



Air quality environmental epidemiology studies are unreliable



S. Stanley Young

CGStat, 3401 Caldwell Drive, Raleigh, NC 27607-3326, United States

ARTICLE INFO

Article history:

Received 17 December 2016

Received in revised form

6 March 2017

Accepted 7 March 2017

Available online 8 March 2017

Keywords:

Environmental epidemiology

Air quality

Mortality

Observational studies

Multiple testing

Multiple modeling

Analysis search space

ABSTRACT

Ever since the London Great Smog of 1952 is estimated to have killed over 4000 people, scientists have studied the relationship between air quality and acute mortality. There are many hundreds of papers examining the question. There is a serious statistical problem with most of these papers. If there are many questions under consideration, and there is no adjustment for multiple testing or multiple modeling, then unadjusted p-values are totally unreliable making claims unreliable. Our idea is to determine the statistical reliability of eight papers published in Environmental Health Perspectives that were used in meta-analysis papers appearing in Lancet and JAMA. We counted the number of outcomes, air quality predictors, time lags and covariates examined in each paper. We estimate the multiplicity of questions that could be asked and the number of models that could be constructed. The results were that the median numbers of comparisons possible for multiplicity, models and search space were 135, 128, and 9568 respectively. Given the large search spaces, finding a small number of nominally significant results is not unusual at all. The claims in these eight papers are not statistically supported so these papers are unreliable as are the meta-analysis papers that use them.

© 2017 Elsevier Inc. All rights reserved.

1. Background and introduction

Epidemiology exhibits a notoriously poor record with a serious lack of reproducibility of published findings going back at least as far as [Feinstein \(1988\)](#) with continuing complaints: [Taubes and Mann \(1995\)](#), [Ioannidis \(2005\)](#), [Kaplan et al. \(2010\)](#), and [Young and Karr \(2011\)](#), to name only a few. Even the popular press is taking notice of the problems; [Taubes \(2007\)](#) and [Hughes \(2007\)](#) are two examples. See also [Wikipedia \(2017\)](#) Replication crisis. Ominously, there may be actual misuse and/or even deliberate abuse of model fitting methods; see [Clyde \(2000\)](#), [Glaeser \(2006\)](#), [Young and Karr \(2011\)](#). In 2002, Norman Breslow noted that students with the same training and the same data set produced statistical models with vastly different claims, [Breslow \(2003\)](#). In 2010, two groups of researchers using the same data base of observational data found that a treatment both caused, [Cardwell et al. \(2010\)](#), and did not cause, [Green et al. \(2010\)](#), cancer of the esophagus. A Nature survey reported that 90% of scientists responding said there is crisis in science: a serious, 52%, or minor, 38%, crisis, [Baker \(2016\)](#).

The state of science is bad enough that a consumer of a science paper should start with the premise that any claim made is more

likely than not to be wrong (it will fail to replicate).

The current US Environmental Protection Agency, EPA, paradigm is that PM_{2.5} is *causal* of acute human deaths. The then head of the EPA, Lisa Jackson, said “Particulate matter causes premature death. It doesn’t make you sick. It’s directly causal to dying sooner than you should.” She went on to say “If we could reduce particulate matter to levels that are healthy we would have an identical impact to finding a cure for cancer.” Cancer causes ~570,000 deaths per year.

This report unapologetically takes the position that the current paradigm, air quality is a killer, is not supported by statistical analysis that take multiple testing and multiple modeling into account and claims made in these papers may not replicate. Papers supporting the current paradigm are many. Google Scholar, “air pollution, mortality”, returns over 900,000 hits; [Schwartz et al. \(2017\)](#) is typical. These studies are almost always associational studies, and of course, association is not proof of causation. To examine our claim that the EPA paradigm is wrong, we start with two recent meta-analysis papers that look at air quality and mortality effects, [Nawrot et al. \(2011\)](#), [Mustafic et al. \(2012\)](#), hereafter Lancet and JAMA. Eight of the base papers used in these meta-analysis studies were published in Environmental Health Perspective, EHP; we examine those papers. Our thesis is that these papers are statistically flawed and that they may be part of a publication bias.

E-mail address: genetree@bellsouth.net.

A major contribution of this research is to show that a seriously flawed analysis strategy is used in these eight EHP papers rendering claims made in these papers unsupported.

2. Methods

In randomized clinical trials, RCTs, there is very careful attention given to the statistical analysis. A statistical protocol is developed and agreed to by the interested parties, often a drug company and the US FDA, before the study starts. One of the major concerns is the control of statistical false positive results. Statistical, experimental and managerial strategies are employed to control the false positive rate. Often replication of a finding is required. Contrast a RCT with the typical environmental observational study, EO. Environmental epidemiology essentially has few, if any, analysis requirements. In an EO study, the researcher can modify the analysis as the data is examined. Multiple outcomes can be examined, multiple variables (air components) can be used as predictors. The analysis can be adjusted by putting multiple covariates into and out of the model. It is thought that effects can be due to events on prior days so different lags can be examined. For example, PM2.5 yesterday or the day before can cause deaths today. Seldom, if ever, is there a written, statistical protocol prior to examination of the data. With these factors (outcomes, predictors, covariates, lags), there is no standard analysis strategy. The strategy can be try-this-and-try-that. Our method is simple counting and computing the size of the available analysis space.

3. Results

In [Table 1](#), we give the numbers of outcomes, predictors, lags and covariates, for each of the eight papers. Functions of these counts can be used to estimate the number of questions, models and search space available analysis. The product of outcomes, predictors and lags gives the number of questions at issue. For example, three outcomes (AllCause Deaths, heart attacks, and stroke) can be paired with six predictors (CO, NO₂, SO₂, PM2.5, PM10, ozone) to give 18 possible questions. The number of models is given by $2^{\text{Covariates}}$, taking the position that each covariate can be in the model or not. The search space is the number of questions times the number of models.

The median sizes of questions, models and search space are 135, 128, and 9568 respectively. See [Table 2](#). None of the eight papers mention correcting for multiple testing or multiple modeling. All papers appear to test at the level of 0.05. Given the multiple testing and multiple modeling, none of these papers provide strong evidence for their claims. Any claim made could easily be due to chance, a false positive. Note that each of these eight papers should be examined separately for strength of evidence. They must stand on their own before they can be considered for combining in a meta-analysis. As the base papers do not appear reliable, the meta-analysis papers, *Lancet* and *JAMA*, also appear unreliable.

Table 2

Number of questions, models, and total search space, medians and quartiles.

	Median	25%	75%
Multiple Questions	135	66	168
Multiple Models	128	20	448
Total Search Space	9568	2920	40,704

4. Discussion

There are many ways to increase the number of analysis options beyond our simple counting. We count two genders and two possible analyses (gender is in the analysis or not), but the analysis could be male, female and combined giving three options. In examination of a dose response, mortality versus PM2.5 level, logistic regression could be used, one model. Doing a transformation of the dose, say log, points the way to trying multiple transformations, [Ginevan and Watkins \(2010\)](#). Often the dose is cut into several groups, which offers further opportunities for model searching. Age can be treated as a continuous variable or cut into groups with an analysis in each group. [Mann et al. \(2003\)](#) do an analysis for each of three age groups so it could enter the counting process as three rather than two, in or out of the model. Temperature is obviously cyclical. It can be treated in any of several ways. Temperature effects can be controlled by use of a spline curve with differing degrees of stiffness. Or analysis can be within seasons. If case crossover analysis is used, comparisons are often within a month. Each of the analysis options could be changed from outcome to outcome and differ for each of the air components. Multi-component models could be computed, e.g. PM2.5 and ozone together in a model. These various methods could be explored giving the analyst many options for analysis.

After the dramatic increase in deaths after the Great London Smog, there was considerable search for the causative agent. The current paradigm, PM2.5 is a killer, essentially starts with [Dockery et al. \(1993\)](#). That paper now has over 7000 citations. In effect, their association claim is usually taken that PM2.5 is causative of deaths. The dramatic claim of Dockery fell upon very fertile ground. Dockery has been much criticized; the data set has been examined, but it is not publicly available.

Arguably a contemporaneous study was better, [Styer et al. \(1995\)](#). The sample size was much larger and the statistical analysis was sound. They tried a wide range of models and they found no consistent air quality effect on mortality. That paper is cited only just over 100 times. Both Dockery and Styer were funded by EPA.

The positive Dockery paper was take as valid and became the operational paradigm. Once a new paradigm is accepted (in this case by the EPA), it is expected that scientists will come in to fill in the gaps, [Kuhn \(1962\)](#), (and take advantage of funding opportunities). Subsequently many positive association studies were published. An editor commented to me, "The issue addresses (sic) was laid to rest in the mid 1990s by a large reanalysis report sponsored

Table 1

Counts of questions at issue in eight Environmental Health Effect papers used in two meta-analysis papers.

Reference	Outcomes	Predictors	Lags	Covariates	Questions	Models	Search Space
1 Koken et al., 2003	5	6	5	5	150	32	4800
2 Linn et al., 2000	10	4	3	7	120	128	15,360
3 Mann et al., 2003	4	4	6	9	96	512	49,152
4 Ye et al., 2001	16	7	5	3	560	8	4480
5 Zanobetti and Schwartz, 2005	1	1	3	7	3	128	384
6 Rich et al., 2010	5	5	7	10	175	1024	179,200
7 Zanobetti and Schwartz, 2009	5	6	5	4	150	16	2400
8 Barnett et al., 2006	7	4	2	8	56	256	14,336

Download English Version:

<https://daneshyari.com/en/article/5561261>

Download Persian Version:

<https://daneshyari.com/article/5561261>

[Daneshyari.com](https://daneshyari.com)