Cure models based on univariate and bivariate random effects

Andreas Wienke

imebi

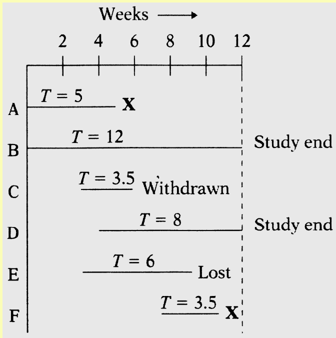
Institute of Medical Epidemiology, Biostatistics, and Informatics
Medical Faculty, University Halle-Wittenberg

content

Introduction

What is special in survival analysis?

- non-negative values, censored observations
- mixture of discrete and continuous variables: (T, Δ)
- T=min(T*,C), Δ =1(T*≤C)
- T observation time,
  T* event time,
  C censoring time,
  Δ censoring indicator



Introduction

Two typical assumptions in survival analysis:

- All individuals are susceptible to the event under study.
- All observations are independent.

cure models

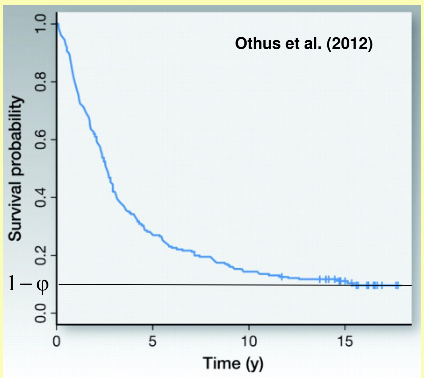
Cure models

- mixture cure models

$$S(t) = (1-\varphi) + \varphi S_0(t)$$

- nonmixture cure models

$$S(t) = (1-\varphi)^{1-S_0(t)}$$



$S_0(t)$ proper survival function

$1-\varphi$ size of the cure fraction

cure models

mixture cure models

- assumption: not everybody is susceptible to the event under study and will eventually experience the event if follow-up is sufficiently long
- sometimes individuals are not expected to experience the event of interest
- those individuals are cured or non-susceptible
- two sub-populations

cure models

- individuals are either cured with probability 1-φ or have proper survival function with probability φ
- examples: genetically influenced diseases or individuals may be vaccinated against infectious diseases
- often used model (Farewell 1982, Sy and Taylor 2000)
- incidence and latency

$$S(t \mid X) = (1 - \varphi(X)) + \varphi(X) S_0(t \mid X)$$
$$= \frac{1}{1 + e^{\alpha' X}} + \frac{e^{\alpha' X}}{1 + e^{\alpha' X}} S_0(t)^{e^{\beta' X}}$$

cure models

TABLE III. Multivariate Analysis for Risk of Recurrence, using Both Conventional Method Cox PH Model and Cure Model

	Conventional analysis using Cox PH model		Analysis using semi parametric cure model method			
Variable	HR (95% CI)	*P*-value	Incidence of recurrence OR (95% CI)	*P*-value	Time to recurrence for uncured patients HR (95% CI)	*P*-value
Age						
<50	Reference		Reference		Reference	
≥50	1.9 (1.4, 2.7)	<0.001	2.9 (1.5, 5.4)	<0.001	0.8 (0.5, 1.8)	0.50
Margin						
Negative	Reference		Reference		Reference	
Positive	1.9 (1.4, 2.8)	<0.001	3.2 (1.4, 7.2)	0.006	0.9 (0.5, 1.9)	0.85
Grade						
Low	Reference		Reference		Reference	
High	4.0 (2.6, 6.1)	<0.001	4.2 (2.1, 8.0)	<0.001	2.0 (1.0, 4.1)	0.05
Size						
≤5 cm	Reference		Reference		Reference	
>5 cm	2.2 (1.5, 3.2)	<0.001	1.7 (0.8, 3.3)	0.15	2.3 (1.3, 4.1)	0.004
Depth						
Superficial	Reference		Reference		Reference	
Deep	1.3 (0.9, 1.9)	0.25	1.8 (0.9, 3.6)	0.10	0.7 (0.4, 1.5)	0.39
Gender						
Male	Reference		Reference		Reference	
Female	1.2 (0.9, 1.6)	0.31	1.3 (0.5, 2.2)	0.31	1.0 (0.6, 1.5)	0.85
Primary site						
Lower extremity	Reference		Reference		Reference	
Upper extremity	0.8 (0.5, 1.2)	0.26	1.2 (0.6, 2.5)	0.69	0.5 (0.3, 1.0)	0.05
Histology						
ALT/WDLS[a]	Reference		Reference		Reference	
Others	2.9 (1.5, 5.9)	0.002	2.4 (1.0, 5.9)	0.059	2.4 (1.2, 5.0)	0.02

[a]ALT, atypical Lipomatous tumor; WDLS, well differentiated liposarcoma.

Soft tissue sarcoma, time to recurrence after surgery, n=682, Jia et al. (2013)

imebi A. Wienke, Cure models based on univariate and bivariate random effects, ROeS 2013

cure models

- frailty cure models by Longini & Halloran (1996), Price & Manatunga (2001)
- Y susceptible status (binary 0/1) – random effect

$$\mu(t \mid Y) = Y\mu_0(t)$$

$$P(Y=1) = \varphi, \quad P(Y=0) = 1 - \varphi$$

$$\mu(t \mid Y, Z) = YZ\mu_0(t)$$

compound Poisson frailty model

univariate compound Poisson frailty by Aalen (1992)

N Poisson distributed r.v.

V_i gamma distributed r.v.

$Z = V_1 + V_2 + \ldots + V_N$

$$\mu(t \mid Z) = Z\mu_0(t)$$

$$S(t) = e^{-\frac{1-\gamma}{\gamma\sigma^2}\left(\left(1+\frac{\gamma\sigma^2}{1-\gamma}M_0(t)\right)^{\gamma}-1\right)}$$

$$S(\infty) = e^{\frac{1-\gamma}{\gamma\sigma^2}} \text{ if } \gamma < 0$$

correlated compound Poisson frailty model

$Z_1=Y_0+Y_1$, $Z_2=Y_0+Y_2$,
Y_0, Y_1, Y_2 univariate compound Poisson distributed r.v.

$$\mu(t_1 \mid Z_1) = Z_1 \mu_0(t_1)$$

$$\mu(t_2 \mid Z_2) = Z_2 \mu_0(t_2)$$

$$S(t_1,t_2) = S(t_1)^{1-\rho} S(t_2)^{1-\rho} e^{-\frac{\rho(1-\gamma)}{\gamma\sigma^2}\left(1-\left(\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1))\right)^{1/\gamma}+\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2))\right)^{1/\gamma}-1\right)^{\gamma}\right)}$$

correlated PVF (three parameter family) frailty model
Yashin et al. (1999) when $0 \le \gamma \le 1$

correlated compound Poisson frailty model

$$S(t_1,t_2) = S(t_1)^{1-\rho} S(t_2)^{1-\rho} e^{-\frac{\rho(1-\gamma)}{\gamma\sigma^2}\left(1-\left(\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1))\right)^{1/\gamma}+\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2))\right)^{1/\gamma}-1\right)^{\gamma}\right)}$$

$\rho = 0$ (univariate frailty model)

$$S(t_1,t_2) = S(t_1)\,S(t_2) \quad \text{with} \quad S(t) = e^{-\frac{1-\gamma}{\gamma\sigma^2}\left(\left(1+\frac{\gamma\sigma^2}{1-\gamma}M_0(t)\right)^{\gamma}-1\right)}$$

frailty distribution		
gamma	$\gamma = 0$	Vaupel et al. 1979
inverse Gaussian	$\gamma = 0.5$	Hougaard 1984
PVF	$0 \le \gamma \le 1$	Hougaard 1986a
compound Poisson	$-\infty \le \gamma \le 0$	Aalen 1992

correlated compound Poisson frailty model

$$S(t_1,t_2)=S(t_1)^{1-\rho}S(t_2)^{1-\rho}e^{-\frac{\rho(1-\gamma)}{\gamma\sigma^2}\left(1-\left(\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1))\right)^{1/\gamma}+\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2))\right)^{1/\gamma}-1\right)^{\gamma}\right)}$$

$\rho = 1$ (shared frailty model)

$$S(t_1,t_2)=e^{-\frac{1-\gamma}{\gamma\sigma^2}\left(1-\left(\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1))\right)^{1/\gamma}+\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2))\right)^{1/\gamma}-1\right)^{\gamma}\right)}$$

frailty distribution		
gamma	$\gamma = 0$	**Clayton 1978**
inverse Gaussian	$\gamma = 0.5$	
PVF	$0 \le \gamma \le 1$	**Hougaard 1992**
compound Poisson	$-\infty \le \gamma \le 0$	**Hougaard 2000**

correlated compound Poisson frailty model

$$S(t_1,t_2)=S(t_1)^{1-\rho}S(t_2)^{1-\rho}e^{-\frac{\rho(1-\gamma)}{\gamma\sigma^2}\left(1-\left(\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1))\right)^{1/\gamma}+\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2))\right)^{1/\gamma}-1\right)^{\gamma}\right)}$$

$0 \le \rho \le 1$ (correlated frailty model)

frailty distribution		
gamma	$\gamma = 0$	**Yashin et al. 1993**
inverse Gaussian	$\gamma = 0.5$	**Zahl 1994**
PVF	$0 \le \gamma \le 1$	**Yashin et al. 1999**
compound Poisson	$-\infty \le \gamma \le 0$	**Wienke et al. 2010**

twin example

Swedish Twin Registry

- study population: all Swedish twin pairs born 1886-1925, both partners were still alive in 1961
- data include age at death and information about whether the twin developed breast cancer or not
- pairs with incomplete information about zygosity or breast cancer were excluded
- follow-up: January 1, 1961 - October 27, 2000
- data are bivariate right censored
- n=5857 female pairs, 715 cases of breast cancer

twin example

Breast cancer of Swedish twins (n=5857 pairs)
715 cases (Wienke et al. 2006, 2010)

	gamma frailty	**inverse Gaussian frailty**	**compound Poisson frailty**
γ	0	0.5	-0.62 (0.90)
σ^2	32.78 (7.78)	17.62 (14.22)	15.94 (8.27)
ρ_{MZ}	0.15 (0.05)	0.34 (0.13)	0.15 (0.05)
ρ_{DZ}	0.13 (0.04)	0.30 (0.11)	0.13 (0.04)
φ	**1.00**	**1.00**	**0.15**
log-L	-5122.32	-5130.59	-5121.23

discussion

- Chatterjee and Shih (2001) and Wienke et al. (2003) found 22% woman to be susceptible to breast cancer
- overall lifetime risk of breast cancer is 8 - 12 % in current western population (Harris et al. 1992; Feuer et al. 1993; Rosenthal and Puck 1999; Ries et al. 1999)
- overall lifetime risk of breast cancer is increasing (less competing risks)
- influence of genetic factors on susceptibility to breast cancer is small (5 - 10 %)

discussion

correlated compound Poisson/PVF frailty model

- more elegant than frailty cure models
- very flexible and general
- includes gamma, inverse Gaussian and PVF frailty models as special cases
- explicit form of the survival function allows traditional ML parameter estimation

- identifiability problems as in all cure models
- parametric model

references

Aalen, O.O. (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. Annals of Applied Probability 4, 951- 972

Chatterjee, N., Shih, J. (2001) A bivariate cure-mixture approach for modeling familial association in diseases. Biometrics 57, 779 – 786

Farewell, V.T. (1982) The use of mixture models for the analysis of survival data with long-term survivors.. Biometrics 38, 1041 – 1046

Longini, I.M., Halloran, M.E. (1996) A frailty mixture model for estimating vaccine efficacy. Applied Statistics 45, 165 – 173

Othus, M., Barlogie,, B., LeBlanc, M., Crowley, J. (2012) Clinical Cancer Research 18, 3731 – 3736

Price, D.L., Manatunga, A.K. (2001) Modelling survival data with a cured fraction using frailty models. Statistics in Medicine 20, 1515 – 1527

Sy, J., Taylor, J. (2000) Estimation in a Cox proportional hazards cure model. Biometrics 56, 227- 236

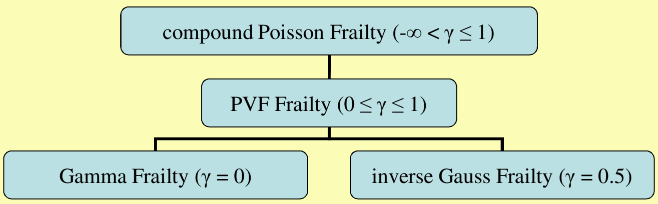
Wienke, A., Lichtenstein, P., Yashin, A.I. (2003) A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. Biometrics 59, 1178 – 1183

Wienke, A., Locatelli, I., Yashin, A. (2006) The modelling of a cure fraction in bivariate time-to-event data. Austrian Journal of Statistics 35, 67 – 76

Wienke, A. (2010) Frailty Models in Survival Analysis. Chapman & Hall/CRC, Boca Raton

Yashin, A.I., Begun, A.Z., Iachine, I.A. (1999) Genetic factors in susceptibility to death: a comparative analysis of bivariate survival models. J. Epidemiology & Biostatistics 4, 53-60

correlated compound Poisson frailty model

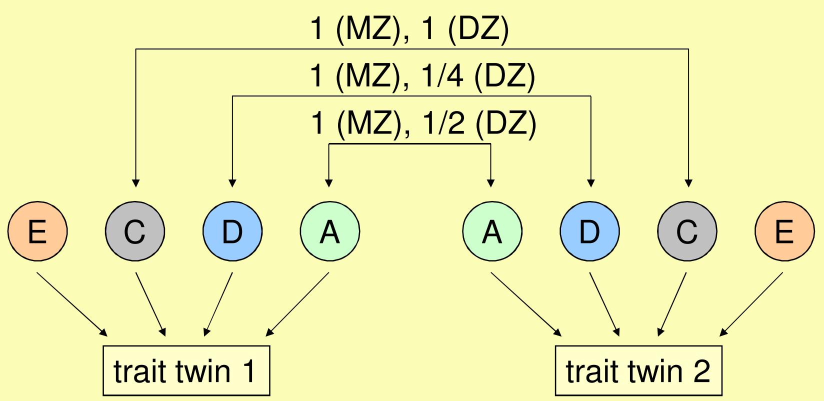


twin example

Methodology for genetic studies of twins

- twin data provide a powerful tool to assess the overall genetic influence in the variation of specific traits
- separation of the impact of genetic and environmental factors
- comparing MZ and DZ twins
- if MZ twins are more similar than DZ twins this indicates the influence of genetic factors (heritability)
- Neale & Cardon (1992)

twin example



A - additive genetic effects
D - dominant genetic effects
C - common environment
E - non-shared environment

imebi A. Wienke, Cure models based on univariate and bivariate random effects, ROeS 2013

twin example

phenotypic variance

$$V = V_A + V_C + V_D + V_E$$

$$a^2 = \frac{V_A}{V}, c^2 = \frac{V_C}{V}, d^2 = \frac{V_D}{V}, e^2 = \frac{V_E}{V}$$

$$\rho_{MZ} = a^2 + c^2 + d^2$$
$$\rho_{DZ} = 0.5a^2 + c^2 + 0.25d^2$$
$$1 = a^2 + c^2 + d^2 + e^2$$

genetic models: ACE, ADE, AE, DE, CE

twin example

Simulation study 1000 runs, 5000 twin pairs each

	true value	mean of estimates	standard error
a	2.00e-5	2.07e-5	5.60e-6
b	0.100	0.100	0.007
γ	-0.600	-0.638	0.273
σ	4.000	3.991	0.184
ρ_{MZ}	0.150	0.152	0.046
ρ_{DZ}	0.150	0.153	0.046

frailty cure models

bivariate extensions by Chatterjee & Shih (2001), Wienke et al. (2003)

$$\mu(t_1 \mid Y_1, Z_1) = Y_1 Z_1 \mu_0(t_1)$$

$$\mu(t_2 \mid Y_2, Z_2) = Y_2 Z_2 \mu_0(t_2)$$