Right censored data ('cohort bias') in veterinary life span studies

S. R. URFER

Veterinary Record (2008)
163, 457-458

S. R. Urfer, BS, DVM,
Division of Animal Housing and Welfare, Vetsuisse Facility, University of Berne, Switzerland

IT is common for veterinary life span and survival estimates to be derived purely from death data and thus grouped by time of death during analysis. This leads to an artificial decrease of the estimated life span due to right censored data ('cohort bias'). This short communication uses practical examples from the literature to emphasise the necessity of applying appropriate statistical methods (such as Kaplan-Meier estimation) to life span and survival research in veterinary medicine.

With regard to survival analysis, right censored data are defined as data from the population under consideration in which death has not yet occurred at the time of study. When the sample consists of dead animals only and is sorted by death date instead of birth date, this causes a misleading decrease in life span estimates due to the fact that a certain percentage of individuals from a birth cohort are still alive at the time of sampling and thus do not appear in the death statistics. This bias is obvious when considering birth cohorts, but becomes masked when considering death cohorts.

In order for such life span evaluations to be representative, all individuals within the birth cohorts studied must already be dead. If this is not the case, and if only data of dead individuals are available, the measured life span will be influenced by the fact that death data of the individuals who are still alive will be censored. Given that these will eventually die at an older age than the dead individuals from the same birth cohort, the estimated life span will be too low. The distribution of causes of death may also be affected.

For these reasons, statistical methods have been developed to correct right censored data in survival analyses, which are commonly applied in studies from human medicine (Kalbfleisch and Prentice 2002). The most commonly used of these methods is Kaplan-Meier estimation (also known as the 'product limit estimator'), which can be used to estimate the actual survival function of the population while taking censored data into account. In some cases, right censored data can also be corrected by eliminating individuals from the data that come from birth cohorts where a significant number of individuals can be assumed to still be alive.

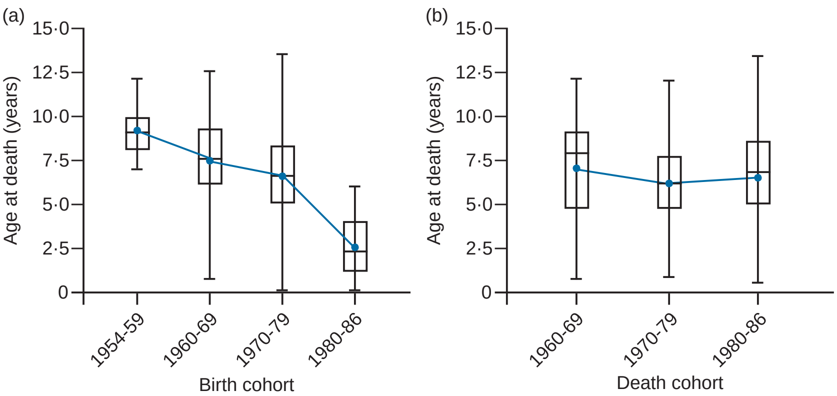

FIG 1: Death data of Irish wolfhounds used in Bernardi (1986). Dogs are grouped by (a) birth cohorts, and (b) death cohorts, as is common in life span studies

TABLE 1: Statistical measures of the Irish wolfhound death data provided by Bernardi (1986), demonstrating the influence of right censored data through elimination of the censored birth cohorts*

Parameter†	All dogs	Dogs from Uncensored birth cohorts	Dogs from Censored birth cohorts
Number	577	326	251
Maximum	13·50	13·50	10·00
95 per cent dead	10·42	11·25	8·88
3rd quartile	8·25	9·23	7·04
Median	6·58	7·46	5·42
1st quartile	4·83	6·00	3·75
5 per cent dead	1·73	3·17	0·96
Minimum	0·50	0·75	0·50
Mean (sd)	6·47 (2·62)	7·37 (2·47)	5·31 (2·34)

* Dogs from uncensored birth cohorts died at a significantly older age than dogs from censored birth cohorts ($P<0{\cdot}0001$, $R^2=0{\cdot}15$, $F=102{\cdot}94$, type III sum of squares).
† Ages are given in years

However, veterinary life span studies commonly analyse a population of dead animals (for example, coming from animal cemetery records, clinical databases or postmortem examination records) and simply use their ages at death to calculate a survival function as well as other statistical parameters regarding life span. Given that data from such sources are usually sorted by time of death rather than time of birth, right censored data are masked by this approach.

If, for example, the average life expectancy of a hypothetical group of dogs was calculated by taking the population of individuals who died in the year 2004 as the sample, the result would be too low due to right censored data. For instance, no members of the 2000 birth cohort would be recorded to have reached over four years of age. Obviously, this result does not imply that the maximum life expectancy for this cohort was four years. However, this fact becomes masked when considering these individuals as members of a death cohort.

Fig 1a shows the distributions of age at death grouped by year of birth for individuals from a study of 577 Irish Wolfhound dogs that died between 1966 and 1986 (Bernardi 1986). The original data were made available for re-analysis during the present study.

Statistical analysis of the data provided by Bernardi (1986) was carried out using linear regression by the general linear model procedure in the SAS System, version 8.02, program package. Fig 1 was created using the BOXPLOT procedure from the same package. Differences in measured life span by birth cohort due to right censored data are presented. Measured age at death is significantly different between birth cohorts ($P<0{\cdot}0001$, $R^2=0{\cdot}25$, $F=63{\cdot}25$, type III sum of squares [SS]). However, as Fig 1b demonstrates, when year of death is used as the grouping criterion, this bias becomes masked ($P=0{\cdot}09$, $R^2=0{\cdot}008$, $F=2{\cdot}41$, type III SS).

Given that very few Irish wolfhounds die aged over 12 years (Urfer and others 2007), it is possible to remove right censored data from this sample by simply removing all dogs that were born less than 12 years before the end of the data collection period. A comparison of statistical parameters before and after this removal is provided in Table 1.

During work on a review paper on life span and causes of death in Irish Wolfhounds (Urfer and others 2007), a number of articles affected by this bias in one way or another were identified in the veterinary and biological literature (Bronson 1981, 1982, Bernardi 1986, Hayashidani and others 1988, 1989, Deeb and Wolf 1994, Eichelberg and Seine 1996, Li and others 1996, Patronek and others 1997, Michell 1999, Proschowsky and others 2003, Casal and others 2006, Galis and others 2006). Their common denominator is that life span and survival were calculated purely from death data and grouped by death cohorts, without giving due concern to the problem of right censored data.

While several possibilities for bias in veterinary life span research have already been mentioned in the literature (Reid and Peterson 2000), the problem of right censored death data has not been addressed. Life span and survival calculations based on death cohorts are widespread throughout the scientific literature. In species with long life expectancy (such as humans), this problem is particularly relevant, but its effects will be encountered whenever a significant percentage of individuals in a study come from birth cohorts of which members are still alive. In human medicine, statistical methods to correct right-censored data are commonly employed (Kalbfleisch and Prentice 2002).

In order for cohort bias due to right censored data to be eliminated from life span and survival data, populations should be analysed in cohorts grouped by their year of birth. To arrive at representative life span estimates, it is necessary to apply appropriate statistical methods to data from birth cohorts of which an important number of individuals are still alive at the time of data collection.

The data used in veterinary life span studies can be subject to some limitations – for example, grouping in very large age categories, which occurs in the Veterinary Medical Database, or lack of information on the maximum life span in the studied population. Nevertheless, the problem of right-censored data should be considered whenever a sample of life span data is analysed in veterinary science.

Using Kaplan-Meier estimation is one possible solution to the problem. In some cases, right-censored data can also be corrected sufficiently well by eliminating individuals from the data that come from birth cohorts where a significant number of individuals can be assumed to still be alive. This approach is not mathematically complex; however, depending on the structure of the studied data, it can potentially lead to the loss of an important part of the database and thus diminish the overall power of the analysis.

Given that the domestic dog is a useful model organism for life span and aging research (Austad 1993, Patronek and others 1997, Greer and others 2007), the bias caused by right-censored data deserves more consideration in veterinary and biological life span and survival studies.

ACKNOWLEDGEMENTS

The author would like to thank Sabine and Martin Gebhardt for their suggestions and critique, and Mrs Gretchen Bernardi for supplying the original data of her life span study. The work as a result of which this paper was written was partially funded by the Veterinary Competence Centre of the Swiss Army.

References

AUSTAD, S. N. (1993) FRAR course on laboratory approaches to aging. The comparative perspective and choice of animal models in aging research. *Aging (Milano)* **5,** 259-267

BERNARDI, G. (1986) Longevity and morbidity in the Irish Wolfhound in the United States – 1966 to 1986. *AKC Gazette* **105,** 70-78

BRONSON, R. T. (1981) Age at death of necropsied intact and neutered cats. *American Journal of Veterinary Research* **42,** 1606-1608

BRONSON, R. T. (1982) Variation in age at death of dogs of different sexes and breeds. *American Journal of Veterinary Research* **43,** 2057-2059

CASAL, M. L., MUNUVE, R. M., JANIS, M. A., WERNER, P. & HENTHORN, P. S. (2006) Epilepsy in Irish Wolfhounds. *Journal of Veterinary Internal Medicine* **20,** 131-135

DEEB, B. & WOLF, N. (1994) Studying longevity and morbidity in giant and small breeds of dogs. *Veterinary Medicine* **89,** 702-713

EICHELBERG, H. & SEINE, R. (1996) Life expectancy and cause of death in dogs. I. The situation in mixed breeds and various dog breeds. *Berliner und Münchener Tierärztliche Wochenschrift* **109,** 292-303 (In German)

GALIS, F., VAN DER SLUIJS, I., VAN DOOREN, T. J., METZ, J. A. & NUSSBAUMER, M. (2006) Do large dogs die young? *Journal of Experimental Zoology. B* **308,** 119-126

GREER, K. A., CANTERBERRY, S. C. & MURPHY, K. E. (2007) Statistical analysis regarding the effects of height and weight on life span of the domestic dog. *Research in Veterinary Science* **82,** 208-214

HAYASHIDANI, H., OMI, Y., OGAWA, M. & FUKUTOMI, K. (1988) Epidemiological studies on the expectation of life for dogs computed from animal cemetery records. *Nippon Juigaku Zasshi* **50,** 1003-1008

HAYASHIDANI, H., OMI, Y., OGAWA, M. & FUKUTOMI, K. (1989) Epidemiological studies on the expectation of life for cats computed from animal cemetery records. *Nippon Juigaku Zasshi* **51,** 905-908

KALBFLEISCH, J. D. & PRENTICE, R. L. (2002) The Statistical Analysis of Failure Time Data. 2nd edn. Hoboken, Wiley-Interscience

LI, Y., DEEB, B., PENDERGRASS, W. & WOLF, N. (1996) Cellular proliferative capacity and life span in small and large dogs. *Journal of Gerontology. A* **51,** B403-408

MICHELL, A. R. (1999) Longevity of British breeds of dog and its relationships with sex, size, cardiovascular variables and disease. *Veterinary Record* **145,** 625-629

PATRONEK, G. J., WATERS, D. J. & GLICKMAN, L. T. (1997) Comparative longevity of pet dogs and humans: implications for gerontology research. *Journal of Gerontology.* A **52,** B171-178

PROSCHOWSKY, H. F., RUGBJERG, H. & ERSBØLL, A. K. (2003) Mortality of purebred and mixed-breed dogs in Denmark. *Preventive Veterinary Medicine* **58,** 63-74

REID, S. W. & PETERSON, M. M. (2000) Methods of estimating canine longevity. *Veterinary Record* **147,** 630-631

URFER, S., GAILLARD, C. & STEIGER, A. (2007) Lifespan and disease predispositions in the Irish wolfhound: a review. *Veterinary Quarterly* **29,** 102-111