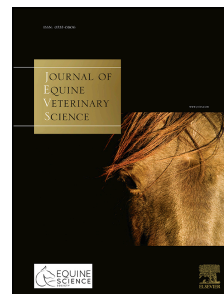


Journal Pre-proof

The impact of horse age, sex and number of riders on horse performance in British Eventing horse trials

K. Hanousek, M. Salavati, B. Dunkel



PII: S0737-0806(20)30341-5

DOI: <https://doi.org/10.1016/j.jevs.2020.103250>

Reference: YJEVS 103250

To appear in: *Journal of Equine Veterinary Science*

Received Date: 10 June 2020

Revised Date: 22 August 2020

Accepted Date: 31 August 2020

Please cite this article as: Hanousek K, Salavati M, Dunkel B, The impact of horse age, sex and number of riders on horse performance in British Eventing horse trials, *Journal of Equine Veterinary Science* (2020), doi: <https://doi.org/10.1016/j.jevs.2020.103250>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Inc.

The impact of horse age, sex and number of riders on horse performance in British Eventing horse trials

K Hanousek^{*1}, M Salavati², B Dunkel¹

¹Dept of Clinical Science and Services and Research Support Office, The Royal Veterinary College, North Mymms, Herts, AL9 7TA, UK.

²The Roslin Institute, Royal (Dick) School of Veterinary Studies, Easter Bush campus, Midlothian, EH25 9RG, UK

* Author to whom correspondence should be addressed

Email: khanousek5@rvc.ac.uk

Phone: 01707 666008

Fax: 01707 666304

Word count: 2943

Ethical Considerations: Approval for the study was granted from the Clinical Research Ethics Board of the Royal Veterinary College (SR2018-1642)

Competing Interests: None

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

All raw data are available at <https://www.britisheventing.com/compete/fixtures-and-results>.

The datasets analysed in this study are available in the FigShare repository,

<https://doi.org/10.6084/m9.figshare.11888769>.

Journal Pre-proof

Background: Limited research has been undertaken to determine the impact of horse age, sex and number of riders on horse performance in British Eventing (BE) horse trials.

Improved understanding of this can aid professionals in planning a competition horse's career.

Objectives: To investigate the impact of age, sex and number of riders on the peak performance of horses at each of the main levels of BE competitions.

Methods: The best score from each horse competing in BE horse trials in the years 2008 to 2018 were recorded, principal component and hierarchical cluster analysis was performed. Basic data analysis was used to identify variables associated with particular better-performing clusters of horses. The interplay of the combinatory variables was then used to map out the trends in career trajectory for horses competing at each level of competition in the best and worst performing clusters.

Results and conclusions: The peak performance of mares was worse than geldings and stallions at all levels. At Novice to Advanced, stallions did not perform as consistently with multiple riders as geldings. The age at which the best performing groups peaked was similar for mares and geldings in all classes, although stallions peaked at an older age than mares and geldings at Novice and Intermediate level. All horses were a minimum of four years old at the time of competition, as per British Eventing rules.

Keywords: equine, competition, show jumping, dressage, training

Abbreviations:

BE	British Eventing
HP	Horse performance dataset
HRP	Horse-rider performance dataset
BPC	Best-performing cluster

WPC Worse-performing cluster

1. Introduction

Eventing is a popular equestrian sport in the UK among amateur and professional riders and has featured in the summer Olympic games since 1912 [1]. It involves three phases of competition: Dressage, show jumping and cross country. Each horse/rider combination will take part in each phase over 1–3 days, and performance across all phases is judged to give a final score. Performance is scored by penalties incurred at each phase. The final score is a cumulation of all penalties incurred, with the lowest numerical score ranking the highest. The mean final score varies between 47.7 and 84.2 [2], with this value representing the sum of all penalties accumulated in the three phases of competition [3].

British Eventing (BE) is the governing body for the sport in Britain, with five main classes of competition; ‘BE90’, ‘BE100’, ‘Novice’, ‘Intermediate’, and ‘Advanced’, in ascending order of difficulty. The technical difficulty and height of the fences are increased with each class [4].

As with any equestrian sport, riders are motivated to select horses with attributes that might improve performance. This will likely involve considering horse age, breed, size, temperament, and sex, as well as the horse’s competition history, training and price. However, currently, there is only limited information available as to whether any of these factors truly influence performance.

Previous studies on eventing are limited, but one report documented better performance in geldings and stallions than mares [2]. The permanent environmental effect (non-genetic repeatable contributors to phenotypic variance, such as training and nutrition) has been shown to be the most important component, followed by the rider and horse genetics, which become more important as competition classes become more difficult [5]. Correlation

between classes of competition has been shown to be high [2,5], indicating that performance at lower levels can be used to forecast performance at higher levels. However, this is limited by the horse's potential ability and does not account for horses which have already peaked at a higher level of competition and now compete at a lower level due to age, soundness, change of rider, or other factors.

O'Brien et al. previously calculated wastage in BE horse trials at 33.7 percent, and cited veterinary problems, sale of horse, and lack of ability as reasons listed by owners of horses who did not re-register with BE. Of the horses not re-registered, lack of ability accounted for 28 percent and poor selection of horses was listed as a potential cause for this. O'Brien et al. describe the need for studies to investigate the selection and training of horses being used for eventing.

Currently, no study has investigated the interplay of multiple variables simultaneously. While the impact of the rider has been measured, there is no data indicating the impact of the number of riders on horse performance. This is relevant as it might aid in understanding the importance of the horse-rider relationship on athletic performance in eventing. Information about the influence of age, sex and the number of previous competing riders on performance in BE horse trials may aid riders and trainers in selecting appropriate horses and planning an animal's athletic career.

This study aims to determine how the measured factors influence horse performance in BE horse trials at each class from BE90 to Advanced, by investigating any relationship that horse age, sex, and number of competing riders may have with one another and with peak horse performance.

2. Materials and Methods

Data for every horse competing in BE horse trials between 2008 and 2018 were collected; this information is publicly accessible on the BE website [4]. Most horses appeared multiple times in the data, depending on how many times they have competed in their career. In order to eliminate multiple entries for the same horse, the data were split into two further datasets; horse performance (HP) included the best rank of each horse in every class in which it competed; Horse-rider performance (HRP) included the best performance of each unique horse/rider combination in each class in which they competed. The performance outputs gathered from the BE website included placing, final score, dressage penalties, show jumping penalties, show jumping time penalties, cross country penalties and cross country time penalties. The horse inputs included horse sex, horse year of birth (year foaled), year when scored, age when scored, class, rider name and number of riders (for HRP). The data were analysed by individual class. Scores where horses were eliminated or withdrawn from competition were not used in either dataset, even when this was the only (and therefore best) score available for that horse, as these horses were not given a finishing placing.

2.1 Statistical analysis

2.1.1 Principal component analysis:

All analyses were performed in R (version ≥ 3.4)¹ unless stated otherwise. The data were scaled and centred prior to principal component analysis being performed on each class. Principle component analysis was performed on both HP and HRP in order to establish which performance output contributed most to the variation. The performance outputs included

¹ R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

placing, final score, dressage penalties, show jumping penalties, show jumping time penalties, cross country penalties and cross country time penalties.

2.1.2 Hierarchical clustering analysis:

Hierarchical clustering was performed based on the principal component analysis to establish the structure of inherent clustering within the datasets. Because of this, cluster membership takes into account all measured variables. Cluster membership was then added as a categorical variable to all datasets.

2.1.3 Model design:

A generalised linear model was produced, and a 50-fold cross validation performed, to assess how well this model would generalise to an independent dataset.

Generalised linear model:

$$\text{Placing} \sim \text{AgeScored} + \text{Sex} + \text{RiderName} + \text{NoRiders} + \text{YearFoaled} + \text{YearScored} + \text{Cluster}$$

2.1.4 Strength of the models:

A Wald test was performed on each model, the output of which shows the extent to which removing each variable would damage the predictive power of the model.

For each model an R^2 value, root mean standard error and mean absolute error were calculated. R^2 is a measure of the variation which is not explained by the model, calculated as a result of a 50-fold cross validation. Higher values, indicate a better fit. However, with higher values of R^2 there is a risk of 'over-fitting' in which case the value may not represent a true relationship [6]. Root mean standard error is the square root of the variance of residuals. It indicates the absolute fit of the model and can be interpreted as the standard deviation of

the unexplained variance; Root mean standard error accounts for very large errors in the model, so a lower value indicates that large errors are unlikely to have occurred [7]. The result is in the same unit as the response variable, in this case placing [7]. Mean absolute error measures the difference between the observed and predicted values, with lower values indicating a better fit [8]. The results of the R^2 , root mean standard error, and mean absolute error are presented in Supplementary Section 1.

3. Results

3.1 Descriptive statistics

As a dataset, HP describes the peak performance for each horse, and HRP describes the peak for each horse/rider combination. As such, HRP is a larger dataset and comprised a total of 105,828 scores, and HP a total of 75,292 scores. Table 1 shows a description of the data for each individual class and all classes for both HP and HRP, which provides background information on the average best performance. The mean placing ranged from 6th to 11th place, with the best mean placings within each dataset at Novice level (6th in HP and 8th in HRP) and the worst at Advanced (9th in HP and 11th in HRP). The mode year scored for all classes was 2008, except BE90 which was 2017. This suggests that either more horses competed at BE90 in 2017, or more horses achieved their best score in 2017. The mean age scored was 9 years old for BE90 to Novice, 10 years old for Intermediate and 11 years old for Advanced.

3.1.1 Mean placing and final score

HP: Horses achieved the best final score at BE100 and the worst at Advanced. In terms of placing, horses performed best at Novice followed by BE100, BE90, Intermediate and worst in Advanced.

HRP: Horses achieved the best mean final score, at BE100 and the worst at Advanced. In terms of placing, horses performed best in Novice, followed by BE100, then BE90 and Intermediate and worst in Advanced.

3.1.2 Sex status

HP: Stallions were the lowest represented sex status ($n = 520$) and geldings were the most common ($n = 53514$), mares were less common than geldings ($n = 21258$). As one score was recorded from each horse with available scores in HP, this can be considered to be a direct representation of the population of horses competing which finish at least one competition without withdrawal or elimination.

HRP: As in HP, stallions were the lowest represented sex status ($n = 1337$) and geldings were the most common ($n = 75597$), mares were less common than geldings ($n = 288294$). As one score was taken for each unique horse/rider combination at each level this is not a direct representation of the competing population.

3.2 Principal Component Analysis

Table 2 shows the top three contributing variables for the first three principal components for each class. This shows which of the measured variables contributed to the most variation within each dataset. The first contributing variable in all cases except HP BE90 was finishing

placing, which is why this was chosen as the output variable used to measure performance in the generalised linear model and cluster analysis.

3.2 Wald test

The results of the Wald test can be seen in Table 3.

HP: For all models the variable which would damage the predictive power of the model the most is the cluster, and the sex status of the horse would damage the power of the model the least.

HRP: The variable which would damage the predictive power of the models most is cluster, and the variable which would damage the models the least is sex status, with the exception of Intermediate, in which the number of riders is the variable which damages the model least.

3.3 Cluster membership

Figure 1 shows the distribution of the peak performance of mares, geldings and stallions across each cluster for all classes. Clustering is inherent within a dataset, and cluster analysis allows exploration of the characteristics of any clustering present within a dataset. In this case, principal component analysis indicated that the variable contributing to the most variation in the dataset was placing, by plotting the distribution of the placing of horses in each cluster it is possible to identify groups of horses which, for some reason, perform better or worse than others. Placing is on the Y axis and the width of the plot indicates the density of horses in that group at each placing.

In HP classes BE100 to Advanced cluster 1 has a higher density of horses finishing in a better placing compared to clusters 2, 3 and 4 making cluster 1 the best-performing cluster (BPC) for these classes. Clusters 3 and 4 have a high distribution of horses finishing in worse placings, making these the worse-performing clusters (WPC) for these classes. Cluster 1 has

the greatest density of horses finishing in worse placing for HP BE90, making this the WPC for HP BE90, and cluster 3 has the highest density of horses finishing in better placings making cluster 3 the BPC.

In all HRP classes cluster 3 has a higher density of horses finishing in first place compared to clusters 1 and 2, making cluster 3 the best-performing cluster (BPC) for these classes. In all HP classes cluster 2 has the greatest density of horses finishing in worse placings, making this the low-performing cluster (WPC) for these classes.

3.3.1 Sex

The distribution of mares, geldings and stallions across all clusters is detailed in Supplementary Section 2. Comparison of this and Figure 1 illustrates the relationship between horse sex and cluster membership. This provides insight into the relevance of horse sex on peak performance in BE horse trials. A greater proportion of mares, geldings, or stallions in a particular better-performing or worse-performing cluster in Supplementary Section 2 indicates an association between that sex and better or worse peak performance.

HP: For BE100 to Advanced cluster 3 has the highest proportion of mares and lowest proportion of geldings compared to other sexes. Cluster 1 always has a greater proportion of stallions, compared to other sexes, as demonstrated in Supplementary Section 2. Cluster 1 is the BPC for HP BE100 to Advanced, and clusters 3 and 4 are the WPCs for these classes. This means that the BPC for HP BE100 to Advanced has more stallions than other clusters, so stallions are associated with better performance. The BPC for these classes always has the lowest number of mares, proportionally, and the WPC always has the highest number of mares. This suggests that mares are associated with worse performance than stallions and geldings.

At BE90 there is a greater proportion of stallions in cluster 1 and 3 compared to mares and geldings. For BE90, cluster 3 is the BPC and cluster 1 is the WPC. This indicates that at this level, stallions are over-represented in both the best and worst performing clusters, and less in the mid-performing clusters. The difference between mares and geldings in the clusters at BE90 is less noticeable than at the higher levels, but there are still more geldings in the BPC.

HRP: Cluster 2, the WPC for all classes in HRP, has the highest proportion of mares and the greatest proportion of stallions (Supplementary Section 2). Cluster 3, the BPC, has the most stallions at BE90 and BE100, and the fewest at Novice to Advanced. This is in contrast to HP, where stallions are consistently associated with BPC, suggesting that at Novice to Advanced levels stallions may perform very well with their best rider, but not as well with other riders.

There are generally more geldings than mares in cluster 3 (BPC).

At Advanced level in cluster 3 stallions perform better than mares and geldings, with all stallions placing 1st to 5th (figure 1).

3.3.2 Age

Table 4 shows the ages associated with best and worst performance for each sex in each class in HP. This is based on cross-referencing of cluster performance and the distribution of ages in each cluster. The distribution of mares and geldings was similar in all classes. There is relatively little data available for stallions, meaning that interpretation of the distribution should be cautious. In HP, age contributed approximately 20% of the variation within the dataset, and was in the top three contributing variables for all classes except Intermediate. In HRP age contributed approximately 15% in Novice and Advanced only. In all cases for HP

age is the most important measurable variable for predicting performance, according to the results of the Wald test (Table 3). Due to the nature of the HRP dataset the discrepancies between the distribution of ages in each cluster is reduced. Interpretation of age is more accurate in HP. Mares in the WPC in all classes tend to be aged 10 years old or over, and the same is true for geldings at Novice level. For all classes and all sexes the age of peak performance ranged from 5 years old (with the exception of BE90 stallions in the BPC) and 10 years old (with the exception of BPC mares at Advanced). In general, horses in the BPCs achieved their peak score at a younger age than horses in the WPCs, except for stallions at Novice level and geldings at Intermediate.

3.3.3 Number of riders

In BPC, more mares and geldings had one rider at BE90, and more geldings had one rider at Advanced. Conversely, more stallions had two riders at Advanced. Overall this suggests that mares and geldings perform better with one rider at BE90 and Advanced, and stallions perform better with two riders at Advanced.

Stallions in WPC had one rider from Novice to Advanced level, mares and geldings had two riders at BE90 and Novice, and geldings had two riders at BE100 and Advanced. Low performing stallions tended to have one rider at Novice to Intermediate, and low performing mares and geldings had two riders at BE90 and Novice, and BE100 and Advanced (geldings only).

4. Discussion

BE horse trials, the equestrian's triathlon, is a popular equestrian sport in the UK designed to test the horse and rider in a range of skills [4]. Wastage in the sport has been calculated to be around 33.7% over one year [9] and improving the selection of horses for eventing may help

to reduce the level of wastage. Performance in this study was measured as the finishing placing. Final score is a less comparable measure of performance; it has been demonstrated that the dressage phase contributes a majority (63%) of the final score at BE90 and BE100 [10], and is subjectively judged [11]. This may not be true for higher levels, as the cross country becomes more technically challenging and the optimum times ranges become narrower [3], the dressage phase may contribute less to the final score. The dressage judges are required to be more qualified as the levels increase, and for Advanced level there are two judges for the dressage phase [3], which may reduce the effect of subjective judging. Using finishing placing also lessens the effect of the subjective dressage judging as horses in the same class at a competition are scored by the same judge [12]. Principle component analysis also revealed that finishing placing accounts for the most variation in the dataset. The skill of the rider could not be controlled for within the scope of this study, but is likely to be a confounding factor.

4.1 Clusters

Hierarchical clustering indicated inherent grouping within the datasets (N=4). The predictive models were reliant on these clusters, which were retrospectively assigned, to predict peak performance. Therefore, the models can only be used retrospectively to quantify the importance of variables, not to forecast horse performance. However, the factors which differentiate the clusters can be explored. It may be possible for future studies to predict performance if more variables are introduced to the models, one of which could include analysis of a horse/rider's previous scores, as this may well be useful in prediction of peak performance.

4.2 Sex

When characterising the importance of sex on performance, the imbalanced proportions of mares, geldings, and stallions must be considered. Overall, there is a pattern of stallions and geldings outperforming mares, which is supported by previous work [2,9,13]. In HP, fewer stallions were in the middle cluster than in the best and worst clusters suggesting that stallions have more polarising performance patterns than mares and geldings. This could be related to distractions, or indicate they are more challenging to ride; stallions have been shown to have increased salivary cortisol during the breeding season, which is not seen in mares and geldings, and could indicate a higher propensity for stress [14]. Stallions were also less common in the best performing cluster for Novice to Advanced in HRP, which is in contrast to HP, suggesting that they do not perform as well at these levels with every rider, but with the right rider-horse combination they will perform very well. Despite this, sex is the generally the weakest variable in predicting performance (Supplementary Section 2).

4.3 Age

Age has been demonstrated to significantly affect horse performance in eventing [5] and other disciplines [15–17]. It is logical that age plays a greater role in HP, as this is the lifetime best score. In HRP, it may be less relevant as horses are likely to be ridden by one rider in each season, and therefore only have a set period in which to achieve their best score with each rider. This study investigates a horse's career best score, but there is no set age at which horses begin or end their competitive career. Younger horses in the population may not yet have achieved their career peak, which may be causing information bias relating to the age of peak performance.

Previous authors [18] have explored the impact of age on the genetic potential of show jumping horses, characterising the age-genotype interaction. This highlighted a phenotypic plasticity, or environmental sensitivity and demonstrated that horses can be broadly split into those with a precocious response (decreasing genetic potential with age and peaking early), a robust response (no change in genetic potential with age), and a gradual response (increasing genetic potential with age and peaking late). As each score analysed was the peak in that horse's career, the data suggest that in most classes, horses with a precocious response will achieve a higher final rank at their peak than those with a gradual or robust response. This is based on comparison of the mean age of peak performance for horses in the dataset, with the age at which horses in the best performing clusters peak. At Novice level younger stallions perform worse than older stallions (Table 4), suggesting that a precocious response is not an advantage at this level. It is also possible that the younger stallions in this situation are moving up the levels, and the older stallions are more experienced and have stepped down from higher levels of competition due to age, change of rider, or soundness. It is not possible to establish if this is the case without analysing competition history, which was beyond the scope of this study.

Overall, the best performing horses peak at similar ages, regardless of sex status. However, in general the worst performing mares peak at an older age than the worst performing geldings and stallions. This may indicate that mares with a gradual response are likely to be the worse performers. Betros et al. have demonstrated no change in the maximal heart rate or aerobic capacity of young (~7 year old) and middle aged (~15 year old) mares [19], suggesting that fitness is unlikely to be a component in the reduced performance of older mares who respond gradually to their genetic potential.

Studies investigating age of peak performance in eventing horses previously have found horses to peak at older ages than those described in this study [5]. This discrepancy is likely

due to the fact that cluster analysis allows investigation not of the peak performance of all horses, but the peak performance of the best horses in the dataset. To produce a more robust analysis of the impact of age the previous experience of each horse could be incorporated. It might then be possible to chart horses' careers, gaining more insight into the dynamics of the variables impacting performance.

4.3 Number of riders

The measure of number of riders in this study includes the total number of riders each horse had at that class of competition in its career. This does not account for the number of riders each horse had at the date the scores were taken. This makes the measure of number of riders less specific, but an overall trend can still be seen.

The results suggest that all sexes are likely to perform better with one rider. The exception are stallions that perform better with multiple riders at Intermediate and Advanced level competitions. As previously mentioned, stallions are also less common in the best performing cluster for Novice - Advanced in HRP, which is in contrast to HP. This may indicate that the horse-rider combination plays a more important role for stallions at these higher levels; While their lifetime best score is likely to be better than that of mares (HP), performance is likely to improve if a stallion has multiple riders. However, with some of these riders performance will be poor (HRP). The impact of the rider has been demonstrated to be greater at higher levels of competition for eventing [5].

Hypothetically, multiple riders increase the chance of the best possible horse-rider combination being achieved, which may explain why stallions with one rider are seen in worse performing clusters at Novice to Advanced. However, having only one rider is likely to strengthen the horse-rider relationship, but this concept is difficult to quantify. This may be

useful information to riders considering a stallion, as the data suggests that performance with a different rider is a less accurate indicator of performance in stallions than in mares and geldings. The rider has been shown to directly influence the gait of horses [20], and the impact extends beyond ridden work to management and training [21]. Work has been done to quantify this [22], which is likely to prove valuable in future assessment of the horse-rider relationship.

One of the limitations of this study is that the experience of the rider has not been quantified, the impact of an inexperienced rider has been well analysed by Williams and Tabor [21]. To incorporate this, future research on this topic could include the number of BE events attended by each rider, and analysis of their previous scores.

Despite its limitations, removing the number of riders as a variable damaged the predictive power of the models for HRP to about the same extent as age scored. Future research in this area could use the highest level at which each rider has competed, or the best score for each rider in that particular class as variables for quantifying rider skill.

1. Conclusion

Overall geldings and stallions perform better than mares in all classes. Stallions appear to have more rider-specific performance; their performance is more polarising (features strongly in the best and worst clusters) in both datasets. Stallions are more common in the best performing cluster in HP, but not as common in HRP. The age at which each sex performs best is class-dependent, but relatively consistent for mares, geldings, and stallions. Age may be less important at Advanced level which could indicate that experience can compensate for age.

Legends for tables, figures and supplementary sections:

Table 1 shows descriptive statistics for the best score of all horses competing in each of the British Eventing classes from BE90 to Advanced in both datasets (HP and HRP), between 2008 and 2018. HP includes the best score each horse achieved in each class, and HRP includes the best score each unique horse/rider combination achieved in each class. Table 1 describes the mean placing, final score, and age scored, the mode year scored and year foaled (year of birth), the median number of riders that competed a particular horse in each class, and the frequencies of mares, geldings and stallions. Frequency is abbreviated to 'freq'.

Table 2 shows contribution of principal component analysis of the peak performance of British Eventing horses competing in all classes from BE90 to Advanced, between the years 2008 and 2018. The table shows the top three variables contributing to the principal components, and their absolute percentage contribution. Variables include placing, Place; show jumping penalties, SJ; show jumping time penalties, SJT; cross country jumping penalties, XCJ; and age. For all classes except BE90 in HP placing is the variable contributing to the most variation in the dataset (~33%), and age scored contributes the most for BE90 (18.1%). Age and cross country jumping penalties are the second and third contributing variables in all cases except BE90 where show jumping is second (15.4%) and Intermediate where show jumping time is third, contributing 13.1%. For all classes in HRP placing contributes the most variation (~33%), and cross country jumping is the third biggest contributor (~14%). The second biggest contributor for BE90, BE100 and Intermediate is show jumping time, and for Novice and Intermediate it is age scored.

Table 3 shows the results of the Wald test for the generalised linear model produced using the peak performance of horses competing in British Eventing between 2008 and 2018. The

colour gradient indicates which variables would damage the model the most (green) and the least (red) if removed.

Table 4 shows the ages (years) of peak performance for horses in clusters (as assigned by hierarchical cluster analysis) associated with best and worst performance for each sex in each class of British Eventing competition. Data included the best score from all horses competing at British Eventing horse trials in the years 2008 to 2018, from each class of competition from BE90 to Advanced. Where no value is shown it indicates that the discrepancy in the data was not enough to draw conclusions for that category. A summary of the impact of sex on performance is described on each row.

Figure 1 shows the distribution of placing across sexes for each cluster in each class for HP and HRP. Data includes the peak scores for each horse (HP), and each horse rider combination (HRP) competing at British Eventing horse trials between 2008 and 2018, in classes BE90 to Advanced. The sex of the animal has been plotted (x axis) against the placing (y axis) in each class, and the plots are split into groups along the Y axis based on cluster number. The distribution of horses is shown in 3 colours (geldings, mares and stallions) as a violin plot overlaid by the interquartile regions as a boxplots. The width of the violin plot indicates the density of horses in that cluster and sex at that placing. Lower numerical values on the Y axis are associated with better performance (first place is better than tenth).

Supplementary Section 1 displays the R^2 value, root mean standard error, and mean absolute error of each model in datasets HP and HRP. These parameters are used to assess how well the model fits the data. The models are designed to retrospectively predict the best final rank of horses in each class.

Supplementary Section 2

Supplementary section 2 shows the percent of each sex of horse that falls into each cluster for all classes in both datasets. This describes the distribution of each sex across clusters. The datasets are based on the peak horse performance (HP) and horse-rider performance (HRP) in all horses competing in British Eventing between 2008 and 2018, and are split into the main classes of competition ranging from BE90 to Advanced.

Manufacturers' details: R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Acknowledgements

The authors would like to thank Renee Groeninx, Executive Assistant at British Eventing Head Office for her assistance in accessing the data and Stephen Papaloizou for his assistance with data handling.

References

- [1] EQUESTRIAN / EVENTING - SUMMER OLYMPICS n.d.
<https://www.olympic.org/equestrian-eventing> (accessed August 7, 2019).
- [2] Hanousek K, Salavati M, Fouladi-Nashta A. Effect of horse gender on British Eventing competition performance: an observational study between 1998 and 2016. Vet Rec 2018.
- [3] British Eventing Members Handbook 2019. <http://www.britisheventing.com/asp->

- net/page.aspx?section=681&itemTitle=Rules+and+Scoring.
- [4] British Eventing 2009. <http://www.britisheventing.com/> (accessed January 3, 2017).
- [5] Kearsley CGS, Woolliams JA, Coffey MP, Brotherstone S. 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* 2008;118:72–81. <https://doi.org/10.1016/J.LIVSCI.2008.01.009>.
- [6] Sapra RL. Using R2 with caution. *Curr Med Res Pract* 2014;4:130–4. <https://doi.org/10.1016/J.CMRP.2014.06.002>.
- [7] Grace-Martin K. Assessing the Fit of Regression Models - The Analysis Factor n.d. <https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/> (accessed May 29, 2019).
- [8] Watson N. Using Mean Absolute Error to Forecast Accuracy - Contemporary Analysis 2012. <https://canworksmart.com/using-mean-absolute-error-forecast-accuracy/> (accessed June 24, 2019).
- [9] O'Brien E, Stevens KB, Pfeiffer DU, Marr D, Hall J, Marr CM. Factors associated with the wastage and achievements in competition of event horses registered in the United Kingdom 2005. <https://doi.org/10.1136/vr.157.1.9>.
- [10] Whitaker T, Hill J. Scoring analysis of completing pre-novice event horses at six selected events. *Equine Comp Exerc Physiol* 2004;1:185–92. <https://doi.org/10.1079/ECEP200417>.
- [11] Whitaker T, Hill J. Dressage scoring patterns at selected British Eventing novice events. *Equine Comp Exerc Physiol* 2005;2:97–104. <https://doi.org/10.1079/ECP200554>.
- [12] Seo JW, Couteau E, Umek P, Hernadi K, Marcoux P, Damjanovic M, et al. BRITISH EVENTING - MEMBERS' HANDBOOK 2016. vol. 22. 2016.

- [13] Whitaker TC, Olusola O, Redwin L. The influence of horse gender on eventing competition performance. *Comp Exerc Physiol* 2008;5:67.
<https://doi.org/10.1017/S1478061508017039>.
- [14] Aurich J, Wulf M, Ille N, Erber R, von Lewinski M, Palme R, et al. Effects of season, age, sex, and housing on salivary cortisol concentrations in horses. *Domest Anim Endocrinol* 2015;52:11–6. <https://doi.org/10.1016/j.domaniend.2015.01.003>.
- [15] Albertsdóttir E, Eriksson S, Näsholm A, Strandberg E, Árnason T. Genetic analysis of competition data on Icelandic horses. *Livest Sci* 2007;110:242–50.
<https://doi.org/10.1016/J.LIVSCI.2006.11.008>.
- [16] Maršálek, Sedláčková, Secká. The influence of the age, sex and performance level of horses on their success in the show jumping competition. *J Cent Eur Agric* 2006;6.
<https://doi.org/10.5513/JCEA.V6I4.334>.
- [17] Oki H, Sasaki Y, Willham RL. Genetics of racing performance in the Japanese Thoroughbred horse: *J Anim Breed Genet* 1994;111:128–37.
<https://doi.org/10.1111/j.1439-0388.1994.tb00446.x>.
- [18] Bartolomé E, Menéndez-Buxadera A, Valera M, Cervantes I, Molina A. Genetic (co)variance components across age for Show Jumping performance as an estimation of phenotypic plasticity ability in Spanish horses. *J Anim Breed Genet* 2013;130:190–8. <https://doi.org/10.1111/jbg.12001>.
- [19] Betros CL, McKeever KH, Kearns CF, Malinowski K. Effects of ageing and training on maximal heart rate and $\dot{V}O_{2max}$. *Equine Vet J* 2002;34:100–5.
<https://doi.org/10.1111/j.2042-3306.2002.tb05399.x>.
- [20] Peham C, Licka T, Schobesberger H, Meschan E. Influence of the rider on the variability of the equine gait. *Hum Mov Sci* 2004;23:663–71.
<https://doi.org/10.1016/J.HUMOV.2004.10.006>.

- [21] Williams J, Tabor G. Rider impacts on equitation. *Appl Anim Behav Sci* 2017;190:28–42. <https://doi.org/10.1016/J.APPLANIM.2017.02.019>.
- [22] Schöllhorn WI, Peham C, Licka T, Scheidl M. A pattern recognition approach for the quantification of horse and rider interactions. *Equine Vet J* 2006;38:400–5. <https://doi.org/10.1111/j.2042-3306.2006.tb05576.x>.

Journal Pre-proof

TABLE 1:

<i>Class</i>		<i>Mean placing</i>	<i>Mean final score</i>	<i>Mean age scored</i>	<i>Mode year foaled</i>	<i>Mode year scored</i>	<i>Median number of riders</i>	<i>Freq. mares</i>	<i>Freq. geldings</i>	<i>Freq. stallions</i>	<i>Total</i>
<i>HP</i>	BE90	8	40.39	9	2005	2017	–	8106	17522	132	25760
	BE100	7	37.97	9	2003	2008	–	7612	18620	186	26418
	Novice	6	42.86	9	2002	2008	–	3756	10548	126	14430
	Intermediate	8	50.77	10	2002	2008	–	1457	5150	59	6666
	Advanced	9	61.39	11	2002	2008	–	327	1674	17	2018
	All	7	41.5	9	2003	2008	–	21258	53514	520	75292
<i>HRP</i>	BE90	9	42.57	9	2004	2017	1	10625	23625	165	34415
	BE100	9	40.48	9	2003	2008	2	10728	27708	248	38684
	Novice	8	45.37	9	2002	2008	2	5285	15454	632	21371
	Intermediate	9	52.48	10	2001	2008	1	1866	6780	239	8885
	Advanced	11	63.45	11	2002	2008	1	390	2030	53	2473
	All	9	43.69	9	2003	2008	1	28894	75597	1337	105828

TABLE 2:

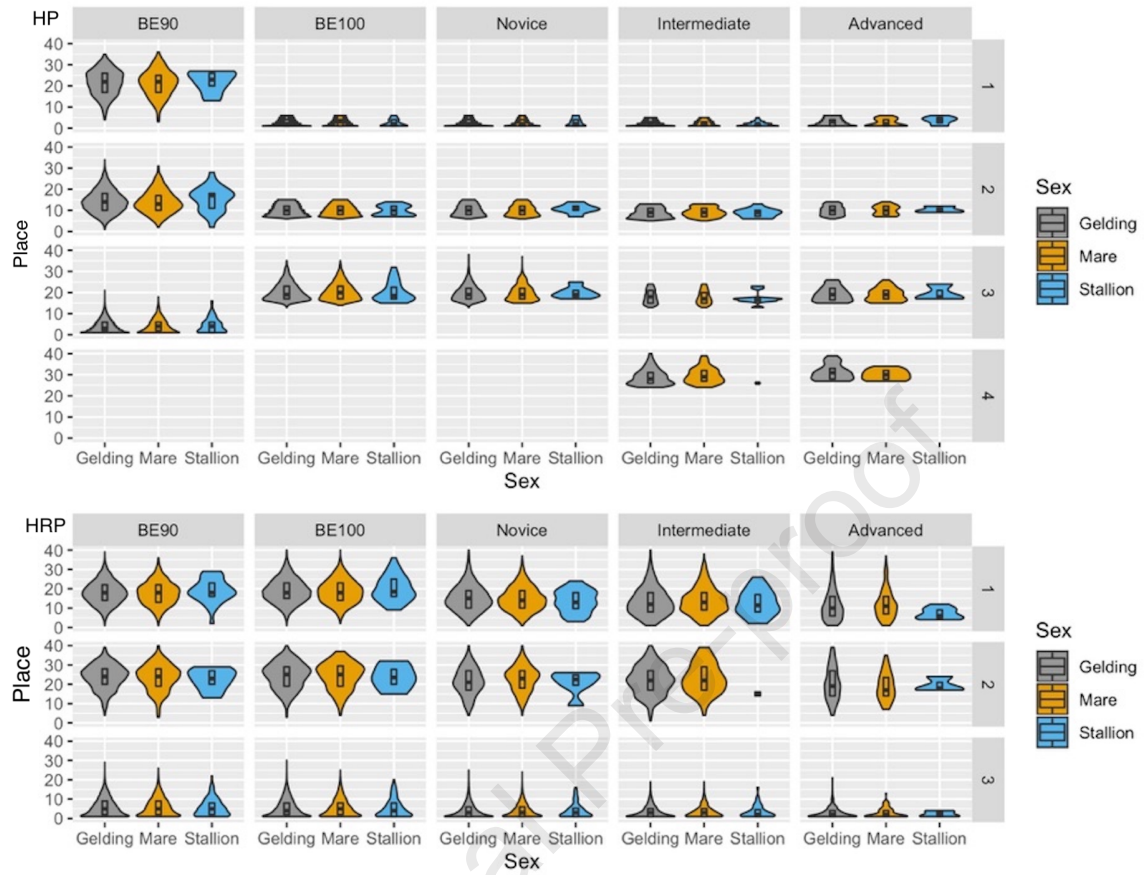
	Class	<i>Variable (absolute % contribution)</i>		
		First	Second	Third
<i>HP</i>	BE90	Age (18.1)	SJ (15.4)	XCJ (14.8)
	BE100	Place (32.9)	Age (25.5)	XCJ (13.3)
	Novice	Place (33.0)	Age (22.6)	XCJ (15.0)
	Intermediate	Place (33.2)	XCJ (14.7)	SJT (13.1)
	Advanced	Place (33.3)	XCJ (15.7)	Age (11.9)
<i>HRP</i>	BE90	Place (33.0)	SJT (14.5)	XCJ (13.7)
	BE100	Place (33.0)	SJT (19.6)	XCJ (14.0)
	Novice	Place (33.2)	Age (16.0)	XCJ (14.1)
	Intermediate	Place (33.2)	SJT (16.8)	XCJ (14.8)
	Advanced	Place (33.3)	Age (14.6)	XCJ (11.3)

TABLE 3:

	<i>Model</i>	<i>Age Scored</i>	<i>Sex</i>	<i>Clusters</i>	<i>Number of Riders</i>
<i>HP</i>	BE90	45.02	9.84	22238.48	–
	BE100	2882.95	58.84	86018.34	–
	Novice	1481.40	95.08	44351.07	–
	Intermediate	212.09	105.74	17782.28	–
	Advanced	18.62	10.43	5732.59	–
<i>HRP</i>	BE90	69.97	6.76	110170.00	1367.90
	BE100	2410.90	4.97	152000.00	1861.00
	Novice	687.98	15.67	55754.51	533.95
	Intermediate	104.00	31.82	18833.48	21.81
	Advanced	13.96	6.47	4349.33	261.37

TABLE 4:

Sex	Performance	BE90	BE100	Novice	Intermediate	Advanced
Mares Associated with overall worse performance	<i>Best cluster</i>	5-7	5-6	6-7	8-9	8-9, 12
	<i>Worst cluster</i>	-	10+	10+	5-6, 10+	10+
Geldings Associated with overall better performance	<i>Best cluster</i>	5	5-6	6-8	7-8	8-9
	<i>Worst cluster</i>	5-8	8+	10+	6	-
Stallions Associated with overall better, more polarising performance	<i>Best cluster</i>	4, 6	5	9-10	8+	-
	<i>Worst cluster</i>	5	6-7	6-8	-	-



Highlights:

- Geldings and stallions perform better than mares at British Eventing Horse Trials
- Stallions may have more rider-specific performance, particularly at Advanced
- Age is less important at Advanced, suggesting experience may compensate for age
- The age of peak performance is relatively consistent for each sex in each class, including BE90, BE100, Novice, Intermediate and Advanced