Strengthening the Reliability of Information to be Combined

S. Stanley Young, CGStat, Raleigh, NC 27607, USA

Warren B. Kindzierski, Independent Consultant, St Albert, Alberta, T8N5R1, Canada

Abstract

Combining information from multiple sources (studies) in environmental epidemiology is a common task for decision makers in inferring causality. There is an important step before combining information; the reliability of each information source should be evaluated. Reliability should not be assumed as claimed risk factor−disease relationships may fail to replicate. Also, because many hypotheses can be tested in a study, researchers may be more inclined to publish positive relationships with many negative relationships remaining unpublished. The reliability of information to be combined should not be taken at face value. Environmental epidemiology methods require a strong statistical component to develop useful and interpretable causal relationships. Our idea is to use two techniques, one ancient (simple counting) and one relatively new (p-value plots) to evaluate statistical reliability. A source can be examined to determine the analysis search space (number of hypotheses tested). How many hypotheses were open to the researcher to search for a positive relationship. The larger the search space the greater the opportunity that a claimed relationship (or its size) could have been influenced by chance. A p-value plot is simple. The p-value linked to each source is determined, and the ranked p-values are plotted against the integers. If the p-values fall on a roughly 45-degree line (they roughly are uniformly distributed), then there is evidence that chance is at play. The benefit of examining the reliability of the information to be combined is that the decision maker can be more confident chance is not driving the decision process.

Introduction

Julia Galef in her 2021 book, *The Scout Mindset*, gives strategies for trying to sort reality from non-reality. She wants to know why some people see things clearly and others don't. One of her rules is to back off from full belief in something, keeping somewhat of an open mind, and let new evidence move your opinion up or down from where you start. Thus, the need for use of different/independent approaches to gather evidence on a research question and for triangulation of the evidence.

On the surface of the sea of evidence, we may just see the tips of icebergs through some mist and fog. Just how many icebergs are there and how much large are they under the surface is what we want to try understand to sort out reality. Why is this important? It turns out that in research today, there is a case made about questioning reliability of evidence – 50 to 100% of science statements made cannot be reproduced depending on the discipline.

Epidemiology largely relies on statistics to make meaningful statements about causality in observational risk factor−disease association studies. Here we explore here some alternative statistical approaches to attempt to triangulate/lend independent support to statistical evidence derived from observational studies combined in meta-analysis. Our framework is straightforward – given the importance of statistics, we promote use of independent statistical approaches to improve causal inference.

We can count and approximate the number of questions (hypotheses) at issue in an observational study to understand whether examining multiple hypotheses is a key source of bias. Also, we can examine a meta-analysis, a study of studies, focusing on a specific research question. Here we can examine the statistical reliability of the included base studies in the meta-analysis.

This poster is based on our Shifting Sands report.

Methods and Materials

The methods we promote are based on theory and can offer strong statistical support in triangulation of epidemiological evidence. A meta-analysis is an accepted formal way to combine evidence from multiple base studies (in our case observational studies). Typically, a computer literature search is done and the found papers evaluated and screened for suitability. Risk ratios and confidence limits are extracted and are used to estimate the overall effect including confidence bounds.

Counting – Observational studies generally use a direct statistical analysis strategy on data collected – e.g., what causes, or risk factors are related to what outcomes (health effects). If an observational data set contains "C" causes and "O" outcomes, C $\times$ O possible hypotheses can be investigated.
An adjustment factor "A" (also called a covariate) can be included as a yes/no adjustment – such as income or education – to see how it can modify each of the C $\times$ O hypotheses. Here an adjustment factor is included or excluded; and a multiplier of 2 is assumed for each adjustment factor considered.

We can count causes (C), outcomes (O), and yes/no adjustment factors (A); where the number of hypotheses can be approximated as = C $\times$ O $\times$ 2^A. Observational studies with large counts (numbers of hypotheses examined) have an increased likelihood of registering a false positive finding.

P-value plots – From risk ratios and confidence limits a p-value can be computed for each base study in a meta-analysis. A p-value plot can be constructed by rank ordering p-values from smallest to largest and plotting them against the integers, 1, 2, 3, . . .

If p-values roughly fall on a 45-degree line, they support randomness (no real effect). If the p-values are mostly smaller than 0.05, they support a real effect. A bilinear, hockey stick, shaped p-value plot indicates ambivalence (uncertainty) in an effect.

Results

Figure 1. Counting, example 1

Figure 2: Estimated Size of Analysis Search Space, Eight Environmental Epidemiology Papers

RowID	Author	Year	Questions	Models	Search Space
1	Zanobetti	2005	3	128	384
2	Zanobetti	2009	150	16	2,400
3	Ye	2001	560	8	4,480
4	Koken	2003	150	32	4,800
5	Barnett	2006	56	256	14,336
6	Linn	2000	120	128	15,360
7	Mann	2003	96	512	49,152
8	Rich	2010	175	1,024	179,200

Figure 2. Counting, example 2

Table 2. Authors, variable counts, and analysis search spaces for the 34 case study base papers.

Cit #	Author	Outcome	Predictor	Covariate	Lag	Space1	Space2	Space3
7	Braga	4	1	6	4	16	64	1,024
8	Koken	5	5	6	5	125	64	8,000
9	Barnett	7	5	10	1	35	1,024	35,840
10	Berglind	1	4	10	2	8	1,024	8,192
11	Cendon	2	5	5	8	80	32	2,560
12	Linn	3	4	8	3	36	256	9,216
19	Ye	8	5	3	5	200	8	1,600
20	Peters	1	8	11	2	16	2,048	32,768
21	Rich	1	5	9	6	30	512	15,360
22	Sullivan	4	4	8	3	48	256	12,288
23	Eilstein	1	12	5	6	72	32	2,304
24	Lanki	1	5	3	6	30	8	240
25	Mate'	4	6	7	6	144	128	18,432
26	Medina	15	6	8	6	540	256	138,240
27	Poloniecki	7	5	5	1	35	32	1,120
28	Stieb	6	6	7	3	108	128	13,824
29	Zanobetti	5	2	11	3	30	2,048	61,440
30	Zanobetti	5	18	8	3	270	256	69,120
31	Zanobetti	5	2	9	2	20	512	10,240
32	Hoek	4	8	9	3	96	512	49,152
33	Cheng	1	5	6	3	15	64	960
34	Hsieh	1	5	6	3	15	64	960
35	Pope	1	2	13	7	14	8,192	114,688
36	D'Ippoliti	3	4	11	3	36	2048	73,728
37	Henrotin	4	5	14	14	280	16,384	4,587,520
38	Ueda	3	1	7	3	9	128	1,152
39	Mann	4	4	9	7	112	512	57,344
40	Sharovsky	4	3	10	8	96	1,024	98,304
41	Belleudi	4	3	8	13	156	256	39,936
42	Nuvolone	1	3	9	8	24	512	12,288
43	Peters	4	5	10	4	80	1,024	81,920
44	Ruidavets	4	3	8	4	48	256	12,288
45	Zanobetti	2	6	7	3	36	128	4,608
46	Bhaskaran	1	5	7	5	25	128	3,200

Note: Cit # before author name is the case study reference number; author name is first author listed (refer to our Supplement).

Figure 3. P-value plot

Figure 8: P-value plot, All-Cause Mortality and $PM_{2.5}$[152]

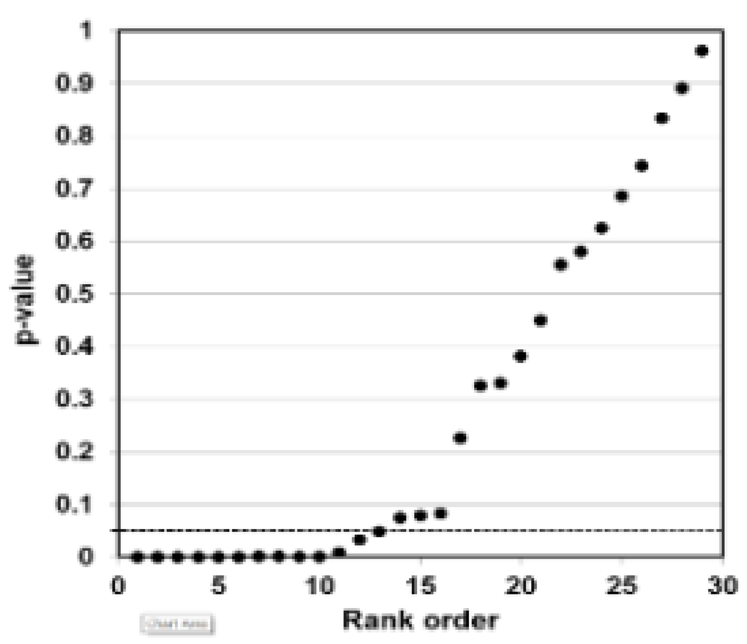


Figure 4. California Air Basins

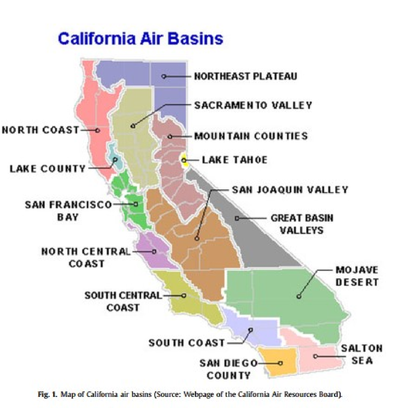
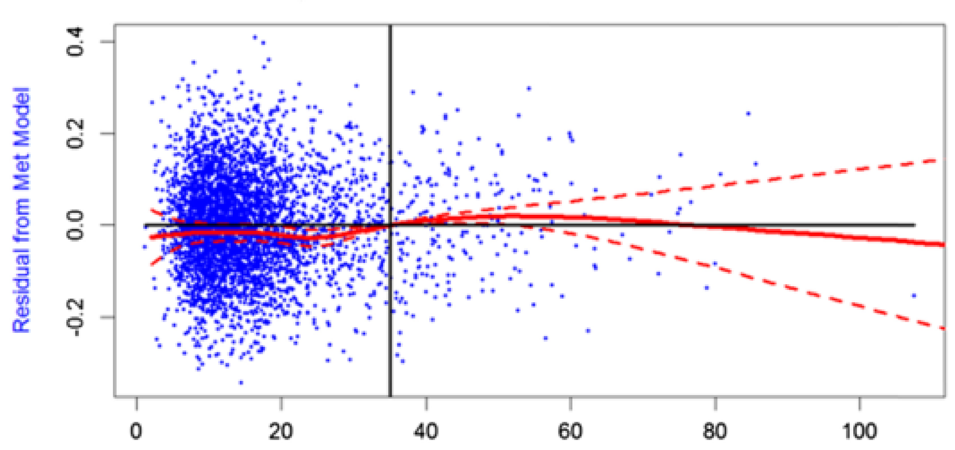


Figure 5. Mortality (model corrected) vs $PM_{2.5}$



Discussion

For the counting we have presented here and many other base studies that we have examined, we find that large numbers of hypotheses tend to be typically examined in environmental epidemiology observational studies. The median number of analyses we observe elsewhere in this discipline is on the order of ten thousand, which has important implications on the possible reporting of false positive findings.

The p-value plot offers an independent look at the reliability of a meta-analytic statistic and the strength of evidence for a research question. In our example, the presented p-value plot is in the shape of a hockey stick. The p-values on the blade are small and suggest a real effect. Those on the handle suggest randomness (chance findings). We have a mixture. Very small p-values (i.e., < 0.001) merit special comment. Do they indicate a real effect? Might these be due to large numbers of hypotheses examined (multiple testing bias)? In our example, there are also many p-values on a 45-degree line suggesting no effect. Overall, the hockey stick, shaped p-value plot indicates ambivalence (uncertainty) in an effect.

From our analysis we can see that a set of base studies in a meta-analysis whose possible numbers of hypotheses examined are large and whose p-values demonstrate bilinearity in a p-value plot should not be accepted as reliable evidence and should require closer independent inspection.

Conclusions

Meta-analysis is an accepted and standard way to combine evidence for a specific research question over many studies. The studies can be similar in form, or they can be from different kinds of studies, human, animal, in vitro. In the case of observational epidemiology, triangulation of evidence has an opportunity of strengthening causal inference. Here we suggest that use of counting and p-value plots can assist in strengthening causal inference in meta-analytic evidence. Counting of numbers of hypotheses examined in observational studies used in meta-analysis can aid in understanding the importance of multiple testing bias and possible reporting of false positive findings. Also, p-value plots can be used to evaluate reliability of multiple similar or disparate studies in a meta-analysis.

Contact

Young, genetree@Bellsouth.net
Kindzierski, wbk@shaw.ca

References, chronological

Mustafic, H., Jabre P., Caussin, C., Murad, M. H., Escolano, S., Tafflet, M., Périer, M.-C., Marijon, E., Vernerey, D., Empana, J.-P., Jouven, X. **2012**. Main air pollutants and myocardial infarction: A systematic review and meta-analysis. *Journal of the American Medical Association* 307, 7: 713-21. https://doi.org/10.1001/jama.2012.126.

Young, S. S. **2017**. Air quality environmental epidemiology studies are unreliable. *Regulatory Toxicology and Pharmacology* 86: 177-80. http://dx.doi.org/10.1016/j.yrtph.2017.03.009 .

Young, S. S., Smith, R. L., Lopiano, K. K. **2017**. Air quality and acute deaths in California, 2000-2012. *Regulatory Toxicology and Pharmacology* 88:173-84. https://doi.org/10.1016/j.yrtph.2017.06.003.

Young, S. S., Kindzierski, W. B. **2019**. Evaluation of a meta-analysis of air quality and heart attacks, a case study. *Critical Reviews in Toxicology* 49,1: 85–94. https://doi.org/10.1080/10408444.2019.1576587.

Young SS, Kindzierski WB, Randall D. **2021**. Shifting Sands, Unsound Science and Unsafe Regulation Report 1. National Association of Scholars. https://www.nas.org/reports/shifting-sandsreport-i

Galef J. **2021**. *The Scout: Why some people see things clearly and others don't*. Penguin Random House