It is Time to Say "Goodbye" to Poisson RegressionIII

Application of Hazard model with Exposure History

Yutaka Hamaoka
hamaoka@fbc.keio.ac.jp
Faculty of Business and Commerce, Keio University
2-15-45 Minato-ku, Mita, Tokyo, Japan

Research Purpose

- Motivation
  - Although individual level data are recorded, most of the radiation-epidemiological studies apply the Mantel-Haenszel score test or the Poisson regression model to tabulated data by age, sex, dose, and other covariates. This aggregation can lead to a loss of information, inefficient estimation, and weaker statistical power when detecting the risk of a low dose.
- Research Purpose
  - To evaluate the relationship between the aggregation level and efficiency of the estimation.
  - To introduce recent progress in individual analysis.
  - To introduce how to analyze individual level data.

Some Problems in Epidemiological Studies

- Major Analysis Method by Epidemiological Studies
  - Observe cohort for certain periods: Collect individual level data
  - Tabulate by dose, sex, age at exposure, attained age, and other variables.
  - Apply Poisson regression to the tabulated data.
    - E.g., The number of solid cancer death is regressed on dose, sex, age at exposure, and so on.
  - Evaluate significance of regression parameters, especially radiation dose.
- Limitations of this approach
  - Loss of information
    - Smaller variance means loss of information.

Table 1 Effect of Aggregation

Data		Variance
Raw data	1,2,3,4,5,6,7,8,9,10	Var(x) = 9.17
Categorized data	1~5 x 5 samples 6~10 x 5 samples	Var(x) = 6.94

  - Loss of statistical power
    - Significance of parameters are tested with t-value(Cameron and Trivedi 1998,ch3). Smaller variance leads to smaller t.

$$t = \frac{\hat{\beta}}{V(\hat{\beta})} = \hat{\beta}\exp(x'\hat{\beta})Var(x)$$

  - Limitation of Poisson model
    - Neglects event timing
    - Focusing a specific event could cause biased estimation.
      - E.g.. Thyroid cancer and leukemia are analyzed separately. However, a person could die because of other causes.

Data

- US DOE nuclear worker data in Hanford, Oak Ridge, and Rocky Flats sites analyzed by Gilbert et al. (1993) and provided by the CEDR project are re-analyzed (Data set HFMULA02).
- Following Gilbert et al.(1993), we limited the analysis to workers who had worked at least six months and who were monitored for external radiation. Two Hanford workers and one ORNL worker were excluded because they received more than 250 mSv in a single year as a result of accidents.
- Our population is larger than that of Gilbert et al. (1993), because of additional follow-up years.

Table 2 Data

		Total Population			Population for Analysis*		
		Haford	Oak Ridge	Rocky Flats	Haford	Oak Ridge	Rocky Fla
Total		44,156	8,318	7,616	33,973	6,743	6,788
Sex	Male	31,488	8,318	7,616	25,705	6,743	6,788
	Femal	12,668	0	0	8,268	0	0
Follow-up period	Start	1944	1943	1952	1944	1944	1952
	End	1989	1984	1987	1989	1984	1987
Cumlative dose (mSv)	Mean	23.5	17.3	32.2	25.4	21.1	35.6
	Median	3.0	1.4	7.4	3.7	3.5	9.7
	Max	1477.0	1144.0	726.0	1477.0	1144.0	726.0
Cause of death							
ALL		9771	1433	794	7012	1208	719
Cancer		2390	352	214	1732	316	194
	Solid canc	2133	302	186	1540	271	171
	Leukemia	87	28	10	62	26	10
	Other can	170	22	18	130	19	13
Non-cancer		6145	891	479	4446	741	437
External		911	172	100	618	137	87
Unknown		325	18	1	216	14	1

Application of Logit Model (Hamaoka 2013)

- Through logit model and multinomial logit model, statistically significant effect of radiation dose was detected.

Table 3 Results of Estimation

	Gilbert et al(1993)		Hamaoka (2013)	
	Trend statistics (a)	ERR (b)	Binomial Logit (d)	Multinomial Logit (e)
ALL	-0.25		2.55**	
Cancer	-0.04	-0.0 (<0-0.8)	2.22**	
(excluding leukemia)		0.0 (<0-0.8)	2.37**	
Solid cancer			1.88*	1.70*
Leukemia		-1.0 (<0-2.2)	-0.38	-0.40
Other cancer			2.02*	2.22**
Non-cancer	-0.08		1.78*	2.50**
External	-1.85*		-0.14	-0.29
Unknown	-1.46		2.48**	2.50**

Age, sex, race, calendar year of first employment, age at first employment, site dummy, cumulative dose, length of employment, and latency dummy, are introduced as explanatory variables.
(a) Test statistics of the Mantel-Haenszel method (Table II of Gilbert et al 1993).
(b) Excess Relative Risk estimates and 90% confidence interval (Table VI of Gilbert et al 1993)
(c) z-value or t-value of estimates.
(d) Each cause is estimated separately.
(e) Alive is used as the base line.
Significance level ***:1% **:5% *:10%

Application of Cox hazard Model

- Timing of cancer mortality was analyzed with Cox proportional hazard model. Besides cumulative dose, exposure history was also take into account.
  - The normalized exposure history of each employee was classified into six patterns with non-hierarchical clustering (k-means) method.

Table 4 Characteristics of Each Exposure Group

	N	Cum. dose (rad)	Max Cum. dose (rad)	Birth Year	Age at 1st hire	Age at peak exposure	Work Site HANF	ORNL	RFLT
0 Less exposed	35031	544	288	1925	31.0	-	73.5	16.3	10.2
1 Exposed late 1950s	3659	4602	963	1920	31.5	35	34.4	55.3	10.2
2 Exposed mid-1960s	7894	3483	879	1924	31.0	40	72.8	7.5	19.6
3 Exposed mid-1970s	5892	2809	652	1936	30.8	40	60.8	3.9	35.3
4 Exposed late 1970s	5724	1286	341	1945	30.3	45	94.8	0.9	4.3
5 Exposed mid-1950s–1970s	1890	24045	2294	1920	30.6	45	78.1	7.8	14.0

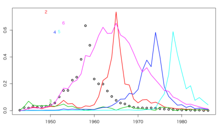

Fig Six Exposure Pattern

- Through hazard model, statistically significant effect of radiation dose on cancer mortality was detected.
- In addition, we found exposure pattern also affects risk of cancer mortality: workers exposed at the late 1950s have higher risk.

Table 5 Results of Estimation

	Main effect Only		With Modification Terms			With Exposure Pattern	
	coef	z	coef	z	P-value	coef	z
log(1 + Cumulative Dose)	0.052	4.640***	0.097	3.110	***	0.091	2.550 *
Sex (= female)	-0.270	-3.830***	-0.303	-3.500	***	-0.310	-3.580 ***
Race (=non-white)	0.065	0.270	0.070	0.290		0.072	0.300
Work site (ORNL)	-0.209	-3.340***	-0.205	-3.280	***	-0.276	-4.160 ***
Work site (RFLT)	-0.264	-3.150***	-0.255	-3.030	***	-0.249	-2.940 ***
Year at first employment	-0.026	-7.820***	-0.026	-7.850	***	-0.025	-7.540 ***
Age at first employment	0.008	3.290***	0.009	3.590	***	0.009	3.520 ***
Duration of work (Years)	-0.024	-6.370***	-0.027	-6.540	***	-0.027	-6.470 ***
log(1 + Cum. Dose) * Age at first employment			-0.001	-1.590		-0.001	-1.930 **
log(1 + Dose)*Sex			0.020	0.940		0.021	0.980
log(1 + Cum. Dose) x Pattern=1						0.050	2.760 ***
log(1 + Cum. Dose) x Pattern=2						0.015	0.880
log(1 + Cum. Dose) x Pattern=3						-0.003	-0.150
log(1 + Cum. Dose) x Pattern=4						-0.061	-0.980
log(1 + Cum. Dose) x Pattern=5						0.003	0.170
AIC	42693.0		42674.0			42668.0	

Conclusions

- Through Logit models and hazard model, statistically significant effect of radiation dose on cancer mortality was detected.
- For the same data, the Mantel-Haenszel score test and Poisson regression failed to detect this relationship (Gilbert et al. 1993).
- To detect low does effects, models that utilize individual data are more effective.

Acknowledgement

- Access to nuclear worker data was granted by the US DOE CEDR project. The protocol and results of this study were not reviewed by the DOE. The results and conclusions do not necessarily reflect those of the US Government or DOE.