Article

Application possibilities of survival analysis for time-to-event data in animal science

Boróka BÁRDOS [1,2], Ildikó BENEDEK [iD][2], Oleksandr KODAK [2],
István NAGY [iD] [2*]

[1]Institute for Wildlife Management and Nature Conservation, Hungarian University of Agriculture and Life Sciences, Kaposvár Campus, Kaposvár, 7400 Guba S. 40.

[2]Institute of Animal Science, Hungarian University of Agriculture and Life Sciences, Kaposvár Campus, Kaposvár, 7400 Guba S. 40.

ABSTRACT – The application efficiency of several statistical methods was tested based on an open field behavior test of mice. The examined trait was the duration time until the animals approached the experimenter's hand. The available time was fixed in 300 seconds. There were monitored 80 mice belonging to two different species of the Mus genus in equal proportion. Besides, male and female and young and adult animals have also represented the evaluated groups in equal proportions. The data of the examined trait was analyzed with Generalized Linear Models, Kaplan-Meier survival curves, and with Cox Proportional hazard model. The applied statistical procedure provided completely discordant results. According to the GLM results none of the examined factors (species, sex, and age) had significant effects on the examined variable. On the contrary, all factors proved to be significant using a procedure based on the survival analysis. Kaplan-Meier survival curves indicated a higher proportion of individuals successfully approaching the experimenter's hand in all of the compared groups representing different species sexes and ages, respectively. The estimated Cox regression coefficients were significant indicating the significant effects of the species, sex, and age on the investigated trait. Based on the estimated Hazard ratios the probability that in the next time unit a successful approach of the experimenter's hand would occur is three times more likely for one species than the other, twice as much for males and the juveniles than for the females and for the adults. Based on the present study it could be concluded that the successful approach of the experimenter's hand by the mice is clearly "time to event data" thus it is suitable to be analyzed with survival analysis procedures. It could be concluded that applying conventional GLM was not adequate because due to the lack of the successful approach and approach times the censored data should not be used and thus the sample size would largely be reduced.

Keywords: behavior studies, time-to-event data, survival analysis

INTRODUCTION

In any statistical analysis (depending on the research hypothesis) type and distribution of the utilized data will determine the applied statistical method. In animal science, most of the economically important traits are quantitative so that they can be measured and they can be characterized by real numbers providing also the unit of measurement. For these cases, the applied statistical

*CORRESPONDING AUTHOR

Magyar Agrár- és Élettudományi Egyetem (MATE), Kaposvári Campus

✉ *7400 Kaposvár, Guba Sándor u. 40.,* ☎ *82/505-800; 82/505-900*

E-mail: nagy.istva.prof@uni-mate.hu

methods are generally straightforward and widely known. There are however some other data types that are much less common and consequently, adequate analyzing methods are less frequently used. One of these method is survival analysis which is a statistical method targeting the so called time-to-event data. The meaning of the term "event" could be death (*Moreno-Bentancur et al.,* 2017), developing a disease (*Lánczky et al.,* 2016; *Li et al.,* 2018), or any other well-defined action that can unambiguously be detected on a time scale. Conventionally from the human aspect, the occurrence of the event was usually negative therefore it is often called "failure" but the character of the event can also be positive such as the necessary time for the unemployed people to find a new job (*Brück-Klingberg et al.,* 2011). One of the most peculiar characteristics of the survival analysis is that at the end of the predefined examination period generally not all of the monitored subjects had achieved the event (i.e. they are not yet died from cancer, they may not have found a new job, etc). This phenomenon is called censoring which means that at the end of the study the fate of these subjects is simply unknown. Probably the most invaluable characteristic of survival analysis is that the data of these subjects are not wasted but also utilized in the statistical analysis by including a binomial variable providing the status of every subject (i.e. if the event happened or not). Interestingly for certain diseases like AIDS, the related data is left-censored because of the detection limits of this disease (*Chen et al.,* 2014).

In animal science, the most obvious application possibility is to analyze longevity (*Baginé Hunyadi et al.,* 2016a; *Soltész et al.,* 2016; *Török et al.,* 2021) where longevity is describing productivity and usually covers the period between first parturition and culling. Longevity has high economic importance as culling of the animals is involuntary and related to developing certain diseases or leg problems (*Balogh et al.,* 2015). However, depending on the utilization of the animals' longevity can also describe the length of their successful sport carrier (*Sole et al.,* 2017). Nevertheless, other time-to-event data can also be examined such as farrowing interval (*Baginé Hunyadi et al.,* 2016b). Another area where survival analysis is the adequate method for statistical analysis is the animal behavior studies especially those describing the animals' exploratory behavior (*Hansen et al.,* 2015). In earlier studies, exploratory behavior was investigated based on either novel object test (*Fox et al.,* 2009) or open field test (*Fitzgibbon*, 1994), however according to more recent findings these tests measure several dimensions of the behavior and characterize several areas of personality (*Coleman et al.,* 1998). One of the main benefits of these tests is that results are generally repeatable both in novel and in accustomed environments. Performing these tests provides insight to the area of exploration

while in the accustomed environment it characterizes boldness (*Coleman et al.,* 1998). The present study aimed to demonstrate the adequacy of survival analysis using the example of an open field behavior test in mice where its efficiency is compared to an other conventional statistical method. Besides, according to the intention of the authors, the detailed explanation of the survival analysis procedure could also increase its recognition and more frequent use in animal science.

MATERIAL AND METHODS

In the present study, an open field was used to describe mouse behavior. The experiment was performed in the rodent house of the Kaposvár Campus of the Hungarian University of Agriculture and Life Sciences. The rodent house has its mouse breeds, where individuals of known ages, sexes, and backgrounds are kept. The current stock came from several wild populations captured from different parts of the country. The study was performed on 80 laboratory-born offspring of wild animals belonging to two different species of the Mus genus in equal proportion. Besides, male and female and young and adult animals have also represented the evaluated group in equal proportions. Animals were housed in standard laboratory plastic rodent boxes at standard laboratory temperatures of 20-22 ° C. For littering, purified wood chips (LIGNOCELL J. Rettenmaier & Söhne GmbH, Rosenberg, D) were used, complete rodent feed (Ssniff S8106-SO11 Spezialdiäten GmbH, Soest, D) and water were available to the animals ad libitum. For the study, selected individuals were tested individually in a 36.5 x 21 x 18 cm white plastic box. The time of the test is 300 seconds, during which the behavior of the animal was recorded (approach, number of approaches, first approach). Mice were grouped by species, sex and age group. Young mice were selected from 28 to 35 days of age while adult mice from around 500 days.

The statistical analysis of the duration of approaching the experimenter's hand was performed by several methods. After showing with the Shapiro-Wilk normality test that the normal distribution of the data was not given (p<0.001) Generalized Linear Model was applied using the SAS 9.4 software (PROC GLM), where species (one and two) sex (male and female) and age (juvenile, adult) were treated as fixed factors. Besides, the duration of approaching the researcher's hand was analyzed separately for the species, sex and age using the Survival function (*Kaplan and Meier*, 1958) where the survival function (S (t)) estimates the probability (P) that the object of the experiment survives longer (T) than some specified time (t): *S(t) = P(T > t)* in this study, it means that the

animal did not yet approach the hand of the experimenter. Significant differences between the various survival function curves were determined by applying the Log-rank test (*Kaplan and Meier*, 1958). Finally, the effects of species, sex and age on the duration of approaching the experimenter's hand was also analyzed applying Cox Proportional Hazard model (*Cox*, 1972) (also called Cox Regression) where the hazard is the probability that after a certain period the event would occur during the next time unit (i.e. the successful approaching will occur in the next second). The Cox regression will provide the regression coefficients (b) for the breed and the age and raising the coefficients to the power of the exponential constant (appr. 2.718) will provide the hazard ratio or the relative hazard:

$$\text{Hazard ratio} = \exp(b)$$

The hazard ratio is constant for the whole examination period (5 minutes in this case). In the present study, the relative hazard shows the ratio of the probabilities that during the next unit of time a successful approach of the experimenter's hand will occur in the examined species, sex, and age groups, respectively (*Woodward*, 2014). These statistical analyses were performed using the SAS 9.4 software package applying the LIFEREG (*SAS Inst. Inc.,* 2013) and PHREG (*SAS Inst. Inc.,* 2014) procedures. To guarantee the repeatability of these procedures all the SAS script with the embedded data is provided as supplementary material (S1.sas).

RESULTS AND DISCUSSION

The descriptive statistics of the analyzed variable are provided in *Table 1*. It has to be mentioned that in 28 cases the animal did not approach the experimenter's hand at all. According to the data pattern of the groups, the animals representing the first species generally approached the experimenter's hand slower than that of the mice in the second group. Besides, the males and juveniles showed shorter approaching times than that of the females and adults, respectively, however, all group means coincided with large standard deviations. Consequently, according to the GLM results none of the examined factors (species, sex, and age) had a significant effect on the examined variable ($p > 0.05$).

Table 1
The descriptive statistics of the duration of approaching the experimenter's hand

Factor	N	Mean	Standard deviation	Min.	Max.
Species1 Sex 1 Age 1	10	140,20	30,982	0	234
Species 1 Sex 1 Age 2	10	121,50	27,237	0	204
Species 1 Sex 2 Age 1	10	61,30	31,232	0	212
Species 1 Sex 2 Age 2	10	82,00	27,395	0	177
Species 2 Sex 1 Age 1	10	89,80	19,994	0	152
Species 2 Sex 1 Age 2	10	52,90	7,288	0	78
Species 2 Sex 2 Age 1	10	102,80	28,082	0	187
Species 2 Sex 2 Age 2	10	55,40	12,387	0	94

The survival function curves for the different breed and age groups were provided in *Figures 1-3*, respectively. The lower curves of a species, sex, or age group indicate a higher proportion of individuals successfully approaching the experimenter's hand compared to the animals of the other group. Log-rank test showed significant effects of all factors ($p < 0.05$).
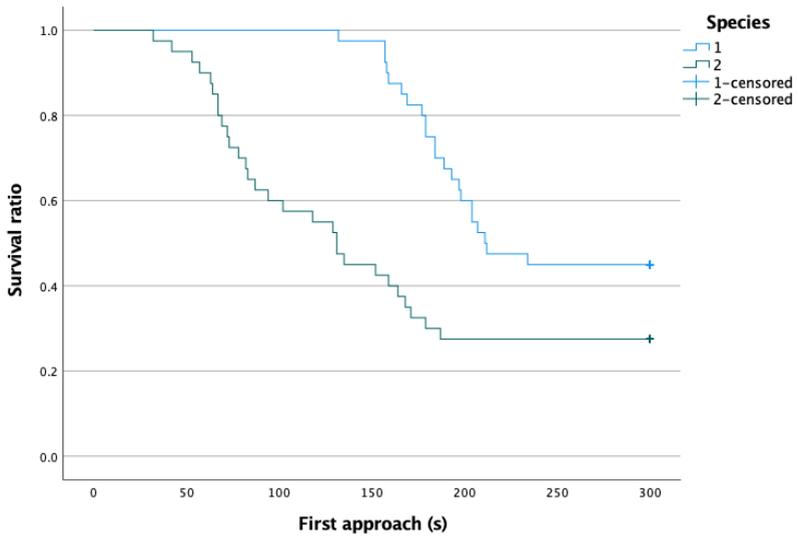
Figure 1. Survival curve of the species
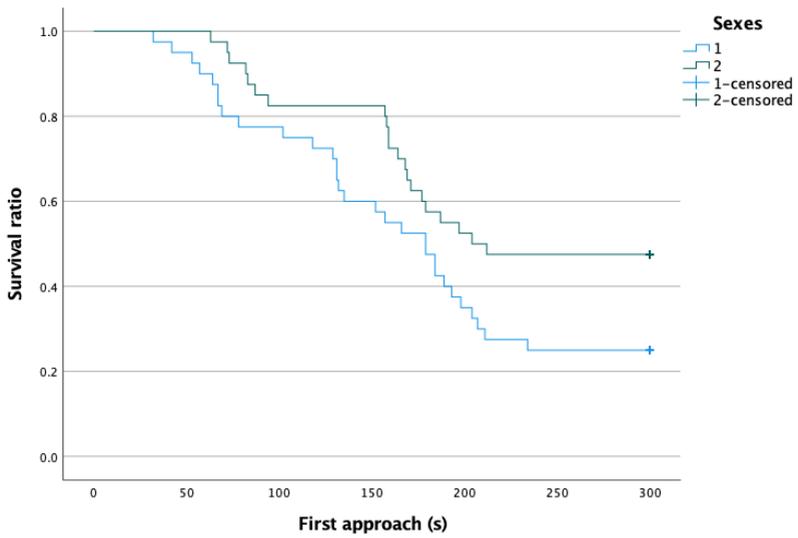
**Figure 2.** Survival curves for the sexes
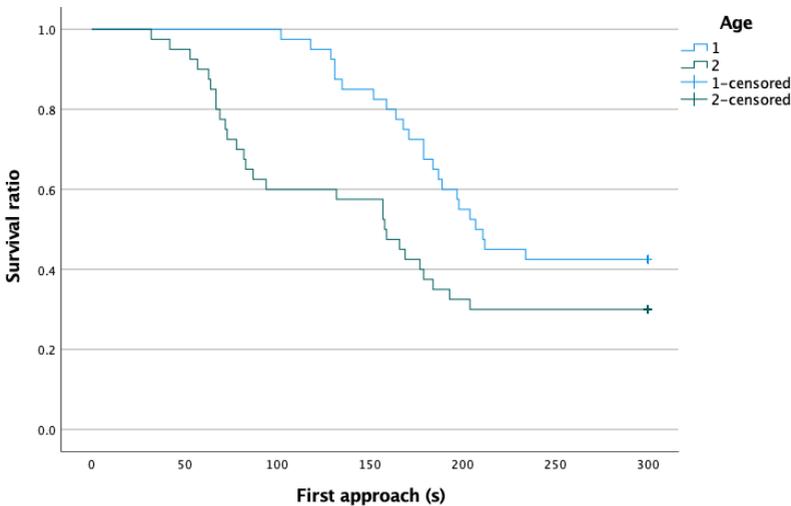
Figure 3. Survival curves for the age groups

In *Figure 1.* it is shown that for the mice of the first group it took appr. 130 seconds until the first animal managed to approach the hand of the experimenter while for the second group this time was 100 seconds faster. The necessary time by which 50% of the mice completed their first hand approach was 211 and 130 seconds, in the first and second groups respectively. The proportion of animals failing to perform any approach was substantially higher in the first species (42,5%) than in the second group (25%). Interpretation of Figures 2-3. are the same and it can be seen that the necessary time by which 50% of the male and female and juvenile and adult mice performed their first approach were 179 and 204 and 158 and 207 seconds, respectively. Besides, the proportion of animals failing to perform any approach was substantially higher in the female (45%) than in the male (22.5%) and in the adult (40%) than in the juvenile (27.5%) group.

The Cox regression parameters estimated by the PHREG procedure are presented in *Table 2.*

The estimated Cox regression coefficients were significant indicating the significant effects of the species, sex, and age on the investigated trait. Based on the estimated Hazard ratios the probability that in the next time unit a successful approach of the experimenter's hand would occur is three times more likely for one species than the other, twice as much for males and the juveniles than for the females and for the adults.

Table 2
Maximum Likelihood estimates of Cox regression coefficients (b) and Hazard ratios examining the effects of species, sex, and age of mice on the duration of their first successful approach to the experimenter's hand.

Factor	b	SE	p	Hazard ratio
Species	1.12	0.30	0.0001	3.08
Sex	-0.77	0.29	0.008	0.46
Age	0.81	0.28	0.005	2.25

Looking at the relevant literature, it is clear that the conventional ANOVA is a frequent evaluation method analyzing behavior tests (*Brantsater et al.,* 2017, *Lee et al.,* 2021). *Brantsater et al.* (2017) even emphasized that ANOVA (rather than Survival Analysis) was used analyzing the approaching time of poultry to a novel object during 120 seconds which trait was a right-censored time to event data. Looking at the observations, of the present study the main problem with ANOVA/GLM was that 28 out of 80 animals did not approach at all the experimenter's hand (S1.sas) which is 35% of the investigated sample. However, a general approach from the researchers that the so-called censored data is often coded as the possible maximal value of the examined trait, and the data is analyzed with ANOVA afterward. This procedure is not appropriate however, as for the censored data the exact value is not known so it should not be included in the analysis at all. The survival function curves are not affected by this problem since the status of the animals (data censoring) has been taken into account. Another advantage of the survival curves is that they are highly informative and provide a clear overview of the number of completed approaches of the experimenter's hand throughout the whole examination period in the examined groups. The only unfavorable characteristic of the Kaplan Meier curves is that they can only describe one factor at a time. Examples for applying Kaplan Meier survival curves can be found in various studies examining for example pig longevity (*Baginé Hunyadi*, 2016a) or mink behavior (*Haage et al.,* 2017). The problem of investigating one factor at a time was not apparent in the Cox regression where the species, sex, and age effects could be examined simultaneously. Using the same methodology as in this study similar results were received in a rabbit behavior study where the running rejection behavior of rabbits on a treadmill was investigated (*Nagy et al.,* 2004). Another example of applying Cox Proportional Hazard was reported in a behavior study of spotted hyenas (*Greenberg et al.,* 2017) where it was observed that the Cox proportional hazards model was significant indicating that subjects in the high-disturbance area were significantly less neophobic than subjects in low-disturbance areas.

CONCLUSIONS

Based on the present study it could be concluded that the successful approach of the experimenter's hand by the mice is clearly "time to event data" thus it is suitable to be analyzed with survival analysis procedures. Applying conventional GLM would not be adequate because due to the lack of successfully approaching the experimenter's hand and approach times the censored data should not be used and thus the sample size would largely be reduced.

Instituional Review Board statement: This research was approved by the Committee on the Ethics of Animal Experiments of Kaposvár Campus (permit number: MATE KC MÁB/21-2/2022). The authors declare that all experiments were performed in accordance with approved guidelines and regulations.

REFERENCES

Balogh, P., Kapelanski, W., Jankowiak, H., Nagy, L., Kovács, S., Huzsvai, L., Popp, J., Posta, J., Soltész, A. (2015). The productive lifetime of sows on two farms from the aspects of reasons for culling. Ann. Anim. Sci., 15, 747-758. DOI: *10.1515/aoas-2015-0020*

Baginé Hunyadi Á., Kusza Sz., Balogh P. (2016a). Magyar nagyfehér, duroc és pietrain kocák túlélés elemzése. Agrártudományi Közlemények, 69, 31-36. DOI: *10.34101/actaagrar/69/1784*

Baginé Hunyadi Á., Kusza Sz., Balogh P. (2016b). Examination of the interval between litters (IBL) of different genotype HLW sows using survival analysis. Agrártudományi Közlemények, 70, 13-17. DOI: *10.34101/actaagrar/70/1810*

Brantsæter, M., Tahamtani, F.M., Nordgreen, J., Sandberg, E., Hansen, T.B., Rodenburg, T.B. Oppermann Moe, R., Janczak, A.M. (2017). Access to litter during rearing and environmental enrichment during production reduce fearfulness in adult laying hens. Appl. Anim. Behav. Sci., 189, 49-56. DOI: *10.1016/j.applanim.2017.01.008*

Brück-Klingberg, A.C., Burkert, A.G., Holger, S., Wapler, R. (2011). Does higher education help immigrants find a job? IAB-Discussion Paper Nr. 6. Nürnberg: IAB.

Chen, Q., May, R.C., Ibrahim, J.G., Chu, H., Cole, S.R. (2014). Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates. Stat. Med. 33(26), 4560–4576. DOI: *10.1002/sim.6242*

Coleman, K., Wilson, D.S. (1998). Shyness and boldness in pumpkinseed sunfish: individual differences are context-specific. Anim. Behav., 56: 927-936. DOI: *10.1006/anbe.1998.0852*

Cox, D.R. (1972). Regression Models and Life Tables. J. Roy. Stat. Soc., 34, 187-220. DOI: *10.1111/j.2517-6161.1972.tb00899.x*

Fitzgibbon, C.D. (1994). The costs and benefits of predator inspection behaviour in Thomson Gazelles Behav. Ecol. Sociobiol., 34, 139-148. DOI: *10.1007/bf00164184*

Fox, R.A., Ladage, L.D., Roth, T.C., Pravosudov, V.V. (2009). Behavioral profile predicts dominance status in mountain chickadees. Anim. Behav., 77, 1441-1448. DOI: *10.1016/j.anbehav.2009.02.022*

Greenberg, J.R., Holekamp, K.E. (2017). Human disturbance affects personality development in a wild carnivore. Anim. Behav., 132, 303–312. DOI: 10.1016/j.anbehav.2017.08.023

Haage, M., Maran, T., Bergvall, U.A., Elmhagen, B., Angerbjörn, A. (2017). The influence of spatiotemporal conditions and personality on survival in reintroductions–evolutionary implications. Oecologia, 183, 45-56. DOI: 10.1007/s00442-016-3740-0

Hansen, M.J., Schaerf, T.M., Ward, A.J.W. (2015). The effect of hunger on the exploratory behaviour of shoals of mosquitofish Gambusia holbrooki Behaviour, 152, 1659-1677. DOI: 10.1163/1568539x-00003298

Moreno-Betancur, M., Sadaoui, H., Piffaretti, C., Ray, G. (2017). Survival analysis with multiple causes of death: extending the competing risks model. Epidemiology 28, 12–19. DOI: 10.1097/ede.0000000000000531

Kaplan, E.L., Meier, P. (1958). Nonparametric Estimation from Incomplete observations. J. Am. Stat. Assoc., 53, 457-481. DOI: 10.1080/01621459.1958.10501452

Lánczky, A., Nagy, Á, Bottai, G., Munkácsy, G., Szabó, A., Santarpia, L., Győrffy, B. (2016). MiRpower: a web-tool to validate survival-associated MiRNAs utilizing expression data from 2178 breast cancer patients. Breast Cancer Res. Treat., 160, 439–446. DOI: 10.1007/s10549-016-4013-7

Lee, K.E., Kim, J.G., Lee, H., Kim, B.S. (2021). Behavioral and cardiac responses in mature horses exposed to a novel object. J. Anim. Sci.Technol. 63, 651-661. DOI: 10.5187/jast.2021.e51

Li, J., Wang, W., Xia, P., Wan, L., Zhang, L., Yu, L., et al. (2018). Identification of a five-LNCRNA signature for predicting the risk of tumor recurrence in breast cancer patients. Int. J. Cancer, 143, 2150-2160. DOI: 10.1002/ijc.31573

Nagy, I., Szabó, A., Romvári, R., Szendrő, Zs. (2004a). Brief description of the survival analysis procedure using the running rejection behaviour of young rabbits as a model trait. Agr. Consp. Sci., 69, 29-33.

SAS Institute Inc. (2013): SAS/STAT® 13.1 User's Guide, The Lifereg Procedure, Cary, NC.

SAS Institute Inc. (2014): SAS/STAT® 13.2 User's Guide, The Phreg Procedure, Cary, NC.

Solé, M., Sánchez, M.J., Valera, M., Molina, A., Azor, P.J., Sölkner, J. Mészáros, G. (2017). Assessment of sportive longevity in Pura Raza Espãnol dressage horses.Livest. Sci., 203, 69–75. DOI: 10.1016/j.livsci.2017.07.007

Soltész, A., Baginé Hunyadi, Á., Kusza, Sz., Balogh, P. (2016). Survival analysis of sow longevity and lifetime reproductive performance – Review. Agrártudományi Közlemények, 70, 75-80. DOI: 10.34101/actaagrar/70/1821

Török, E., Komlósi, I., Szőnyi, V., Béri, B., Mészáros, G., Posta, J. (2021). Combinations of Linear Type Traits Affecting the Longevity in Hungarian Holstein-Friesian Cows. Animals, 11 (11), 3065. DOI: 10.3390/ani11113065

Woodward, M. (2014). Epidemiology, Study Design and Data Analysis. 3rd Edition. CRC Press Boca Raton.