.04
$e^2$	0.88	0.85	0.84	0.87	0.85	0.84	0.84	0.84
$\rho$	0.5864 ± 0.0353	0.6010 ± 0.0424	0.6018 ± 0.0301	0.5127 ± 0.0391	0.6114 ± 0.0305	0.5725 ± 0.0358	0.5869 ± 0.0351	0.5682 ± 0.0369
<b>Threshold</b>								
$h^2$	0.14	0.10	0.09	0.10	0.08	0.07	0.08	0.06
$r^2$	–	0.06	–	–	0.06	–	0.05	0.04
$m^2$	–	–	0.10	–	–	0.08	0.06	0.06
$p^2$	–	–	–	0.05	0.03	0.03	–	0.03
$e^2$	0.86	0.84	0.82	0.86	0.83	0.81	0.81	0.81
$\rho$	0.2349 ± 0.0129	0.2438 ± 0.0126	0.2486 ± 0.0123	0.2382 ± 0.0128	0.2437 ± 0.0128	0.2453 ± 0.0129	0.2506 ± 0.0126	0.2492 ± 0.0129
<b>Thurstonian</b>								
$h^2$	0.21	0.15	0.11	0.11	0.09	0.08	0.10	0.08
$r^2$	–	0.11	–	–	0.11	–	0.06	0.06
$m^2$	–	–	0.20	–	–	0.19	0.15	0.14
$p^2$	–	–	–	0.11	0.07	0.05	–	0.03
$e^2$	0.79	0.74	0.69	0.78	0.73	0.69	0.71	0.69
$\rho$	0.5974 ± 0.0131	0.6033 ± 0.0133	0.5920 ± 0.0184	0.5975 ± 0.0139	0.6038 ± 0.0133	0.5967 ± 0.01832	0.6014 ± 0.0139	0.6067 ± 0.0134

A = no additional random effects; R = rider effect; M = match effect; P = permanent environmental effect; RP = rider and permanent environmental effects; MP = match effect and the permanent environmental effect; RM = rider and match effects; RMP = rider, match and permanent environmental effects.

Thurstonian models (0.06 to 0.11). The highest estimated value for random effects was the match across models, which was greater than heritabilities in threshold and Thurstonian models; its value ranged from 0.14 to 0.20 for Thurstonian models, from 0.06 to 0.10 in threshold models and from 0.03 to 0.06 in linear models. The Thurstonian approach had the highest ratios for all random effects and lower residual ratio (Table 2).

As for model predictability, the average correlations between real and predicted performances in threshold models were lower (0.24 to 0.25) than in linear (0.51 to 0.61) and Thurstonian models (0.59 to 0.61) (Table 2). The correlations showed low variability within an approach but the differences were observed among approaches. Despite the differences between models within approaches were small, it was possible to find higher correlations in models which included the rider and permanent effect (RP) within a linear approach. Regarding the Thurstonian approach the best model was that which included the RMP effect. Finally, for the threshold approach it was RM. Note that despite non-significant differences were found between models within approaches, we chose the model with the highest correlation. The lowest correlations in the cross-validation were found for models that did not include rider random effect. Model P in particular had lower correlation values in linear models.

Table 3 shows the Pearson's and Spearman's correlations (above and below the diagonal, respectively) among predicted breeding values obtained with the three best models within each approach (RP for linear, RM for threshold and RMP for Thurstonian approaches) for all animals, top 20% in the genetic ranking and top 5%. Note that the negative values obtained when comparing Thurstonian with both threshold and linear models were because of the opposite definitions of the trait in Thurstonian models: the lower the rank in the race, the best its performance. Therefore, the values through text are presented in absolute values. The highest Pearson's correlation was found between threshold and Thurstonian approaches in all cases. They were 0.90, 0.91 and 0.88 for all animals, top 20% and top 5% best animals, respectively. The lowest Pearson's correlations were between threshold-linear pair: with all animals (0.75), top 20% (0.70) and top 5% (0.58). Regarding Spearman's correlation a similar trend was observed. The best percentage of coincidence was between Thurstonian and threshold

with all animals (0.85), top 20% (0.84) and top 5% (0.86), and the worst values with the linear-threshold pair (0.65, 0.62 and 0.51). In all cases, the Thurstonian model had more in common with the threshold model than with linear models. In order to compare the same model for the three approaches, Table 3 was also computed with the RMP model across approaches and the magnitude of correlations was similar; there was no changes in the conclusion.

## Discussion

Ranking allows a horse entering an event to be compared with the other horses. However, the level of the event has to be determined first (Tavernier, 1991). The estimation of race level has always been an important issue in horse breeding programs. Several models were fitted to deal with this requirement, such as the Tobit model (Bayesian model with data augmentation). Bugislaus *et al.* (2011) concluded that Trotters without earnings did not show their real racing potential and should be regarded as censored observations. In their study, the Tobit model and Threshold models with censored race results were shown to be very suitable for genetic evaluation. However, they are useful to include in the genetic evaluation of individuals without data, but they are not able to account for the competition level of the race. Here we use a Thurstonian approach that implies the assumption of a continuous underlying variable that represents the racing performance of the horse. This underlying variable is restricted by the ranking position of horses within a race, but as horses perform in several races, it takes into account the ranking at the same time that the competition level of the races. Thurstonian models are able to predict breeding values according to their racing abilities in more detail since the same horses are not ranked in each race.

In cross-validation analysis, the Thurstonian models had the highest prediction ability for ranking trait in endurance horses with an average across models of 0.60. Although the connection between races by related animals could partially separate the genetic component and estimate a breeding value independent of the level of the race, only in Thurstonian models does the prediction of a horse breeding value take into account the influence of the level of the other horses participating in the same race. This seems to be a better fit. Thus, an additional participating horse ranked

**Table 3** Pearson's and Spearman's correlations coefficients<sup>1</sup> between model RP linear (Lin) and models RM for threshold (Thr) and RMP for Thurstonian (Thu) approaches (with all animals, 20% top and 5% top animals) breeding values for ranking trait

	All data			20% Top			5% Top		
	Lin	Thr	Thu	Lin	Thr	Thu	Lin	Thr	Thu
Lin		0.75	-0.89		0.70	-0.76		0.58	-0.65
Thr	0.65		-0.90	0.62		-0.91	0.51		-0.88
Thu	-0.80	-0.85		-0.71	-0.84		-0.58	-0.86	

RP = rider and permanent environmental effects; RM = rider and match effects; RMP = rider, match and permanent environmental effects.

<sup>1</sup>Pearson above and Spearman below of the diagonal.

before would not imply moving backward an extra position of the horse (linear) nor moving the animal to the next category (threshold). The Thurstonian approach just assesses an intermediate liability for the extra participant without modifying the position of the rest of participants.

Correlations between expected and predicted values for ranking trait were also high for the linear approach (0.58) models, but they were lower for the threshold approach of around 0.25. Although there were no significant differences, the models with highest correlations were the RP among the linear approach, RM for threshold and RMP for Thurstonian approach. The low predictability of the threshold approach showed the difficulty to establish the number of thresholds that could explain the real data. The threshold approach assumes the presence of an underlying normal distribution. A threshold demonstrates the discontinuity in the visible scale, but some categories included several categories of rankings in order not to increase the number of thresholds. Note that by increasing the number of participants in one race, it also increases the number of animals in the seventh category. Many of these horses could have performed at a lower rank in races with a lower number of participants. From a different perspective, the sixth category is the last in a race with six participants, but this result is in the top 20% in a race with 40 participants. This would increase the unexplained phenotypic variance. With reference to the linear approach, despite the correlation being similar including one or more additional random effects (environmental permanent effect, rider and match effect), the predictability of the linear approach seems to be more affected by the inclusion or not of random effects. In this approach the inclusion of rider effect and environmental permanent effect gave the best results. In fact the inclusion of just rider or match gave very similar results. On the other hand the inclusion of match effect with any of the other two effects decreased the predictability, and the inclusion of only environmental permanent effect gave the lowest correlation in the linear approach. Therefore, the rider/match effect seemed to be important to include (Kearsley *et al.*, 2008; Bartolomé *et al.*, 2013; Sánchez-Guerrero *et al.*, 2014). Nevertheless, the linear and Thurstonian models gave a high predictability as in the trot at hand in dressage, accumulated points in show jumping traits for Swedish horses (Viklund *et al.*, 2010) or for racing performance of Norwegian and North Swedish Coldblooded trotters (Olsen *et al.*, 2012). Otherwise the predictabilities were lower than in previous cross-validation studies for dressage and show jumping (Rovere *et al.*, 2016), for dressage scores in Spanish trotter horses (Sánchez-Guerrero *et al.*, 2014), and for racing time in German trotters (Bugislaus *et al.*, 2005). These differences might be explained by a different data structure in which the relevance of the systematic effects on the phenotype became more or less high and by a larger number of records per horse (Sánchez-Guerrero *et al.*, 2014), with an average of 7.2, whereas in the present study the average number of records per horse was 4.2. As for other random effects the average number of records per match level was 5.7 (Sánchez-Guerrero *et al.*, 2014) and 2.5 in the present study. The average for different

riders for one horse was around 2 as in this study, but the average for different horses ridden by one rider was lower (1.3) in the study by Sánchez-Guerrero *et al.* (2014) than in this endurance data set.

Heritability in general was higher in Thurstonian models (0.08 to 0.21) than in linear (0.06 to 0.13) or threshold models (0.06 to 0.14), as in a recent study for ranking trait in Thoroughbred horses, with 0.05 to 0.29 for Thurstonian and 0.03 to 0.23 for linear models (da Gama *et al.*, 2016). Estimated heritabilities for models with more random effects in the present study were in the range found in the literature for ranking traits (transformed or not) in other disciplines: Blom-scores ranking transformation for jumping Belgian sport horses (0.09; Janssens *et al.*, 1997), rank at finish German trotters horses (0.08; Roehle *et al.*, 2001), the square roots of rank at finish in Thoroughbred horses (0.18; Sobczynska and Lukaszewicz, 2004), probit function transformation of ranking for Thoroughbred horses in France and United Kingdom (0.18 and 0.06; Bokor *et al.*, 2005), square root rank at finish German trotters horses (0.06; Bugislaus *et al.*, 2005), ranking Arabian horses (0.07; Ekiz and Kocak, 2005) and ranking Spanish trotters horses (0.09; Gómez *et al.*, 2011). The lowest values were similar to those reported by Gómez *et al.* (2011; 0.09 and 0.05 for Thurstonian and linear approaches, respectively). Compared with studies for endurance but using different traits, the resulting heritability was similar as placing (0.10), but lower than speed (around 0.20) (Ricard and Touvais, 2007). The heritabilities were overestimated when the additional random effects were not included: i.e. rider, match and environmental permanent effect, which was also observed by Sánchez-Guerrero *et al.* (2014).

The use of a high predictability model could increase the genetic response. The reliability of breeding values is directly affected by the heritability values. Here higher values were found using Thurstonian models. Also the genetic progress could be faster if heritability and correlation between real and predicted breeding value is high. In the Bayesian analysis, the breeding values reliabilities can be computed using the standard deviation of posterior distribution. The average of reliabilities for participants were ranged between 0.15 (Thurstonian) and 0.17 (threshold) across approaches, with no important differences. Note that the data set was too small to obtain high breeding value reliabilities. Models improving the accuracy will give a better response to selection as it is proportional to the accuracy. As correlations between breeding values was high genetic progress seems to be warranted whatever is the most real model.

The match effect was greater than rider or permanent effect in all approaches, note that the match effect includes the rider effect when this was not present in models M and MP. This reflects the importance of interaction rider–horse effect in models for ranking trait in endurance races. However, for final score of dressage Spanish horses, the rider effect had a higher ratio than the match effect (Sánchez-Guerrero *et al.*, 2014). Concerning the environmental permanent effect, the rider ratio effect was greater than the environmental permanent ratio in the Thurstonian approach for the RP model, but lower in the linear approach. A similar trend occurred in Gómez *et al.* (2011),

where the trainer–driver variance was observed to be greater than environmental permanent ratio in the Thurstonian approach for the RP model but lower in linear approach. Although a direct comparison cannot be done because of the different methodology, this also occurred for working on traits from show jumping in Spanish Sport horses using random regression model: environmental permanent effect ratio had the biggest ratio followed by the match ratio and the rider ratio (Bartolomé *et al.*, 2013). On the other hand, driver component, was lower than environmental permanent effect in German trotters (Roehle *et al.*, 2001) using a linear model. Unlike the rider effect, the environmental permanent and animal effects are usually confounded.

Summarizing, models were similar within approach, but the differences across approaches were meaningful. The best predictability was obtained by Thurstonian models followed by linear models and the threshold model had the worst predictability. The heritability magnitude in Thurstonian models had the highest values with linear and threshold approaches very similar. And finally, regarding the correlation between predicted breeding values the threshold approaches had higher correlations with the Thurstonian approaches than with the linear approaches. Given, therefore, that this last criterion, the correlation between predicted breeding values, is the least powerful (the true model is unknown), it makes not recommendable using threshold models. The predictability criteria favored the Thurstonian approach over the linear approach because of the variability of correlations between the real data set and the resulting data set after reconstructing the real data set in the linear approach. Consequently, the Thurstonian approach can be recommended for routine genetic evaluation instead of the linear or threshold models. Nevertheless, the Pearson's and Spearman's correlation between predicted breeding values using best models for each threshold and Thurstonian approaches were high, suggesting that in the absence of both possibilities (Thurstonian and linear approaches) the threshold approach would not reach such incongruent predicted breeding values, despite the predictability being very low.

## Conclusions

The results of the present study encourage the application of the Thurstonian approach for genetic evaluations for the ranking of endurance horses. The Thurstonian approach is the best by accommodating the number of participants and the level of the race, and, therefore, reaching more accurate predictions for breeding values in this trait for this population. Despite the threshold cross-validation predictability was lower, the threshold–Thurstonian predicted breeding value correlations were higher than linear–Thurstonian. The match effect for ranking trait should be included in endurance horse models since it had the largest ratio thus making it more important than heritabilities in the threshold and Thurstonian models.

## Acknowledgments

The authors thank the Spanish Arabian Horse Breeder Association and to the 'Subsecretaría de Agricultura, Alimentación y Medio

Ambiente' of the Spanish Ministry of Agriculture, Food and the Environment for their support. Also authors are grateful to the Spanish Royal Equestrian Federation for providing the data.

## References

- Aldridge LI, Kelleher DL, Reilly M and Brophy PO 2000. Estimation of the genetic correlation between performances at different levels of show jumping competitions in Ireland. *Journal of Animal Breeding and Genetics* 117, 65–72.
- Bartolomé E, Menéndez-Buxadera A, Valera M, Cervantes I and Molina A 2013. Genetic (co)variance components across age for Show Jumping performance as an estimation of phenotypic plasticity ability in Spanish horses. *Journal of Animal Breeding and Genetics* 130, 190–198.
- Bokor Á, Blouin C, Langlois B and Stefler J 2005. Genetic parameters of racing merit of Thoroughbred horses in steeplechase races. *Italian Journal of Animal Science* 4, 43–45.
- Bugislaus AE, Roehle R and Kalm E 2005. Comparison of two different statistical models considering individual races or racetracks for evaluation of German trotters. *Livestock Production Science* 92, 69–76.
- Bugislaus AE, Stamer E and Reinsch N 2011. The use of a Tobit-like-classification in genetic evaluation of German Trotters. Paper presented at 62nd Annual Meeting of the European Federation for Animal Science, 29 August–2 September 2011, Stavanger, Norway, pp. 132.
- da Gama M, Aspilcueta Borquis R, de Araújo Neto F, de Oliveira H, Fernandes G and da Mota M 2016. Genetic parameters for racing performance of Thoroughbred horses using Bayesian linear and Thurstonian models. *Journal of Equine Veterinary Science* 42, 39–43.
- Efron B and Tibshirani RJ 1993. Cross-validation and other estimates of prediction error. An introduction to the Bootstrap. Chapman & Hall, New York.
- Ekiz B and Kocak O 2005. Phenotypic and genetic parameter estimates for racing traits of Arabian horses in Turkey. *Journal of Animal Breeding and Genetics* 122, 349–356.
- Gianola D 1982. Theory and analysis of threshold characters. *Journal of Animal Science* 54, 1079–1096.
- Gianola D and Foulley JL 1983. Sire evaluation for ordered categorical data. *Genetic Selection Evolution* 15, 201–223.
- Gianola D and Simianer H 2006. A Thurstonian model for quantitative genetic analysis of ranks: a Bayesian approach. *Genetics* 174, 1613–1624.
- Gómez MD, Cervantes I, Bartolomé E, Molina A and Valera M 2006. Genetic evaluation of show jumping performance in young Spanish Sport Horse Breed. In *Book of abstracts of the 57th Annual Meeting of the EAAP*, 17–20 September 2006 (ed. Y van der Honing), pp. 351. Wageningen Academic Publishers, Antalya, The Netherlands.
- Gómez MD, Varona L, Molina A and Valera M 2011. Genetic evaluation of racing performance in trotter horses by competitive models. *Livestock Science* 140, 155–160.
- Gutiérrez JP and Goyache F 2005. A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding and Genetics* 122, 172–176.
- Hausberger M, Roche H, Henry S and Visser EK 2008. A review of the human–horse relationship. *Applied Animal Behaviour Science* 109, 1–24.
- Janssens S, Geysen D and Vandepitte W 1997. Genetic parameters for show jumping in Belgian sporthorses. In *Book of abstracts of the 48th Annual Meeting of the EAAP*, 25–28 August 1997 (ed. Arendonk, J A M Van), p. 5. Wageningen Publishers, Vienna, The Netherlands.
- Kearsley CGS, Woolliams JA, Coffey MP and 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. *Livestock Science* 118, 72–81.
- Legarra A, Varona L and Lopez de Maturana E 2008. TM Threshold model. Retrieved on 26 October 2008 from <http://snp.toulouse.inra.fr/~alegarramanualtm.pdf>.
- Lührs-Behnke H, Röhe R and Kalm E 2006. Genetic analyses of riding test and their connections with traits of stallion performance and breeding mare tests. *Züchtungskunde* 78, 119–128.
- McLean AN and McGreevy PD 2010. Ethical equitation: capping the price horses pay for human glory. *Journal of Veterinary Behavior: Clinical Applications and Research* 5, 203–209.

- Olsen HF, Klemetsdal G, Ødegård J and Árnason T 2012. Validation of alternative models in genetic evaluation of racing performance in North Swedish and Norwegian cold-blooded trotters. *Journal of Animal Breeding and Genetics* 129, 164–170.
- Ricard A and Chanu I 2001. Genetic parameters of eventing horse competition in France. *Genetics Selection Evolution* 33, 175–190.
- Ricard A and Legarra A 2010. Validation of models for analysis of ranks in horse breeding evaluation. *Genetics Selection Evolution* 42, 3–10.
- Ricard A and Touvais M 2007. Genetic parameters of performance traits in horse endurance races. *Livestock Science* 110, 118–125.
- Roehe R, Savas T, Brka M, Willms F and Kalm E 2001. Multiple-trait genetic analyses of racing performances of German trotters with disentanglement of genetic and driver effects. *Archiv Tierzucht* 44, 579–587.
- Rovere G, Ducro BJ, van Arendonk JAM, Norberg E and Madsen P 2016. Analysis of competition performance in dressage and show jumping of Dutch Warmblood horses. *Journal of Animal Breeding and Genetics* 133, 503–512.
- Ruhlmann C, Janssens S, Philipsson J, Thorén-Hellsten E, Crollly H, Quinn K, Manfredi E and Ricard A 2009. Genetic correlations between horse show jumping competition traits in five European countries. *Livestock Science* 122, 234–240.
- Sánchez-Guerrero MJ, Cervantes I, Valera M and Gutiérrez JP 2014. Modelling genetic evaluation for dressage in Pura Raza Español horses with focus on the rider effect. *Journal of Animal Breeding and Genetics* 131, 395–402.
- Snell EJ 1964. A scaling procedure for ordered categorical data. *Biometrics* 20, 592–607.
- Sobczynska M and Lukaszewicz M 2004. Genetic parameters of racing merit of thoroughbred horses in Poland. *Journal of Animal Breeding and Genetics* 121, 302–306.
- Sorensen D and Gianola D 2002. Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer, New York, NY, USA.
- Tavernier A 1991. Genetic evaluation of horses based on ranks in competitions. *Genetics Selection Evolution* 23, 159–173.
- Viklund Å, Braam Å, Näsholm A, Strandberg E and Philipsson J 2010. Genetic variation in competition traits at different ages and time periods and correlations with traits at field tests of 4-year-old Swedish Warmblood horses. *Animal* 4, 682–691.
- Visser EK, Van Reenen CG, Blokhuis MZ, Morgan EKM, Hassmén P, Rundgren TMM and Blokhuis HJ 2008. Does horse temperament influence horse-rider cooperation? *Journal of Applied Animal Welfare Science* 11, 267–284.