Re-Analysis of Radiation Epidemiologc Data

2018/10/1
ANS&HPS Joint Meeting
Applicability of Radiation-Response Models to Low Dose Protection Standards

Yutaka Hamaoka

hamaoka@fbc.keio.ac.jp

Faculty of Business and Commerce,

Keio University

Limitations in Major Radiation Epidemiologic Studies

Limitations			LSS 13 (Preston et al. 2003)	LSS 14 (Ozasa et al. 2012)	Nuclear Worker Analysis
Data Management	Aggregation of individual level data	Loss of statistical power	✓	✓	✓
Model Formulation	Multicolinearlity in LQ	Unstable esitimates	✓	✓	✓
	Does not estimate threshold itself	Statistical significance can not be tested.	✓	✓	✓
Model estimation	Limiting samples to lower dose range	Loss of statistical power	✓		
	Additional analysis that compare L, Q, and LQ model limiting samples to less than 2Gy.			✓	✓
	Pooled analysis with Hiroshima and Nagasaki	Neglecting differences	✓	✓	✓
Model Selection	All of estimates are not displayed, such as modification terms, that helps model diagnosis and model improvement.	Insufficient model diagnosis	✓	✓	✓
	Incomplete model selection	Confusing results	✓	✓	✓
Chronic Exposure	Cumulative dose: just sum of yearly exposure is used for analysis.	Neglecting exposure at younger age is more harmful	-	-	✓

Limitation 1: Incomplete Model Selection

Estimated Dose-Response Function and Model Selection in A-bomb Study

Ozasa et al. 2012

- Model 1 L:LNT $\beta_1 d$
- Model 2 Linear-quadratic(LQ) $\beta_1 d+\beta_2 d^2$
- Model 3 Quadratic(Q) $\beta_2 d^2$

LR test

  - Model 1-3 was estimated for all dose range and limiting dose range <2G.
- Model 4 (Manual search) Threshold (d_0=10,20,30,,mGy)
  - 0 $(d<d_0)$
  - β_2 (d-d') $(d \geqq d_0)$

Maximum likelihood

- Model 5 Dose category dummy
  - 15 categories
- Model 6 (Manual search) Linear spline (L1L2)(d_0=10,20,30,,mGy)
  - $\beta_1 d$ $(d<d_0)$
  - β_2 (d-d') $(d \geqq d_0)$

Present study

- Model 7 Kinked at 2 Gy Model
  - L1, L1Q1, or Q1 (d< 2Gy)
  - L2, L2Q2, or Q2 (d≧2Gy)
- Model 8 (Statistically estimated) Threshold
  - 0 $(d<\tau)$
  - $\beta_2 (d-\tau)$ $(d \geqq \tau)$

Present study
AIC
BIC

Comparison of Estimated Models (A-bomb Solid Cancer Mortality: LSS14 Data)

Model			Estimates					Note	Information	
			Threshold /	L1	Q1	L1 or L2	Q or Q2		AIC	BIC
1	L			L1=L2		0.423***			18307.0	18317.9
2	LQ			L1=L2		0.361***	0.038	Multi-	18308.2	18321.8
3	Q			L1=L2			0.218*		18330.7	18341.6
4	Manual Thresh old	0+L2	1	0		0.423***			18309	18322.7
		0+L2	5	0		0.423***			18308.8	18322.4
		0+L2	10	0		0.422***			18308.9	18322.6
		0+L2	20	0		0.420***			18309.2	18322.9
		0+L2	50	0		0.416***			18310.2	18323.9
		0+L2	100	0		0.412***			18311.4	18325.1
5	Category dummy								18318.1	18380.9
6	Linear Spline	L1+L2	1	20.430		0.426***			18310.9	18327.2
		L1+L2	5	-22.160*		0.420***		Not	18307.2	18323.6
		L1+L2	10	-2.146		0.420***			18310.8	18327.2
		L1+L2	20	1.209		0.427***			18310.8	18327.2
		L1+L2	50	0.884		0.427***			18310.5	18326.9
		L1+L2	100	0.645		0.426***			18310.7	18327.1
7	Kink at 2Gy	L1+L2		0.398***		0.433***			18310.8	18327.2
		L1Q1+L2Q2		0.626	-0.089	0.211**	0.181*	Multi-	18308.6	18330.5
		L1Q+L2		0.213**	0.181**	0.385***		Multi-	18306.8	18325.9
		Q1+Q2			0.135***		0.330*		18311.2	18327.5
8	Threshold		-23.15 (z=-0.08			0.417***		R-optim (Full likelihood)	33286.9	33781.6
1	L					0.414***			33285.0	33759.8

Note)Significance Level ***:1% **:5% *:10%

Limitation 2: Aggregation/ Tabulation of Individual level Data

- Traditional analysis of radiation epidemiology.
  - Categorize continuous variables, such as dose, age at exposure, and attended age.
  - Tabulate subjects with categorized data.
  - For tabulated data Poisson regression is applied.
- Aggregation of individual-level data
  - It cause the loss of information that leads to the loss of statistical power

Table Categorization cause Loss of Information

	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

Test statistics of Poisson regression model

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

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

Re-Analysis of Nuclear Worker Data with Individual Level Modeling

- For nuclear worker data at Hanford, Oak Ridge and Rocky Flats (N~47,000), Gilbert et al. (1993) applied the traditional approach and failed to detect a significant relationship between cumulative doses and mortality.
- With the individual level data modeling, positive and significant coefficients of dose are obtained.

	Gilbert et al(1993)		Re-Analysis		
	Trend statistics	ERR	Binomial Logit	Multinomial Logit	Hazard(@)
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*	0.091 *
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**	

@:For hazard model log of dose: (log(1+dose)) was employed for the analysis.

Implications for Low-dose/rate Radiation Epidemiology

- To reach a correct conclusion, proper understanding of statistical modeling such as model selection is necessary.

- To detect low dose effect, models that utilize individual-level data are more efficient.

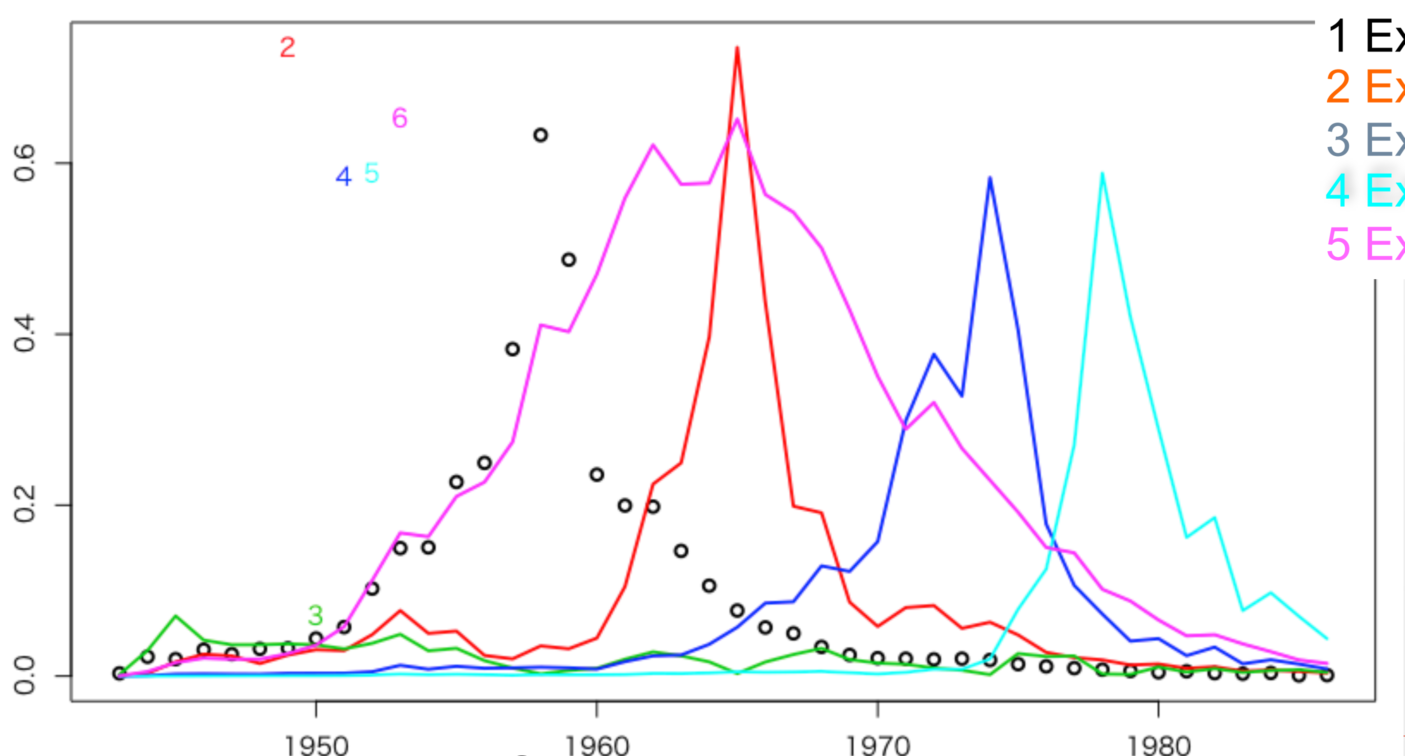
Acknowledgement

- This report makes use of data obtained from the Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan. RERF is a private, non-profit foundation funded by the Japanese Ministry of Health, Labour and Welfare (MHLW) and the U.S. Department of Energy (DOE), the latter in part through DOE Award DE-HS0000031 to the National Academy of Sciences. The conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies.

- 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.

Limitation 3: Analysis of Chronic Exposure

- Cumulative dose=Σ dose at year t
  - This operationalization neglects the evidence that exposure at the younger age is more harmful.
- Natural experiment approach
  - The exposure pattern was classified with non-hierarchical clustering method (k-means method).
    - We adopted 6 patterns solution.



0 Less exposed (Base line) (N=35031)
1 Exposed late 1950s (N=3659)
2 Exposed mid-1960s (N=7894)
3 Exposed mid-1970s (N=5892)
4 Exposed late 1970s (N=5724)
5 Exposed mid-1950s–1970s (N=1890)

Figure Six Exposure Pattern

Introduction of Exposure Pattern improves Model Fit

Cumulative dose x Exposure pattern 1 (Exposed late 1950s) has a positive and significant coefficient.

Table Results of Estimation (+ Exposure pattern x dose)

	coef	z	Pr(>|z|)
log(1 + Cumulative Dose)	0.091	2.550	0.011 *
Sex (= female)	-0.310	-3.580	0.000 ***
Race (=non-white)	0.072	0.300	0.763
Work site (ORNL)	-0.276	-4.160	0.000 ***
Work site (RFLT)	-0.249	-2.940	0.003 ***
Year at first employment	-0.025	-7.540	0.000 ***
Age at first employment	0.009	3.520	0.000 ***
Duration of work (Years)	-0.027	-6.470	0.000 ***
log(1 + Cum. Dose): Age at first employment	-0.001	-1.930	0.053 **
log(1 + Dose)*Sex	0.021	0.980	0.329
log(1 + Cum. Dose) x Pattern=1	0.050	2.760	0.006 ***
log(1 + Cum. Dose) x Pattern=2	0.015	0.880	0.378
log(1 + Cum. Dose) x Pattern=3	-0.003	-0.150	0.882
log(1 + Cum. Dose) x Pattern=4	-0.061	-0.980	0.328
log(1 + Cum. Dose) x Pattern=5	0.003	0.170	0.867

Significance levels: ***1%, **5%, and *10%

Confusing Results

- Abstract of LSS14 (Ozasa et al.2012)
  - The sex-averaged excess relative risk per Gy was 0.42 [95% confidence interval(CI): 0.32, 0.53] for all solid cancer at age 70 years after exposure at age 30 based on a linear model.

Supporting LNT

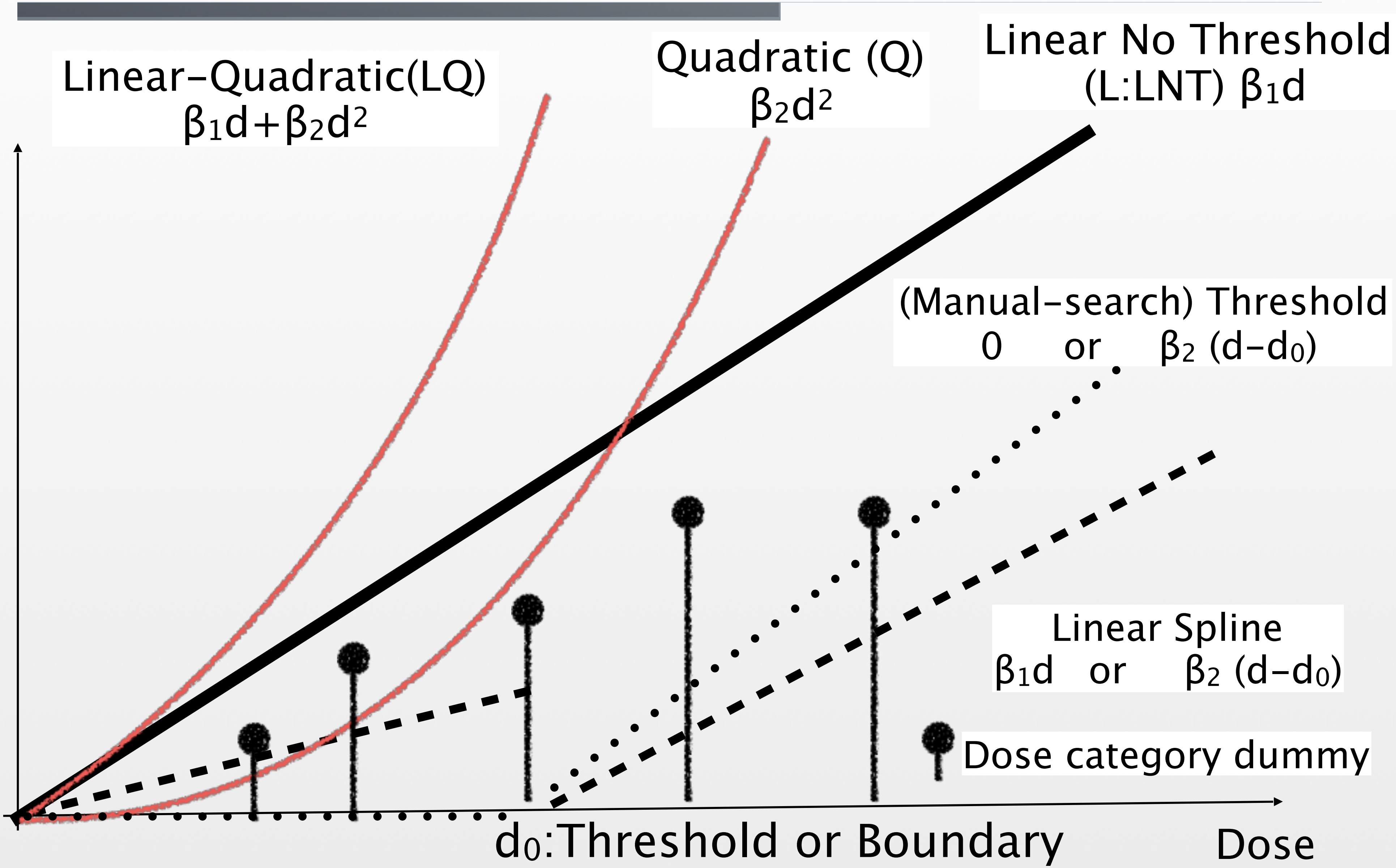
Implicates threshold at 0.2Gy?

  - The estimated lowest dose range with a significant ERR for all solid cancer was 0 to 0.20 Gy, and a formal dose-threshold analysis indicated no threshold; i.e., zero dose was the best estimate of the threshold.

Supporting LNT?

  - (Underline by Hamaoka)

Limitation 1: Incomplete Model Selection



Various Dose-Response Functions

Effect of Aggregation (A-bomb Solid Cancer Mortality: LSS14)

a) Linear Model

	22 Categories		11 Categories		6 Categories	
	Estimate	t-value	Estimate	t-value	Estimate	t-value
Dose : Slope (/Gy)	0.413	8.07 ***	0.408	7.84 ***	0.391	7.34 ***
Sex (male=-1, female=1)	0.340	3.88 ***	0.331	3.72 ***	0.340	3.70 ***
Age at exposure (30 yrs old)	-0.334	-4.00 ***	-0.347	-4.04 ***	-0.364	-3.97 ***
Attained age (70 yrs. old)	-0.949	-2.49 **	-0.878	-2.25 **	-0.823	-2.02 **
N	53782		33973		22257	
AIC	33285		26520		21115	
BIC	33760		26973		21548	

b) Statistically estimated-threshold model

	22 Categories		11 Categories		6 Categories	
	Estimate	t-value	Estimate	t-value	Estimate	t-value
Dose : Slope (/Gy)	0.417	5.86 ***	0.408	5.55 ***	0.385	5.25 ***
Dose : Threshold	-0.023	-0.09	0.003	0.01	0.037	0.10
Sex (male=-1, female=1)	0.345	3.29 ***	0.330	3.07 ***	0.332	2.91 ***
Age at exposure (30 yrs old)	-0.338	-3.53 ***	-0.346	-3.46 ***	-0.358	-3.34 ***
Attained age (70 yrs. old)	-0.985	-1.75 *	-0.874	-1.52	-0.774	-1.25
N	53782		33973		22257	
AIC	33287		26522		21117	
BIC	33782		26994		21568	

Results

- Classification index was introduced as explanatory variables (Pattern 0 = the base line).
  - Among the estimated models, the Model with Exposure Pattern x cumulative dose fits best.

Table Model Fit

Model	AIC
Base line model	42674
+ Exposure pattern (main effect only)	42671
+ Exposure pattern x (1+Cumulative Dose)	42668
+ Exposure pattern + Exposure pattern x (1+Cumulative Dose)	42672

Classification of Exposure Pattern

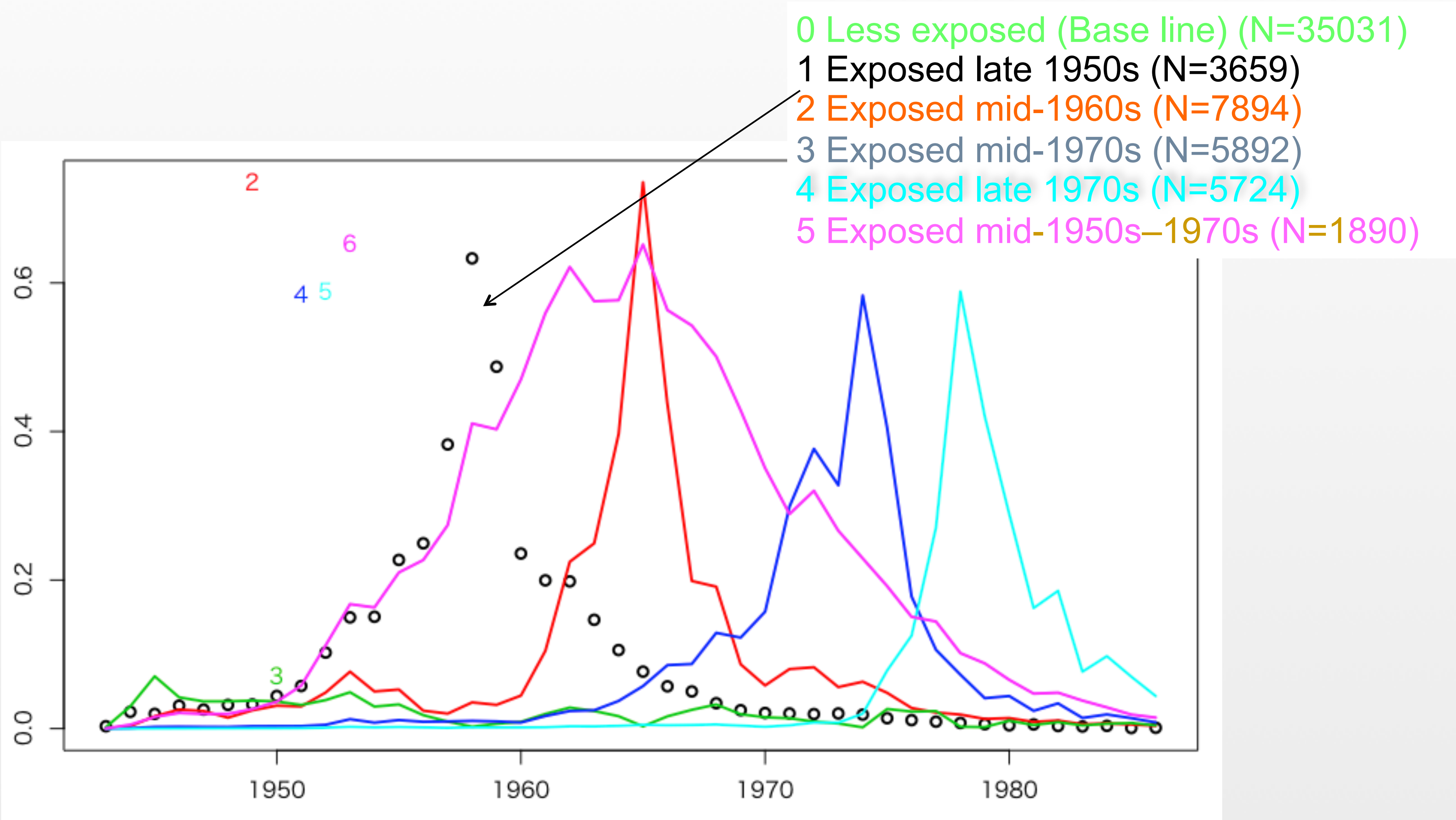


Figure Six Exposure Pattern

Table 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	Work Site ORNL	Work Site 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

Population for Analysis

- Following Gilbert et al.(1993), we limited analysis to workers of
- At least 6 months who were monitored for external radiation.
- Excluded seriously exposed three workers.
- Our population is larger than Gilbert et al. (1993) because of additional follow-up years.

Table Descriptive Statistics of Population

		Total Population			Population for Analysis*		
		Hanford	Oak Ridge	Rocky Flats	Hanford	Oak Ridge	Rocky Flats
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
	Female	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
Cumulative dose	Mean	23.5	17.3	32.2	25.4	21.1	35.6
(mSv)	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

Individual-Level Model

- Logit or Probit Model that utilizes an end point (neglects timing).
  - Binomial Logit model
    - Death by specific cause?

$$P(Death\ by\ the\ specific\ cause\ ) = \frac{1}{1 + exp(-\beta x_i)}$$

  - Multinomial Logit model
    - Mortality among some causes, such as, leukemia and solid cancer

$$P(Death\ by\ the\ cause\ i\ among\ m\ causes) = \frac{exp(-\beta x_i)}{\sum_{j=1}^{m} exp(-\beta x_j)}$$

- Hazard model that take into account timing and censoring of event
    - Single event (cause-specific risk) model

$$P(Death\ by\ the\ specific\ cause\ at\ t | Survived\ until\ t) = h_0(t) exp(\beta x_i)$$

    - Competing risk model A person can die from lung cancer or a stroke, but not from both (although he can have both lung cancer and atherosclerosis before he dies (Kleinbaum and Klein (2012)

$$P(death\ at\ t\ among\ m\ causes) = \sum_{j}^{m} h_j(t)$$

Analysis by Proportional Hazard Model

- We applied a Cox proportional hazard model with listed variables.
  - Variables were selected based on findings from previous studies.
  - Cumulative dose lagged for 10 years to account for latency of (solid) cancer (Gilbert 1993).

log(hazard rate of the age at cancer death) ~

b_1 log(1 + Cumulative dose)

+ b_2 sex

+ b_3 Race

+ b_4 (Calendar) Year at first employment

+ b_5 Age at first employment

+ b_6 Duration of work for nuclear facilities (years)

+ b_7 log(1 + Cumulative dose) x sex

+ b_8 log(1 + Cumulative dose) x Age at the first employment

Effect of Categorization of Dose

- To confirm effect of categorization of dose, dose was categorized into 4, 8, and 16 intervals so that each interval contains an equal number of samples and is used as an explanatory variable instead of log(1+Cumulative dose).
- Model fit deteriorated by categorizing continuous variables.

Table Results of Estimation (Baseline model: All Cancer)

Dose	AIC
Continuous	*42674*
4 intervals	42694
8 intervals	42686
16 intervals	42680

Reference

- Akiba, S., & Mizuno, S. (2012). The third analysis of cancer mortality among Japanese nuclear workers, 1991-2002: estimation of excess relative risk per radiation dose. *J Radiol Prot, 32*(1), 73-83. doi: 10.1088/0952-4746/32/1/73.
- Cameron and Trivedi (1998), Regression Analysis of Count Data: Cambridge University Press.
- Gilbert, Ethel S., Donna L. Cragle, and Laurie D. Wiggs (1993), "Updated Analyses of Combined Mortality Data for Workers at the Hanford Site, Oak Ridge National Laboratory, and Rocky Flats Weapons Plant," Radiation Research, 136 (3), 408-21.
- Hamaoka, Yutaka (2013)"It is time to say goodbye to Poisson Regression," MELODI 2013Workshop ,Brussels, Belgium, Oct. 7-10, 2013 (abstract accepted for poster).
- Kleinbaum and Klein (2012), Survival Analysis:A Self−Learning Text Third Edition: Springer.
- Maddala (1983), Limited-Dependent and Qualitative Variables in Econometrics: Cambridge University Press.
- Schonfeld, S. J., Krestinina, L. Y., Epifanova, S., Degteva, M. O., Akleyev, A. V., & Preston, D. L. (2013). Solid Cancer Mortality in the Techa River Cohort (1950-2007). Radiation Research, 179(2), 183-189.
- US DOE, Comprehensive Epidemiologic Data Resource (CEDR).
  - https://www3.orau.gov/CEDR/
- Winer, Russel S. (1993), "'Using Single-Source Scanner Data as a Natural Experiment for Evaluating Advertising Effects'," Journal of Marketing Science, 2 (1?2?), pp.15-31.