Local Control Analysis of Radon and Ozone

S. Stanley Young, CGStat LLC
Robert L. Obenchain, Risk Benefit Statistics LLC
Goran Krstic, Fraser Health Authority

Abstract

Large (observational) data sets typically present research opportunities, but also problems that can lead to false claims. In Big Data, the standard error of an average effect estimate goes to zero as sample size increases, so even small biases can lead to declared (but false) claims. In addition, the average of a treatment difference, a so called main effect, can be almost meaningless when there are interactions with confounders that create local variation in effect-sizes. Data miners need statistical methods that can deal simply and efficiently with these sources of bias. Here, we demonstrate use of a JMP add-in, Moving Median, and a new JMP platform, Local Control, for the analysis of two data sets. Our first case study illustrates reduction of bias in an environmental epidemiology data set. Our second study uses Local Control on a time series air quality example. By detecting interactions, data miners can produce more realistic and more relevant analyses that reduce the bias typically implied by the variety and heterogeneity of Big Data. Our results are somewhat surprising. Radon appears to protect against lung cancer and ozone appears to have little or no effect on acute mortality.

Introduction

We will conduct what is call a Local Control Analysis, LCA, **Obenchain (2009**, **2010)**, **Obenchain and Young (2013)**, of two air quality data sets. Local Control Analysis first clusters the data set and then conducts a simple analysis within each cluster. The statistics coming from each cluster are then analyzed further. The whole analysis process can be subjected to a sensitivity analysis by varying how the steps of the analysis are conducted. We will use a simple difference within each cluster with our first data set, county levels of radon and lung cancer. We will use simple linear regression within the clusters for our second data set, daily deaths, air quality (ozone) and temperature. The second data set comes from London England. Once the simple statistics are computed, Local Treatment Differences for radon and Intercept and Slope for air quality, we determine if these treatment effects vary significantly across clusters. For that evaluation we use recursive partitioning. Finally, the entire process can be examined with a sensitivity analysis. There is a SAS JMP Add-In, Local Control, for automating steps in the analysis process, **Wolfinger and Obenchain (2014)**.

There is some controversy on the question of radon causing lung cancer. The Environmental Protection Agency cites two meta-analysis papers of case control studies, **Darby et al. (2004)** and **Krewski et al. (2005)**, that support the claim that radon causes lung cancer. On the other hand, hormesis, https://en.wikipedia.org/wiki/Hormesis, is pointed to in support of the observation that low dose radon is protective against lung cancer. Of course smoking is the important variable lurking in the radon data set. Local Control Analysis allows us to address smoking and other confounding variables.

There are multiple papers on both sides of the radon/lung cancer question, **Appleton (2007)**, **Cohen (1989**, **1995, 1997**, **2008)**, **National Research Council (1999)**, **Internatinal Agencey for Research on Cancer (1998)**. For radon, using public sources, we built a data set where each record is for one US County; we have data for 2,881 counties. The variables used in our study are given in **Table 1**. Included are Mean values, standard deviation and p-values for simple linear regression predicting mortality.

There are multiple papers on all sides of the ozone/acute deaths question, **Bell et al. (2004)**, **Young and Fogel (2013)**, **Milojevic et al. (2014)**, **Smith et al. (2009)**.

The historic "London Smog" of 1952 and the associated deaths galvanized the public health interest in air quality and health effects. Much research finds an association between ozone and acute mortality, e.g. **Bell (2004)** and PM2.5 and acute mortality with temperature being an important covariate for both. Extreme temperatures are associated with increased mortality. There are papers asserting no association if covariates are controlled, most notably **Chay et al. (2003)** and **Greven et al. (2011).** Overall the literature is mixed. Our impression is that the results depend heavily on how well covariates are controlled. We turn out attention to a London ozone and temperature data set.

Data Radon

The presented study includes epidemiological and environmental data from 2,881 counties of the continental United States. The data was collected from various public sources. The data set includes U.S. county-level lung cancer mortality, current smoking, ever smoking, household income, obesity, population and housing, and the indoor radon data (**NCI, 2015a**; **NCI, 2015b**; **U.S. Census Bureau, 2000**; **U.S. Census Bureau, 2015**; **Max Masnick, 2011**; **U.S. EPA, 2014**). The most current version of the radon data set can be obtained from Stan Young, stan.young@omicsoft.com.

Data Ozone

For air quality and mortality, we obtained the data set used by **Bhaskaran et al. (2013)** in their Education Corner on Time Series Regression. The data set covers the years 2002 to 2006, 5 years, and includes daily ozone, temperature, relative humidity and deaths. We are in the process of rebuilding this data set.

For mortality, ozone, temperature and relative humidity, we also compute deviations from a time series smoother and also lag variables where the value is the value of the variable for the previous day.

In the study of acute mortality, usually the interest is in changes in mortality, ozone and temperature. So each time series is smoothed and deviations from the time series smoother are computed. We use a JMP Add-In, Time Series Smoother written by Paul Fogel, **Young and Fogel (2014).** The smoother uses a window that moves across the time series. Within the window the middle day is the day at issue. We omit a window around the day at issue and use the median of the remaining days as the smoother. After some thought and experimentation we use a 21-day window and a five day window. **Figure 1** shows the time series before and after the application of the smoother. It is clear that the seasonal and yearly trend has been removed. There is some literature support that the conditions of yesterday might have an effect today, so we build one day lag variables for Deaths, Ozone, Temperature and Relative Humidity, and Temperature. A regression analysis points to variables useful for clustering the data set, **Figure 2**. D_Temperature (today), Deaths-1 (yesterday), and D_Temperature-1 (yesterday). The analysis also points to the possible importance of ozone-1 (yesterday), but does not indicate that today's ozone level has much influence. If the ozone effect is real, it is very small.

The most current London Ozone data set can be obtained from Stan Young, stan.young@omicsoft.com.

Methods

Local control, LC, analysis strategy, **Obenchain (2009, 2010)**, **Obenchain and Young (2013)**, for large, observational data sets is easily explained as it is a series of simple steps that together provide a coherent analysis strategy. Nontechnical audiences with an understanding of clustering, simple differences, single predictor linear regression and histograms already know the basic technologies of the analysis strategy. Originally LC was designed to look at two treatments or a treatment and a control. LC starts by dividing data points without regard for their status as either treated or control, into many subgroups. The point is to assure that objects within a cluster are as alike as possible for their observed baseline x-characteristics. We will use hierarchical clustering. Next a simple difference between the two treatments is computed within each cluster:

$$E[(Y|t=1) - (Y|t=0)|X],$$

so that we have a single degree of freedom comparison, given X. We call this a "fair treatment comparison" and a local treatment difference, LTD. Next we display the LTDs in a histogram. Any cluster that contains only treated or only control patients is considered non-informative and omitted from further analysis.

The clustering and computing of LTDs can be thought of as “nonparametric preprocessing”. Specifically, for the observed y-outcome variable, LC creates a new, corresponding LTD variable that estimates the unknown, local, true counterfactual difference within each cluster. The objects within the clusters can be viewed as statistical clones.

We can use simulation to evaluate the LTD distribution. Simply place the data points at random into an equal number of clusters with the same sample size and re-compute the LTD distribution. Repeat the simulation many times to get a smooth null distribution. There are a number of ways to statistically compare the two distributions, but usually a visual inspection is informative.

The LTDs can be appended to the observations of their cluster thereby creating a new data set where the LTD is the variable of interest and the predictors, those used in clustering and other predictors, now become an analysis data set. This data set can be subjected to analysis by any suitable data analysis method, e.g. multiple linear regression, recursive partitioning, etc. We favor recursive partitioning.

In the data from 2,881 US Counties, the Radon median is 2.01 and the mean is 3.0. We chose a cut point to define low and high radon exposure at 2.6 pCi/L. Our choice is arbitrary. The EPA indicates that the average home radon level is 1.3 pCi/L and they recommend remediation if the level is 4 pCi/L or higher.

With the ozone data set, we replace the simple difference computation within each cluster with single variable linear regression and obtain an intercept and slope for each informative cluster. In our case the deviation of ozone, D_ozone, from a time series smoother is the predictor and D_mortality is the response. One can filter out clusters that are too small or do not have a good range for the predictor variable.

A SAS JMP Add-In to do Local Control Analysis is available at https://community.jmp.com/docs/DOC-7453.

A SAS JMP Add-In for time series smoother is available from Stan Young, stan.young@omicsoft.com.

Results Radon

In a preliminary analysis, we examined radon as a predictor of lung cancer and its overall regression coefficient was negative and highly significant (p-value $< 10^{-44}$). The preliminary analysis was unadjusted for confounding variables.

We now present the Local Control Analysis. Obesity(%), Age (%>=65) and Current Smoker(%) were used to cluster the data. These variable were selected for use in clustering as they are highly predictive of Lung Cancer Mortality. After clustering and computing the LTDs, we present histograms of the actual LTDs and the randomized LTDs, **Figure 3**. The distribution of lung cancer incidence is different from the randomized distribution: the observed distribution is much more compact than the random distribution.

To confirm that the difference in the distributions is real, we used recursive partitioning to see if the LTDs were predictable we computed a recursive partitioning analysis, JMP/analyze/modeling/partition. The terminal nodes of a regression tree are numbered 1 through 7, **Table 4**. The first split is on age; those with a higher %>65 have a larger reduction in lung cancer. Counties with fewer current smokers have a greater reduction in lung cancer, Terminal Node 1. The trend continues for Nodes 2 and 3. So for Nodes 1-3, the lower the % current smokers the greater the reduction in lung cancer. Next we turn to the counties with lower than 17.5% people over 65, Terminal Nodes 4 through 7. The lower the % currently smoke, the greater the reduction in lung cancer, reduction of 7.7% versus 4.5%. The lower the %Obese, the greater the reduction in lung cancer. The split giving rise to Terminal Nodes 4 and 5 indicates that the higher the %>65 the greater the reduction in lung cancer, although the difference is modest, -8.6 versus -7.1. In none of the terminal nodes, do we see an increase in lung cancer; the radon treatment effects are all negative.

All of the p-values in the recursive partitioning analysis are remarkably small, 10^{-37} to 10^{-461}.

The correlation structure of this data set is remarkably complex. Partial correlations were computed using the SAS JMP Add-in Partial Correlation and are displayed in **Figure 5**. There is the expected strong positive partial correlation between Current Smoking and Lung Cancer Mortality. The surprise, of course, is the negative partial correlation of Radon and Lung Cancer Mortality. A bit confusing/potentially troubling is the small, positive partial correlation between Radon and Current Smoking (0.1370; p<0.0001). This observation suggests that regions with higher prevalence of Smoking tend to have higher indoor Radon levels which, according to the published literature (Band et al., 1980; U.S. EPA, 2015), should be expected to act in synergy and potentiate the adverse effects of Radon leading to a significantly elevated lung cancer risk (i.e., a strong positive relationship between indoor Radon and Lung Cancer Mortality). Our analyses show the opposite: a strong negative relationship between indoor Radon and Lung Cancer Mortality and, hence, no apparent evidence of potentiation/synergism between Radon and Current Smoking. Indeed, in Figure 4, when current smoking is high the effect of radon is still negative.

In summary, the effect of radon on lowering Lung Cancer Mortality varies, but it holds up within clusters (that control for level of smoking). These results appear to be in contradiction to the two papers cited by the EPA.

Results Ozone

The ozone data set is clustered, 5 years and ~1,800 days, using the variables, Temperature (today), Deaths-1 (yesterday), and Temperature-1 (yesterday) into 35 clusters. In effect we are saying there are 35 kinds of days and within each day-type, the clustering variables are very similar. Of course as D_Deaths, D_ozone and D_ozone-1 were not used in the clustering, they are available to see if mortality and ozone are related. Within each cluster we compute a simple linear regression where D_Deaths is the dependent variable and yesterday's ozone, D_ozone-1, is the independent variable. We output Intercepts and Slopes for each cluster.

Figure 6 shows a histogram of the 35 intercepts, a p-value plot of their p-values testing for zero intercept, and a plot of the D_deaths versus D_ozone-1 for the one cluster where something appears to be going on. Intercept histogram has one outlier. Within that cluster as yesterday's ozone goes up, the number of deaths goes down!

Figure 7 gives a histogram of the slopes, and a p-value plot of the p-values testing that a slope is different from zero. There is no evidence that today's daily deaths are affected by yesterday's ozone.

LCA of ozone and mortality is divided into two parts, intercept and slope, **Figures 6 and 7**. The intercept from simple linear regression within the clusters measures the base mortality effect across clusters. Recursive partitioning trees were computed for intercept and slope, **Figures 8 and 9.** Complex relationships are uncovered for both, but ozone does not figure into the relationships with mortality.

Discussion

The effect of radon on lung cancer depends on age, currently smoking and obesity, but in essentially all cases the effect of higher radon is to lower incidence of lung cancer. These finding are in stark contrast to papers cited by the EPA.

In this data set there is no effect of current ozone on current mortality. Simple linear regression indicates that Ozone-1 influences mortality and Local Control Analysis shows no evidence that yesterday's ozone affects today's mortality.

Traditional statistical methods for observational data focus, almost exclusively, on the so-called average or "main-effect" of treatment. Because this measure is an average across potentially diverse situations, main effects can have quite limited utility. In medicine we have the current tension between the public health of populations in contrast to individualized medicine. In the real world, the answer to any complex question is most likely to be "it depends". Local Control Analysis shows how the answers to the radon and the ozone questions change with the situation. In both cases we get surprising results. Radon appears to be protective against lung cancer. Any ozone effect on acute more mortality, if present at all, is small relative to other factors.

References

Appleton JD. (2007) Radon: sources, health risks and hazard mapping. AMBIO: A Journal of the Human Environment 36(1):85-89. doi: http://dx.doi.org/10.1579/0044-7447(2007)36[85:RSHRAH]2.0.CO;2.

Band P, Feldstein M, Saccomanno G, Watson L, King G (1980). Potentiation of cigarette smoking and radiation: evidence from a sputum cytology survey among uranium miners and controls. *Cancer* **45**:1273-1277,

Bell ML, McDermott A, Zeger SL, Samet JM, Dominici F. (2004) Ozone and Short-term Mortality in 95 US Urban Communities, 1987-2000, *Journal of the American Medical Association* 292(19):2372-2378.

Bell ML, Peng RD, Dominici F (2006) The exposure-response curve for ozone and risk of mortality and the adequacy of current ozone regulations. *Environmental Health Perspectives* **114**:532-536.

Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. (2013) Time series regression studies in environmental epidemiology. *International Journal of Epidemiology* 42:1187–1195

Cohen BL. (1989) Expected indoor 222 Rn levels in counties with very high and very low lung cancer rates. Health Physics, 57(6):897-907.

Cohen BL. (1995) Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon decay products. Health Physics. 68:157-174.

Cohen BL. (1997) Lung cancer rate vs. mean radon level in U.S. counties of various characteristics. Health Physics. 72:114-119.

Cohen BL. (2008) The linear no-threshold theory of radiation carcinogenesis should be rejected. J. Amer. Physicians and Surgeons. 13(3):70-76.

Darby S, Hill D, Deo H, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, Falk R, Farchi S, Figueiras A, Hakama M, Heid I, Hunter N, Kreienbrock L, Kreuzer M, Lagarde F, Mäkeläinen I, Muirhead C, Oberaigner W, Pershagen G, Ruosteenoja E, Rosario AS, Tirmarche M, Tomásek L, Whitley E, Wichmann HE, Doll R. (2006) Residential radon and lung cancer: detailed results of a collaborative analysis of individual data on 7148 persons with lung cancer and 14,208 persons without lung cancer from 13 epidemiologic studies in Europe. Scandinavian Journal of Work, Environment and Health 32(1):1–83.

International Agency for Research on Cancer (IARC). Man-made Mineral Fibres and Radon. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Volume 43, 1998. World Health Organization (WHO), Lyon, France.

Krewski D, Lubin JH, Zielinski JM, Alavanja M, Catalan VS, Field RW, Klotz JB, Létourneau EG, Lynch CF, Lyon JL, Sandler DP, Schoenberg JB, Steck DJ, Stolwijk JA, Weinberg C, Wilcox HB. (2006) A combined analysis of North American case-control studies of residential radon and lung cancer. Journal of Toxicology and Environmental Health, 69(7):533–597.

Lopiano KK, Obenchain RL, Young SS (2014) Fair treatment comparisons in observational research. Statistical Analysis and Data Mining 7:376-384.

Max Masnick, 2011: U.S. 2008 obesity rates at the county level. Available at: http://www.maxmasnick.com/2011/11/15/obesity_by_county/ (accessed in July 2015).

Milojevic A, Wilkinson P, Armstrong B, Bhaskaran K, Smeeth L, Hajat S. (2014) Short-term effects of air pollution on a range of cardiovascular events in England and Wales: case-crossover analysis of the MINAP database, hospital admissions and mortality. *Heart* 100:1093-1098.

National Cancer Institute (NCI). Cancer Mortality Maps – U.S. National Institutes of Health (NIH), 2015a. Available at: http://ratecalc.cancer.gov/ratecalc/ (accessed in July 2015).

National Cancer Institute (NCI). Small Area Estimates for Cancer Risk Factors and Screening Behaviors – Ever Smoking Prevalence (Age 18+). U.S. National Institutes of Health (NIH), 2015b. Available at: http://sae.cancer.gov/estimates/lifetime.html (accessed in July 2015).

National Research Council (NRC). Committee on Health Risks of Exposure to Radon: BEIR VI. Health Effects of Exposure to Radon. Washington, DC: National Academy Press, 1999.

Obenchain, RL. (2009) SAS macros for local control (phases one and two), Observational Medical Outcomes Partnership (OMOP), Foundation for the National Institutes of Health (Apache 2.0 License). http://members.iquest.net/~softrx

Obenchain, RL. (2010) The local control approach using JMP. In Analysis of observational health care data using SAS, ed. D. E. Faries, A. C. Leon, J. M. Haro, and R. L. Obenchain, 151–192. Cary, NC, SAS Press

Obenchain RL, Young SS. (2013) Advancing Statistical Thinking in Observational Health Care Research, Journal of Statistical Theory and Practice 7:2:456-469.

SAS JMP Add-In Moving Median. Personal communication, S. S. Young.

SAS JMP Partial Correlation. https://community.jmp.com/docs/DOC-6173

Smith RL, Xu B, Switzer PP. (2009) Reassessing the relationship between ozone and short-term mortality in U.S. urban communities. *Inhal Toxicol* 29(S2):37–61.

U.S. Census Bureau – American Fact Fider: 2000 Census of Population and Housing. Available at: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_00_SF1_DP1 (accessed in July 2015)

U. S. Census Bureau. Small Area Income and Poverty Estimates. U.S. Department of Commerce, 2015. Available at: http://www.census.gov/did/www/saipe/data/statecounty/data/ (accessed in July 2015).

U.S. Environmental Protection Agency (EPA). Screening indoor radon data from the State Residential Radon Survey (SRRS), 2014 (obtained through personal communication from the U.S. EPA – Radiation & Indoor Environments Division).

U.S. Environmental Protection Agency (EPA). Health Risk of Radon, 2015. Available at: http://www2.epa.gov/radon/health-risk-radon (accessed in October 2015).

Wikipedia. Hormesis. https://en.wikipedia.org/wiki/Hormesis

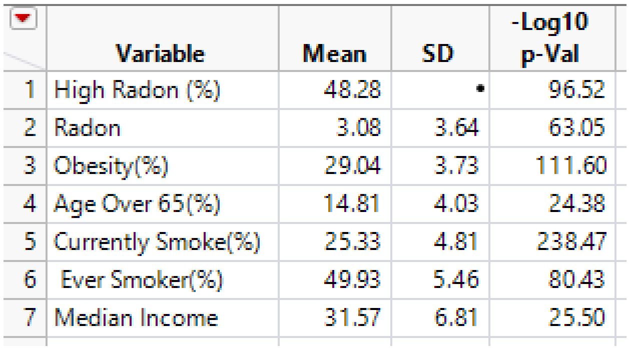
Wolfinger and Obenchain (2014) https://community.jmp.com/docs/DOC-7453

Young SS, Fogel P. (2014) Air pollution and daily deaths in California. Proceedings, 2014 Discovery Summit. https://community.jmp.com/docs/DOC-6691/

Young SS, Xia JQ. Assessing geographic heterogeneity and variable importance in an air pollution data set. *Statistical Analysis and Data Mining.* 2013;6, 375-386.

Tables and Figures

Table 1. Variables used in radon analysis, n=2881.



	Variable	Mean	SD	-Log10 p-Val
1	High Radon (%)	48.28	•	96.52
2	Radon	3.08	3.64	63.05
3	Obesity(%)	29.04	3.73	111.60
4	Age Over 65(%)	14.81	4.03	24.38
5	Currently Smoke(%)	25.33	4.81	238.47
6	Ever Smoker(%)	49.93	5.46	80.43
7	Median Income	31.57	6.81	25.50

Figure 1. Time series plots of daily deaths before and after application of a moving median time series smoother.

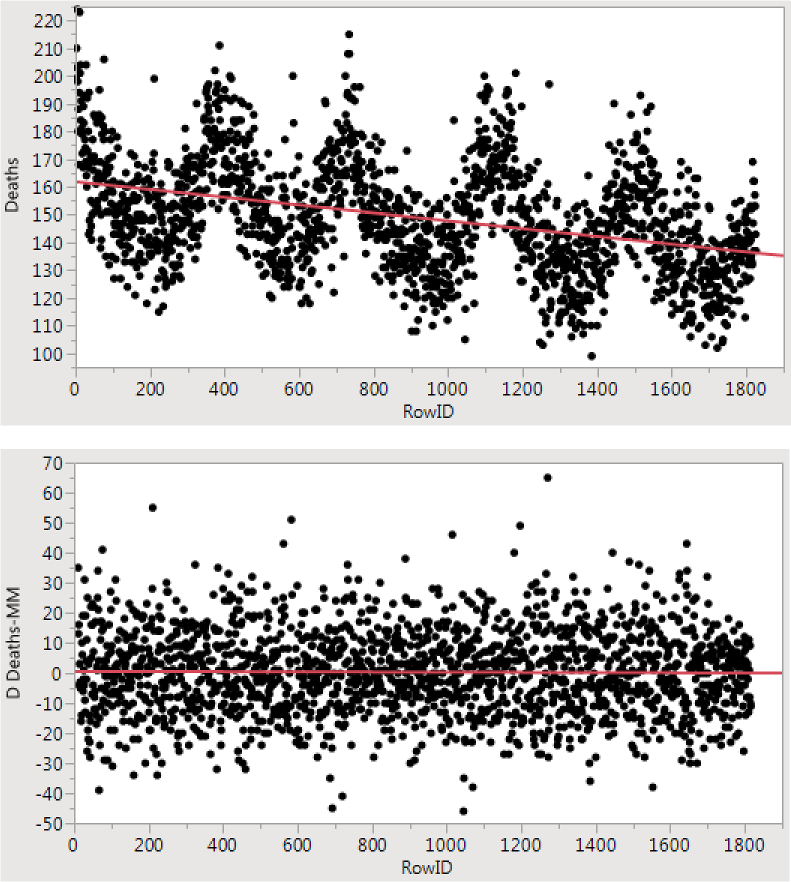
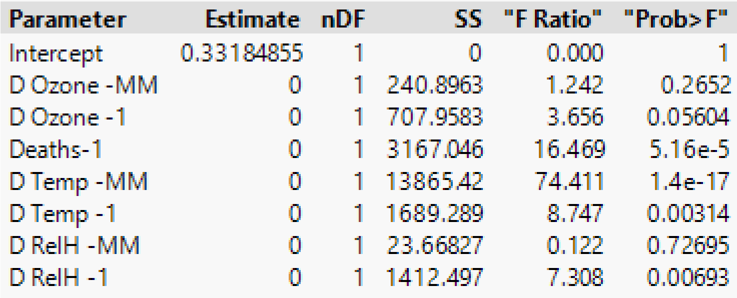


Figure 2. Simple Linear Regression predicting D_Deaths, deviation of daily mortality from moving median.



Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	0.33184855	1	0	0.000	1
D Ozone -MM	0	1	240.8963	1.242	0.2652
D Ozone -1	0	1	707.9583	3.656	0.05604
Deaths-1	0	1	3167.046	16.469	5.16e-5
D Temp -MM	0	1	13865.42	74.411	1.4e-17
D Temp -1	0	1	1689.289	8.747	0.00314
D RelH -MM	0	1	23.66827	0.122	0.72695
D RelH -1	0	1	1412.497	7.308	0.00693

Figure 3. Random and observed Local Treatment Differences. Note that virtually all of the observed LTDs are negative, indicating a protective effect.

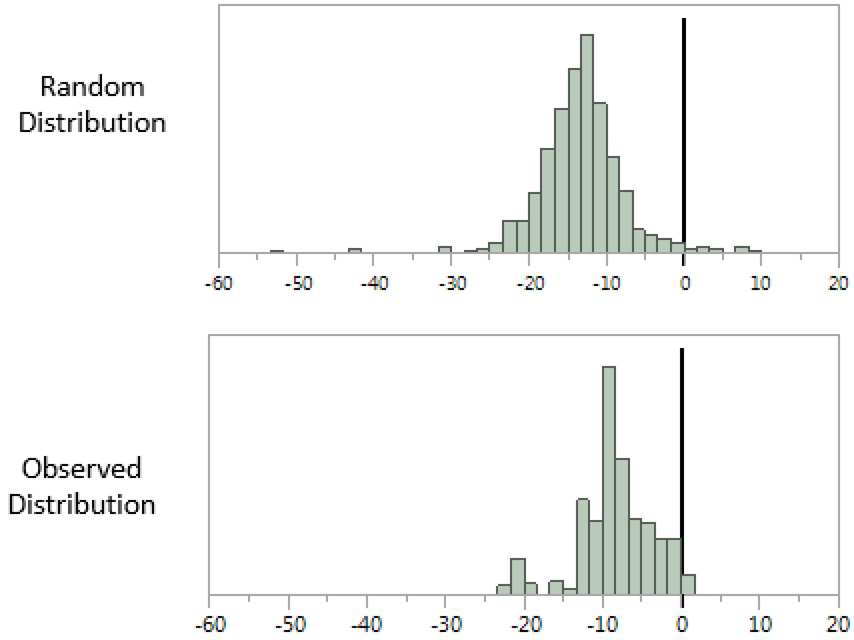
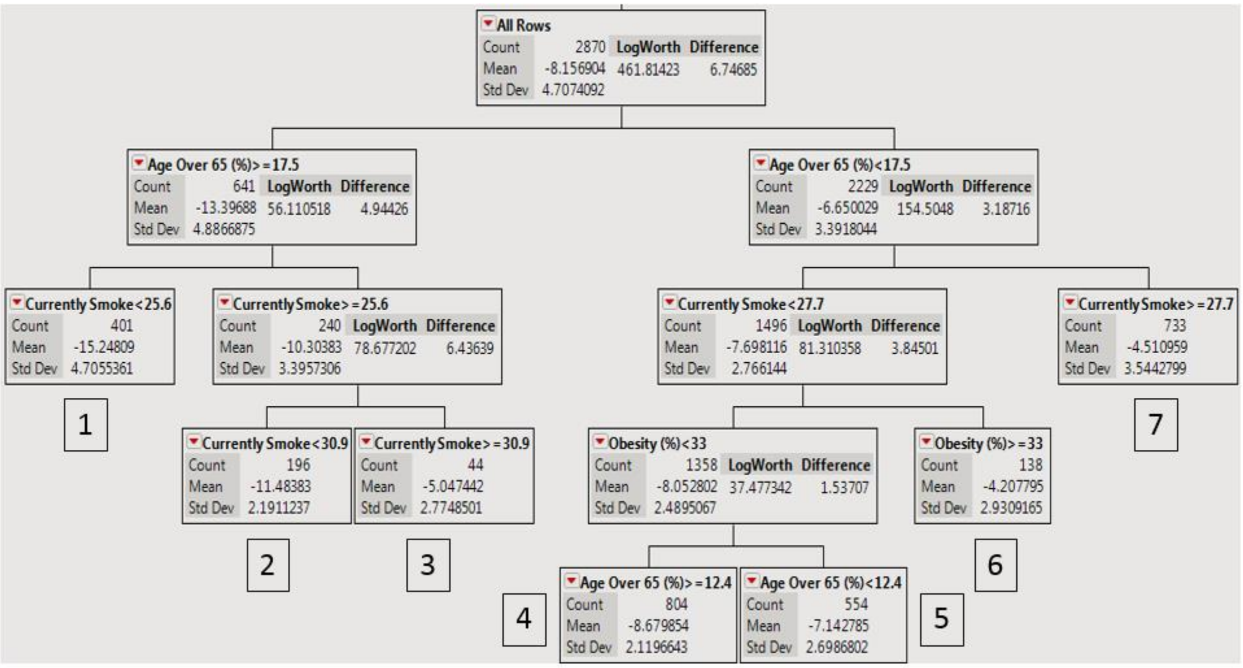


Figure 4. Recursive Partitioning analysis of radon data.



All Rows
Count 2870 LogWorth 461.81423 Difference 6.74685
Mean -8.156904
Std Dev 4.7074092

Age Over 65 (%)>=17.5
Count 641 LogWorth 56.110518 Difference 4.94426
Mean -13.39688
Std Dev 4.8866875

Age Over 65 (%)<17.5
Count 2229 LogWorth 154.5048 Difference 3.18716
Mean -6.650029
Std Dev 3.3918044

Currently Smoke<25.6
Count 401
Mean -15.24809
Std Dev 4.7055361

Currently Smoke>=25.6
Count 240 LogWorth 78.677202 Difference 6.43639
Mean -10.30383
Std Dev 3.3957306

Currently Smoke<27.7
Count 1496 LogWorth 81.310358 Difference 3.84501
Mean -7.698116
Std Dev 2.766144

Currently Smoke>=27.7
Count 733
Mean -4.510959
Std Dev 3.5442799

Currently Smoke<30.9
Count 196
Mean -11.48383
Std Dev 2.1911237

Currently Smoke>=30.9
Count 44
Mean -5.047442
Std Dev 2.7748501

Obesity (%)<33
Count 1358 LogWorth 37.477342 Difference 1.53707
Mean -8.052802
Std Dev 2.4895067

Obesity (%)>=33
Count 138
Mean -4.207795
Std Dev 2.9309165

Age Over 65 (%)>=12.4
Count 804
Mean -8.679854
Std Dev 2.1196643

Age Over 65 (%)<12.4
Count 554
Mean -7.142785
Std Dev 2.6986802

Figure 5. Partial correlation diagram for radon data.

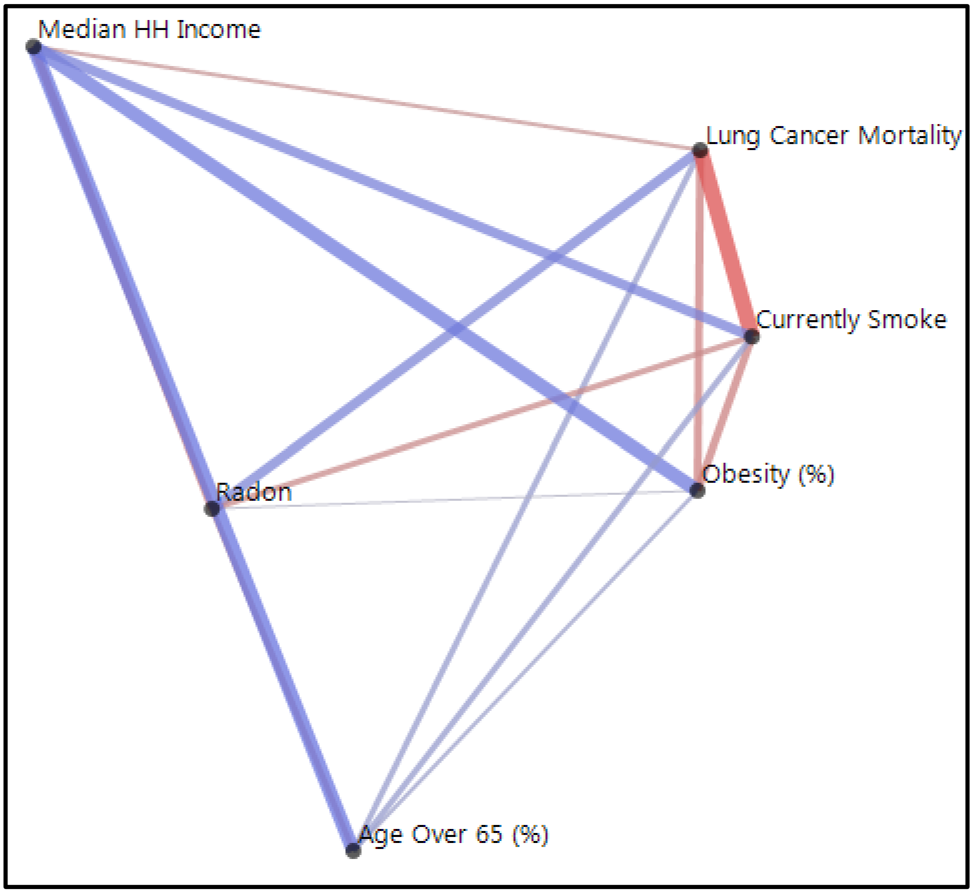


Figure 6. Results for ozone intercept analysis.

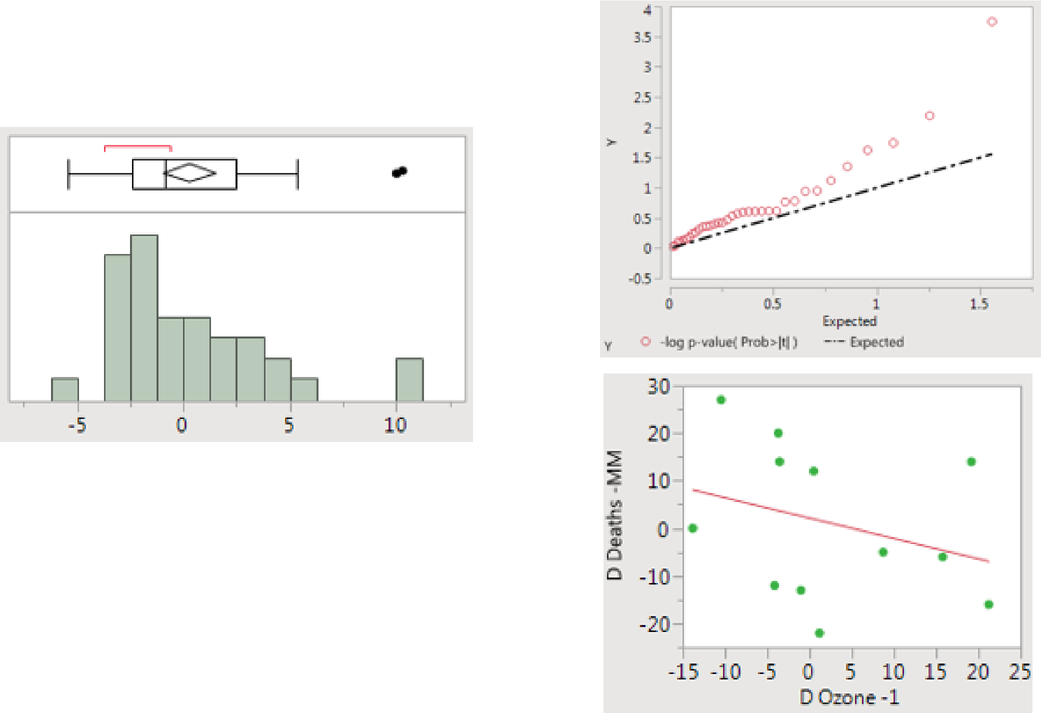


Figure 7. Results for ozone slope analysis.

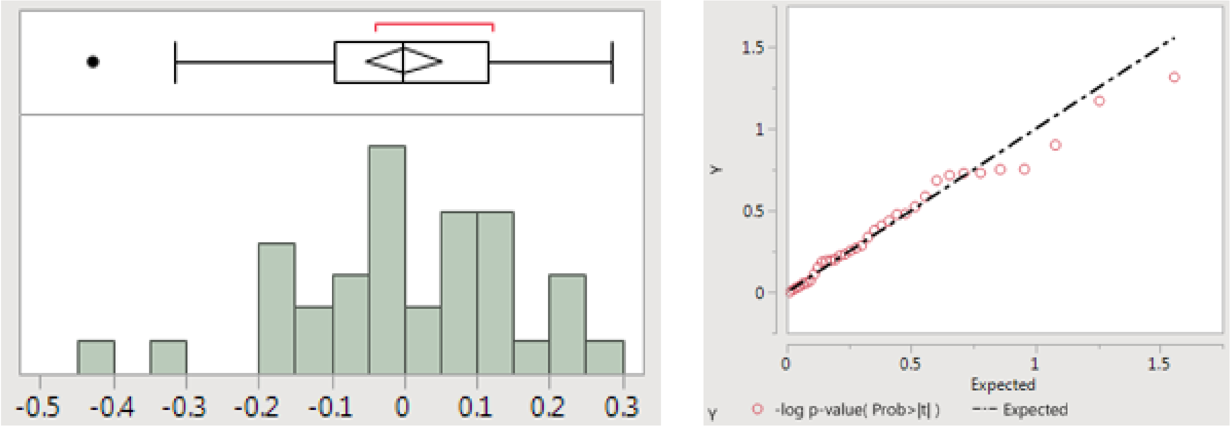
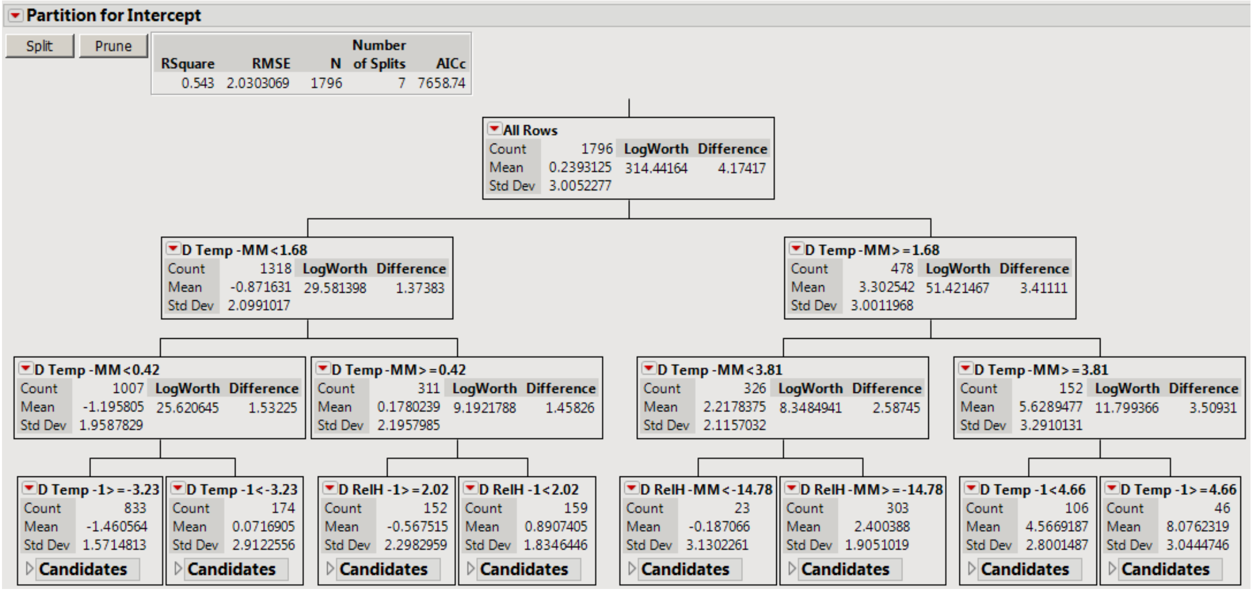


Figure 8. Recursive Partitioning analysis of ozone for intercept.



Partition for Intercept

RSquare	RMSE	N	Number of Splits	AICc
0.543	2.0303069	1796	7	7658.74

All Rows

Count	1796	LogWorth	Difference
Mean	0.2393125	314.44164	4.17417
Std Dev	3.0052277		

D Temp -MM<1.68

Count	1318	LogWorth	Difference
Mean	-0.871631	29.581398	1.37383
Std Dev	2.0991017		

D Temp -MM>=1.68

Count	478	LogWorth	Difference
Mean	3.302542	51.421467	3.41111
Std Dev	3.0011968		

D Temp -MM<0.42

Count	1007	LogWorth	Difference
Mean	-1.195805	25.620645	1.53225
Std Dev	1.9587829		

D Temp -MM>=0.42

Count	311	LogWorth	Difference
Mean	0.1780239	9.1921788	1.45826
Std Dev	2.1957985		

D Temp -MM<3.81

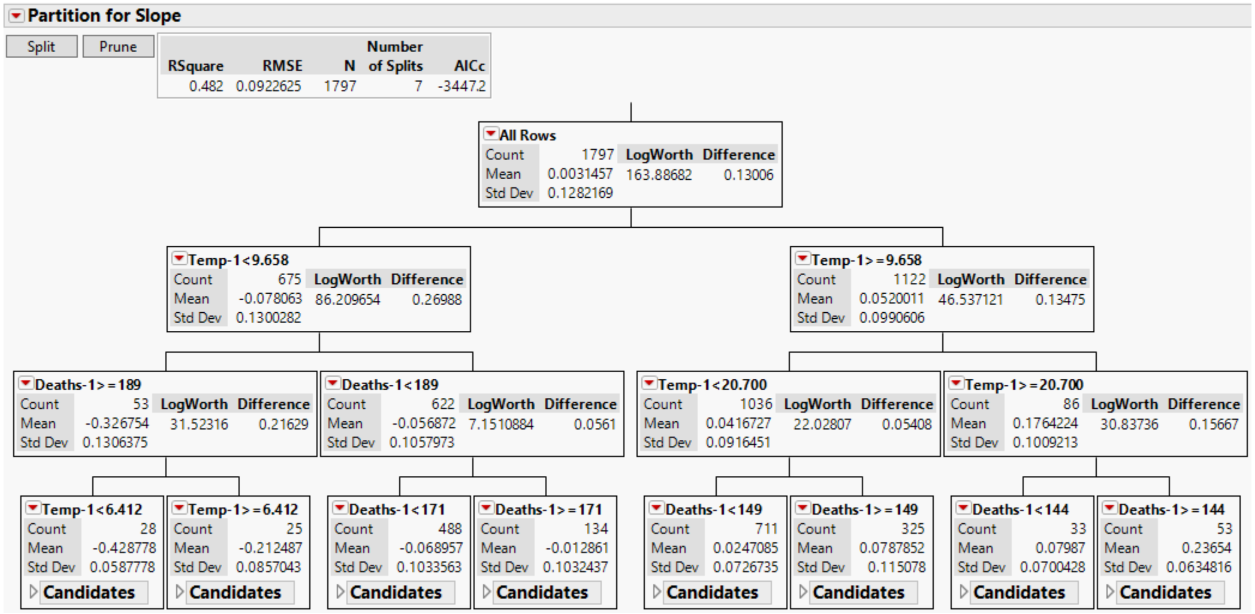
Count	326	LogWorth	Difference
Mean	2.2178375	8.3484941	2.58745
Std Dev	2.1157032		

D Temp -MM>=3.81

Count	152	LogWorth	Difference
Mean	5.6289477	11.799366	3.50931
Std Dev	3.2910131		

	D Temp -1>=-3.23	D Temp -1<-3.23	D RelH -1>=2.02	D RelH -1<2.02	D RelH -MM<-14.78	D RelH -MM>=-14.78	D Temp -1<4.66	D Temp -1>=4.66
Count	833	174	152	159	23	303	106	46
Mean	-1.460564	0.0716905	-0.567515	0.8907405	-0.187066	2.400388	4.5669187	8.0762319
Std Dev	1.5714813	2.9122556	2.2982959	1.8346446	3.1302261	1.9051019	2.8001487	3.0444746
	Candidates	Candidates	Candidates	Candidates	Candidates	Candidates	Candidates	Candidates

Figure 9. Recursive Partitioning analysis of ozone data for slope.



Partition for Slope

RSquare	RMSE	N	Number of Splits	AICc
0.482	0.0922625	1797	7	-3447.2

All Rows

Count	1797	LogWorth	Difference
Mean	0.0031457	163.88682	0.13006
Std Dev	0.1282169		

Temp-1<9.658

Count	675	LogWorth	Difference
Mean	-0.078063	86.209654	0.26988
Std Dev	0.1300282		

Temp-1>=9.658

Count	1122	LogWorth	Difference
Mean	0.0520011	46.537121	0.13475
Std Dev	0.0990606		

Deaths-1>=189

Count	53	LogWorth	Difference
Mean	-0.326754	31.52316	0.21629
Std Dev	0.1306375		

Deaths-1<189

Count	622	LogWorth	Difference
Mean	-0.056872	7.1510884	0.0561
Std Dev	0.1057973		

Temp-1<20.700

Count	1036	LogWorth	Difference
Mean	0.0416727	22.02807	0.05408
Std Dev	0.0916451		

Temp-1>=20.700

Count	86	LogWorth	Difference
Mean	0.1764224	30.83736	0.15667
Std Dev	0.1009213		

	Temp-1<6.412	Temp-1>=6.412	Deaths-1<171	Deaths-1>=171	Deaths-1<149	Deaths-1>=149	Deaths-1<144	Deaths-1>=144
Count	28	25	488	134	711	325	33	53
Mean	-0.428778	-0.212487	-0.068957	-0.012861	0.0247085	0.0787852	0.07987	0.23654
Std Dev	0.0587778	0.0857043	0.1033563	0.1032437	0.0726735	0.115078	0.0700428	0.0634816
	Candidates	Candidates	Candidates	Candidates	Candidates	Candidates	Candidates	Candidates