Scottish Government
Riaghaltas na h-Alba
gov.scot

Potential Utility of Survival Analysis to Evaluate Mortality Arising from Multiple Infections in Aquaculture Production Cycles

Scottish Marine and Freshwater Science Vol 12 No 9

L M Hall

marinescotland

Potential Utility of Survival Analysis to Evaluate Mortality Arising from Multiple Infections in Aquaculture Production Cycles

Scottish Marine and Freshwater Science Vol 12 No 9

L M Hall

Published by Marine Scotland Science

ISSN: 2043-7722
DOI: 10.7489/12379-1

Marine Scotland is the directorate of the Scottish Government responsible for the integrated management of Scotland's seas. Marine Scotland Science (formerly Fisheries Research Services) provides expert scientific and technical advice on marine and fisheries issues. Scottish Marine and Freshwater Science is a series of reports that publishes results of research and monitoring carried out by Marine Scotland Science. It also publishes the results of marine and freshwater scientific work that has been carried out for Marine Scotland under external commission. These reports are not subject to formal external peer-review.

This report presents the results of marine and freshwater scientific work carried out by Marine Scotland Science.

Potential Utility of Survival Analysis to Evaluate Mortality Arising from Multiple Infections in Aquaculture Production Cycles

L M Hall

Marine Scotland Science, 375 Victoria Road
Aberdeen, AB11 9DB

Executive summary

Simulations indicate that the Cox proportional hazards model can be used to estimate approximate animal-level multiplicative mortality effects arising from multiple infections within aquaculture production cycles.

Introduction

Commercial aquaculture farm production data have long been used to investigate the causes of stock mortality (Menzies *et al*., 1996) with analyses of disease focussing on individual presumptive causes (e.g. Menzies *et al*., 1996; Soares *et al*., 2011; Kilburn *et al*., 2012). Substantial differences in mortality rates attributed to a specific disease have been reported (e.g. Menzies *et al*., 1996; Kilburn *et al*., 2012; Escobar-Dodero *et al*., 2019) and an interaction with other infections may explain a part of this variation (reviewed by Kotob *et al*., 2016).

Aquarium-based disease challenge experiments can provide information on the potential effect of multiple infections on mortality (e.g. Hedrick *et al*., 1994; Johansen & Sommer, 2001), although the estimates of magnitude are not necessarily applicable to commercial production conditions. A practicable approach to estimating the size of interactive effects between infections on mortality using farm production data is therefore desirable.

The Cox proportional hazards model (Cox, 1972) represents one possible approach. Specifically the model can estimate a hazard ratio (HR) for mortality attributed to a specified disease when another preceding or concurrent specified infection occurs in the same production cycle, relative to mortality attributed to the same specified disease when the preceding or concurrent infection does not occur in a cycle. This approach was recently used to identify infection risk factors associated with mortality attributed to infectious pancreatic necrosis (IPN) within Atlantic salmon, *Salmo salar* L., production cycles (Escobar-Dodero *et al*., 2019). The investigation, however,

focussed on estimating the HR for the occurrence of one or more IPN mortalities during a production cycle, rather than estimating a HR for all IPN mortalities; a production cycle rather than animal-level estimate. No aquaculture based animal-level analyses using a proportional hazards model have been published and it would be reassuring to ensure that such analyses are computationally feasible and valid before embarking on such investigations. This present report therefore describes *in-silico* simulations generating animal-level disease mortality HR estimates intended to help build the capacity within Marine Scotland Science of carrying out analyses of mortality associated with multiple infections.

Methods

Data, similar in structure to the marine-stage production database of one of Scotland's Atlantic salmon aquaculture companies (Kilburn *et al*., 2012), were simulated *in-silico*.

The simulated data (Figure 1) comprised five variables (columns) for:

- identity of production cycle ('production-cycle'; categorical comprising '1', '2', '3', …, '100');
- production day comprising at least one record (row) for each production cycle day ('day'; numeric comprising 1, 2, 3, …, 550 with stocking at the beginning of day 1 and harvest at the end of day 550);
- number of livestock at the start of each production cycle day ('opening-count'; numeric $\leq 1 \times 10^6$);
- at least one record of the number of deaths on each production cycle day ('mortality'; numeric $< 1 \times 10^6$);
- cause of death for each mortality record ('cause', categorical comprising 'production-loss', 'disease-1', 'disease-2' and 'harvest').

The simulated production cycles were categorised into four type groups based on whether they had experienced:

- production-loss and harvest only (TG0);
- production-loss, disease-1, and harvest (TG1);
- production-loss, disease-2, and harvest (TG2);
- production-loss, disease-1, disease-2, and harvest (TG12).

The estimation of the HR only requires TG2 and TG12 production cycles.

Production-cycle	Day	Opening count	Mortality	Cause
1	1	1000000	5572	production-loss
1	2	994428	5413	production-loss
1			...	
1	251	911815	91	production-loss
1	251	911815	54	disease-1
1	252	911670	90	production-loss
1	252	911670	130	disease-1
1			…	
1	354	807064	88	production-loss
1	354	807064	253	disease-2
1	355	806723	93	production-loss
1	355	806723	753	disease-2
1			...	
1	550	732923	732923	harvest
2	1	1000000	4753	production-loss
2			...	
2	550	638428	638428	harvest
		…		
100	1	1000000	6032	production-loss
100			….	
100	550	925248	639025	harvest

Figure 1: Example of simulated production data. Production-cycle 1 experiences both disease-1 & disease-2 mortality and is categorised as type group TG12.

An example of the simulated survivals is presented in Figure 2. This involved 100 production cycles, each stocked with $1x10^6$ fish with surviving fish harvested on day 550. Mortalities attributed to disease-2 affected 2.5-7.5% of the stock for TG2 production cycles and 10-30% for TG12 cycles, a multiplicative effect of four. Additional details of the production data simulation are given in the Appendix.

The production database was converted into a survival database (Figure 3), comprising five variables (columns) for:

- production-cycle (described previously);
- day (described previously);
- type-group (described previously) comprising TG2 and TG12 only;
- indicator categorising mortality attributed to disease-2 as ‘1’ and all other causes as ‘0’ (‘indicator’; categorical: ‘1’, ‘0’);
- number of deaths (‘mortality’; numeric $<1x10^6$).

Mortalities for each indicator variable within each production-cycle day combination are summed.

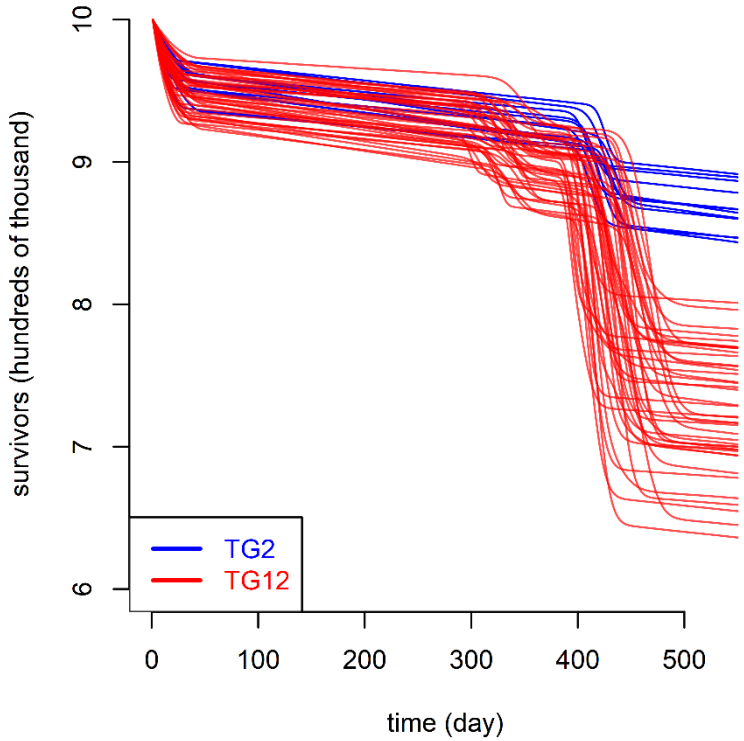

Figure 2: Example of simulated survival for production cycles (individual lines) experiencing mortality attributed to disease-2 only (TG2) or disease-1 & disease-2 (TG12) assuming an interactive effect of disease-1 on disease-2 mortality of four. Mortality attributed to disease-1 precedes disease-2 in TG12 cycles. Harvest on day 550 is not plotted.

The HR was estimated using the Cox proportional hazards model:

$$\frac{h(t)}{h_0(t)} = e^{\beta X}$$

where: $h(t)$ = probability of disease-2 mortality at time *t* for type group TG12 given that it had not occurred previously; $h_0(t)$ = probability of disease-2 mortality at time *t* for type group TG2 given that it had not occurred previously; β = fixed effect coefficient for type group (X) TG12 relative to TG2. The method of Efron (1977) was used to correct for multiple daily mortalities. Individual fish within production cycles share a more similar experience than individuals between cycles and a robust variance estimate (Lin & Wei, 1989) was used to correct 95% confidence intervals for this model misspecification. Finally operationalising the model by replacing the dependent variable of mortality number with the indicator (1 or 0) and weighting this by mortality number is computationally more efficient.

Multiple sets of simulated data with defined HR values ranging from 1 to 4, were generated within the R Statistical Environment version 3.6.3 (R Core Team, 2020) and the estimated HR calculated using the 'coxph' function of 'survival' 3.1-11 (Therneau, 2020). Hazard ratios of greater than 1 represent an increased risk of

disease-2 mortality if disease-1 is present in a production cycle relative to cycles where disease-1 is not present. The relationship between the estimated and defined HR was evaluated by linear regression assuming a normal error distribution, and the parameter values compared to the expected intercept of 0 and gradient of 1.

Production-cycle	Day	Type-group	Indicator	Mortality
7	354	TG2	0	86
7	...	TG2	0	...
7	384	TG2	0	77
7	384	TG2	1	12
7	385	TG2	0	85
7	386	TG2	0	73
7	386	TG2	1	54
7	…	TG2	…	
7	550	TG2	0	832576
11	354	TG12	0	85
11	...	TG12	0	...
11	383	TG12	0	76
11	383	TG12	1	3
11	384	TG12	0	78
11	385	TG12	0	80
11	385	TG12	1	212
11	386	TG12	0	73
11	386	TG12	1	319
11	…	TG12	…	
11	550	TG12	0	639025
		...		

Figure 3: Example of a simulated survival database comprising production cycles experiencing disease-2 (TG2),and disease-1 & disease-2 (TG12) only. The indicator codes disease-2 deaths as '1' and all others, including harvest, as '0'.

Results

There is agreement between estimated and defined HR (Figure 4) with no evidence that the linear regression equation intercept (-0.04±0.08) or gradient (1.00±0.03) differ from the expected values of 0 and 1 respectively.

The estimation of the HR from a single simulated database took less than one second using a 64 bit laptop with a 2.4 GHz Intel Core i5-6200U CPU and 8 Gb of RAM running under Windows 10 Enterprise.

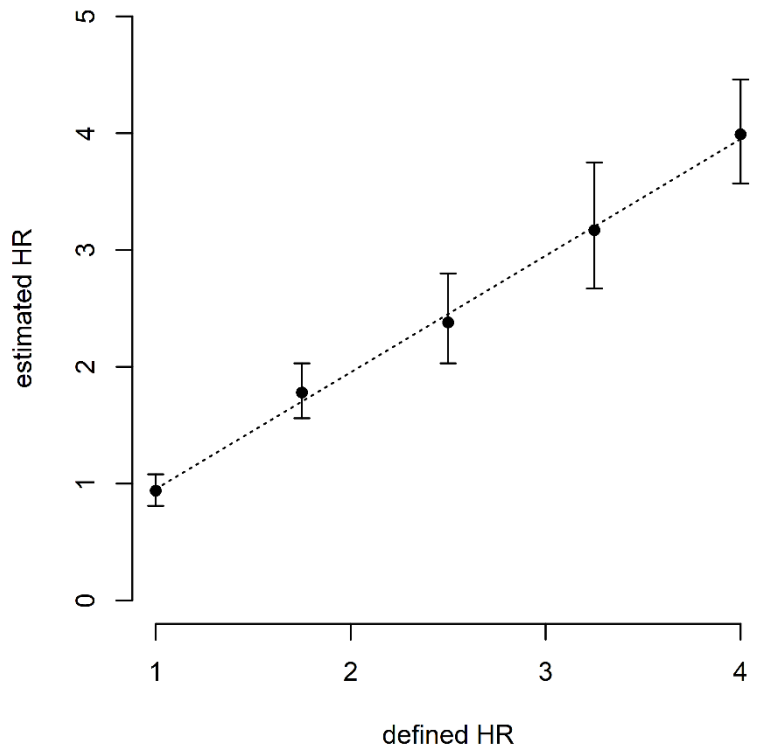

Figure 4: Relationship of estimated and defined hazard ratios (HR). Estimates are shown as points with 95% confidence intervals and the relationship as a dotted line.

Discussion

The results of these simulations indicate that the Cox proportional hazards model can be used to estimate interactive animal-level disease mortality effects within aquaculture production cycles. This approach is not necessarily limited to investigations of disease mortality because disease-1 and/or disease-2 can be replaced by environmental events or production management variables.

The simulation results do not, however, demonstrate the practical utility of the Cox proportional hazard model. There are several reasons for this. First, actual production data are not as conditioned as simulated data. For example additional stocking could disrupt analyses depending on when it occurs. In contrast stock movements from a production cycle can be accommodated by coding the indicator variable as '0' (Figure 3). Second, the simulations assumed that the timing and duration of disease-2 mortalities are similar for both TG2 and TG12 production cycles. Differences would violate a modelling assumption of constant HR over time. Model modifications incorporating change over time are available, the simplest being to estimate the HR for stratified time periods (e.g. Zhang *et al*., 2018). Compliance of actual production data with other modelling assumptions such as independent censoring would also require investigation. Third, actual production cycles are not

independent as assumed by the simulations and share similarities arising from geographical location, operator and disease occurrence on nearby sites. A robust variance was used for the simulations to correct the confidence intervals for model misspecification, and it is very likely that this approach would be applicable to actual production data. The point estimates are however based on production cycle type group only and are therefore approximate. An investigation to confirm that the approximate HR are satisfactory using a subset of actual production data is desirable. In summary, while the simulations indicate the utility of the proportional hazard model to aquaculture production cycle data, it will be necessary to pilot its use on actual animal-level production data to demonstrate this.

There are no published animal-level proportional hazard analyses associated with disease in aquaculture. It is likely that a perceived problem of computational resource is, at least in part, responsible for deficiency. The simple approach described in this report, using an ordinary laptop to rapidly generate HR estimates, indicates that the Cox proportional hazards model can be used to estimate approximate animal-level multiplicative mortality effects arising from multiple infections within aquaculture production cycles.

References

Cox, D. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society* B 34, 187-220.

Efron, B. 1977. The Efficiency of Cox's Likelihood Function for Censored Data. *Journal of the American Statistical Association* 72, 557-565.

Escobar-Dodero, J., Kinsley, A., Perez, A.M., Ibarra, R., Tello, A., Monti, G., Mardones, F.O. 2019. Risk factors for infectious pancreatic necrosis in farmed Chilean Atlantic salmon (*Salmo salar* L.) from 2010 to 2013. *Preventive Veterinary Medicine* 167, 182-189.

Hedrick, R.P., LaPatra, S.E., Yun, S., Lauda, K.A., Jones, G.R., Congleton, J.L. & de Kinkelin, P. 1994. Induction of protection from infectious heatopoietic necrosis virus in rainbow trout by pre-exposure to the avirulent cutthroat trout virus (CTV). *Diseases of Aquatic Organisms* 20, 111-118.

Johansen, L. & Sommer, A. 2001. Infectious pancreatic necrosis virus infection in Atlantic salmon *Salmo salar* post-smolts affects the outcome of secondary infections

with infectious salmon anaemia virus or *Vibrio salmonicida*. *Diseases of Aquatic Organisms* 47, 109-117.

Kilburn, R., Murray, A.G., Hall, M. Bruno, D.W., Cockerill, D. & Raynard, R.S. 2012. Analysis of a company's production data to describe the epidemiology and persistence of pancreas disease in Atlantic salmon (*Salmo salar* L.) farms off Western Scotland. *Aquaculture* 368-369, 89-94.

Kotob, M.H., Menanteau-Ledouble, S., Kumar, G., Abdelzaher, M., & El-Matbouli, M. 2016. The impact of co-infections on fish: a review. *Veterinary Research* 47, 98.

Lin, D.Y. & Wei, L.J. 1989. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* 84, 1074-1078.

Menzies, F.D., McLoughlin, M.F., Wheatley, S.B. & Goodall, E.A. 1996. Development of a computerized information retrieval system for Atlantic salmon, *Salmo salar* L., production. *Aquaculture Research* 27, 183-190.
R Core Team 2020. *R: A Language and Environment for Statistical Computing*. Vienna, R Foundation for Statistical Computing. https://www.R-project.org/.

Soares, S., Green, D.M., Turnbull, J.F., Crumlish, M. & Murray, A.G., 2011. A baseline method for benchmarking mortality losses in Atlantic salmon (*Salmo salar*) production. *Aquaculture* 314, 7-12.

Therneau T (2020). *A Package for Survival Analysis in R*. R package version 3.1-11. https://CRAN.R-project.org/package=survival.

Zhang, Z., Reinikainen, J., Adeleke, K.A., Pieterse, M.E., Groothuis-Oudshoorn, C.G.M., 2018. Time-varying covariates and coefficients in Cox regression models. *Annals of Translational Medicine* 6(7), 121.

Appendix – Simulation of Production Data

The simulated production data comprised 100 production cycles each stocked with $1x10^6$ fish at the beginning of day one. No further stocking occurred after this time point. Each production cycle then experienced continuous mortality until the remaining fish were harvested at the end of day 550. No harvesting or fish movements from production cycles occurred before day 550. Mortalities between stocking and harvest were attributed to production losses or disease.

The production mortality for each individual fish was:

- a probability of daily mortality sampled individually for each production cycle from ~U($4.54x10^{-5}$,$1.36x10^{-4}$) and constant for that cycle from day 1 to day 550 inclusive and;
- a probability of daily mortality sampled individually for each production cycle from ~U($5.00x10^{-4}$,$3.75x10^{-3}$), and constant for that cycle from stocking to a day sampled individually for each cycle from ~U(20,50).

Disease mortality attributed to disease-1 or disease-2 was assigned to individual fish. Some production cycles experienced no disease (TG0), other cycles disease-1 only (TG1), other cycles disease-2 only (TG2), and the remaining cycles disease-1 and disease-2 (TG12). The number of production cycles experiencing each disease comprised:

- TG1: Production cycles affected by disease-1 were randomly assigned with a probability of 0.100. The total disease-1 mortality was sampled from ~U($1.25x10^{-2}$, $3.75x10^{-2}$) with a start day sampled from ~U(275, 325) and a duration sampled from ~U(30, 100) for individual production cycles.
- TG2: Previously unassigned production cycles affected by disease-2 were assigned with a probability of 0.222. The total disease-2 mortality was sampled from ~U($2.50x10^{-2}$, $7.50x10^{-2}$) with a start day sampled from ~U(375, 425) and a duration sampled from ~U(30, 100) for individual production cycles.
- TG12: Previously unassigned production cycles affected by both disease-1 and disease-2 were assigned with a probability of 0.571. The total disease-1 mortality for individual production cycles was sampled as described for TG1. The total disease-2 mortality for individual production cycles was sampled as described for TG2 except the mortality rate was multiplied by a factor ranging from >1 to 4, corresponding to an interactive mortality effect of disease-1 on disease-2.

The remaining production cycles, which did not experience disease mortality, were assigned to TG0.

The number of deaths on each production cycle was determined stochastically sequentially for each day from the number of survivors at the start of that day using the cycle specific mortality values. All surviving fish were harvested at the end of day 550.

The resulting production data (Figure 1 and 2) were then converted into survival data (Figure 3) and subjected to the Cox proportional hazards model analysis described on pages 4-5.