Assessing Variable Importance in Environmental Observational Studies

S. Stanley Young and Jessie Q. Xia
National Institute of Statistical Sciences
Research Triangle Park, NC 27709

Abstract: In environmental observational studies, often authors do not address the relative importance of variables under consideration, choosing instead to concentrate on specific claims of significance. Yet good policy decisions require knowledge of the magnitude of relevant effects. In this paper we examine data on the relationship between air quality and mortality in the United States. The analysis uses two methods for determining variable importance, regression analysis and recursive partitioning, showing how this puts predictor variables into a context that supports better environmental policy-making. In particular, using both regression and recursive partitioning, we are able to confirm a spatial interaction with the air quality variable PM2.5, a critical variable in this application domain. We also determine the relative importance of this variable in comparison to others used in air pollution research. We show that there is no association between PM2.5 and mortality west of Chicago and that where there is an association between decreased PM2.5 and increase longevity, it is much less important than other variables such as income and smoking. Our findings point to somewhat different policy recommendations from those developed by previous researchers.

Introduction

The current policy paradigm is that air pollution, as measured by small particles (those less than 2.5 micrometers, or PM2.5), is killing people and that it needs to be brought under further regulatory control. The Environmental Protection Agency, EPA, speaks of more than 160,000 annual deaths attributable to PM2.5; see News Release (2011). The EPA bases its case almost entirely on statistical analysis of observational data.

Pope *et al.* (2009) cited eight studies (their References 4-11), saying,

> "Associations between long-term exposure to fine particulate air pollution and mortality have been observed...more recently, in cohort-based studies. ... all support the view that relatively prompt and sustained health benefits are derived from improved air quality."

On the other hand, Enstrom (2005), after citing papers supporting an association says, "Other cohort studies have also examined mortality associations with PM2.5 and other pollutants … with somewhat different findings." There were eight papers that Dr. Pope - referred to supporting an association between pollution, PM2.5, and statistical deaths, and four papers that Enstrom referred to that cast doubt on the claim. Peng *et al.* (2006) commented, "For example, in air pollution epidemiology, the national relative risk of increased mortality is estimated to be 1.005 per 10 parts per billion of 24-hour ozone. …. Nevertheless, the potential for unexplained confounding cannot be denied for such a small relative risk."

When this controversy was breaking in the early 1990s, the EPA asked the National Institute of Statistical Sciences to evaluate data from two cities, see Stayer *et al.* (1995). Stayer *et al.* commented on some of the difficulties, saying "The data used in the analyses (meteorological conditions, particulate levels, death counts) are observational; that is, data that are measured and recorded without control or intervention by researchers. Deducing causal relationships from observational data is perilous. A practical approach described by Mosteller and Tukey involves considerations beyond regression analysis. In particular, consideration should be given to whether the association between particulate levels and mortality is consistent across 'settings,' whether there are plausible common causes for elevated particulate levels and mortality, and whether the derived models reflect reasonable physical relationships." They then concluded, "...that the reported effects of particulates on mortality are unconfirmed." Essentially noting the same and additional difficulties, Smith *et al.* (2009) agreed that the case for a significant association of low-level air pollution with statistical deaths was unproven: "In summary, it is our view that estimates of the association between ozone and mortality, based on time-series epidemiologic analyses of daily data from multiple cities, reveal important still-unexplained inconsistencies and show sensitivity to modeling choices and data selection. These inconsistencies and sensitivities contribute to serious uncertainties when epidemiological results are used to discern the nature and magnitude of possible ozone-mortality relationships or are applied to risk assessment."

Krewski *et al.* (2000) noted that if there are effects, they are heterogeneous, i.e., varying across the US. Smith *et al.* (2009), using complex methods for ozone levels, also noted that the effects were not constant across the U.S. In geo-maps, from both groups, there are hot spots and vast areas where any affect of air pollution on mortality appears minimal to non-existent. Even Krewski (2010) could find no association between PM2.5 and mortality in California.

Dr. Pope generously provided the data used in his 2009 *New England Journal of Medicin*e paper. Using this data set, we address two questions: Is there evidence for effect geographic heterogeneity and what is the relative importance of air pollution relative to other factors? In particular, is there evidence for differential effects in the western U.S. as opposed to the eastern U.S. Enstrom (2005) finds no effect in California, with a relative risk of 1.00 and confidence limits of 0.98-1.02. His results are confirmed by CARB consultant Professor Jerrett (2010), with a relative risk of 1.00 and confidence limits of 0.97-1.03. We computed multiple analyses sweeping across the county from west to east and show that one can "cut" along the longitude passing just west of Chicago and find no effect of PM2.5 to the west and a small association of PM2.5 on statistical deaths to the East. Both Stayer *et al.* (1995) and Smith *et al.* (2009) make the logical point that if the effect of the pollutant is not consistent, then it is unlikely it is a causative agent.

The Pope *et al.* (2009) based their interpretation on a main effects only analysis, i.e. PM2.5 is a statistically significant cause of deaths uniformly across the US. Based on our analysis, PM2.5 exhibits different associations with mortality in the eastern and western U.S., which suggests that PM2.5 per se is not a causative agent and that a single national policy is therefore not appropriate across the entire country. In any case, the relative importance of PM2.5 to statistical mortality, as compared to other factors, should be taken into account by decision-makers. We will show that PM2.5 exerts much less influence on longevity than income, for example.

Data

Pope *et al.* (2009) started with 2068 county units from which 215 county units in metropolitan areas were selected that had matching PM2.5 data available. New York areas were consolidated, so ultimately there were 211 records for 51 metropolitan areas within the US. The response variable was the change in age-adjusted mortality from the early 1980s to the late 1990s. There were ten predictor variables; see Table 1. Note that the change in PM2.5 is the same for each unit within a metropolitan area.

Table 1. Variables use by Pope *et al.* for regression analysis. All variable are given as change from years ~1980 to ~2000.

Variable	Comment
Life Expectancy, life-table methods	Response variable.
Per capita income (in thousands of $)	Inflation adjusted to the year 2000.
Lung Cancer (Age standardized death rate)	Surrogate for smoking.
COPD (Age standardized death rate)	Surrogate for smoking. COPD denotes chronic obstructive pulmonary disease
High-school graduates (proportion of population)	
PM2.5 (μg/m3)	Particulate matter, aerodynamic diameter < or = 2.5 μm.
Black population (proportion of population)	Self reported.
Population (in hundreds of thousands)	
5-Year in-migration (proportion of population)	Five-year in-migration refers to the proportion of the population who did not reside in the county 5 years earlier.
Hispanic population (proportion of population)	Self reported.
Urban residence (proportion of population)	

Methods

Introduction

We use two methods of model fitting: linear regression and recursive partitioning. Regression makes relatively strong modeling assumptions whereas recursive partitioning is more non-parametric. For example it is robust to non-linear relationships.

Regression

The linear regression model of the following form is considered:

$$Y=\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_{10} X_{10}+e \quad (1)$$

where Y represents the change of life expectancy from the early 1980s to the late 1990s; X_1 to X_{10} represent the ten covariates given in Pope *et al.* (2009), including the changes of PM2.5, income, high school graduate rate, and two proxy indicators for smoking etc, which are listed in Table 1. The residuals ε are assumed to have an independent, identical Gaussian distribution with mean of 0 and variance of σ^2.

Step-wise regression

For the purpose of either variable selection or variable importance assignment, step-wise regressions are often conducted. In the forward selection mode, the simplest model without any regression variables is first fitted: $Y=\beta_0+e$. Then one regression variable is added to the model, forming a second model: $Y=\beta_0+\beta_i X_i+\varepsilon$. The decrease of the residual sum of squares r_i is assigned to the regression variable X_i. This procedure is repeated, until all ten variables enter the 11th model. Eventually, there will be a vector of residual sum of square (r_1, r_2, …, r_{10}) for the 10 regression variables. If all regression variables are independent, the vector will be unique regardless of the sequence of the variables entering the linear model in Equation (1). However, if there are correlated variables, their r_i values will depend on the order that the variable enter the linear model.

Regression variable importance
Variable importance estimates are achieved by decomposing *var(Y)* into the parts attributable to the individual X_i's. There are several methods of variable importance assignment based on the linear regression models shown in Equation (1), as described in Lindeman *et al.* (1980) and Grömping (2006, 2007, 2009). In our paper, we use the method proposed by Lindeman, Merenda, and Gold (LMG). It considers all the 10! = 3,628,800 permutations of the stepwise regression using the 10 regression variables. The method is computationally intensive, but there is free code to do the analysis in R, see Grömping (2006). Let ($r_1^{(k)}$, $r_2^{(k)}$, …, $r_{10}^{(k)}$) represents the k^{th} variable importance assignment for the regression variables, the final variance importance assignment is just the average importance over all the permutations:

$$\bar{r}_i = \frac{1}{10!}\sum_{k=1}^{10!} r_{i^{(k)}},\ i=1,\ldots,10 \quad . \tag{2}$$

For two correlated variables, a single stepwise regression will diminish the relative importance of the variable that enters the regression model at a later time, whereas the LMG method averages across all possible full-term stepwise regressions and assigns a more balanced value of importance to both variables.

Single tree
Recursive partitioning (RP) is a data mining method useful for uncovering complicated relationships in large, complex data sets. These relationships may involve thresholds, interactions, and nonlinearities. Any or all of these relationships hinder an analysis based on the standard assumptions in multiple linear regression. RP was originally designed for automatic interaction detection; see Morgan and Sonquest (1963). The method has been subject to much development and is widely used for complex modeling situations; see Hawkins (2009) for a short review. The basic analysis strategy of recursive partitioning is simple and easily understood by example. Consider an analysis of the Pope data set for the eastern US; see Figure 1.

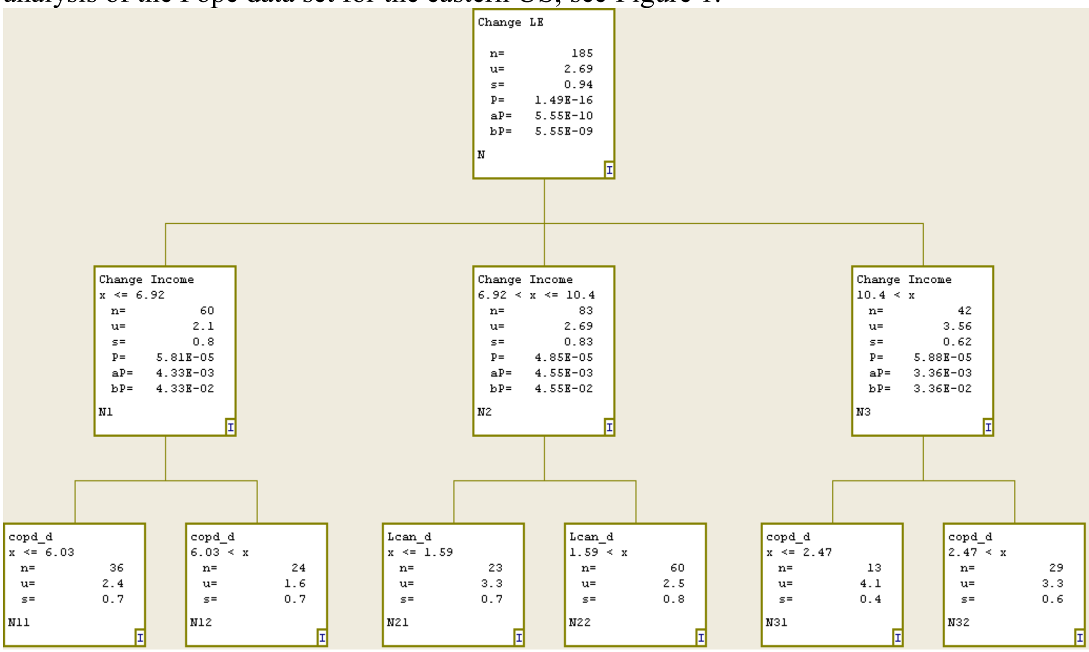

Figure 1. Recursive Partitioning analysis selects the best predictor, Change in Income, and makes two "cuts" splitting the predictor into three groups with Life Expectancy increasing with increased income. Each of the three nodes is split in turn by variables that are surrogates for smoking, Lung Cancer and COPD. The difference of Life Expectance from the node with lowest increase in income to the highest is about 1.5

years. Lung Cancer and COPD confer about 0.8 years in increased life expectancy. Three p-values are given: the raw p-value, P, is unadjusted; aP is adjusted for the number of ways to cut the predictor into categories; bP is adjusted for cuts and variables available for making a cut.

The 185 observations from the Eastern US are in the top node, denoted by N. Also given within a node are summary statistics that show the mean (u), standard deviation (s) and p-values used in the splitting process. All potential predictor variables are examined and the variable with the smallest adjusted p-value is used to split the node into two or more daughter nodes. In this case, Change Income is the variable with the smallest adjusted p-value. Segmentation is used to find the optimal "cut points", making in this case three daughter nodes, denoted by N1, N2 and N3 respectively. The p-value for this cut is adjusted to reflect the number of variables available and the number of ways the segmentation can be done, as well as the number and placement of the cuts. Each of the daughter nodes is examined in turn and is split if significant. Nodes N1 and N3 use COPD to split and Node N2 is split using Lung Cancer. Pope *et al.* (2009) used both COPD and Lung Cancer as surrogates for smoking. Each node is split in turn and the recursive splitting stops when there are no statistically significant splits to be made. Notice that at each level of the tree building that the standard deviation in each node gets smaller as splitting progresses. Tracing from N to N1 to N11 we see the standard deviations decrease as 0.94, 0.8 and 0.7, respectively.

Multiple trees

There are advantages (more accurate predictions and the ability to assess variable importance) to computing and using multiple trees in the analysis of a data set; see Hawkins and Musser (1999, 2001). Multiple trees can be computed by sampling with replacement multiple random samples from the data set and computing a tree for each such sample; see Breiman (2001). Alternatively, at a split, one can randomly sample one of the valid split variables to make the split; see Hawkins and Musser (1999, 2001). Once one has multiple trees, they can be used to determine variable importance. One can compute how often a variable is used over all the multiple trees. Alternatively, the split variable controls all the samples below it so, across the multiple trees, the fraction of the observations controlled by a variable can be computed. The latter method is used by Optimus Recursive Partitioning (Optimus RP) from Golden Helix (Bozeman, MT) and we report its results.

Results

It is perhaps not appreciated by the general scientific community, but it is well-known among experts that air quality has a differential effect on mortality in eastern and western U.S. with essentially no effect in the west; see Krewski *et al.* (2000), Enstrom (2005), Smith *et al.* (2009), Jerrett (2010). As these results are based on several data sets with analyses done by several teams of investigators independently, the no-detectable-effect on mortality of PM2.5 in the West appears to be real. One explanation is that PM2.5 is based on physical particle size, not specific chemical composition. Bell *et al.* (2007) report that there is both temporal and spatial variation in the chemical composition of PM2.5. With the dataset of Pope *et al.* (2009), we confirm the geographic heterogeneity of PM2.5 health effects, and that there is no detectable effect in the western U.S. Figure 2 gives scatter plots of change in Life Expectancy versus PM2.5 for the eastern and western U.S.

(a) East (b) West

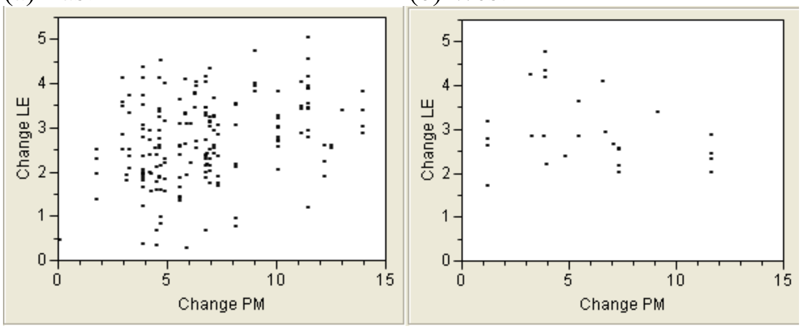

Figure 2. Change in Life Expectancy(Change LE), in (a) East US is positive with respect to Change PM2.5 whereas it is not statistically significant in (b) West US.

A linear regression for the eastern and western subsets finds a significant increase in mortality for the East, but not for the West; the slopes for the two regression lines (not shown) are significantly different from one another, with p-value 0.0063. To better understand the effect of PM2.5 across the US, we computed the regression of longevity on PM2.5, stepping across the U.S. from west to east and from east to west and we give the slope of the regression line as we go (figure not shown). So the Pope *et al.*'s claim that life expectancy increases with a decrease in PM2.5 is supported in eastern U.S., but is not supported in western U.S.

Note that Change in Income has a more dramatic affect on increase in life expectancy than reduction in PM2.5, Figure 3.

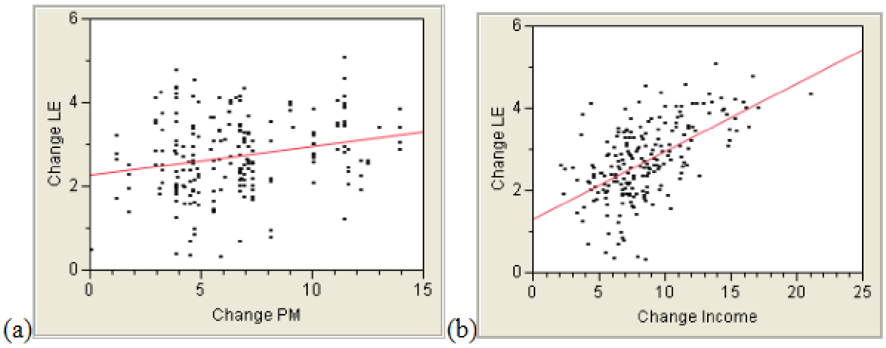

Figure 3. Change in Life Expectancy(Change LE), versus (a) change in PM2.5 and (b) Change in Income.

Also note change in income appears equally effective in East and West, Figure 4.

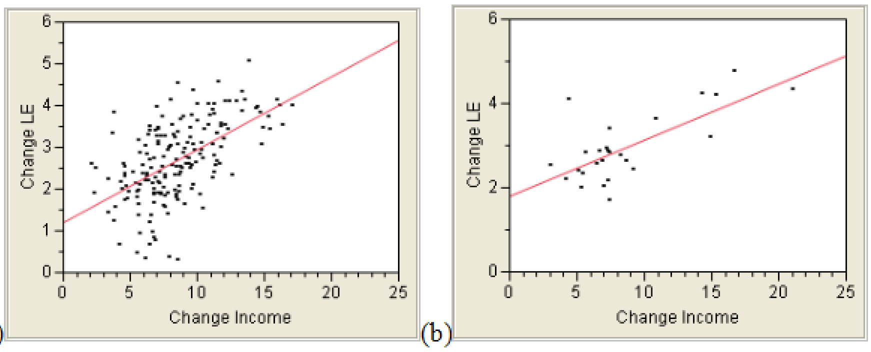

Figure 4. Change in Life Expectancy in (a) eastern and (b) western US increases with increased income.

Variable importance for the eastern U.S. and western U.S. is computed using the regression method of LMG and the recursive partitioning method in Optimus RP. The variable importance results are given in **Table 2**.

Variable importance for the eastern U.S. and western U.S. is computed using the regression method of LMG and the recursive partitioning method in Optimus RP. The variable importance results are given in Table 2.

	Regression			Recursive Partitioning		
Variable	East	West	US	East	West	US
Income	0.2792	0.3996	0.3390	0.2865	1.0000	0.2108
COPD	0.1789	0.0216	0.1621	0.2298	0.0000	0.1199
LungCancer	0.1697	0.1806	0.1768	0.2385	0.0000	0.1467
PM2.5	**0.1095**	**0.0299**	0.0732	**0.1118**	**0.0000**	0.1302
HighSchool	0.1013	0.0859	0.0997	0.1097	0.0000	0.1066
%Black	0.0620	0.1250	0.0537	0.0000	0.0000	0.0319
PopDensity	0.0370	0.1171	0.0418	0.0000	0.0000	0.0793
%Hispanic	0.0281	0.0065	0.0177	0.0237	0.0000	0.0136
Migration	0.0240	0.0120	0.0228	0.0000	0.0000	0.0202
Urban	0.0103	0.0217	0.0133	0.0000	0.0000	0.0105

Table 2. Variable importance, ranked by importance in East US. Importance by linear regression uses all 10! permutations of the order of the variables. “Recursive Partitioning” is importance by Recursive Partitioning using 1000 trees. Note, in bold, the differences in importance of PM2.5 in East and West. PM2.5 has little or no importance in the West.

The predictor variables are given in the order of their importance in multiple linear regression in the eastern U.S. Increase in income is the most important variable for predicting improved mortality, in both eastern and western U.S., and for both the regression and the RP variable importance methods. Lung Cancer and COPD are about equally important in the eastern U.S. COPD and PM2.5 are relatively unimportant in the western U.S. The Percent Graduating from High School and PM2.5 are about equally important in the eastern U.S. Regression analysis indicates that %Black and Population Density are important in the western U.S., but not very important in the eastern U.S. Both regression and RP put the importance of PM2.5 in fourth place among the predictors, and roughly equal in importance to a high school education. Both linear regression and recursive partitioning indicate that PM2.5 is unimportant in the western U.S.

Discussion

The problems of observational studies have been well-known for many years; see Mayes *et al.* (1988) and Feinstein (1988) for discussion. But there has been little or no progress in adopting better methods; see Pocock *et al.* (2004) and Boffetta *et al.* (2008). The end result is that most claims that are based on observational data fail to replicate on retesting; see Ioannidis (2005) and Young and Karr (2011).

The association between PM2.5 with mortality, when compared to the associations between other variables and mortality shows that the importance of PM2.5 is relatively small. There is no measurable association in the western U.S., although it accounts for about 11% of the variance in the eastern U.S.

All analysis indicates that changes in income and several other variables are more influential than PM2.5, so policy makers might better focus on improving the economy, reducing cigarette smoking, and encouraging people to pursue education.

However, there are bigger lessons in this analysis. In 1985, Richard Feynman (1997) said, "In summary, the idea is to try to give all the information to help others to judge the value of your contribution; not just the information that leads to judgment in one particular direction or another." In sharp contrast, Glaeser (2006), describing what he considers the current state of affairs, says, "Certainly, there is no sense in which standard techniques have been adjusted to respond to researcher incentives. … The same incentives that induce researchers to data mine will induce them to avoid techniques that appropriately correct for that data mining." And there is a need to know "how to adjust statistical inference for researcher initiative." Researcher initiative for Glaeser is the use of data selection, statistical methods and language to support the author's narrative. Ioannidis (2008) points to similar researcher initiative issues in epidemiology studies, saying "However, this is counter-intuitive to the discovery process. One makes exploratory analyses specifically to find something. The effects selected for presentation are likely to be among the largest observed, if not the largest possible."

Others discuss the problems with complex modeling. Friedrich Hayek in his Nobel Prize lecture of 1974 described the situation of complex modeling outside the area of the physical sciences where theory offers guidance on which variables need to be measured. In non-physical sciences, one might simply use available measurements. In physical sciences the number of relevant variables can be small and the relationships simple, whereas in complex biological systems both the number of variables and how they are related can be very complex, which Hayek called essential complexity.

With a large number of variables and complex relationships, laymen have essentially no ability to discern the validity of the model. Even experts will have trouble evaluating claims based on models. Debunking invalid models is difficult because the models are complex and because people and institutions tend to become invested in those models. We summarize the Hayek argument: There are multiple factors that are likely to impinge on a phenomenon of interest and many of these factors may not be measured or even measurable. Outside of the physical sciences we have little theory to guide us on what needs measuring. These unavailable factors can lead to biases that may be on the same order of magnitude as the phenomenon under study. In our case, the study of factors associated with mortality, the mechanism is one of *essential* complexity. The statistical modeling process is not simple, so even experts find it difficult to judge the validity of the analysis.

It makes good sense that any risk mitigation plan should take alternative regulatory strategies into account, Science and Decisions (2008), so knowing the relative importance of variables associated with risk is important. In this case, would money be better spent on education and anti-smoking education? Also, general economic development would likely contribute more to longevity than costly attempts to reduce PM2.5.

There are technical and operational fixes to control for researcher initiative. A good start would be for journals to require authors to clearly state how many questions are under consideration, to evaluate their relative importance, and to make data used in the publication publicly available. For simple and rather complex multiple testing questions, resampling-based adjusted p-values can be computed to adjust for the comparisons; see Westfall and Young (1993). Adjusted p-values should be given in addition to the usual unadjusted p-values. P-value plots are often helpful, Schweder and Spjøtvoll (1982). For complex modeling, the use of a hold out sample or cross-validation can be used to help judge whether the final model is reproducible. As Glaeser would predict, this simple and effective strategy is seldom used in the analysis of environmental observational studies. Bias is a bit more complicated as unmeasured variables can push the results in one direction or another. The standard defense against bias has been to require large effects, say a risk ratio over 2.0, Federal Judicial Center, (2000), or even over 3 to 4, Temple (1999). Our message is that journal reviewers and editors, consumers of scientific information, including regulators, science advisory committees, and the press, should be aware of the problem of "researcher initiative" and should not rely on papers which do not make systematic efforts to address such issues.

Data: The data used in this paper was obtained from Professor Pope and can be downloaded from www.niss.org/content/s-stanley-young

References

Bell ML, Dominici F, Ebisu K, Zeger SL, Samet JM. 2007. Spatial and temporal variation in PM2.5 chemical composition in the United States for health effects studies. Env Health Pers 115 989-995.

Boffetta P, McLaughlin JK, La Vecchia C, *et al*. 2008. False-positive results in cancer epidemiology: A plea for epistemological modesty. *J Natl Cancer Inst* **100**: 988–995.

Breiman L. 2001. Random Forests. *Machine Learning* **45**:5–32.

Enstrom JE. 2005. Fine particulate air pollution and total mortality among elderly Californians, 1973-2002. *Inhalation Toxicology* **17**: 803-816.

Federal Judicial Center, *Reference Manual on Scientific Evidence*, 2nd ed., (Federal Judicial Center, 2000), www.fjc.gov/public/home.nsf/pages/610

Feinstein AR. 1988. Scientific standards in epidemiologic studies of the menace of daily life. *Science* **242**:1257-1263.

Feynman RP. 1997. *Surely You're Joking, Mr. Feynman!* Norton paperback: p. 341.

de la Fuente A, Bing N, Hoeschele I, Mendes P. 2004. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **20**:3565–3574.

Glaeser EL. 2006. Researcher Incentives and Empirical Methods www.economics.harvard.edu/pub/hier/2006/HIER2122.pdf

Grömping, U. 2006. Relative importance for linear regression in R: The Package relaimpo. *J Statistical Software* **17**: 1. Available at *http://www.jstatsoft.org/ v17/ i01/*.

Grömping, U. 2007. Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician* **61**:139–147.

Grömping U. 2009. Variable Importance Assessment in Regression: Linear Regression versus Random Forest *The American Statistician* **63**:308-319.

Hawkins DM. 2009. Recursive partitioning. *Computational Statistics* **1**:290-295.

Hawkins DM, Musser BJ. 1999. One tree or a forest? Alternative dendrographic models. *Computing Science and Statistics* **30**:534–542.

Hawkins DM, Musser BJ. 2001. Feature selection with nondeterministic recursive partitioning, *Proceedings of the American Statistical Association* [CD-ROM] Alexandria, VA: ASA.

Hayek FA. 1974 Nobel Prize Lecture: The Pretence of Knowledge. http://nobelprize.org/nobel_prizes/economics/laureates/1974/hayek-lecture.html

Huber PW. 1991.*Galileo's Revenge, Junk science in the court room.* BasicBooks. New York, NY.

Ioannidis JPA. 2005. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* **294**:218-228.

Ioannidis JPA. 2008. Why most discovered true associations are inflated. *Epidemiology* **19:** 640-648

Jarrett M. 2010. California-specific Studies on the PM2.5 Mortality Association. See slides 12 and 13, no increase in “All causes” death rate. http://www.arb.ca.gov/research/health/pm-mort/jerrett.pdf

Kendall MG, Stuart A. 1973. *The Advanced Theory of Statistics*, Volume 2 (3rd Edition), Section 27.22.

Krewski D, Burnett RT, Goldberg MS, *et al*. 2000. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality. Part II: Sensitivity Analysis, HEI Publications. http://pubs.healtheffects.org/view.php?id=6, See Figure 21, in particular.

Krewski D. 2010. Letter from Krewski to HEI President Greenbaum. http://www.arb.ca.gov/research/health/pm-mort/HEI_Correspondence.pdf

Lindeman RH, Merenda PF, Gold RZ. 1980. *Introduction to Bivariate and Multivariate Analysis.* Glenview, IL: Scott, Foresman, New York, NY.

Mayes LC, Horwitz RI, Feinstein AR. 1988. A collection of 56 topics with contradictory results in case-control research. *Int J Epidemiolgy* **17**:680-685.

Morgan JA, Sonquest JN. 1963. Problems in the analysis of survey data, and a proposal. *J Am Stat Assoc* **58**:415-434.

News Release - Air. March 1, 2011. EPA Report Underscores Clean Air Act's Successful Public Health Protections. *Landmark law saved 160,000 lives in 2010 alone* http://yosemite.epa.gov/opa/admpress.nsf/6424ac1caa800aab85257359003f5337/f8ad3485e788be5a8525784600540649!OpenDocument

OptimusRP Predictive Analytics Software. **www.helixtree.com/Predictive_Analytics/optimus_rp**

Pocock SJ, Collier TJ, Dandreo KJ, *et al*. 2004. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ* **329**:883-888.

Pope CA III, Ezzati E, Dockery DW. 2009. Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med* **360**:376-86.

Schweder T, Spjøtvoll E. 1982. Plots of p-values to evaluate many tests simultaneously. Biometrika 69, 493–502.

Science and Decisions: Advancing Risk Assessment, National Academies, Washington D.C., 2008 http://books.nap.edu/catalog/12209.html

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

Styer P, McMillan N, Gao F, Davis J, Sacks J. 1995. Effect of outdoor airborne particulate matter on daily death counts. *Environmental Health Perspectives* **103**:490-497.

Temple R. 1999. Meta-analysis and epidemiologic studies in drug development and postmarketing surveillance. *JAMA* **281**:841-844.

Westfall PH, Young SS. 1993. *Resampling-Based Multiple Testing. Wiley, New York.*

Young SS, Karr A. 2011. Deming, data and observational studies. *Significance* **8**: 116-120.