Forward, backward and time-adjusted recurrent event processes in the presence of a terminal event

Mei-Cheng Wang
Dept. of Biostatistics, Johns Hopkins University

The Joint Pepper Centers Biostatistics Symposium

September 9, 2011

Outline

- **Description of data**
  recurrent events, time to terminal event (failure), marker measurements
- **Standard recurrent event data**: fundamental tools/models
  rate functions, conditional and marginal regression models
- **Current data**: three approaches
- **Descriptive examples** in mental health, cancer, AIDS
  Some data analysis, if time allows.

Recurrent events, markers, and terminal event

- **Recurrent event process:** $N(t)$
  - multiple tumors; repeated hospitalizations
- **Markers given occurrence of recurrent events**: $X(t)|_{dN(t)=1}$
  - health or prognostic measurement; biomarkers, medical cost
  - marker could be treated as covariate or outcome measurement
- **Terminal event**, T: time to terminal event
  - time to death, disease, failure
- **Data**: observe recurrent events, markers and covariates until terminal event or censoring, whichever occurs first.

Examples

Example 1. Denmark schizophrenia registry data

$N(t)$ is hospitalizations for schizophrenia

$X(t)|_{dN(t)=1}$ = prognostic measurements

T= time to death

Example 2. HIV-AIDS trials

$N(t)$: opportunistic diseases for HIV-infected patients

$X(t)|_{dN(t)=1}$ = prognostic measurement, severity score

T= time to AIDS

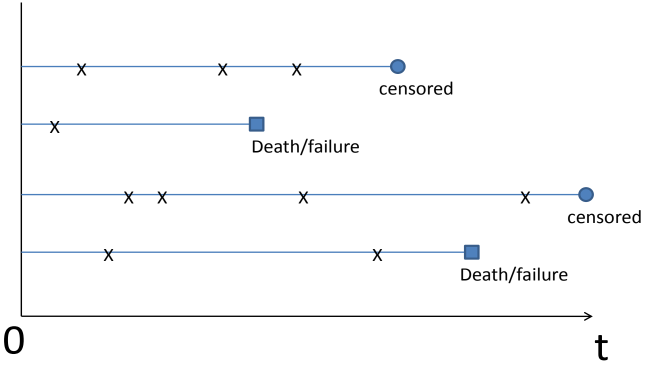
Example 3. Ovarian cancer SEER study

$N(t)$ is hospitalizations for ovarian cancer

$X(t)|_{dN(t)=1}$ = medical cost for hospitalization

T= time to death

Plot of observed data



Statistical challenges

- $N(t)$ is typically correlated with T; observation of recurrent events is subject to dependent censoring.
- When terminal event = death, survivor population changes over time - which population to focus on?
- Proper/interesting models? Accompanied statistical methods?
- Need methodology for analyzing integrated data:
  recurrent event data + survival data (+ marker data)

Fundamental functions for recurrent event process

- **Conditional rate function** is also called the **intensity function**, which is the rate function *conditional* on the event history,

$$\lambda(t|N_H(t)) = \lim_{\Delta t \to 0^+} \frac{\Pr(N(t+\Delta) - N(t) > 0 \mid N_H(t))}{\Delta},$$

$N_H(t) = \{N(u) : 0 \leq u \leq t\}$: recurrent event history before or at t

- **Marginal rate function** is the rate function *unconditional* on the event history $N_H(t)$:

$$\lambda(t) = \lim_{\Delta \to 0^+} \frac{\Pr(N(t+\Delta) - N(t) > 0)}{\Delta}.$$

Useful semiparametric models for recurrent events

- **Conditional rate model** $\lambda(t \mid N_H(t), Z_H(t)) = \lambda_0(t) exp\{X(t)\beta\}$
  $X(t) = \phi(N_H(t), Z_H(t))$: a transformation of $(N_H(t), Z_H(t))$

  Prentice, Williams and Peterson (1981); Andersen and Gill (1982)

  Marginal rate model $\lambda(t \mid Z_H(t)) = \lambda_0(t) exp\{Z(t)\beta\}$

  Lin et al. (2000)

- **Critiques.** These two useful models are defined over time (t) subject to the same population. Methods for these two models heavily rely on independent censoring assumption.
- **If terminal event is death, the models and methods become inappropriate!**

Our research

Data: observe recurrent events, possibly marker measurements until terminal event (death or failure) or censoring, whichever occurs first.

Proposed approaches: either nonparametric or semiparametric models

Approach 1. Joint model of recurrent events and time-to-failure

Approach 2. Failure-time-adjusted recurrent event models

Approach 3. Backward recurrent marker process model

1. Joint model of recurrent events and time-to-failure

Subject-specific joint modeling on $(N(t), T)$

$Z = z$: subject-specific latent variable

$X = x$: covariate vector

- **model**

  hazard function of T: $h(t) = zh_0(t)\exp(\beta' x)$

  rate function for $N(t)$: $\lambda(t) = z\lambda_0(t)\exp(\alpha' x)$

- **independent censoring assumption**

  $(N(\cdot), T, C)$ are independent conditioning on (x, z)

  C: censoring time for other reasons

- **methodological approach** conditional likelihood, estimating equations

  A nice feature is that distribution of Z is a nonparametric component.

1. Joint model of recurrent events and time-to-failure

Data application

- Denmark national-based registry data; long-term disease data recorded from 1st hospitalization to death

  - repeated psychiatric measurements
  - recorded hospitalization information over lifetime

- Recurrent events: schizophrenia-related hospitalizations
  - terminal event = death
  - administrative censoring

2. Failure-time-adjusted recurrent event process

Use T to adjust time-scale for recurrent event process $N(t)$

- **model**

  $\ln T = \beta' x + \ln T_0$

  $\mathrm{E}[dN(t\exp(\beta' x)) \mid T, X = x] = \exp(\alpha' x)\mathrm{E}[dN_0(t) \mid T_0],\ 0 \le t \le T_0$

  assuming: $(T, N(\cdot))$ is independent of C conditional on $X = x$

- **example.** If $x = 0, 1;\ \alpha = 0, \beta = 0.5$, then

  $\mathrm{E}[dN_{ctrl}(t) \mid T_{ctrl}] = \mathrm{E}[dN_0(t) \mid T_0]$

  $\mathrm{E}[dN_{tr}(1.65t) \mid T_{tr}] = \mathrm{E}[dN_0(t) \mid T_0]$

  example. If $x = 0, 1;\ \alpha = 0, \alpha = 0.2, \beta = 0.5$, then

  $\mathrm{E}[dN_{ctrl}(t) \mid T_{ctrl}] = \mathrm{E}[dN_0(t) \mid T_0]$

  $\mathrm{E}[dN_{tr}(1.65t) \mid T_{tr}] = 1.22 \cdot \mathrm{E}[dN_0(t) \mid T_0]$

2. Failure-time-adjusted recurrent event process

Data application

- Terry Beirn Community Programs for Clinical Trials on AIDS (CPCRA)

  - HIV-infected patients
  - didanisine (ddI) v.s. zalcitabine (ddC) treatments

- - recurrent events: repeated opportunistic diseases
  - terminal event: diagnosis of AIDS
  - censoring: loss to follow-up or administrative censoring

3. Backward recurrent marker process model

- **Backward model**
  - using death as time-origin and counting time backward
  - model $N(t)$ and $M(t) = \int_0^t X(t)dN(t)$
  - $X(t)$= cost per hospitalization, $M(t)$= total cost in t-unit time
- **Research interest**
  - pattern changes of $N(t)$, $M(t)$ within k months before death
  - estimation of E$[M(t)]$, rate of $M(t)$,... etc.
    (average total cost within t months before death, rate of cost at t months before death,... etc)
- **Methodological challenge**
  - methods developed to deal with 'censoring for backward data'
  - nonparametric, semiparametric

3. Backward recurrent marker process model

Data application.

- SEER-Medicare linked data for ovarian cancer
  - Ovarian SEER data provide information of ovarian cancer incidence and follow-up.
  - Medicare data provide information of medical costs for 97% US population 65 years or older
- - Recurrent events: multiple hospitalizations due to o.c.
  - Marker measurements: hospitalization cost
  - Terminal event: death
  - Censoring: administrative censoring
- Research interest: pattern of hospitalization cost before death, in relation to diagnosis category and treatment plans.

Discussion for aging research

- For aging research, it is not uncommon to deal with

  - recurrent events (repeated falls, injuries)

  - marker measurements (biomarkers, medical cost)

  - terminal event (death, MCI, AD, ...)

- For 'backward process', statistical methods in literature are ad hoc and some in error!

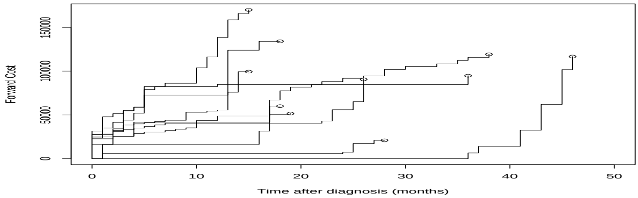
  - change-point for biomarkers, in relation to failure event

- Integration of different data sources

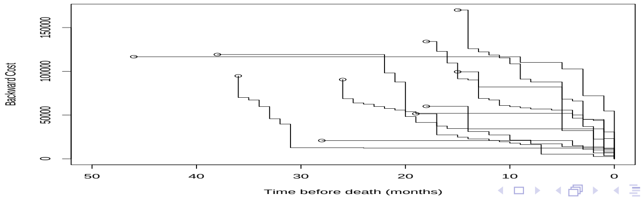
THANK YOU !

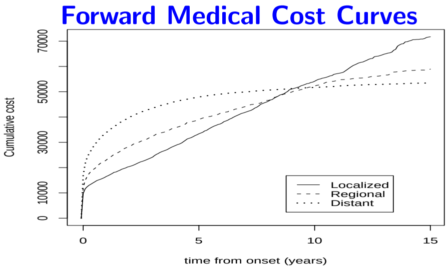
Approach 3: Data analysis

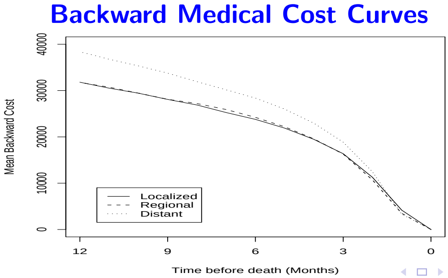
Forward Medical Cost Data

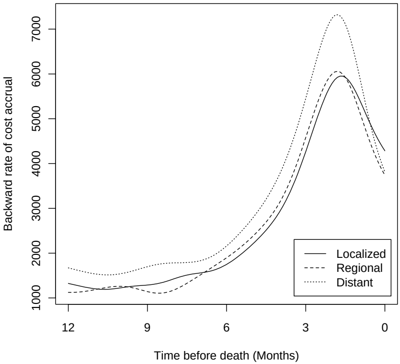


Backward Medical Cost Data




Forward Medical Cost Curves
Cumulative cost
0
10000
30000
50000
70000
0
5
10
15
time from onset (years)
Localized
Regional
Distant


Backward Medical Cost Curves
Mean Backward Cost
0
10000
20000
30000
40000
12
9
6
3
0
Time before death (Months)
Localized
Regional
Distant


Backward rate of cost accrual
1000
2000
3000
4000
5000
6000
7000
12
9
6
3
0
Time before death (Months)
Localized
Regional
Distant

- Estimated Pearson correlation coefficient between $M^B(1\ year)$ and T, conditioned on $T \geq 1$ year, is

localized	regional	distant
-0.65	-0.31	-0.46

- Estimated final-year medical cost is

localized	regional	distant
\$31802	\$31752	\$38377

(s.e.: \$1229, \$2205, \$896)

Approach 2: Analysis of CPCRA Trial

	$\widehat{\beta}_T$	Model A: $\widehat{\beta}_R^A$	Model B: $\widehat{\beta}_R^B$
Estimate	.179	.153	.195
95% Wald-type CI	(−.023, .381)	(−.199, .504)	(−.056, .447)
95% bootstrap CI	(−.015, .375)	(−.217, .508)	(−.061, .458)
P-value	.082	.394	.129

NOTE: Treatment is coded 0 for ddI and 1 for ddC.

Approach 1: too much for 30 min. presentation (see our papers for schizophrenia data analysis

References

Conditional and marginal regression models for recurrent event data

Prentice, R.L., Williams, B.J. and Peterson, A.V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373-379.

Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* 10, 1100–1120.

Chang S.-H. and Wang, M.-C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association,* 94:1221-1230.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000) Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society* Ser. B 62, 711–730.

Our approaches 1, 2 & 3 in this talk

Wang, M.-C, Qin, J., and Chiang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* **96**, 1057–1065.

Wang, M.-C. and Chiang, C.-T. (2002). Nonparametric methods for recurrent event data with informative and non-informative censorings. *Statistics in Medicine,* (21):445-456.

Huang, E. and Wang, M.-C. (2003). Frequency of recurrent events at failure time: Modeling and inference. *Journal of the American Statistical Association* 98, 663–670.

Huang C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation of recurrent event processes and failure time. *Journal of the American Statistical Association.Association,* Vol. 99, No. 468, pp.1153-1165.

Huang CY, Qin J, Wang MC (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. Biometrics, 2010 Mar; 66(1):39-49.

Chan, G. and Wang MC. (2010). Backward estimation of medical cost in the presence of a failure event. Annals of Applied Statistics. 2010, Vol. 4, No. 3, 1602-1620.

Gap time models for recurrent event data

Aalen, O. O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine,* **10**, 1227-1240.

Wang, M.-C. and Chang, S.-H. (1999). Nonparametric estimation of a recurrent survival function. *Journal of the American Statistical Association,* 94:146-153.

Wang, M.-C. (1999). Gap time bias in incident and prevalent cohorts. *Statistica Sinica* 9:999-1010.

Wang MC and Chen Y.-Q. (2000). Nonparametric and semiparametric trend analysis for stratified recurrence time data *Biometrics* 56(3):789-794.

Chen Y. Q., Wang, M.-C. and Huang E. (2004). Semiparametric regression analysis on longitudinal pattern of recurrent gap times. *Biostatistics,* 5, 2, pp. 277-290.

Huang C.-Y. and Wang, M.-C. (2005). Nonparametric estimation of a bivariate distribution of recurrence times. *Biometrics* 61, 392-402.

Other relevant references

Eaton, W. W., Mortensen, P. B., Herrman, H., Freeman, H., Bilker, W., Burgess, P., and Wooff, K. (1992a), "Long-term Course of Hospitalization for Schizophrenia: Part I. Risk for Hospitalization," *Schizophrenia Bulletin*, 18, 217–228.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure

Lawless, J. F. and Nadeau, C. (1995) Some simple robust methods for the analysis of recurrent events. *Technometrics* 37, 158–168.

Lancaster, T. and Intrator, O. (1998). Panel Data with Survival: Hospitalization of HIV-positive patients. *Journal of the American Statistical Association,* **93**, 46-53.

Ghosh, D. and Lin, D. Y. (2003). Semiparametric analysis of recurrent events in the presence of dependent censoring. Biometrics 59, 877-885.

Fine, J. P., Yan, J. and Kosorok, M. R. (2004) Temporal process regression. *Biometrika* 91, 683–703.