Effect Size Magnification and Epidemiologic Design Calculations

Use in EPA's Office of Pesticide Programs in evaluating study size in epidemiology studies

ISEE Poster ID: P-0807

David J. Miller and James T. Nguyen
Health Effects Division, Office of Pesticide Programs, Environmental Protection Agency

32nd Annual Conference of the International Society for Environmental Epidemiology
ISEE 2020
AUGUST 24-27, 2020 | VIRTUAL CONFERENCE
FULLY VIRTUAL CONFERENCE

Abstract

Most researchers recognize issues associated with low powered (and generally small) studies *vis-a-vis* their lessened ability to detect true effects. Fewer, however, recognize issues associated with low powered studies and their tendency to produce inflated estimates if those estimated effects are required to pass a statistical (e.g., p<0.05) or other threshold to be judged important, relevant, or "discovered" (Ioannidis, 2008). Effect size magnification (ESM) is a term used to refer to this phenomenon. Specifically*: low-powered studies that find evidence of an effect often provide inflated estimates of the size of that effect*.

This poster discusses the implications of ESM with respect to epidemiological study conclusions and our efforts in EPA's Office of Pesticide Programs to understand, reproduce, and finally apply this knowledge to better evaluate the reliability of reported (statistically significant) effect sizes in epidemiology studies and put these into a fuller context. Routinely performing such ESM calculations (aka "*post-hoc* design calculations" per Gelman and Carlin (2014)) in epidemiology can assist in determining the extent to which ESM may be a concern or should be otherwise accounted for in interpretation of epidemiological results. While such design calculations do not change a statistically significant result to a nonsignificant result, they do allow regulatory staff to consider that a reportedly large effect in a study may in fact be much lower, to a degree that the effect may have less influence on EPA's conclusions and decisions.

What is Effect Size Magnification (ESM)?

- ESM refers to the phenomenon that low-powered studies that find evidence of an effect often provide *inflated estimates* of the size of that effect.
- The amount of ESM is inversely related to power which, in turn, is dependent upon:
  - Sample Size
  - True Effect Size
  - Background or Control or Reference Rate
- ESM is expected when an effect has to pass a certain threshold — such as reaching statistical significance — in order for it to have been 'discovered'.
- ESM is worst for low-powered studies that can only detect effects that are large.
- In practice, this means that research findings of small studies can often be biased in favor of finding inflated effects.

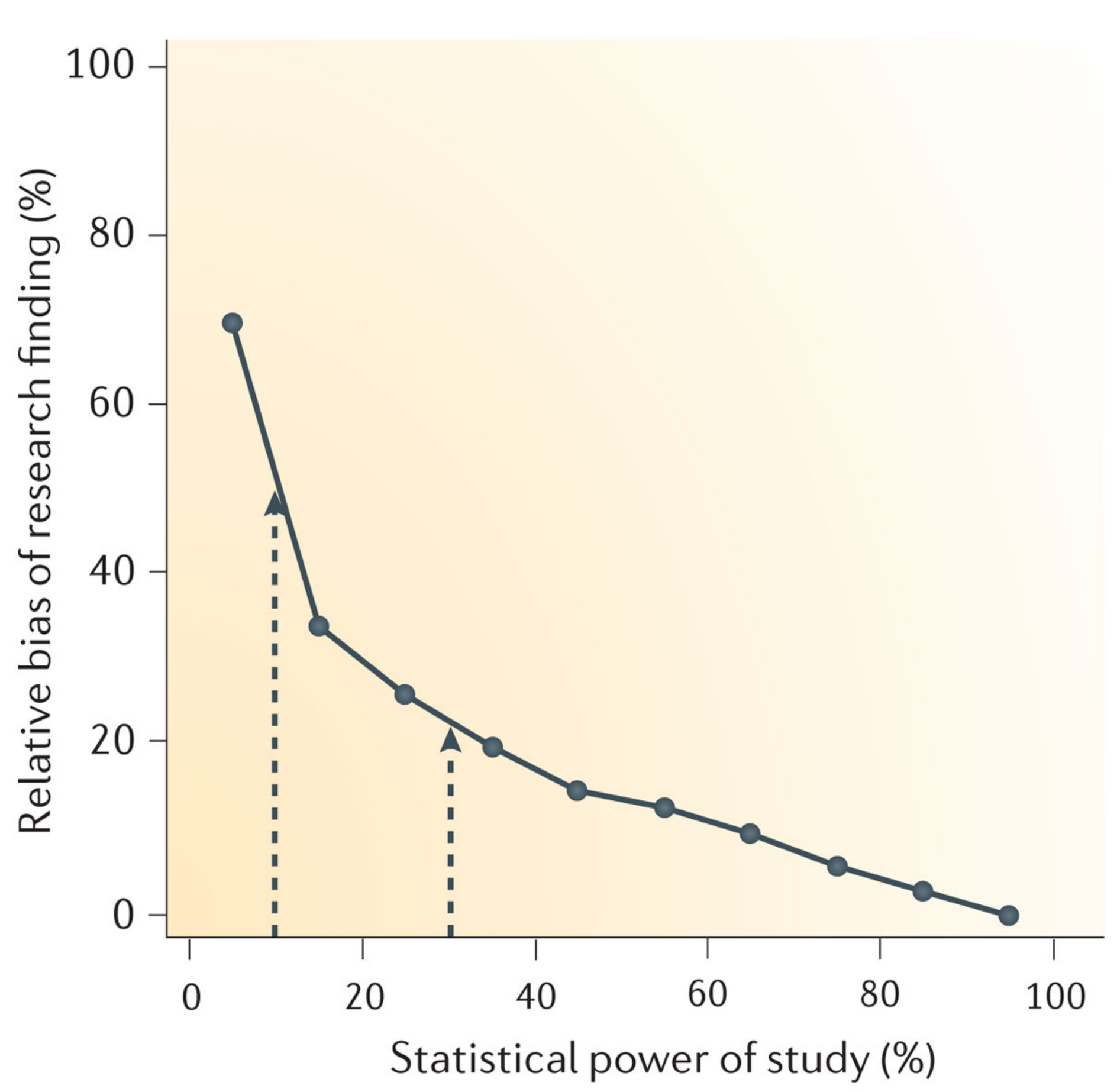


Figure excerpted from: Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nature reviews. Neuroscience. 2013 May;14(5):365-376. https://www.nature.com/articles/nrn3475#citeas

- Most researchers recognize issues associated with low powered studies *vis-a-vis* the failure to detect true effects. However, fewer recognize issues associated with low powered studies and their tendency to produce inflated estimates.

The analysis described in this poster has been reviewed by US EPA's Office of Chemical Safety and Pollution Prevention (OCSPP) and approved for release as reflective of one component of OPP's current practices in review and interpretation of the epidemiologic literature. The contents do not necessarily reflect the views, policies, or determinations of the Agency.

Case Study*

- An epidemiological example uses a study published by Montgomery et al. (2008) in AJE (**167**(10):1235-46)). In this study, rates of diabetes were compared among cases and controls from occupational exposure to trichlorfon in the Agricultural Health Study (See table below).

TABLE 2. Ever use of specific pesticides comparing incident diabetics and nondiabetics among applicators enrolled in the Agricultural Health Study, 1993–2003

Pesticide name	No. of diabetics (*n* = 1,176)	% exposed	No. of nondiabetics (*n* = 30,611)	% exposed	Age-adjusted odds ratio*	95% confidence interval	Adjusted odds ratio†	95% confidence interval
Trichlorfon	13	1	169	1	2.03	1.15, 3.60	1.85	1.03, 3.33

- **Question:** To what extent might effect size magnification be important here if one were interested in a statistically significant result?

1 Setup analysis

- **number of subjects** in the:
  - *reference* (or control) group
  - *comparison* group
- **proportion of interest** in the *reference* group

```
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        13         169  |        182      0.0714
        Controls |      1163       30422  |      31585      0.0368
-----------------+------------------------+------------------------
           Total |      1176       30591  |      31767      0.0370
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         2.01217        |    1.141184     3.54792
 Attr. frac. ex. |         .5030241       |    .1237168     .7181448
 Attr. frac. pop |         .0359303       |
                 +-------------------------------------------------
                               chi2(1) =     6.08  Pr>chi2 = 0.0137
```

2 Simulation

Here, the authors "discovered" an odds ratio of 2.01 for an association between trichlorfon exposure and diabetes.

...but the (low) power of the study suggests an OR of 2.01 could be readily attributed to effect size magnification at a true OR of as low as 1.1 ...

...and thus the study sample size is small and the study will receive less weight in any epidemiological WoE evaluations re: trichlorfon and diabetes due to concerns about reliability of the effect size estimate.

Simulations for Effect Sizes Passing a Threshold of Formal Statistical Significance (p = 0.05) for Montgomery *et al.* (2008) Epidemiology Study

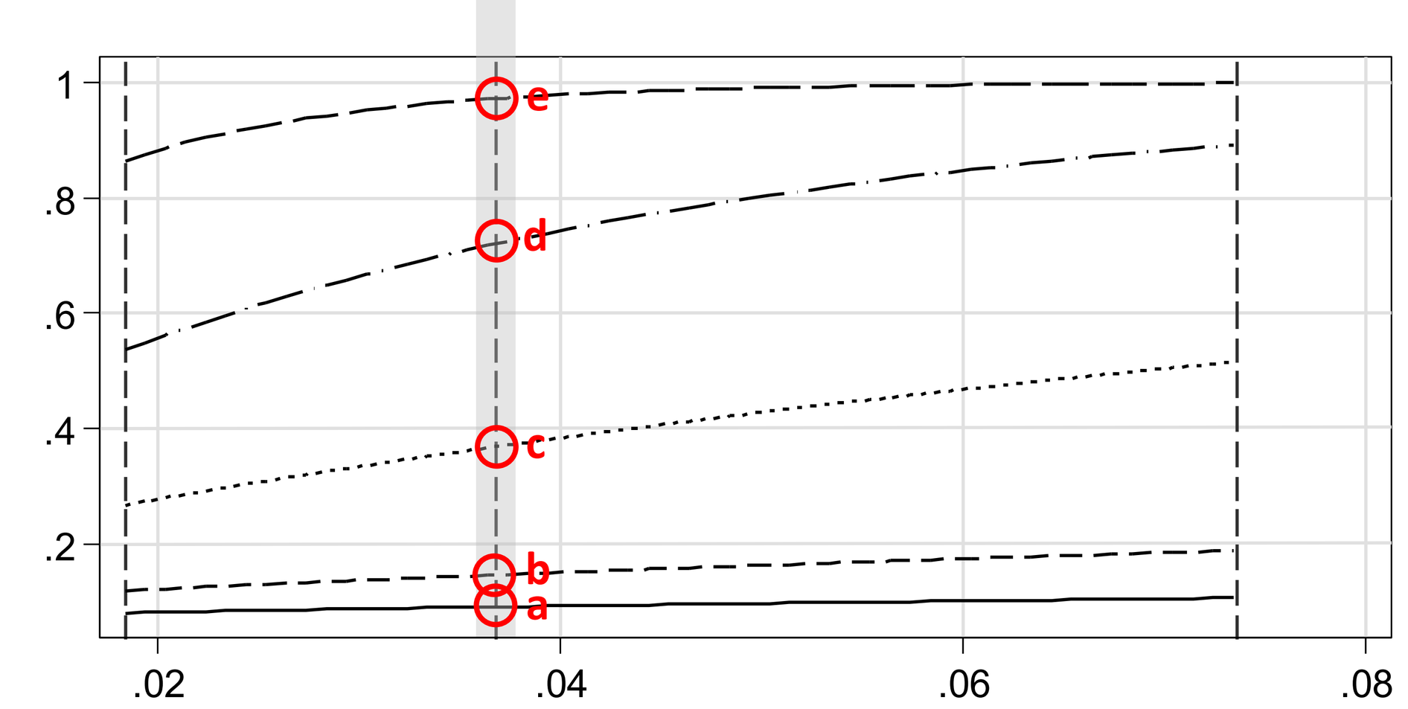
True OR	Control Group Rate, p_0 (%)	Sample n Per Group (n_0/n_1)	Observed OR in Significant Associations: Median (10^{th}-90^{th})[a]	Observed OR in Significant Associations: Median Fold Inflation	
1.1	3.68	31605/182	1.852 (1.694 – 2.242)	1.68	11% power a
1.2	3.68	31605/182	1.881 (1.697– 2.247)	1.57	16% power b
1.5	3.68	31605/182	1.931 (1.697– 2.487)	1.29	35% power c
2	3.68	31605/182	2.201 (1.776 – 2.928)	1.1	74% power d
3	3.68	31605/182	2.999 (2.166 – 4.015)	1.0	98% power e

[a] 10^{th} to 90^{th} indicates the 10^{th} and 90^{th} percentiles of the statistically significant results.

3 Sensitivity Analysis

"Proportion Exposed in Control Group" can be an important parameter in a sensitivity analysis. It is useful to vary this to determine how sensitive power is to this (observed) quantity.

Results suggest that above conclusion regarding ESM's potential role at a true OR of as low as 1.1 is robust and not sensitive to observed proportion of exposed in control group.



Parameters: α = .05, N = 31787, N_1 = 31605, N_2 = 182

Estimated power for a two-sample proportions test
Pearson's χ^2 test
H_0: $p_2 = p_1$ versus H_a: $p_2 > p_1$

Case Study Conclusions

- Adequately powered studies are necessary to be able to have at least some minimal degree of confidence in the estimate of the effect size, particularly in "discovery" phases with effect sizes that are statistically significant.
- *Post-hoc* epidemiologic design calculations can assist in determining if effect size magnification may be present and the extent to which it may be an issue or should be accounted for in interpretation of results.

***For more technical detail, examples, and analytical code (SAS and Stata), see working paper at: http://www.imm.ki.se/biostatistics/emagnification**

Why is ESM Important?

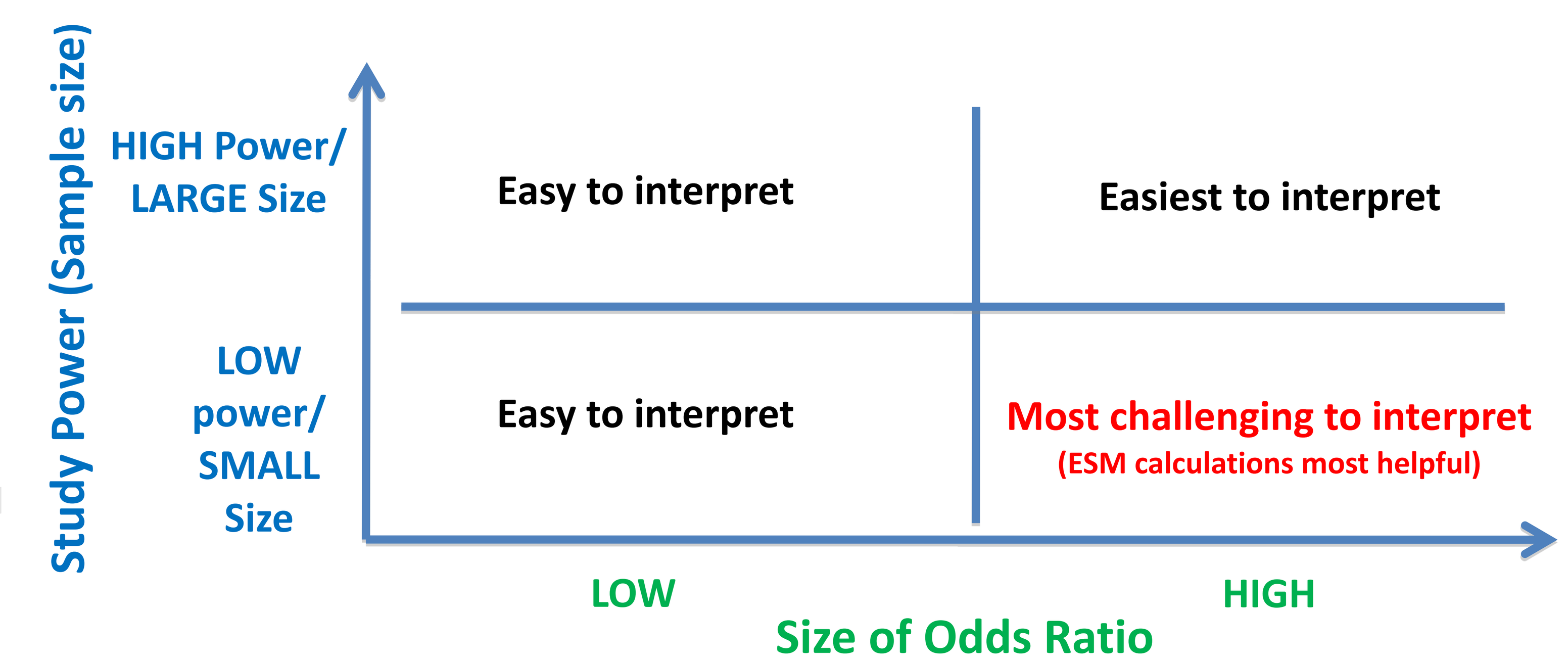
Key Questions

- If the results of a study or studies of interest cannot – in theory or practice – be reliably replicated and might reflect systematically inflated effect sizes, how much confidence can we have in decisions that rely upon them?
- Can we understand, reproduce, and finally apply the ESM work to better understand (epidemiological) studies available in the literature?
- Can we use ESM to better evaluate the reliability of reported (statistically significant) effect sizes and put these into a fuller context with respect to epidemiological study conclusions?

Essential Input to Assess ESM

- In order to determine the potential degree of effect size magnification for any given study, the reviewer needs to perform various "design effect" calculations. This, in turn, requires that we know four values:
  1. the number of subjects in the *reference* (or control) group
  2. the number of subjects in the *comparison* group
  3. the proportion of interest in the *reference* group
     e.g., the proportion of **exposed** subjects in control group for case-control studies
  4. a target value of interest (typically an OR or RR in epidemiology studies) to detect a difference of a given (pre-determined) size in a comparison of two groups (e.g., exposed vs. not exposed)

The first three listed values are provided in or must be obtained from the publication while the target value of interest is selected by the risk managers (and is ultimately a policy decision).



Key Messages

1. Effect Size Magnification refers to the phenomenon that studies that find evidence of an effect often provide inflated estimates of the size of that effect
   - Occurs when studies have low power
   - Such magnification is expected when an effect has to pass a certain threshold — such as reaching statistical significance — in order for it to have been 'discovered'
2. Many epidemiological studies are under-powered to find low to moderate effects, which can lead to exaggerated or inflated effect size estimates if primary interest is in "discovered" effects.
3. If an epidemiological study has low power, we must be suspect of 'large' or 'significant' ORs since these values may be inflated.
4. Don't rely just on p-values, as these may only be meaningful or reliable in adequately powered studies.
5. If an epidemiological study does have low power and a 'large' discovered effect size, then a *post-hoc* design calculation should be performed to assist in quantitatively evaluating how reliable the effect size estimate may be.

Where Can I Learn More?

Gelman, A. 2017. "Yes, it makes sense to do design analysis ('power calculations') after the data have been collected" at https://statmodeling.stat.columbia.edu/2017/03/03/yes-makes-sensedesign-analysis-power-calculations-data-collected/ 3 March.

Gelman, A. and J. Carlin. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives in Psychological Science*. Vol 9(6): 641-651. https://pubmed.ncbi.nlm.nih.gov/26186114/

Ioannidis, J. P. A. 2008. Why most discovered true associations are inflated. *Epidemiology* 19:640-648. https://pubmed.ncbi.nlm.nih.gov/18633328/

Miller, D. J., Nguyen, J. T., and Bottai, M. emagnification: a tool for estimating effect size magnification and performing design calculations in epidemiological studies. 2020. *Stata Journal* 20:3 *(forthcoming)*

For additional questions, contact David Miller (Miller.DavidJ@epa.gov)