The reliability of two meta-analysis studies

S. Stanley Young, CGStat

genetree@bellsouth.net

10th International Conference on Multiple Comparison Procedures in Riverside, California from June 20th – 23th, 2017

1

MCP Conference 2017

10th International Conference on Multiple Comparison Procedures in Riverside, California from June 20th – 23th, 2017

http://www.mcp-conference.org/hp/2017/submit-abstracts/

The reliability of two meta-analysis studies

S. Stanley Young, CGStat, genetree@bellsouth.net

Many regulatory decisions are based on meta-analysis of observational studies. There is a need to understand the reliability of meta-analysis studies. Our idea is to examine the reliability of the base studies used in two meta-analysis studies, one appearing in Lancet and the other in JAMA. Both of these studies examine the claimed causal effect of air quality on heart attacks. We count the number of outcomes, predictors, covariates and lags used in each base paper. Lags are of interest as the air quality yesterday might have a health effect today. Outcomes, predictors, covarites and lags are used to estimate multiplicity. Covariates are used to estimate the number of possible models. Together they can be used to estimate the analysis search space available to the researcher. Altogether we examined 21 base papers. We find a median of 11,520 possible analyses with an interquartile range of 1,440 to 81,920. We conclude that the base papers do not support their claims due to a very large model search space and that therefore the meta-analysis paper claims are not supported either. The benefit of our work is to inform regulatory bodies that previous regulations are not supported by papers using sound statistical analysis.

2

Meta-analysis studies based on observational data set papers are called into question.

Meta-Analysis - review

1. Define the research objective
2. Criteria of selected papers
3. Searching for and collecting papers
4. Extract information from selected papers
5. Compute combined estimate
6. Quality check

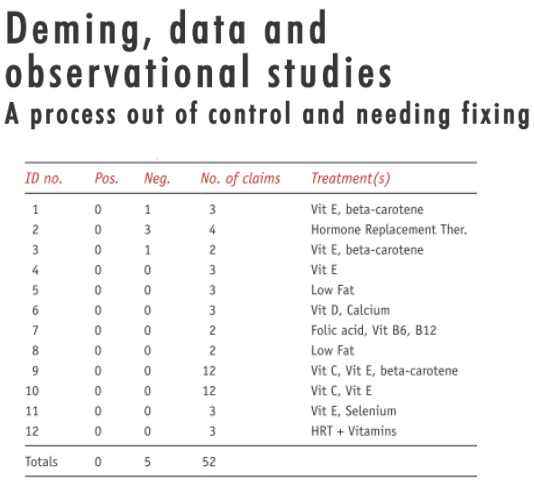
3

It is important that any meta-analysis be performed in a scientifically rigorous way; See Cheng and Peace (2013) and Ehm (2016), for example. Each step should be carefully thought out and not be subject to manipulation, conscious or otherwise; see Strup et al. (2000). Rules for selection of papers should be documented. A data collection form should be used to collect extracted information for each paper.

Our contribution is that each base (primary) paper should be assessed for quality, multiple testing and multiple modeling.

Expect base papers to fail!

Reproducibility – observational studies



Deming, data and observational studies

A process out of control and needing fixing

ID no.	Pos.	Neg.	No. of claims	Treatment(s)
1	0	1	3	Vit E, beta-carotene
2	0	3	4	Hormone Replacement Ther.
3	0	1	2	Vit E, beta-carotene
4	0	0	3	Vit E
5	0	0	3	Low Fat
6	0	0	3	Vit D, Calcium
7	0	0	2	Folic acid, Vit B6, B12
8	0	0	2	Low Fat
9	0	0	12	Vit C, Vit E, beta-carotene
10	0	0	12	Vit C, Vit E
11	0	0	3	Vit E, Selenium
12	0	0	3	HRT + Vitamins
Totals	0	5	52	

4

In many areas of science there is great consternation on the reproducibility of claims made. Routinely, even in experimental science, claims fail to usefully replicate 60-90% of the time.

Young and Karr (2011) found 12 papers where the claims came from observational studies and these claims were tested in randomized clinical trial. The observational papers appeared in high-impact journals and reported very small p-values. None of the claims replicated in the claimed direction and five were statistically significant in the opposite direction. Astounding!

Positive Obs studies, Negative RCTs

ID no.	*Pos.*	*Neg.*	*No. of claims*	*Treatment(s)*
1	0	1	3	Vit E, beta-carotene
2	0	3	4	Hormone Replacement Ther.
3	0	1	2	Vit E, beta-carotene
4	0	0	3	Vit E
5	0	0	3	Low Fat
6	0	0	3	Vit D, Calcium
7	0	0	2	Folic acid, Vit B6, B12
8	0	0	2	Low Fat
9	0	0	12	Vit C, Vit E, beta-carotene
10	0	0	12	Vit C, Vit E
11	0	0	3	Vit E, Selenium
12	0	0	3	HRT + Vitamins
Totals	0	5	52	

5

Here are the results from Young and Karr (2011).

Lancet, Nawrot, 2011

Public health importance of triggers of myocardial infarction: a comparative risk assessment

Tim S Nawrot, Laura Perez, Nino Künzli, Elke Munters, Benoit Nemery

Lancet 2011; 377: 732–40

"In view of both the magnitude of the risk and the prevalence in the population, air pollution is an important trigger of myocardial infarction…"

6

The Lancet, Nawrot et al. (2011) is the first of two meta-analysis papers we examine. They claim that air quality, small particles, PM2.5, trigger heart attacks.

The reported effect is very small, even though they claim the magnitude is large.

Note well, there could well be publication bias. It is well-known that editors and referees favor "positive effect" papers.

We will make the case that multiple testing and multiple modeling can lead to falsely small p-values.

Lancet, Nawrot, 2011

	Design	n	Hazard period before MI episode	OR (95% CI) for 10 μg/m³ increase
Linn[41]	Time series	~51 465	24 h	1·01 (1·00–1·01)
Peters[43]	Case-crossover	772	24 h	1·18 (1·04–1·36)
Ye[46]	Time series	~7380	24 h	NS
Mann[42]	Time series	19 690	24 h	1·00 (0·99–1·01)
Koken[40]	Time series	~4073	24 h	NS
Sullivan[45]*	Case-crossover	5793	24 h	1·01 (0·99–1·05)
Zanobetti[47]	Case-crossover	302 453	24 h	1·01 (1·00–1·01)
Peters[15]	Case-crossover	851	24 h	1·02 (0·97–1·06)
Pope[44]	Case-crossover	4818	24 h	1·02 (1·01–1·05)
Zanobetti[13]*	Case-crossover	15 578	24 h	1·10 (1·01–1·20)
Cendon[21]	Time series	724 717	24 h (ICU) 24 h (infirmary)	1·03 (1·02–1·09) 1·05 (1·00–1·10)
Lanki[20]	Time series	26 854	24 h	1·00 (0·99–1·01)
Barnett[12]*	Case-crossover	~30 660	24 h (age ≥65 years)	1·05 (1·02–1·08)
Zanobetti[14]*	Time series	121 652	48 h	1·02 (1·01–1·02)
Combined estimate		593 480	..	1·02 (1·01–1·02)

7/14 significant
7/14 no effect

7

Here are the Nawrot et al. results for air quality as measured by PM2.5.

Note well the very small effects. They are highly statistically significant, but could well be due to publication bias, negative studies are not reported, and multiple testing and/or multiple modeling.

JAMA, Mustafic, 2012

Main Air Pollutants and Myocardial Infarction
A Systematic Review and Meta-analysis

JAMA, February 15, 2012—Vol 307, No. 7

"All the main air pollutants, with the exception of ozone, were significantly associated with a near-term increase in MI risk."

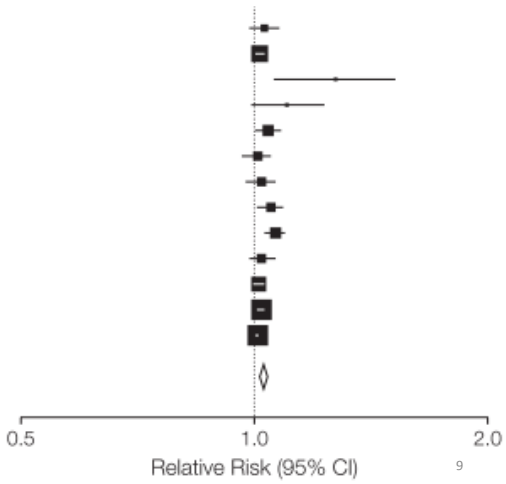
8

The 2nd study we examine is Mustafic et al. (2012) that appeared in JAMA. They too conclude that air quality, as measured by PM2.5, leads to heart attacks.

The two papers were done at about the same time, the endpoints were very similar, yet the found papers differed. See later slides. In both cases, a computer search was winnowed down to a small number of papers used in the study. Nawrot reduced 538 papers to 36. Mustafic reduced 1667 papers to 34. Rather obviously there is potential for selection bias. Note that we only examined the studies that examine PM2.5 on CV effects.

JAMA PM2.5

Source	Relative Risk (95% CI)
Barnett et al,[9] 2006	1.031 (0.983-1.082)
Belleudi et al,[41] 2010	1.018 (1.001-1.036)
Peters et al,[20] 2001	1.272 (1.061-1.525)
Peters et al,[43] 2005	1.105 (0.991-1.232)
Pope et al,[35] 2006	1.042 (1.003-1.083)
Rich et al,[21] 2010	1.010 (0.966-1.056)
Sullivan et al,[22] 2005	1.020 (0.976-1.066)
Zanobetti and Schwartz,[45] 2006	1.051 (1.010-1.094)
Maté et al,[25] 2010	1.066 (1.033-1.101)
Stieb et al,[28] 2009	1.024 (0.983-1.066)
Ueda et al,[38] 2009	1.013 (0.991-1.035)
Zanobetti et al,[30] 2009	1.022 (1.011-1.034)
Zanobetti and Schwartz,[31] 2009	1.011 (1.004-1.018)
Combined	1.025 (1.015-1.036)



One custom is to consider risk ratios less than 0.50 or greater than 2.00 as possibly indicative of causal effects. The current custom in environmental epidemiology is to consider any RR with a p-value <0.05 as indicative of an effect however small in magnitude.

Combining the p-values, weighted by variability, gives a RR of 1.025 with a p-value of < 0.001.

Again, publication bias might be an issue.

Simple counting

In each paper count

Outcomes

Predictors

Covariates

Lags

Model search space = out x pred x lags x 2^{covar}

10

This is as complicated as it gets. We examine each paper and count the number of outcomes, predictors, covariates, and lags. It can take a skilled counter ½ to 1 day to count as the authors might mention some of the variables more or less anywhere in the paper. Covariates can be particularly elusive.

The product of Outcomes and Predictors is one measure of the search space. In air quality studies the air quality some days before might have an effect on mortality today. A covariate can be in or out of the model.

The total modeling search space can be surprisingly large.

None of the papers examine did any correction for multiple testing or multiple modeling. In that sense, all of the papers might be considered exploratory. In each meta-analysis the authors took the papers as confirmatory, which we think is a grave mistake.

Lancet Only
Outcomes, predictors, covariates, lags

Paper	Out	Pred	Covar	#Lags	#models
Linn WS 2000 EHP	10	4	7	3	15360
Ye F 2001 EHP	16	6	1	5	960
Mann JK 2002 EHP	4	4	9	6	49152
Koken PJ 2003 EHP	5	6	5	5	4800
Peters 2004 NEJM	4	5	10	4	81920
Zanobetti A 2005 EHP	1	1	7	3	384
Cendon S 2006 RSP	2	5	5	7	2240
Lanki T 2006 OEM	3	5	3	3	360

11

These 8 papers appeared in the Lancet paper only. They were not “found” by Mastafic. The number of possible models is staggering. Any testing at 0.05 is a very weak screen indeed.

Note well, a very small random p-value will be paired with a biased treatment effect.

Note also, it is the responsibility of a researcher making a claim to provide strong evidence for their claim. Strictly speaking the claim of the author could be correct or not, but the claim is certainly without statistical support.

JAMA Only
Outcomes, predictors, covariates, lags

Paper	Out	Pred	Covar	#Lags	# Models
Stieb 2009 JAMA	6	6	7	3	13824
Ueda 2009 JC	20	1	5	3	1920
Zanobetti 2009 EHP	5	6	4	5	2400
Belleudi 2010 Epi	6	3	6	7	8064
Mate 2010 STE	28	6	9	7	602112
Rich 2010 Epi	5	5	10	7	179200

12

These six papers appeared in JAMA only. They were not “found” by Nawrot.

Both Lancet and JAMA
Outcomes, predictors, covariates, lags

Paper	Out	Pred	Covar	#Lags	# Models
Peters 2001 Cir	2	7	2	2	112
Peters 2005 HEI	4	5	10	5	102400
Sullivan 2005 Epi	4	12	10	3	147456
Barnett 2006 EHP	7	4	8	2	14336
Pope 2006 Cir	1	2	13	5	81920
Zanobetti 2006 JECH	8	6	3	2	768
Zanobetti 2009 EH	5	18	7	1	11520

13

These 7 papers were found by both sets of authors.

Summary Statistics

Variable	min	Median	Max
Outcomes	1	5	28
Predictors	1	5	18
Covariates	1	7	13
Lags	1	4	7
Models	112	11,520	602,112
		12,800	

14

The number of models available ranged from 112 to 602,112 with a median of 11,520. If you put into the formula the median for each variable, the search space is 12,800.

It is clear that none of the papers have any statistical rigor and that none of the claims should be considered reliable. The two meta-analysis studies should be considered unreliable.

Again, note well, a very small random p-value will be paired with a biased treatment effect.

Comments

1. Exploratory versus Confirmatory (kind)
   a. p-value plots (1981)
   b. Adjusted p-values (1993)
2. p-Hacking (multiple testing and modeling)
3. HARKing (Hypothesis After the Results are Known)
4. Publication bias, editors often reject negative studies

15

Methods to deal with multiple testing have been available for some time. There have been many warnings about false positives due to multiple testing and multiple modeling.

Authors of observational studies are either woefully ignorant or they are responding to incentives other than solid science.

Two recent books cover the current problem of poor reproducibility:

Harris, R. 2017 Rigor Mortis.

Chandler, C. 2017. The Seven Deadly Sins of Psycoholgy.

Harris covers experimental biology with a false positive rate of 89%.

Chandler covers experimental psychology with a false positive rate of at least 60%.

Young, 2017

Air quality environmental epidemiology studies are unreliable

S. Stanley Young

Regulatory Toxicology and Pharmacology 86 (2017) 177e180

"The claims in these eight papers are not statistically supported so these papers are unreliable as are the meta-analysis papers that use them."

Essentially every meta-analysis that uses observational studies is suspect.

16

Young, 2017, discusses eight papers that appeared in Environmental Health Perspectives, the environmental epidemiology journal with the highest impact factor. All eight papers have massive multiple modeling and multiple testing issues.

It is somewhat amazing that the problems appear so pervasive and yet have not been acknowledge/addressed in the meta-analysis literature on base observational studies.

Anyone that wants a publication is welcome to jump in.

Heart (BMJ) 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

Ai Milojevic,[1] Paul Wilkinson,[1] Ben Armstrong,[1] Krishnan Bhaskaran,[2] Liam Smeeth,[2] Shakoor Hajat[1]

Massive study. All of England and Wales.
400k MIs, 600k deaths, 2M CVD admissions.

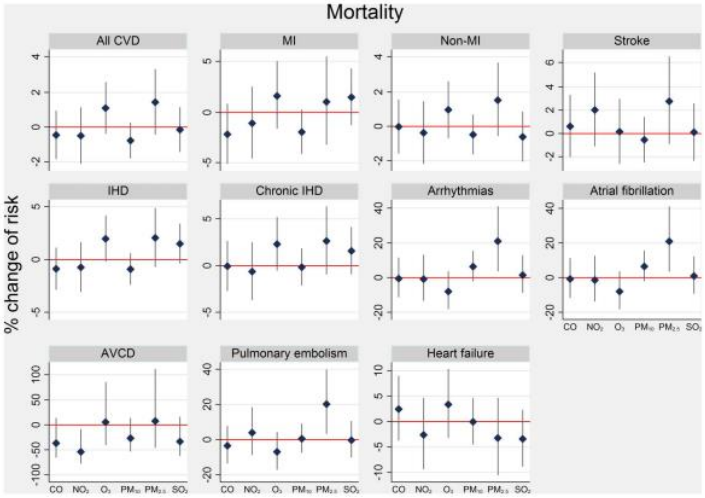
17

So where is the truth?

So we show that two meta-analysis papers are without statistical support; the studies are not reliable.

We now report two massive studies have been reported. Both are negative.

Heart (BMJ) continued



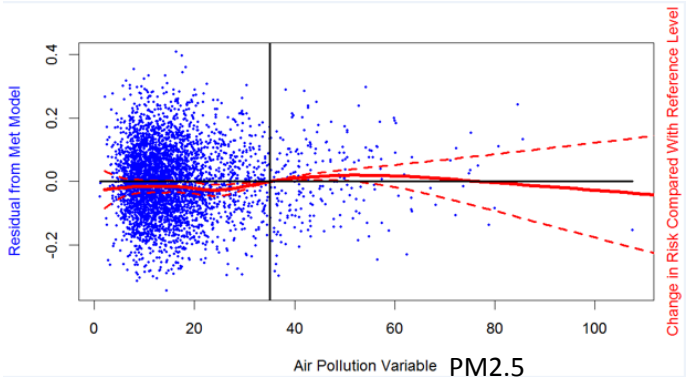
+	24
0	17
-	25
Total	66

18

Milojevic et al. (2014) report on six air components and eleven outcomes. The results are consistent with random.

Young, Smith and Lopiano, 2015

California, 8 air basins, 2000-2012, 37,000 exposure days.



19

Young et al. (2015) placed a technical report at arXiv, https://arxiv.org/abs/1502.03062

Here is one figure from that report. After removing seasonal and weather effects, there is no relationship between PM2.5 and mortality.

The study is massive. The data set and analysis code is public.

Contact information

Stan Young, genetree@bellsouth.net

I'm willing to collaborate on

1. Examination of a meta-analysis studie or
2. Letter to editor on an observational studies.

20

This reliability of meta-analysis using observational studies is wide open. This is an area where statistician can have high impact.

I'm willing to collaborate.

Some additional references

1. Applied Meta-Analysis with R, Chen and Peace.
2. Reproducibility from the Perspective of Meta-analysis, Werner Ehm
3. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. Shah et al., JAMA, April 19, 2000—Vol 283