Original article

Comparison of methods for analysis of selective genotyping survival data

Joseph P. McElroy[a], Wuyan Zhang[b], Kenneth J. Koehler[b],
Susan J. Lamont[a], Jack C.M. Dekkers[a*]

[a] Department of Animal Science and Center for Integrated Animal Genomics, Iowa State
University, Ames, USA
[b] Department of Statistics, Iowa State University, Ames, USA

Abstract – Survival traits and selective genotyping datasets are typically not normally distributed, thus common models used to identify QTL may not be statistically appropriate for their analysis. The objective of the present study was to compare models for identification of QTL associated with survival traits, in particular when combined with selective genotyping. Data were simulated to model the survival distribution of a population of chickens challenged with Marek disease virus. Cox proportional hazards (CPH), linear regression (LR), and Weibull models were compared for their appropriateness to analyze the data, ability to identify associations of marker alleles with survival, and estimation of effects when all individuals were genotyped (full genotyping) and when selective genotyping was used. Little difference in power was found between the CPH and the LR model for low censoring cases for both full and selective genotyping. The simulated data were not transformed to follow a Weibull distribution and, as a result, the Weibull model generally resulted in less power than the other two models and overestimated effects. Effect estimates from LR and CPH were unbiased when all individuals were genotyped, but overestimated when selective genotyping was used. Thus, LR is preferred for analyzing survival data when the amount of censoring is low because of ease of implementation and interpretation. Including phenotypic data of non-genotyped individuals in selective genotyping analysis increased power, but resulted in LR having an inflated false positive rate, and therefore the CPH model is preferred for this scenario, although transformation of the data may also make the Weibull model appropriate for this case. The results from the research presented herein are directly applicable to interval mapping analyses.

survival / Cox proportional hazards / Weibull / quantitative trait loci

1. INTRODUCTION

Genetic association analyses are becoming a common approach in animal breeding to identify genes or genomic regions that affect quantitative traits

---

(*e.g.*, [17, 29, 33]). Most analyses utilize statistical models that assume normality of phenotypes [8, 9, 12, 14, 32]. Many phenotypic traits of interest in agriculture, however, do not follow a normal distribution (*e.g.* ordinal traits such as conformation scores, binary traits such as calving/not calving, or time to success/failure traits such as survival time).

Many agriculturally important traits follow a survival distribution (*e.g.* survival after infection, length of productive life, or days open). To analyze such traits, the Weibull and Cox proportional hazards models are commonly employed [1, 2, 4, 6, 7, 11, 16, 23, 25, 27]. The Weibull model, a generalization of the exponential model, is parametric and is therefore appropriate for only specific distributions; the Cox model is a rank-based semi-parametric method and, therefore, should be appropriate for all distributions for which the hazards between groups are proportional [26]. Both models can appropriately handle datasets that include data from individuals without a recorded time of death, *i.e.*, censored data.

Another distributional difficulty that is often encountered in marker/phenotype association analyses is created by the use of selective genotyping. Selective genotyping is a method to reduce the costs of an experiment by genotyping only individuals from the extremes of the phenotypic distribution [14, 15]. Individuals from the phenotypic extremes provide most of the power for a marker-trait analysis, so power can be maximized for a limited number of genotypes by selective genotyping. This, however, also causes the distribution of phenotypes to be non-normal. When the phenotype of all individuals in the experiment is normally distributed, using linear regression to analyze selective genotyping data results in the effects being overestimated [14]. Darvasi and Soller [3] and Ronin *et al.* [24] derived an approximation to correct for this bias when the complete phenotypic data follows a normal distribution. Maximum likelihood can also be used to appropriately analyze such data [14, 20], if the proper distribution can be identified. Henshall and Goddard [10] showed that logistic regression of genotype onto phenotype also resulted in unbiased estimates of effects; however, the original phenotype must be normally distributed for this method as well.

Moreno *et al.* [19] used simulation to compare Cox, Weibull, and Gaussian models for QTL analysis with a survival phenotype. They found that when the data are censored at a fixed date, the survival models improved power of QTL detection and the accuracy of estimates of QTL location and effect. They did not, however, evaluate the models for analysis of selective genotyping survival data, which was the objective of the current study.

McElroy *et al.* [18] used the Cox proportional hazards model and linear regression to identify markers associated with survival in selectively genotyped layer chickens infected with Marek disease virus. Using simulation under the null hypothesis of no QTL, they found that both linear regression and the Cox proportional hazards model resulted in valid false positive rates for tests of association. They did not, however, compare the power of these two models, nor did they evaluate the use of a Weibull model. Because the Weibull model is a parametric survival model, it should be more powerful that the Cox model if the data follow a Weibull distribution. Therefore, the objective of this study was to compare the validity of false positive rates and the power of employing the Weibull model, the Cox proportional hazards model, and the linear regression model to analyze QTL associations under full and selective genotyping of survival data.

2. MATERIALS AND METHODS

2.1. Data simulation

Survival data were simulated for a backcross under the null hypothesis of no QTL, or the alternative hypothesis that a QTL affecting survival resides at the marker under analysis. To mimic real data, the simulated data were generated to represent actual data from the backcross population of Marek disease virus-challenged layer chickens described by McElroy *et al.* [18]. Survival times in the real experimental population over the recording time ranged from 33 to 140 d (Fig. 1). Survival times followed a right-skewed distribution with a mean of 65.5 d, a median of 59.0 d, and a standard deviation of 23.9 d. Twenty-eight individuals (4.3%) survived to the end of the study and were considered censored in the analyses. A survival function ($S(t)_{KM}$) was estimated from the real data using the non-parametric Kaplan-Meier estimator [13]:

$$\hat{S}(t)_{KM} = \prod_{j=1}^{k} \frac{n_j - d_j}{n_j}$$

for $t_{(k)} \leq t < t_{(k+1)}$, where j is the rank of a particular day (t) among all chronologically ordered days in which death occurred, k is the rank of the day in which the survival function is being evaluated, n_j is the number of animals still alive on day j, and d_j is the number of animals dying on day j. This resulted in an estimate of $S(t)_{KM}$ for each time of death (t).

Survival data for 700 backcross animals were simulated following methods described by Moreno *et al.* [19] by generating a survival probability, $S_0(t)_i$, for
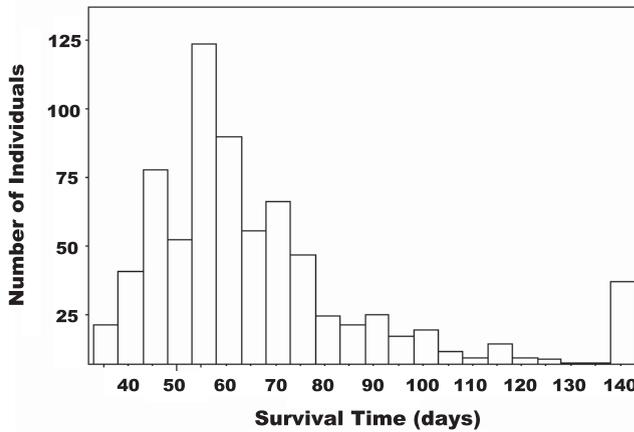
Figure 1. Distribution of survival times from the real population of chickens (data reported in McElroy *et al.* [18]). Recording terminated at 140 days.

each individual *i* based on a proportional hazards model:

$$S(t)_i = S_0(t)_i^{\exp(\beta_s x_i)}$$

where $S(t)_i$ is a draw from a random uniform [0,1], β_s is the simulated QTL allele substitution effect [5], and x_i indicates the QTL allele received from the F1 parent, which was either zero or one and was randomly selected from a Bernoulli distribution with a 50% chance of getting either allele. The value of $S_0(t)_i$ obtained for an animal was then cross-referenced with the estimated Kaplan-Meier function obtained from the real data to get the corresponding time of death. To simulate selective genotyping, only the 20% short (140 animals) and 20% long (140 animals) surviving individuals were considered for analysis and their genotype for the marker (QTL) was assumed known.

The simulation was performed for each of four allele substitution effect levels of the QTL: $\beta_s = 0, 0.1, 0.2,$ or 0.3. The largest effect was chosen to be 0.3 because all analyses had high power for this effect. It should be noted that a larger β_s results in an increased risk of death and, therefore, corresponds to a more negative effect on days of survival. Two additional censoring scenarios (0 and 20%) were also considered, in addition to the 4.3% censoring that was present in the real data. For the no censoring scenario, the Kaplan-Meier function from the real data set was extended past day 140 by assuming the same risk of dying for each day past 140 as the average risk from day 101 and 139. For 20% censoring, day 77 was considered as the last day of the study.

2.2. Models of analysis

Simulated data were analyzed for associations of marker and phenotype using the Weibull, Cox proportional hazards, and linear regression models. The linear regression model (*e.g.*, [28]) was:

$$T_i = \mu_i + \beta_{LR}p(Q)_i + \varepsilon_i,$$

where T_i is the survival time of animal i, in days; β_{LR} is the increase in the mean survival time associated with inheriting allele 1 *versus* allele 0; $p(Q)_i$ is the probability of inheriting one *versus* the other QTL allele for animal i; μ_i is the average survival time when $p(Q)_i = 0$; and ε_i is the residual for animal i.

The Cox model used, following Smith [25], was:

$$S(t;n)_i = [S_0(t)]^{\exp(\eta_i)},$$

where $S(t;n)_i$ is the probability that animal i survived at least until time t, $S_0(t)$ is the baseline survival, $S_0(t) = e^{-\int h_o(t)dt}$, where $h_o(t)$ is the baseline hazard, $\eta_i = \beta_{PH}p(Q)_i$, where β_{PH} is the allele substitution effect of the QTL on the natural log of the hazard ratio for the Cox model, and $p(Q)_i$ is as defined previously.

The Weibull model [30] used was:

$$S(t)_i = \exp\left\{-\theta t^\alpha e^{[\beta_W p(Q)_i]}\right\},$$

where $S(t)_i$ is the probability that animal i survived at least until time t, θ is the shape parameter of the Weibull distribution, α is the scale parameter of the Weibull distribution, β_W is the allele substitution effect of the QTL on the natural log of the hazard ratio for the Weibull model, and $p(Q)_i$ is as defined previously. For all models, $p(Q)_i = 0$ or 1 for genotyped individuals, depending on which QTL allele they inherited.

Selective genotyping data were analyzed in two ways: (i) including only genotypic and phenotypic data from genotyped individuals (SG) in the analysis, and (ii) also including phenotypic data from the ungenotyped individuals (SGI) in the analysis. For the SGI scenario, $p(Q)_i = 0.5$ for ungenotyped individuals because, with unknown genotypes, these individuals were equally likely to have received either allele. These data were then analyzed with the three models of interest. For the 4.3 and 20% censoring scenarios, animals dying on day 140 and 77, respectively, were considered censored for the Weibull and Cox models, and to have died on day 140 and 77 for the linear regression model.

The proper method for including ungenotyped individuals in the survival models, in which the allele inherited is unknown, is to use the average of the hazards for the two possible alleles in the partial likelihood. However, standard statistical packages do not permit modification of the likelihood equations. To allow implementation of standard methods, $p(Q)_i = 0.5$ was used as the explanatory variable for the survival models as an approximation of the average of the hazards for the two possible alleles in the partial likelihood. This method is discussed in detail in the appendix of McElroy *et al.* [18] for the Cox model. They found that, under the null hypothesis of no QTL effect, the estimated effects are slightly biased away from zero when the approximation is used, but the standard error of the estimated effects is also higher, resulting in an appropriate type I error rate.

False positive rates (assuming no QTL effect) and power of the analyses were calculated based on 1000 replicates for each allelic effect level for each of two significance levels ($P < 0.05$ and 0.01) for the Full (all individuals genotyped), SG, and SGI genotyping scenarios. A two-tailed binomial test was used to determine if false positive rates differed significantly from expected (*i.e.* the α level) under the null hypothesis and a two-tailed Fisher exact test was used to determine if the power of the alternate models differed significantly.

The Weibull model for all genotyping scenarios and the linear regression model for the SGI scenario resulted in inflated false positive rates. Although transforming the data to fit a Weibull distribution could have made the Weibull model appropriate for analysis in the Full genotyping scenario and may have improved the performance of the Weibull model in the SGI scenario, no transformations of the data were performed for the current study because transformation would not have addressed the SG scenario, which was the main emphasis of the study. Instead, for the Weibull analyses and the linear regression analyses in the SGI scenario, empirical thresholds with appropriate false positive levels were derived and used to compute power for each level of censoring. Empirical thresholds were obtained from 10 000 replicates of data simulated with no QTL effect.

Average allele substitution effect estimates and their variances across each set of 1000 replicates were also calculated. Estimated effects from linear regression describe the average difference of survival in days between the two QTL alleles, and estimates from the Cox and Weibull models describe the allele substitution effect on the natural log of the hazard ratio. To make the effect

estimates from all three models directly comparable, the mean difference of survival in days between groups inheriting one *versus* the other QTL allele was estimated for the Cox and Weibull models for each replicate. For the Weibull model, estimates in terms of the hazard ratio can be directly converted into estimates in days by:

$$E(T) = \Gamma(1/\alpha + 1)(\theta e^{p(Q)_i \beta_W})^{-1/\alpha}$$

where E(T) is the expected mean days of survival, Γ is the gamma function, and $p(Q)_i$, α, β_W, and θ are as described previously. The mean difference in days was calculated by finding the difference between E(T) for $p(Q)_i = 0$ and 1. To obtain estimates in days from the Cox proportional hazards model, the difference in the means of the survival functions for $p(Q)_i = 0$ and 1 was calculated across all times using the Cox proportional hazards estimates for β_{PH}. For this estimation, censored individuals were assumed to have died on the last day of the study. Estimates from the three models were compared based on their correlations, magnitudes, and standard errors. The magnitudes of effect estimates from the three models were compared by using a t-test assuming unequal variances. All simulations and analyses were performed using the R [22] statistical package.

3. RESULTS

3.1. False positive rates

Table I shows the false positive rates for the Weibull, Cox, and linear regression models for $\alpha = 0.05$. False positive rates for the Cox model were not significantly ($P > 0.05$) inflated for any genotyping scenario. False positive rates for linear regression were not significantly ($P > 0.05$) inflated in the full and SG cases, but were significantly ($P < 0.05$) inflated for every censoring level when SGI was used. The Weibull model had a significantly ($P < 0.05$) higher number of false positives than expected for every censoring and genotyping scenario. Significance of differences from the expectations for the three models with $\alpha = 0.01$ were similar to those when $\alpha = 0.05$.

3.2. Power between models

Powers of the Cox, Weibull, and linear regression models for different effects and censoring for $\alpha = 0.05$ are shown in Figure 2. Although not always

Table I. False positive rates for Cox proportional hazards, linear regression, and Weibull models for different censoring levels and genotypic scenarios at $\alpha = 0.05$ based on 1000 replicates.

Censoring (%)	Genotyping[1]	False Positive Rate[2]		
		CPH[3] (%)	LR[3] (%)	WEI[3] (%)
0	Full	4.7	4.3	16.9*
4	Full	4.6	4.9	13.9*
20	Full	5.4	5.8	7.9*
0	SG	4.8	4.9	6.8*
4	SG	4.8	5.0	8.1*
20	SG	4.5	5.0	6.8*
0	SGI	4.5	20.7*	26.3*
4	SGI	4.7	18.4*	23.4*
20	SGI	5.0	15.7*	16.3*

[1] Full = all individuals genotyped and used in analysis; SG = selective genotyping with ungenotyped individuals excluded from the analysis; SGI = selective genotyping with ungenotyped individuals included in the analysis.
[2] * indicates significant difference ($P < 0.05$) between observed and expected false positive rates (0.05) based on a two-tailed binomial test.
[3] CPH = Cox proportional hazards; LR = linear regression; WEI = Weibull.

significant, for the Full case ($\alpha = 0.05$), the Cox model was more powerful than both linear regression and the Weibull model, and the linear regression model was more powerful than the Weibull model, except with 20% censoring. The linear regression model had lower power when censoring was high for the Full scenario, but performed similarly to the Cox and Weibull models for the SG and SGI scenarios with 20% censoring. For the SG low-censoring (0 and 0.04%) scenarios, linear regression tended to have more power than the Cox and Weibull models. Power was similar between the Weibull and Cox models for the SG scenario. For the SGI scenario, the Cox model tended to have higher power than the Weibull model across different censoring levels, and slightly lower power than the linear regression model with low censoring. Linear regression tended to have higher power than the Weibull model with the SGI scenario and 0 and 0.04% censoring. The results for the power between models presented in Figure 2 are for $\alpha = 0.05$, but trends were similar for $\alpha = 0.01$.
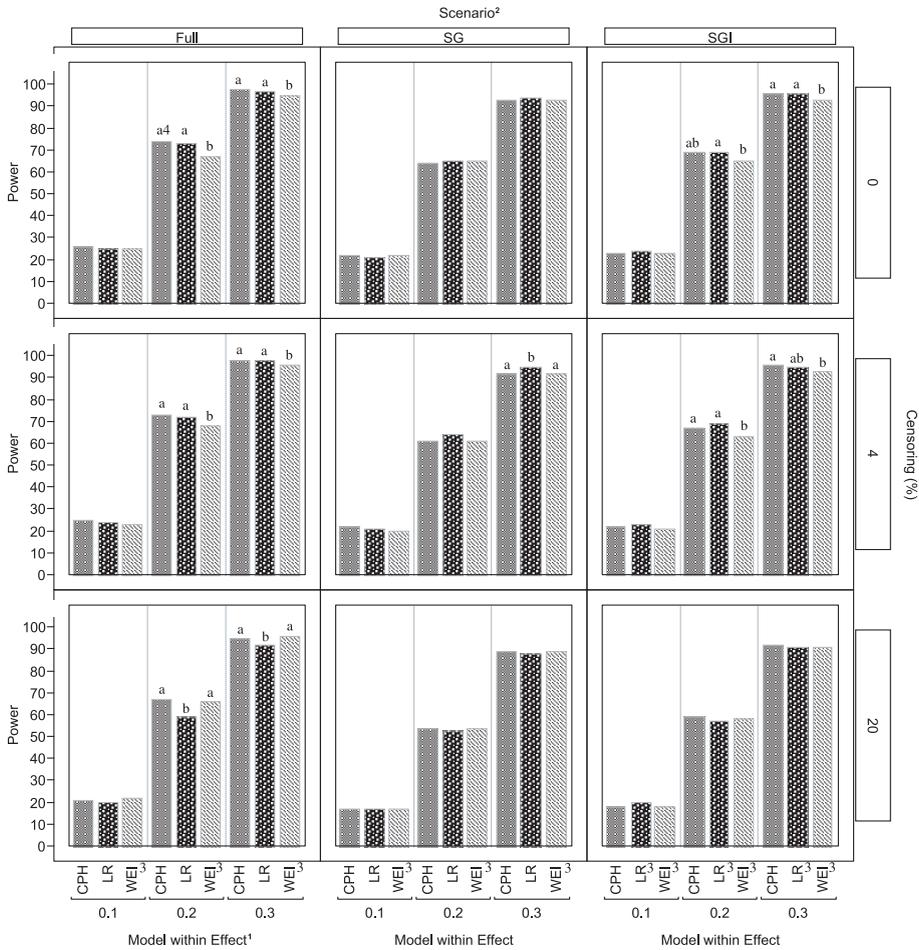
Figure 2. Power of Cox proportional hazards, linear regression, and Weibull models for different censoring levels and genotypic scenarios at $\alpha = 0.05$ based on 1000 replicates. All powers are in %.

[1] CPH = Cox proportional hazards; LR = linear regression; WEI = Weibull.

[2] Full = all individuals genotyped and used in the analysis; SG = selective genotyping with ungenotyped individuals excluded from the analysis; SGI = selective genotyping with ungenotyped individuals included in the analysis.

[3] P-value levels determined empirically for these analyses.

[4] Bars with different letter superscripts (a, b, or ab) within scenario, censoring, and effect are significantly different at $P < 0.05$. No letter indicates no significant differences ($P > 0.05$).

3.3. Power among genotyping scenarios

Although not always significant, definite trends in differences of power were apparent when comparing genotyping scenarios (Tab. II). For all three models,

Table II. Differences in power between genotyping scenarios for different censoring levels and simulated effects using Cox, linear regression, and Weibull models at $\alpha = 0.05$ based on 1000 replicates.

Model[1]	Censoring (%)	QTL Effect	Difference in Power[2]		
			Full[3]-SG (%)	Full-SGI (%)	SG-SGI (%)
CPH	0	0.1	5*	3	−1
CPH	0	0.2	11*	6*	−5*
CPH	0	0.3	5*	2*	−3*
CPH	4	0.1	3	3	−1
CPH	4	0.2	12*	6*	−6*
CPH	4	0.3	6*	1	−4*
CPH	20	0.1	4*	4	−1
CPH	20	0.2	13*	8*	−5*
CPH	20	0.3	7*	3*	−3*
LR[4]	0	0.1	4*	2	−2
LR	0	0.2	8*	4	−4
LR	0	0.3	4*	2	−2
LR	4	0.1	3	1	−2
LR	4	0.2	8*	3	−5*
LR	4	0.3	3*	2*	−1
LR	20	0.1	3	0	−4*
LR	20	0.2	7*	2	−4
LR	20	0.3	5*	2	−3*
WEI[4]	0	0.1	3	2	−1
WEI	0	0.2	3	2	0
WEI	0	0.3	2*	2	0
WEI	4	0.1	3	2	0
WEI	4	0.2	7*	5*	−2
WEI	4	0.3	4*	2*	−1
WEI	20	0.1	5*	4*	−1
WEI	20	0.2	13*	8*	−5*
WEI	20	0.3	7*	5*	−2

[1] CPH = Cox proportional hazards; LR = linear regression; WEI = Weibull.

[2] * indicates significant difference ($P < 0.05$) between genotyping scenarios from a two-tailed Fisher exact test.

[3] Full = all individuals genotyped and used in analysis; SG = selective genotyping with ungenotyped individuals excluded from the analysis; SGI = selective genotyping with ungenotyped individuals included in the analysis.

[4] *P*-values determined empirically for LR SGI and all WEI scenarios.

across QTL effects and censoring, the Full genotyping scenario had more power than SG and SGI, and SGI had more power than SG. Table II shows results for $\alpha = 0.05$ but trends were similar for $\alpha = 0.01$.

3.4. Effect estimates

For the Full genotyping scenario, estimates of allele substitution effects on the ln(hazard ratio) scale from the CPH model were never significantly ($P > 0.2$) different from the simulated values (0.1, 0.2, or 0.3) (data not shown). For the Cox model, the correlation of estimates on the ln(hazard ratio) scale with converted estimates in days (within censoring level) was less than –0.99 for all scenarios, so the Cox estimates in days were adequate to compare estimates in days from all models with the expected estimates that were in terms of the ln(hazard ratio) within censoring level. Effect estimates from the Cox model for the SG and SGI scenarios with non-zero simulated effects were significantly ($P < 0.05$) larger than expected (*i.e.* 0.1, 0.2, and 0.3), but effects were less overestimated for the SGI scenario.

Within a censoring and genotyping scenario, mean estimates from all three models were highly correlated (greater than 0.99). For a given model, correlations of mean effect estimates (in days) from different genotyping scenarios were also greater than 0.99. The rest of the results will only consider simulated effects >0. For the Full genotyping scenario, mean estimates from the Cox and linear regression models did not differ significantly ($P > 0.05$), except for the 0.2 censoring, effect = 0.3, case (Tab. III). For the Full scenario, mean estimates from the Weibull model were significantly ($P < 0.05$) greater than mean estimates from the Cox model. For the SG and SGI scenarios, mean estimates from the linear regression model were significantly ($P < 0.05$) larger than mean estimates from the Cox model for all cases. For the Full and SG scenarios, mean Weibull estimates were significantly ($P < 0.05$) larger than mean estimates from linear regression, with the exception of the SG, no-censoring scenario. For the SGI scenario with low censoring (0 and 0.04), mean estimates from linear regression were significantly ($P < 0.05$) larger than Weibull estimates. However, for the SGI high-censoring (0.2) scenario, mean estimates from the Weibull model were significantly ($P < 0.05$) larger than mean estimates from linear regression.

For all three models, standard errors of the means of the effect estimates were less for the Full scenario than for SG and SGI (Tab. III). For the Cox and Weibull models, standard errors were greater for the SG scenario than for the SGI scenario but they did not differ for the linear regression model.

Table III. Average estimates of effects (standard errors) in days from 1000 replicates for Cox proportional hazards, linear regression, and Weibull models for different censoring levels and different simulated QTL effects.

Censoring	Simulated QTL Effect	CPH[1] Full[2]	LR Full	WEI Full	CPH SG	LR SG	WEI SG	CPH SGI	LR SGI	WEI SGI
0	0	0.03(0.06)[a3]	0.05(0.06)[a]	−0.02(0.09)[a]	0.05(0.14)[a]	0.03(0.15)[a]	0.02(0.15)[a]	0.05(0.11)[a]	0.08(0.15)[a]	−0.04(0.13)[a]
0	0.1	−2.4[a4]	−2.5[a]	−3.2[b]	−4.8[a]	−5.2[b]	−5.0[ab]	−3.9[a]	−5.5[b]	−4.7[c]
0	0.2	−4.6[a]	−4.6[a]	−6.2[b]	−9.2[a]	−9.9[b]	−9.8[b]	−7.5[a]	−10.4[b]	−9.1[c]
0	0.3	−6.7[a]	−6.7[a]	−9.1[b]	−13.3[a]	−14.2[b]	−14.3[b]	−10.8[a]	−14.9[b]	−13.4[c]
0.04	0	0.01(0.06)[a]	0.07(0.06)[a]	−0.03(0.08)[a]	−0.02(0.13)[a]	0.16(0.13)[a]	−0.04(0.17)[a]	0.01(0.10)[a]	0.16(0.13)[a]	−0.06(0.13)[a]
0.04	0.1	−2.2[a]	−2.3[a]	−3.1[b]	−4.3[a]	−4.8[b]	−5.7[c]	−3.7[a]	−5.1[b]	−4.6[c]
0.04	0.2	−4.2[a]	−4.2[a]	−6.0[b]	−8.3[a]	−9.0[b]	−10.7[c]	−7.0[a]	−9.5[b]	−8.8[c]
0.04	0.3	−6.1[a]	−6.2[a]	−8.7[b]	−12.1[a]	−13.3[b]	−15.2[c]	−10.1[a]	−14.0[b]	−12.9[c]
0.20	0	0.02(0.03)[a]	0.02(0.03)[a]	0.02(0.04)[a]	0.02(0.05)[a]	0.05(0.07)[a]	0.13(0.20)[a]	0.03(0.06)[a]	0.05(0.07)[a]	0.04(0.09)[a]
0.20	0.1	−1.0[a]	−1.1[a]	−1.5[b]	−1.8[a]	−2.2[b]	−5.8[c]	−1.9[a]	−2.2[b]	−2.48[c]
0.20	0.2	−2.0[a]	−2.1[a]	−2.9[b]	−3.9[a]	−4.3[b]	−10.5[c]	−3.8[a]	−4.4[b]	−4.83[c]
0.20	0.3	−3.0[a]	−3.2[b]	−4.3[c]	−5.9[a]	−6.6[b]	−14.1[c]	−5.7[a]	−6.6[b]	−7.09[c]

[1] CPH = Cox proportional hazards; LR = linear regression; WEI = Weibull.

[2] Full = all individuals genotyped and used in the analysis; SG = selective genotyping with ungenotyped individuals excluded from the analysis; SGI = selective genotyping with ungenotyped individuals included in the analysis.

[3] Numbers with different letter superscripts (a, b, or c) within a row and genotyping scenario (Full, SG, SGI) are significantly different at $P < 0.05$. No letter indicates no significant differences ($P < 0.05$) within a row and genotyping scenario.

[4] Standard errors were equal across effects within model and genotyping scenario.

4. DISCUSSION

4.1. False positive rates

Significance tests under the Cox model were robust to the distribution of phenotypes that was included in the analysis. The Weibull model had significantly ($P < 0.05$) inflated false positive rates for all scenarios, indicating that the data did not fit a Weibull distribution. This was confirmed by significance ($P < 0.01$) of a lack of fit test of the survival distribution from the real population to the Weibull distribution using the Cramer-von Mises W test [21]. Tests based on linear regression were robust to the lack of normality in the Full and SG scenarios, but had an inflated false positive rate for the SGI scenario. Further research is needed to understand why the addition of phenotypes of the ungenotyped individuals causes linear regression to have an inflated false positive rate using this particular dataset and to determine if this inflated false positive rate occurs with datasets having different distributions.

In their comparison of models for the analysis of a non-selectively genotyped survival dataset, Moreno *et al.* [19] first transformed their data to better fit a Weibull or Gaussian distribution. With the transformed data, Moreno *et al.* [19] found that the Weibull and the Cox models had similar performances. Transformation of our data would likely have resulted in similar results for our Full scenario but for the SG scenario, the transformed dataset would still be statistically inappropriate for the Weibull model, unless all phenotypic records were included in the analysis. The effects of transforming the data on the selective genotyping scenarios are unknown.

4.2. Power differences between models

Differences in power between models varied between genotyping scenarios and censoring levels. In the Full scenario, the Cox model had higher power than the linear regression and Weibull models. This is likely because the Cox model is more appropriate for this distribution. The linear regression model had less power than the Cox and Weibull models when censoring was 20%. This was expected because the censored individuals were considered as dying on the last day of the study in the linear regression model, and therefore the extreme long-surviving individuals (which are more likely to have the favorable genotype) were grouped with less extreme long-surviving individuals that

would be expected to have a lower frequency of the favorable genotype. This situation reduces the mean phenotypic difference between individuals with the favorable allele and individuals with the unfavorable allele, and thereby reduces power.

In contrast to Moreno *et al.* [19], we found significant differences in power between the Cox and the Weibull models in the Full scenario. The discrepancy between the two studies is likely due to the fact that the data were transformed to fit a Weibull distribution in the Moreno *et al.* [19] study, whereas the data were not transformed in the present study.

For the linear regression analyses, Moreno *et al.* [19] either excluded censored individuals or included them as dying on the last day of the study, as in our study. They found that differences in power between the survival models and both linear regression analyses were small when censoring was low or random, but increased as censoring increased. When censoring was greater than 20% and at a fixed date, linear regression resulted in much lower power than the survival models when censored individuals were excluded. Inclusion of censored individuals as dying on the last day resulted in similar power as the survival models at 20% censoring, but in substantially lower power at 40% censoring. The current study included censored individuals in the linear regression analyses and only examined censoring up to 20%. The differences in power we observed between the linear regression model and the Cox model would likely have been much larger if the censored individuals were excluded from the linear regression analyses or at 40% censoring.

4.3. Power differences between genotyping scenarios

For all models and scenarios, power was somewhat (usually less than 10%) lower for the SG and SGI scenarios than for the Full scenario. This is expected because for a normally distributed phenotype, it has been shown that, for a fixed number of *genotyped* individuals, SG is more powerful than Full [3, 14, 15], but when genotyping is not constrained, Full is more powerful than SG. This is because the SG scenario excludes some of the information that is used in the Full scenario, as does the SGI scenario. For the simulated survival data, SGI had slightly (usually less than 5%) higher power than SG for all models and scenarios. To determine which individuals are the most extreme for selective genotyping, phenotypes from the entire population must usually

be obtained, so phenotypes of ungenotyped individuals should routinely be available for inclusion in a selective genotyping analysis to increase power.

4.4. Effect estimation

For the Full scenario, estimates of the hazard ratio from the Cox model were not significantly different from expectations ($P > 0.05$; data not shown). This is likely because the survival data were appropriately distributed for the Cox model. The linear regression model also estimated effects accurately in the Full scenario when compared to estimates from the Cox model after conversion to days, indicating the robustness of linear regression to deviations from normality. Moreno *et al.* [19] also found little differences between estimates from the Cox and linear regression models when censored individuals were included in the linear regression analysis as dying on the last day of the study. The Weibull model overestimated the effects in the Full scenario by approximately 50%, likely due to lack of fit of the data to a Weibull distribution.

Even though in the full scenario the estimates of the hazard ratio from the Cox model were not significantly different from the expected, the estimates in days were smaller when censoring was 20% for all models (Tab. III). With censored data, the actual length of survival of the animals is unknown. The effect in days from linear regression is the difference in the means of the two allelic groups, and it is only known that censored individuals survived *at least* to the date at which they were censored. The difference between the two allelic groups must get smaller since the length of survival time is reduced for the long surviving allelic group as a result of increased censoring from shortening the length of the study.

For the SG scenario, it is well known that linear regression overestimates effects when the phenotypes are normally distributed, and methods have been devised to correct for these biases [3, 20, 24]. The upward bias of estimates from linear regression is due to a positive correlation between QTL effects and residuals when only extreme phenotypes are included in the analysis [3]. The Cox and Weibull models also overestimated effects in the SG scenario by approximately 100%, although the Cox model did so significantly ($P < 0.05$) less than the Weibull and linear regression models. For the SGI scenario, linear regression overestimated the true effects slightly more than for the SG scenario, and the Cox and Weibull models overestimated the true effects slightly less. Skewness of the distribution may cause estimated effects from linear regression

to increase when including the ungenotyped individuals with a $p(Q)_i = 0.5$, since the mean phenotype of these individuals will not fall on the regression line between the mean phenotypes of the two extremes. With a normally distributed phenotype, the mean of the ungenotyped individuals, when given the mean genotypic probability (0.5), is expected to fall on the regression line between the means of the two phenotypically extreme groups, and therefore should not change the slope of the regression line. Inclusion of non-genotyped individuals in a selective genotyping analysis has been examined by several authors [10, 14, 31] as a way to reduce biases in effect estimates. However, all of these studies assumed a normal distribution. Further investigation into the proper statistical models for including non-genotyped individuals in analysis of non-normal, selectively genotyped data is needed.

4.5. Conclusions

Within the scope of the observed data distribution, the Cox model performed better than the linear regression and Weibull models across the tested genotyping scenarios and censoring levels, without the need to derive empirical significance thresholds. This was likely because the distribution of the data met the assumptions of the Cox model better than that of the other two models. However, survival models are much more computationally demanding than linear regression and the effect estimates from survival models are much more difficult to interpret than the effect estimates from linear regression. This, combined with the fact that the linear regression and Cox models had similar power, suggests that with little censoring and Full or SG genotypic scenarios, linear regression can be the model of choice, since false positive rates for linear regression were valid for these two scenarios. The Weibull model had an inflated false positive rate for all scenarios and linear regression had an inflated false positive rate for the SGI scenario, so significance thresholds had to be determined empirically, which is also computationally demanding.

In cases where selective genotyping is employed, more power can be obtained by including phenotypes from non-genotyped individuals in the analysis. These phenotypes should typically be available since they need to be recorded to determine which individuals are extreme. The Cox model was the only model that did not require derivation of empirical thresholds for SGI, and, therefore, may be the best model to use in this situation. Further research is needed to determine if the Cox model would also be appropriate for normally distributed traits with the SGI scenario.

The current study was based on single marker analysis, rather than employing interval mapping but the genotypic data were entered into the models as probabilities of marker genotypes. The genotypic data from an interval mapping analysis are also used to compute probabilities of genotypes at a specific locus. The methods explored herein should therefore also apply to interval mapping analyses but require further validation of model comparisons because probabilities will be different from 0 and 1 for all individuals.

REFERENCES

[1] Beaudeau F., Ducrocq V., Fourichon C., Seegers H., Effect of disease on length of productive life of French Holstein dairy cows assessed by survival analysis, J. Dairy Sci. 78 (1995) 103–117.

[2] Benard H.J., Stark K.D., Morris R.S., Pfeiffer D.U., Moser H., The 1997-1998 classical swine fever epidemic in the Netherlands – a survival analysis, Prev. Vet. Med. 42 (1999) 235–248.

[3] Darvasi A., Soller M., Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus, Theor. Appl. Genet. 85 (1992) 353–359.

[4] Ducrocq V., Besbes B., Protais M., Genetic improvement of laying hens viability using survival analysis, Genet. Sel. Evol. 32 (2000) 23–40.

[5] Falconer D.S., Mackay T.F.C., Introduction to quantitative genetics, 4th edition, Longman Group Ltd., Essex, England, 1996.

[6] Grohn Y.T., Ducrocq V., Hertl J.A., Modeling the effect of a disease on culling: An illustration of the use of time-dependent covariates for survival analysis, J. Dairy Sci. 80 (1997) 1755–1766.

[7] Grohn Y.T., Eicker S.W., Ducrocq V., Hertl J.A., Effect of diseases on the culling of Holstein dairy cows in New York state, J. Dairy Sci. 81 (1998) 966–978.

[8] Haley C.S., Knott S.A., A simple regression method for mapping quantitative trait loci in line crosses using flanking markers, Heredity 69 (1992) 315–324.

[9] Haley C.S., Knott S.A., Elsen J.M., Mapping quantitative trait loci in crosses between outbred lines using least squares, Genetics 136 (1994) 1195–1207.

[10] Henshall J.M., Goddard M.E., Multiple-trait mapping of quantitative trait loci after selective genotyping using logistic regression, Genetics 151 (1999) 885–894.

[11] Hirst W.M., Murray R.D., Ward W.R., French N.P., A mixed-effects time-to-event analysis of the relationship between first-lactation lameness and subsequent lameness in dairy cows in the UK, Prev. Vet. Med. 54 (2002) 191–201.

[12] Jansen R.C., Stam P., High resolution of quantitative traits into multiple loci via interval mapping, Genetics 136 (1994) 1447–1455.

[13] Kaplan E.L., Meier P., Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. 53 (1958) 457–481.

[14] Lander E.S., Botstein D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps, Genetics 121 (1989) 185–199.

[15] Lebowitz R.J., Soller M., Beckmann J.S., Trait-based analyses for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines, Theor. Appl. Genet. 73 (1987) 556–562.

[16] Maizon D.O., Oltenacu P.A., Grohn Y.T., Strawderman R.L., Emanuelson U., Effects of diseases on reproductive performance in Swedish Red and White dairy cattle, Prev. Vet. Med. 66 (2004) 113–126.

[17] Malek M., Dekkers J.C.M., Lee H.K., Baas T.J., Rothschild M.F., A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition, Mamm. Genome 12 (2001) 630–636.

[18] McElroy J.P., Dekkers J.C.M., Fulton J.E., O'Sullivan N.P., Soller M., Lipkin E., Zhang W., Koehler K.J., Lamont S.J., Cheng H.H., Microsatellite markers associated with resistance to Marek's disease in commercial layer chickens, Poultry Sci. 84 (2005) 1678–1688.

[19] Moreno C.R., Elsen J.M., Le Roy P., Ducrocq V., Interval mapping methods for detecting QTL affecting survival and time-to-event phenotypes, Genet. Res. Camb. 85 (2005) 139–149.

[20] Muranty H., Goffinet B., Selective genotyping for location and estimation of the effect of a quantitative trait locus, Biometrics 53 (1997) 629–643.

[21] Phillips D.T., Applied Goodness of Fit Testing, American Institute of Industrial Engineers, Inc., Atlanta, GA, 1972.

[22] R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org (2006).

[23] Rajala-Schultz P.J., Saville W.J.A., Frazer G.S., Wittum T.E., Association between milk urea nitrogen and fertility in Ohio dairy cows, J. Dairy Sci. 84 (2001) 482–489.

[24] Ronin Y.I., Korol A.B., Weller J.I., Selective genotyping to detect quantitative trait loci affecting multiple traits: Interval mapping analysis, Theor. Appl. Genet. 97 (1998) 1169–1178.

[25] Roxstrom A., Ducrocq V., Strandberg E., Survival analysis of longevity in dairy cattle on a lactation basis, Genet. Sel. Evol. 35 (2003) 305–318.

[26] Smith P.J., Analysis of failure and survival data, Chapman & Hall/CRC, Boca Raton, FL, 2002.

[27] Southey B.R., Rodriguez-Zas S.L., Leymaster K.A., Survival analysis of lamb mortality in a terminal sire composite population, J. Anim. Sci. 79 (2001) 2298–2306.

[28] Steel R.G.D., Torrie J.H., Principles and procedures of statistics: A biometrical approach, 2nd edn., McGraw-Hill Publishing Company, New York, 1980.

[29] Van Kaam J.B., Groenen M.A., Bovenhuis H., Veenendaal A., Vereijken A.L., Van Arendonk J.A., Whole genome scan in chickens for quantitative trait loci affecting carcass traits, Poultry Sci. 78 (1999) 1091–1099.

[30] Weibull W., A statistical distribution function of wide applicability, J. Appl. Mech. 18 (1951) 293–297.

[31] Xu S., Vogl C., Maximum likelihood analysis of quantitative trait loci under selective genotyping, Heredity 84 (2000) 525–537.

[32] Zeng Z.B., Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci, Proc. Natl. Acad. Sci. USA 90 (1993) 10972–10976.

[33] Zhou H., Li H., Lamont S.J., Genetic markers associated with antibody response kinetics in adult chickens, Poultry Sci. 82 (2003) 699–708.