Model Choice when Proportionality Assumption is violated: An illustration using infant mortality Data from Kenya

Sam W. Wafula (PhD) [1]

[1]Post Doctoral fellow, Fred H. Bixby program, Population Council

BACKGROUND

- Cox PH model is widely used in health research.
- Unlike standard regression models, Cox PH can model the occurrence of the event of interest as well as characterize its occurrence.
- But most studies, especially in sub Saharan Africa use Cox PH models without testing the *proportionality assumption* even where time dependency is critical.
- Proportionality assumption is flouted when any of the covariates interacts with time
- Consequences: Biased results of the estimated event of interest
- I test for the proportionality assumption using infant mortality data from Kenya and propose alternative modeling procedures in the event that the PH assumption is flouted

METHODS

- Merged KDHS data 1993, 1998 & 2003
- Log log plots and scaled Schoenfeld residuals were used to test for proportionality assumption
- Cox PH models were run to show the bias arising from violation of the PH assumption
- Parametric models (Weibull, exponential, Log- logistic, lognormal, Gompertz and Gamma) were run and compared to ascertain the best model

RESULTS

- From Table 1, *duration of breastfeeding* is the only covariate that flouted the proportionality assumption (p-value<0.05)

Table 1: Testing the proportionality assumption using the Schoenfeld residuals

Covariates	rho	chi2	df	Prob>chi2
Maternal age at birth	-0.096	0.52	1	0.4689
Prec. Birth Interval	0.115	0.71	1	0.4010
Baby size at Birth	0.060	0.24	1	0.6216
Breastfeeding period	0.201	5.04	1	0.0247*
Maternal Education	-0.057	0.19	1	0.6592
Paternal Education	-0.004	0.00	1	0.9797
Marital Status	0.048	0.15	1	0.7027
Water Source	-0.080	0.42	1	0.5193
Toilet Facility type	-0.216	2.41	1	0.1206
Household wealth Index	0.049	0.13	1	0.7175
Place of delivery	0.041	0.11	1	0.7449
Tetanus injection	-0.129	1.06	1	0.3034
Time period	0.038	0.07	1	0.7936
Global test	χ^2 = 12.26 ; df = 15 ; P-value =0.6593			

BIAS INTRODUCED WITH VIOLATION OF PH ASSUMPTION

The use of the Cox regression model when the proportionality assumption is flouted yields wrong results- as shown in Table 2

Table 2: Covariate values that flout the PH assumption- various models

Covariate	OR- Cox PH	β (exponential)	β (Weibull AFT)	β (Log logistic)	β (Lognormal	β (Gamma- AFT)	OR- Gompertz
Breastfeeding period							
10-19 months (RC)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
<10 months	4.36 (45.89)	-4.41 (0.59)	-3.58 (0.52)	-3.55 (0.51)	-3.26 (0.39)	-3.09 (0.32)	4.44 (49.85)
>20+ months	3.63 (22.12)	-3.63 (0.59)	-2.91 (0.51)	-2.87 (0.50)	-2.55 (0.38)	-2.34 (0.31)	3.63 (22.22)

Notes: Standard errors are shown in parentheses;

When Cox PH is flouted, what should guide parametric model selection? Is it theory or the calculated values of the AIC/ BIC?

Theory

•Human mortality is theoretically known to follow a Weibull distribution (Cleves *et al.,* 2010).

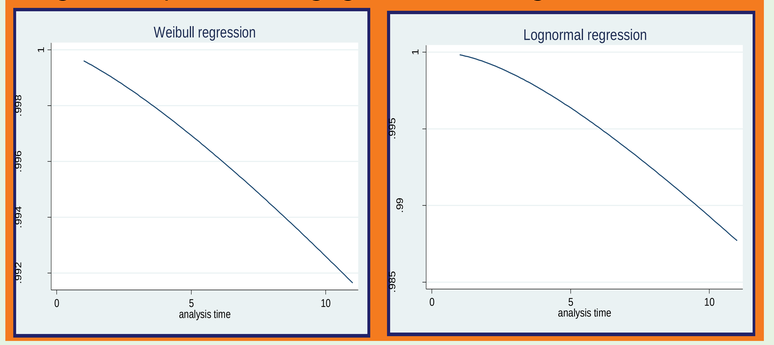
•Weibull models are also appropriate for hazards that are either monotonically increasing or decreasing (Blossfeld and Rowher, 2002).

•**Computed values of AIC/BIC** – Lognormal model is preferred (Table 3)

Table 3: Parametric Model selection based on LL, AIC and BIC

Model	Shape	Obs	Log Likelihood (null)	Log Likelihood (model)	df	AIC	BIC
Exponential	constant	6568	-1058.85	-908.48	24	1864.97	2027.93
Weibull	Monotone	6568	-1057.65	-903.32	25	1856.64	2026.38
Gompertz	Monotone	6568	-1058.65	-907.55	25	1865.11	2034.86
Lognormal	Variable	6568	-1051.87	-897.99	25	1845.97	2015.72
Gamma	Variable	6568	-1042.83	-896.03	26	1844.05	2020.59
Log-logistic	Variable	6568	-1057.36	-902.87	25	1855.73	2025.48

Figure 1: Comparison of the Log log survival curve for Lognormal and Weibull



CONCLUSION

- It is critical to test for proportionality if time dependency is important rather than a statistical nuisance
- When the proportionality assumption is flouted, Cox regression would give very biased results
- Alternative model choice should be based on theory if it exists rather than model comparison based on the lowest value of AIC/BIC.
- If no theory exists choose a model with highest maximum likelihood and smallest AIC and BIC
- Studies should also consider use of the flexible parametric survival analysis proposed by Royston and Parmer (2011)

REFERENCES

1. Box-Steffensmeier, J. M., Jones, S.B. (2004). Event History Modelling: A Guide for Social Scientists, Cambridge University Press.
2. CBS (Kenya), KEMRI, Macro International Inc. (2004). *Kenya Demographic and Health Survey 2003,* Calverton, Maryland
3. Cleves, M.A., Gould W.W, & Roberto G. Gutierrez R.G., Marchenko Y.V., (2010). *An introduction to survival analysis using STATA,* Third edition, College Station, Texas. Stata Press.
4. Blossfeld, H.-P., & Rowher, G. (2002). Techniques of event history modelling: New approaches to causal analysis. London, England: Lawrence Erlbaum Associates.
5. Belle GV, Fisher L.D., Heagerty P.J., Lumley T. (2004). *Biostatistics: A methodology for the Health Sciences*, second edition, John Wiley & Sons, New Jersey.
6. DuPont W. (2009). Statistical Modelling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data, second edition, Cambridge University Press, Cambridge.
7. StataCorp, Stata Statistical Software: Release 11.0. (2009). Stata Corporation: Lakeway Drive, College Station, TX.
8. Royston P., Lambert (2011). Flexible Parametric Survival Analysis using Stata: Beyond the Cox Model, Stata Corporation: Lakeway Drive, College Station, TX.

FOR MORE INFORMATION

I thank Prof. Ikamari and Dr. Boniface K'Oyugi of University of Nairobi for supervising me during my PhD work. This poster is part of a chapter of my PhD Thesis. I also thank my mentor - Dr. Ben Bellows of Population Council and the Fred H. Bixby Post doctoral Fellowship for enabling me to attend this conference.

For further info E-mail: swangila@popcouncil.org

Population Council

www.popcouncil.org