VERIFICATION OF PROPORTIONAL HAZARDS ASSUMPTION IN COST-EFFECTIVENESS ANALYSIS (CEA)

ISPOR 15th Annual European Congress
Berlin, Germany, November 3–7, 2012

Zerda I, Gwiosda B, Plisko R
HTA Consulting, Krakow, Poland

A study conducted by
HTA Consulting www.hta.pl

Objective

Kaplan-Meier survival curves are non-parametric methods often used to evaluate survival distribution in cohort of patients in the clinical trials. In modelling the course of disease in cost-effectiveness analyses (CEAs) the common practice is to digitalize published Kaplan-Maier graphs and to fit parametric model to predict the treatment effects on time-to-event variables in considered population of patients. [1]

One of parametric methods broadly applicable in CEAs is the Cox proportional hazards (PH) model. The wide popularity of this model in the survival analysis included in CEAs follows largely from the fact that it is distribution-free, i.e. no assumption has to be made about the underlying distribution of survival times to make inferences about relative event rates. Convenient features of Cox model cause extensive popularity of this method in applications. Although this methodology has many advantages, it relies on PH assumption that is rarely checked and hardly verifiable in case of lack of individual patient-level data (IPD).

Due to the fact that data available in clinical studies usually are cumulative it is insufficient to get unambiguous and objective results of conventional, statistical PH assumption testing. Research work has been carried out to examine the utility of two alternative algorithms applied to data reported in clinical studies to PH assumption verification.

Methods

To use Cox proportional hazard model in CEAs 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.

The first method applies the algorithm proposed in Guyot 2012 [2] which closely approximates the time-to-event IPD from Kaplan-Meier graphs published in clinical studies. Next advanced, analytical techniques are adopted to estimated IPD to check PH assumption. We apply built-in R statistical software procedure cox.zph which calculates tests of the PH assumption for included covariates (in our case only group assignment), by correlating the corresponding set of scaled Schoenfeld residuals with a suitable transformation of time [3]. It is well-known statistical test for verification of PH assumption based on IPD.

The second algorithm utilizes the Weibull model fitted to digitalized Kaplan-Maier data. Proposed method is a variant of the graphical procedure broadly used to check the PH assumption in the Cox model. In that situation if the model satisfies the PH assumption then the graph log[-log(S(t))] vs log(t), where S(t) denotes the proportion of survivors in time t, for compared study arms should results in parallel lines. Graphical methods can be highly subjective and there are no clear guidelines how to interpret the plots. Therefore, to make it possible to compare the curves for a study arms, a criterion for rejection PH assumption is needed. In the Weibull model PH assumption is satisfied when the shape coefficients of compared curves are the same. In calculation statistical t-test for comparison of the fitted shape coefficients estimated from linear regression of log[-log(S(t))] vs log(t) were applied to verify equality of shape coefficients and consequently proportionality of hazards.

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

Computer simulation study was conducted on the Kaplan-Meier plots generated from the Weibull model with assumption of proportional and non-proportional hazards. The values of the Weibull distribution scale and shape parameters were simulated using uniform distribution on the [0.01, 10] 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 consider groups of respectively 20, 100 and 500 patients to examine the cohort size influence on the results of PH assumption testing. 1000 iterations were conducted for each simulation. In computer simulation censoring has not been 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 for survival data of two or more study arms, and
- description and results of PH assumption testing based on empirical IPD using statistical, analytical tests.

Key words searches were completed using the terms on PH and Kaplan-Meier plot. 100 abstracts were reviewed and 49 full text papers were collected for further analysis. Finally 6 publications with 11 Kaplan-Meier plots were included in analysis. From each study published Kaplan-Meier plots were digitalized and PH assumption testing results were collected.

Both of proposed algorithms were applied to simulated and published data from Kaplan-Meier plots. The results of methods were compared with results of PH testing assumed in computer simulations or conducted based on empirical IPD available in acollected 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.

Summary

OBJECTIVES: The Cox proportional hazards (PH) model is commonly used to describe time-to-event data in CEAs. Although this methodology has many advantages, it requires proportional hazards, a strong assumption that is rarely checked and hardly verifiable in case of lack of individual patient data (IPD). Time-to-event outcomes are usually reported in clinical studies by Kaplan-Meier plots with median time-to-events or hazard ratios. In CEAs, the common practice is to digitalize the published Kaplan-Maier graphs and fit parametric model to predict the treatment effects. However all these data are insufficient to get unambiguous and objective results of conventional PH assumption tests. Our aim was to present two alternative algorithms of how PH assumption may be checked based on data reported in clinical studies.

METHODS: The first method applies the algorithm proposed in Guyot 2012 (BMC Medical Research Methodology 2012, 12:9) which closely approximates the original Kaplan-Meier curves from published graphs. Advanced, analytical techniques were adopted to estimated IPD to check PH assumption. The second algorithm utilizes the Weibull model fitted to digitalized Kaplan-Maier data. Statistical tests for comparison of the fitted shape coefficients were applied to verify PH assumption (in Weibull model if difference between shape coefficients is statistically insignificant PH assumption is accepted). The accuracy of both algorithms was assessed in theoretical computer simulations and by comparing results of published IPD analysis and discussed algorithms on empirical data from trials systematically identified in Medline.

RESULTS: The validation exercise established there was agreement in results of PH testing by IPD analysis and proposed algorithms. The inconsistency areas were specified.

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 CEAs where survival analysis was included should test PH assumption using at least one of proposed methods.

Results

Results of computer simulation indicate that in the case of proportional hazard, method based on the Weibull model performs poorly and often rsesults in conclusion that hazards are not proportional where they were proportional. In case of non-proportional hazards model considered algorithm behaves very good with proportion of correct answers greater than 90% irrespectively of arms sizes.

Method based on Guyot 2012 algorithm performs quite well in both cases. For non-proportional hazards results for that method are strongly associated with the scale parameters. If scale parameters are much different, algorithm performs poorer, for curves with similar scale parameters, as is often the case in practice, method works much better (with proportion of correct answers greater than 60%, irrespective of study arms sizes, data not shown).

Detailed results are presented in Table 1.

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

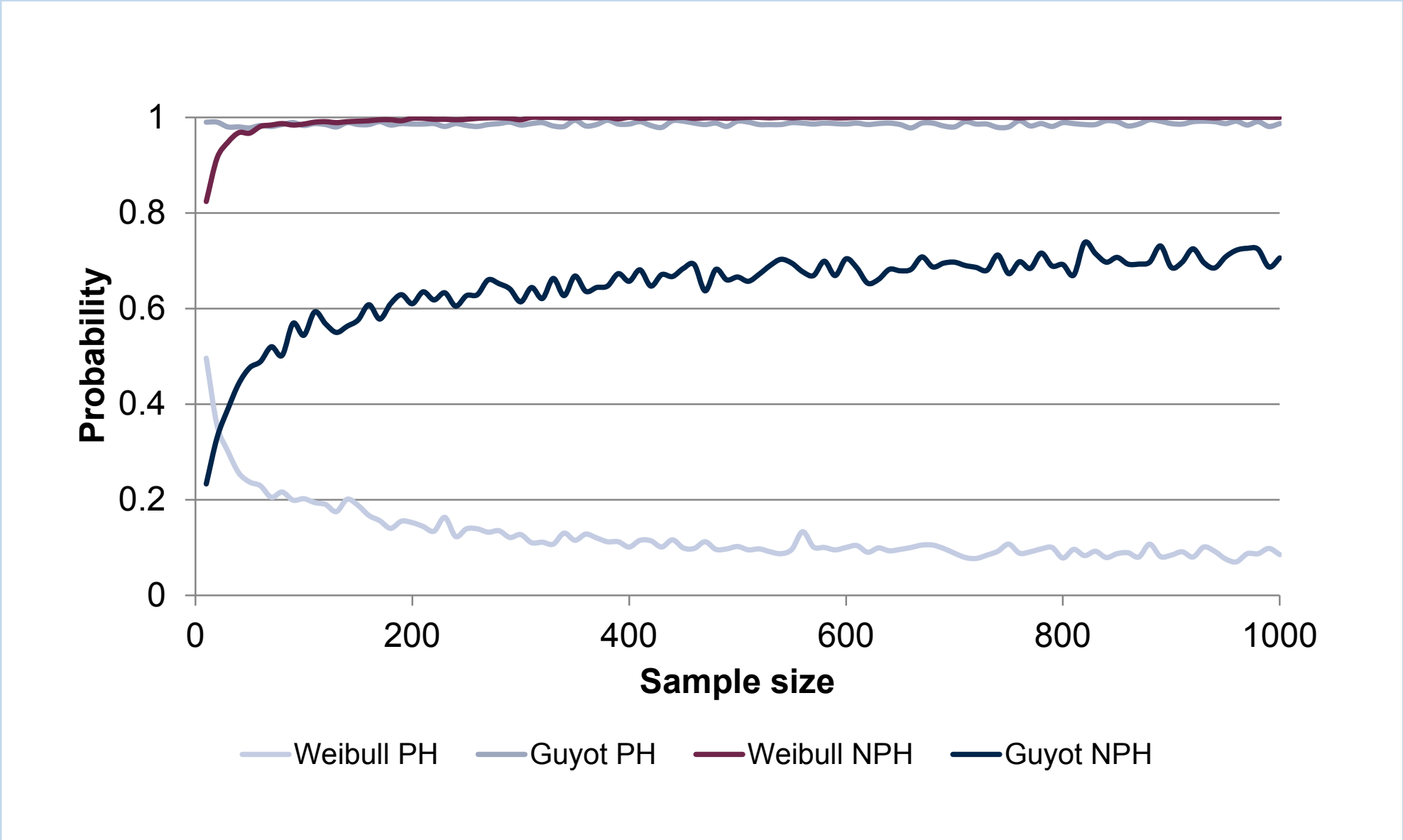
Study arms sizes (n_1, n_2)	Method based on the Weibull model	Method based on Guyot 2012 algorithm
Non-proportional hazard (different shape parameters)		
n_1, n_2 = 20	91.6%	35.9%
n_1, n_2 = 100	98.3%	54.2%
n_1, n_2 = 500	99.9%	67.2%
Proportional hazard (equal shape parameters)		
n_1, n_2 = 20	33.6%	98.9%
n_1, n_2 = 100	18.6%	98.1%
n_1, n_2 = 500	10.4%	98.2%

For both methods the proportion of correct answers increases for non-proportional hazard and decreases for proportional hazard with increasing study arms sizes. It is a result of incorporated statistical tests (t-test and test of scaled Schoenfeld residuals) features, where the power of tests increases with number of data. In computer simulation the correct answers proportions for proposed algorithms with a varying study arms sizes were estimated. The comparison of that results for non-proportional and proportional hazards model 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. Uncertainty of empirical results are also connected with model fitting error (for method based on Weibull model) and digitalization error.

Detailed results based on empirical IPD available in a collected studies are presented in Table 2. The inconsistency areas are bold. Due to small set of a reliable studies with PH assumption testing, assessment of proposed methods needs further consideration, preferably on an IPD.

Figure 1. Comparison of correct answers proportions depending on study arms sizes for proposed algorithms in case of non-proportional and proportional hazard



Conclusions

Investigating if the PH assumption is satisfied should be an integral part of a Cox survival analyses included in CEAs. 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 CEAs, which include survival analysis, should test PH assumption using at least one of proposed or equivalent methods.

References

1. Guyot et al., Survival Time Outcomes in Randomized, Controlled Trials and Meta-Analyses: The Parallel Universes of Efficacy and Cost-Effectiveness, Value in health, 2011, 14, 640 – 646;
2. Guyot et al., Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves, BMC Medical Research Methodology, 2012, 12:9;
3. Abeysekera 2009, Use of Schoenfeld's global test to test the proportional hazards assumption in the Cox proportional hazards model: an application to a clinical study, Journal of the National Science Foundation of Sri Lanka, 2009, 37(1), 41-51;
4. Fang et al., Relapse risk assessment of transplantation for patients with chronic myeloid leukaemia, Chinese Medical Journal, 2003, 116(2), 305-308;
5. Fu et al., Time-dependent effect of non-Hodgkin's lymphoma grade on disease-free survival of relapsed/refractory patients treated with high-dose chemotherapy plus autotransplantation, Contemporary Clinical Trials, 2008, 29, 157–164;
6. Maida et al., Wounds and Survival in Noncancer Patients, Journal of Palliative Medicine, 2010, 13(4), 453-459;
7. Steel et al., Extent of mesorectal invasion is a prognostic indicator in t3 rectal carcinoma, ANZ Journal of Surgery, 2002, 72, 483–487;
8. Bellera et al., Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer, BMC Medical Research Methodology, 2010, 10, 20;
9. Polkinghorne et al., Vascular Access and All-Cause Mortality: A Propensity Score Analysis, Journal of the American Society of Nephrology, 2004, 15, 477–486.

Abbreviations

IPD	Individual patient-level data
PH	Proportional hazards
CEA	Cost-effectiveness analysis

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

Study; considered variable	Prognostic factor	Study arms sizes (n_1, n_2)	PH assumption testing		
			Data from study	Method based on the Weibull model	Method based on Guyot 2012 algorithm
Bellera 2010 [4]; time to metastasis in women treated for breast cancer	Scarff-Bloom-Richardson modified grade (I vs II)	n_1=275, n_2=444	Non-proportionality	Non-proportionality	Non-proportionality
	Scarff-Bloom-Richardson modified grade (I vs III)	n_1=275, n_2=260	Non-proportionality	Non-proportionality	Non-proportionality
	tumor size	n_1=753, n_2=226	Proportionality	Non-proportionality	Proportionality
	peritumoral vascular invasion	n_1=700, n_2=279	Non-proportionality	Proportionality	Non-proportionality
	hormone receptor status (ER- and PR- vs ER+ or PR+)	n_1=801, n_2=178	Non-proportionality	Non-proportionality	Non-proportionality
Fang 2003 [5]; time to relapse after transplantation	type of donor	n_1=2411, n_2=731	Non-proportionality	Non-proportionality	Non-proportionality
Fu 2008 [6]; disease-free survival in patients with Non-Hodgkin's lymphoma	tumor grade	n_1=36, n_2=78	Non-proportionality	Proportionality	Proportionality
Maida 2010 [7]; overall survival	presence of pressure ulcers	n_1=132, n_2=57	Non-proportionality	Non-proportionality	Non-proportionality
	presence of other wounds	n_1=74, n_2=115	Proportionality	Non-proportionality	Proportionality
Steel 2002 [8]; time to local recurrence in patients with rectal cancer	stage of disease	n_1=74, n_2=148	Proportionality	Non-proportionality	Proportionality
Polkinghorne 2004 [9]; overall survival in patients with end-stage renal disease	vascular access of hemodialysis (arteriovenous fistula vs catheter)	n_1=2261, n_2=1120	Non-proportionality	Non-proportionality	Non-proportionality

ER - estrogen receptor, PR - progesterone receptor