Donora, PA, 1948.
(Source: *New York Times*, November 2 2008)

Air Pollution and Mortality

Background of major air pollution incidents

- Donora, PA, 1948
- London, 1952

Consequences:

- British Clean Air Act 1956
- U.S. Clean Air Act 1963
- EPA established 1970
- Current: air pollution standards established by EPA and various national and international regulatory bodies, including WHO

U.S.E.P.A. Standards

National Ambient Air Quality Standard (NAAQS) — defines legal limits for air pollutants

Six *criteria pollutants* (particulate matter, ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, lead)

PM standard originally defined for *total suspended particulates* (TSP)

Changed to PM_{10} (1987)

New standard for $PM_{2.5}$ (1997)

Current standard for $PM_{2.5}$: 35 μg/m^3 as 24-hour max, 15 μg/m^3 as annual mean.

Data from London 1952

- Fig. 1.2: time series plot of deaths and smoke at 1-day lag
- Fig. 1.3: scatterplot of deaths v. smoke at 1-day lag

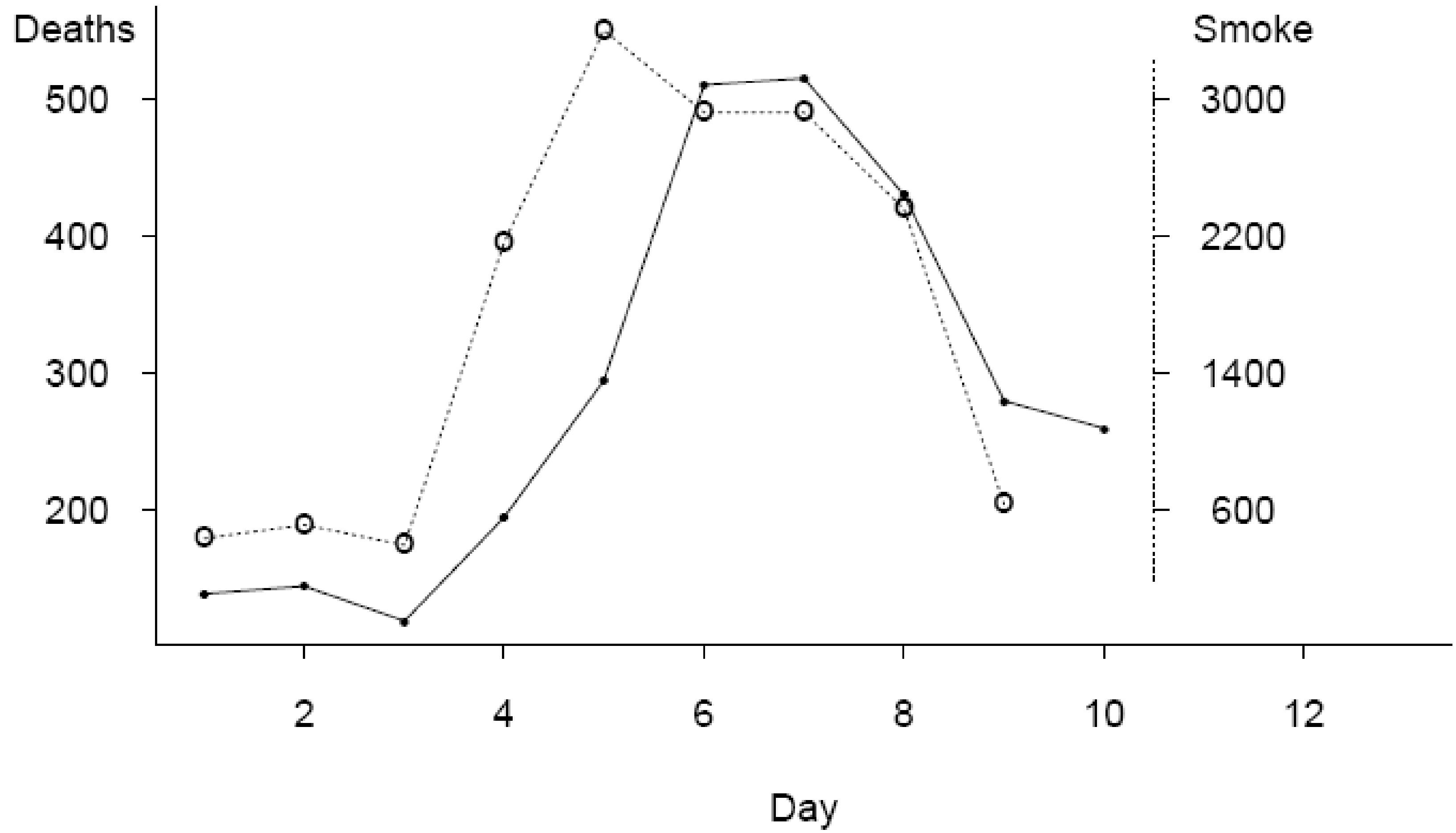

Figure 1.2. Smoke levels and deaths during 1952 London fog. Open circles and dashed lines: smoke levels by day. Closed circles and solid lines: deaths by day.

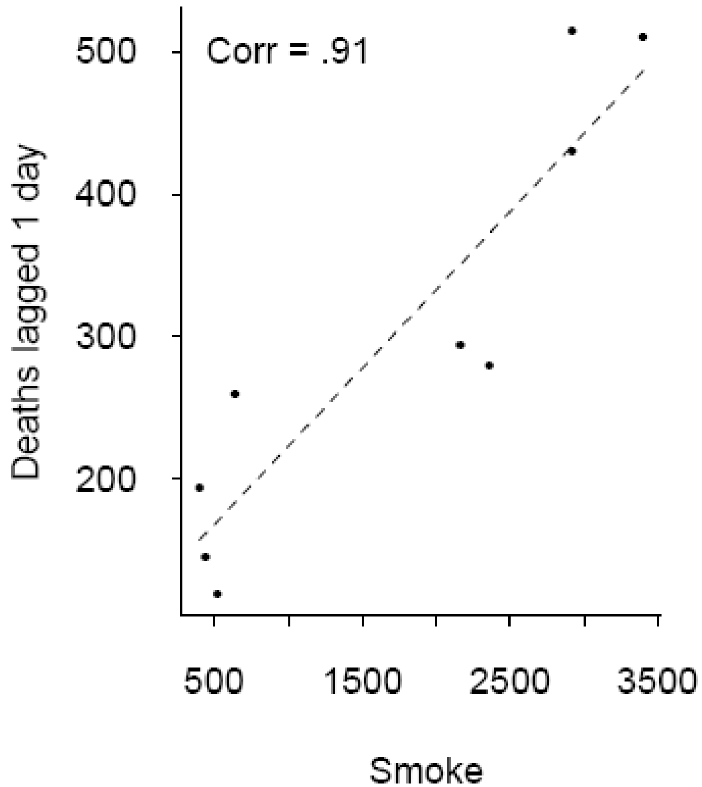

Figure 1.3. Scatter plots of deaths (lagged one day) against Smoke.

Data from London, December 1957

- Table 1.1: Daily deaths + Smoke, SO_2, Max and Min Temperature
- Fig. 1.4: time series plots
- Fig. 1.5: Scatterplots

Conclusion from this: difficult to tell *which* variable is responsible for deaths.

- *Confounder*: A variable that is correlated both with the outcome of interest (deaths) and with the true causal variable (assumed to be Smoke). Example: SO_2.
- *Effect Modifier*: A variable that, though not itself a causal factor, could change the relationship between the true causal variable and the outcome of interest (possibly temperature)

Need larger datasets to decide which variable might be a confounder or an effect modifier.

Day	Smoke ($\mu g/m^3$)	SO_2 (pphm)	Max Temp °F	Min Temp °F	Deaths (All)	Deaths (70+)	Deaths (0–69)
1	530	17	46	43	129	66	63
2	470	15	49	45	112	66	46
3	510	14	52	45	135	71	64
4	490	15	53	44	121	72	49
5	510	12	48	38	133	80	53
6	400	10	39	32	102	58	44
7	800	35	41	27	136	63	73
8	1200	38	41	25	178	101	77
9	1800	51	35	27	182	87	95
10	2300	63	43	29	198	99	99
11	1150	26	45	36	198	107	91
12	500	13	55	42	189	108	81
13	200	3	56	48	186	107	79
14	180	5	48	34	178	105	73
15	550	16	43	27	178	94	84
16	300	8	46	40	147	80	67
17	500	12	45	40	156	99	57

Table 1.1 Pollution levels, temperatures and deaths during the London smog incident of December 1957. Values are estimated from charts in Scott (1958).

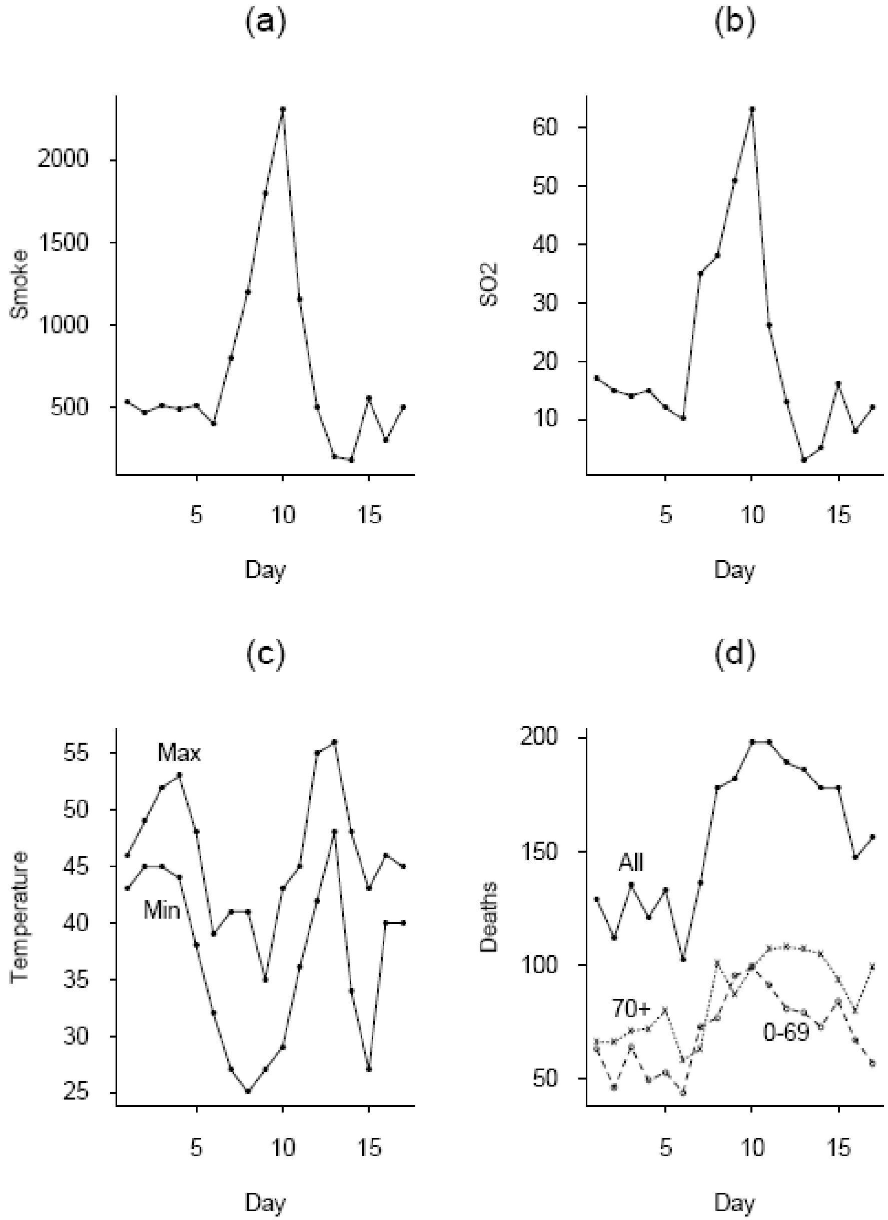

Figure 1.4. Time series plots of Smoke (a), SO_2 (b), Daily maximum and minimum temperature in °F (c) and deaths classified as all deaths, deaths aged 70 and over, and deaths aged under 70 (d), for the London smog of December 1957.

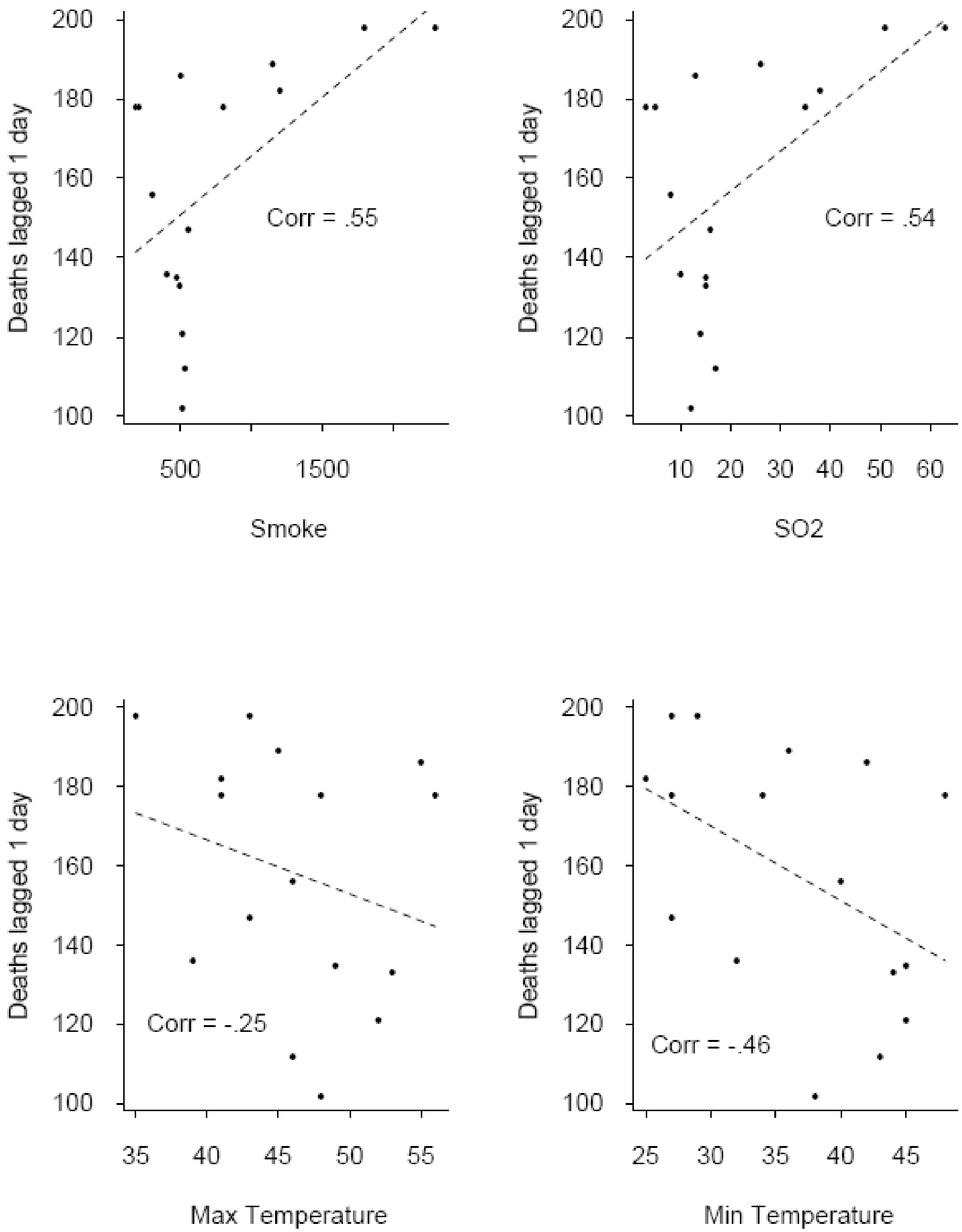

Figure 1.5. Scatter plots of total deaths (lagged one day) against each of the other variables in Fig. 1.4.

First Long-term Analysis

Mazumbar *et al.* (1982) analyzed 14 years daily mortality and air pollution data in London from 1959–1972.

Analyses based on annual averages are misleading — suggest SO_2 stronger effect than smoke.

Issues:

- Effect of 1963 outlier (very cold winter in UK)
- *Ecological bias*: presence of other year to year effects

In fact, this wasn't the analysis they used. They analyzed daily data for each of the 14 years, then combined the results. Their conclusion was that when both variables are included, Smoke and SO_2, only Smoke has a statistically significant effect.

In other words, SO_2 is most likely a confounding variable. Only Smoke appears to have a causal effect. First paper to present convincing evidence of this.

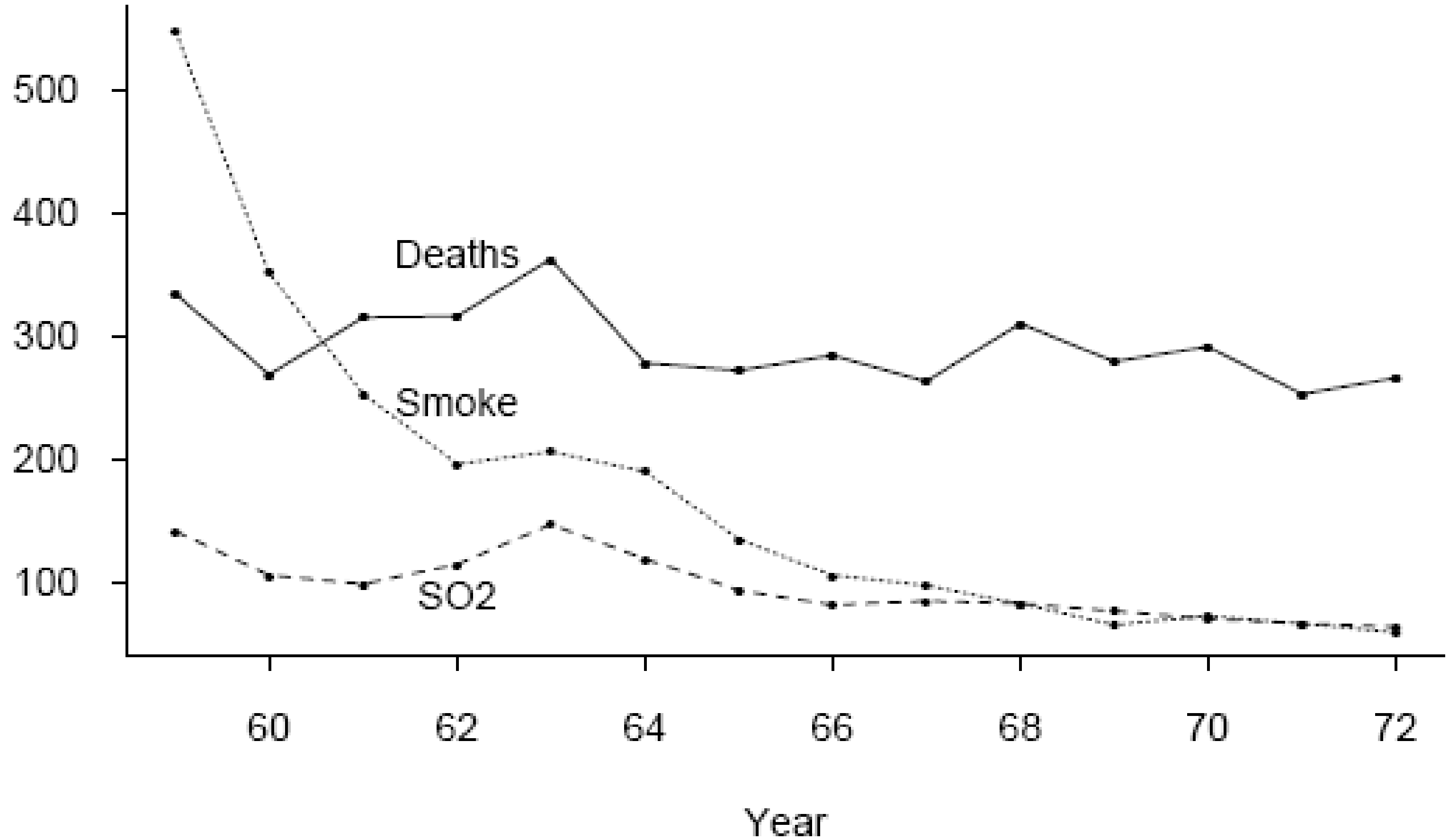

Figure 1.6. Time series plots (against year) of daily mean deaths, Smoke levels and SO_2, for 14 London winters.

Winter	Mean deaths	Mean SO_2	Mean Smoke
58–59	334	142	547
59–60	269	106	351
60–61	315	99	253
61–62	316	115	196
62–63	362	147	206
63–64	278	119	190
64–65	272	94	135
65–66	284	82	106
66–67	263	85	98
67–68	310	83	83
68–69	279	77	66
69–70	291	71	73
70–71	253	67	67
71–72	266	64	60

Table 1.2 Mean deaths per day, SO_2 level (ppb) and Smoke level ($\mu g/m^3$) for 14 London winters.

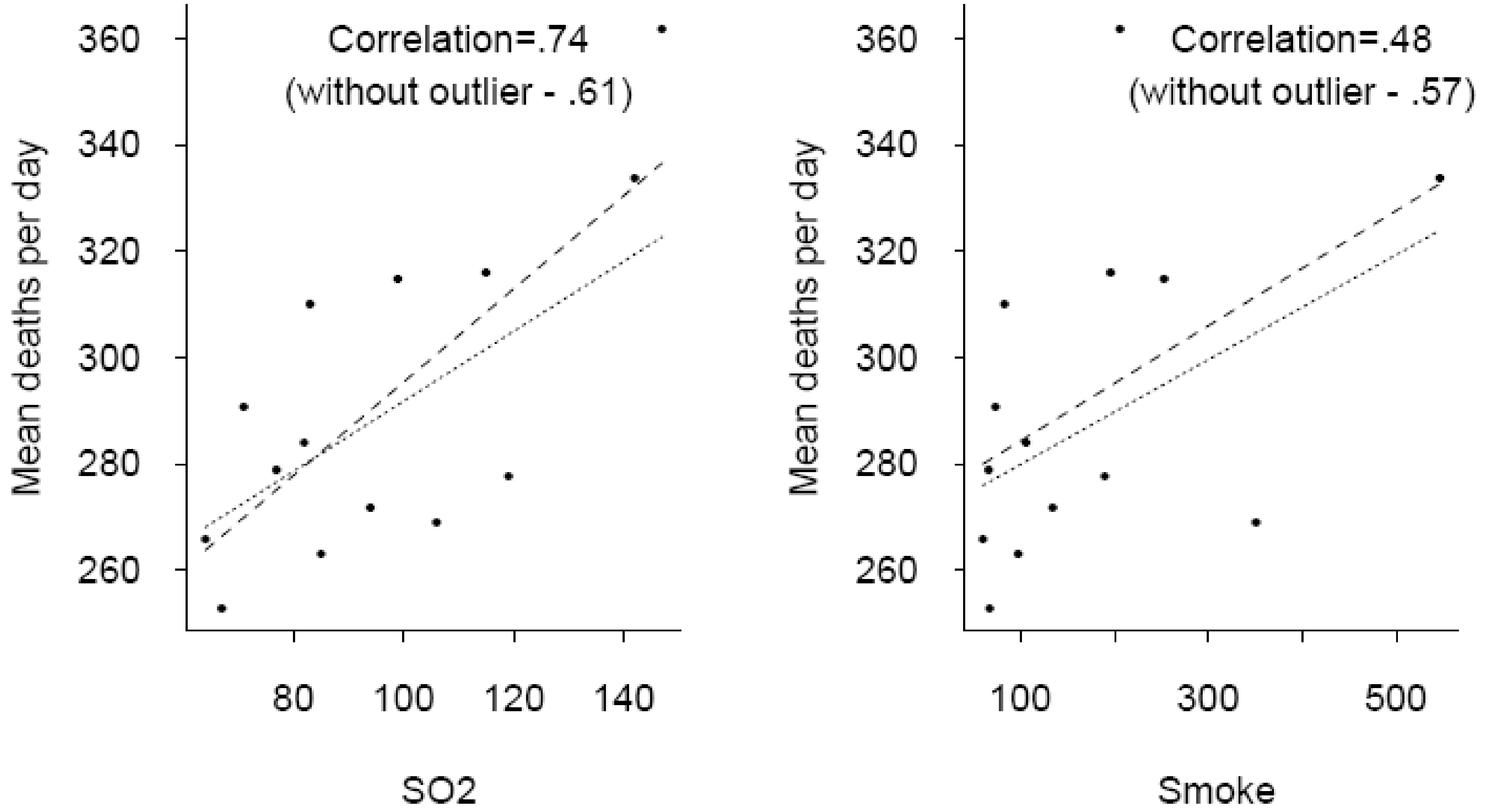

Figure 1.7. Scatterplots of annual mean deaths against mean Smoke levels and mean SO_2, for 14 London winters. Long-dashed lines are fitted straight lines to the whole data; short-dashed lines are fitted straight lines to data omitting suspected outlier for 1962–63.

Data from Philadephia 1974-1988

Here, we give a sample analysis based on daily mortality data from Philadelphia. This mimics an analysis originally given by Samet and Zeger (1997)

- Data
  - 14 years of daily death data (aged 65+)
  - Meteorology — temperature, dewpoint
  - Criteria pollutants — TSP, SO_2, NO_2, CO, O_3

- Weekly deaths with smoothed curve (Fig. 1.8)

- Nonlinear relationships between mortality and other variables (Fig. 1.9). (But warning: only plotting two variables at a time could be misleading)

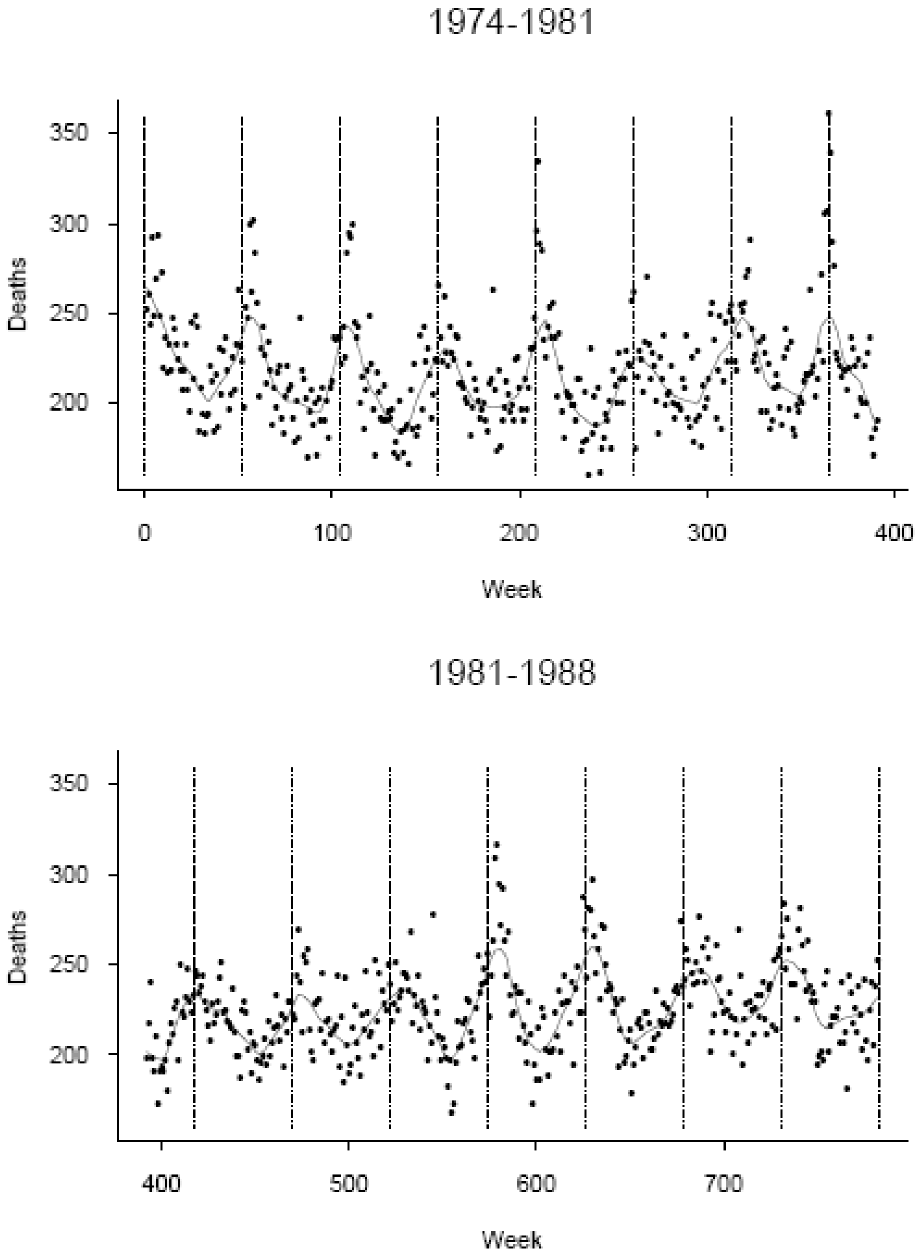

Figure 1.8. Time series plot of weekly deaths in Philadelphia, with smoothed lowess curve. The vertical dotted lines are placed to indicate the ends of years.

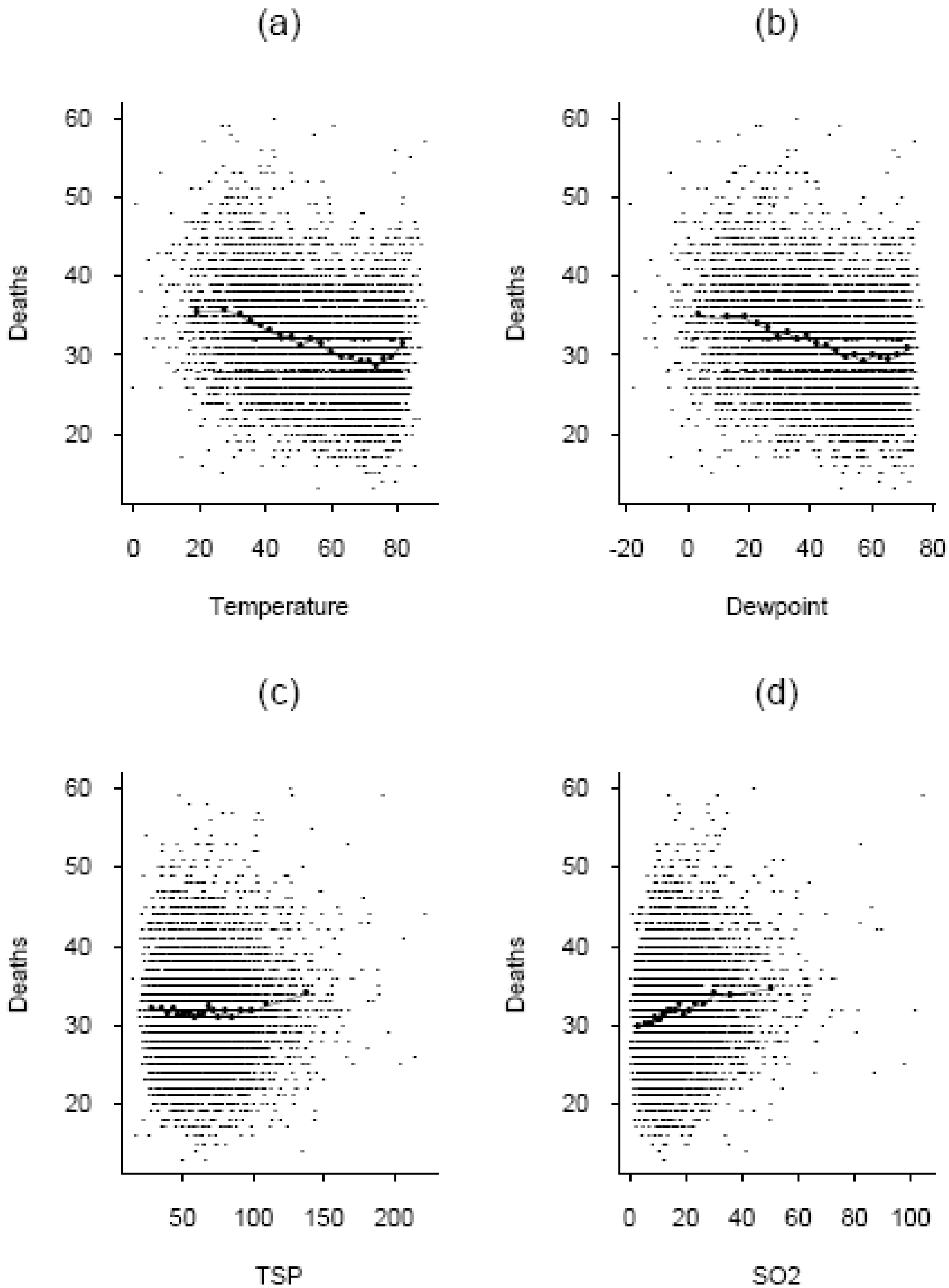

Figure 1.9. Scatterplots of daily deaths against four covariates, with fitted subsample averages (see text for details).

Details of Analysis

- y_t is square root of daily deaths (variance-stabilizing transformation — Chapter 5)
- Seasonal and long-term trend modeled nonlinearly through 180 cubic spline basis functions — Chapter 6
- Meteorological effects — linear and quadratic terms, indicator for exceeding a threshold, also lagged values. Use variable selection to reduce number of terms (Chapter 5)
- 5 pollutants entered singly and in combination — also take account of lagged variables and possible combinations of lags
  - TSP — current day ($t = 3.1$)
  - SO_2 — current day ($t = 3.3$)
  - NO_2 — 4-day lag ($t = 2.1$)
  - CO — average of lags 3 and 4 ($t = 2.8$)
  - O_3 — average of lags 1 and 2 ($t = 2.9$)

Neglects *simultaneous testing* issue (Chapters 2 and 3)

- Put pollution variables in pairwise or all five together (in latter case, only O_3 is significant — $t = 2.7$)

- Correlations between pollutant variables — only O_3 is not highly correlated with at least one other variable (Table 1.3)

  Illustrates problem of *multicollinearity* (Chapter 5)

- Also consider threshold effects (Fig. 1.10(a)) and possible nonlinear analysis (Fig 1.10(b))

	TSP	SO_2	NO_2	CO
SO_2	.61			
NO_2	.09	.07		
CO	.01	.10	.47	
O_3	.16	−.29	−.01	−.19

Table 1.3 Correlation table for the five pollution measures

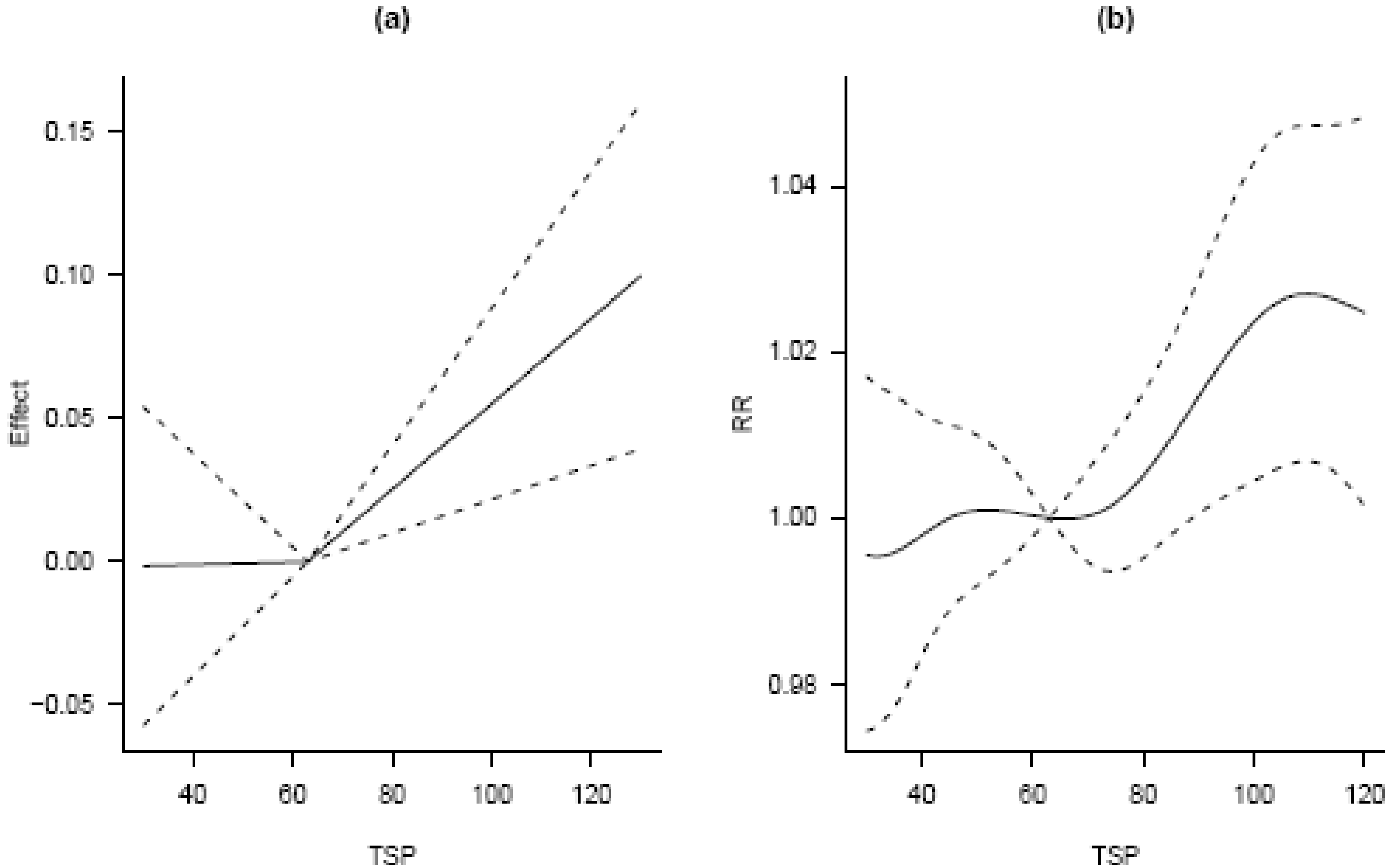

Figure 1.10. Plot of piecewise linear effect derived from equation (1.3) in (a), and a nonlinear representation of relative risk in (b). In each case the plot is normalized at a TSP level of 63 $\mu g/m^3$.

Conclusions

- Each pollutant on its own has a seemingly significant effect (except maybe NO_2)
- When different pollutants are entered in combination, only O_3 is consistently significant (but this could be because of multicollinearity)
- Nonlinear analysis — TSP appears to have little effect below 75 $\mu g/m^3$. Evidence of possible threshold is relevant for determining a standard.