Published in final edited form as:

Stat Med. 2009 April 15; 28(8): 1315–1317. doi:10.1002/sim.3473.

Comment on “Choice of time scale and its effect on significance of predictors in longitudinal studies”

Mitchell H. Gail¹, Barry Graubard¹, David F. Williamson², and Katherine M. Flegal³

¹*Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland* ²*Rollins School of Public Health, Emory University, Atlanta, Georgia, USA* ³*National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, USA*

The paper by Pencina, Larson and D'Agostino [1], abbreviated PLD, raises a number of important practical issues regarding the choice of time scale and adjustment for age at entry into longitudinal cohort studies. PLD compare several approaches to the preferred “entry-age-adjusted age-scale model” with hazard

$$h(a|a_0, z) = h_0(a|a_0) \exp(z'\beta), \quad (1)$$

where a is attained age, a_0 is the age at entry into the cohort, and z is a vector of covariates. In simulations, PLD used the baseline Weibull hazard $h_0(a|a_0) = \lambda \gamma a^{\gamma-1}$, which is independent of a_0 , but sampled from the left truncated Weibull distribution with truncation at a_0 and survival $S(a|a_0, z) = \exp\{-\lambda \exp(z'\beta)(a^\gamma - a_0^\gamma)\}$ for $a \geq a_0$. They found serious bias only for what they called the “unadjusted age scale model”, which does not include age at entry as a covariate and which also ignores left truncation.

In analyzing left truncated data on the attained age scale, one must decide whether or not to “adjust” for age at entry by including a_0 as a covariate in the model or by stratifying the baseline hazard by categories of a_0 . Analysts agree that one must account for left truncation in data based on attained age, either by using a left truncated likelihood for parametric models or by including in the risk sets only persons at risk after their age at entry in a Cox model analysis. Indeed, this latter option is included in widely distributed survival software, such as the ENTRYTIME statement in SAS Proc PHREG, which can be found at <http://v8doc.sas.com/sashtml/stat/chap49/sect9.htm>. Having accounted for left truncation, the analyst must decide whether to allow the baseline hazard to depend on a_0 by stratification on a_0 as in Korn, Graubard and Midthune [2] or by including a_0 as a covariate that affects the nuisance hazard. We call these procedures “covariate adjustment” on a_0 to distinguish them from left truncation adjustment on a_0 . Since the “unadjusted age scale model” in PLD fails to adjust for left truncation, it results in a large bias (Table I in PLD) even for a model that requires no covariate adjustment on a_0 because, by construction in the simulation, $h_0(a|a_0) = \lambda \gamma a^{\gamma-1}$ does not depend on a_0 .

We would insist on the need to adjust for left truncation, but whether or not it is necessary to adjust for age at entry, a_0 , in addition, depends on circumstance. Thiébaud and Bénichou [3] found for models in which $h_0(a|a_0)$ was independent of a_0 , only adjustment for left truncation was required to obtain unbiased estimates of β , even if z and a_0 were correlated. Adjustment on a_0 is needed if it is required for the validity of the proportional hazards assumption on z in equation (1). For example, if the study tends to accrue healthy subjects, the hazard may be reduced initially. Thus a 60 year old man who was accrued at age 40 may have a higher hazard

than a 60 year old man with the same z who was accrued at age 58, because the latter man is more subject to favorable selection bias from recent accrual. In this context, equation (1) may hold only if the baseline hazard is allowed to depend on a_0 .

These issues are of practical importance. For example, our group reported [4] on associations between body mass index (BMI) and mortality in the U.S. Our analyses were based on attained age and adjusted for left truncation, but no covariate adjustment was used for age at entry. We have now repeated the analysis with adjustment for age at entry, a_0 , by including it as an additional covariate in the Cox model. This adjustment had negligible effect, compared to the analysis without covariate adjustment on a_0 , either on estimated relative risks (Table I) or on the estimated numbers of deaths in the U.S. associated with various levels of BMI (Table II).

References

1. Pencina MJ, Larson MG, D'Agostino RB. Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* 2007;26:1343–1359. [PubMed: 16955538]
2. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up for a survey: choice of the time-scale. *American Journal of Epidemiology* 1997;145:72–80. [PubMed: 8982025]
3. Thiébaud AC, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine* 2004;23:3803–3820. [PubMed: 15580597]
4. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. *Journal of the American Medical Association* 2005;293:1861–1867. [PubMed: 15840860]

Table 1
Estimated hazard ratios without and with adjustment for age at entry into cohort, a_0 , according to categories of body mass index (BMI) for three ranges of attained age. Data are from the combined NHANES I-II-III data set used in reference [4].

BMI level	Ages 25-59 years			Ages 60-69 years			Ages 70+ years		
	Not adjusted	Adjusted	Change	Not adjusted	Adjusted	Change	Not adjusted	Adjusted	Change
< 18.5	1.378	1.378	0.000	2.304	2.308	0.004	1.692	1.727	0.035
18.5-<25.0	Reference category								
25.0-<30.0	0.830	0.830	0.000	0.951	0.951	0.000	0.914	0.913	-0.001
30.0-<35.0	1.202	1.202	0.000	1.127	1.128	0.001	1.033	1.032	-0.001
35.0+	1.827	1.827	0.000	1.632	1.633	0.001	1.174	1.177	0.003

Table II

Estimated excess deaths in the U.S. associated with categories of body mass index (BMI) without and with adjustment for age at entry into cohort, a_0 . Estimates are for all ages combined and from the combined NHANES I-II-III data set used in reference [4].

BMI level	Not adjusted	Adjusted	Change
< 18.5	33,746	34,957	1,211
18.5-<25.0	Reference category		
25.0-<30.0	-86,095	-86,978	-883
30.0-<35.0	29,843	29,758	-85
35.0+	82,066	83,391	1,325