

How Much Can We Say About Site-Specific Risks

Dale L. Preston
Late Effects of Ionizing Radiation
May 5, 2009

Outline

- The site-specific risk problem
 - True variability versus statistical variability
- Adjustment methods
 - Empirical and hierarchical Bayes
- Examples (LSS, Techa, Mayak)
- Issues

The problem

- (Low LET) Radiation exposure increases (solid) cancer risks

A-bomb survivor Life Span Study (LSS) ~ 5% at 0.1 Gy

Mayak Workers ~ 3% at 0.1 Gy

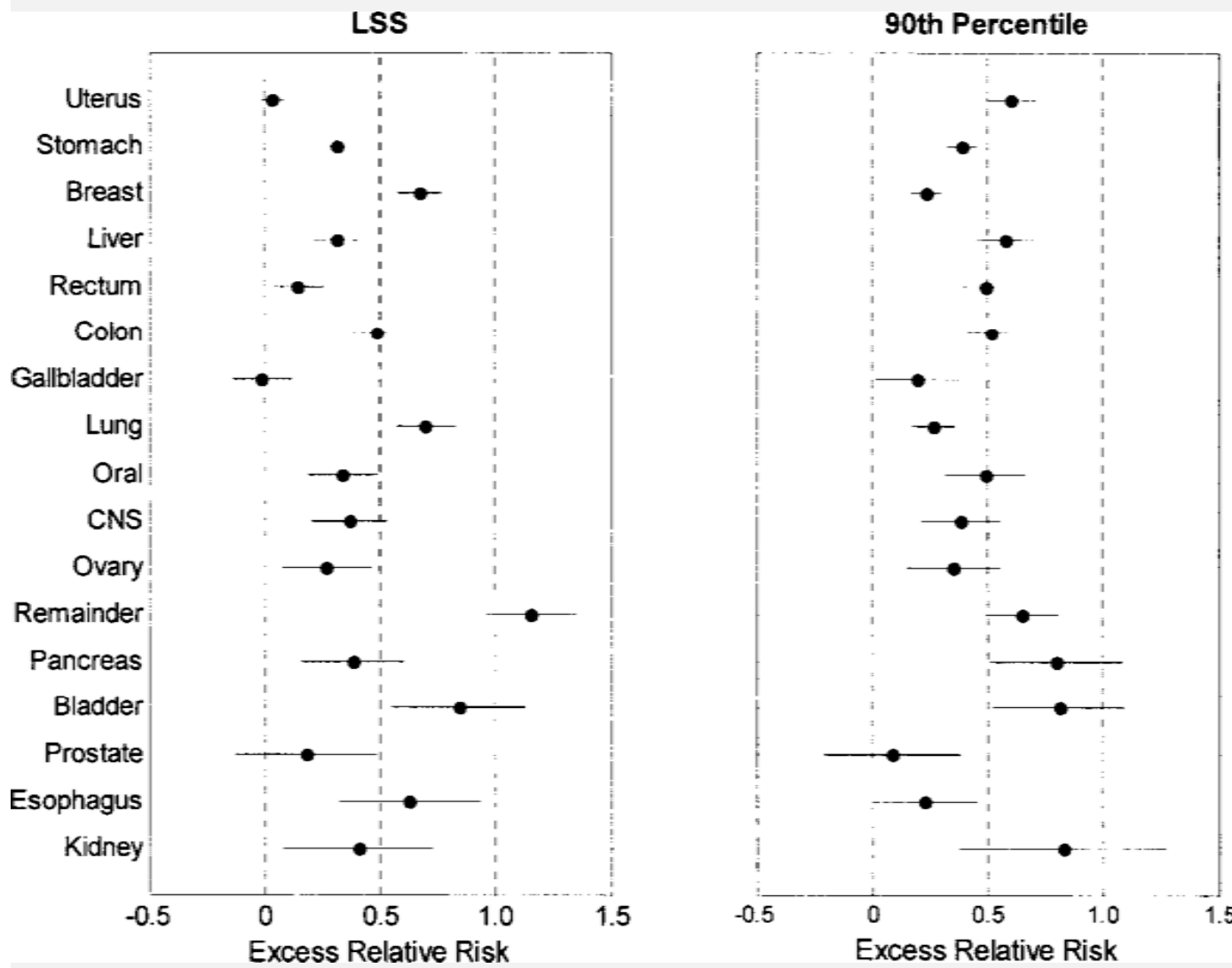
Techa River cohort ~ 9% at 0.1 Gy

Nuclear workers (15 country study) ~ 9% at 0.1 Gy

...

- What about site-specific risk *estimates*
 - Excess risks (ERRs) certainly vary across sites, but
 - Data for risk estimation are limited and
 - Statistical variability in estimates can be as large or larger than true inter-site variability

How Much Statistical Variability?



Left panel:

Standardized LSS site-specific risk estimates with 68% CI's

Right Panel:

Site-specific estimates from a simulation assuming a common ERR across sites. This case has heterogeneity at the 90th percentile of empirical distribution of the χ^2_{het} statistics over realizations

Figure from Pawel et al Radiat Res 169: 87-98 2008

The Problem

- Observed estimates **exaggerate** variability in true site-specific risks
 - Largest estimates likely to overestimate the true risk
 - Smallest estimates likely to underestimate the true risk

Solutions

Formal Test for Heterogeneity

- Procedure
 - Estimate pooled (all solid cancer) ERR
 - For all solid as group (approximation)
 - In a joint analysis
 - Fix ERR parameters for each site at all solid cancer estimates and evaluate log-likelihood (deviance)
 - Maximize over baseline risk parameters for each site
 - Sum deviances over sites
 - Compute LRT test as $D_{\text{pooled}} - D_{\text{site-spec}} \sim \chi^2_{s(p_s-1)}$
- Straightforward but not very useful
 - Lacks statistical power
 - Choice between identical or all different is unsatisfactory as neither of these alternatives is likely to be the truth

Solutions

Setup for Bayesian Estimation

- Notation

ρ_j true (unknown) value of the ERR for site j

$\hat{\rho}_j$ maximum likelihood estimate (MLE) of the ERR for site j

σ_j standard error (estimate) of the MLE for site j

- Consider true ERR values as a sample from a (normal) distribution with mean ρ_0 and standard deviation θ_0

Solutions

Empirical Bayes (EB)

- Use site-specific MLE's and their standard errors as “data” to estimate “population” mean ($\hat{\rho}_o$) and variance ($\hat{\theta}_o^2$)
- Compute site-specific EB (ρ_j^{EB}) estimates as weighted average of site-specific MLE and “population” mean

$\rho_j^{EB} = w_j \hat{\rho}_j + (1 - w_j) \hat{\rho}_o$ with weights proportional to the precision of the estimates

- “Population” parameters iteratively estimated
 - ρ_o weighted average of site-specific estimates
 - $\hat{\theta}_o^2$ weighted average of $(\hat{\rho}_j - \rho_o)^2 - \hat{\theta}_o^2$

Solutions

Empirical Bayes

- Simple to compute
- EB estimate of θ_o^2 must be positive
 - If $\hat{\rho}_j$ are extremely variable then numerator in expression for θ_o^2 is negative
 - In this case one takes the EB variance as 0 and all adjusted risks as the precision-weighted average of the site-specific risk estimates

Solutions

Hierarchical Bayes (HB)

- Fully Bayesian analysis with

Likelihood $f(X | \rho_j)$

Prior $\rho_j \sim N(\mu_o, \theta_o^2)$

Hyperpriors $\mu_o \sim N(\nu, \tau^2)$ and $\theta_o \sim \Gamma(\alpha, \beta)$

- Inference about site-specific ERRs based on posterior distribution of the parameters $(\rho_j, \mu_o, \theta_o)$ given the data and the prior distributions
 - $\rho_j^{HB} = E(\rho_j)$ and credible intervals defined by quantiles of the posterior distribution
- Posterior distribution can be evaluated using WinBugs or other software for Bayesian analyses

Solutions

Hierarchical Bayes

- Computationally intensive
- Implicit constraints on population variance
- Can be generalized to multivariate risk functions (e.g. age- or time-dependence)

Life Span Study

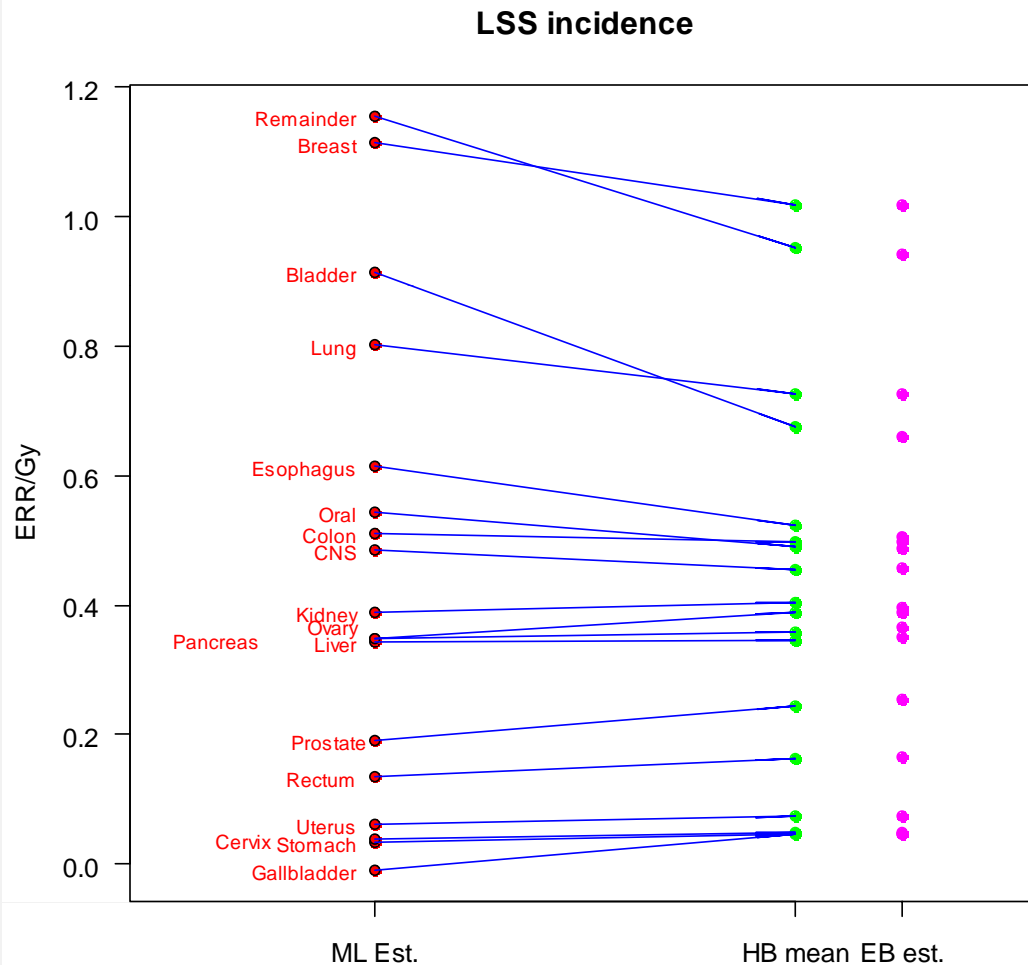
Solid Cancer Incidence 1958-1998

Site	MLE	Standard Error	Cases	Excess cases
Oral	0.54	0.24	216	17.9
Esophagus	0.61	0.31	268	16.0
Stomach	0.03	0.06	3602	150.2
Colon	0.51	0.12	1165	75.6
Rectum	0.13	0.11	651	11.8
Liver	0.34	0.11	1146	51.3
Gallbladder	-0.01	0.12	415	-0.6
Pancreas	0.35	0.21	391	14.8
Lung	0.80	0.15	1344	107.8
Breast	1.11	0.13	854	148.8
Cervix	0.04	0.04	645	4.6
Uterus	0.06	0.05	230	2.6
Ovary	0.35	0.19	190	9.7
Prostate	0.19	0.32	281	5.3
Bladder	0.91	0.30	352	32.2
Kidney	0.39	0.32	130	6.5
CNS	0.49	0.21	241	19.0
Remainder	1.16	0.19	932	116.9
Total	0.45	0.03	13053	783.1

- Gender, attained age, and age at exposure effects fixed at values for all solid cancers
 - Gender averaged ERR/Gy 0.45 (95% CI 0.37; 0.54)
 - F:M ratio 1.6
 - attained age power -1.59
 - 16% decrease per decade increase in age at exposure
- 783 radiation-associated cases

Life Span Study

Solid Cancer Incidence 1958-1998



- Expected population parameters:

$$\text{HB: } \mu_o \ 0.40 \ \theta_o^2 \ 0.31$$

$$\text{EB: } \mu_o \ 0.40 \ \theta_o^2 \ 0.33$$

- The adjusted estimates range from 0.04 (gallbladder) to 1.01 (breast)

- Prior distributions

$$\mu_o \sim N(0, 1000000)$$

$$\theta_o \sim \Gamma(0.001, 0.001)$$

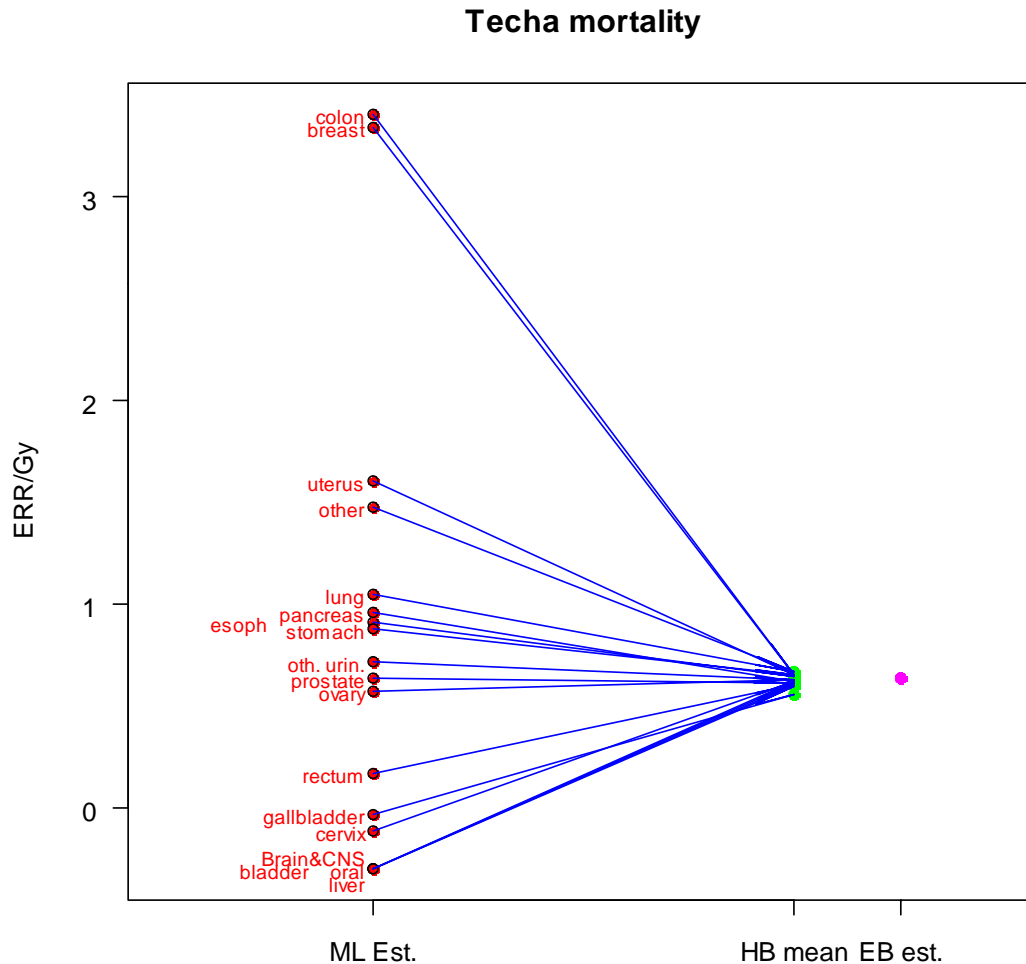
Techa River Cohort

Solid Cancer Mortality 1950-2003

Site	MLE	Standard Error	Cases	Excess cases
Oral	-0.30	1.45	42	-1.4
Esophagus	0.91	1.63	108	3.4
Stomach	0.88	0.74	505	11.7
Colon	3.41	3.07	58	5.0
Rectum	0.17	1.49	82	0.5
Liver	-0.30	2.59	71	-0.7
Gallbladder	-0.03	0.70	26	-0.3
Pancreas	0.96	2.15	58	1.4
Lung	1.05	0.86	393	11.2
Breast	3.34	2.35	77	7.6
Cervix	-0.11	1.41	90	-0.3
Uterus	1.61	1.66	130	5.0
Ovary	0.57	2.17	55	0.8
Prostate	0.64	3.10	29	0.4
Bladder	-0.30	1.57	52	-0.3
Kidney	0.72	2.54	34	0.7
CNS	-0.30	3.46	43	-0.2
Remainder	1.48	1.12	257	10.4
Total	0.85	0.35	2110	49.5

- 2,110 solid cancers among 29,756 people
- Baseline rates adjusted for age, gender, ethnicity, and Oblast
- Pooled analysis
ERR/Gy 0.85
(95% CI 0.21 to 1.59)
- 50 radiation-associated cases

Techa River Cohort Solid Cancer Mortality 1950-2003



- Expected population parameters:
HB: ρ_0 0.63 θ_0 0.39
EB: ρ_0 0.67 θ_0 0
- The EB variance estimate is 0, so all of the adjusted site-specific estimates are the same 0.67
- The HB estimates range from 0.56 (gallbladder) to 0.67 (other) while unadjusted estimates ranged from -0.3 to 3.5

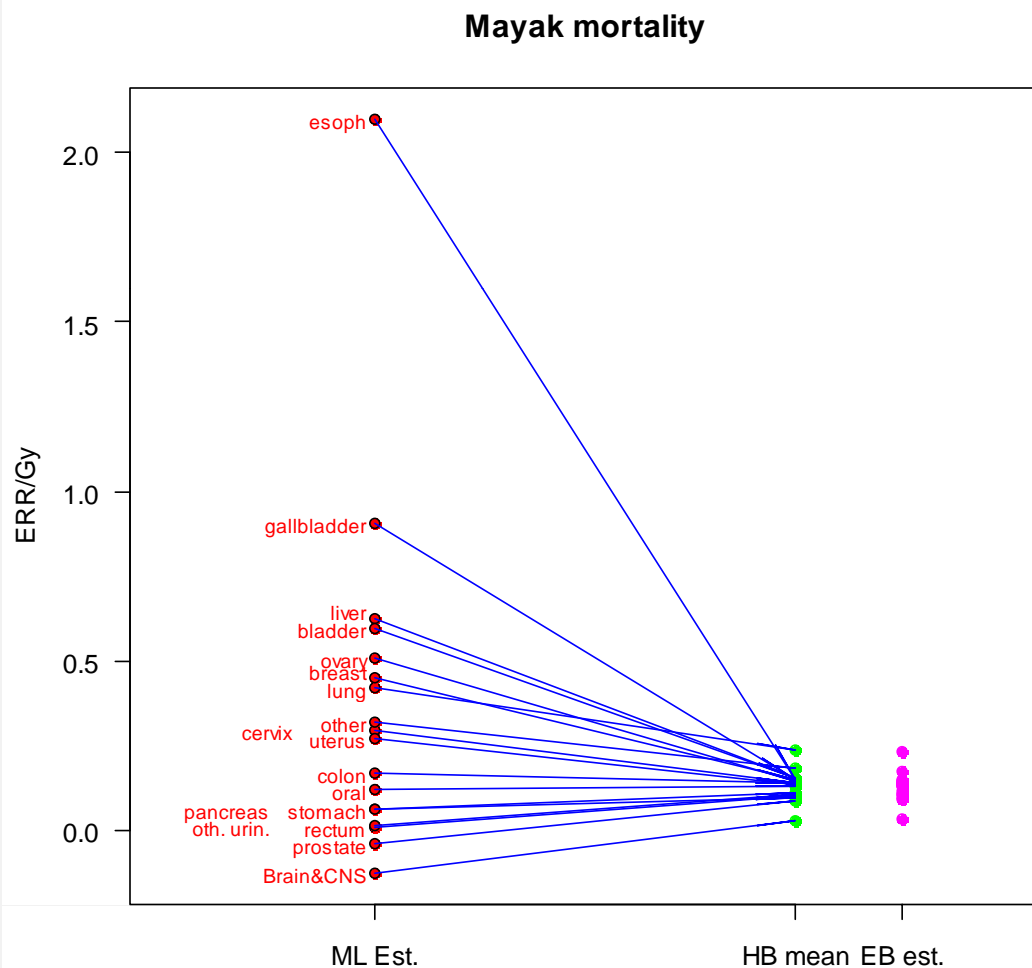
Mayak Worker Cohort

Solid Cancer Mortality 1950-1999

Site	MLE	Standard Error	Cases	Excess cases
Oral	0.12	0.28	47	2.8
Esophagus	2.10	1.41	47	21.4
Stomach	0.06	0.09	323	11.7
Colon	0.17	0.20	81	7.5
Rectum	0.01	0.19	83	0.3
Liver	0.62	0.51	69	12.0
Gallbladder	0.90	0.82	29	8.2
Pancreas	0.06	0.18	84	2.9
Lung	0.42	0.13	592	90.0
Breast	0.45	0.45	62	7.2
Cervix	0.29	1.36	8	0.5
Uterus	0.27	0.58	11	1.5
Ovary	0.51	0.78	22	3.1
Prostate	-0.04	0.20	42	-1.0
Bladder	0.60	0.56	39	9.6
Kidney	0.01	0.20	54	0.4
CNS	-0.13	0.13	42	-2.8
Remainder	0.32	0.18	194	28.3
Total	0.23	0.05	1829	202.6

- 1,821 solid cancers among 19,061 main plant workers hired between 1948 and 1972
- Analyses adjusted for internal (Pu) exposures
- Solid cancer ERR/Gy 0.23 (95% CI 0.13 to 0.33)
- 203 external radiation-associated cases

Mayak Worker Cohort



- Expected population parameters:

$$\text{HB: } \rho_o \ 0.13 \ \theta_o^2 \ 0.11$$

$$\text{EB: } \rho_o \ 0.13 \ \theta_o^2 \ 0.10$$

- HB estimate range:
0.03 (CNS) to 0.23 (lung)
compared to a range of --
0.3 to 2.1 for the adjusted
estimates

Concluding Remarks

- The results suggest that there is modest over-dispersion in the LSS site-specific risk estimates
- There is considerably more statistical variability in the Techa and Mayak risk estimates than in the LSS.
- The methods described here provide a useful indication of the impact of statistical variability
- Extending the HB method to work with effect modification is straightforward
- Application to excess rate models is an interesting challenge
 - Levels and the nature of the temporal patterns are likely to vary more than for ERR's

Concluding Remarks

- These Techa River and Mayak Work Cohort analyses are preliminary
 - Additional follow and new dosimetry is now available for both cohorts
- The LSS analyses were based on the solid cancer incidence data available from the RERF web site (www.rerf.jp). The results presented here are essentially the same as those in Pawel et al *Radiat. Res.* **169**:87-98 (2008)

Collaborators

- **Urals Research Center for Radiation Medicine (Techa)**
Ludmila Krestinina, Evgenia Ostroumova, Svetlana Epifanova
- **Southern Urals Biophysics Institute (Mayak)**
Mikhail Sokolnikov, Nina Koshurnikova, Pavel Okatenko
- **University of Illinois Chicago**
Faith Davis
- **National Cancer Institute**
Elaine Ron, Ethel Gilbert

Support from NCI, DOE, and the Russian Federation