London, 7 October 2019

Excess hazard models in case of insufficiently stratified life tables

Bernard Rachet, Javier Rubio, Aurélien Belot

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Excess mortality hazard regression models

Individual observed hazard, $h_O(.;.)$, decomposed as:

$$h_O(t;\boldsymbol{x}) = h_P(A+t;\boldsymbol{z}) + h_E(t;\boldsymbol{x})$$

where:

- $h_P(A+t;\boldsymbol{z})$ is the population hazard at age (at diagnosis) A,
  - obtained from the lifetables ($\boldsymbol{z} \subset \boldsymbol{x}, age, sex, deprivation, \ldots$),
- $h_E(t;\boldsymbol{x})$ is the excess hazard

Assumptions

- The general population hazard correctly reflects the other-causes hazard in our population of interest
- The excess hazard is interpreted as the hazard due to the cancer under study

Insufficiently stratified life tables

- Official life tables are usually stratified by age, sex, and year
- This implies that patients sharing age, year of diagnosis, and sex, are assigned the same background mortality
  - Most-deprived and least-deprived patients
  - Smokers and non-Smokers
  - …

➢ The corresponding population hazard is either underestimated or overestimated

Sensitivity analyses using modified life tables

- Life tables adjusted for smoking using external information
  - Impact on net survival estimates for lung and laryngeal cancers
  - Small impact on deprivation gap
  - Likely stronger impact on crude probabilities of death, avoidable deaths…

- Life tables adjusted for deprivation using external information
  - Little impact on net survival estimates
  - Impact on deprivation gap in net survival

Single-parameter correction

- Cheuvart and Ryan [1991] proposed a single-parameter correction

$$h_O(t; \boldsymbol{x}|\eta) = \eta h_P(A + t; y + t; \boldsymbol{z}) + h_E(t; \boldsymbol{x})$$

where $\eta \geq 0$ is an unknown parameter

Limitations:

- Proportional excess hazard model
- Correction constant for all the patients,
  - rather unrealistic in population studies

Including covariates in the single-parameter correction

In order to alleviate this assumption, Touraine et al. [2019] proposed modelling η in terms of available covariates

- ➢ This allows for a different individual correction

However:

- It imposes a specific model for the inclusion of these variables

$$\eta_i = exp\left(x_i^T \theta\right)$$

- Not all relevant variables may be available

Correlated Frailty Model

Zahl [1997] proposed a correlated frailty model, by using frailties on both the population hazard and the excess hazard

$$h_O^Z(t; \boldsymbol{x}|\gamma_1, \gamma_2) = h_P(A + t; y + t; \boldsymbol{z})\gamma_1 + h_E(t; \boldsymbol{x})\gamma_2$$

where $(\gamma_1, \gamma_2) \sim G_2$, a bivariate gamma distribution

However, identifiability issues with this model – no maximum likelihood estimators of the parameters

Proposed solution

Rubio et al. [2019a] proposed a solution to that, adding a random correction (frailty)

$$h_O(t;\boldsymbol{x}|\eta) = \eta h_P(A + t; y + t; \boldsymbol{z}) + h_E(t;\boldsymbol{x})$$
$$\eta \sim Gamma(\mu, b)$$

- This correction
  - Is at individual level
  - Is non-specific
  - Only applies to the population hazard

Proposed solution

We can estimate the parameters using likelihood methods

- Closed-form of the marginal survival, therefore known full likelihood
- The information about the frailty parameters comes from the differences in population cumulative hazards
- Key step – model the excess hazard parametrically

- An extensive simulation study suggests that at least 5,000 observations are needed (fortunately, not an onerous condition in cancer epidemiology), and less than 50% censoring rate
- Using simulations, we have explored situations where:
  - Life tables were mismatched
  - No correction was necessary
  - These situations were identified using model selection

Real data example

- 15,688 men diagnosed with Non-Small Cell Lung Cancer (NSCLC) in England in 2012 (only complete cases)
- 13,603 died before 31 December 2015
- Life tables stratified by age, sex, year, and deprivation level

- 3 models
  - M1 = model without correction
  - M2 = single-parameter correction model
  - M3 = frailty-correction model

Real data example

	M1	M2	M3
b	–	–	9.83 (3.03)
$\gamma \mid \mu$	–	2.7 (0.21)	6.54 (0.91)
θ	0.05 (0.01)	0.03 (0.01)	0.03 (0.01)
κ	0.38 (0.01)	0.35 (0.01)	0.34 (0.01)
α	4.64 (0.34)	5.64 (0.48)	5.92 (0.58)
Age-t	0.29 (0.04)	0.29 (0.04)	0.16 (0.05)
Dep-t	0.11 (0.04)	0.12 (0.04)	0.09 (0.04)
Stage 1-t	-2.66 (0.25)	-2.17 (0.32)	-5.4 (1.4)
Stage 2-t	-2.2 (0.2)	-2 (0.22)	-2.69 (0.35)
Stage 3-t	-1.66 (0.11)	-1.57 (0.11)	-1.75 (0.13)
CV-t	0.31 (0.11)	0.31 (0.11)	0.42 (0.11)
COPD-t	0.13 (0.11)	0.08 (0.12)	0.37 (0.14)
Age	0.27 (0.01)	0.23 (0.02)	0.16 (0.02)
Dep	0.06 (0.01)	0.06 (0.01)	0.04 (0.01)
Stage 1	-2.84 (0.06)	-3.13 (0.1)	-3.53 (0.36)
Stage 2	-2.16 (0.06)	-2.32 (0.07)	-2.65 (0.1)
Stage 3	-1.23 (0.03)	-1.27 (0.04)	-1.36 (0.04)
CV	0.24 (0.04)	0.26 (0.04)	0.3 (0.04)
COPD	0.19 (0.04)	0.17 (0.04)	0.25 (0.05)
AIC	20304.69	20241.27	**20213.41**

Real data example

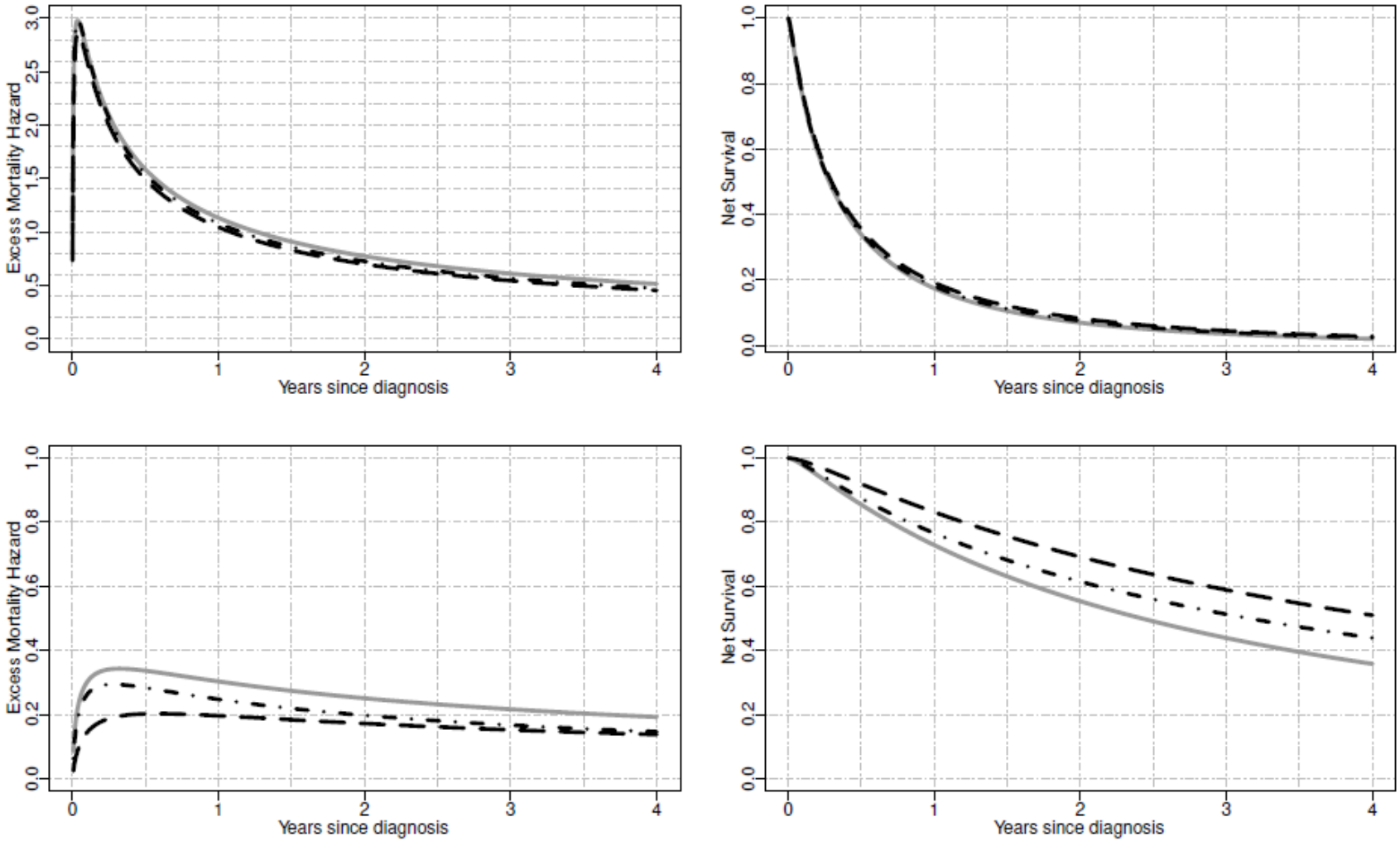

Figure: Men, 70 years old at diagnosis, Least deprived, without Cardiovascular comorbidity nor COPD, and Stage IV (upper panels) or Stage II (lower panels). M1=solid grey lines, M2=dot-dashed black lines, M3=long-dashed black lines

Quick interpretation

- We observe that the frailty distribution, used for correcting the population mortality in M3, cumulates 23% of the probability mass below 1, and 77% above 1
- These values are in fact related to the proportion of smokers (roughly 80%, which would, in principle, require a correction higher than 1) for England lung cancer patients, based on hospital data
- The impact of the presence of a comorbidity is higher in M3 compared to M1 and M2. Thus, correcting the population life table for unobserved predicting variables of background mortality seems to be quite relevant in this example.

Bibliography

1. Y.Q. Chen and N.P. Jewell. On a general class of semiparametric hazards regression models. *Biometrika*, 88(3):687–702, 2001.
2. B. Cheuvart and L. Ryan. Adjusting for age-related competing mortality in long-term cancer clinical trials. *Statistics in Medicine*, 10(1):65–77, 1991.
3. G.S. Mudholkar, D.K. Srivastava, and G.D. Kollia. A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, 91(436):1575–1583, 1996.
4. F.J. Rubio, B. Rachet, R. Giorgi, C. Maringe, and A. Belot. On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables. *Biostatistics*, [Epub ahead of print], 2019a.
5. F.J. Rubio, L. Remontet, N.P. Jewell, and A. Belot. On a general structure for hazard-based regression models: an application to population-based cancer research. *Statistical Methods in Medical Research*, 28:2404–2417, 2019b.
6. C. Touraine, N. Grafféo, R. Giorgi, and the CENSUR working survival group. More accurate cancer-related excess mortality through correcting background mortality for extra variables. *Statistical Methods in Medical Research*, [Epub ahead of print], 2019.
7. P.H. Zahl. Frailty modelling for the excess hazard. *Statistics in Medicine*, 16(14): 1573–1585, 1997.
8. Ellis L, Coleman MP, Rachet B. The impact of life tables adjusted for smoking on the socio-economic difference in net survival for laryngeal and lung cancer. *British Journal of Cancer*, **111**: 195-202, 2014.
9. Ito Y, Nakaya T, Nakayama T, et al. Socioeconomic inequalities in cancer survival: A population-based study of adult patients diagnosed in Osaka, Japan, during the period 1993-2004 [Epub ahead of print]. *Acta Oncologica*, 1-11, 2014.
10. Antunes L, Mendonça D, Bento MJ, Rachet B. No inequalities in survival from colorectal cancer by education and socioeconomic deprivation - a population-based study in the North Region of Portugal, 2000-2002. *BMC Cancer*, **16**: 1-12, 2016.