Survival analysis: a tool in the study of post-harvest diseases in peaches

Cristiano Nunes Nesi¹, Silvia Emiko Shimakura², Paulo Justiniano Ribeiro Junior³, Louise Larissa May De Mio⁴

http://dx.doi.org/10.1590/0034-737X201562010007

ABSTRACT

Survival analysis is applied when the time until the occurrence of an event is of interest. Such data are routinely collected in plant diseases, although applications of the method are uncommon. The objective of this study was to use two studies on post-harvest diseases of peaches, considering two harvests together and the existence of random effect shared by fruits of a same tree, in order to describe the main techniques in survival analysis. The nonparametric Kaplan-Meier method, the log-rank test and the semi-parametric Cox’s proportional hazards model were used to estimate the effect of cultivars and the number of days after full bloom on the survival to the brown rot symptom and the instantaneous risk of expressing it in two consecutive harvests. The joint analysis with baseline effect, varying between harvests, and the confirmation of the tree effect as a grouping factor with random effect were appropriate to interpret the phenomenon (disease) evaluated and can be important tools to replace or complement the conventional analysis, respecting the nature of the variable and the phenomenon.

Key words: time-occurrence, Kaplan-Meier, Cox regression, peach.

RESUMO

Análise de sobrevivência como ferramenta no estudo de doenças na pós-colheita de pêssegos

A análise de sobrevivência é aplicada quando o tempo até a ocorrência de um evento for o objeto de interesse. Em doenças de plantas, dados dessa natureza são rotineiramente coletados, embora aplicações do método sejam pouco comuns. O objetivo deste trabalho foi utilizar dois estudos de doenças em pós-colheita de pêssegos, considerando-se safras conjuntamente e a existência de efeito aleatório, compartilhado por frutos de uma mesma árvore, para descrever as principais técnicas em análise de sobrevivência. Aplicaram-se a técnica não paramétrica de Kaplan-Meier e a estatística log-rank, além do modelo semiparamétrico de riscos proporcionais, de Cox, para estimar o efeito de cultivares e o número de dias após a floração plena sobre a sobrevivência ao síntoma de podridão parda e sobre o risco instantâneo de expressá-lo, em duas safras consecutivas. A análise conjunta com efeito basal, variando entre safras, e a verificação do efeito de árvore como fator de agrupamento com efeito aleatório, mostraram-se adequadas para interpretar o fenômeno avaliado (doença) e podem ser ferramentas importantes para substituir ou complementar as análises convencionais, respeitando-se as naturezas da variável e do fenômeno.

Palavras-chave: tempo-ocorrência, Kaplan-Meier, regressão de Cox, pêssego.
INTRODUCTION

The analysis of the occurrence of an event and the time for its occurrence in a population of individuals is a common statistical problem. In this context, an event is defined as a qualitative change of the individual observed, which occurs at a particular point in time (Scherm & Ojiambo, 2004). In the medical field, often the event of interest is the time until cure or death of the individual, measured from a particular treatment or the onset of the disease (McGilchrist & Aisbett, 1991; Goel et al., 2010). In such situations, a statistical analysis technique known as survival analysis was developed to be applied when the time until the occurrence of an event (the dependent variable) is the object of interest (Carvalho et al., 2011). Techniques of conventional statistical analysis are not appropriate for this type of data, because the time of observation is rarely normally distributed and data can be censored, that is, the study may end before all the individuals undergo the event of interest (right censoring), the response being partly observed (Bewick et al., 2004). Generally, data of this nature are subjected to conventional statistical analysis, which limits the inference capability. For survival data, many researchers use standard statistical techniques, such as logistic regression and ordinary least squares regression, to quantify the importance of covariates on the occurrence of an event (Scherm & Ojiambo, 2004). Logistic regression classifies of individuals into two groups, those who underwent and those who did not undergo the event during the observation period, which causes loss of information, because the differences in the occurrence times are not considered. With the least square regression analysis, the observations whose exact time of the occurrence of the event is not known (censored observations) are discarded, although they bear important information for understanding the phenomenon (Scherm & Ojiambo, 2004; Lima Junior et al., 2012). Discarding censored observations reduces the power of the statistical tests, because losses degrees freedom and introduces bias in survival functions (Colosimo & Giolo, 2006), besides overestimating the risk, because the time until the occurrence of the event is unknown (Carvalho et al., 2011). In contrast, the survival analysis uses the likelihood method for parameter estimation and effectively extract relevant information and reliable estimates, even in situations with censorship (Colosimo & Giolo, 2006). Although the term come from studies in health, survival analysis is applied in many areas of knowledge, such as demographics (Oliveira et al., 2006), economy (Oliveira & Rios-Neto, 2007), entomology (Krüger et al., 2008), agronomy (Couto et al., 2009) and education (Lima Junior et al., 2012).

In the study of plant diseases, although data on occurrence and time are routinely collected in laboratories or field trials, survival analysis is still unusual. Examples of applications of this technique to plant diseases can be found in Dallot et al. (2004), in which the authors identified risk factors in peach infection by Plum pox virus in time and, as a result, affecting the persistence of the disease. Ojiambo & Scherm (2005) used the technique to study the time to abscission of blueberry leaves, as a function of the severity of leaf spot caused by Septoria albobunctata, considering the age and location on the leaf in the canopy together. Copes & Thomson (2008) used this analysis to determine the length of the incubation period of rust (Colletotrichum gloeosporioides) in camellia twigs. The latent period of Mycosphaerella pinodes in peas was estimated using the survival analysis, considering the isolate aggressiveness, leaf wetness duration, concentration of inoculum, plant age and host susceptibility as explanatory variables (Setti et al., 2010).

In the studies cited above, the survival analysis used for plant disease data does not address common situations, such as the possibility to accommodate between-harvest variation in a joint analysis, or the existence of natural groupings as fruit of the same tree, which may influence the time to occurrence of the event (Gorfine et al., 2006) and invalidate the independence assumption of time between individuals (Colosimo & Giolo, 2006). From the foregoing, the objective of this study was to describe the main nonparametric and semiparametric techniques in survival analysis using two case studies on peach post-harvest diseases, considering two harvests together and the existence of random effect shared by fruits of a same tree.

MATERIALS AND METHODS

Two data subsets from experiments on brown rot caused by Monilinia fructicola in post-harvest peaches were used to quantify the effect of covariates related to fruit over time for symptom expression. In the examples used, there are right censored observations because the evaluations ended before all the fruit in the study showed the disease symptom.

Example 1: In the 2009/10 and 2010/11 seasons, peaches from five cultivars at the pit hardening stage were wrapped in paper bags to prevent contact with the pathogen or fungicides. After harvested, peaches were taken to the laboratory and those without any apparent damage were placed in sterilized plastic containers, inoculated with M. fructicola conidial suspension and kept in controlled temperature and humidity. Visual evaluations of brown rot in each fruit was performed
every 12 hours for five days. In this experiment, the aim of survival analysis was to quantify the effect of cultivar (covariate) on the survival of fruit (remain asymptomatic) and the risk of expressing disease symptoms. There were fruits that did not express symptoms until the end of the observation period and was considered a baseline risk for symptom expression, which can vary between seasons in the joint analysis.

Example 2: In peach trees of a same cultivar, ten green fruits per tree (24 trees) were inoculated with M. fructicola suspension at five different times: 17, 24, 49, 64, and 67 days after full bloom. Immediately after inoculation, the peaches were wrapped in wax paper bags and were kept until the harvest time. After harvest, the peaches without symptoms or injuries were incubated at constant temperature with continuous light for ten days in two consecutive seasons (2008/09 and 2009/10).

For each fruit, it was evaluated the time from the harvest until the onset of symptoms, in post-harvest. The survival analysis was used to evaluate the effect of time of inoculation (number of days after full bloom) on the survival and risk of expressing brown rot symptoms in postharvest. It was considered that the baseline risk may vary between seasons and it was evaluated the existence of a common effect shared by fruits of a same tree.

Analysis techniques

a) Survival and hazard function

Time distribution without symptoms of brown rot (survival time) was estimated in the two experimental situations. Among other things, this allows the calculation of derived quantities, such as the median survival time, which in this case means the time at which 50% of the fruit remained without symptoms and compare survival time distributions among treatments (Carvalho et al., 2011), i.e., among cultivars and fruit inoculation times.

The mathematical functions of survival $S(t)$ and instantaneous risk $\lambda(t)$ are essential in the survival analysis (Bewick et al., 2004). The survival function $S(t)$ describes the probability of an individual to have lifetime longer than $t$; in this case, the probability that, after a time $t$, the fruit shows no symptoms. This is defined as $S(t) = Pr(T > t)$, where $T$ is the time until the expression of symptom for the fruit observed. Similarly, the cumulative lifetime distribution $F(t)$ is the probability of the symptom to be expressed before the time $t$. With large number samples, the survival function $S(t)$ can be thought as a fraction of fruits without symptoms as a function of the time (Lima Junior et al., 2012). The function $\lambda(t)$ expresses the instantaneous risk of a fruit to show symptoms at the time $t$, conditional on remaining without symptoms until the time $t$:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{Pr[T \leq t + \Delta t | T \geq t]}{\Delta t}$$

The hazard function may be related to the survival function by means of various functions, among them $\lambda(t) = \frac{f(t)}{S(t)}$, where $f(t)$ is the probability density function of the fruit to show symptom and $S(t)$ is the probability of a fruit to remain without symptoms for more than a determined time.

In general, the survival analysis aims to build estimators of functions used to determine the lifetime, testing the dependence of these functions on covariates. In this study, non-parametric and semiparametric techniques of survival analysis were used.

b) The non-parametric technique

The main non-parametric technique in survival analysis is the Kaplan-Meier estimator (Kaplan & Meier, 1958). In this method, the survival function is recalculated after each fruit has expressed symptoms. The basic idea is that the probability of a fruit to remain without symptom for $k$ or more periods, from the time it entered the study, is the product of $k$ survival rates for each period (Akbar et al., 2009; Goel et al., 2010). It is assumed that symptom expression is independent among fruit and, consequently, the survival function is estimated by the product of the probabilities to remain asymptomatic until time $t$. The survival function $S(t)$ is estimated empirically by

$$S(t) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j}\right)$$

where $d_j$ is the number of fruits that showed symptom in a given time $t_j$ ($j = 1, ..., k$) and $n_j$ is the number of fruits at risk in the time $t_j$, i.e., fruit that showed no symptoms and were not removed from the study until the instant immediately preceding $t_j$.

The graph of $S(t)$ against time $t$ is called the survival curve. The Kaplan-Meier estimates this curve from the survival times, without having to assume a probability distribution, even when there is right censored data in the set of observations (Akbar et al., 2009). By convention, the Kaplan-Meier plots are represented with steps to indicate the time in which the terminal events (symptoms) occur and signs (+) to indicate censored observations (Carvalho et al., 2011).

The non-parametric approach of the survival analysis using the Kaplan-Meier method allows statistical significance tests to compare treatments (Akbar et al., 2009; Goel et al., 2010), such as cultivars and fruit inoculation times addressed in this study. In this context,
the most used test in survival analysis is the log-rank test, but should only be applied to compare groups defined by categorical variables (Akbar et al., 2009) and when the ratio of hazard functions of the compared treatments is approximately constant, which is called proportional hazards. The log-rank test evaluates the null hypothesis that there is no difference between the survival curves of each treatment, i.e., the probability of a fruit to show symptoms at any point in time is the same in all cultivars or at the time of inoculation. The statistical test is calculated by:

\[
\log(-\text{rank}) = \sum_{i=1}^{n} \frac{O_i - E_i}{E_i}, 
\]

where \(O\) and \(E\) are, respectively, the observed and expected number of fruit with symptoms in each treatment, in which the expected number is obtained by assuming that the null hypothesis is true (Carvalho et al., 2011). For the comparison of only two treatments, the value of the log-rank test has a chi-square distribution with one degree of freedom (Akbar et al., 2009).

The log-rank test can be extended to compare more than two treatments, because in many situations, the relationship between lifetime and several explanatory variables is studied simultaneously, but it is necessary to categorize continuous variables, as in Example 2. To circumvent these difficulties and give the analysis more explanatory power, parametric or semiparametric models are used.

c) The semi-parametric technique

Cox regression (Cox, 1972) is a proportional hazards model. It is defined in McGilchrist & Aisbett (1991) as the product of non-parametric and parametric components:

\[
\lambda(t) = \lambda_0(t) \exp(x' \beta),
\]

where \(\lambda(t)\) describes the risk of fruit express symptoms over time \(t\); \(\lambda_0(t)\) is the baseline risk at time \(t\)- and \(g(x' \beta)\) is the multiplicative effect of the explanatory variables (cultivars, inoculation times) combined in the \(x'\) function, which corresponds to the transposed vector \(x\). The non-parametric component \(\lambda_0(t)\) is not specified and is a non-negative function of time, usually called baseline hazard function, because \(\lambda(t) = \lambda_0(t)\), in the absence of covariates \((x = 0)\). The parametric component is often expressed as:

\[
g(x' \beta) = \exp(x' \beta) = \exp(\beta_0 x_1 + ... + \beta_p x_p).
\]

where \(x\) is the vector of explanatory variables and \(\beta\) is the vector of parameters to be estimated. The Cox model is said of proportional hazards, because the ratios between the risk rates for fruits of different cultivars or inoculation times (explanatory variables) are assumed to be constant in the follow-up time of the study (Carvalho et al., 2011).

The Cox regression model is characterized by \(\beta\) coefficients, which measure the effects of covariates on the hazard function. These quantities must be estimated from the sample observations, in order to determine the model. The regression procedure used to adjust the Cox model consists of maximizing the partial likelihood function for the parameter vector \(L(\beta)\):

\[
L(\beta) = \prod_{i=1}^{n} \left[ \frac{\exp(x_i' \beta)}{\sum \exp(x_j' \beta)} \right]^{\delta_i},
\]

where \(\delta_i\) is a random variable indicating the occurrence of the event \((\delta = 1)\) or censorship \((\delta = 0)\).

To consider the possible existence of association among times to express symptoms in fruits of a same tree, we evaluated a model with unobservable random effect and shared by the individuals, which is called frailty model. Using the classical Cox model, the inclusion of random effects was done from an unknown random variable \((Z)\), which reflects the individual heterogeneity or in this case, the fragility of each tree (McGilchrist & Aisbett, 1991; Sargent, 1998), acting multiplicatively on the baseline hazard. Therefore, for a fruit with covariates represented by the vector \(x\) and random effect \(Z = z\), the proportional hazards model became

\[
\lambda(t) = z \lambda_0(t) \exp(x' \beta).
\]

It is assumed that the frailty values are an independent sample of the random variable \(Z\) with known probability distribution, mean equal 1 and unknown variance (Kosorok et al., 2004; Gorfine et al., 2006). To estimate the variance of the random effect, it is necessary to select a statistical distribution for the random variable \(Z\). The gamma distribution was selected and the frailty variance was estimated, using the likelihood profile (EM) and the Akaike information criterion (AIC). Alternatively, frailty was assumed with log-normal distribution, where the standard method of variance estimation is the approximate restricted maximum likelihood (REML) and also the AIC criterion.

The basic assumption for the Cox model is the proportional hazards (Carvalho et al., 2011), which in this work, was evaluated by the Schoenfeld residual plot and the significance of the simple linear correlation coefficient between the Schoenfeld standardized residuals and time for each of the covariates.

Data analyses were performed using the package ‘survival’, version 2.36-12 (Therneau, 2012). R Statistical System. The commands used in the analyses are available at [http://www.leg.ufpr.br/papercompanions].

RESULTS AND DISCUSSION

The survival curves estimated by the Kaplan-Meier estimator in the two examples are shown in Figure 1, separated by seasons and together. In both examples and harvests, the probability of the fruits to remain without brown rot symptoms is reduced with time in all treatments, but the behavior is different for the cultivars among the years. The differentiation of the cultivar effect is not clear in the first harvest (Figure 1A), with less than 50% probability of the fruits to remain without symptoms after 60 hours from the inoculation. In the second harvest (Figure 1B), there is formation of two groups, one above 80% probability to remain without symptoms until the end of the study (120 hours). In the Example 2, the probability of fruit remain healthy is above 85% in the first harvest (Figure 2D) and 75% in the second harvest (Figure 2E).

The median time to symptom expression (Table 1) could not be estimated for some cultivars (Example 1) in the 2010/11 harvest and, together, between harvests and in any treatment, in Example 2, because less than 50% of fruit showed symptoms. The low number of fruit with symptoms of cultivars A, B and C explain the group of cultivars with lower risk (Figure 1B). In both examples,
the value of the log-rank statistic for comparing the Kaplan-Meier curves resulted in significant differences (p<0.05) in both seasons and jointly, indicating differences in the survival function among cultivars or inoculation times. Although the non-parametric Kaplan-Meier technique and the log-rank test do almost no restrictions on the lifetime distribution, the methods are limited because they do not allow to test the effect of different covariates simultaneously (Colosimo & Giolo, 2006; Carvalho et al., 2011). As an example of this limitation, Setti et al. (2010) determined the amount of latent *M. pinodes* in peas, with the factor levels plant age, cultivars, isolate, inoculum concentration and duration of leaf wetness compared separately using the log-rank test. In the same study, the authors used Cox regression to accommodate all the explanatory variables in the same model.

In the present work, the effects of cultivars and inoculation times can be quantified by interpreting the estimates of the Cox model parameters, which were transformed in the hazards rates (Table 2). The hazard function $\lambda(t)$ provides the inverse of the information given by the survival function $S(t)$, so that the larger the $\lambda(t)$ for a given time, the smaller the $S(t)$ (Oliveira & Rios-Neto, 2007).

To adjust the Cox model, in both examples, the data of the two harvests were grouped considering the harvest as a stratification factor, which is to say that the baseline

---

**Figure 2.** Schoenfeld standardized residuals estimated for the semi-parametric Cox model as a function of times for each covariate in the study. Plots A, B, C and D for evaluation of the effect of cultivars. Plots E, F, G and H for evaluation of the effect of inoculation time.
hazard (ëj) is not the same in both years, as one may suspect by observing the Kaplan-Meier curves (Figures 1A, 1B, 1D and 1E). In this case, the change from the baseline hazard between seasons may be related to different field conditions. The assumption of proportional hazards can be evaluated in the survival Kaplan-Meier curves (Figure 1), in which the distance between them must be approximately constant all the time. The correlation coefficient between wastes and the time for each of the treatments and the overall model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Season 1</th>
<th>Season 2</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>e</td>
<td>t</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>e</td>
<td>t</td>
</tr>
<tr>
<td>Cultivar</td>
<td>n</td>
<td>e</td>
<td>t</td>
</tr>
<tr>
<td>A</td>
<td>36</td>
<td>35</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>40</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>22</td>
<td>84</td>
</tr>
<tr>
<td>C</td>
<td>20</td>
<td>20</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>52</td>
<td>25</td>
<td>-</td>
</tr>
<tr>
<td>D</td>
<td>23</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>19</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>E</td>
<td>40</td>
<td>40</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>28</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>82</td>
<td>68</td>
<td>24</td>
</tr>
<tr>
<td>log-rank</td>
<td>34.1 (p = 7.05 x 10^-3)</td>
<td>78.2 (p = 4.44 x 10^-16)</td>
<td>60.4 (p = 2.34 x 10^-12)</td>
</tr>
<tr>
<td>Days after bloom</td>
<td>17</td>
<td>122</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>194</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>199</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>219</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>129</td>
<td>18</td>
</tr>
<tr>
<td>log-rank</td>
<td>20.9 (p = 3.36 x 10^-4)</td>
<td>53.1 (p = 8.05 x 10^-11)</td>
<td>72.7 (p = 6 x 10^-15)</td>
</tr>
</tbody>
</table>

Season 1: 2009/2010 and 2007/2008 for examples 1 and 2, respectively; Season 2: 2010/2011 and 2008/2009 for Examples 1 and 2, respectively; t: median time to symptom expression, measured in hours in example 1 and in days in example 2; It was not possible to estimate the median time; log-rank: log-rank association test.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Parameter Estimates</th>
<th>Fit Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative Risk</td>
<td>CI (95%)</td>
</tr>
<tr>
<td></td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>Cultivar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>0.82**</td>
<td>0.496</td>
</tr>
<tr>
<td>C</td>
<td>1.19**</td>
<td>0.703</td>
</tr>
<tr>
<td>D</td>
<td>1.34**</td>
<td>0.851</td>
</tr>
<tr>
<td>E</td>
<td>3.82*</td>
<td>2.559</td>
</tr>
<tr>
<td>Global</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Days after bloom</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>1.86**</td>
<td>0.582</td>
</tr>
<tr>
<td>49</td>
<td>2.77**</td>
<td>0.972</td>
</tr>
<tr>
<td>64</td>
<td>5.20*</td>
<td>1.883</td>
</tr>
<tr>
<td>67</td>
<td>9.60*</td>
<td>3.501</td>
</tr>
<tr>
<td>Global</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- Not applicable.

LL and UL correspond to the lower and upper limits, respectively, of the confidence interval for the relative risk.
is no treatment that shows a marked tendency, over time, (Colosimo & Giolo, 2006). From the foregoing, there is no evidence against the proportional hazards be a horizontal line, because with zero inclination there is no evidence against the proportional hazards (Colosimo & Giolo, 2006). From the foregoing, there is no treatment that shows a marked tendency, over time, corroborating the assumption of proportionality of hazards required by the Cox model.

In the Example 2, the observed peaches have the tree as a natural grouping, which can lead to non-independent times to symptom expression. The introduction of a random effect for each tree (frailty) makes the estimates of the covariate effects more consistent and increases confidence in the estimates (Carvalho et al., 2011). None of the methods of frailty estimation and distribution had significant variance of the random effect of tree (Figure 3), suggesting that observations of different fruits from the same tree can be considered independent for the time of brown rot symptom expression. Otherwise, the estimation method with better results for making inferences would be chosen. The null hypothesis that the effects of inoculation time are equal to zero was rejected by the Wald test, in the four models, besides providing similar estimates for the parameters. In the plots of frailty confidence intervals for each tree (Figure 3), ordered by the variance point estimate, there is no change in the order of frailties among the models and all the confidence intervals contain the value 1, indicating that there is no tree with a differential effect on the disease. It is noteworthy that trees with frailty above 1 tend to express symptoms in fruits, with faster rate than under the classical Cox model and those with frailties smaller than 1 would have longer time until the symptoms. Therefore, it was considered that the frailty models are equivalent to the classical Cox model, which was used to estimate the effect of inoculation time on the incidence of brown rot.

Regarding the relative risk (Table 2), if the value 1 is in the confidence interval (CI), it indicates no evidence that the risks of expressing symptoms differ between the treatments and of that considered standard, in this case the cultivar A in example 1, for having larger number of fruit, and in example 2, the inoculation time at 17 days after full bloom, for being the lowest value in the range. Thus, the relative risks of expressing symptoms in the cultivar B, or at 24 days after full bloom, do not differ significantly from the standards. However, fruits of the cultivar E and the inoculation time at 67 days after full bloom have 3.82 and 9.60 times more likely of express to symptoms than the standards, respectively. It is noteworthy the good accuracy of the estimates associated with risk reasons in Example 1, because of the narrow confidence intervals, which does not occur in Example 2.

Many studies in the literature use only one evaluation date to compare treatments for post-harvest disease control in fruit. Generally, the data are subjected to analysis of variance after transformation of the original variable to meet the assumptions of the analysis. Moreira & May-De Mio (2006) evaluated the effect of antagonistic fungi and chemicals on the control of brown rot in post-harvest peaches, considering only the mean incidence three days after incubation. The effects of UV-C irradiation and the interval between treatment and inoculation on the incidence of brown rot in peaches were evaluated separately, at three and four days of storage, in a factorial design and examining the data by analysis of variance and mean comparison test, in Bassetto et al. (2007). Sestari et al. (2008) evaluated the effect of physical and chemical treatments on the percentage of rot peaches in nine daily assessments, using analysis of variance for comparison of treatments in each evaluation. Villarino et al. (2010) evaluated the incidence of brown rot in post-harvest at seven days of incubation in peaches from three orchards and three consecutive harvests, in samples collected from ten trees (ten fruits per tree). The authors reported only the incidence on day seven, without comparing orchards, harvests or consider a possible tree effect. Similarly, Pinho et al. (2010) evaluated the incidence of Colletotrichum musae in nine banana genotypes inoculated with six inoculum concentrations in two evaluation dates, in a factorial design allowed by variance.

In this paper we presented the most commonly used techniques of survival analysis, as well as the application in two real problems, which allowed the illustration of a first approach to time-occurrence data. The presented techniques have extensions or different approaches to more complex problems. The occurrence of an event prior to observation in the individual (left censoring) or the occurrence within an interval (interval-censored) require different approaches of those presented here (Scherm & Ojiambo, 2004; Carvalho et al., 2011). Furthermore, when the proportional hazards assumption in the Cox model is not met, could be used models with time-dependent covariates, which are called accelerated life models (Raman & Venkatesan, 2012). In the case of continuous covariates, such as the inoculation time in the Example 2, the relationship between the covariate and the associated risk can be estimated, respecting the functional form for covariates (Gray, 1994) with non-parametric functions, such as local linear regression (lowess) and spline functions (Carvalho et al., 2011).
CONCLUSIONS

In studies to assess the incidence of diseases in post-harvest peaches, the joint analysis of experiments with basal effect varying between harvests, the evaluation of a grouping factor (trees) as a random effect and the fit of the semi-parametric Cox model are tools of the survival analysis that allow making inferences correctly, while respecting the nature of the variable.

REFERENCES


Figure 3. Frailty point estimates and respective with 95% confidence intervals using different distributions for the frailty effect (Gamma and Lognormal) and estimation algorithms (EM, AIC and REML): (A) Gama-EM with variance of 0.078 for the random effect (p = 0.14); (B) Gama-AIC with variance of 0.087 for the random effect (p = 0.14); (C) Gaussian-REML with variance of 0.074 for the random effect (p = 0.14); (D) Gauss-AIC with variance of 0.085 for the random effect (p = 0.13)
Survival analysis: a tool in the study of post-harvest diseases in peaches


