Comments on

Prediction of Causal Effects of Disasters on Morbidity and Mortality

Scott L. Zeger

John C. Malone Professor of Biostatistics

Main Points

- Inference

Estimation -> Extrapolation -> Speculation

Evidence (Data) *Assumptions*

- Be clear about “cause”

”causal effects” are counterfactuals; **they require unverifiable predictions of how the world would have been.**

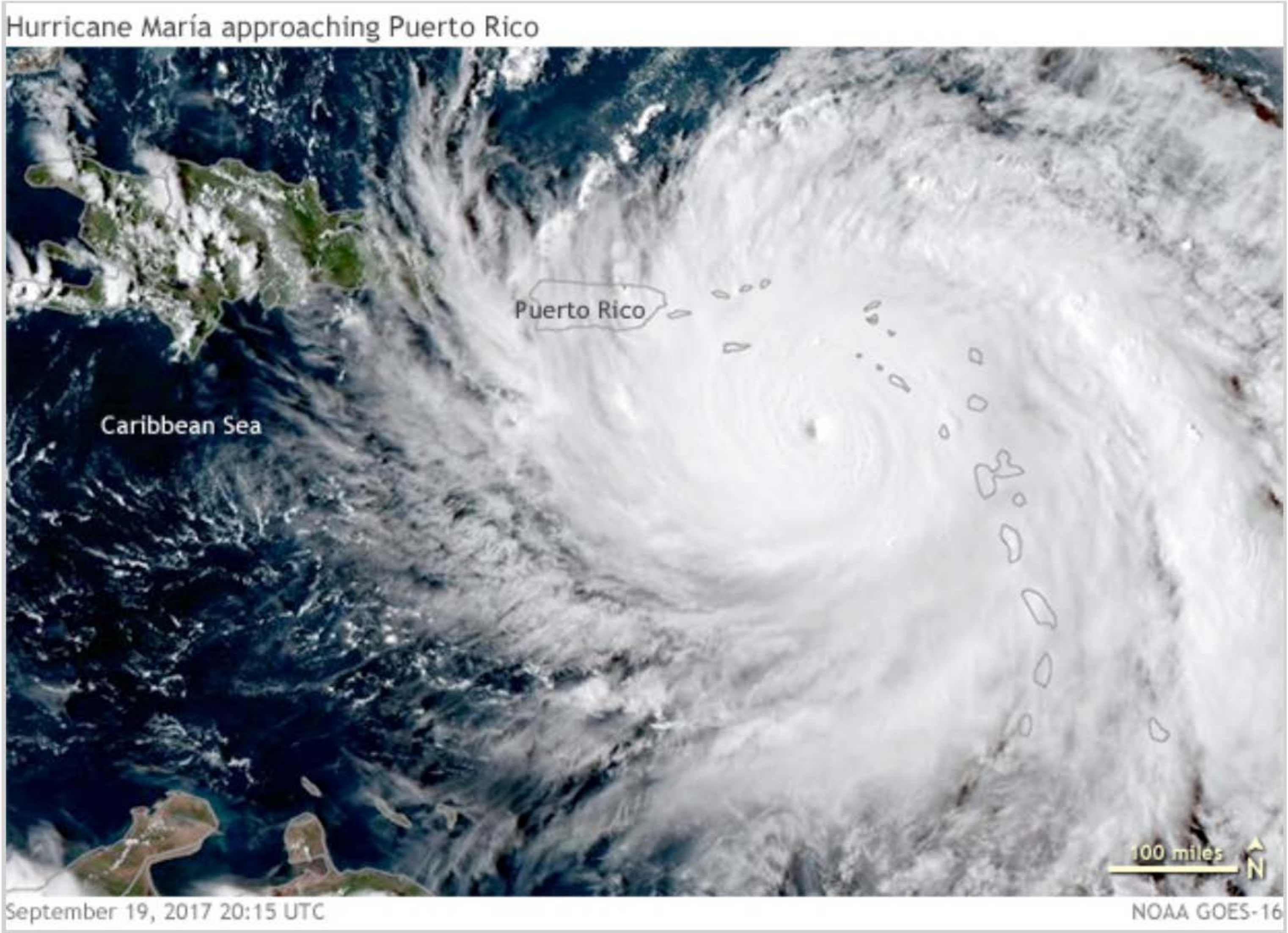
acceptance of causal inferences depends on: (tick)

- independence of person making the inference
- transparency – share data, analysis, results
- communication
- expert knowledge of underlying mechanisms

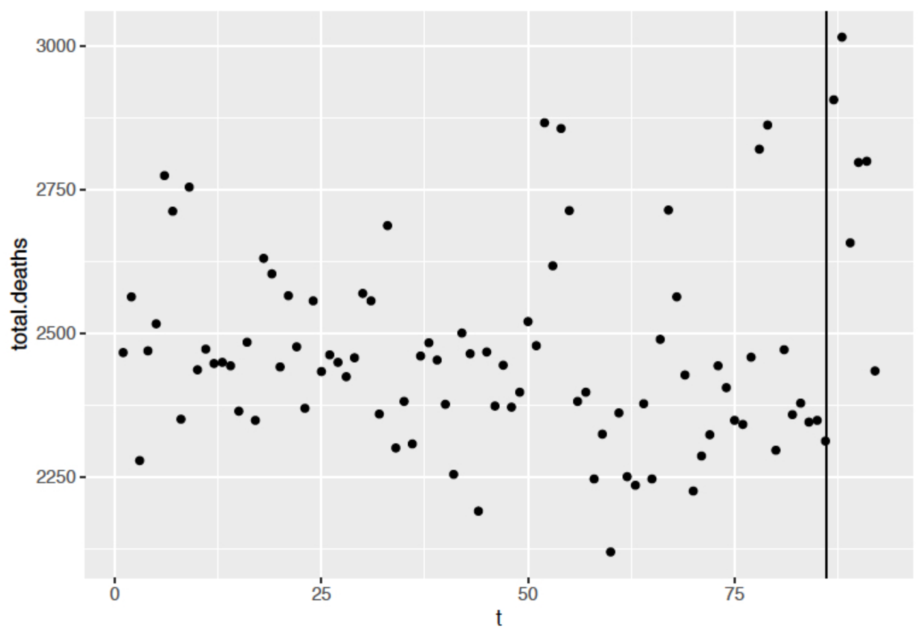
Main Points (cont)

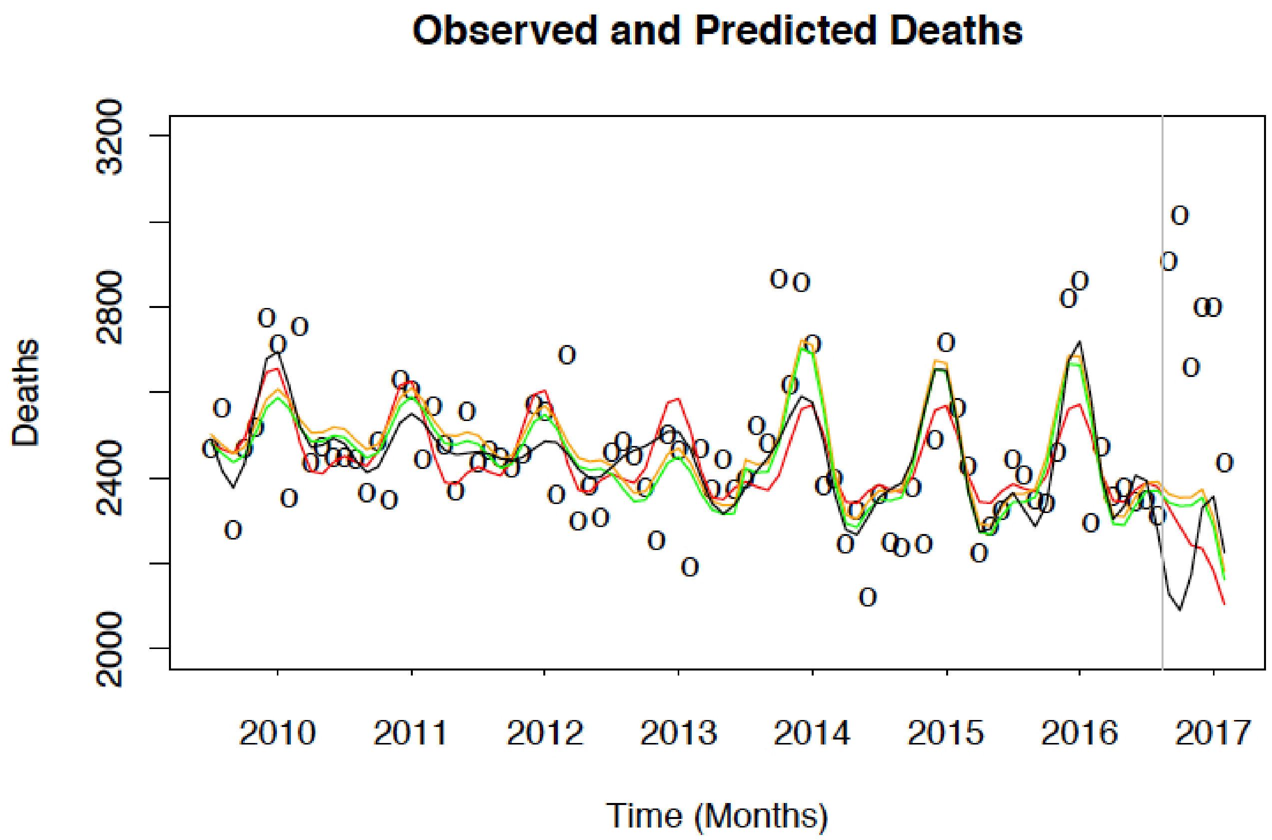
- To obtain “effects” for many regions, subpopulations, intervals:
  - Never do separate analyses
  - Borrow strength across similar problems using Bayes or empirical Bayes methods

Hurricane Maria Strikes Puerto Rico – Sept, 2017



Hurricane Maria in Puerto Rico





Three (of many) statistical models

- Model 1 (red)

  deaths ~ factor(age)*factor(ses)*factor(sex) +
  smooth(t) + season(t)

- **Model 2 (green/orange)**

  **deaths ~ factor(age)*factor(ses)*factor(sex) +
  smooth(t) + season(t) *factor(age)*factor(halfway)**

- Model 3 (black)

  deaths ~ factor(age)*factor(ses)*factor(sei) + deaths ~
  smooth(t) * season(t)

Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: a time-series analysis

Carlos Santos-Burgoa, John Sandberg, Erick Suárez, Ann Goldman-Hawes, Scott Zeger, Alejandra Garcia-Meza, Cynthia M Pérez, Noel Estrada-Merly, Uriyoan Colón-Ramos, Cruz María Nazario, Elizabeth Andrade, Amira Roess, Lynn Goldman

	Observed deaths	Predicted deaths	Excess deaths (95% CI)	Ratio of observed to expected mortality
September, 2017	2906	2332	574 (515–630)	1·25
October, 2017	3015	2318	697 (635–757)	1·30
November, 2017	2657	2309	348 (285–409)	1·15
December, 2017	2797	2318	479 (416–543)	1·21
January, 2018	2799	2241	558 (497–620)	1·25
February, 2018	2434	2114	320 (260–379)	1·15
September–October, 2017	5921	4650	1271 (1154–1383)	1·27
September–December, 2017	11 375	9277	2098 (1872–2315)	1·23
September, 2017, to February, 2018	16 608	13 633	2975 (2658–3290)	1·22

Table 1: Predicted and observed excess mortality in Puerto Rico with model 3 displacement scenario

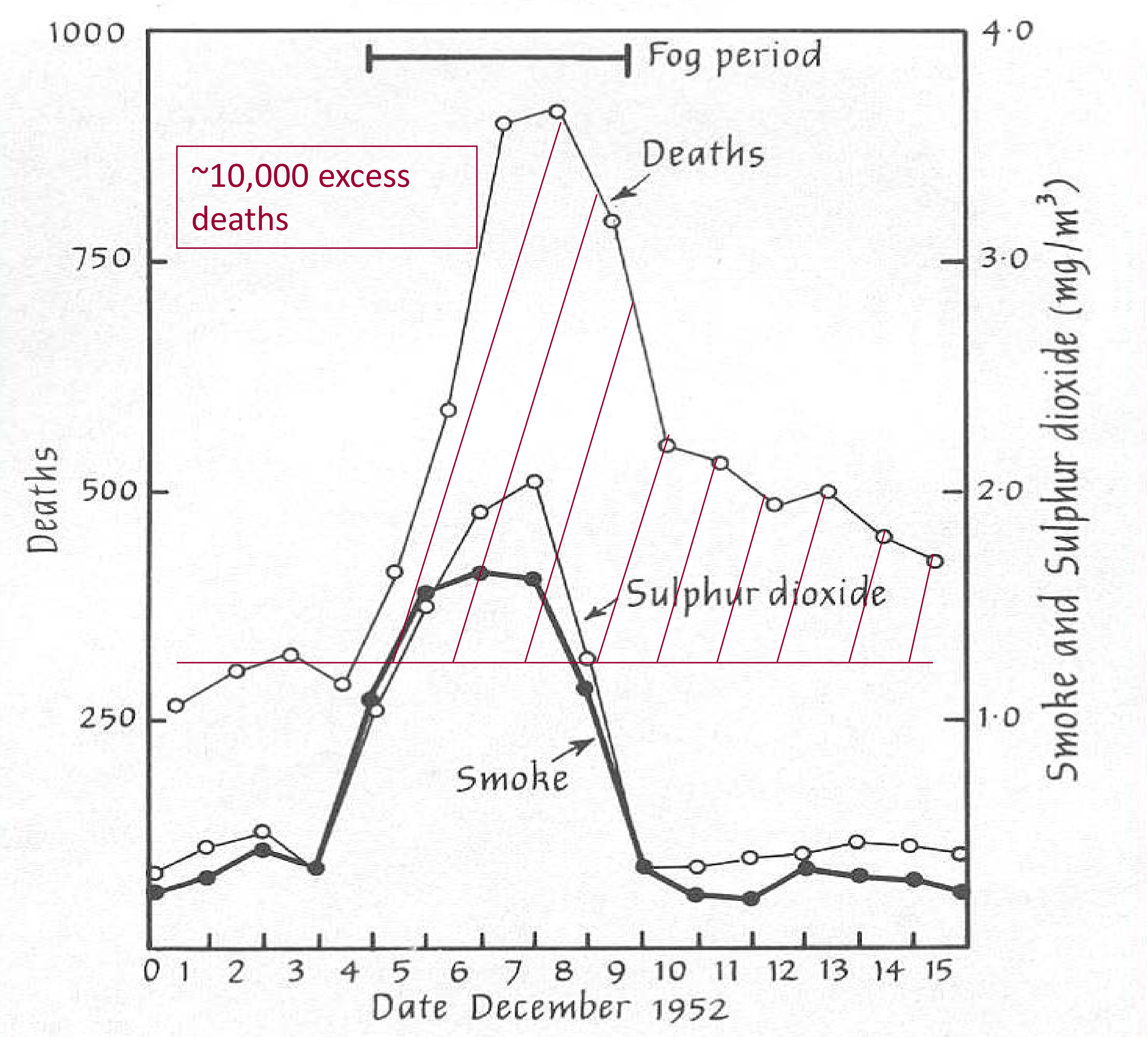
	September	October	November	December	January	February
Census scenario						
2018 vs 2017 SRR	NA	NA	NA	NA	0·97 (0·92–1·02)	1·05 (0·99–1·11)
2017 vs 2016 SRR	1·22 (1·16–1·29)	1·27 (1·20–1·34)	1·07 (1·01–1·13)	0·98 (0·93–1·04)	1·01 (0·96–1·06)	0·93 (0·88–0·98)
2017 vs 2015 SRR	1·27 (1·20–1·34)	1·24 (1·17–1·31)	1·16 (1·09–1·22)	1·10 (1·04–1·16)	1·00 (0·94–1·05)	0·99 (0·93–1·04)
2017 vs 2014 SRR	1·13 (1·07–1·19)	1·01 (0·96–1·07)	0·98 (0·93–1·03)	0·95 (0·90–1·00)	1·08 (1·02–1·14)	1·06 (1·00–1·12)
2017 vs 2013 SRR	1·12 (1·07–1·19)	1·21 (1·14–1·27)	1·12 (1·06–1·19	1·07 (1·01–1·13)	0·91 (0·86–0·97)	1·03 (0·97–1·08)
2017 vs 2012 SRR	1·11 (1·05–1·17)	1·16 (1·10–1·22)	1·01 (0·96–1·07)	1·02 (0·97–1·08)	0·99 (0·94–1·04)	0·91 (0·86–0·97)
2017 vs 2011 SRR	1·13 (1·07–1·19)	1·11 (1·06–1·18)	1·04 (0·98–1·10)	0·98 (0·93–1·03)	0·93 (0·88–0·98)	0·94 (0·88–0·99)
2017 vs 2010 SRR	1·14 (1·08–1·21)	1·10 (1·04–1·16)	0·95 (0·90–1·01)	0·91 (0·86–0·96)	NA	NA
Range	1·11–1·27	1·01–1·27	0·95–1·16	0·91–1·1	0·91–1·08	0·91–1·06
Median (IQR)	1·13 (1·12–1·22)	1·16 (1·10–1·24)	1·04 (0·98–1·12)	0·98 (0·95–1·07)	0·99 (0·93–1·01)	0·99 (0·93–1·05)
Displacement scenario						
2018 vs 2017 SRR	NA	NA	NA	NA	1·16 (1·10–1·22)	1·27 (1·20–1·34)
2017 vs 2016 SRR	1·25 (1·18–1·32)	1·34 (1·27–1·42)	1·19 (1·13–1·26)	1·14 (1·08–1·20)	1·20 (1·14–1·27)	1·12 (1·06–1·19)
2017 vs 2015 SRR	1·29 (1·22–1·37)	1·31 (1·24–1·38)	1·29 (1·27–1·34)	1·27 (1·20–1·34)	1·19 (1·13–1·26)	1·19 (1·13–1·26)
2017 vs 2014 SRR	1·15 (1·09–1·22)	1·07 (1·02–1·13)	1·09 (1·03–1·15)	1·09 (1·04–1·15)	1·29 (1·22–1·36)	1·28 (1·21–1·35)
2017 vs 2013 SRR	1·15 (1·09–1·21)	1·28 (1·21–1·35)	1·25 (1·18–1·32)	1·23 (1·17–1·30)	1·23 (1·16–1·29)	1·17 (1·10–1·24)
2017 vs 2012 SRR	1·13 (1·07–1·19)	1·23 (1·16–1·30)	1·13 (1·07–1·19)	1·18 (1·12–1·25)	1·18 (1·12–1·25)	1·11 (1·05–1·17)
2017 vs 2011 SRR	1·15 (1·09–1·22)	1·18 (1·12–1·24)	1·16 (1·10–1·23)	1·13 (1·07–1·19)	1·12 (1·06–1·18)	1·13 (1·07–1·20)
2017 vs 2010 SRR	1·17 (1·11–1·24)	1·16 (1·10–1·22)	1·06 (1·00–1·12)	1·05 (1·00–1·11)	NA	NA
Range	1·13–1·29	1·07–1·34	1·06–1·29	1·05–1·27	1·12–1·29	1·11–1·28
Median (IQR)	1·15 (1·15–1·25)	1·23 (1·16–1·31)	1·16 (1·09–1·25)	1·14 (1·09–1·23)	1·19 (1·16–1·23)	1·17 (1·12–1·27)

Data are SRR (95% CI), unless otherwise stated. Puerto Rico age distribution in 2015 is the standard population. SRR=standardised rate ratio. NA=not applicable.

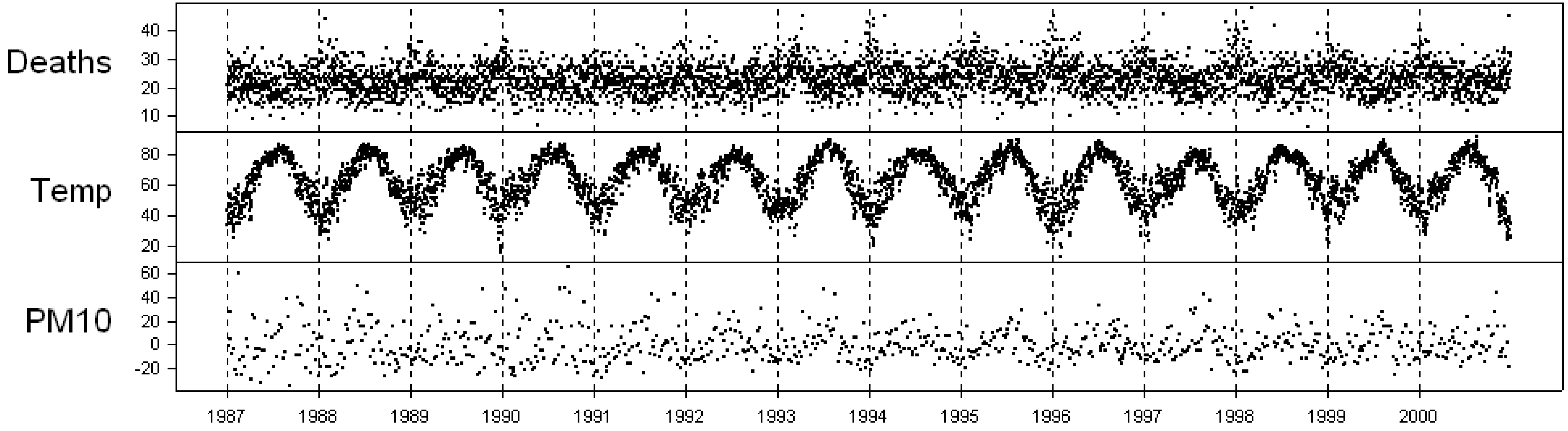
Table 5: Age-standardised mortality rate ratios in Puerto Rico by month

Particulate air pollution and mortality – London Smog

London, 12 noon, December, 1952; Particulate levels – 3,000 μg/m^3


~10,000 excess deaths
Fog period
Deaths
Sulphur dioxide
Smoke
Deaths
Smoke and Sulphur dioxide (mg/m^3)
1000
750
500
250
4·0
3·0
2·0
1·0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Date December 1952

National **M**orbidity **M**ortality **A**ir **P**ollution **S**tudy



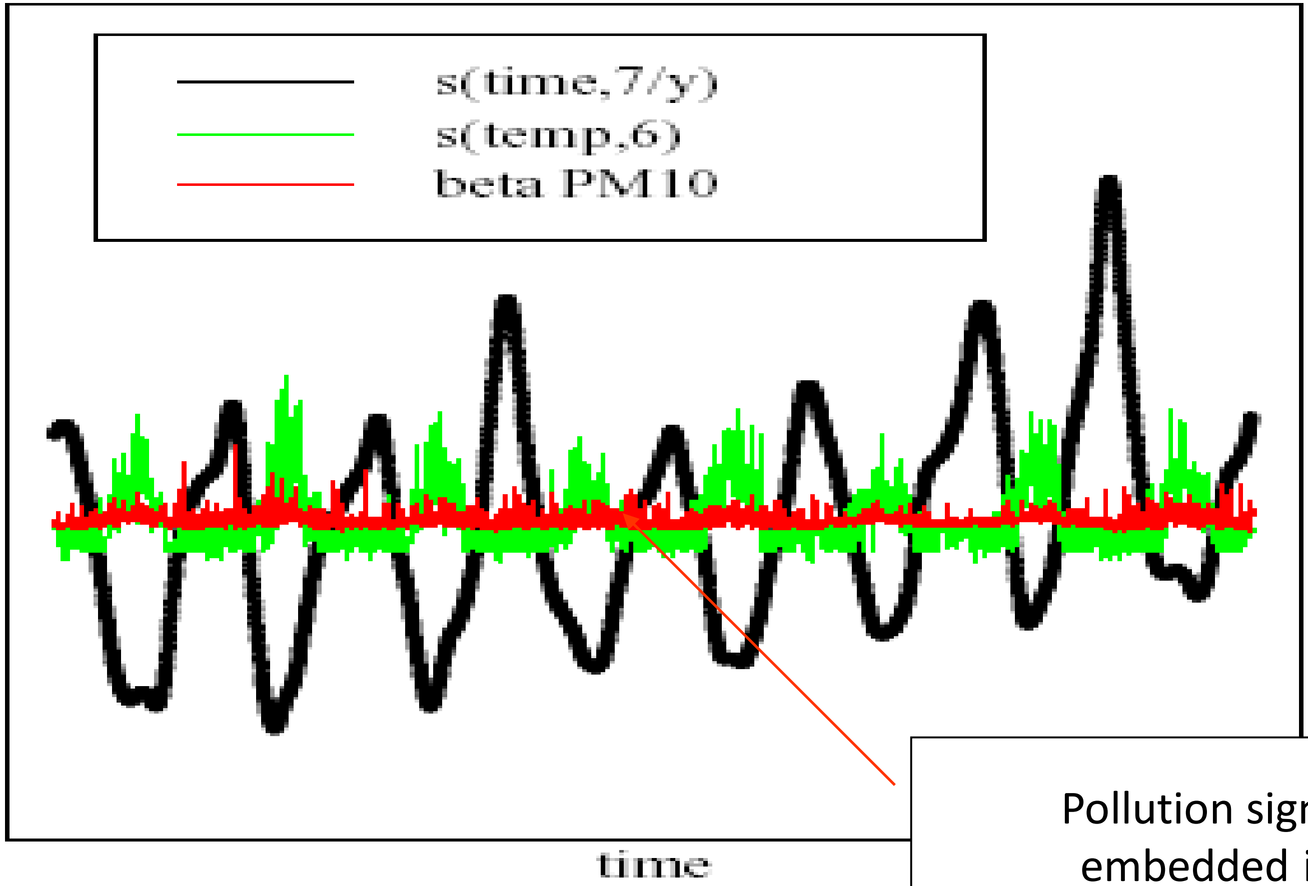
Semi-parametric Regression Model for Each City (c)

- Y_t^c is the mortality count on day t at location c
- PM_{10t}^c is the level of particulate matter on day t at location c
- β^c is the relative rate of mortality, e.g. the percentage increase in mortality associated with 10 $\mu g/m^3$ increase in PM_{10}
- $s(\text{temp}, df)$ are smooth functions (smoothing splines or regression splines)

$$\begin{aligned}\log E[Y_t^c] &= \text{age-specific intercept} + \boxed{\beta^c PM_{10t}^c} + \boxed{s(\text{time}, 7/\text{year})} + \\ &+ \boxed{s(\text{temp}, 6)} + s(\text{dewpoint}, 3) + \text{age} \times s(\text{time}, 8) + \ldots\end{aligned}$$

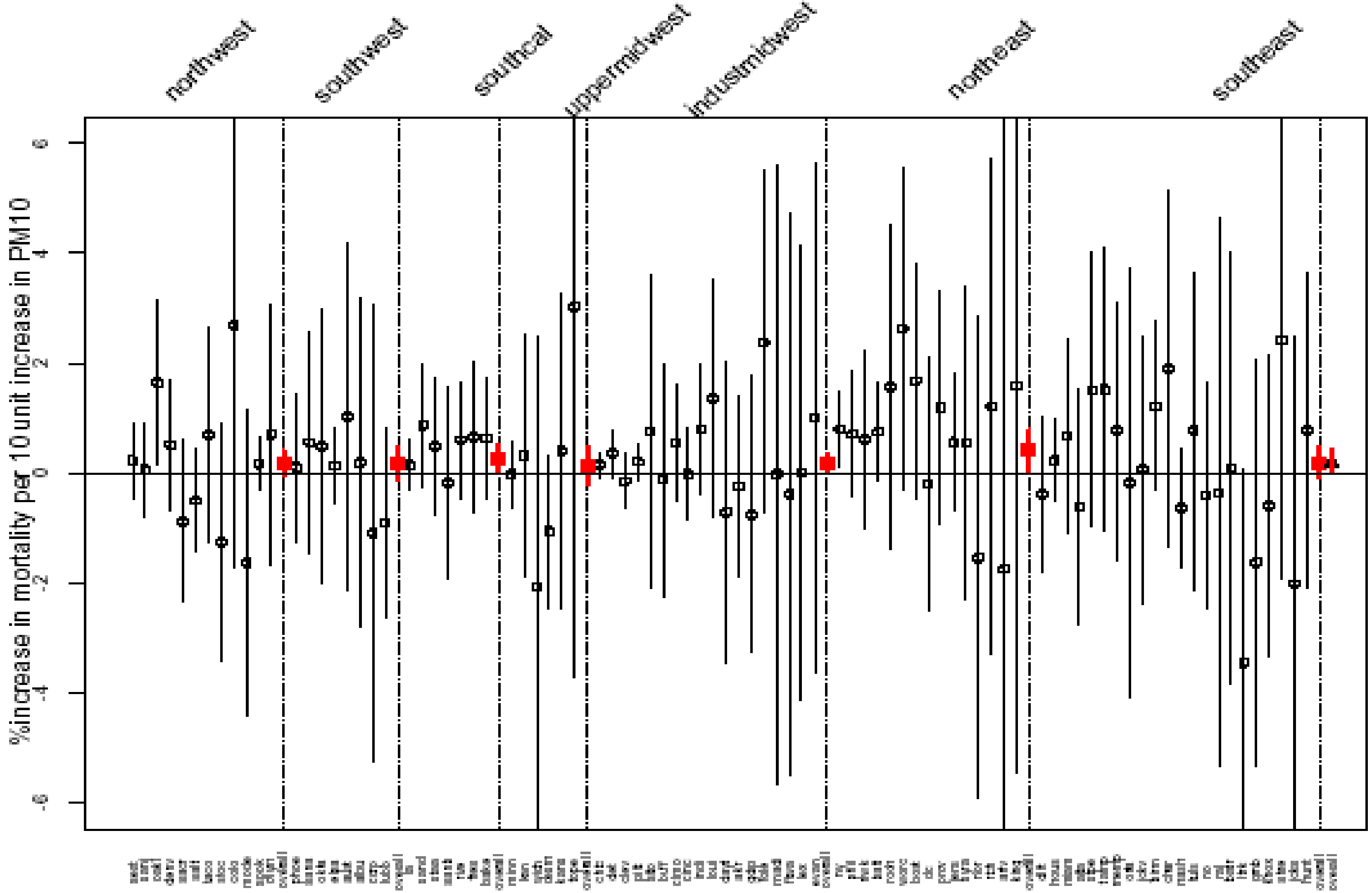
- estimate $\hat{\beta}^c$ and its statistical variance v^c within each location
- Generalized additive models (GAM) with smoothing splines or Generalized Linear Models (GLM) with natural cubic splines are the methods of choice

Statistical Problem



Pollution signal embedded in correlated "noise"

City-specific Estimates

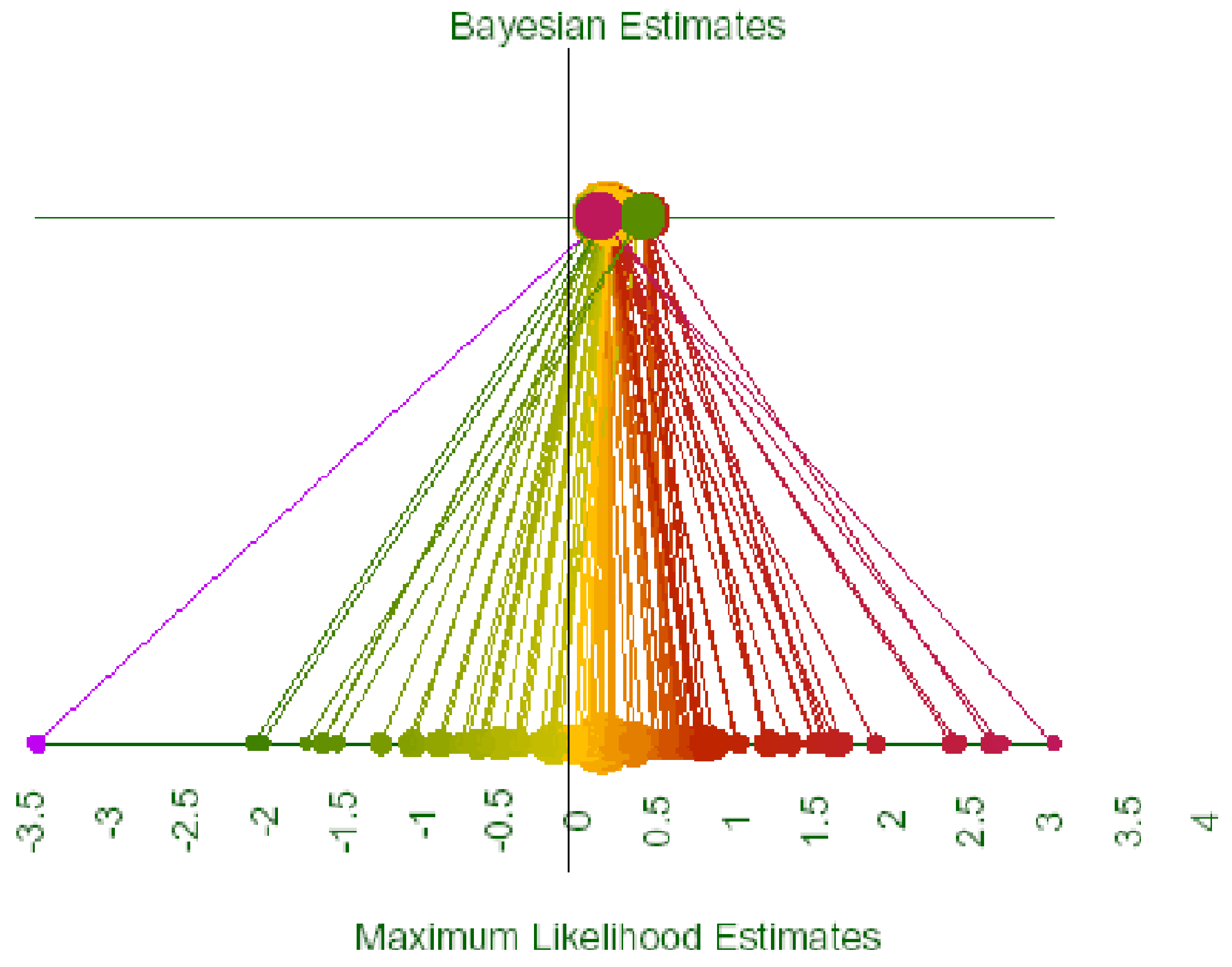


Spatial Model for Relative Rates

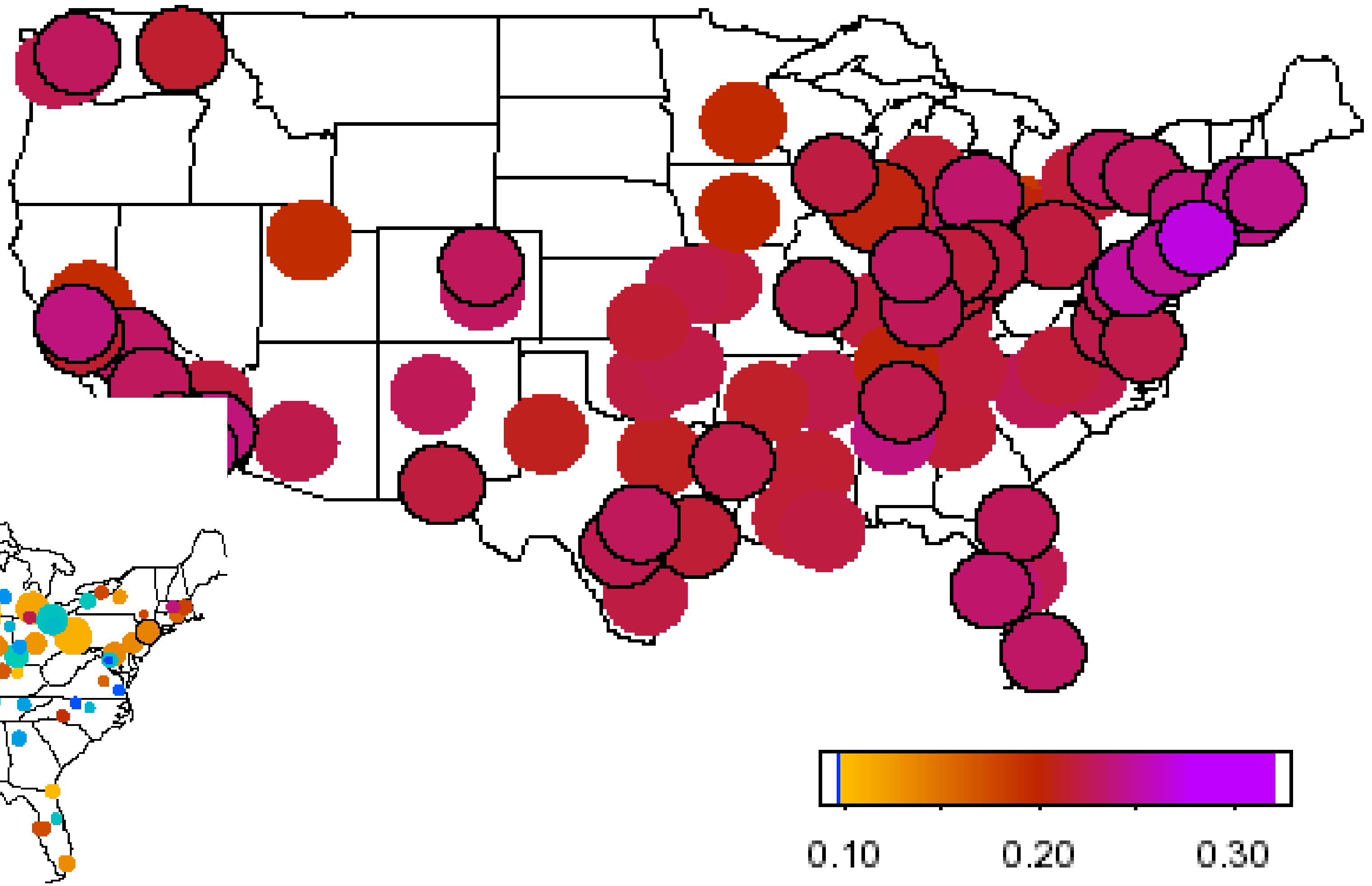
- β_r^c = log relative rate of mortality associated with PM_{10} in city c in region r
- α_r = average log relative rate in region r
- α^* = national average log relative rate

$$\hat{\beta}_R^c = \alpha^* + \underbrace{(\hat{\beta}_r^c - \beta_r^c)}_{\text{statistical error}} + \underbrace{(\beta_r^c - \alpha_r)}_{\text{heterogeneity within region}} + \underbrace{(\alpha_r - \alpha^*)}_{\text{heterogeneity across regions}}$$

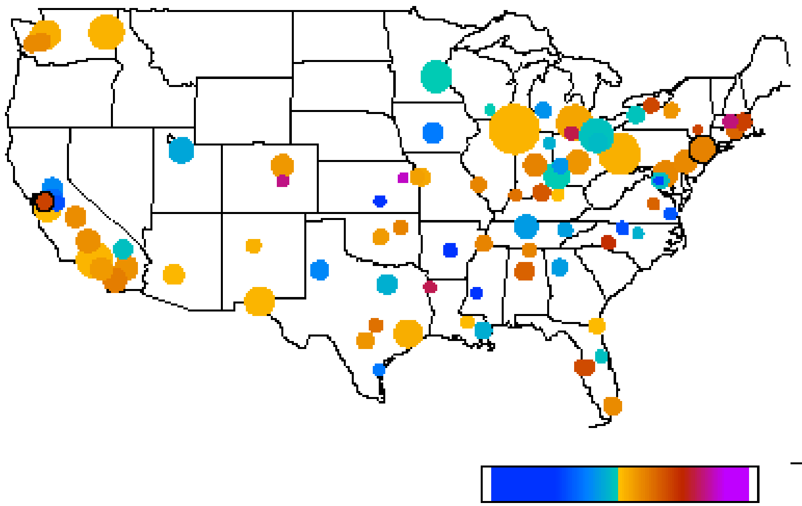
Shrinkage


Bayesian Estimates
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
Maximum Likelihood Estimates

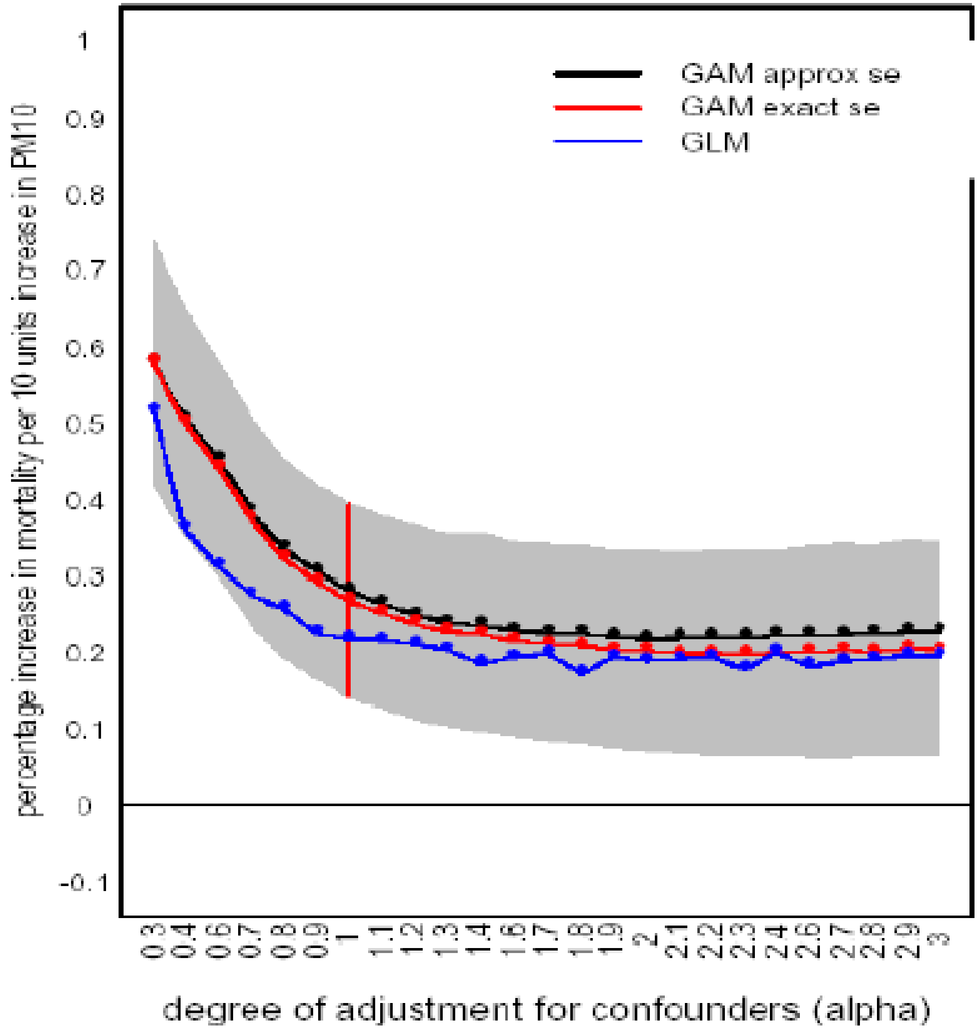
Smoothed Estimates



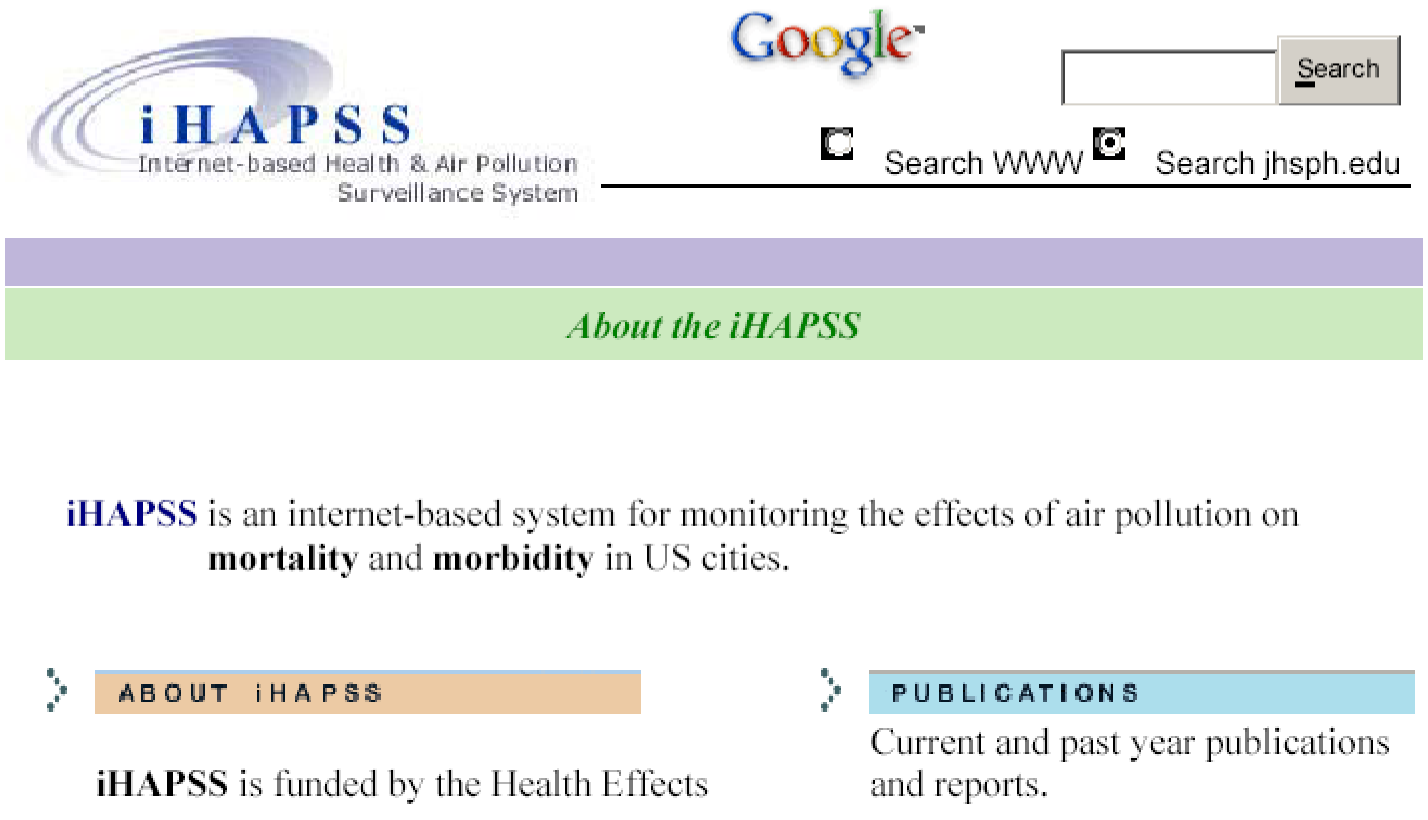
MLE



Statistical Issues: Degree of Adjustment



internet Health and Air Pollution Surveillance System (iHAPSS) + R package




THIS WEEK

NEWS

PAGE 28 Sputtering radio galaxy

BUDGET

31 Mixed U.S. budget outlook

AIR POLLUTION

Evidence Mounts That Tiny Particles Can Kill

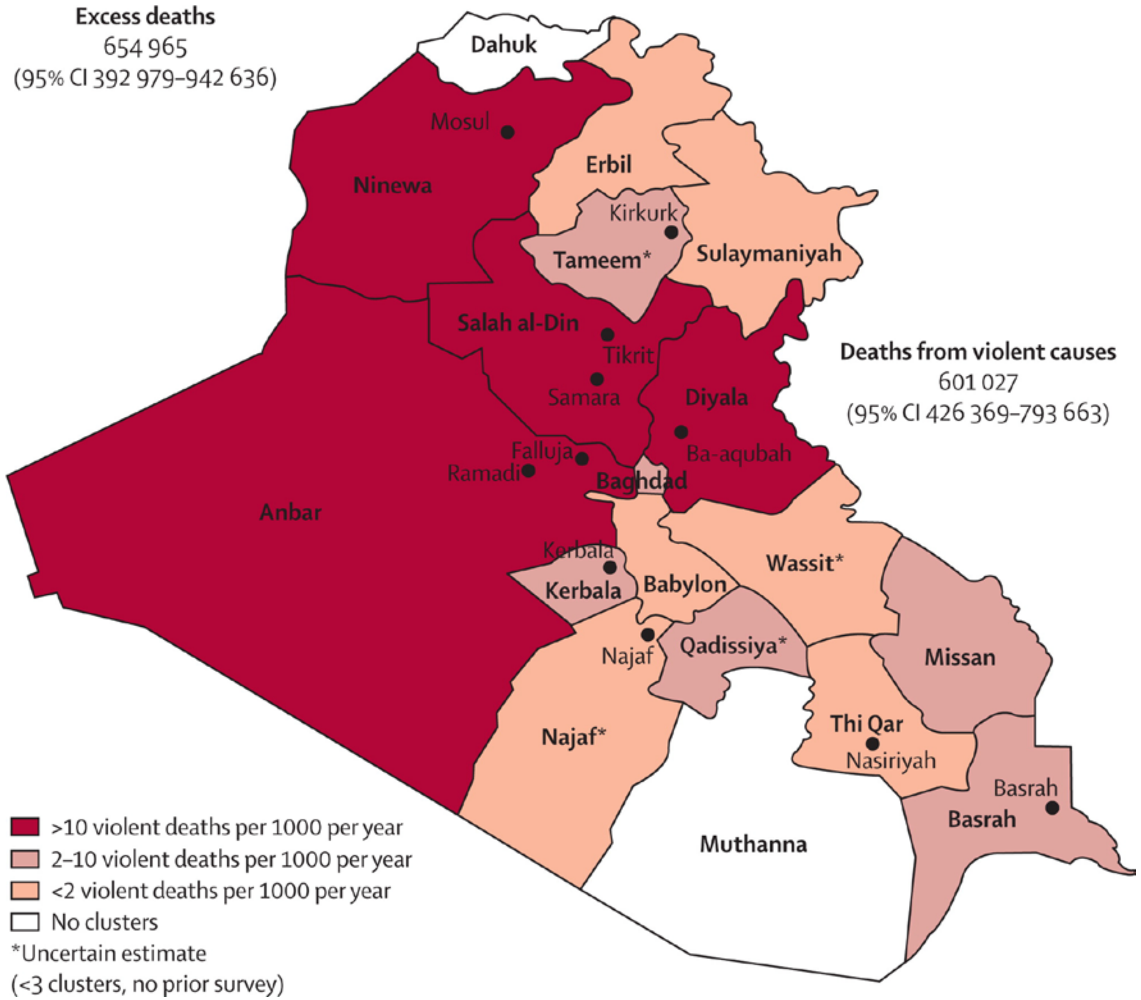
Profile of a killer.

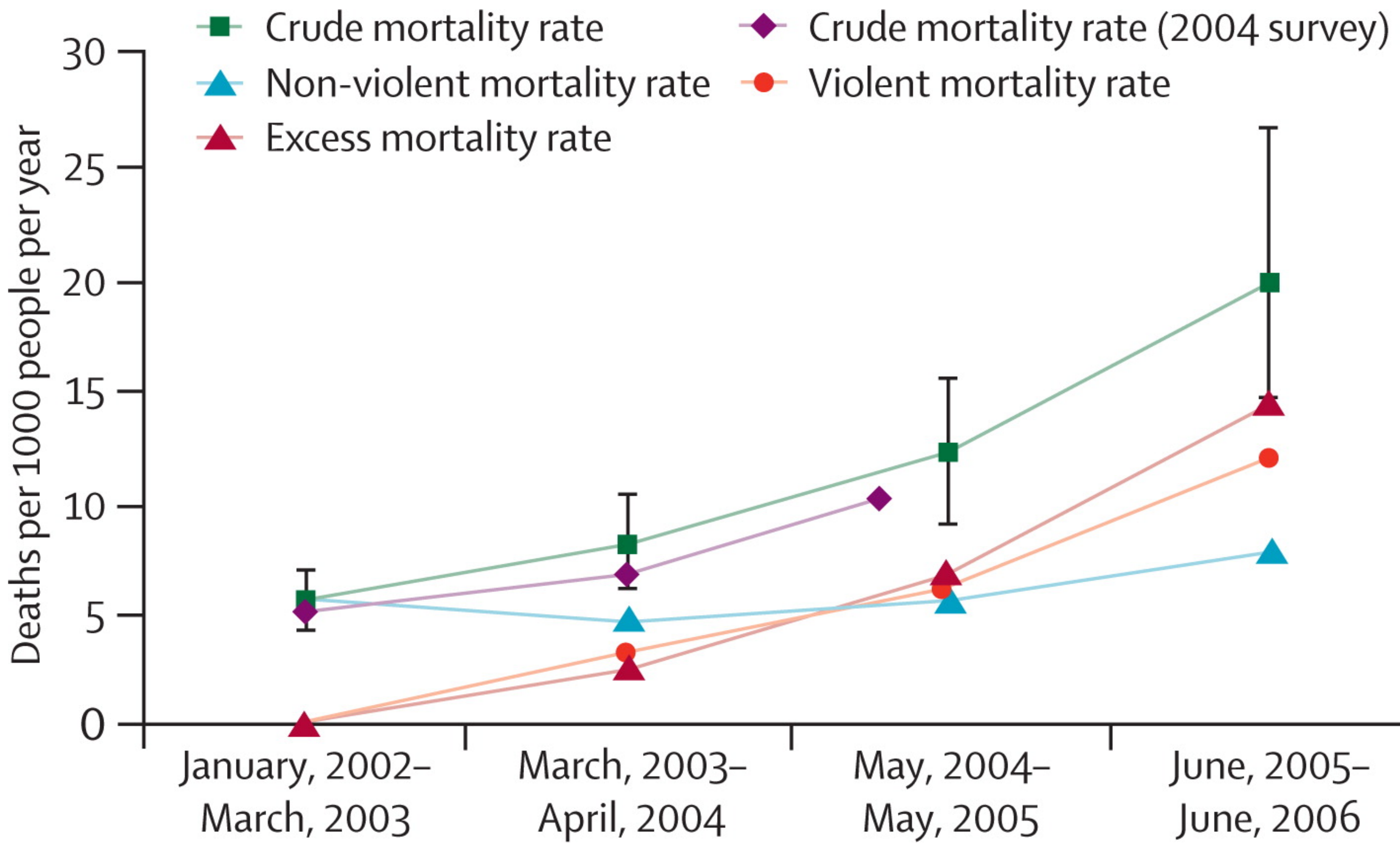
22 7 JULY 2000 VOL 289 SCIENCE www.sciencemag.org

“(A)lthough many questions remain about how fine particles kill people, the NMMAPS study shows *there’s no mistaking that PM is the culprit*

Understatement of statistical uncertainty in the press

Iraq Invasion Civilian Mortality – Burnham, et al. 2006. Lancet

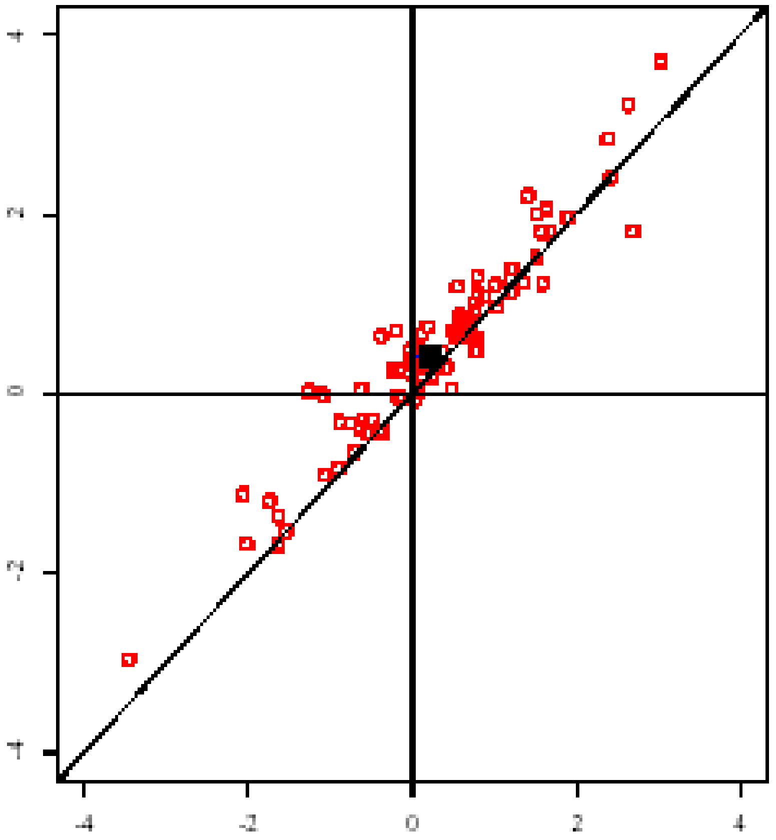




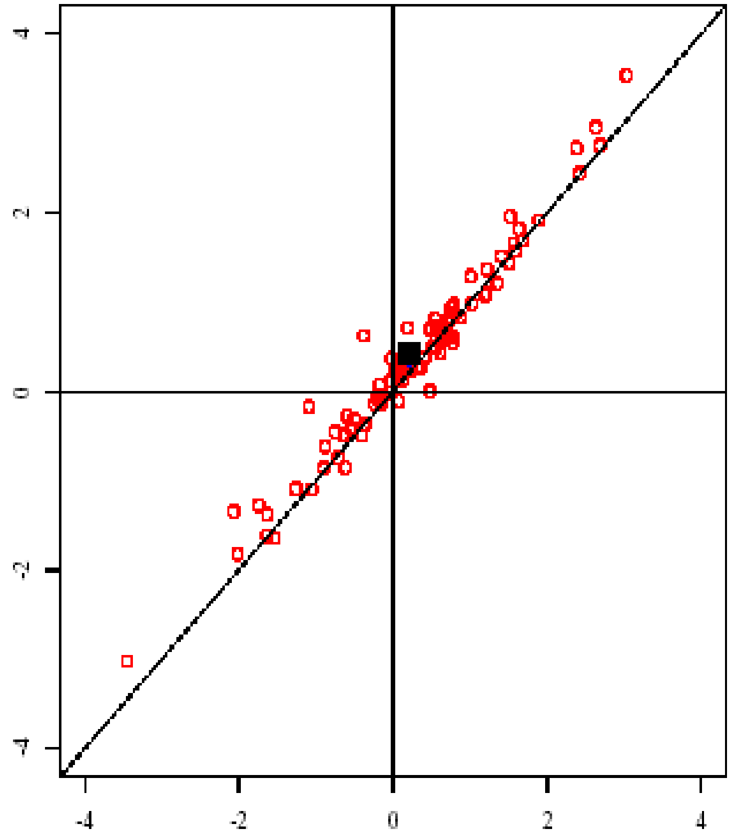
Crude mortality rate
Crude mortality rate (2004 survey)
Non-violent mortality rate
Violent mortality rate
Excess mortality rate
Deaths per 1000 people per year
30
25
20
15
10
5
0
January, 2002–
March, 2003
March, 2003–
April, 2004
May, 2004–
May, 2005
June, 2005–
June, 2006

Model Sensitivity: relative rate estimates for GAM (default and strict) versus GLM

GAM (default) versus GLM estimates



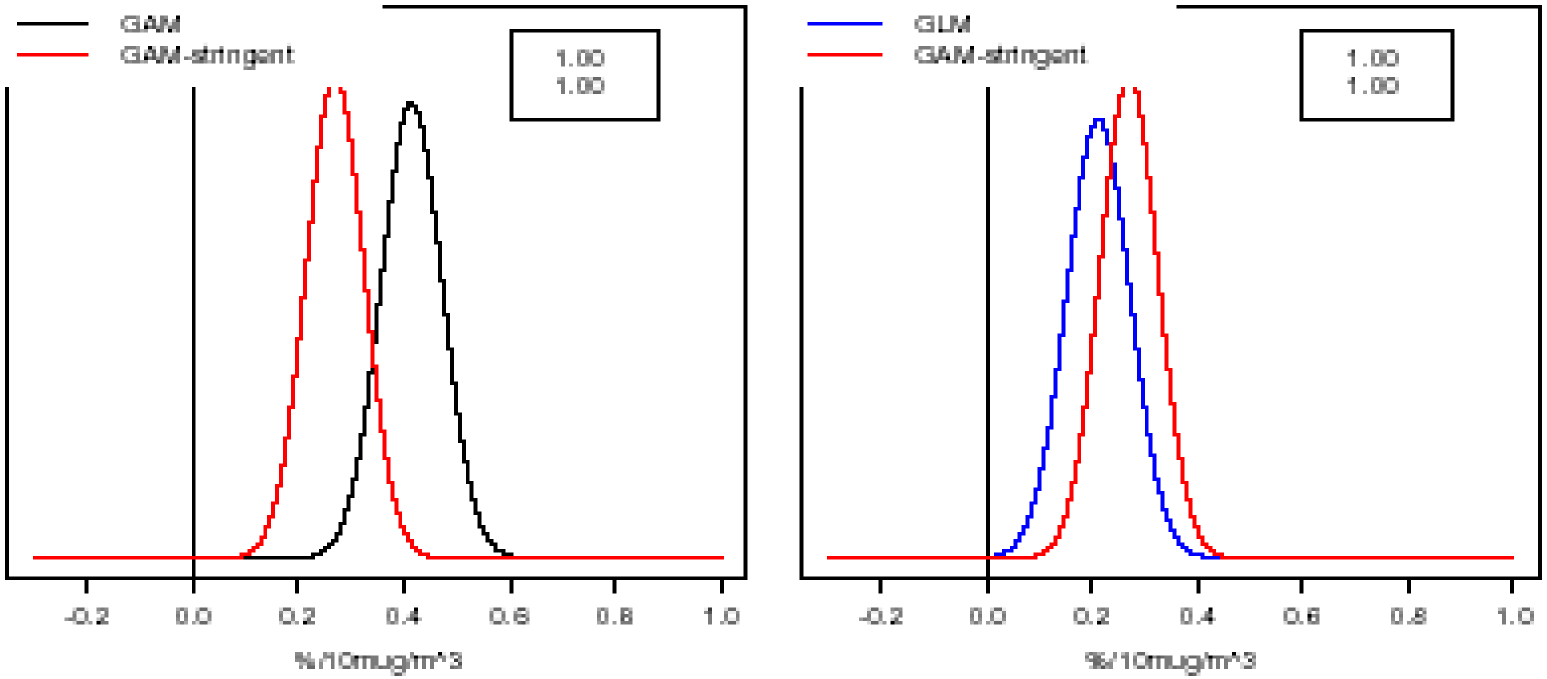
GAM(strict) versus GLM estimates



Dominici, McDermott, Zeger, Samet AJE 2002

What difference did it make?

Pooled effects at lag 1 for GAM, GAM-stringent, GLM



$0.27\% \leftarrow 0.41\%$ $\qquad$ $0.21\% \leftarrow 0.27\%$

The Press: The New York Times (June 2002)

Data Revised on Soot in Air and Deaths

Scientists Lower Their Estimate of Risk From Bad-Air Days

Revisiting their own data with new methods, scientists who conducted influential studies that linked sooty air pollution with higher death rates have lowered their estimate of the risk posed by bad-air days.

New research may delay a review due next year on small-particle pollution.

The researchers, at the Johns Hopkins University, have been distributing their new analysis to scientists and government officials by fax and e-mail. Yesterday, they set up a Web site, biosun01.biostat.jhsph.edu/~fdominic/research.html, that details their new findings.

Dr. Zeger and Mr. Greenbaum stressed that the new findings did not overturn the basic link between soot and illness. They also pointed to the recent publication of other studies on the long-term effects of soot that do not use the same analytical tools.

Still, industry officials said the new findings called into question the validity of some research underlying the new federal standards.