ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIII^th SISMEC Congress - Vol. 20 Suppl. 1

A Comparison of Baseline and Time-Dependent Approaches in Cox Model

Schiattarella Paola[1], Smimmo Annafrancesca[1], Speranza Teresa[1], Di Gennaro Piergiacomo[1], Nicolao Giovanni[1], Fordellone Mario[1], Simeon Vittorio[1], Signoriello Simona[1], Massimo Stafoggia[2], Andrea Ranzi[3], Laura Bonvicini[4], Serena Broccoli[5], Nicola Caranci[6], Nicolás Zengarini[7], Paola Angelini[5], Paolo Giorgi Rossi[4], Chiodini Paolo[1]

(1) Unità di Statistica Medica, Università degli Studi della Campania "Luigi Vanvitelli"
(2) Dipartimento Epidemiologia del S.S.R. - ASL Roma1 Regione Lazio
(3) Unità Epidemiologia Ambientale, Arpae – Agenzia Prevenzione Ambiente Energia dell'Emilia-Romagna
(4) Servizio di Epidemiologia, Azienda USL-IRCCS di Reggio Emilia
(5) Settore Prevenzione collettiva e Sanità pubblica, Regione Emilia-Romagna
(6) Settore innovazione nei Servizi sanitari e sociali, Regione Emilia-Romagna
(7) S.C. Servizio Sovrazonale di Epidemiologia ASL TO3, Regione Piemonte

CORRESPONDING AUTHOR: *Schiattarella Paola, paola.schiattarella@unicampania.it*

INTRODUCTION

In clinical and epidemiological research, risk factors vary throughout the observation period. Time-dependent covariates reflect evolving exposures and provide a dynamic view of the individual. In longitudinal settings, repeated observations within a subject are typically correlated and this correlation often decreases as the time interval between measurements increases.

Handling variables that change over time can be challenging in survival analysis framework, especially when their values are influenced by the outcome process itself. This interdependence may limit their application during specific phases of the analysis [1,2]. The extended Cox model provides a robust approach to incorporate such variables under a properly specified time-dependent structure [3,4] and is widely used in epidemiological settings. However, it is common practice, especially in cohort studies, to simplify time-varying covariates by relying on baseline measurements, potentially introducing bias when assessing the instantaneous risk.

OBJECTIVES

This contribution aims to assess the impact of modeling time-varying covariates as fixed at baseline within Cox model. A series of simulations was conducted to quantify the resulting information loss and identify key factors driving the discrepancy in results between baseline and time-dependent specifications.

METHODS

Simulation data were generated for 1.000 individuals, with a single time-dependent covariate drawn from a standard multivariate normal distribution. The covariance matrix was modeled with a first-order autoregressive structure, setting the coefficient to 0.3,0.6 and 0.9. The repeated measurements, defining the number of change-points, were evenly spaced over a fixed maximum follow-up, with the number of measures M set to 4, 8, or 16. Event times were generated using a Weibull distribution with shape parameter k set to 0.5,1 and 2, reflecting decreasing, constant and increasing hazard rates. The covariate effect size (log-hazard) was set to 0,0.2,0.4 or 1. Survival and censoring times were simulated using the permutational algorithm described in [5]. The expected censoring rate was approximately 50%. Baseline and time-dependent models were implemented and compared across 1000 repetitions for each scenario. For both models, the distributions of the estimated coefficients and their difference (, bias, and empirical statistical power were assessed over the simulation runs.

RESULTS

The highest discrimination between the models (median was observed in lower correlation and higher measurement frequency scenarios. Baseline models underestimated the true covariate effect, as shown in Table 1. Other scenarios based on a different number of measurements reflected these patterns.

DOI: 10.54103/2282-0930/29526

Table 1. Bias of estimated coefficients from the baseline model - 8 measurements per subject

			κ = 1			κ = 0.5				κ = 2			
β	ρ	$\hat{\beta}$	Bias	Rel. Bias	$\widehat{HR}$	$\hat{\beta}$	Bias	Rel. Bias	$\widehat{HR}$	$\hat{\beta}$	Bias	Rel. Bias	$\widehat{HR}$
0.2 (HR=1.22)	0.3	0.06	-0.14	-0.68	1.07	0.13	-0.07	-0.36	1.14	0.01	-0.19	-0.93	1.01
	0.6	0.10	-0.10	-0.51	1.10	0.14	-0.06	-0.29	1.15	0.04	-0.16	-0.79	1.04
	0.9	0.16	-0.04	-0.20	1.17	0.18	-0.02	-0.09	1.20	0.13	-0.07	-0.33	1.14
0.4 (HR=1.49)	0.3	0.12	-0.28	-0.69	1.13	0.26	-0.14	-0.35	1.30	0.03	-0.37	-0.93	1.03
	0.6	0.18	-0.22	-0.55	1.20	0.29	-0.11	-0.28	1.34	0.08	-0.32	-0.81	1.08
	0.9	0.32	-0.08	-0.19	1.38	0.36	-0.04	-0.09	1.44	0.27	-0.13	-0.32	1.31

Note: mean values across the repetition were reported

Across all scenarios, correlation decrease between repeated measurements led to a progressive increase in the relative bias of the estimates. This pattern can be relieved especially in the increasing hazard rate (k=2) settings, meanwhile the underestimation was mitigated in decreasing hazard rate scenarios (k=0.5).

Moreover, baseline model power improved with higher correlation between measures; conversely, low correlation reduces power, especially when measurement frequency is high. This relationship holds across various effect sizes. Type I error remained controlled in all conditions.

CONCLUSIONS

Our study shows that relying on baseline covariates may lead to underestimation of the true association between variables and outcomes, particularly in scenarios with frequent measurements and low correlation between repeated values. In these settings, the negative bias grows, revealing that the baseline value fails to represent the trajectory of the covariate adequately and in some extreme configurations indicates a near-complete failure to detect the effect. A limitation of our work is using a simulation framework based on controlled assumptions, which may not fully capture the variability and complexity of real-world data. Future research will move towards a formalization of these relationships, while also examining broader settings and additional factors to strengthen these findings. Moreover, future efforts will focus on applying these insights and pursuing further explorations within environmental epidemiology framework, which is characterized by a large amount of data, as well as numerous challenges in individual exposure assessment that may interact in various ways with the time-dependent dynamics of exposure.

ACKNOWLEDGMENTS

This project was carried out with the technical support and funding of the Italian Ministry of Health – PNC PREV-A-2022-12376981 Aria outdoor e salute: un atlante integrato a supporto delle decisioni e della ricerca.

REFERENCES

1. Kalbfleisch J.D., Prentice R.L., The statistical analysis of failure time data. John Wiley & Sons, 2002
2. Fisher L.D., Lin D.Y., Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health, 1999; 20(1):145-57
3. Therneau T.M., Grambsch P.M., Modeling Survival Data: Extending the Cox Model. Statistics for Biology and Health, Springer, 2000
4. Therneau T., Crowson C., Atkinson E., Using time dependent covariates and time dependent coefficients in the Cox model. Survival Vignettes, 2017; 2(3):1-25 (Updated December 2024; accessed on April 14, 2025)
5. Sylvestre M.P., Abrahamowicz M., Comparison of algorithms to generate event times conditional on time-dependent covariates. Stat Med., 2008; 27(14):2618-34

DOI: 10.54103/2282-0930/29526