Objective

Kaplan-Meier survival curves are non-parametric methods often used to evaluate survival distribution in cohort of patients in clinical trials. In modelling the course of disease in cost-effectiveness analyses (CEA) Kaplan-Meier survival curves are popularly used as a basis for the simulation. To illustrate the usefulness of the model in survival analysis in CEA, Figura 2 shows results of the methods of estimation of survival curves for a hypothetical treatment group. Figures of the Kaplan-Meier survival curve and to fit parametric model to predict the treatment effects in time-to-event data presented in conventional population of patients. [1]

One of parametric methods broadly applicable in CEA is the Cox proportional hazards model. The advantage of this model in the survival analysis included in CEA follows largely from the fact that it is based on the assumption that the hazard function is proportional and does not change over time. Cox proportional hazards model causes extreme popularity of this method in applications. Although this method has many advantages, i.e. it is flexible and can handle censored data in clinical trials. It is not in our view that Kaplan-Meier survival curves in clinical trials are commonly used and available in clinical studies.

To use Cox proportional hazard model in CEA PH assumption must be satisfied. To check PH assumption in data reported by Kaplan-Meier plots in clinical trials two methods are proposed and compared.

Methods

The first method adopts the algorithm proposed in Guyot 2012 (2012) which closely approximates the time-to-event IPD from Kaplan-Meier graph published in clinical studies. Next, alternative analytical techniques are adopted to estimate IPD to check PH assumption. We apply unified statistical software package coxph which calculates the PH assumption for included covariates (in our case only a group effect), by censoring the corresponding observed Schoenfeld residuals at a suitable transformation of time [2]. It is well established test for verification of PH assumption based on IPD.

The second algorithm utilizes the Weibull model fitted to Kaplan-Meier data. Proposed methods require a graph of the log-logistic hazard function estimated from linear regression of log[- log(S(t)) ] vs log(t) then the graph log-logistic vs log(t), where S(t) denotes the survival function. The hazard ratio is assumed to be constant over the entire range of survival times.

The accuracy of both methods was assessed in computer simulations and by comparing results of published IPD analysis and discussed algorithms on empirical data from trials systematically identified in Medline.

Computer simulation study was conducted on the Kaplan-Meier plots generated from the Weibull model with a proportional and non-proportional hazards. The values of the Weibull distribution scale and shape parameters were simulated using uniform random number generator (U(0,1) interval. Due to the hazard function in the Weibull model equal shape parameters in both study arms were selected for proportional hazard and different for non-proportional hazard. Each simulation included IPD for two study arms, where we computed groups of respectively 25, 100 and 500 patients to examine the cohort size influence on the results of the PH assumption testing. 1000 replications were conducted for each simulation. In computer simulation censoring was not considered.

Additionally the assessment of discussed methods in application to empirical data from clinical trials was performed. A systematic Medline search was conducted to identify studies that have included:

- Kaplan-Meier plot survival data for two or more study arms, and
- description and results of PH assumption testing based on empirical IPD using statistical, analytical techniques.

Key words searches were compiled using the terms on IPD and Kaplan-Meier plots. The results of methods were compared with results of PH assumption testing in computer simulations or conventional methods used in acclimated studies. Results of conducted simulations were presented in terms of proportion of correct answers, e.g. non-rejections of the hypothesis that hazard is proportional when it is true and rejections of this hypothesis when alternative is true (hazard is non-proportional). For comparison based on the empirical data case study were conducted.

Results

Results of computer simulation indicate that in the case of proportional hazard the Weibull model performs poorly and often results in conclusion that hazards are not proportional where they were proportional. In case of non-proportional hazards model considered the algorithm behaves very well with good proportion of correct answers greater than 50% irrespectively of study arms sizes, data not shown.

Table 1. Results of computer simulation, proportion of correct answers

<table>
<thead>
<tr>
<th>Shape comparison variable</th>
<th>Guyot 2012 algorithm</th>
<th>Weibull model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportionality</td>
<td>Non-proportionality</td>
<td>Proportionality</td>
</tr>
<tr>
<td>n = 25</td>
<td>100%</td>
<td>99%</td>
</tr>
<tr>
<td>n = 100</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>n = 500</td>
<td>100%</td>
<td>96%</td>
</tr>
</tbody>
</table>

For both methods the proportion of correct answers increases for increasing study arms sizes. In case of non-proportional hazards the results are much different, algorithm performs better, for curves with similar scale parameters, as a common practice, methods works worse (with better, of course) with increasing study arms sizes. The inconsistency areas are bold. Due to small set of tests increases with number of data. In computer simulation the correct answer proportions for proportional hazards with a varying number of samples were estimated. The comparison of the test results for non-proportional and proportional hazards is presented on Figure 1.

Results of evaluation on empirical IPD from collected studies indicate that the method based on Guyot 2012 algorithm performs much better than that based on the Weibull model. Both methods are sensitive to small study arms sizes, that causes significantly worse statistical tests performance.

Table 2. Results of comparison on empirical data from trials systematically identified in Medline

<table>
<thead>
<tr>
<th>Study arms</th>
<th>Patients</th>
<th>Study arms</th>
<th>Patients</th>
<th>Study arms</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 25</td>
<td>n = 100</td>
<td>n = 500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Francis et al, 2008</td>
<td>n = 48</td>
<td>n = 96</td>
<td>n = 240</td>
<td>0.15</td>
<td>0.75</td>
</tr>
<tr>
<td>Fang 2007</td>
<td>n = 28</td>
<td>n = 56</td>
<td>n = 168</td>
<td>0.16</td>
<td>0.78</td>
</tr>
<tr>
<td>Fang et al, 2007</td>
<td>n = 70</td>
<td>n = 140</td>
<td>n = 440</td>
<td>0.18</td>
<td>0.78</td>
</tr>
</tbody>
</table>

In conclusion, the proposed algorithms are a reliable tools for testing PH assumption of time-to-event data in case of lack of IPD. It is recommended that all CEA where survival analysis was included should test PH assumption using at least one of proposed methods.

Conclusions

The algorithms are a reliable tools for testing PH assumption of time-to-event data in case of lack of IPD. It is recommended that all CEA where survival analysis was included should test PH assumption using at least one of proposed methods.

References

7. typr of痴者si用arc