Training Characteristics of Athletes in Golden Cheetah Open Data

Background: Over the last decade more and more tools for collecting and analysing data have been developed. These tools have also made sharing data with others easier. While this feature is mostly used to share data between athletes and coaches it also provides the possibility to share data with the world. In the past several studies such as (Metcalfe et al., 2017) and (van Erp, Sanders, & de Koning, 2019) have investigated the training characteristics of professional world tour level athletes little research is available about recreational and non-professional athletes. With the rise of open data, however, also large-scale investigations into the training of non-professional athletes becomes possible. One open data set is provided by the maintainers of the software package "Golden Cheetah" (**GC**) (Open Science Framework, 2018). This data set (GCD) contains anonymised data of cyclists, runners, swimmers, triathletes and other unspecified athletes. The metrics about the training within it are calculated with GC.

Methods: The GCD was downloaded from Open Science Framework in March 2020. It currently contains data from 4885 athletes. In a first step corrupted data was removed. Next, remaining data was manually inspected for outliers. Additionally, athletes with fewer than one year of data or less than 50 rides per year were removed as well. A total of 3001 athletes (2989 male, 92 female) were included in the analysis. The athletes were assigned into groups using k-means clustering.

Results: The athletes in the dataset completed 167 rides per year on average. For the male athletes three clusters could be identified while two clusters have been found for females. Table 1 shows the characteristics of the male athletes while Table 2 highlight those of the female athletes.

Discussion: Similar to the findings of van Erp (2019) the results show differences between male and female riders. Since all data is anonymous no conclusions about the quality of the computed metrics can be drawn. Consequently, most metrics provided by GC were not included in this analysis. For male athletes, the difference in mean power between cluster 1 and 3 is small. However, athletes in cluster 1 exhibit a far greater training frequency.

Limitations: While at first glance a data set such as GCD seems to provide a deeper insight into training habits of non-professional athletes the anonymisation and data quality prohibits this.

References

Metcalfe, A. J., Menaspà, P., Villerius, V., Quod, M., Peiffer, J. J., Govus, A. D., & Abbiss, C. R. (2017). Within-season distribution of external training and racing workload in professional male road cyclists. *International Journal of Sports Physiology and Performance*, *12*, 142–146. https://doi.org/10.1123/ijspp.2016-0396

Open Science Framework. Golden Cheetah OpenData Project (2018). https://doi.org/10.17605/OSF.IO/6HFPZ

van Erp, T., Sanders, D., & de Koning, J. J. (2019). Training Characteristics of Male and Female Professional Road Cyclists: A 4-Year Retrospective Analysis. *International Journal of Sports Physiology and Performance*, (July), 1–7. https://doi.org/10.1123/ijspp.2019-0320

Table 1 Characteristics of male athletes. PO mean Power Output during training (mean over all sessions). Work per Week: mean work for weeks training. Work per Day: Mean work per day

training. Sessions per Year: mean number of sessions for an athlete. N: number of athletes in this cluster.

Cluster	PO [W/kg]	Work per Week [MJ]	Work per Day [kJ]	Sessions per Year	N
1	2.43±0.38	5.4±1.59	1914±5014	233±116	620
2	1.89±0.35	1.1±0.81	1029±494	117±53	1120
3	2.49±0.37	2.2±1.02	1325±463	179±83	1077

Table 2 Characteristics of female athletes. PO mean Power Output during training (mean over all sessions). Work per Week: mean work for weeks training. Work per Day: Mean work per day training. Sessions per Year: mean number of sessions for an athlete. N: number of athletes in this cluster.

Cluster	PO [W/kg]	Work per Week [MJ]	Work per Day [kJ]	Sessions per Year	N
1	2.33±0.28	3.7±0.9	1525±465	267±111	19
2	2.15±0.55	1.2±0.7	970±364	157±67	39
3	1.44±0.46	0.4±0.3	432±128	99±35	18