Multiplicative effects and frailty intercepts on the survival for fertile and subfertile men: A piecewise constant exponential survival model for sperm count data

Ronny Westman[1], Katharina Belting[1], Hanna Seydel[1], Markus Thiede[1], Walter Krause[2], Ulrich O.Mueller[1]

[1]Institute of Medical Sociology and Social Medicine, Philipps-University Marburg

[2] Department for Andrology, Clinic of Dermatology and Allergology
Philipps-University Marburg

Key words: piecewise constant exponential survival model, infertility, frailty, multilevel analysis

Introduction:

Considering micro- and macroscopic- level information in health data analysis have become more important to understand the effects of interaction between group-level and the individual-level variables and their impact on the individual-level outcomes. Over the past few years several multilevel-approaches have been established for allowing the simultaneous modelling of group-to-group with individual-to-individual variation, just well as the inclusion of group-level properties with individual-level information (Diex-Roux, 2000). These techniques will not be only sophisticated procedures for hierarchical structured data more broadly regarding to the multiplicity view of qualitatively different levels (e.g. society, groups, individuals, organ systems, cells or genes) and provide more precisely clues to understand the substantial mechanism of health or disease (Schwartz et al. 1999).

The main purpose of our work will be to differentiate the “real” (independent) and the reciprocal associations between factors at multi levels for sperm count data. With previous analysis (Westerman et al. 2009) using the Perks-Model for analysing the survival for fertile and subfertile men we could define some intensity for unobserved heterogeneity, indeed quiet low but need to be considered for parameter estimation. We also found that age-specific effects and within-group correlation in cohorts can be assert as the major reason for the disparity in survivsl for subfertile and fertile men and might be better predictors for the population hazards.

It's also known that the quality and quantity of sperms are associated with age, while the sperm counts (decrease of 3.6% per year) and motility (decrease of 0.7% per year) is decreasing with age (Eskenazi et al. 2001) corresponding with higher risks of mortality.
Investigating the macroscoping-level point of view for the responsibility of some environmental agents (e.g. the patient's residence) and their possible toxic effects could be beneficial for the detection of changes in male reproductive function (Hellstrom et al. 2006). Age-specific referred values for sperm counts and motility could be also helpful to determine the real effect of environmental compounds evoking different risk sets for spermagonesis and the survival. Thus lead us to suppose some interaction between individual or group-level information. Just imagine that lower sperm counts or motility more likely occur in some geographic units (e.g. in counties) would be more effective to predict differential in survival but need to adjust for age or cohort-specific determinants because of different demographic composition.
If one could detect some appropriate findings often leads to some new challenges to frame some new theoretical determinants making those results more plausible. Latter seem not always feasible while successive multilevel procedures need to separate out the "real" independent effects and distinguishing between individual and group-level effect. In face of all those drawbacks, which need to be considered for we want to relax our survival model with information of different qualitative levels.

Methods:

For analysis we use the piecewise exponential survival model taking account the precise time for failure that subjects due to experiencing as the event or censoring. In contrast to the discrete-time survival model, assuming risk setting throughout the intervals, the piecewise exponential survival model allows continuous-time analysis with the drawback of supposing constant hazards for each time interval by given covariates.
Within the first step one has to estimate the hazard rates for the covariates by performing the Poisson regression. One could specify the Poisson regression for county *s*, individual *i*, and cohort *j*,

$$\ln(\mu_{sij}) = \alpha_1 \mathrm{d}_{1\mathrm{sij}} + \ldots + \alpha_x d_{xsij} + \ln(t_{sij}) \qquad (1.1)$$

with $(\mathrm{d}_{1\mathrm{ij}}, \ldots, \mathrm{d}_{\mathrm{xsij}})'$ as dummy variable for county (patient's residents) and $\mathrm{t}_{\mathrm{sij}}$ is the failure time for individual *i* in cohort *j* respond to county *s*. Alternatively one could derive the model in terms of the piecewise-constant hazard rate λ_{sij}

$$\lambda_{sij} \equiv \ln\left(\frac{\mu_{sij}}{t_{sij}}\right) = \alpha_1 d_{1sij} + ... + \alpha_n d_{nsij} \qquad (1.2)$$

The designation “piece” in the piecewise-constant Model corresponds with group-variable county. For some reason the Poisson model can be dedicated as equivalent procedure for the exponential model for the time intervals between events (Rabe-Hesketh and Skrondal, 2008). The hazarde rate is valid as the continuous-time hazard allowing the consideration of censoring because it will be also defined as instantaneous probability for every event that has not already occurred. The conditioning of the model is realized then individuals drop off the risk set as they have experienced the event or with adjustment of their exposure time until they have experienced these occasions.

Within the second step we will generate the piecewise exponential survival model with covariates and the intercept frailty term. More over we use non-aggregated data offering the opportunity to regard important covariate information.

Optional for the piecewise-exponential model one can introduce the between-cohort heterogeneity as an intercept ξj for cohort j either as random intercept being normal distributed with E(V) = 0 and $\sigma = \psi$ or the frailty $\exp(\xi j)$ specified by the gamma distribution (alternatively log-normal frailty) with E(V) = 0 and $\sigma = \alpha$

The final piecewise-exponential survival model can de derived following as:

$$\ln(\lambda_{sij}) = \beta_0 + \sum_{h=1}^{p} \alpha_h^{*} (d_{sij})^h + (\beta X) + \xi_j \qquad (1.3)$$

again with d_{sij} as dummy variable for county, α_h the “block-factor” assuming the underlying hazard function at time t, βX as vector for the covariates and ξ_j is the effect for $j-th$ cohort being typically assumed normal-distributed with zero mean and variance t.

Four our special case we introduce frailty-term $\exp(\xi j)$ exploring the between-cohort heterogeneity. The frailty is specified as the gamma-distribution. The use of the gamma-distribution for modelling the frailty can be justified with its special characteristics being most flexible in allowing the approximation for every other parametric distribution.

Designated variables

Outcome variable:	**hazard rates**
Group-level variable:	**county**, the patient's residence ("City of Marburg", "County of Marburg-Biedenkopf "Outside of County of Marburg-Biedenkopf")
Individual-level variable:	**age at date of first examination** in years, **fertility status** (subfertil/fertile, azospermic, oligospermic and normospermic) **cohort** (year of birth)
Other Covariates:	**prelimary diseases** (i.e. mumps, gonorroe affecting the system of genitourinary, **sperm-motility**

Data

The data set includes all infertility patients who had attended the fertility and sterility office of the department of andrology at Marburg University Hospital for semen analysis between 1949 and 1998 who were born before January 1st in 1942. Until now we have analyzed more than 2.000 medical records. The assignment for status of fertility was carried out by the analysis of semen samples according to the WHO declaration classifying the status of infertility (subfertility) by sperm counts of less than 20 Mio. per mL and fertility with more than 20 Mio. per mL. Otherwise it is almost necessary to designate the subfertile cases into azospermics with none sperms in ejaculate and into oligospermics with more than zero but less than 20 Mio. sperms in ejaculate because we would expect differences in survival within the subfertile subgroup. For actual analysis we include patients who have died until December 31st, 2006 so that the youngest probands were at an age of 65 years. Our data set contains 2297 cases thereof 890 will be lost to follow up because it was not possible to identify the status of vitality for all case.

Table 1 Descriptives for azoosperm, oligosperm and normosperm patients

	lost-to-follow up quotes	row death rates	mean age date of first exam.	mean age end of obs.
azoospermic	0,4202	0,3756	34,09	70,79
oligospermic	0,3889	0,3193	34,20	69,73
normospermic	0,3777	0,2898	34,58	70,22

vitality status	azoosperm	oligosperm	normosperm	Missing	Total
alive	128	194	647	2	971
dead	77	91	264	1	433
total	205	285	911	3	1404

Research statement:

The main issue of our analysis will be examining multiplicative effects between group characteristics and the composition of the individuals in groups and their impact on the incidence rates as the outcome variable. Therefore we have to respond following questions:

1. Do groups differ in average outcomes after controlling for the characteristics of individuals within them?
2. How are group-level variables related to outcomes after controlling for the individual-level information?
3. Could one assign the group to group variation for individual-level and how could they effect the function of the group-level variables?
4. Do the effects of the individual level will be modified by the group level characteristics?
5. How are the quantities of σ^2 within groups and between groups
6. Could one detect an evidence for unobserved heterogeneity between cohorts effecting the survival estimates?

References

Albright, JJ (2007) Estimating Multilevel Models using SPSS, Stata and SAS, Stat/Math Center, University of Indiana.

Diez-Roux, A.V. (2000) Multilevel Analysis in Public Health Research. Annual Review Public Health, 21,pp.171-92.

Eskenazi, B., Wyrobek, A.J., Sloter, E., Kidd, S.A., Moore, L., Young, S.S. and Moore, D. (2001)

Fahrmeier, L., Tutz, Gerhard (2001) Multivariate statistical modelling based on generalized linear models, 2nd ed., New York: Springer

Friedman M. (1982) Piecewise Exponetial Models for survival data with covariates, The Annals of Statistics, Vol.10, No.1, 101-113

Hellstrom W.J.G., Overstreet, J.W., Sikka, S.C., Denne, J., Ahuja, S., Hoover, A.M., Gregory D.S., Cordell, W.H., Harrison L.M. and Whitakers J.S.(2006): Journal of Andrology, Vol.27, No.3, pp.421-428.

Jensen, T.K., Jacobsen, R., Christensen, K., Nielsen, N.C. and Bostofte, E. (2009): Good Semem Quality and Life Expectancy: A Cohort Study of 43, 277 Men, American Journal of Epidemiology; 170 pp. 559-565.

Goldstein, H. (1986) Multilevel mixed linear model analysis using iterative generalized least squares, Biometrika 73. No.1, pp.43-56.

Goldstein, H. (2003) Multilevel Statistical Models. Third edition. Kendall library series.

Karrison, T. (1996) Confidence Intervals for median survival times under a piecewise exponential model with proportional hazards covariates effects. Statistics in Medicine, 15,pp. 171-182.

Lee, Y. (2002) Fixed-effect versus random-effect models for evaluating therapeutic preferences. Statistics in Medicine, 21, pp. 2325-2330.

Rabe-Hesketh, S., Skrondal A. (2008) Multilevel and Longitudinal Modeling Using Stata, Second Edition. A Stata Press Publication, StataCorp LP College Station Texas.

Rodriguez, G., Elo, I. (2003) Intra-class correlation in random-effects models for binary data. The Stata Journal, 3, No. 1, pp.32-46.

Recknor, J.C.; Gross, A.J.(1994) Fitting Survival Data to a Piecewise Linear Hazard Rate in the Presence of Covariates. Biometrical Journal 36, 6, pp. 729-7440

Schwartz, S, Susser E, Susser, M. (1999) A future for epidemiology? Annual Review Public Health 20, pp.1-19

Verbeke, G., Spiessens, B., Lesaffre, E. (2001) Conditional Linear Mixed Models. The American Statistican, Vol, 55, No.1 pp. 25-34.

Westerman, R, Pyka, K., Seydel, H., Krause W., Mueller, Ulrich (2009): Fertility biomarkers predict differential survival in a long-term follow up study of male infertility Capturing and treating heterogeneity, presented as Paper at Annual Meeting of Population Association of America 2009 , Detroit MI.