Applying Systematic Review to Assessments of Health Effects of Chemical Exposures

EPA Workshop

August 26, 2013

Washington, DC

Using Systematic Review Methods to Strengthen Risk Assessments

Jonathan M. Samet, M.D., M.S.

Professor and Flora L. Thornton Chair,
Department of Preventive Medicine
USC Keck School of Medicine
Director, USC Institute for Global Health

EPA Workshop: Applying Systematic Review to Assessments of Health Effects of Chemical Exposures

EPA East Facility, Washington DC

August 26, 2013

What I am going to talk about!

- Systematic review/meta-analysis 101
- Extending systematic review to risk assessment
  - Hazard identification and weight-of-evidence
  - Dose-response
- Systematic review and the future of risk assessment

Searching for Truth: The Episcope”

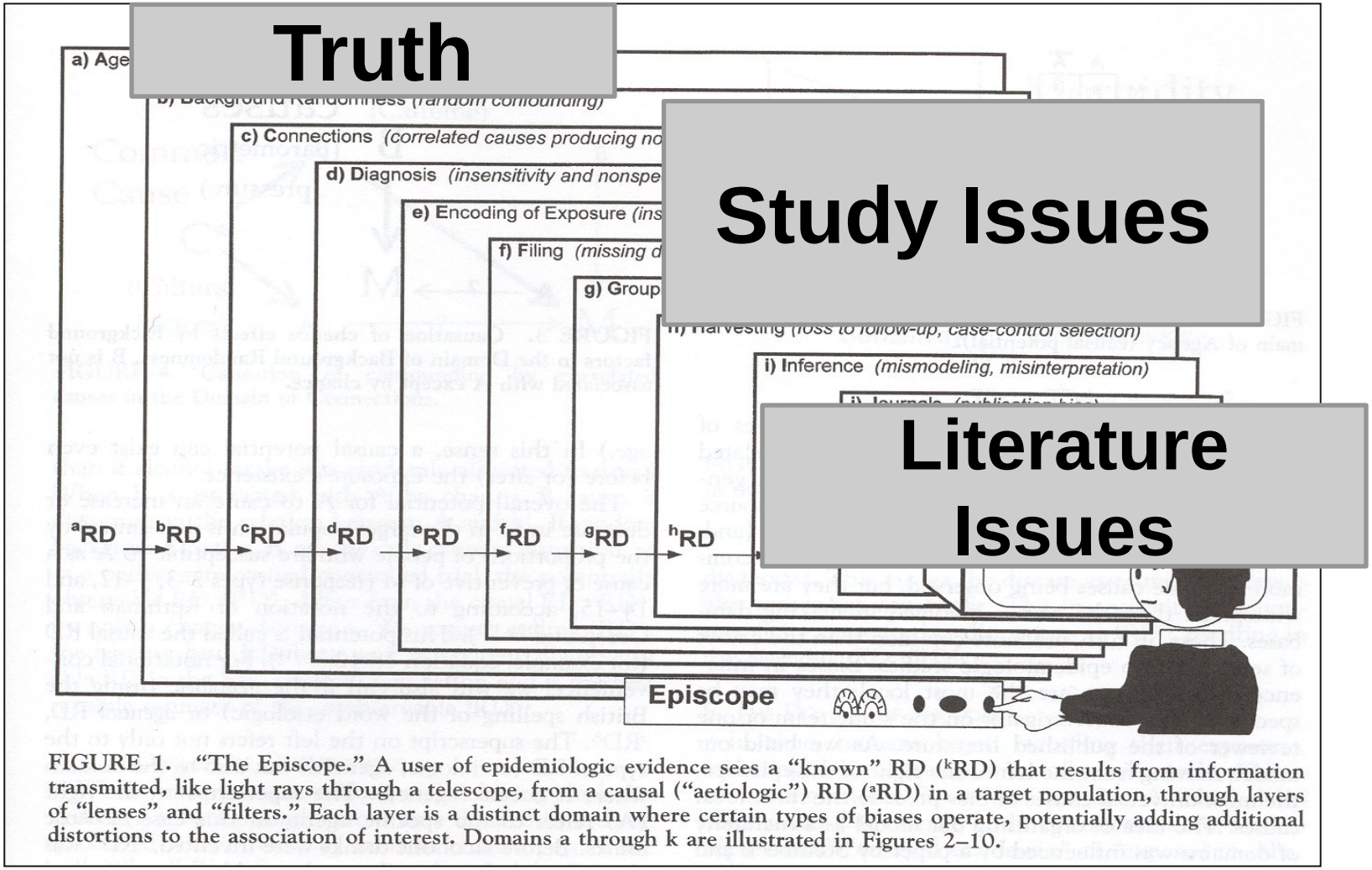
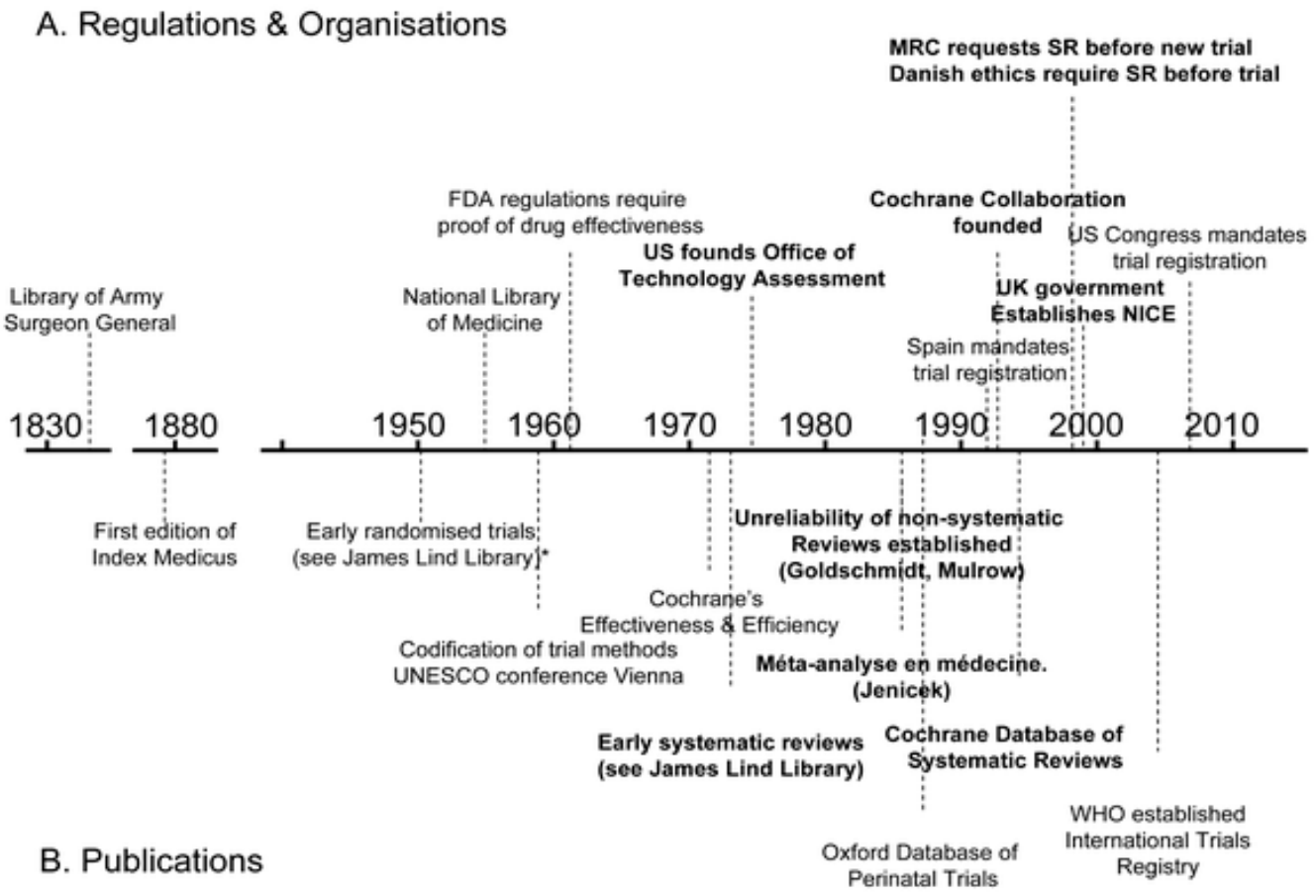


FIGURE 1. “The Episcope.” A user of epidemiologic evidence sees a “known” RD (^{k}RD) that results from information transmitted, like light rays through a telescope, from a causal (“aetiologic”) RD (^{a}RD) in a target population, through layers of “lenses” and “filters.” Each layer is a distinct domain where certain types of biases operate, potentially adding additional distortions to the association of interest. Domains a through k are illustrated in Figures 2–10.

(Maclure and Schneeweiss, Epidemiology, 2001)

“SYSTEMATIC REVIEW AND META-ANALYSIS: 101”

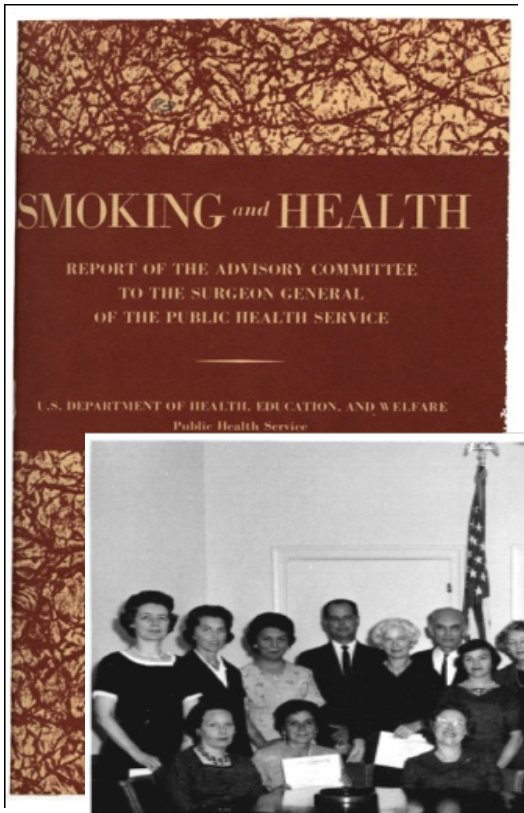
Figure 1. Milestones in the development of trials and the science of reviewing



Bastian H, Glasziou P, Chalmers I (2010) Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?. PLoS Med 7(9): e1000326. doi:10.1371/journal.pmed.1000326
http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.1000326

PLOS | MEDICINE

1964 Surgeon General's Report

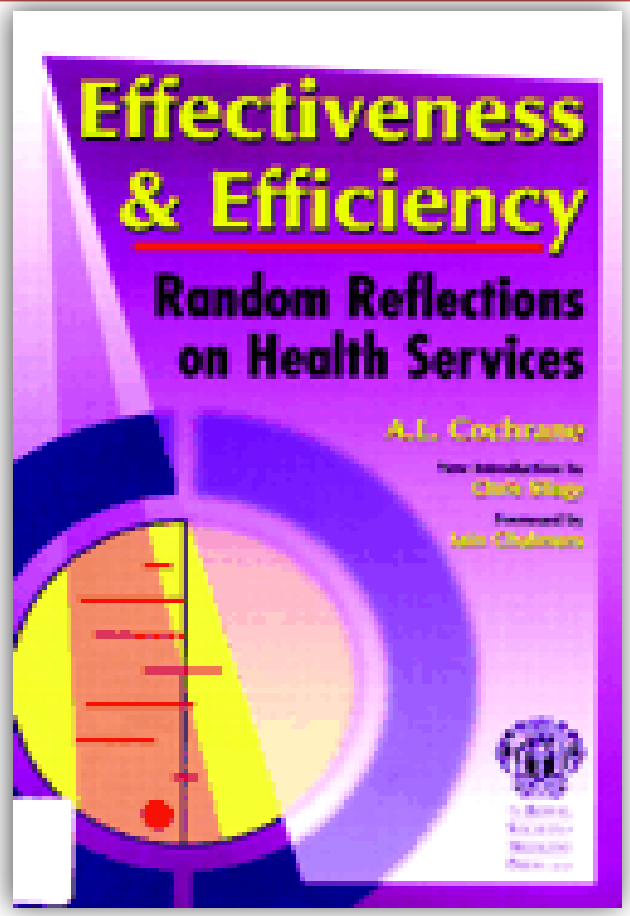


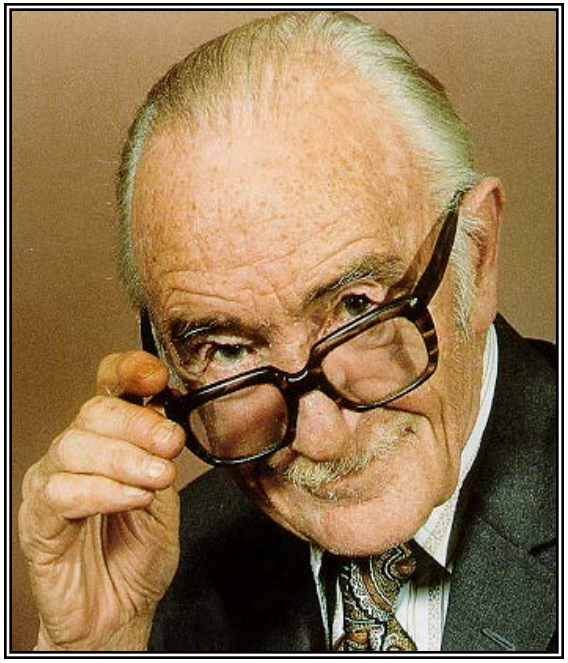
Causal Criteria

Statistical methods cannot establish proof of a causal relationship in an association. The causal significance of an association is a matter of judgment which goes beyond any statement of statistical probability. To judge or evaluate the causal significance of the association between the attribute or agent and the disease, or effect upon health, a number of criteria must be utilized, no one of which is an all-sufficient basis for judgment. These criteria include:

a) The consistency of the association
b) The strength of the association
c) The specificity of the association
d) The temporal relationship of the association
e) The coherence of the association

1964 U.S. Surgeon General's Report, p. 20





Archie Cochrane: Physician and respiratory epidemiologist who asked about evaluating the National Health Service

Cochrane: Systematic Review

"A systematic review is a high-level overview of primary research on a particular research question that tries to identify, select, synthesize and appraise all high quality research evidence relevant to that question in order to answer it."

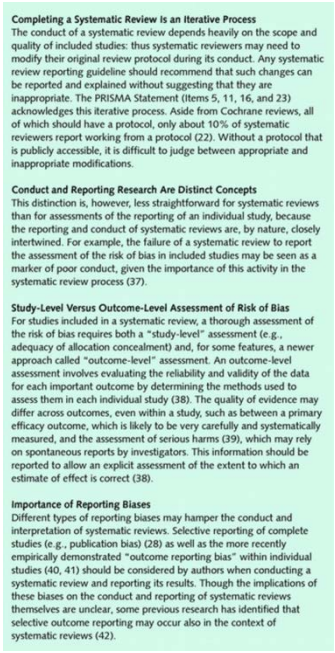
Cochrane Review: Key Points

- “Systematic reviews seek to collate all evidence that fits pre-specified eligibility criteria in order to address a specific research question
- Systematic reviews aim to minimize bias by using explicit, systematic methods
- The Cochrane Collaboration prepares, maintains and promotes systematic reviews to inform healthcare decisions: Cochrane Reviews”

Annals of Internal Medicine
Established in 1927 by the American College of Physicians

From: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

Ann Intern Med. 2009;151(4):264-269. doi:10.7326/0003-4819-151-4-200908180-00135



Completing a Systematic Review Is an Iterative Process
The conduct of a systematic review depends heavily on the scope and quality of included studies: thus systematic reviewers may need to modify their original review protocol during its conduct. Any systematic review reporting guideline should recommend that such changes can be reported and explained without suggesting that they are inappropriate. The PRISMA Statement (Items 5, 11, 16, and 23) acknowledges this iterative process. Aside from Cochrane reviews, all of which should have a protocol, only about 10% of systematic reviewers report working from a protocol (22). Without a protocol that is publicly accessible, it is difficult to judge between appropriate and inappropriate modifications.

Conduct and Reporting Research Are Distinct Concepts
This distinction is, however, less straightforward for systematic reviews than for assessments of the reporting of an individual study, because the reporting and conduct of systematic reviews are, by nature, closely intertwined. For example, the failure of a systematic review to report the assessment of the risk of bias in included studies may be seen as a marker of poor conduct, given the importance of this activity in the systematic review process (37).

Study-Level Versus Outcome-Level Assessment of Risk of Bias
For studies included in a systematic review, a thorough assessment of the risk of bias requires both a "study-level" assessment (e.g., adequacy of allocation concealment) and, for some features, a newer approach called "outcome-level" assessment. An outcome-level assessment involves evaluating the reliability and validity of the data for each important outcome by determining the methods used to assess them in each individual study (38). The quality of evidence may differ across outcomes, even within a study, such as between a primary efficacy outcome, which is likely to be very carefully and systematically measured, and the assessment of serious harms (39), which may rely on spontaneous reports by investigators. This information should be reported to allow an explicit assessment of the extent to which an estimate of effect is correct (38).

Importance of Reporting Biases
Different types of reporting biases may hamper the conduct and interpretation of systematic reviews. Selective reporting of complete studies (e.g., publication bias) (28) as well as the more recently empirically demonstrated "outcome reporting bias" within individual studies (40, 41) should be considered by authors when conducting a systematic review and reporting its results. Though the implications of these biases on the conduct and reporting of systematic reviews themselves are unclear, some previous research has identified that selective outcome reporting may occur also in the context of systematic reviews (42).

Figure Legend:

Conceptual Issues in the Evolution From QUOROM to PRISMA

Annals of Internal Medicine
Established in 1927 by the American College of Physicians

From: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

Ann Intern Med. 2009;151(4):264-269. doi:10.7326/0003-4819-151-4-200908180-00135

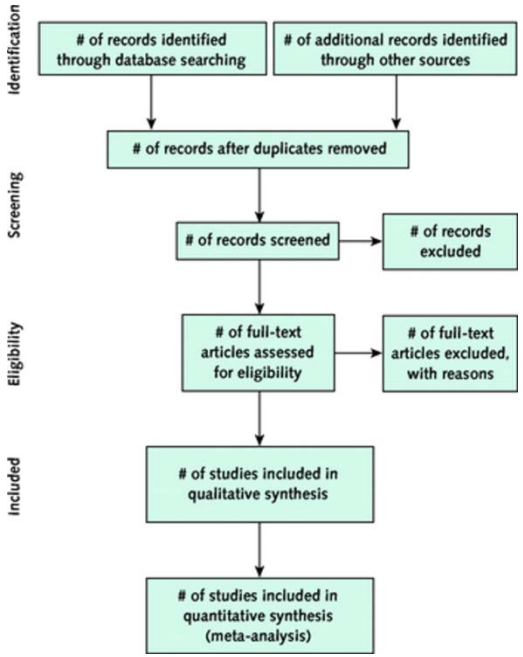


Figure Legend:

Flow of information through the different phases of a systematic review.

Cochrane: Meta-Analysis

"Meta-analysis is the use of statistical methods to summarize the results of independent studies (Glass 1976).

By combining information from all relevant studies, meta-analyses can provide more precise estimates of the effects of health care than those derived from the individual studies included within a review. They also facilitate investigations of the consistency of evidence across studies, and the exploration of differences across studies."

Evidence Table

Table 1. Cardiovascular Events and Outcomes by Randomized Treatment (cont)

Source	No. of Subjects	Mean Follow-up, y	Intervention	No. of Subjects					
				CHD*	Stroke*	CHF*	Major Cardiovascular Events	Total Mortality	Cardiovascular Disease Mortality†
FACET,[58] 1998	191	2.5	Dihydropyridine CCB	13	10	0	23	5	NA
	189		ACE inhibitor	10	4	0	14	4	NA
UKPDS,[59,60] 1998	400	8.4	ACE inhibitor	61	21	12	94	75	48
	358		ß-Blocker	46	17	9	72	59	34
CAPPP,[13] 1999	5492	6.1	ACE inhibitor	162	189	75	363	0.93 (0.76-1.14)‡	76
	5493		ß-Blockers or diuretics	161	148	66	335	1.00	95
NICSEH,[61] 1999	204	4.2	Dihydropyridine CCB	2	8	0	11	2	2
	210		Diuretics	2	8	3	12	2	0
STOP-2,[14] 1999	2196	5.0	Dihydropyridine CCB	179	207	186	450	362	212
	2213		ß-Blockers or diuretics	154	237	177	460	369	221
	2205		ACE inhibitor	139	215	149	437	380	226
INSIGHT,[9] 2000	3157	3.5	Dihydropyridine CCB	77	67	26	200	153	60
	3164		Diuretics	61	74	12	182	152	52
NORDIL,[10] 2000	5410	4.5	Nondihydropyridine CCB	183	159	63	466	231	131
	5471		ß-Blockers or diuretics	157	196	53	453	228	115
ALLHAT,[12] 2000	9067	3.3	α-Blockers	365	244	491	1592	514	130
	15 268		Diuretics	608	351	420	2245	851	218
AASK,[62-64] 2001 and 2002	436	3.0	ACE inhibitor	NA	NA	NA	0.59 (0.40-0.83)‡	18	NA
	217		Dihydropyridine CCB	NA	NA	NA	1.00	13	NA
	441		ß-Blocker	NA	NA	NA	0.52 (0.35-0.74)‡	NA	NA
PROGRESS,[65] 2001	1281	3.9	ACE inhibitor	48	157	NA	227	NA	93
	1280		Placebo	52	165	NA	237	NA	77
	1770		ACE inhibitor and diuretics	67	150	NA	231	NA	88
	1774		Placebo	102	255	NA	367	NA	121
IDM,[66] 2001	194	2.0	High-dose ARB	NA	NA	NA	9	3	NA
	195		Low-dose ARB	NA	NA	NA	NA	0	NA
	201		Placebo	NA	NA	NA	17	1	NA
Lewis et al,[67] 2001	579	2.6	ARB	NA	NA	NA	138	87	NA
	567		Dihydropyridine CCB	NA	NA	NA	128	83	NA
	569		Placebo	NA	NA	NA	144	93	NA
LIFE,[11] 2002	4605	4.7	ARB	198	232	153	508	383	204
	4588		ß-Blocker	188	309	161	588	431	234
CONVINCE,[70] 2002§	8179	3.0	Nondihydropyridine CCB	133	133	126	364	NA	152
	8297		ß-Blockers or diuretics	166	118	100	365	NA	143
ELSA,[71] 2002	1157	3.8	ß-Blocker	17	14	NA	33	17	8
	1177		Dihydropyridine CCB	18	9	NA	27	13	4
ALLHAT,[15] 2002	15 255	4.9	Diuretics	1362	675	870	3941	2203	992
	9048		Dihydropyridine CCB	798	377	706	2432	1256	592
	9054		ACE inhibitor	796	457	612	2514	1314	609
ANBP2,[68,69] 2002 and 2003	3044	4.1	ACE inhibitor	173	112	69	490	195	84
	3039		Diuretics	195	107	78	529	210	82

Abbreviations: AASK, African American Study of Kidney Disease and Hypertension; ABCD, Appropriate Blood Pressure Control in Diabetes; ACE, angiotensin-converting enzyme; ALLHAT, Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial; ANBPS, Australian National Blood Pressure Study; ANBP2, Australian National Blood Pressure 2 Trial; ARB, angiotensin II type 1 receptor blockers; CAPPP, Captopril Prevention Project; CCB, calcium channel blockers; CHD, coronary heart disease; CHF, congestive heart failure; CONVINCE, Controlled Onset Verapamil Investigation of Cardiovascular Endpoints; Dutch TIA, Dutch Transient Ischemic Attack Trial Study Group; ELSA, European Lacidipine Study on Atherosclerosis; EWPHE, European Working Party on High Blood Pressure in the Elderly; FACET, Fosinopril versus Amlodipine Cardiovascular Events Trial; HAPPHY, Heart Attack Primary Prevention in Hypertension Trial Research Group; HDFP, Hypertension Detection and Follow-up Program Cooperative Group; HSCSG, Hypertension Stroke Cooperative Study Group; IDM, Irbesartan in Patients with Type 2 Diabetes and Microalbuminuria study; INSIGHT, Intervention as a Goal in Hypertension Treatment; LIFE, Losartan Intervention For Endpoint Reduction in Hypertension Study; MIDAS, Multicenter Isradipine Diuretic Atherosclerosis Study; MRC, Medical Research Council Working Party; NA, not available; NICSEH, National Intervention Cooperative Study in Elderly Hypertensives; NORDIL, Nordic Diltiazem Study; PATS, Post-Stroke Antihypertensive Treatment Study; PROGRESS, Perindopril Protection Against Recurrent Stroke Study; SHEP, Systolic Hypertension in the Elderly Program; STOP, Swedish Trial in Old Patients with Hypertension; SYST-EUR, Systolic Hypertension in Europe Trial; TEST, Tenormin After Stroke and Transient Ischemic Attack; UKPDS, UK Prospective Diabetes

Psaty et al. JAMA 2003; 289:2534-2544

Forest Plots

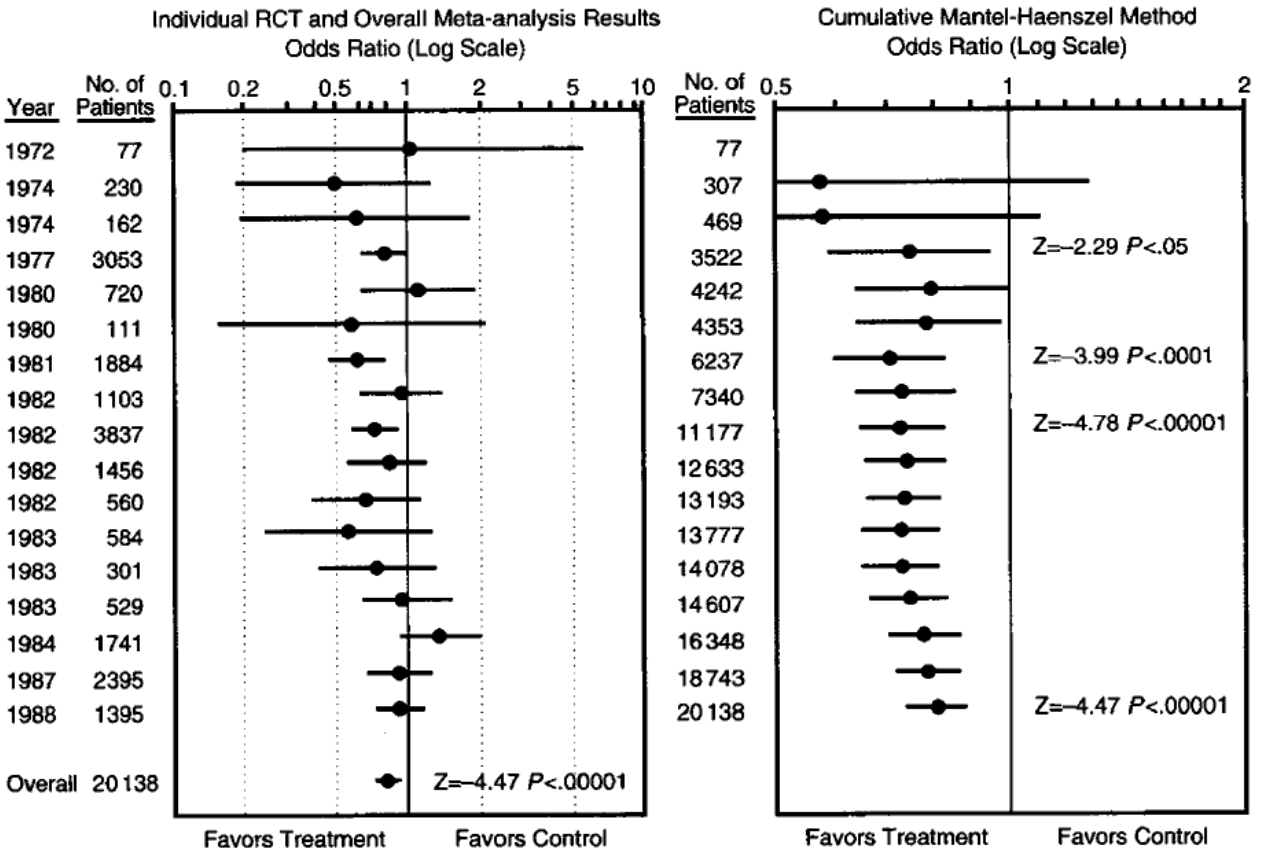
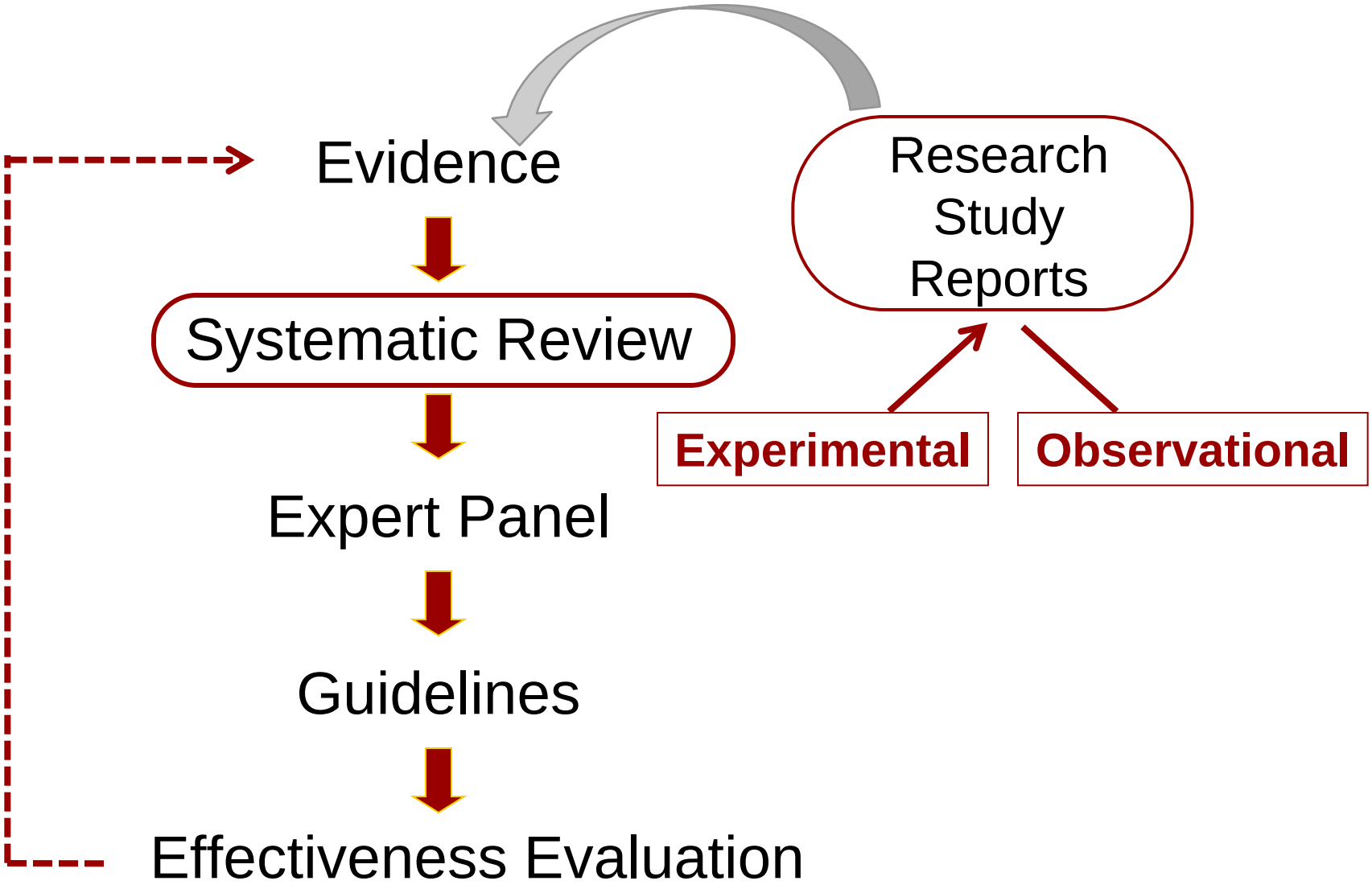


Fig 1.—Results of 17 randomized control trials (RCTs) of the effects of oral β-blockers for secondary prevention of mortality in patients surviving a myocardial infarction presented as two types of meta-analyses. On the left is the traditional one, revealing many trials with nonsignificant results but a highly significant estimate of the pooled results on the bottom of the panel. On the right, the same data are presented as cumulative meta-analyses, illustrating that the updated pooled estimate became statistically significant in 1977 and has remained so up to the present. Note that the scale is changed on the right graph to improve clarity of the confidence intervals.

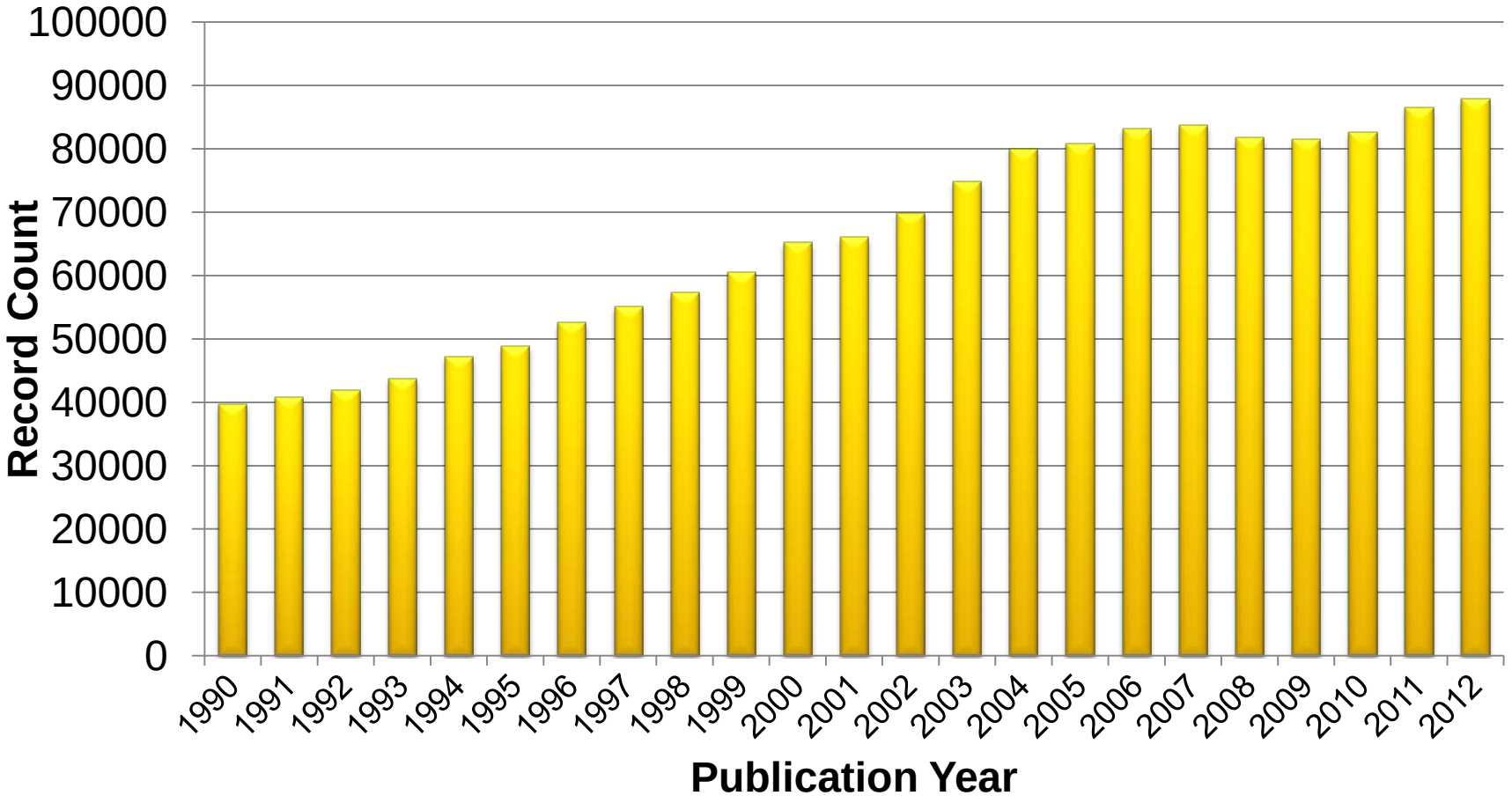
Antman *et al.* JAMA 1992;268:240-248.

From Evidence to Guidelines



PubMed Citation Analysis:

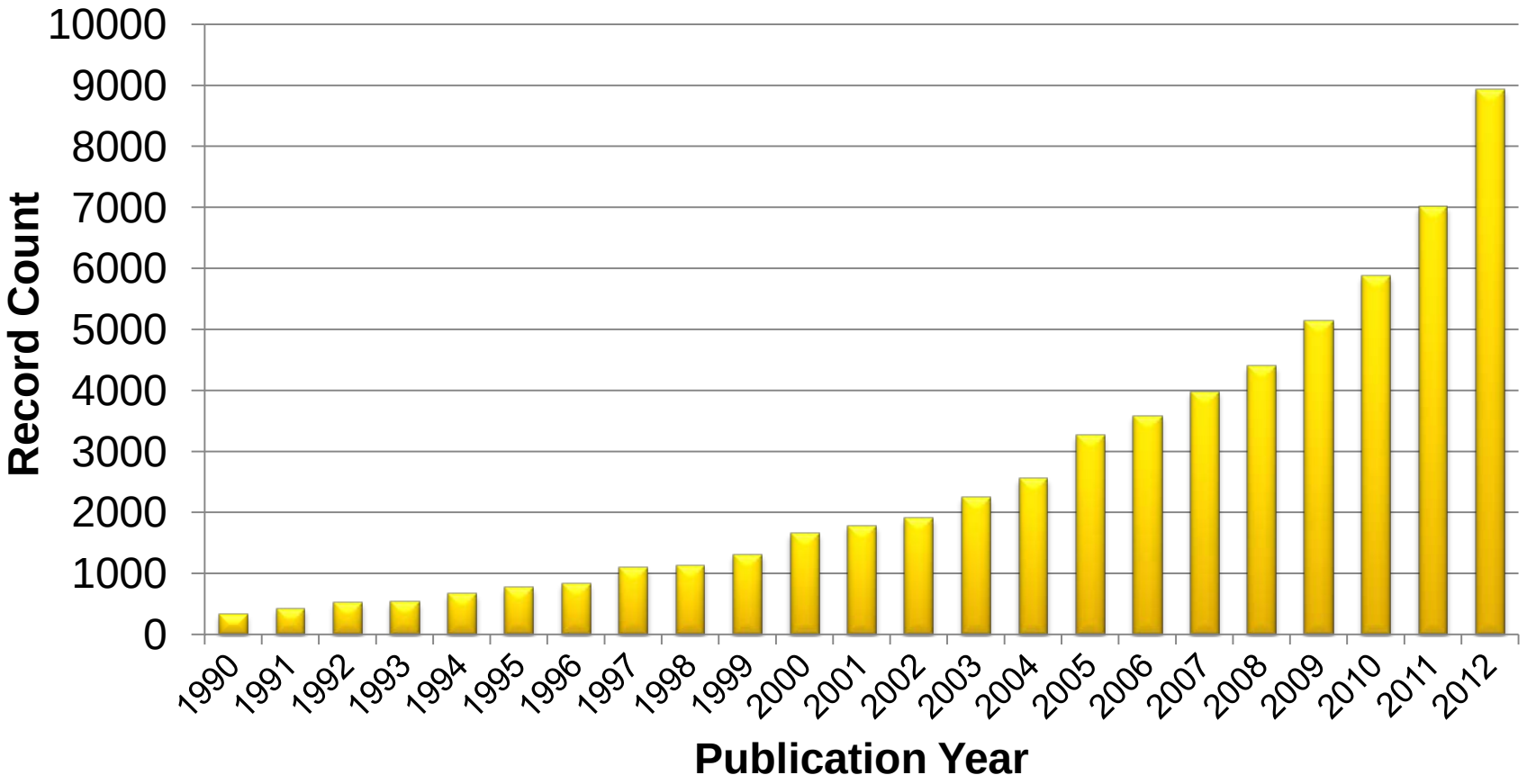
“Systematic review”



Search string: "review"[Publication Type] OR "review literature as topic"[MeSH Terms] OR "systematic review"[All Fields]

PubMed Citation Analysis:

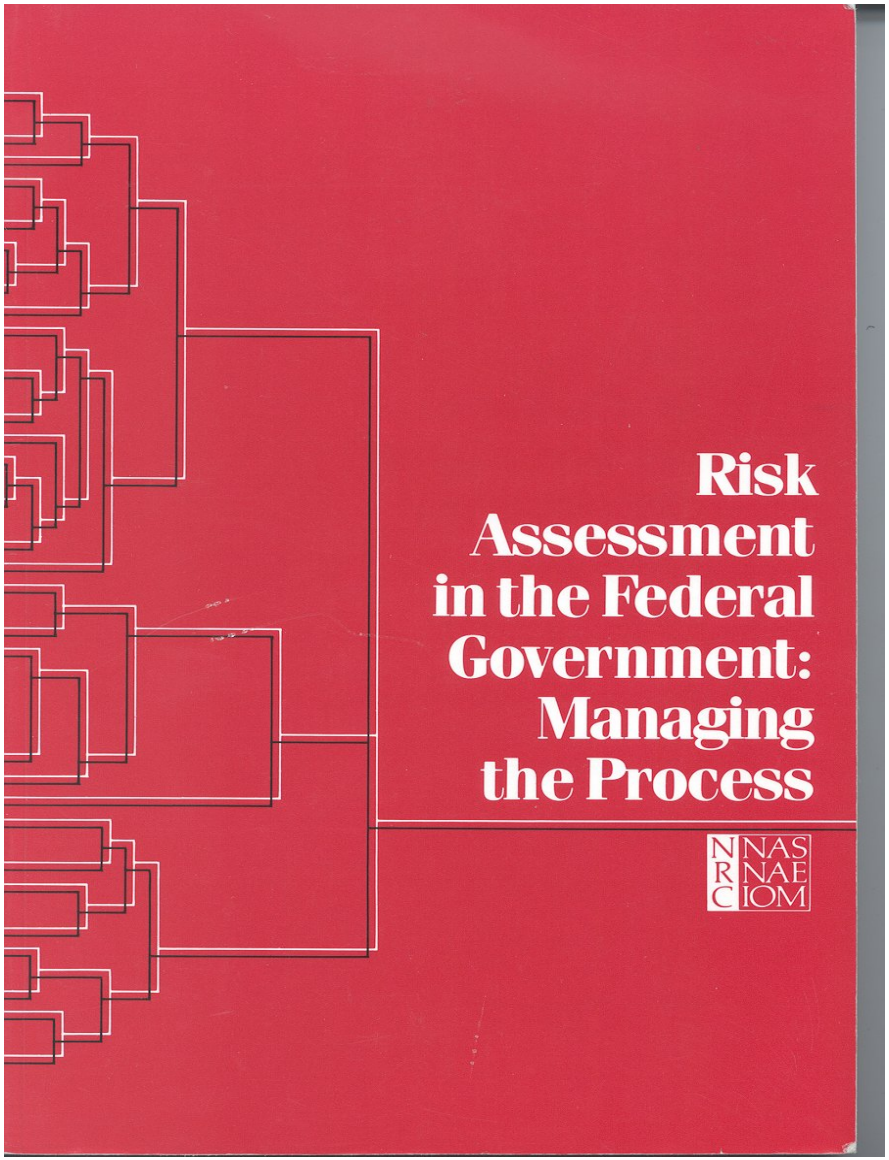
“Meta-analysis”



Search string: "meta-analysis"[Publication Type] OR "meta-analysis as topic"[MeSH Terms] OR "meta-analysis"[All Fields]

SYSTEMATIC REVIEW, META-ANALYSIS AND RISK ASSESSMENT

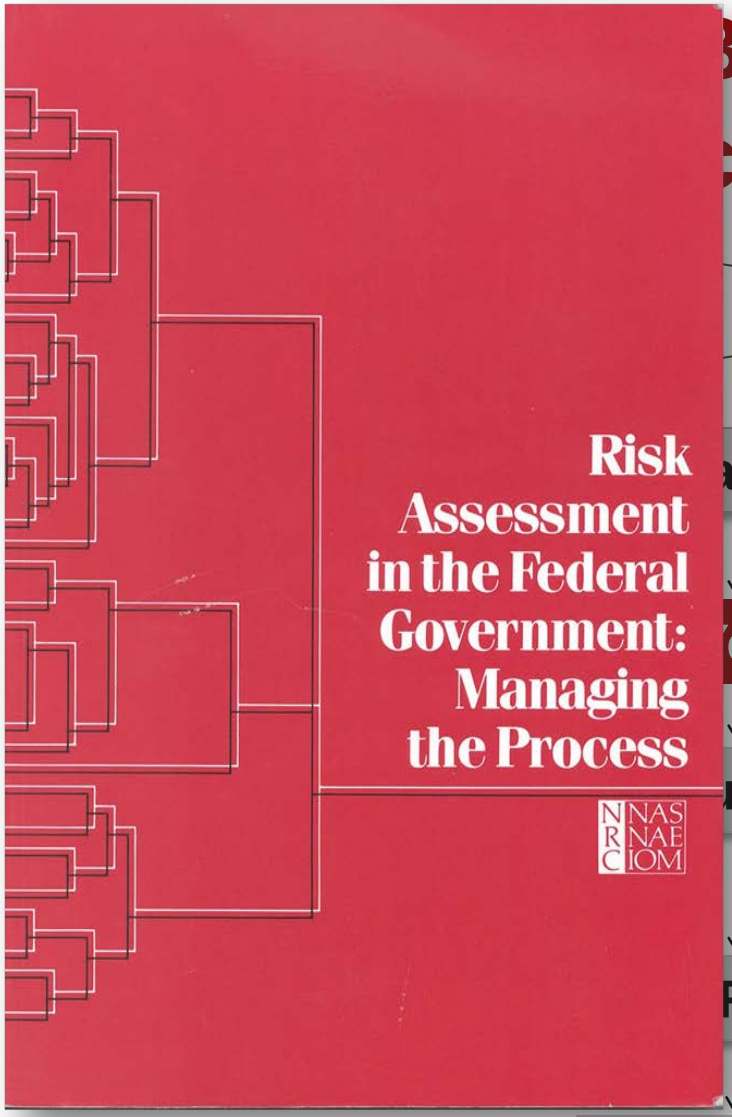
The "Red Book"

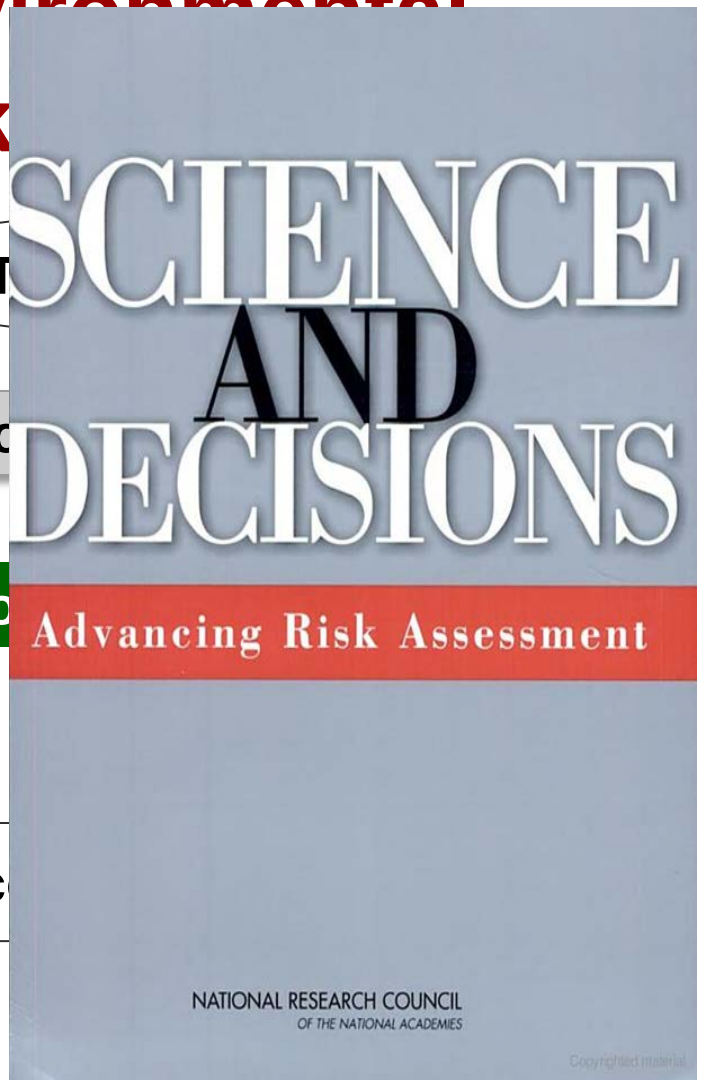


Elements of QRA

- Hazard ID
- Dose-response
- Exposure assessment
- Risk characterization

Based Envi
cision Mak





T

azard Identificatio

es

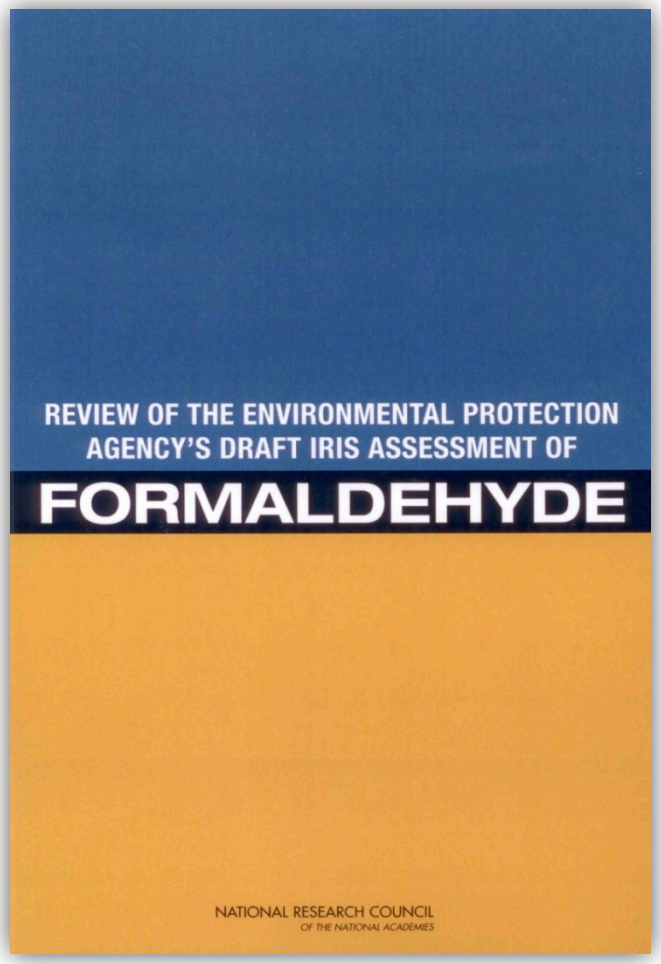
No

ch risk?

Toxic

RA

Decisions
Regulation

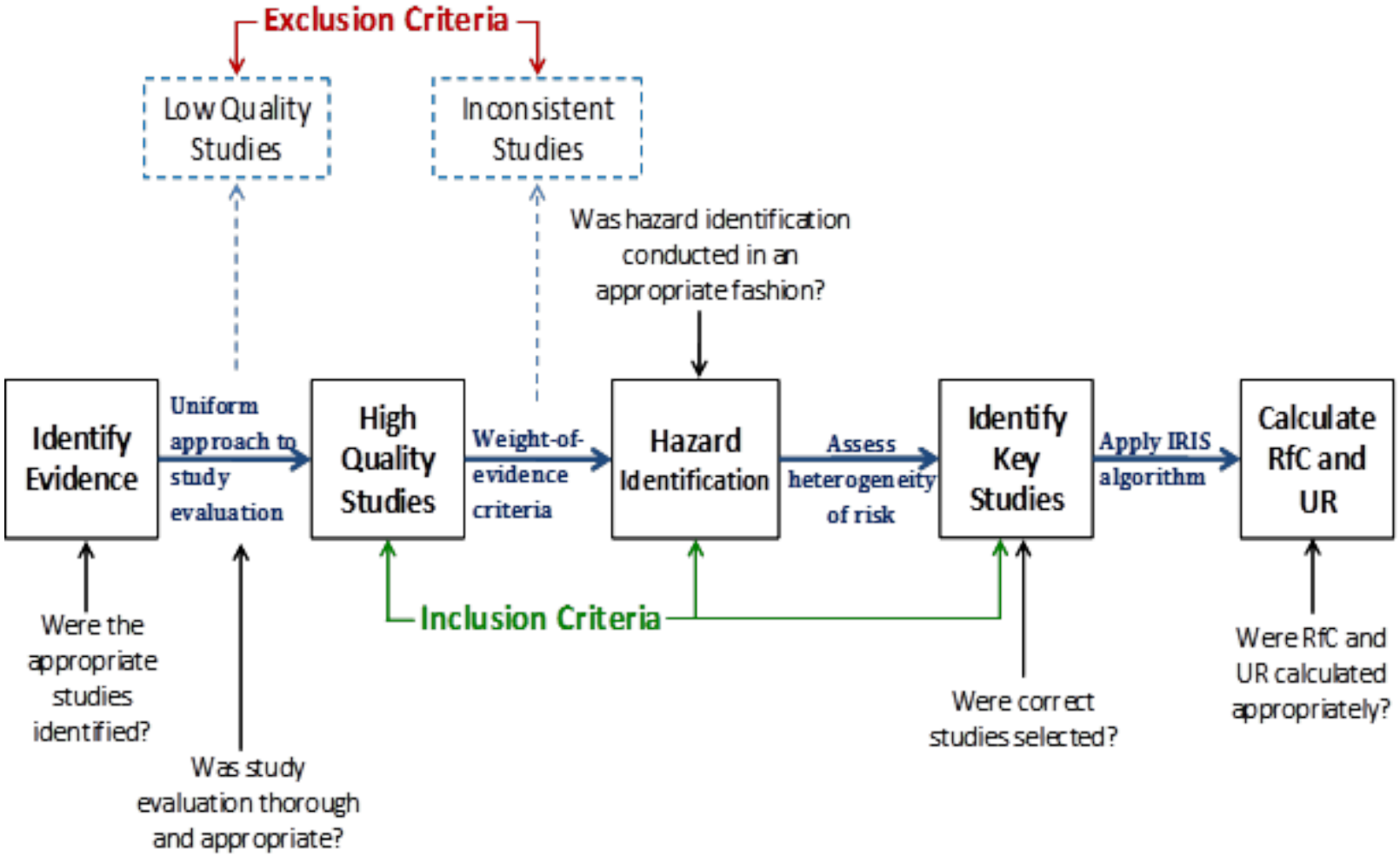


Chapter 7: A Roadmap for Revision

- Need to fully reassess and revise the IRIS process
- Problems with formaldehyde noted in prior reviews
- State-of-Art processes not followed throughout
- Lack of transparency in review and evidence evaluation
- Weight of evidence analyses inadequate

National Academies' National Research Council (NRC), April 2011

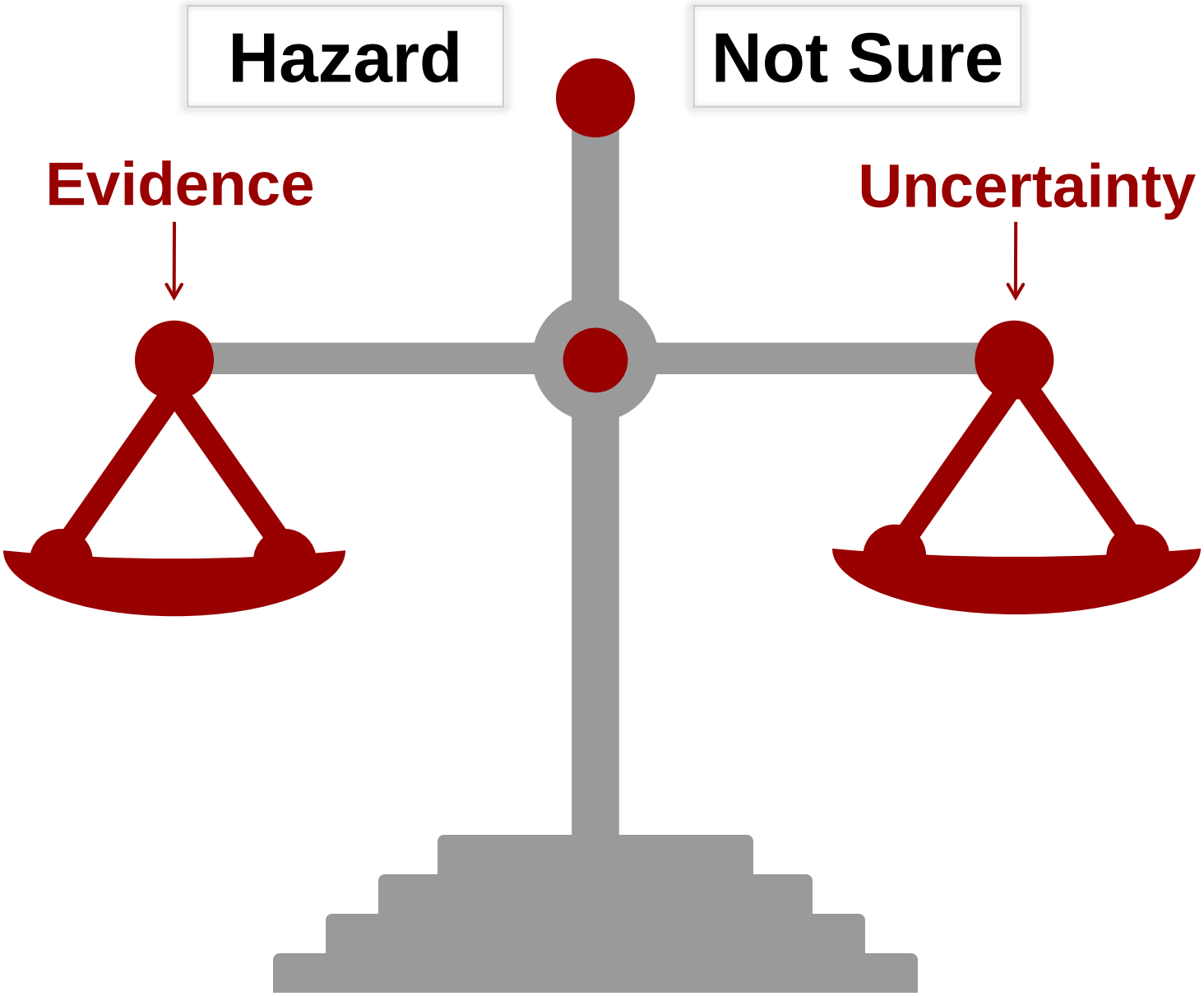
Steps of IRIS Assessments



Hazard Identification

- Description of the underlying question
- Identification of *all* relevant evidence in a transparent way
- Systematic capture of the evidence
- Evaluation of the evidence
- Documented use of weight-of-evidence criteria

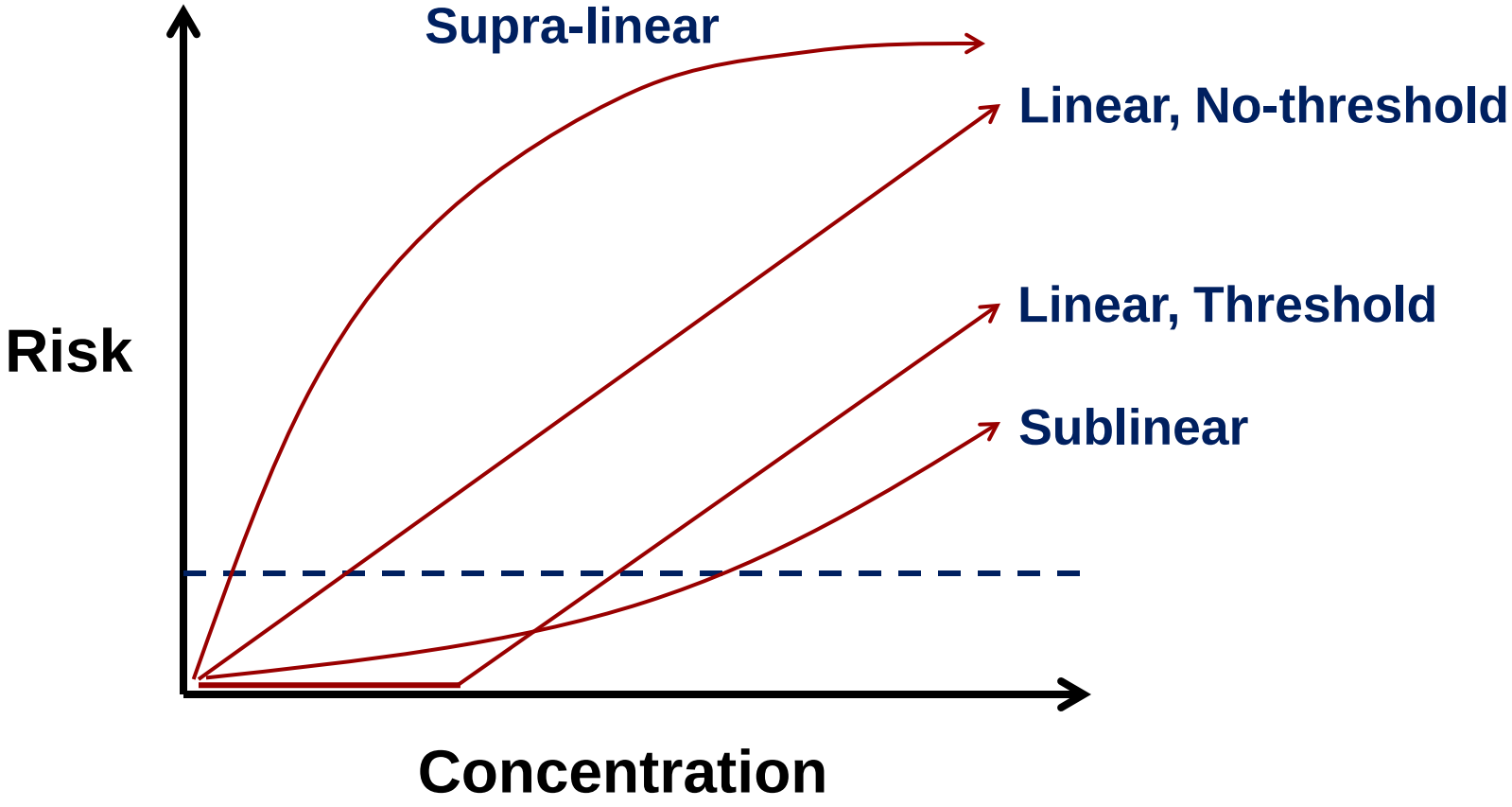
The Evidence Scale

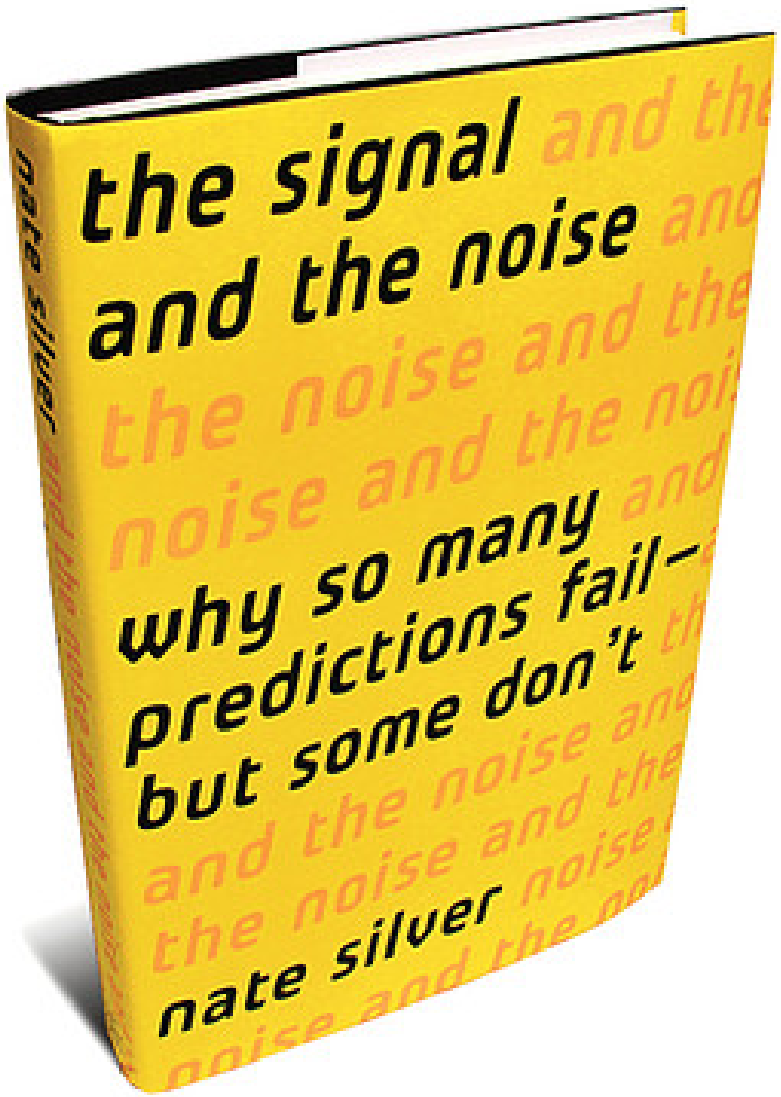


Dose-Response Assessment

- Role of systematic review:
  - Identify the suite of relevant dose-response relationships
  - Examine heterogeneity
  - Characterize the range of risk estimates and determinants of heterogeneity

What is the Form of the Relationship?





As the statistician George E.P. Box wrote,

> ***"All models are wrong, but some models are useful."***

What he meant by that is that all models are simplifications of the universe, as they must necessarily be. As another mathematician said,

> ***"The best model of a cat is a cat."***

RADON AND LUNG CANCER RISK:

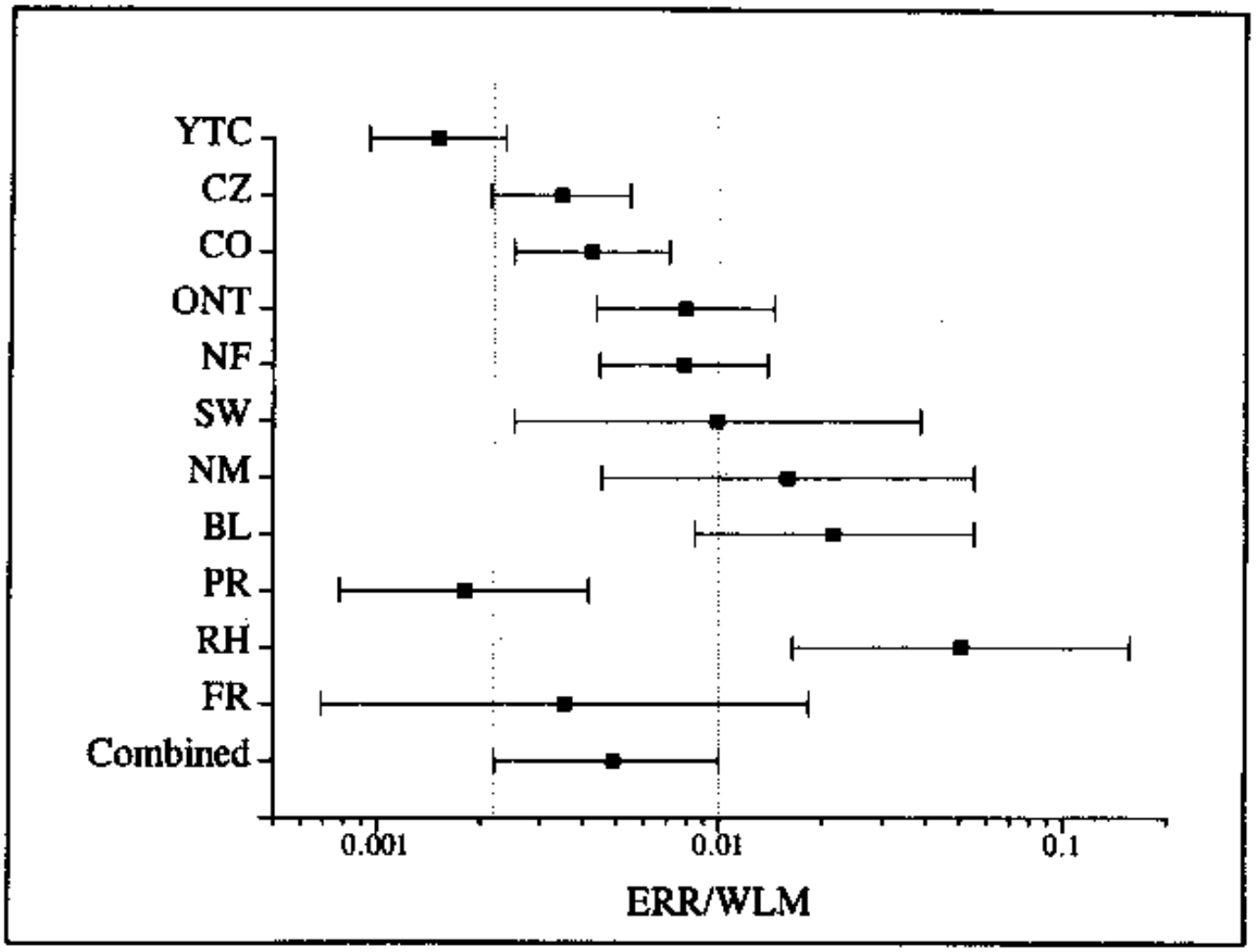
A JOINT ANALYSIS OF 11 UNDERGROUND MINERS STUDIES

JANUARY 1994

JAY H. LUBIN, JOHN D. BOICE, JR., CHRISTER EDLING, RICHARD W. HORNUNG, GEOFFREY HOWE, EMIL KUNZ, ROBERT A. KUSIAK, HOWARD I. MORRISON, EDWARD P. RADFORD, JONATHAN M. SAMET, MARGOT TIRMARCHE, ALISTAIR WOODWARD, YAO SHU XIANG, DONALD A. PIERCE

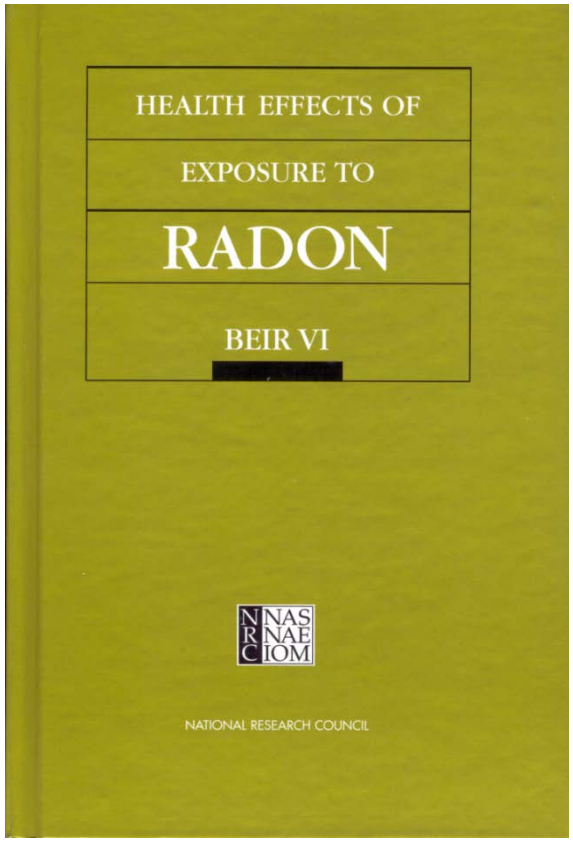
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health

NIH Publication No. 94-3644



Estimates of excess relative risk of lung cancer per WLM and 95% confidence limits for each cohort and for all data combined. Data taken from Table 5. Dotted lines show 95% CI for the combined ERR/WLM estimate based on random effects model.

BEIR VI: Assessing Radon's Risks



A Risk Model* For Lung Cancer and Radon

TSE/AGE/WL-cat model:

$$RR = 1 - \beta \times (w_{5\text{-}14} + \theta_2 w_{15\text{-}24} + \theta_3 w_{25+}) \times \phi_{age} \times \gamma_{WL}$$

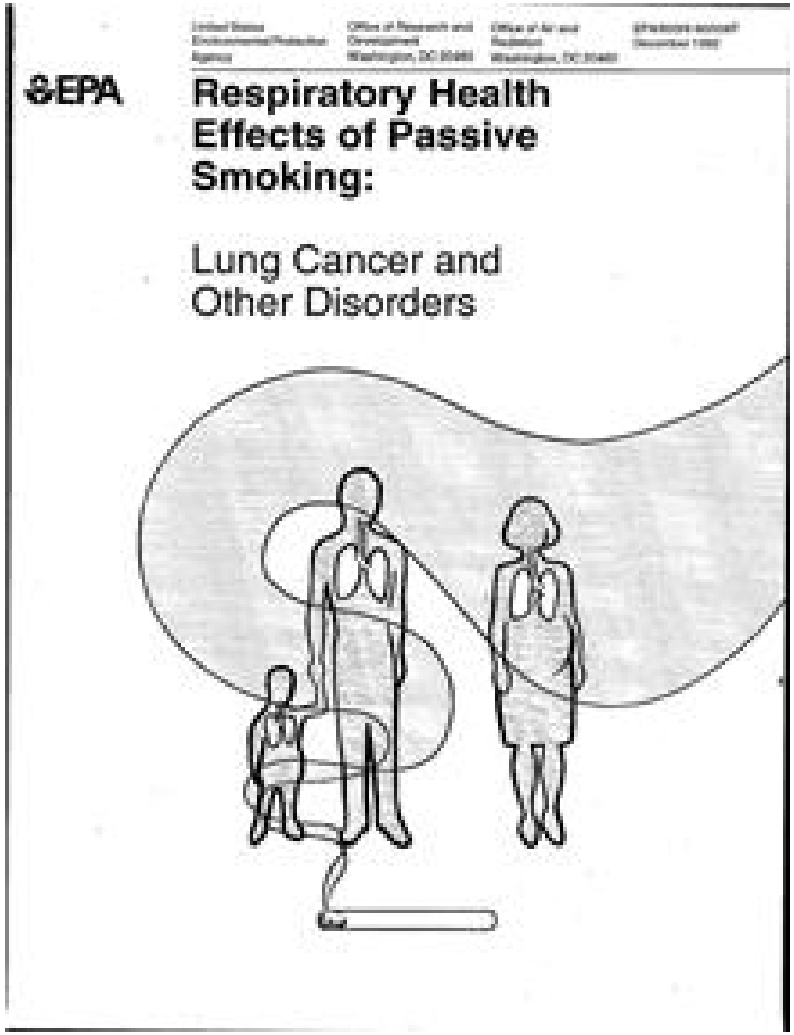
where $\beta = 0.0611$, $\theta_2 = 0.81$, $\theta_3 = 0.40$,

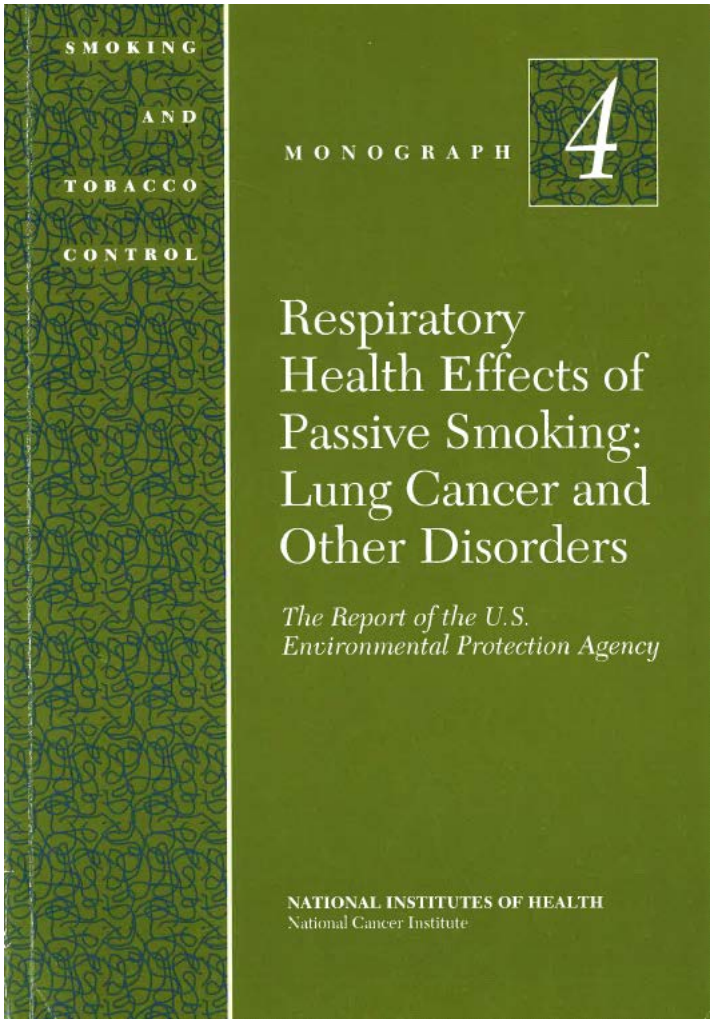
$$\phi_{age} = \begin{cases} 1.00 & \text{for age} < 55 \\ 0.65 & \text{for } 55 \le \text{age} < 65 \\ 0.38 & \text{for } 65 \le \text{age} < 75 \\ 0.22 & \text{for } 75 \le \text{age} \end{cases}$$

$$\gamma_{WL} = \begin{cases} 1.00 & \text{for WL} < 0.5 \\ 0.51 & \text{for } 0.5 \le \text{WL} < 1.0 \\ 0.32 & \text{for } 1.0 \le \text{WL} < 3.0 \\ 0.27 & \text{for } 3.0 \le \text{WL} < 5.0 \\ 0.13 & \text{for } 5.0 \le \text{WL} < 15.0 \\ 0.10 & \text{for } 15.0 \le \text{WL} \end{cases}$$

* Based on pooled analysis of 11 cohorts of miners.

1992 EPA review of ETS





Development of the EPA report

Due to the serious health concerns that have arisen regarding ETS, a virtually ubiquitous indoor air pollutant, and the wealth of new information that has become available since the extensive 1986 reviews, the EPA has performed its own analytical hazard identification and population risk assessment for the respiratory health effects of passive smoking, based on a critical review of the data currently available, with an emphasis on the abundant epidemiologic evidence. The number of lung cancer studies analyzed in this document is more than double the number reviewed in 1986 (31 vs. 13), with a total of about 3,000 lung cancer cases in female nonsmokers now reported in case-control studies and almost 300,000 female nonsmokers followed by cohort studies. Furthermore, the database on passive smoking and respiratory disorders in children contains more than 50 new studies, including 9 additional studies on acute lower respiratory tract illnesses, 10 on acute and chronic middle ear diseases, 18 on respiratory symptoms, 10 on asthma, and 8 on lung function. This report also discusses six recent studies of the effects of passive smoking on adult respiratory symptoms and lung function. Finally, eight studies of maternal smoking and sudden infant death syndrome (SIDS), which was not addressed in the NRC report or the Surgeon General's report, are reviewed. (Although the cause of SIDS is unknown, the most widely accepted hypotheses suggest that some form of respiratory pathogenesis is usually involved.)

EPA (1992). EPA/600/6-90/006F, p. 2-4

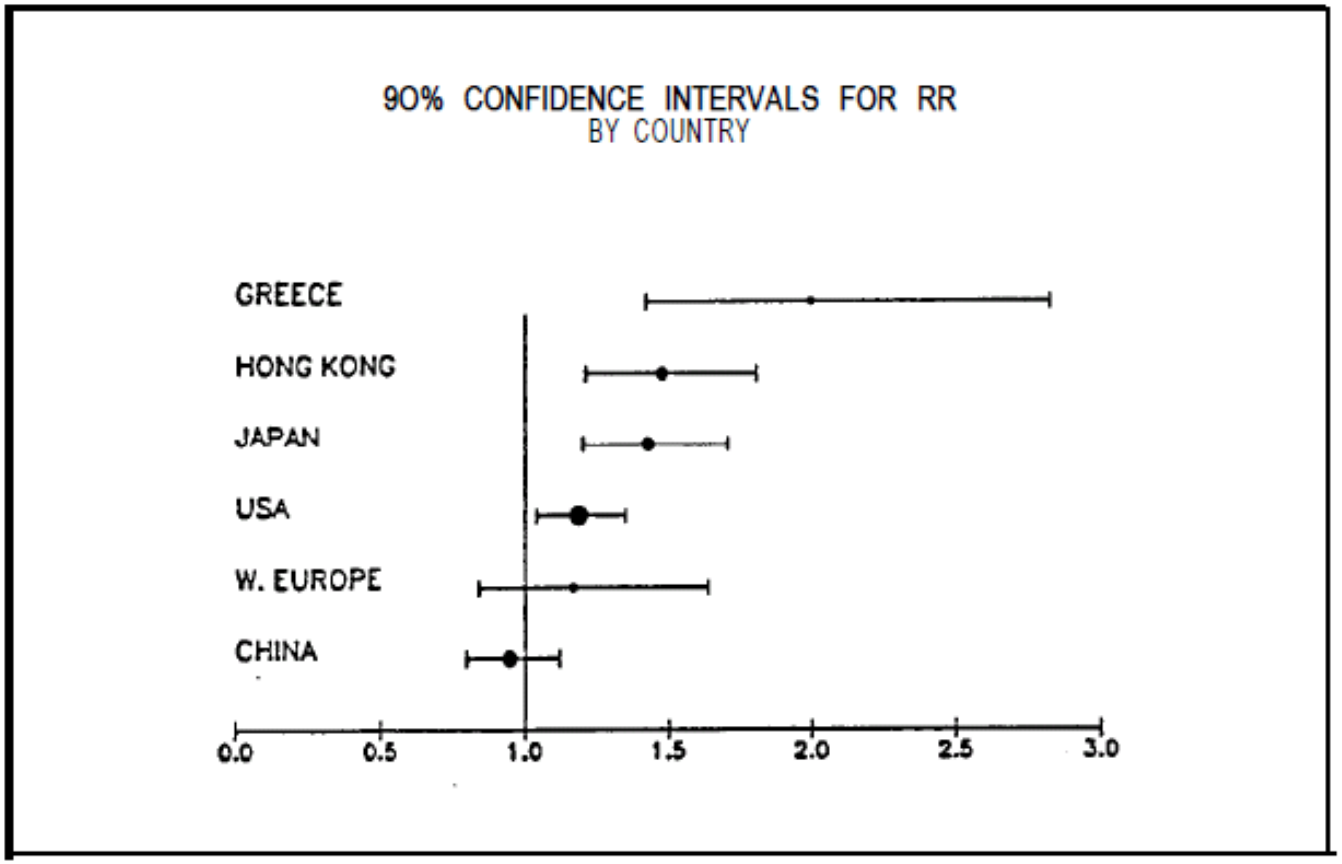


Figure 5-5. 90% confidence intervals, by country.

EPA (1992). EPA/600/6-90/006F, p. 5-35

The Attack on Meta-analysis

STATISTICS IN MEDICINE, VOL. 14, 545–569 (1995)

13321

META-ANALYTIC APPROACHES TO DOSE-RESPONSE RELATIONSHIPS, WITH APPLICATION IN STUDIES OF LUNG CANCER AND EXPOSURE TO ENVIRONMENTAL TOBACCO SMOKE

R. L. TWEEDIE AND K. [illegible]

Department of Statistics, Colorado State Unive

SUMMAR

This paper outlines several meta-analytic approaches relationships; that is, to the evaluation of an increase in the relative risk of a disease when this is investigated over a nu two levels: first, a consistent method of evaluating the dos and second, an overall picture is obtained by comparing stage, for an individual study, dose–response assessment in trend, which are influenced by such issues as dose measu second stage, different methods for pooling results across choices made in the first stage of analysis, with additional due to studies included in meta-analysis. We describe thes evaluating dose response. The approaches are illustrated b and levels of exposure to environmental tobacco smoke (a point of debate in recent assessment of evidence for an ov little indication of a consistent dose response, a result exp passive smoking developed by Darby and Pike, the curre current studies in Tweedie and Mengersen, and misclassi Environmental Protection Agency (EPA).

J Clin Epidemiol Vol. 44, No. 2, pp. 127–139, 1991
Printed in Great Britain

0895-4356/91 $3.00 + 0.00
Pergamon Press plc

META-ANALYSIS IN EPIDEMIOLOGY, WITH SPECIAL REFERENCE TO STUDIES OF THE ASSOCIATION BETWEEN EXPOSURE TO ENVIRONMENTAL TOBACCO SMOKE AND LUNG CANCER: A CRITIQUE

JOSEPH L. FLEISS[1] and ALAN J. GROSS[2]

[1]Columbia University, School of Public Health, 600 West 168 Street, New York, NY 10032 and
[2]Medical University of South Carolina, Charleston, SC 29425, U.S.A.

(Received in revised form 29 August 1990)

association, although as we have indicated, at least an equal burden of care is required for its valid implementation and interpretation.

ACKNOWLEDGEMENTS

This work was largely carried out at Bond University and at the University of Central Queensland. We are grateful for the input and assistance of Professor John Eccleston and Ms. Samantha Low Choy in the early stages of the ETS analysis. The input of Bill duMouchel and Tom Chalmers at the CDC Atlanta meeting is also gratefully acknowledged.

The paper was completed at Colorado State University, with partial support from several tobacco companies; the methods and analysis here are however entirely those of the authors and should not be otherwise ascribed.

We are grateful to David Williamson and the organizers of the 1993 CDC symposium on statistical methods, and to the referees of this paper, for encouragement and many valuable suggestions in its presentation.

CCC 0277–4715/95/060545–25

is, a set of statistical tools for combining and integrating the tudies of a given scientific issue, can be useful when the stringent such integration is valid are met. In this report we point out ning sound meta-analyses of either controlled clinical trials or . We demonstrate that hastily or improperly designed meta- ults that may not be scientifically valid. We note that much care meta-analysis is applied to the results of clinilcal trials. The Food ion, for example, requires strict adherence to the principles we fore it allows a drug's sponsor to use a meta-analysis of separate ort of a New Drug Application.

lways carry over to epidemiological studies, as demonstrated by National Research Council concerning the purported association nvironmental tobacco smoke and the risk of lung cancer. On the s of 13 studies, 10 of which were retrospective and the remaining e, the Council concluded that non-smokers who are exposed to o smoke are at greater risk of acquiring lung cancer than posed. In our opinion, this conclusion in unwarranted given the udies on which it is based.

1. INTRODUCTION

A working definition of meta-analysis is given

"yes." The criteria for reaching this affirmativ answer are now considered.

In applications of meta-analysis to clinica

mineral,[5] suggesting that stimuli associated with suckling are important in regulating calcium metabolism. Breast-feeding women have elevated serum concentrations of prolactin and parathyroid hormone–related peptide and low serum estradiol concentrations, all of which are likely to influence calcium handling and bone metabolism.

Overall, the evidence suggests that optimal lactation and maternal bone health do not depend on an increase in calcium intake by the breast-feeding mother. This tentative conclusion does not imply that good nutrition, including the maintenance of adequate calcium intake, is unimportant during lactation. However, the accumulating scientific data suggest that breast-feeding women need not consume extra calcium.

ANN PRENTICE, PH.D.

Medical Research Council Dunn Nutrition Unit
Cambridge CB4 1XJ, United Kingdom

REFERENCES

1. Prentice A, Jarjou LMA, Laskey MA. Lactation and bone development: implications for the calcium requirements of infants and lactating mothers. In: Tsang RC, Bonjour J-P, eds. Nutrition and bone development. New York: Raven Press (in press).
2. Prentice A. Maternal calcium requirements during pregnancy and lactation. Am J Clin Nutr 1994;59:Suppl:4775-483S.
3. Kalkwarf HJ, Specker BL, Bianchi DC, Ranz J, Ho M. The effect of calcium supplementation on bone density during lactation and after weaning. N Engl J Med 1997;337:523-8.
4. Cross NA, Hillman LS, Allen SH, Krause GF. Changes in bone mineral density and markers of bone remodeling during lactation and postweaning in women consuming high amounts of calcium. J Bone Miner Res 1995; 10:1312-20.
5. Laskey MA, Prentice A, Hanratty LA, et al. Bone changes after 3 months of lactation: influence of calcium intake, breast-milk calcium output and vitamin-D receptor genotype. Am J Clin Nutr (in press).
6. Sowers M, Corton G, Shapiro B, et al. Changes in bone density with lactation. JAMA 1993;269:3130-5.
7. Laskey MA, Prentice A. Effect of pregnancy on recovery of lactational bone loss. Lancet 1997;349:1518-9.
8. Prentice A, Jarjou LMA, Cole TJ, Stirling DM, Dibba B, Fairweather-Tait S. Calcium requirements of lactating Gambian mothers: effects of a calcium supplement on breast-milk calcium concentration, maternal bone mineral content, and urinary calcium excretion. Am J Clin Nutr 1995;62: 58-67.
9. Kalkwarf HJ, Specker BL, Heubi JE, Vieira NE, Yergey AL. Intestinal calcium absorption of women during lactation and after weaning. Am J Clin Nutr 1996;63:526-31.
10. Sowers M, Eyre D, Hollis BW, et al. Biochemical markers of bone turnover in lactating and nonlactating postpartum women. J Clin Endocrinol Metab 1995;80:2210-6.

PROMISE AND PROBLEMS META-ANALYSIS

ETA-ANALYSIS has acquired a substantial fol- lowing among both statisticians and clinicians. echnique was developed as a way to summarize esults of different research studies of related ems. Meta-analysis may be applied even when the studies are small and there is substantial variation in the specific issues studied, the research methods applied, the source and nature of the study subjects, and other factors that may have an important bearing on the findings. In this issue of the *Journal*, LeLorier et al.[1] compare the findings of 12 large randomized, controlled trials with the results of meta-analyses of the same problems. They find important discrepancies. When a large randomized, controlled trial — commonly considered the gold standard for determining the effects of medical interventions — disagrees with a meta-analysis, what should the reader conclude? Perhaps more important, when only one of the two tools is used, how much uncertainty should the reader add to the confidence limits and other statistical measures of uncertainty reported by the author?

The core of meta-analysis is its systematic approach to the identification and abstracting of critical information from research reports. Doing a meta-analysis correctly demands expertise in both the method and the substance and hence almost always requires collaboration between clinicians and an experienced statistician. The questions must be defined carefully to maximize the relevance of the reports to be included and to reduce uncertainties about procedures. The investigators must then try to find every relevant report by searching data bases, reviewing bibliographies, and asking widely about unpublished work. The collected reports are then winnowed to the few (often less than 10 percent) that meet the requirements for the meta-analysis. The reports must be searched carefully to identify problems and validate the quantitative findings of interest. These findings must be expressed on a common scale (often as odds ratios), and some reports may have to be dropped for lack of information. Those doing a meta-analysis may also abstract information from each report to produce a quantitative measure of research quality. Each of the individual quantitative estimates must be scrutinized for problems, and this may require the efforts of a range of specialists. When the analysis is completed and submitted for publication, the editor and the reviewers must assure themselves of its quality. A rigorous technical review of a meta-analysis requires the reviewer to identify, reabstract, and interpret a fair sample of the original papers. Very few editors and reviewers will do this, which may be one reason why there are so many poor meta-analyses in the literature.

Although some meta-analyses stop with the presentation and discussion of the results of the individual studies, many others proceed further and combine the results into a single, comprehensive "best" estimate, generally with statistical confidence bounds, that is meant to summarize what is known about the clinical problem. This last step — preparing and presenting a single estimate as the distillation of all that

Acknowledgements—This research was supported by a grant from The Tobacco Institute, Washington, D.C., U.S.A.

We thank Dr Myron Weinberg, President of the Weinberg Group/WASHTECH, for encouraging us to develop this critique.

Report of the Committee to Review the IRIS Process

Coming Attractions

Applying Systematic Review to Assessments of Health Effects of Chemical Exposures

Session 1

The Newcastle-Ottawa Scale: A Springboard for Evaluating Epidemiology

Glinda S. Cooper, Ph.D.
US EPA – ORD – NCEA - IRIS

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

United States Environmental Protection Agency

Outline

- Motivation for the talk
- The Newcastle-Ottawa Scale
  - description
  - as springboard
    - What do we want to know?
    - Documentation
    - Use

EPA
United States
Environmental Protection
Agency

- Different diseases, exposures, journals
- Each used Newcastle-Ottawa Scale
- "Used the scale" ✓ ….
  …but never mentioned it again

How do we evaluate methods/quality/strengths/ limitations/bias of a study (or a set of studies)?

How do we incorporate information on methods/quality/strengths/limitations/bias in our evaluation of a study (or a set of studies)?

The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analysis

Developed by George Wells, Beverley Shea, Peter Tugwell et al.

http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
COHORT STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability

Selection

1) Representativeness of the exposed cohort
   a) truly representative of the average _____ (describe) in the community **p**
   b) somewhat representative of the average _____ in the community **p**
   c) selected group of users eg nurses, volunteers
   d) no description of the derivation of the cohort
2) Selection of the non exposed cohort
   a) drawn from the same community as the exposed cohort **p**
   b) drawn from a different source
   c) no description of the derivation of the non exposed cohort
3) Ascertainment of exposure
   a) secure record (eg surgical records) **p**
   b) structured interview **p**
   c) written self report
   d) no description
4) Demonstration that outcome of interest was not present at start of study
   a) yes **p**
   b) no

Comparability

1) Comparability of cohorts on the basis of the design or analysis
   a) study controls for _____ (select the most important factor) **p**
   b) study controls for any additional factor **p** (This criteria could be modified to indicate specific control for a second important factor.)

Outcome

1) Assessment of outcome
   a) independent blind assessment **p**
   b) record linkage **p**
   c) self report
   d) no description
2) Was follow-up long enough for outcomes to occur
   a) yes (select an adequate follow up period for outcome of interest) **p**
   b) no
3) Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for **p**
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > ____ % (select an adequate %) follow up, or description provided of those lost) **p**
   c) follow up rate < ____% (select an adequate %) and no description of those lost
   d) no statement

Cohort Studies

NOTE:
Short! (8 items)

"Stars"

3 categories:
Selection
Comparability
Outcome

NEWCASTLE - OTTAWA QUALITY ASSESSMENT SCALE
CASE CONTROL STUDIES

Note: A study can be awarded a maximum of one star for each numbered item within the Selection and Exposure categories. A maximum of two stars can be given for Comparability.

Selection

1) Is the case definition adequate?
   a) yes, with independent validation **p**
   b) yes, eg record linkage or based on self reports
   c) no description

2) Representativeness of the cases
   a) consecutive or obviously representative series of cases **p**
   b) potential for selection biases or not stated

3) Selection of Controls
   a) community controls **p**
   b) hospital controls
   c) no description

4) Definition of Controls
   a) no history of disease (endpoint) **p**
   b) no description of source

Comparability

1) Comparability of cases and controls on the basis of the design or analysis
   a) study controls for _____ (Select the most important factor.) **p**
   b) study controls for any additional factor **p** (This criteria could be modified to indicate specific control for a second important factor.)

Exposure

1) Ascertainment of exposure
   a) secure record (eg surgical records) **p**
   b) structured interview where blind to case/control status **p**
   c) interview not blinded to case/control status
   d) written self report or medical record only
   e) no description

2) Same method of ascertainment for cases and controls
   a) yes **p**
   b) no

3) Non-Response rate
   a) same rate for both groups **p**
   b) non respondents described
   c) rate different and no designation

Case-Control Studies

NOTE:
Also Short! (8 items)
"Stars"

3 categories:
- Selection
- Comparability
- Exposure

Cohort Study: Outcome Assessment

1. Assessment of outcome
   a) independent blind assessment ★
   b) record linkage ★
   c) self report
   d) no description

2. Was follow up long enough for outcomes to occur
   a) yes (select an adequate follow up period for outcome of interest) ★
   b) no

3. Adequacy of follow up of cohorts
   a) complete follow up - all subjects accounted for ★
   b) subjects lost to follow up unlikely to introduce bias - small number lost - > ___ % (select an adequate %) follow up, or description of those lost) ★
   c) follow up rate < ___% (select an adequate %) and no description of those lost
   d) no statement

Case-Control Study: Selection Assessment

1. Is the case definition adequate?

   a) yes, with independent validation (e.g. >1 person/record/time/process to extract information, or reference to primary record source such as x-rays or medical/hospital records) ★

   b) yes, e.g. record linkage or based on self reports (ICD or self-report with no reference to primary record or no description)

   c) no description

2. Representativeness of the cases

   a) consecutive or obviously representative series of cases ★

   b) potential for selection biases or not stated

3. Selection of Controls

   a) community controls ★

   b) hospital controls

   c) no description

4. Definition of Controls

   a) no history of disease (endpoint) ★

   b) no description of source

Thoughts About the Newcastle-Ottawa Scale

- **Focused questions**; applied to all studies
  - Different sets for different types of studies

- **Categories that make sense**
  - Selection (population)
  - Measurements
  - Comparability (confounding)

How well does (this/any) instrument address each of these categories?

What We Want To Know: Selection (Population)

- Inclusion and exclusion criteria
- Recruitment strategies
- Participant knowledge of study hypotheses
- Participation rates (defined)
- Loss to follow-up (reasons)
- Differences between individuals who did and did not participate, or were or were not lost to follow-up

Am I worried about selection bias; if so, why, and in what way (i.e., direction)?

- Description of the study population

What We Want To Know: Measurements

- Validity (sensitivity/specificity) of outcome measure
- Validity (sensitivity/specificity) of exposure measure
- Blinding of outcome assessment to exposure status (or vice versa)
- Timing of measurement in relation to relevant time window for exposure - effect

Am I worried about information bias (misclassification); if so, why, and in what way (i.e., direction)?

- Levels (and range) of exposures in study setting

What We Want To Know: Confounding

Strong risk factors for the outcome that are also associated with the exposure (but not in pathway)

- What are strong risk factors for the outcome?
- Did (do) these factors vary between groups (cases and controls, exposed and unexposed)?
- How were potential (relevant) differences addressed in the study design or analysis?

Am I worried about confounding; if so, why, and in what way (i.e., direction)?

More Thoughts About Evaluating Epidemiology

- **Documentation (transparency) of relevant information**
- **How do you use the evaluation?**
- **Additional sources of information**

Documentation

- What do you need to know about how the study was designed and conducted?
- What are you worried about?

what was done ——— *worries*

Reference	Participant Selection	Exposure Measure and Range	Outcome Measure	Consideration of Likely Confounding	Data Presentation and Analysis	Comments

How Do You Use the Evaluation of Study Methods?

- "Scoring" or "ranking" [counting the stars] not likely to be useful
- Using evaluation to exclude studies is not likely to be optimal approach
- Stratification (grouping) by methodological features may allow assessment of influence on results

White RH et al. Workshop Report: Evaluation of Epidemiological Data Consistency for Application in Regulatory Risk Assessment. *Open Epidemiology Journal*, 2013; 6:1-8

Additional Sources of Information ("Background Research")

- Exposure measures
  - Validation/reliability studies, probability and levels of exposure in different situations or settings
- Outcome measures
  - Validation/reliability studies, prevalence in different populations, incidence versus mortality, relation between access to health care and survival
- Confounders
  - What is related to the outcome? Is it related to exposure (in a specific type of setting/population)? How strongly?

Springing Forward

http://sports-illustration.com/56-team/ 118-team.html

- **Focused questions**; applied to all studies (but may differ by type, exposure, and outcome)
- **Categories that make sense**
  - Selection
  - Measurements
  - Comparability (confounding)
- **Inclusive**: "rating" system used not to eliminate studies, but rather to understand potential limitations that would affect interpretation of results
- **Documentation of "input" and of "worries"** (separate from "evidence table" (results), but incorporated into evaluation of results)
- **Background research incorporated into review process**

NTP
National Toxicology Program

Evaluating Observational Human Studies in Draft OHAT Systematic Review Framework

Kristina Thayer, Ph.D.
Office of Health Assessment and Translation
National Institute of Environmental Health Sciences

EPA Workshop: Applying Systematic Review to Assessment of Health Effects of Chemical Exposures
August 26, 2013

Outline

- Philosophy
- Steps in process where aspects of “study quality” are considered
- Current risk of bias tool for individual studies (draft)
- Consideration of observational studies within a body of evidence

Philosophy

Separately Consider Different Aspects of Study Quality

- Risk of bias (“internal validity”) – Are findings credible based on design and conduct of study?
- Directness/applicability – Does the study address topic under review?
- Reporting quality – How well was study reported?
- Separating risk of bias from directness/applicability should facilitate use of risk of bias assessments for projects that have different directness & applicability considerations

Use State of Science Approaches to Assess Study Quality

- Single summary scores of studies strongly discouraged
- Endpoint specific
- Update approach and tools as best practices are identified

Goal to Develop a Risk of Bias (RoB) Tool For Use Across Evidence Streams

- Issues for controlled human exposure studies ≈ experimental animal studies
- Can experimental guidance for animal studies be used as a starting point to develop RoB tool for *in vitro* and mechanistic studies?
  - Future phase of work

Steps in Draft OHAT Framework Where "Study Quality" is Considered

When possible consider critical aspects of study design or applicability limitations in eligibility criteria during STEPS 1 & 2

Step 1: Prepare topic

Step 2: Search for and select studies

Step 3: Extract data from studies

Step 4: Assess individual study quality

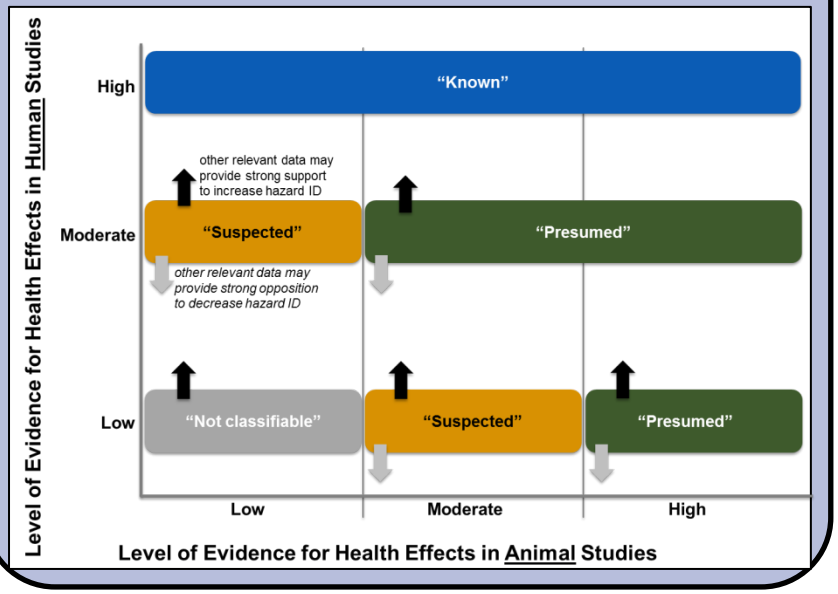
Step 5: Rate confidence in body of evidence

Initial Confidence by Key Features of Study Design	Factors Decreasing Confidence	Factors Increasing Confidence	Confidence in the Body of Evidence
High (++++) 4 Features	❖ Risk of Bias ❖ Unexplained Inconsistency ❖ Indirectness ❖ Imprecision ❖ Publication Bias	❖ Large Magnitude of Effect ❖ Dose Response ❖ All Plausible Confounding • Studies report an effect and residual confounding is toward null • Studies report no effect and residual confounding is away from null ❖ Consistency • Across animal models or species • Across dissimilar populations • Across study design types ❖ Other e.g., particularly rare outcomes	High (++++)
Moderate (+++) 3 Features			Moderate (+++)
Low (++) 2 Features			Low (++)
Very Low (+) ≤1 Features			Very Low (+)

Features
- Controlled exposure
- Exposure prior to outcome
- Individual outcome data
- Comparison group used

Step 6: Translate confidence ratings into level of evidence for health effect

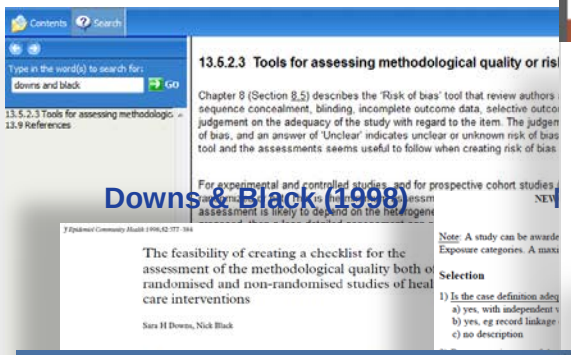
Step 7: Integrate evidence to develop hazard identification conclusions



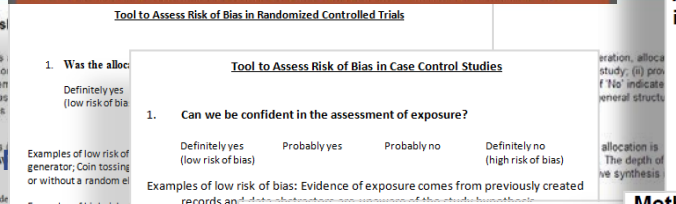
Risk of Bias for Individual Studies

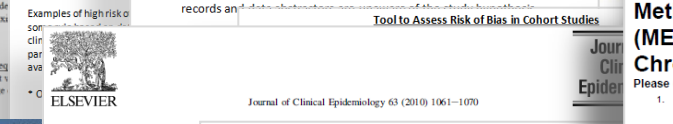
Survey of Methods

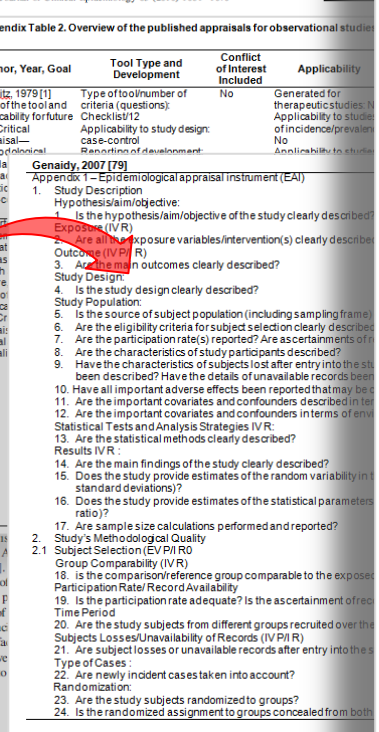
Cochrane Handbook



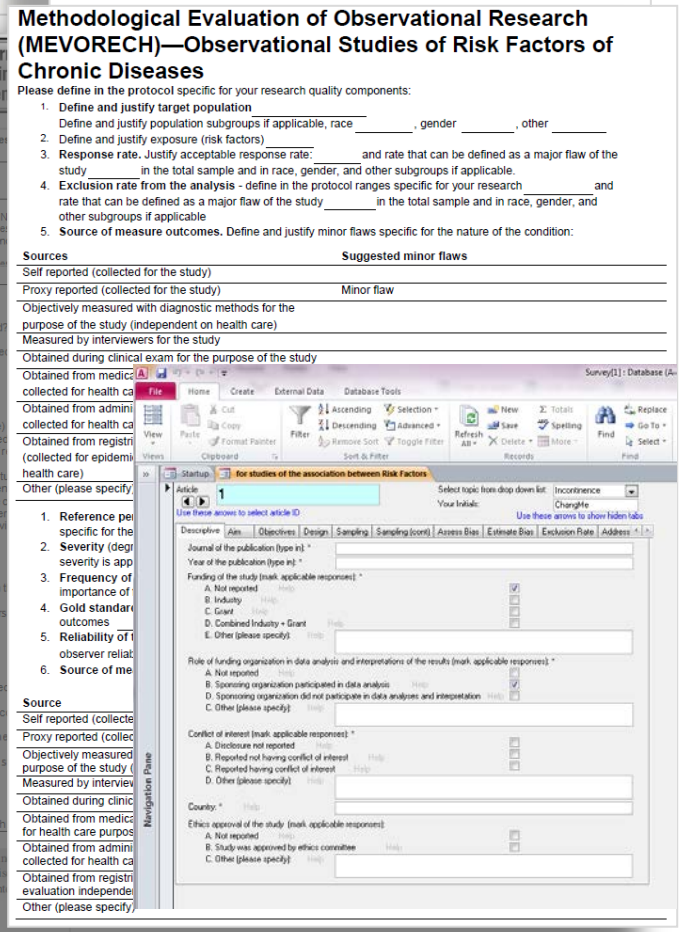
Downs & Black (1998)







March 2012. AHRQ Publication No. 12-EHC047-EF. Available at: www.effectivehealthcare.ahrq.gov/



Assessment of Existing Study Quality Tools

- Often mix internal validity with directness/applicability and reporting quality items
- Range of complexity and detail, e.g., 1 page to 67 items
- Human observational tools often oriented towards cohort or case-control designs
- Format of recent AHRQ guidance useful (March 2012)

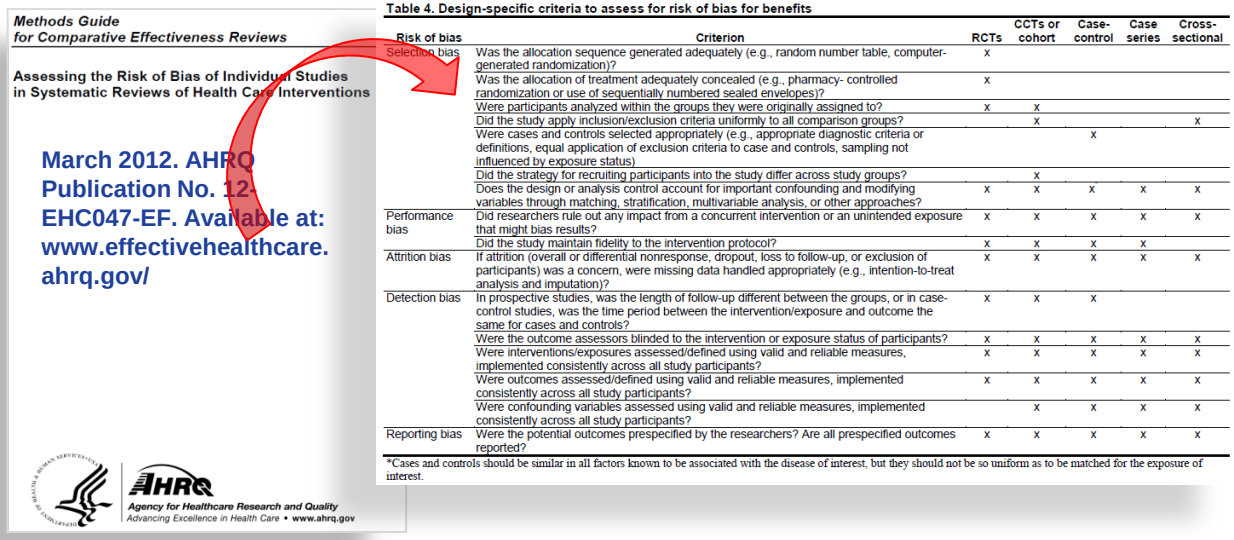


Table 4. Design-specific criteria to assess for risk of bias for benefits

Risk of bias	Criterion	RCTs	CCTs or cohort	Case-control	Case series	Cross-sectional
Selection bias	Was the allocation sequence generated adequately (e.g., random number table, computer-generated randomization)?	x				
	Was the allocation of treatment adequately concealed (e.g., pharmacy- controlled randomization or use of sequentially numbered sealed envelopes)?	x				
	Were participants analyzed within the groups they were originally assigned to?	x	x			
	Did the study apply inclusion/exclusion criteria uniformly to all comparison groups?		x			x
	Were cases and controls selected appropriately (e.g., appropriate diagnostic criteria or definitions, equal application of exclusion criteria to case and controls, sampling not influenced by exposure status)			x		
	Did the strategy for recruiting participants into the study differ across study groups?		x			
	Does the design or analysis control account for important confounding and modifying variables through matching, stratification, multivariable analysis, or other approaches?	x	x	x	x	x
Performance bias	Did researchers rule out any impact from a concurrent intervention or an unintended exposure that might bias results?	x	x	x	x	x
	Did the study maintain fidelity to the intervention protocol?	x	x	x	x	
Attrition bias	If attrition (overall or differential nonresponse, dropout, loss to follow-up, or exclusion of participants) was a concern, were missing data handled appropriately (e.g., intention-to-treat analysis and imputation)?	x	x	x	x	x
Detection bias	In prospective studies, was the length of follow-up different between the groups, or in case-control studies, was the time period between the intervention/exposure and outcome the same for cases and controls?	x	x	x		
	Were the outcome assessors blinded to the intervention or exposure status of participants?	x	x	x	x	x
	Were interventions/exposures assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were outcomes assessed/defined using valid and reliable measures, implemented consistently across all study participants?	x	x	x	x	x
	Were confounding variables assessed using valid and reliable measures, implemented consistently across all study participants?		x	x	x	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x	x	x

*Cases and controls should be similar in all factors known to be associated with the disease of interest, but they should not be so uniform as to be matched for the exposure of interest.

Consideration of New Castle Ottawa

- Major advantage: short
- Disadvantages*
  - Use of star system to rate studies
  - Blending of risk of bias with applicability
    - Representativeness of cohort with respect to community – Results may be unbiased assessment within cohort, but not applicable to more representative sample
    - Duration of follow-up may be less than optimal to address question of interest, but the results of study may be accurate

*Guyatt G, Busse JW. Methods Commentary: Risk of Bias in Cohort Studies. http://distillercer.com/resources/methodological-resources/ [accessed 19 August 2013]

Current Risk of Bias Tool for Individual Studies (Draft)

Uses AHRQ approach for same set of questions applied to different study designs

Bias Domain	Criterion	Animal	Controlled	Cohort	Case-Control	Cross-sectional	Case Series
Selection	Was administered dose or exposure level adequately randomized?	X	X				
	Was allocation to study groups adequately concealed?	X	X				
	Were the comparison groups appropriate?			X	X	X	
Confounding	Did the study design or analysis account for important confounding and modifying variables?	X	X	X	X	X	X
	Did researchers adjust or control for other exposures that are anticipated to bias results?	X	X	X	X	X	X
Performance	Were experimental conditions identical across study groups?	X	X				
	Did deviations from the study protocol impact the results?	X	X	X	X	X	X
	Were the research personnel and human subjects blinded to the study group during the study?	X	X				
Attrition	Were outcome data incomplete due to attrition or exclusion from analysis?	X	X	X	X	X	
Detection	Were the outcome assessors blinde						X
	Were confounding variables assesse measures						X
	Can we be confident in the exposure characterization?	X	X	X	X	X	X
	Can we be confident in the outcome assessment?	X	X	X	X	X	X
Reporting	Were all measured outcomes reported?	X	X	X	X	X	X
Other	Were there any other potential threats to internal validity (e.g., inappropriate statistical methods)?	X	X	X	X	X	X

Some items seem unlikely to be useful in short-term but may be useful in long-term, i.e., changes in reporting quality, develop empirical data to assess potential risk of bias of item

Current Tool: Response Format & Review Process

- Uses responses recommended by the Clarity Group
  - “definitely no” (●) risk of bias
  - “probably no” (●) risk of bias
  - “probably yes” (●) risk of bias
  - “definitely yes” (●) risk of bias
- Rationale for selecting a response is noted
  - Based on instructions and expert judgment (e.g., members of review team, technical advisors)
- Risk of bias is independently assessed by 2 members of review team
  - Independent reviews discussed to develop draft response for report
- Risk of bias conclusions assessed by review team, technical advisors, and undergo external public peer-review

Current Tool: Impact of Non-Reporting

- Reporting quality not separately assessed but will impact risk of bias assessment for individual studies
  - Studies penalized for non-reporting: Assigned “probably yes” (●)
  - Will attempt to contact author to gather unreported information
- Willing to consider collecting reporting quality data
  - e.g., STROBE (human observation); ToxRTool (animal, *in vitro*)
  - Many reporting quality elements already embedded in our risk of bias instructions and data extraction
  - Need to determine how information would be used, e.g., should studies that have a significant degree of under-reporting be excluded?

Presenting Risk of Bias for a Single Study (Example Appendix Summary)

Risk of bias response options for individual items:			
Bias Domain	**Criterion**	**Response & Rationale**	
Selection	Was administered dose or exposure level adequately randomized?	n/a	not applicable
	Was allocation to study groups adequately concealed?	n/a	not applicable
	Were the comparison groups appropriate?	++	yes, based on quartiles of exposure
Confounding	Does the study design or analysis account for important confounding and modifying variables?	+	yes (sex, age, race, urinary creatinine, education, smoking), but no adjustment for nutritional quality, e.g., soda consumption
	Did researchers adjust or control for other exposures that are anticipated to bias results?	+	no, but not considered to present risk of bias in general population studies
Performance	Were experimental conditions identical across study groups?	n/a	not applicable
	Did deviations from the study protocol impact the results?	+	no deviations reported
	Were the research personnel and human subjects blinded to the study group during the study?	n/a	not applicable
Attrition	Were outcome data incomplete due to attrition or exclusion from analysis?	+	not considered a risk of bias, excluded observations (≤ 87 for any analysis) based on missing BMI or covariate data
Detection	Were the outcome assessors blinded to study group or exposure level?	++	yes, BPA levels not known at time of outcome assessment
	Were confounding variables assessed consistently across groups using valid and reliable measures?	++	yes, used standard NHANES methods
	Can we be confident in the exposure characterization?	++	yes, NHANES methods are considered “gold standard” for urinary BPA
	Can we be confident in the outcome assessment?	++	yes, used standard diagnostic criteria
Selective Reporting	Were all measured outcomes reported?	++	yes, primary outcomes discussed in methods were presented results section with adequate level of detail for data extraction
Other	Were there any other potential threats to internal validity (e.g., inappropriate statistical methods)?	++	none identified

Risk of Bias Ratings Across Individual Studies

++ Definitely Low risk of bias
\+ Probably Low risk of bias
\- Probably High risk of bias
-- Definitely High risk of bias
○ Not applicable due to study design

Draft OHAT Risk of Bias Questions	NotGood ., 2010	Bucher et al., 1999	Wolfe et al., 2000	Boyles et al., 2011	Thayer et al., 2008
Selection Bias					
Was administered dose or exposure level adequately randomized?	○	○	○	○	○
Was allocation to study groups adequately concealed?	○	○	○	○	○
Were the comparison groups appropriate?	+	++	+	--	+
Confounding Bias					
Did the study design or analysis account for important confounding and modifying variables?	-	++	-	--	+
Did researchers adjust or control for other exposures that are anticipated to bias results?	-	++	-	-	--
Performance Bias					
Were experimental conditions identical across study groups?	○	○	○	○	○
Did deviations from the study protocol impact the results?	-	++	+	-	-
Were the research personnel and human subjects blinded to the study group during the study?	○	○	○	○	○
Attrition / Exclusion Bias					
Were outcome data incomplete due to attrition or exclusion from analysis?	--	++	-	+	+
Information / Detection Bias					
Were outcome assessors blinded to study group or exposure group?	+	++	+	+	+
Were confounding variables assessed consistently across groups using valid and reliable measures?	--	++	+	++	++
Can we be confident in the exposure characterization?	-	++	--	-	+
Can we be confident in the outcome assessment?	-	++	-	+	-
Selective Reporting Bias					
Were all measured outcomes reported?	+	++	+	-	+

Visualizing Risk of Bias Strengths and Weaknesses Across a Collection of Studies

Table 11. Visual summary of risk of bias ratings for each outcome (hypothetical summary for a set of 10 observational human studies)										
Questions	**20%**		**40%**		**60%**		**80%**		**100%**	
Was administered dose or exposure level adequately randomized?	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Was allocation to study groups adequately concealed?	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Were the comparison groups appropriate?	++	++	++	++	+	+	-	-	--	--
Does the study design or analysis account for important confounding and modifying variables?	++	++	+	+	+	-	-	-	-	--
Did researchers adjust or control for other exposures that are anticipated to bias results?	++	+	+	+	+	-	-	-	-	--
Were experimental conditions identical across study groups?	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Did deviations from the study protocol impact the results?	++	++	+	+	+	+	+	+	+	-
Were the research personnel and human subjects blinded to the study group during the study?	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Were outcome data incomplete due to attrition or exclusion from analysis?	++	+	+	+	+	+	+	-	-	-
Were the outcome assessors blinded to study group or exposure level?	++	++	++	++	++	+	+	+	-	-
Were confounding variables assessed consistently across groups using valid and reliable measures?	++	++	++	++	+	+	+	-	-	-
Can we be confident in the exposure characterization?	+	+	+	-	-	-	-	-	--	--
Can we be confident in the outcome assessment?	++	++	++	++	+	+	+	+	-	-
Were all measured outcomes reported?	++	+	+	+	+	+	+	+	+	+

++	**definite low risk of bias**
+	**probably low risk of bias**
-	**probably high risk of bias**
--	**definitely high risk of bias**
n/a	**not applicable**

Using Risk of Bias to Potentially Exclude Studies

- Tier studies based on risk of bias

Guidance for developing risk of bias categories for individual studies												
		Risk of Bias Criteria & Responses										
Catagory	Guidance	key criteria #1	key criteria #2	other criteria	other criteria	other criteria	other criteria	other criteria	other criteria	other criteria	other criteria
1st tier	"definitely low" or "probably low" risk of bias for key criteria **AND** "definitely low" or "probably low" risk of bias for ≥50% of other criteria	●	●	●	●	●	●	●	●	●	●
2nd tier	study does not meet criteria for "low" or "high" example 1	●	●	●	●	●	●	●	●	●	●
	example 2	●	●	●	●	●	●	●	●	●	●
	example 3	●	●	●	●	●	●	●	●	●	●
3rd tier	"definitely high" or "probably high" risk of bias for key criteria **AND** "definitely high" or "probably high" risk of bias for ≥50% of other criteria			●	●	●	●	●	●	●	●

- Base conclusions on studies in 1st or 2nd tier only?
  - Conduct "sensitivity" analysis with high risk of bias studies included to assess impact

Consideration of Observational Studies Within a Body of Evidence

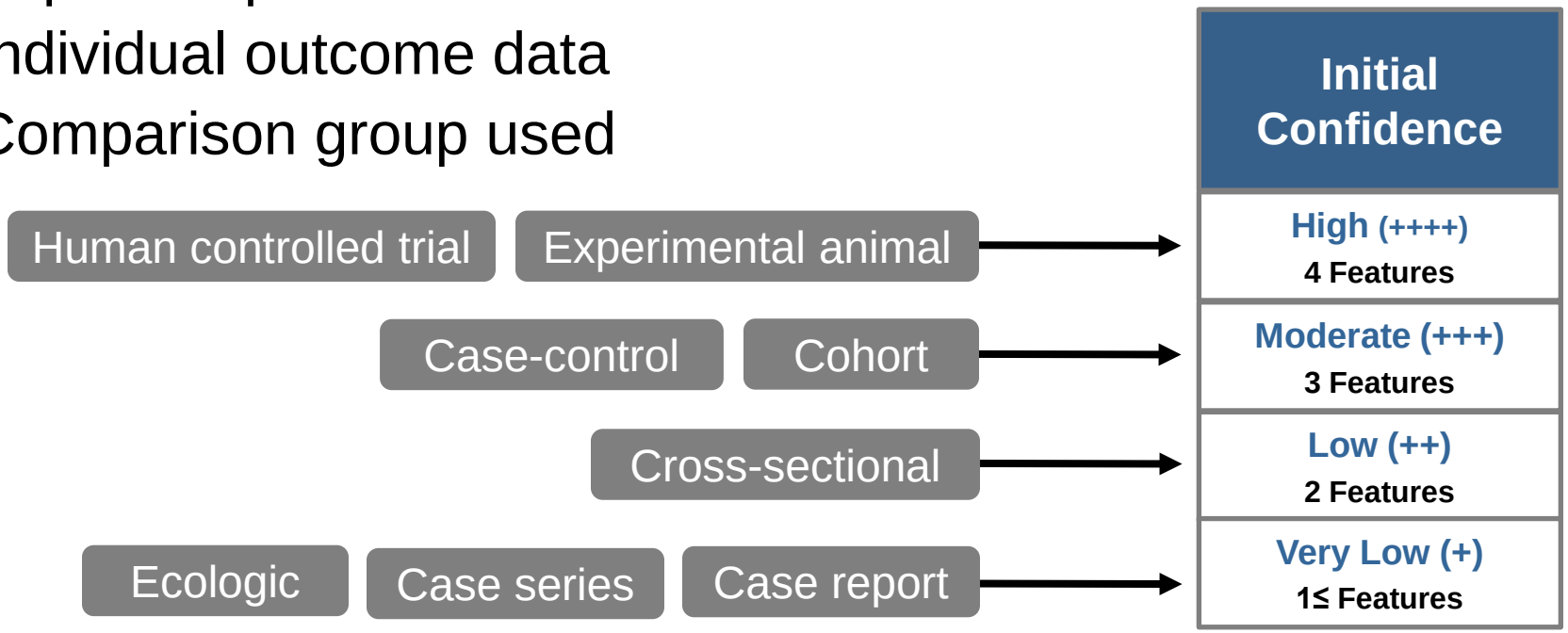
Framework to Assess Confidence in a Body of Evidence

Initial Confidence by Key Features of Study Design	Factors Decreasing Confidence	Factors Increasing Confidence	Confidence in the Body of Evidence
High (++++) **4 Features**	❖ **Risk of Bias** ❖ **Unexplained Inconsistency** ❖ **Indirectness** ❖ **Imprecision** ❖ **Publication Bias**	❖ **Large Magnitude of Effect** ❖ **Dose Response** ❖ **All Plausible Confounding** • Studies report an effect and residual confounding is toward null • Studies report no effect and residual confounding is away from null ❖ **Consistency** • Across animal models or species • Across dissimilar populations • Across study design types ❖ **Other** e.g., particularly rare outcomes	High (++++)
Moderate (+++) **3 Features**			Moderate (+++)
Low (++) **2 Features**			Low (++)
Very Low (+) **≤1 Features**			Very Low (+)

Features
- Controlled exposure
- Exposure prior to outcome
- Individual outcome data
- Comparison group used

Initial Confidence Based on Key Study Design Features

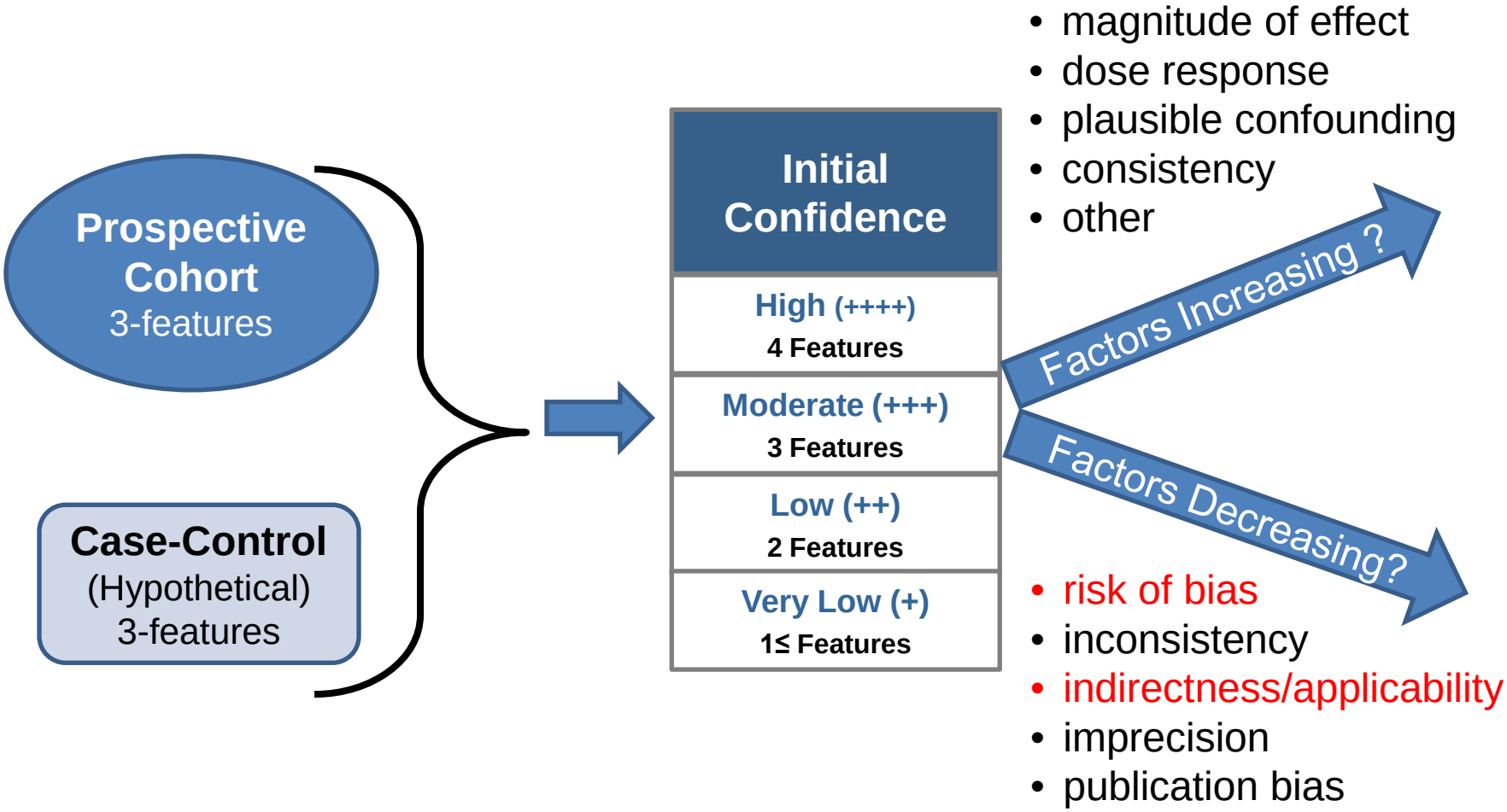
- Controlled exposure
- Exposure prior to outcome
- Individual outcome data
- Comparison group used



- Differs from GRADE (all observational studies start as low) and Navigation Guide (all observational studies start as moderate)

Initial Confidence by Study Design Features

- Starting point for evaluating confidence in a collection of studies in same initial confidence category and evaluate as a group for the same outcome (or set of related outcomes)



Next Steps: Assess OHAT Approach in Case Studies

- Evaluate overall approach in 2 case studies: BPA & obesity; PFOS/PFOA & immunotoxicity
  - Clarity and transparency of current approach
  - Consider providing reporting quality report
  - Evaluate consistency of assessment among reviewers
  - Consider issues identified in public and interagency comments
- Complete case studies during next calendar year
- Two public webinars
  - Clarification of issues raised in public comments & update: Sept 26, 2013, 1-4 pm (http://ntp.niehs.nih.gov/go/40490)
  - Lessons learned from case studies (2014)

Acknowledgements

- **Office of Health Assessment and Translation**
  - Abee Boyles
  - Kembra Howdeshell
  - Andrew Rooney, Deputy Director
  - Michael Shelby
  - Kyla Taylor
  - Kristina Thayer, Director
  - Vickie Walker
- **Office of Liaison, Policy and Review**
  - Mary Wolfe, Director
  - Lori White
- **Office of Library and Information Services**
  - Stephanie Holmgren
- **Approach Technical Advisors and Experts**
  - **Lisa Bero,** Director, San Francisco Branch, United States Cochrane Center at UC San Francisco
  - **Gordon Guyatt,** Co-chair, GRADE Working Group, McMaster U
  - **Malcolm Macleod,** CAMARADES Centre, University of Edinburgh
  - **Karen Robinson,** Co-Director, Evidence-Based Practice Center, The Johns Hopkins Bloomberg School of Public Health
  - **Holger Schünemann**, Co-chair, GRADE Working Group, McMaster U.
  - **Tracey Woodruff,** Director, Program on Reproductive Health and the Environment, UCSF
- **NTP Board of Scientific Counselors**
- **NTP BSC Working Group**
  - **Lynn Goldman, Chair,** Dean, School of Public Health and Health Services, George Washington U.
  - **Reeder Sams, Vice-chair,** Acting Deputy Director, NCEA/RTP Division, USEPA
  - **Lisa Bero,** Director, San Francisco Branch, United States Cochrane Center at UC San Francisco
  - **Edward Carney,** Senior Science Leader, Mammalian Toxicology, Dow Chemical Company
  - **David Dorman,** Professor, North Carolina State University
  - **Elaine Faustman**, Director, Institute for Risk Analysis and Risk Communication, U. Washington
  - **Dale Hattis**, Research Professor, George Perkins Marsh Institute, Clark University
  - **Malcolm Macleod,** CAMARADES Centre, University of Edinburgh
  - **Tracey Woodruff,** Director, Program on Reproductive Health and the Environment, UCSF
  - **Lauren Zeise,** Chief, Reproductive and Cancer Hazard Assessment Branch, OEHHA, California EPA
- **Protocol Technical Advisors**

EVALUATING OBSERVATIONAL EPIDEMIOLOGY STUDIES

Panel Discussion

Evaluating Observational Epidemiology Studies

1. What gives you confidence in a study or set of studies? [i.e., what do you look for in a study that makes you comfortable in interpreting the observed risk estimate to be an accurate estimate; what makes you worried that the observed risk estimate is an over estimate or spurious finding; what makes you worried that the observed risk estimate is an underestimate of the actual risk; what criteria would you use to "downgrade" a study (because you're worried it's overestimating, underestimating, or because you don't know how to interpret the results…?]

Evaluating Observational Epidemiology Studies

2. What type of or level of detail (with respect to decisions by the evaluators, and with respect to descriptions of individual studies) would you want to see in an evaluation of study methods/limitations/biases?

Evaluating Observational Epidemiology Studies

3. What thoughts or advice can you offer on addressing the tension between balancing transparency and reproducibility in evaluation of study methods/limitations/biases with the need for flexibility and professional expertise or judgment?

Evaluating Observational Epidemiology Studies

4. Quantitative methods to estimate the extent of specific sources of bias in epidemiology (e.g., misclassification of exposure, selection bias) and the impact on risk estimates have been developed, but are not widely used. What role should quantitative bias assessment play in the systematic review of individual studies and of groups of studies? What minimum data are necessary in order to attempt quantitative bias assessment?

Instruments for Assessing Risk of Bias and Other Methodological Criteria of Published Animal Studies: A Systematic Review

August 26, 2013

David Krauth[1], Tracey Woodruff[2], Lisa Bero[1, 3, 4]

[1] University of California, San Francisco, Department of Clinical Pharmacy, San Francisco, CA

[2] University of California, San Francisco, Department of Obstetrics, Gynecology, and Reproductive Sciences, San Francisco, CA; Program on Reproductive Health and the Environment (PRHE), Oakland, CA

[3] Institute of Health Policy Studies , UCSF School of Medicine

[4] San Francisco Branch of United States Cochrane Center

Funding Source: National Institute of Environmental Health Sciences (Grant # R 21ES 021028)

ehp ENVIRONMENTAL HEALTH PERSPECTIVES
http://www.ehponline.org

Instruments for Assessing Risk of Bias and Other Methodological Criteria of Published Animal Studies: A Systematic Review

David Krauth, Tracey J. Woodruff and Lisa Bero

http://dx.doi.org/10.1289/ehp.1206389

Online 14 June 2013

NIH National Institute of Environmental Health Sciences

Disclosure Statement

- All authors declare that they have no conflicts of interest to disclose.

Risks of Bias **IS NOT** Reporting or Quality

- **Risks of bias**

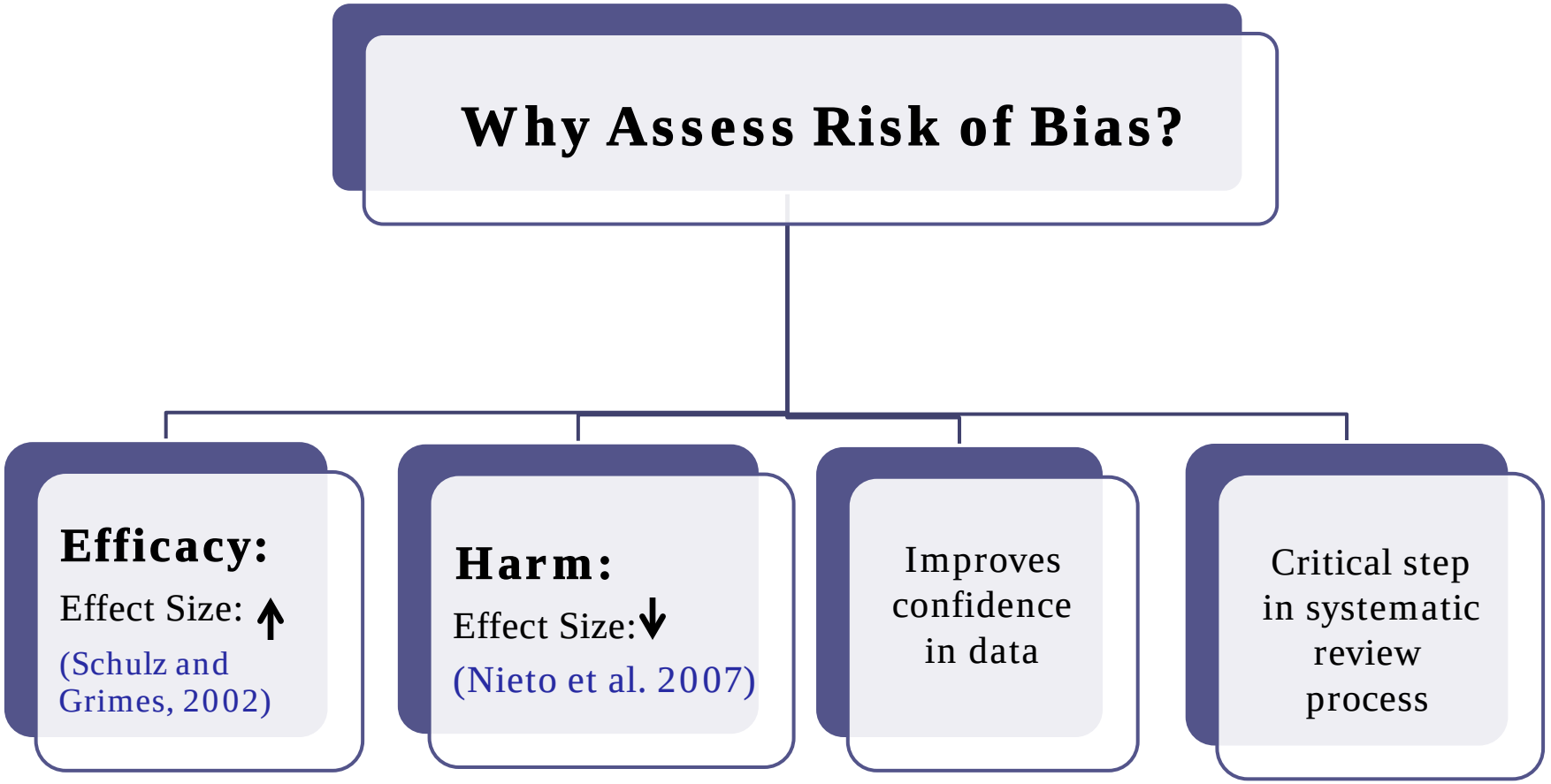
  Methodological criteria that can introduce a systematic error in the magnitude or direction of the results (Higgins and Green 2008)

- **Quality**

  Study criteria related to how a study is conducted (e.g., in compliance with human subjects guidelines)

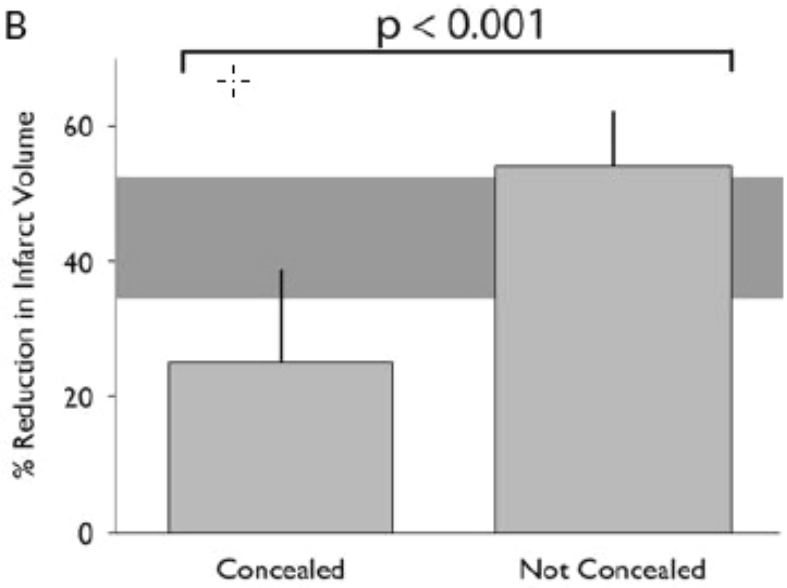
- **Reporting**

  Completeness of information (e.g. study population described)


Why Assess Risk of Bias?
Efficacy:
Effect Size: ↑
(Schulz and Grimes, 2002)
Harm:
Effect Size:↓
(Nieto et al. 2007)
Improves confidence in data
Critical step in systematic review process

Example of High Risk of Bias

Reported drug efficacy was significantly lower in studies that reported measures taken to conceal treatment allocation from the time of cerebral ischemia up to the time of outcome assessment (25.1% versus 54.0%; *P < 0.001)*



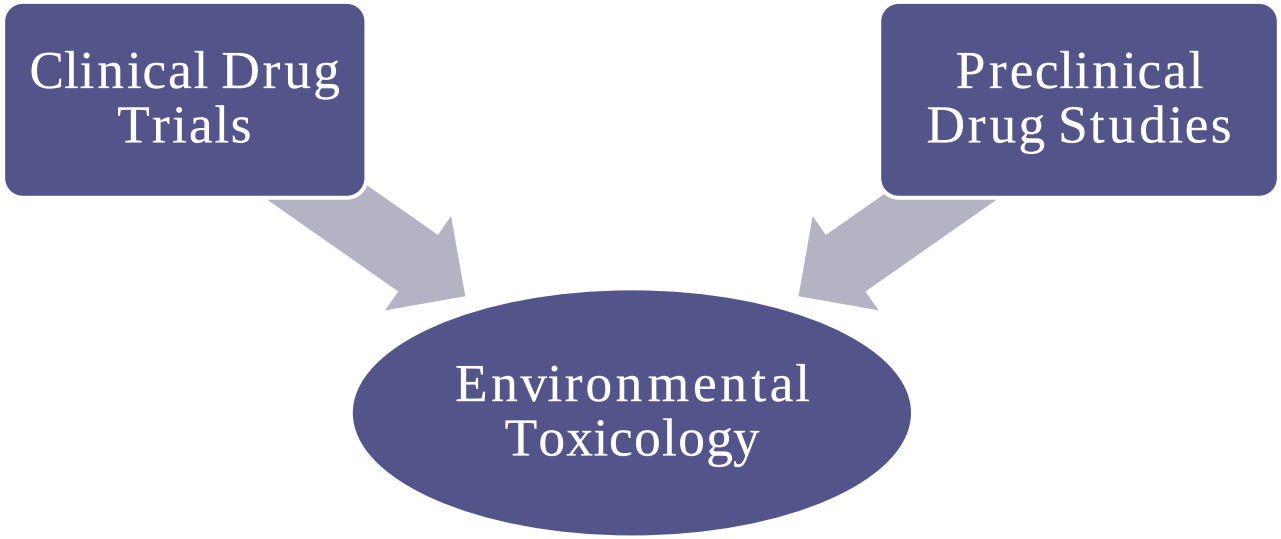
Macleod et *al.* 2008

Systematic Review Protocol

1. State objective
2. Selection criteria
3. Search strategy
4. Apply selection criteria
   - In duplicate, reproducible, transparent
5. Assess risk of bias of included studies
6. Analyze results, using meta-analysis if appropriate

Study Objective

Identify and summarize existing instruments for animal studies



Methods

Search Strategy*	• Medline (January 1966 - November 2011) • Reference lists
Inclusion Criteria	• Instruments for assessing risk of bias in animal studies • English
Exclusion Criteria	• Review articles • Application of an instrument

Krauth et al, 2013

*http://ehp.niehs.nih.gov/wp-content/uploads/121/6/ehp.1206389.pdf

Methods

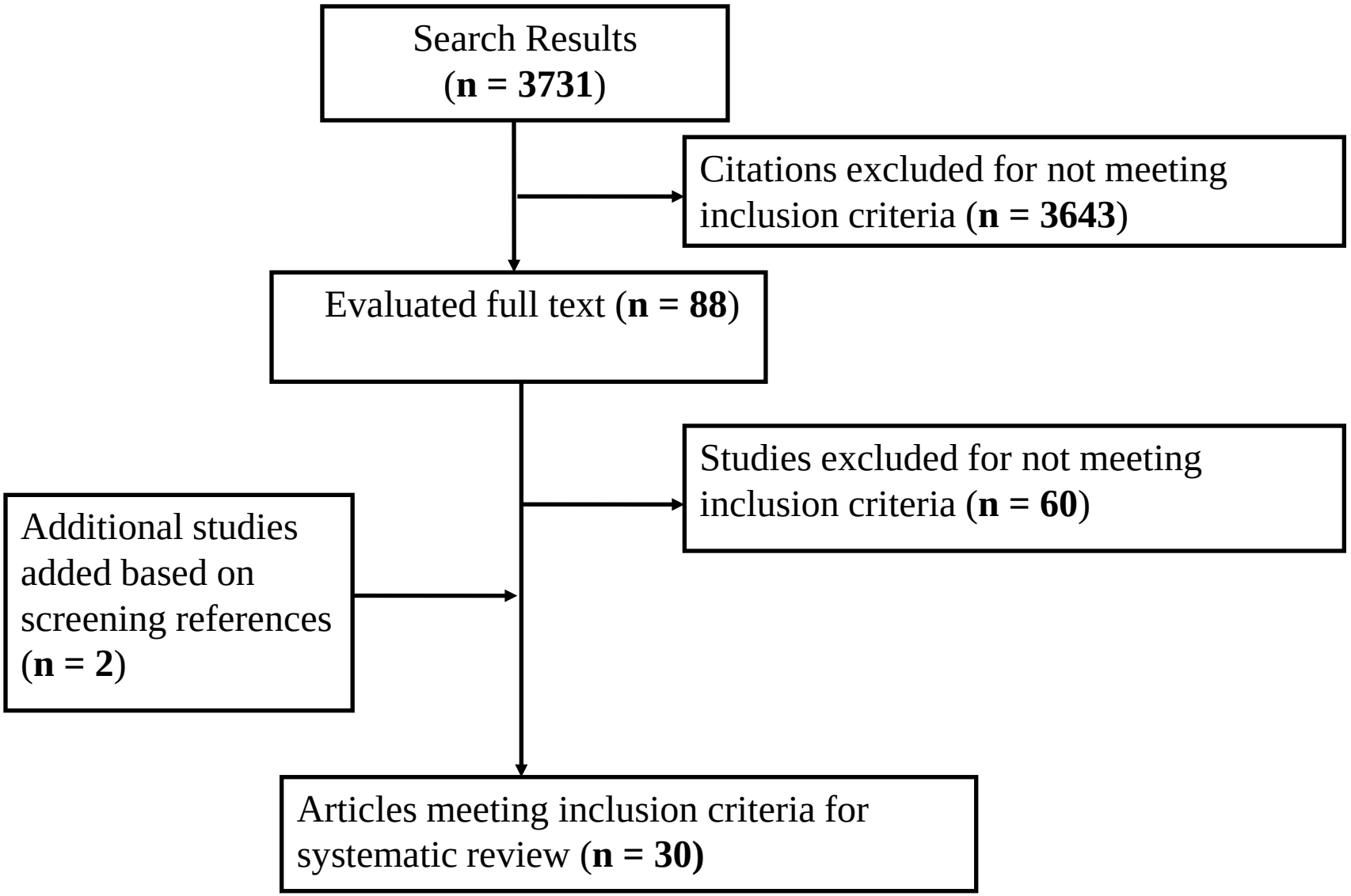
Data Extraction – Instrument Characteristics

- Animal model
- Number of criteria
- Date of publication
- Tested for reliability
- Tested for validity

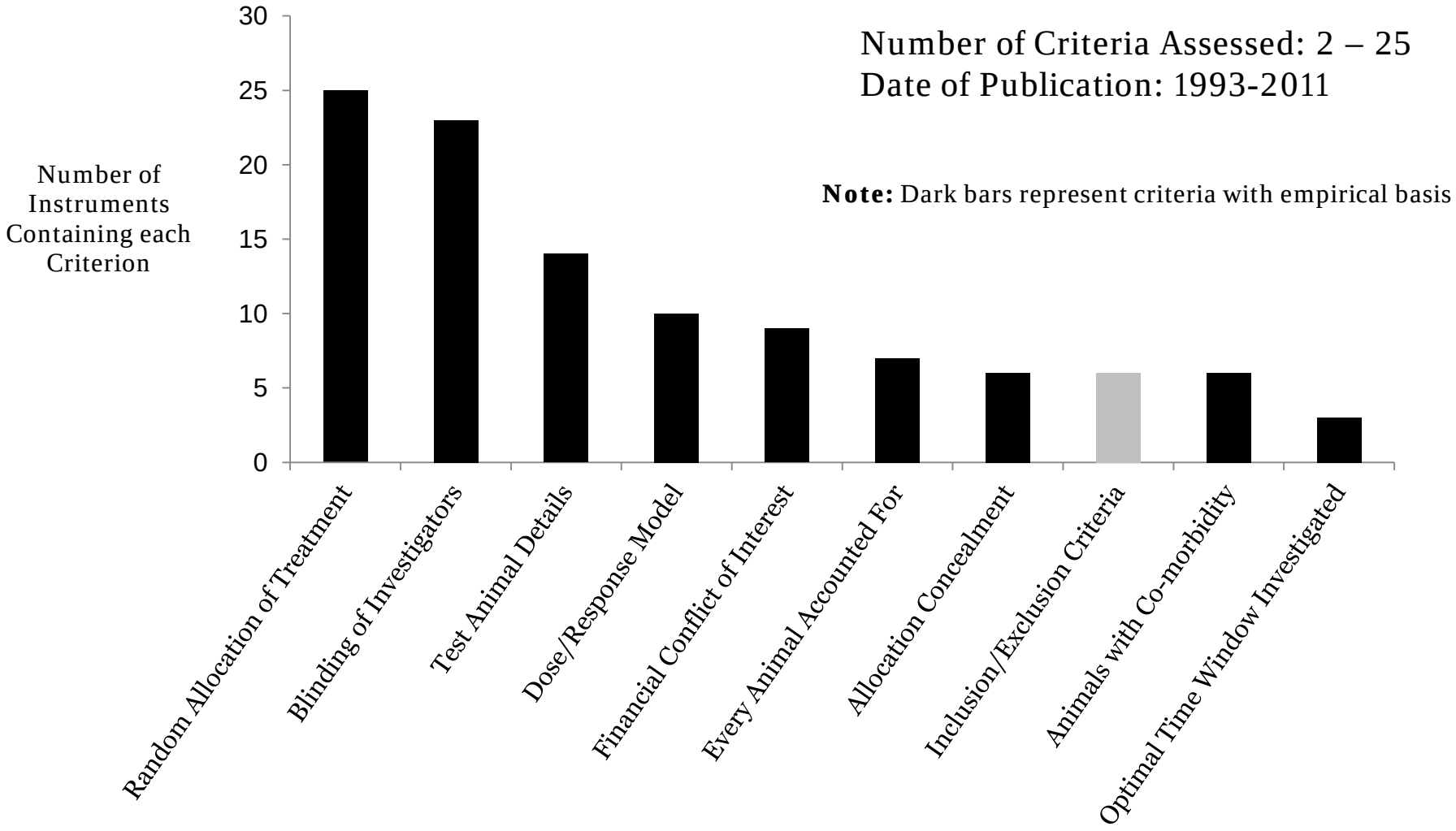
Methods

We extracted risk of bias criteria, reporting criteria, and other methodological characteristics

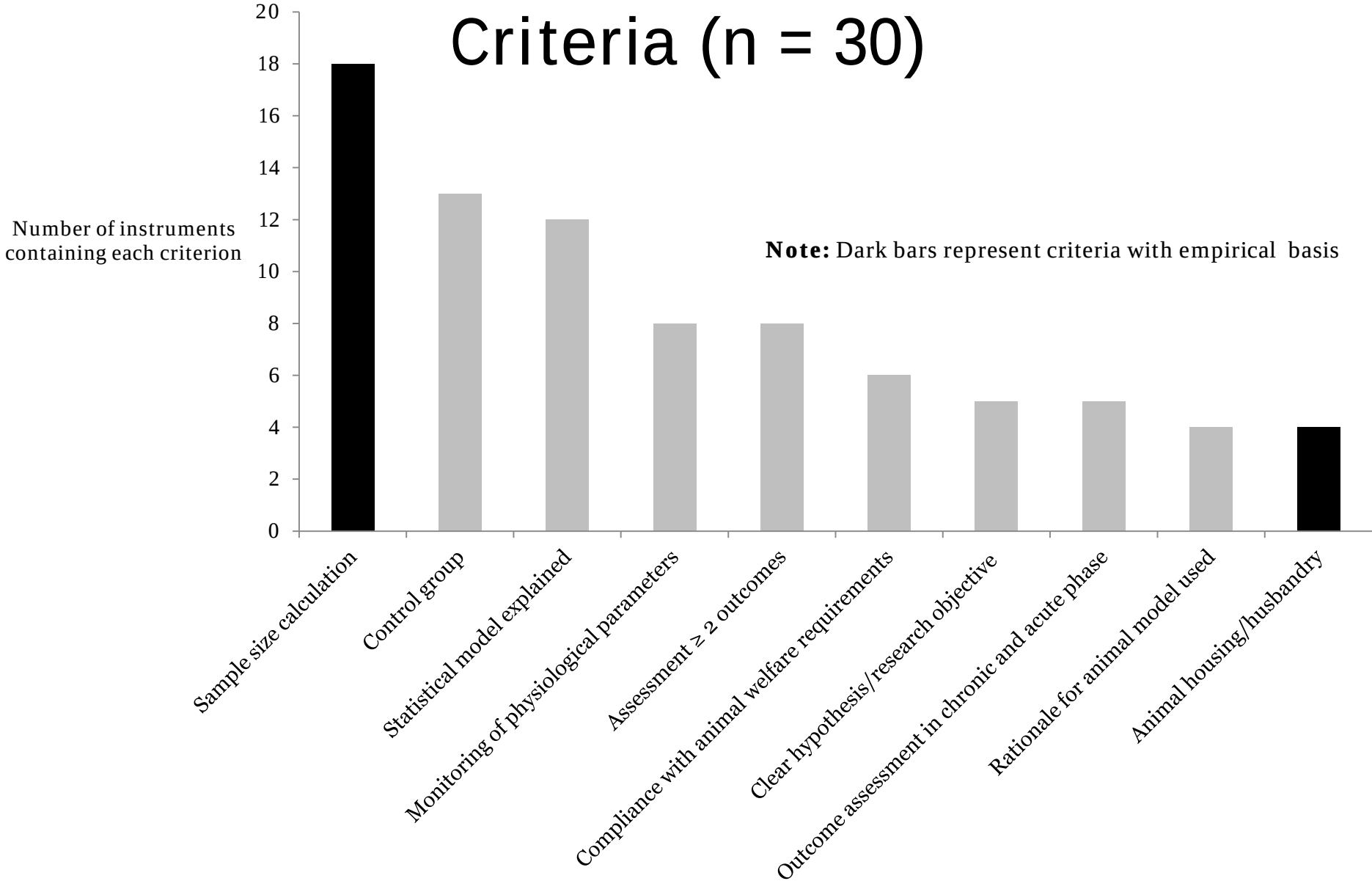
Flow Chart for Study Inclusion



Results: Risk of Bias (n=30)



Results: Reporting and Other Methodological Criteria (n = 30)



Limitations of the Instruments (n = 30)

- Few instruments developed for animal toxicology (4)
- Most instruments not tested for validity and reliability
- Most instruments mix reporting, risk of bias, and other methodological criteria

Limitations of our Study

- Searched Medline database and articles published in English

Recommendation

*Use of **empirically** based criteria for assessing risk of bias in animal toxicology studies*

Acknowledgments

- Our funding source, the National Institute of Environmental Health Sciences

- Gloria Won (UCSF Mount Zion Campus) for assistance with developing the search strategy

- Dr. Dorie Apollonio (UCSF), Dr. David Dorman (North Carolina State University), and Rose Philipps (UCSF) for reviewing the manuscript

Questions?

Supplemental Slides NEXT

SUPPLEMENTAL SLIDE #1

A Priori List of Study Design Elements Aimed at Reducing Bias and other Methodological Characteristics

1. Treatment allocation/Randomization
2. Concealment of Allocation
3. Blinding of Investigators
4. Inclusion/Exclusion Criteria
5. Sample Size Calculation
6. Compliance with Animal Welfare Requirements
7. Financial Conflicts of Interest
8. Statistical Model Explained
9. Use of Animals with Comorbidity
10. Test Animal Descriptions
11. Dose/Response (D/R) Model
12. All Animals Accounted for
13. Optimal Time Window Investigated

SUPPLEMENTAL SLIDE #2
Randomization

- 25 of 30 instruments include random allocation of treatment

- A systematic review of multiple sclerosis interventions in animal research has shown that non-randomized studies report significantly higher treatment efficacy (41.6%, 95% CI 36.7-46.5%) than randomized studies (20.6%, 95% CI 11.4-29.7%)

  (Vesterinen et *al.* 2010)

- In emergency medicine, animal studies lacking randomization were over three times more likely to show a statistically significant result relative to studies that included these attributes

  (Bebarta et *al.* 2003)

SUPPLEMENTAL SLIDE #3
Blinding of Investigators

- 23 0f 30 instruments include blinding

- Blinding in experimental stroke studies significantly alters the effectiveness of an intervention with effect sizes ranging by 10% in studies with or without this feature (Crossley et *al.* 2008)

- A systematic review of multiple sclerosis interventions has shown that studies performed without blinded assessment of outcome report higher efficacy estimates (41.0%, 95% CI 36.2–45.8%) compared to blinded studies (29.8%, 95% CI 19.8–39.8%) (Vesterinen et *al.* 2010)

SUPPLEMENTAL SLIDE #4
Financial Conflict of Interest

- 9 of 25 instruments include disclosure of conflicts of interest

- Reviews of clinical studies have shown that study funding sources and financial ties of investigators (including university or industry affiliated investigators) are associated with favorable research outcomes for the sponsors [efficacy results risk ratio (RR): 1.32; harm results RR: 1.87] even when controlling for other risks of bias.

(Lundh et *al.* 2012)

SUPPLEMENTAL SLIDE #5
Animals with Co-morbidity

- 6 of 30 instruments state the need to use animals with pre-existing co-morbidity.

- Using co-morbid animals in experimental stroke studies was found to significantly alter the effectiveness of an intervention with effect sizes ranging by 10% in studies with or without these features

(Crossley et *al.* 2008)

SUPPLEMENTAL SLIDE #6

Test Animal Details

- 14 of 30 instruments state the need to include detailed reporting of test animal characteristics

- In a meta-analysis containing 14 animal studies, it was determined that the efficacy of using nicotinamide to treat stroke outcomes depends on animal species and sex. Drug efficacy was effective in rats but not mice ($p < 0.0001$) and male species performed better than females ($p = 0.012$).

(Macleod et *al.* 2004)

SUPPLEMENTAL SLIDE #7
Was every animal accounted for?

- 7 of 30 instruments include assessing whether all animals were accounted for

- In a study comparing clinical data from 14 meta-analyses that addressed therapeutic treatments for cancer, it was shown that not accounting for all patients leads to more favorable research outcomes (p-value = 0.03) relative to studies that do account for all patients.

(Tierney and Stewart 2005)

SUPPLEMENTAL SLIDE #8: Criteria with Empirical Evidence

Type of Bias	Risk of Bias Criteria
Selection Systematic differences between baseline characteristics in treatment and control groups	*Empirically tested in animal models* **Randomization** (Macleod et al 2008, Bebarta et al. 2003, Sena et al. 2007, Vesterinen et *al*. 2010) **Concealment of allocation** (Macleod et *al*. 2008)
Performance Systematic difference between treatment and control groups with regard to care or other exposure besides the intervention (Higgins and Green, 2008).	*Empirically tested in animal models* **Blinding** (Bebarta et *al*. 2003, Sena et *al*. 2007, Vesterinen et *al*. 2010) **Use of animals with identical co-morbid conditions** (Crossley et *al*. 2008; Macleod et *al*. 2004; Macleod et *al*. 2008; Sena et *al*. 2007) **Identical housing/husbandry conditions between treatment groups** (Duke et *al*. 2001; Gerdin et *al*. 2012)
Detection Systematic differences between treatment and control groups with regards to how outcomes are assessed	*Empirically tested in animal models* **Blinding** (Bebarta et *al*. 2003; Vesterinen et *al*. 2010) **Optimal time window investigated for outcome assessment** (EPA 2009)
Exclusion Systematic difference between treatment and control groups in the number of animals that were included in and completed the study.	*Empirically tested in clinical trials* **Data on whether all animals are accounted for** (Tierney and Stewart 2005) **Intention-to-treat analysis performed** (Melander et *al*. 2003; Porta et *al*. 2007)
Other Bias	*Empirically tested in animal models* **Sample size calculation** (Vesterinen et *al*. 2010) **Test animal details** (Macleod et al. 2004; Sniekers et *al*. 2008) **Appropriateness of dose selection (validated by use of a dose/response model)** (Bucher et *al*. 1996) **Timing of exposure** (Benatar 2007; van der Worp et *al*. 2010; Vesterinen et *al*. 2010) **Measurement of outcomes that are sensitive to the exposure** (Wood 2000) *Empirically tested in clinical trials* **Type of funding source** (Lundh et *al*. 2012) **Financial conflicts of interest stated** (Lundh et *al*. 2012) **Selective outcome reporting** (Hart et *al*. 2012; Rising et *al*. 2008)

SUPPLEMENTAL SLIDE #9

Summary of Commonly Used Instruments

CHECKLIST	INSTRUMENT DESCRIPTION
Agerstrand et al 2011	• 25 item instrument • Not empirically tested • No methodological score is used • Intended use of instrument is environmental toxicology research
Kilkenny et al, 2010 The ARRIVE Guidelines	• 13 item instrument • Not empirically tested • No methodological score is used is used • No specific disease modeled • Developed using the CONSORT criteria as a foundation, and consensus and consultation from scientists, statisticians, journal editors, and research funders
Sena et al, 2007	• 21 item instrument • No methodological score is used • Provide empirical data for randomization and blinding • Disease modeled is stroke • Instrument derived from 4 previous checklists: STAIR, Amsterdam Criteria (Horn et *al*. 2001), CAMARADES, Utrecht Criteria (van der Worp et *al*. 2005) • Instrument appears to have validity

SUPPLEMENTAL SLIDE #10: References Cited

- Bebarta V, Luyten D, Heard K. 2003. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med 10(6):684-687.
- Benatar, M. 2007. Lost in translation: Treatment trials in the SOD1 mouse and in human ALS. Neurobiology of Disease 26(1):1-13.
- Bucher JR, Portier CJ, Goodman JI, Faustman EM, Lucier GW. 1996. Workshop overview. National Toxicology Program Studies: principles of dose selection and applications to mechanistic based risk assessment. Fundam Appl Toxicol 31(1): 1-8.
- Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM, et al. 2008. Empirical evidence of bias in the design of experimental stroke studies: a metaepidemiologic approach. Stroke 39(3):929-934.
- Duke JL, Zammit TG, Lawson DM. 2001. The effects of routine cage-changing on cardiovascular and behavioral parameters in male Sprague-Dawley rats. Contemp Top Lab Anim Sci 40(1):17-20.
- EPA Committee on Improving Risk Analysis Approaches Used by the U.S. EPA National Research Council. 2009. Science and Decisions: Advancing Risk Assessment: The National Academies Press.
- Gerdin AK, Igosheva N, Roberson LA, Ismail O, Karp N, Sanderson M, et al. 2012. Experimental and husbandry procedures as potential modifiers of the results of phenotyping tests. Physiol Behav 106(5):602-611.
- Hart B, Lundh A, Bero L. 2012. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ 344:d7202.
- Krauth D, Woodruff T, Bero L. 2013. Instruments for Assessing Risk of Bias and Other Methodological Criteria of Published Animal Studies: A Systematic Review. Environmental Health Perspectives. http://dx.doi.org/10.1289/ehp.1206389
- Lundh A, Lexchin J, Sismondo S, Busuioc O, Bero L. 2012. Industry sponsorship and research outcome . Cochrane Database Syst Rev 12:MR000033
- Macleod MR, O'Collins T, Howells DW, Donnan GA. 2004. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke 35(5):1203-1208.

SUPPLEMENTAL SLIDE #11: References Cited

- Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. 2008. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39(10):2824-2829.
- Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. 2003. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 326(7400): 1171-1173.
- Nieto A, Mazon A, Pamies R, Linana JJ, Lanuza A, Jimenez FO, et al. 2007. Adverse effects of inhaled corticosteroids in funded and nonfunded studies. Arch Intern Med 167(19):2047-2053.
- Porta N, Bonet C, Cobo E. 2007. Discordance between reported intention-to-treat and per protocol analyses. J Clin Epidemiol 60(7): 663-669.
- Rising K, Bacchetti P, Bero L. 2008. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. PLoS Med 5(11):e217; discussion e217.
- Schulz KF, Grimes DA. 2002a. Allocation concealment in randomised trials: defending against deciphering. Lancet 359(9306):614-618.
- Sena E, van der Worp HB, Howells D, Macleod M. 2007. How can we improve the pre-clinical development of drugs for stroke? Trends Neurosci 30(9):433-439.
- Sniekers YH, Weinans H, Bierma-Zeinstra SM, van Leeuwen JP, van Osch GJ. 2008. Animal models for osteoarthritis: the effect of ovariectomy and estrogen treatment - a systematic approach. Osteoarthritis Cartilage 16(5):533-541.
- Tierney JF, Stewart LA. 2005. Investigating patient exclusion bias in meta-analysis. International Journal of Epidemiology 34(1): 79-87.
- van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. 2010. Can animal models of disease reliably inform human studies? PLoS Med 7(3):e1000245.
- Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. 2010. Improving the translational hit of experimental treatments in multiple sclerosis. Mult Scler 16(9):1044-1055.
- Wood, PA. 2000. Phenotype assessment: are you missing something? Comp Med. 50 (1):12-5.

Applying Systematic Review to Assessments of Health Effects of Chemical Exposures

Session 2

Survey of Existing Frameworks and Insights on Integration Challenges

Lorenz Rhomberg, PhD FATS
Gradient

EPA Systematic Review Workshop
26 August 2013
Washington

GRADIENT

Critical Reviews in Toxicology

http://informahealthcare.com/txc
ISSN: 1040-8444 (print), 1547-6898 (electronic)

Crit Rev Toxicol, Early Online: 1–32
 DOI: 10.3109/10408444.2013.832727

informa healthcare

REVIEW

A survey of frameworks for best practices in weight-of-evidence analyses

Lorenz R. Rhomberg[1], Julie E. Goodman[1], Lisa A. Bailey[1], Robyn L. Prueitt[1], Nancy B. Beck[2], Christopher Bevan[3], Michael Honeycutt[4], Norbert E. Kaminski[5], Greg Paoli[6], Lynn H. Pottenger[7], Roberta W. Scherer[8], Kimberly C. Wise[2], and Richard A. Becker[2]

[1]*Gradient Corporation, 20 University Road, Cambridge, MA, USA,* [2]*American Chemistry Council, NE, Washington, USA,* [3]*CJB Consulting LLC,* [4]*Texas Commission on Environmental Quality, MC-168, Austin, USA,* [5]*Michigan State University, East Lansing, USA,* [6]*Risk Sciences International,* [7]*The Dow Chemical Company, Midland, USA, and* [8]*Johns Hopkins School of Public Health*

Abstract

The National Academy of Sciences (NAS) *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* proposed a "roadmap" for reform and improvement of the

Keywords

Data integration, human relevance, mode of action, risk assessment, systematic review

GRADIENT

Survey

- NAS “Roadmap” recommendation
- 50+ frameworks
  - information in online supplement to paper
  - “scored” for features in common and different
- White Paper, then Workshop Discussion
- Not reviews or evaluations, but source of insight into how WoE structures try to meet challenges

WoE "Frameworks" aimed at Specific Evaluations

- Guidance-like, procedural, specified operations and structured evaluations based on stated rules
- Aim at capturing principles of valid scientific inference into rules that apply to the question at hand
    - Rules become standards that analysts can be held to
    - Aim at objective, operational analysis independent of the judge
    - Often with lists of "principles" or "considerations"
- Challenge: Automating "judgment"
    - Too prescriptive → lose credibility, become conventionalized
    - Too unstructured → lose warrant, question whose judgment?

GRADIENT

Phase 1: Define Causal Question and Develop Criteria for Study Selection

- Define causal question or hypothesis
- Define criteria for study inclusion
- Plan literature search
- Design literature search strategies
- Select studies and extract data

Phase 2: Develop and Apply Criteria for Review of Individual Studies

- Assess study quality
- Characterize study quality
- Characterize study relevance

Systematic Presentation and Review of Relevant Data

- Not just positive results from positive studies
  - Also null results from same and other studies
  - Selection / Omission criteria explicit
- Consistent evaluation criteria
  - Design soundness, rigor, statistical power
  - Reliability (aka "internal validity")
    - According to standards of field
    - According to needs of the application
  - Relevance (aka "external validity")
    - … largely a question of interpretation, so intermediate between Phase 1 and Phase 2
- Other "relevant" data – historical controls, understanding of endpoints and MoA, basis for understanding biology, similar agents, etc.

Phase 3 – Integrate and Evaluate Evidence

- Evaluate data within and across realms of evidence
- Integrate negative/null Data into assessment
- Assess adversity of effects
- Assess mode of action (MoA)
- Assess human relevance of MoA

Phase 4 – Draw Conclusions Based on Inferences

- Summary and communication of WoE findings
- Alternative interpretations and uncertainties
- Choices?
- Categories of sufficiency of evidence?
- Are conclusions ultimately justified by soundness of judgment or by following the process?
- "Fit for purpose" assessments -- How do risk management decisions to be made affect categories and evaluation of sufficiency of evidence?

INTEGRATION:

Two Kinds of Inferences from Multiple Studies

- Multiple observations of the thing of interest itself
  - e.g., multiple epidemiologic studies; Evidence-Based Medicine on studies of treatment efficacy
  - Main question is consistency and reliable observation
  - "Weight" from methodologically and statistically reliable measurements
- Indirect evidence of related or relevant phenomena in other systems
  - e.g., animal bioassays, MoA information
  - Main question is relevance and how to generalize
  - Need to integrate across evidence that is relevant in different ways
  - "Weight" from support of relevance arguments

General Kinds of Evidence

- Observed toxicity process that represents an instance of a more general one that would operate in parallel in the target population
- Observed biological perturbation or effect that represents a candidate element of a possible MoA that might operate in the target population
- Evidence by correlation of the study outcome with the target population toxicity of concern in other cases
- Evidence by analogy with other similar cases

Sailing between Scylla and Charybdis

"JUDGMENT"

"RULES"

A ***"Known Human Carcinogen"*** is one for which the evidence is sufficient to conclude that it is a human carcinogen.

A ***"Known Human Carcinogen"*** is one for which, following the framework, one ends up in the "Known Human Carcinogen" box.

Sailing between Scylla and Charybdis

"JUDGMENT"

"RULES"

A ***"Known Human Carcinogen"*** is one for which the evidence is sufficient to conclude that it is a human carcinogen.

A ***"Known Human Carcinogen"*** is one for which, following the framework, one ends up in the "Known Human Carcinogen" box.

"STRUCTURED JUDGMENT"

- guided evaluations with recorded results
- Judgments are proposed explanations of the array of results
- Judgments are justified by citing basis and showing superiority over alternatives

GRADIENT

The Span of Generalization

- We observe particular instances, but what makes them relevant is the potential for *generalization* – that other settings (including the target population) might have similar causal processes.
- What is the span of generalization? What are its limits? Assessing this is part of the WoE.

Key WoE Questions

- Based on observed positives, what hypothesized causal processes are necessary? Sufficient?
- How do they generalize? What *other* manifestations should they have?
- If hypothesis were wrong, how *else* would one explain the array of outcomes?

For Observed Outcomes that are Candidates for "Evidence"

- Why we think they happened where they did.
- Why we think they *didn't* happen where they *didn't*.
- Why we think the "did-happen" factors would also apply to the target population.
    - Might apply? Probably apply? Known to apply?
- Are there discrepant observations, and if so, how do we account for them?
- Are our "whys"
    - Observable underlying causes?
    - Reasonable guesses based on wider knowledge, other cases?
    - *Ad hoc* assumptions without evidence, needed to explain otherwise puzzling phenomena?

GRADIENT

Relative Credence in Competing "Accounts"

- "Account" = an articulated *set* of proposed explanations for the *set* of observations
  - Relevant Causation – but also chance, error, confounding factors, general-knowledge possibilities, plausible assumptions, assertions of irrelevance, and "unknown reasons"

Certain Findings Indicate Target-Population Risk	Those Findings Do Not Indicate Target-Population Risk
• reasoning why • how contradictions resolved • why assumptions reasonable	• reasoning why *not* • how findings are *otherwise* explained • why assumptions reasonable

Can we measure the weights?

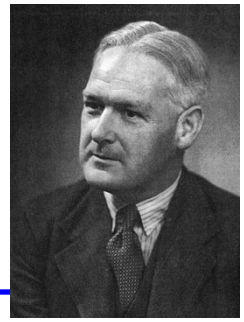
Phase 3 Best Practices

- Evaluate what types of data are being considered and what makes these data evidence.
- Assess data relevant to MoA, human relevance, and dose-response.
- Evaluate negative, null, and positive results.
- Integrate these data across all lines of evidence, so that interpretation of one will inform interpretation of another.
- Ask, if the proposed causative process were true, what other observable consequences should it have, and are these in fact seen?

Phase 3 Best Practices

- Note assumptions, especially when they are *ad hoc* in that they are introduced to explain some phenomenon already seen.
- Evaluate, compare, and contrast alternative explanations of the same sets of results.
- Present conclusions (in text, tables, and figures) not just as the result of judgments but with their context of reasons for coming to them and choosing them over competitors.
- Recognize that applying specific study results to address a more general causation question is an exercise in generalization.
- Based on results of the WoE evaluation, identify data gaps and data needs, and propose next steps.

Sir Austin Bradford Hill on the Hill Criteria



"... ***the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?"*** A. Bradford Hill (1965) *Proc Roy Soc Medicine* **58**:295.

"set of facts" =

- all the epi (+ and -)
- mode of action
- animal studies
- other potential explanations

NCEA Causal Frameworks
Focus on Integrated Science Assessments

Mary A. Ross

Office of Research and Development
National Center for Environmental Assessment

August 26, 2013

Disclaimer: The views expressed are those of the authors and do not necessarily reflect the views or policies of the US EPA.

Integrated Science Assessments

- Synthesis of the most policy-relevant science to provide scientific support for periodic review of national ambient air quality standards (NAAQS) for criteria air pollutants -- O_3, PM, CO, NO_X, SO_X, Pb
- Assess the body of relevant literature, building upon evidence available during previous NAAQS reviews, to draw conclusions on the causal relationships between relevant pollutant exposures and health or environmental effects. Also, evaluate:
  - concentration-, exposure- or dose-response relationships and exposure conditions (dose or exposure, duration and pattern) that are important
  - populations and lifestages that may be more at risk of experiencing effects from pollutant exposure
- Causal framework used in ISAs since 2008
- Provides transparency through structured framework and establishes uniform language concerning causality and brings more specificity to our findings

Informed by Existing Decision-making Frameworks

- EPA Guidelines for Carcinogen Risk Assessment (EPA, 2005)
  - Carcinogenic to Humans
  - Likely to Be Carcinogenic to Humans
  - Suggestive Evidence of Carcinogenic Potential
  - Inadequate Information to Assess Carcinogenic Potential
  - Not Likely to Be Carcinogenic to Humans
- Surgeon General's Report on Smoking (CDC, 2004)
- Improving the Presumptive Disability Decision-Making Process for Veterans (IOM, 2008)

EPA United States Environmental Protection Agency

Data Available for Assessments Varies

	Pharma-ceuticals	Pesticides	Criteria air pollutants	IRIS chemicals
Randomized control trials	Required	--	--	--
Guideline-based animal studies	Required	Required	Sometimes	Sometimes (*e.g.* NTP)
Epidemiology studies at ambient exposure levels	--	Sometimes	Extensive	Sometimes
Other epidemiology studies	Post-market surveillance	Sometimes	Yes	Sometimes
Other animal studies	Sometimes	Sometimes	Yes	Usually

Integrated Risk Information System: Preamble

- ***Carcinogenic to humans:*** There is convincing epidemiologic evidence of a causal association (that is, there is reasonable confidence that the association cannot be fully explained by chance, bias, or confounding); or there is strong human evidence of cancer or its precursors, extensive animal evidence, identification of key precursor events in animals, and strong evidence that they are anticipated to occur in humans.
- ***Likely to be carcinogenic to humans:*** The evidence demonstrates a potential hazard to humans but does not meet the criteria for *carcinogenic*. There may be a plausible association in humans, multiple positive results in animals, or a combination of human, animal, or other experimental evidence.
- ***Suggestive evidence of carcinogenic potential:*** The evidence raises concern for effects in humans but is not sufficient for a stronger conclusion. This descriptor covers a range of evidence, from a positive result in the only available study to a single positive result in an extensive database that includes negative results in other species.
- ***Inadequate information to assess carcinogenic potential:*** No other descriptors apply. *Conflicting evidence* can be classified as *inadequate information* if all positive results are opposed by negative studies of equal quality in the same sex and strain. *Differing results*, however, can be classified as *suggestive evidence* or as *likely to be carcinogenic*.
- ***Not likely to be carcinogenic to humans:*** There is robust evidence for concluding that there is no basis for concern. There may be no effects in both sexes of at least two appropriate animal species; positive animal results and strong, consistent evidence that each mode of action in animals does not operate in humans; or convincing evidence that effects are not likely by a particular exposure route or below a defined dose.

Draft IRIS assessment for ammonia:
http://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=200305

Causal Framework - ISAs

- Five categories based on overall weight of evidence:
  - Causal relationship
  - Likely to be a causal relationship
  - Suggestive of a causal relationship
  - Inadequate to infer a causal relationship
  - Not likely to be a causal relationship
- Availability and relative importance of different types of evidence varies by pollutant or assessment

Causal Framework for Integrated Science Assessments (ISAs)

Table II **Weight of evidence for causal determination.**

	Health Effects	Ecol
Causal relationship	Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (i.e., doses or exposures generally within one to two orders of magnitude of current levels). That is, the pollutant has been shown to result in health effects in studies in which chance, bias, and confounding could be ruled out with reasonable confidence. For example: a) controlled human exposure studies that demonstrate consistent effects; or b) observational studies that cannot be explained by plausible alternatives or are supported by other lines of evidence (e.g., animal studies or mode of action information). Evidence includes multiple high-quality studies	Evider relatio i.e., do orders pollut in whic out wit studie studie but the determ by mul consid usually lines o
Likely to be a causal relationship	Evidence is sufficient to conclude that a causal relationship is likely to exist with relevant pollutant exposures, but important uncertainties remain. That is, the pollutant has been shown to result in health effects in studies in which chance and bias can be ruled out with reasonable confidence but potential issues remain. For example: a) observational studies show an association, but copollutant exposures are difficult to address and/or other lines of evidence (controlled human exposure, animal, or mode of action information) are limited or inconsistent; or b) animal toxicological evidence from multiple studies from different laboratories that demonstrate effects, but limited or no human data are available. Evidence generally includes multiple high-quality studies.	Evider causal That is pollut bias a remain but sus and ot Gener multiple researc
Suggestive of a causal relationship	Evidence is suggestive of a causal relationship with relevant pollutant exposures, but is limited. For example, (a) at least one high-quality epidemiologic study shows an association with a given health outcome but the results of other studies are inconsistent; or (b) a well-conducted toxicological study, such as those conducted in the National Toxicology Program (NTP), shows effects in animal species,	Evidence is sug relevant polluta confounding ca one high-quality of other studies
Inadequate to infer a causal relationship	Evidence is inadequate to determine that a causal relationship exists with relevant pollutant exposures. The available studies are of insufficient quantity, quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of an effect.	The available s consistency, or regarding the p
Not likely to be a causal relationship	Evidence is suggestive of no causal relationship with relevant pollutant exposures. Several adequate studies, covering the full range of levels of exposure that human beings are known to encounter and considering at-risk populations, are mutually consistent in not showing an effect at any level of exposure.	Several adequa relevant exposu effect at any lev

Causal relationship

Evidence is sufficient to conclude that there is a causal relationship with relevant pollutant exposures (i.e., doses or exposures generally within one to two orders of magnitude of current levels). That is, the pollutant has been shown to result in health effects in studies in which chance, bias, and confounding could be ruled out with reasonable confidence. For example: a) controlled human exposure studies that demonstrate consistent effects; or b) observational studies that cannot be explained by plausible alternatives or are supported by other lines of evidence (e.g., animal studies or mode of action information). Evidence includes multiple high-quality studies

From Preamble to ISAs

Evaluation of evidence

- **Types of health studies:**
  - **Controlled human exposure studies:** Controlled exposures and conditions; small sample size, generally healthy subjects, short exposure time
  - **Epidemiologic studies:** Real-world exposures and human populations; need to consider potential confounders, exposure error, design factors
  - **Animal toxicological studies:** Controlled exposures, exposure pathways or mechanisms; consider homology to effects in humans
- **Bradford-Hill "aspects" aid in judging causality:**
  - Consistency
  - Strength
  - Specificity
  - Temporal relationship
  - Biological gradient
  - Biological plausibility
  - Coherence
  - Experimental evidence
  - Analogy

Example: Application of Causal Framework in the Pb ISA

United States Environmental Protection Agency

Table 4-17 Summary of Evidence Supporting Nervous System Causal Determinations.

Attribute in Causal Framework[a]	Key Evidence[b]	References[b]	Pb Biomarker Levels Associated with Effects[c]
Cognitive Function Decrements in Children - Causal			
Consistent associations from multiple, high quality epidemiologic studies with relevant blood Pb levels	Evidence from prospective studies for decrements in FSIQ in association with prenatal, earlier childhood, peak, concurrent, lifetime average blood Pb levels and tooth Pb levels in children ages 4-17 yr in multiple U.S. locations, Mexico, Europe, Australia	Canfield et al. (2003a), Bellinger et al. (1992), Jusko et al. (2008), Dietrich et al. (1993b), Schnaas et al. (2006), Wasserman et al. (1997), Tong et al. (1996), Lanphear et al. (2005) Plus Table 4-3, Section 4.3.2.1	Blood Pb (various time periods & lifestages): Means 3-16 µg/dL With consideration of peak or early childhood blood Pb levels: Means 3-8 µg/dL for concurrent (age 4, 5 yr), age 2 yr
	Evidence from prospective studies for lower scores on tests of executive function and academic performance in association with earlier childhood or lifetime average blood Pb levels or tooth Pb levels in children ages 5-20 yr in multiple U.S. locations, U.K, New Zealand. Associations less consistent for learning and memory.	Canfield et al. (2004), Stiles and Bellinger (1993), Miranda et al. (2009; 2007a), Fergusson et al. (1997, 1993), Leviton et al. (1993), Chandramouli et al. (2009) Sections 4.3.2.3, 4.3.2.4, 4.3.2.5	Blood Pb (various time periods & lifestages): Means 4.8-7.2 µg/dL, Groups with early childhood blood Pb 2-16 µg/dL and 5-10 µg/dL Tooth Pb (ages 6-8 yr): means 3.3, 6.2 µg/g
	Supporting evidence from cross-sectional studies of children ages 3-16 yr, but most did not consider potential confounding by parental caregiving quality. Includes large NHANES III analysis.	Surkan et al. (2007), Kim et al. (2009b), Roy et al. (2011), Lanphear et al. (2000), Froehlich et al. (2007), Chiodo et al. (2007; 2004)	Concurrent (ages 3-16 yr) blood Pb : Means 1.7-12 µg/dL, Group (ages 6-10 yr) with blood Pb 5-10 µg/dL
	Outcomes assessed using widely-used, structured questionnaires.		
	Several studies indicate supralinear C-R relationship, with larger decrements in cognitive function per unit increase in blood Pb at lower blood Pb levels in children ages 5-10 yr	Canfield et al. (2003a), Bellinger et al. (1992), Jusko et al. (2008), Kordas et al. (2006), Lanphear et al. (2005) Plus Table 4-16	Groups with peak blood Pb <10 µg/dL: concurrent mean 3.3 µg/dL, age 2 year mean 3.8 µg/dL

Transparent Application of Causal Framework (cont'd)

Table 4-17 (Continued): Summary of Evidence Supporting Nervous System Causal Determinations.

Attribute in Causal Framework[a]	Key Evidence[b]	References[b]	Pb Biomarker Levels Associated with Effects[c]
Additional epidemiologic evidence to help rule out chance, bias, and confounding with reasonable confidence	Several epidemiologic studies found associations with adjustment for SES, maternal IQ and education, HOME score. Several adjust for birth weight, smoking. A few, nutritional factors. Epidemiologic studies had population-based recruitment, most with moderate to high follow-up participation not conditional on blood or tooth Pb level or cognitive function.	Table 4-3, Table 4-5; Table 4-8, Table 4-9, Sections 4.3.2.1, 4.3.2.3 4.3.2.4, and 4.3.2.5	
	Pooled and meta-analyses demonstrate the consistency of association	Lanphear et al. (2005), Pocock et al. (1994), Schwartz (1994)	
Consistent evidence in animals with relevant exposures to help rule out chance, bias, and confounding with reasonable confidence	Impaired learning and associative ability in juvenile and adult animals as indicated by performance in tasks of visual discrimination, water maze, y maze, and operant conditioning with schedules of reinforcement with relevant dietary Pb exposure.	Stangle et al. (2007), Niu et al. (2009), Cory-Slechta et al. (2010), Altmann et al. (1993), Section 4.3.2.3	Blood Pb (after prenatal/ lactation, lactation only, prenatal/lifetime Pb exposure): 10-25 µg/dL
	Impaired learning, memory, executive function in adult monkeys as indicated by poorer performance on delayed spatial alternation and spatial discrimination reversal learning tasks with dietary Pb exposures.	Gilbert and Rice (1987), Rice and Karpinski (1988), Sections 4.3.2.3 and 4.3.2.4	Blood Pb (after lifetime Pb exposure from birth): 15, 25 µg/dL

Transparent Application of Causal Framework (cont'd)

Table 4-17 (Continued): Summary of Evidence Supporting Nervous System Causal Determinations.

Attribute in Causal Framework[a]	Key Evidence[b]	References[b]	Pb Biomarker Levels Associated with Effects[c]
Evidence describes mode of action:	Decreased neurogenesis in hippocampus DG (involved in LTP and learning). Decreased NMDAR (involved in integration of new neurons into existing neuronal pathways). Decreased neurite outgrowth.	Sections 4.3.10.9 and 4.3.10.10	
Impaired neuron development	Found in animals with dietary gestational-lactational, lactational, post-lactational (3-8 weeks), lifetime from gestation Pb exposures.		
Synaptic changes	Decreased synaptic development. Changes in synaptic protein composition. Decreased ATP and AchE, which both mediate neurotransmission.	Section 4.3.10.4	
	Found in animals with dietary gestational with or without additional lactational Pb exposures.		
LTP	Decreased magnitude, increased threshold of LTP with gestational-lactational or lifetime Pb exposure.	Sections 4.3.12, 4.3.10.7, 4.3.10.8	
Neurotransmitter changes	Changes in dopamine metabolism. Increased sensitivity of dopamine receptor. Increased catecholamine transmission in cerebral cortex, cerebellum, hippocampus. Decreased glutamate and expression of glutamate receptor, NMDAR.	Section 4.3.10.8	
	Found in animals with dietary gestational-lactational, lactational, or post-lactational Pb exposure.		

2013 Pb ISA:
http://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=255721

Example: Short-Term O_3 Exposure and Cardiovascular Effects

Likely Causal determination supported by:

– Strong toxicological evidence from a small body of recent and past studies for systemic oxidative stress and inflammation which may promote progression of atherosclerosis and enhance ischemia-reperfusion injury.

– Controlled human exposure studies showed evidence of systemic oxidative stress. One key new study provided evidence of systemic inflammation, a prothrombogenic environment, and altered heart repolarization.

– Epidemiologic evidence:

- Consistent, positive associations between short-term exposure and cardiovascular mortality

- Inconsistent findings for cardiovascular morbidity (e.g., heart rhythm, physiological biomarkers, and hospital admissions or emergency department visits)

Hospital Admissions and ED visits

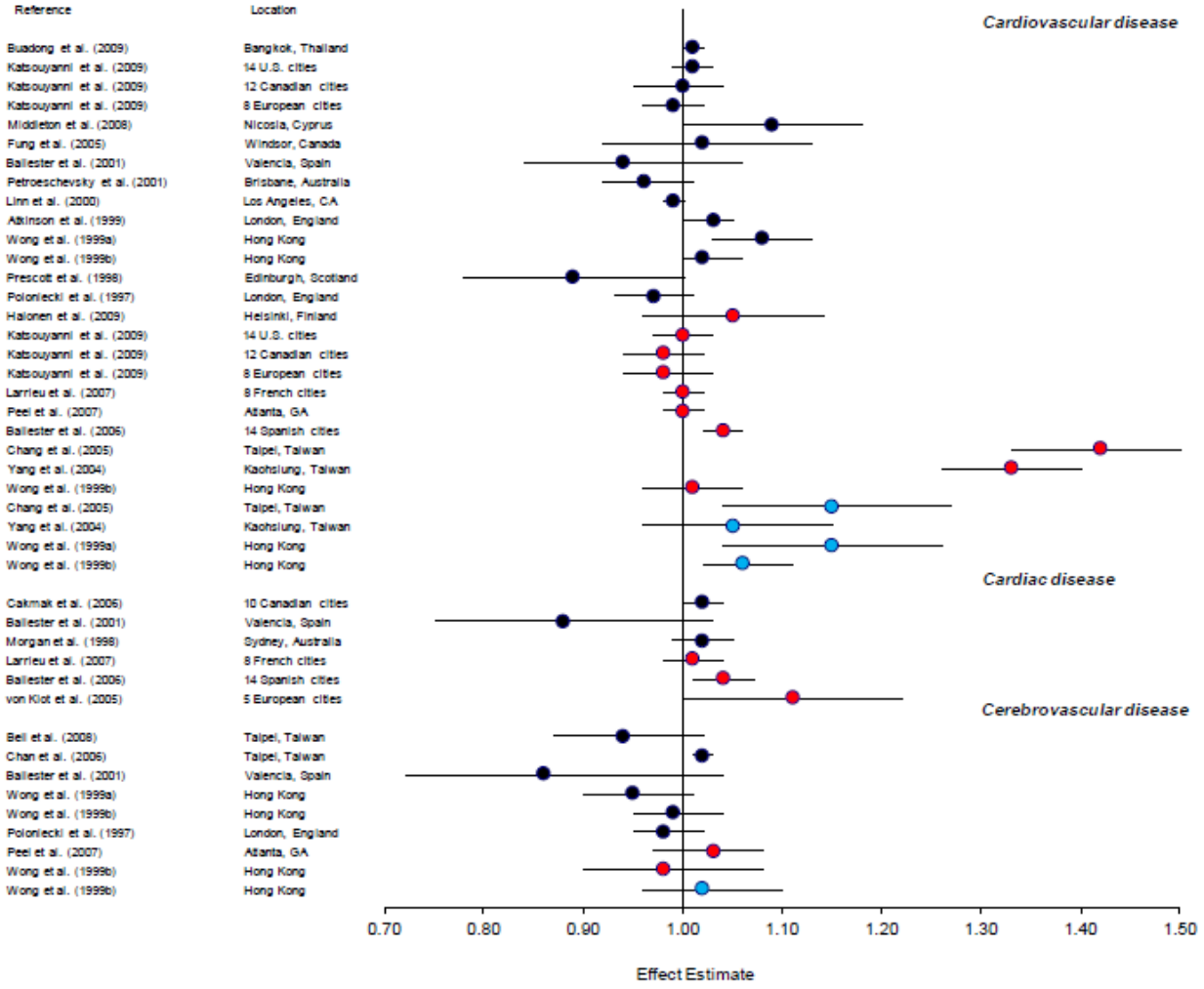


Figure 6-22. Odds ratio (95% CI) per increment ppb increase in ozone for overall cardiovascular ED visits or HAs.

Note: Increase in O_3 standardized to 20 ppb for 24-h avg period, 30 ppb for 8-h avg period, and 40 ppb for 1-h avg period. Ozone concentrations in ppb. Seasons depicted by colors – black: all year; red: warm season; light blue: cold season. Age groups of study populations were not specified or were adults with the exception of Fung et al. (2005), Wong et al. (1999b), and Prescott et al. (1998), which included only individuals aged 65+.

Cause-Specific Mortality

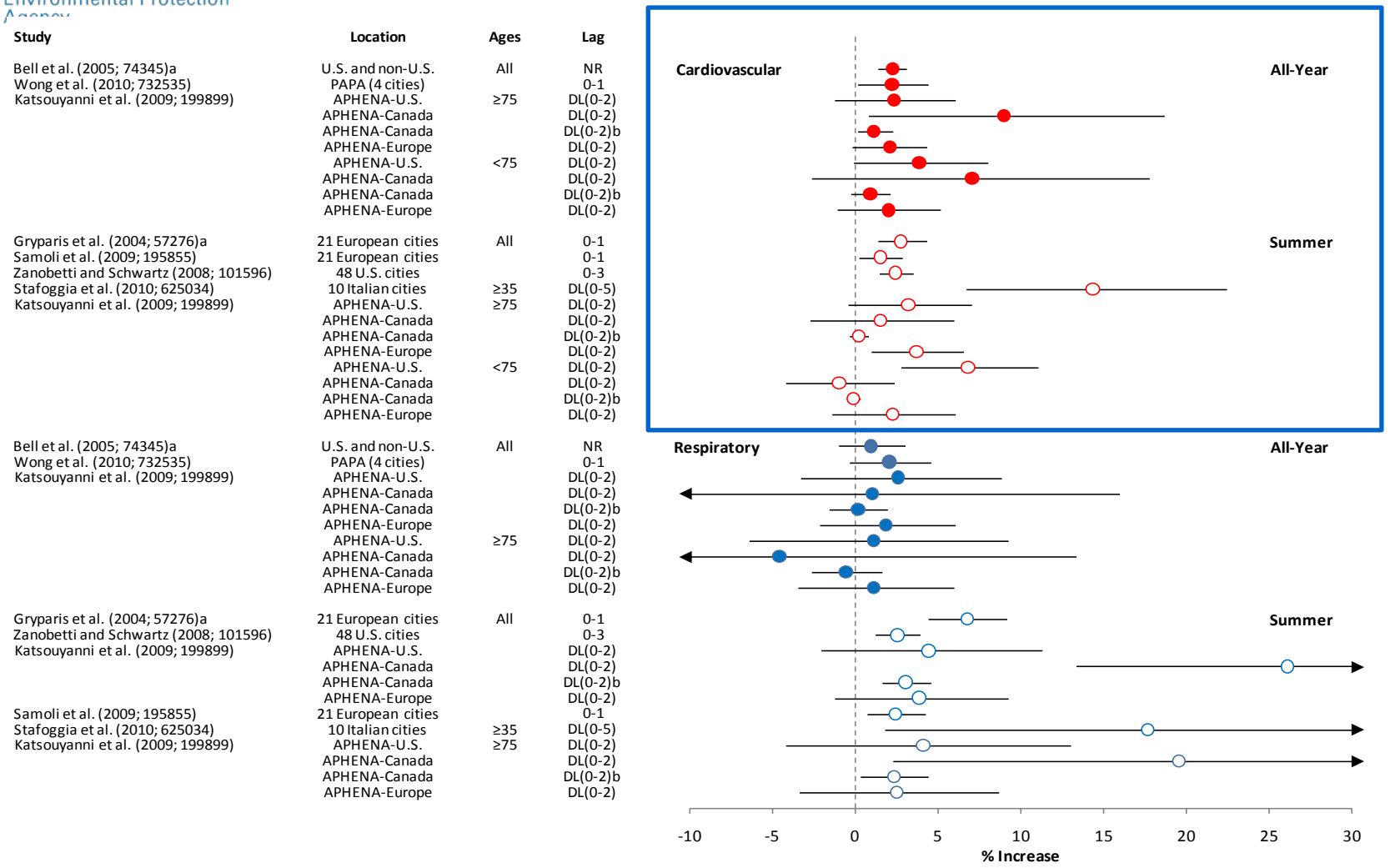


Figure 6-37 Percent increase in cause-specific mortality.

Effect estimates are for a 20 ppb increase in 24-h avg; 30 in 8-h max; and 40ppb increase in 1-h max ozone concentrations. Red = cardiovascular; blue = respiratory; closed circles = all-year analysis; and open circles = summer-only analysis. An "a" represents studies from the 2006 ozone AQCD. A "b" represents risk estimates from APHENA-Canada standardized to an approximate IQR of 5.1 ppb for a 1-h max increase in ozone concentrations (Section 6.2.7.2).

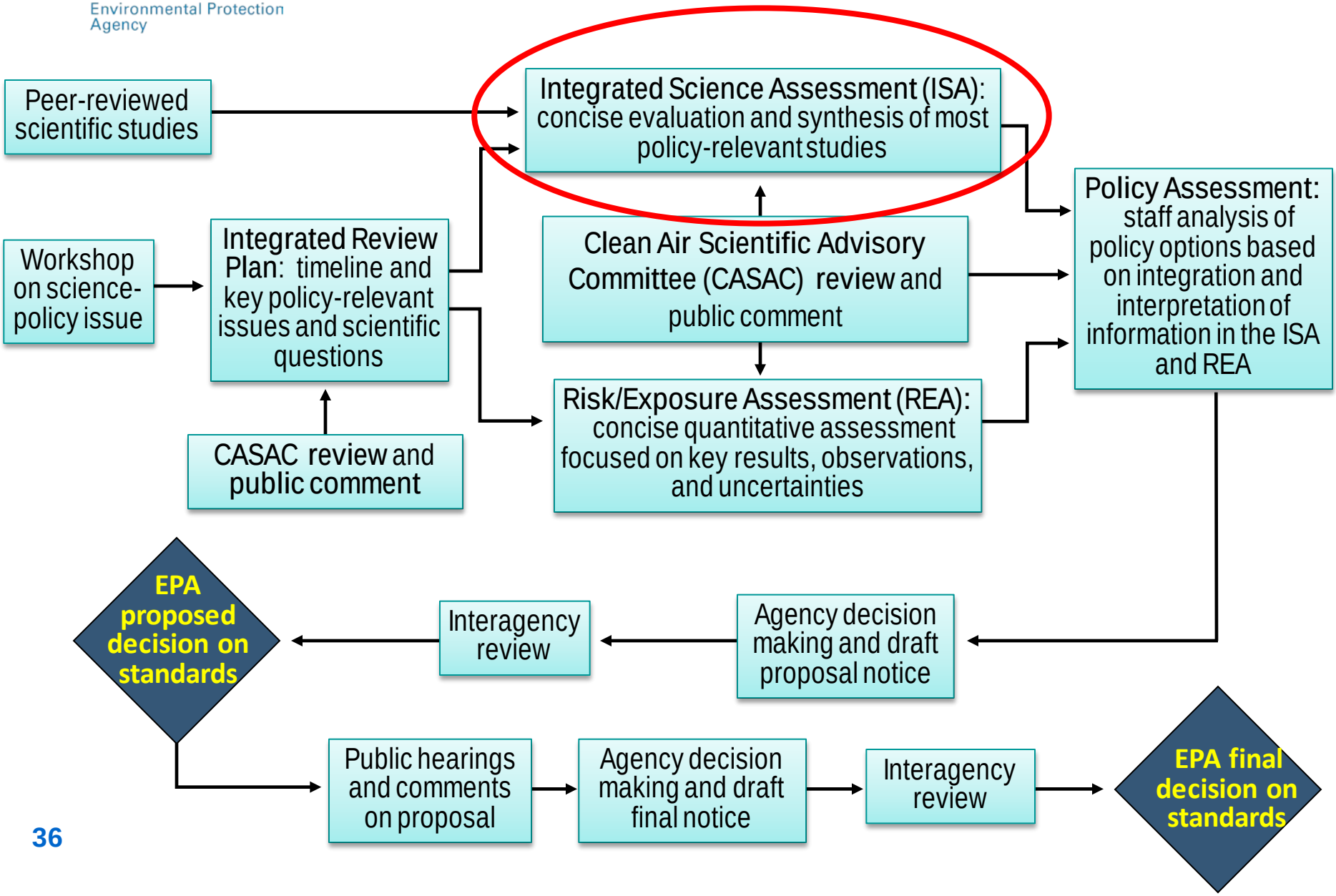
Summary

- Causal framework supports transparency and consistency in evaluation of scientific evidence and conclusions in ISAs
  - Clean Air Scientific Advisory Committee support for use of framework
- Weight of evidence and availability of evidence from different disciplines varies for pollutants and health outcomes, for example:
  - Controlled human exposure studies provide evidence for respiratory effects of gaseous pollutants such as O_3; not conducted for Pb or effects such as mortality
  - Large body of epidemiologic evidence available for pollutants such as PM

United States Environmental Protection Agency

Supplemental

National Ambient Air Quality Standard Review Process



Putting the Pieces Together

Navigation Guide Proof of Concept: A Systematic Review of Human and Non-Human Evidence for PFOA and Fetal Growth

US EPA

August 26, 2013

Tracey J. Woodruff, PhD, MPH
UCSF Program on Reproductive Health and the Environment

What Is A Systematic Review?

- Transparent and systematic approach to evaluating available evidence
- Developed to prevent harm from treatment decisions being made without strong basis in the evidence

THE COCHRANE COLLABORATION®

Model for Navigation Guide

GRADE

Navigation Guide Work Group

Systematic and transparent methodology

Provides uniform, simple, and transparent summaries

Integrates the best practices of evaluation in **environmental** and **clinical** health sciences

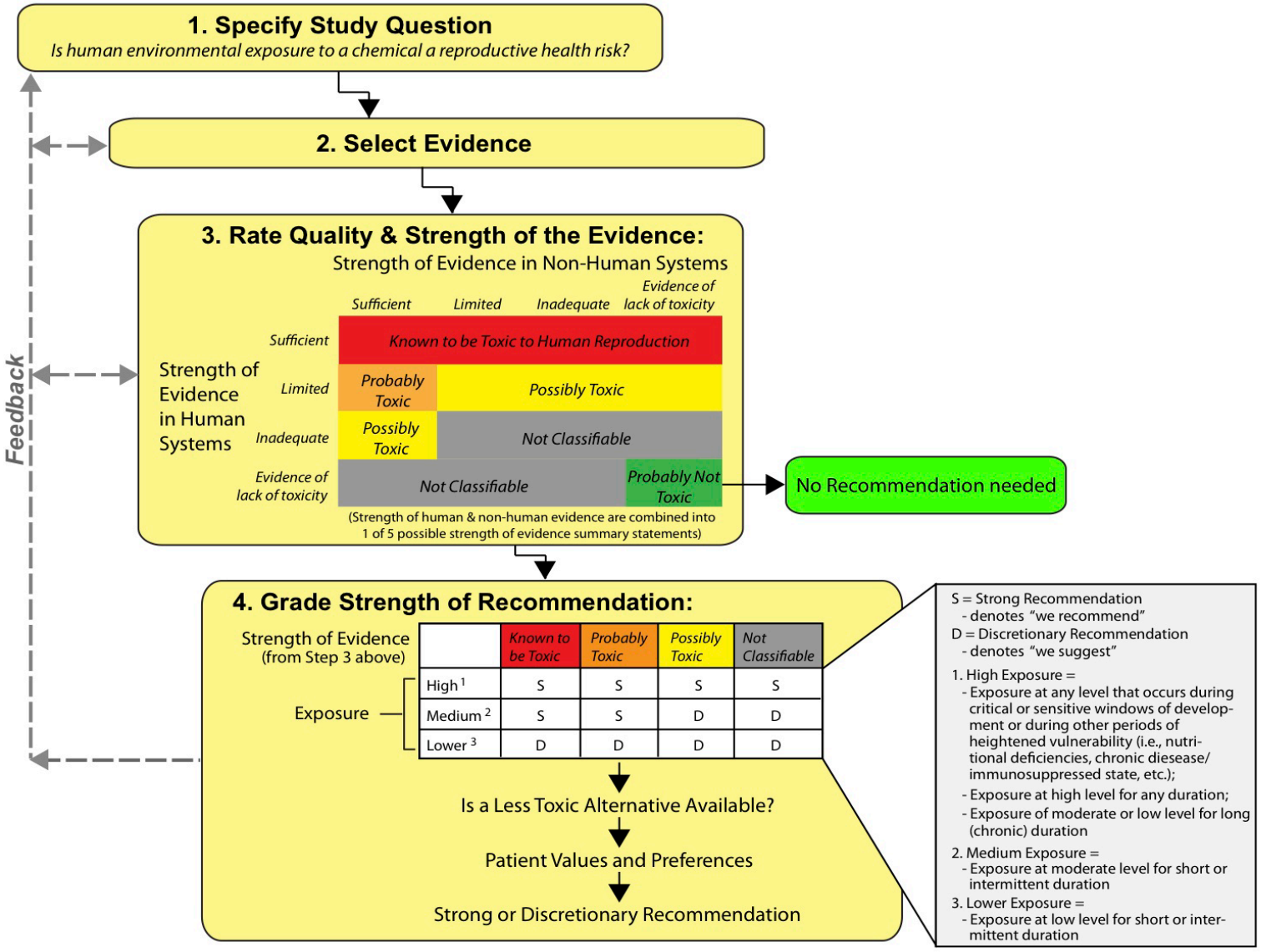
BRIDGING CLINICAL & ENVIRONMENTAL HEALTH

By Tracey J. Woodruff, Patrice Sutton, and The Navigation Guide Work Group

An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences

ABSTRACT Physicians and other clinicians could help educate patients about hazardous environmental exposures, especially to substances that could affect their reproductive health. But the relevant scientific evidence is voluminous, of variable quality, and largely unfamiliar to health professionals caring for people of childbearing age. To bridge this gap between clinical and environmental health, we created a methodology to help evaluate the quality of evidence and to support evidence-based decision making by clinicians and patients. The methodology can also support professional societies, health care organizations, government agencies, and others in developing prevention-oriented guidelines for use in clinical and policy settings.

Overview of the Navigation Guide Methodology



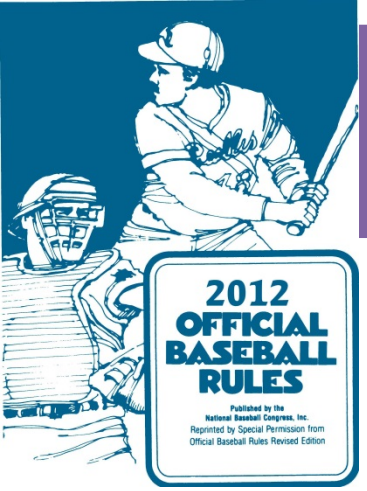
Establishing Proof-of Concept

Methods

Top priorities:

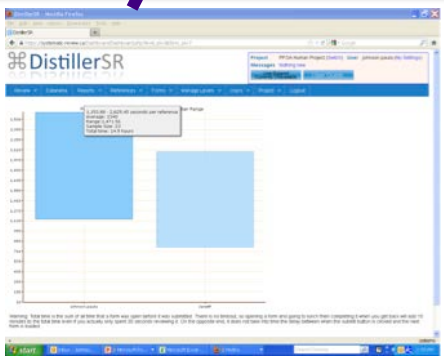
Systematic, Transparent & Reproducible

- *GRADE and Cochrane Handbook for Systematic Reviews of Interventions* as a guide
- Multiple reviewers independently perform several steps of process to ensure accuracy/consensus/reproducibility
- *A priori* protocol development essential for guiding systematic review

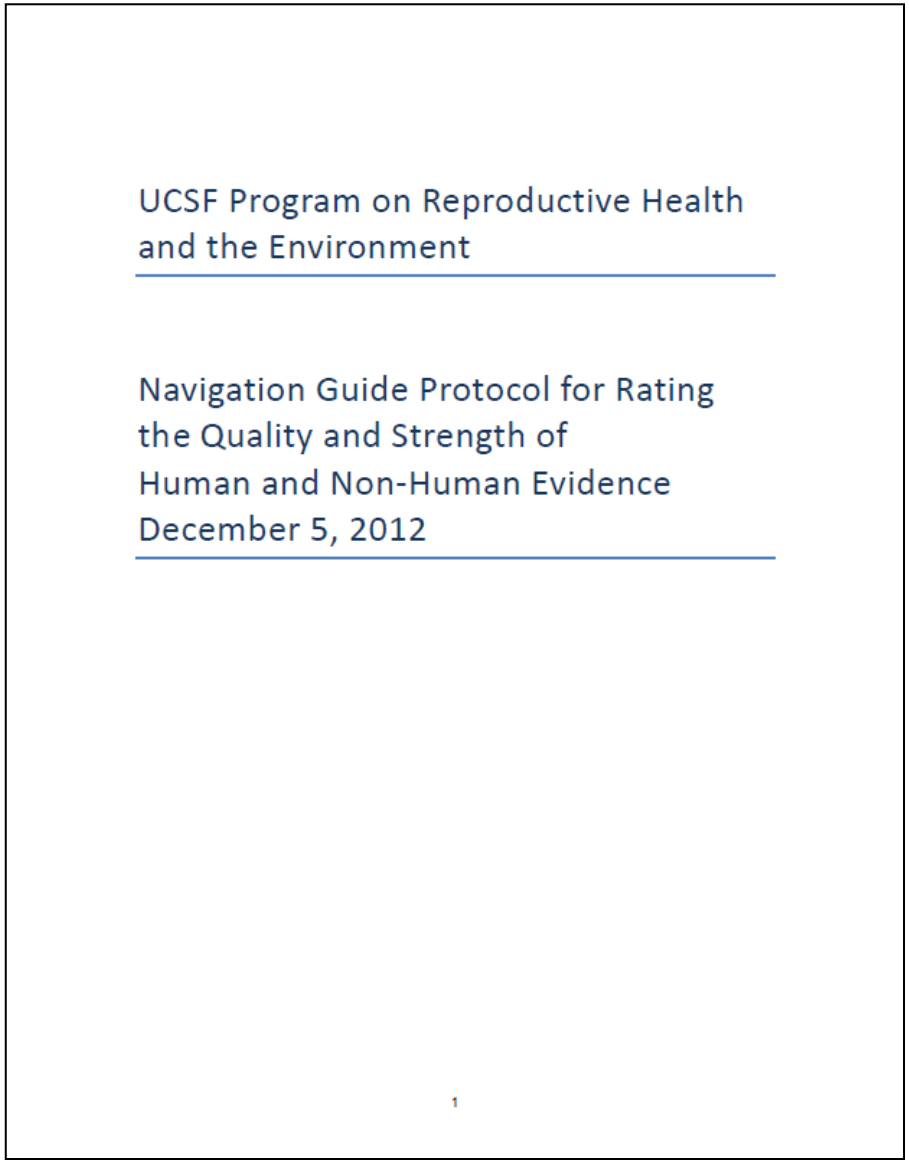


A Priori Written Protocol

- "PECO" Statement
- Criteria for Selecting Studies
- Select the Studies
- Extract data
- Data Analysis
- Risk of Bias
- Rating the Quality of Evidence
- Rating the Strength of Evidence



Results



UCSF Program on Reproductive Health and the Environment

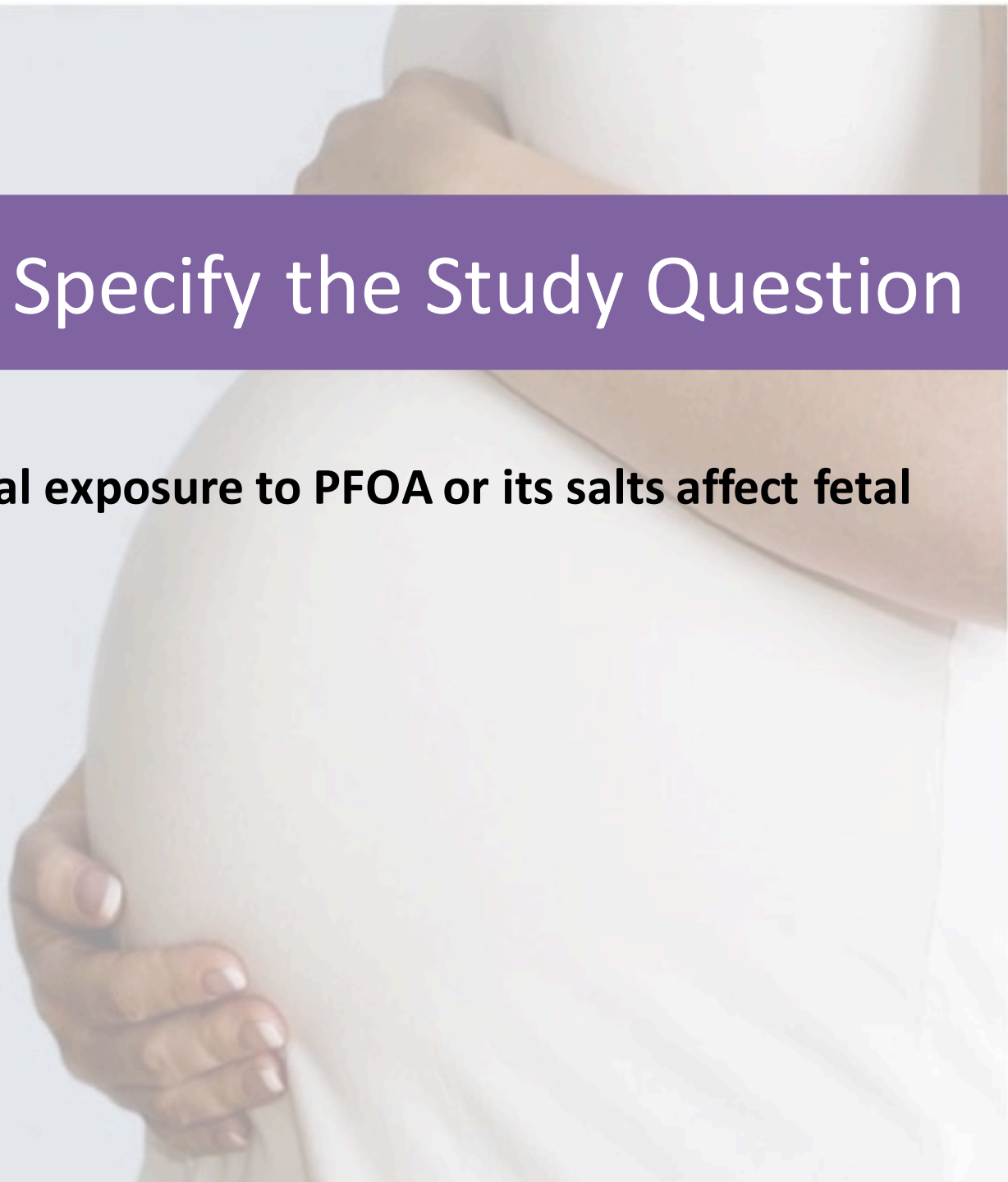
Navigation Guide Protocol for Rating the Quality and Strength of Human and Non-Human Evidence December 5, 2012

1

Results Step 1. Specify the Study Question

Does fetal developmental exposure to PFOA or its salts affect fetal growth?

P opulation
E xposure
C omparator
O utcome



PECO

Population: Animals from non-human species that are studied during reproductive/developmental time period (before and/or during pregnancy for females or during development for embryos).

Exposure: One or more oral, subcutaneous or other treatment(s) of any dosage with perfluorooctanoic acid (PFOA), CAS# 335-67-1, or its salts during the time before pregnancy and/or during pregnancy for females or directly to embryos.

Comparator: Experimental animals receiving different doses of PFOA or vehicle-only treatment.

Outcome: Changes in fetal weight near term (for example, embryonic day 18 for mice and embryonic day 21 for rat); birth weight; and/or other measures of size at term or birth, such as length.

PECO

Population: Humans that are studied during reproductive/developmental time period (before and/or during pregnancy or development).

Exposure: Exposure to perfluorooctanoic acid (PFOA), CAS# 335-67-1, or its salts during the time before pregnancy and/or during pregnancy for females or directly to fetuses.

Comparator: Humans exposed to lower levels of PFOA than the more highly exposed humans.

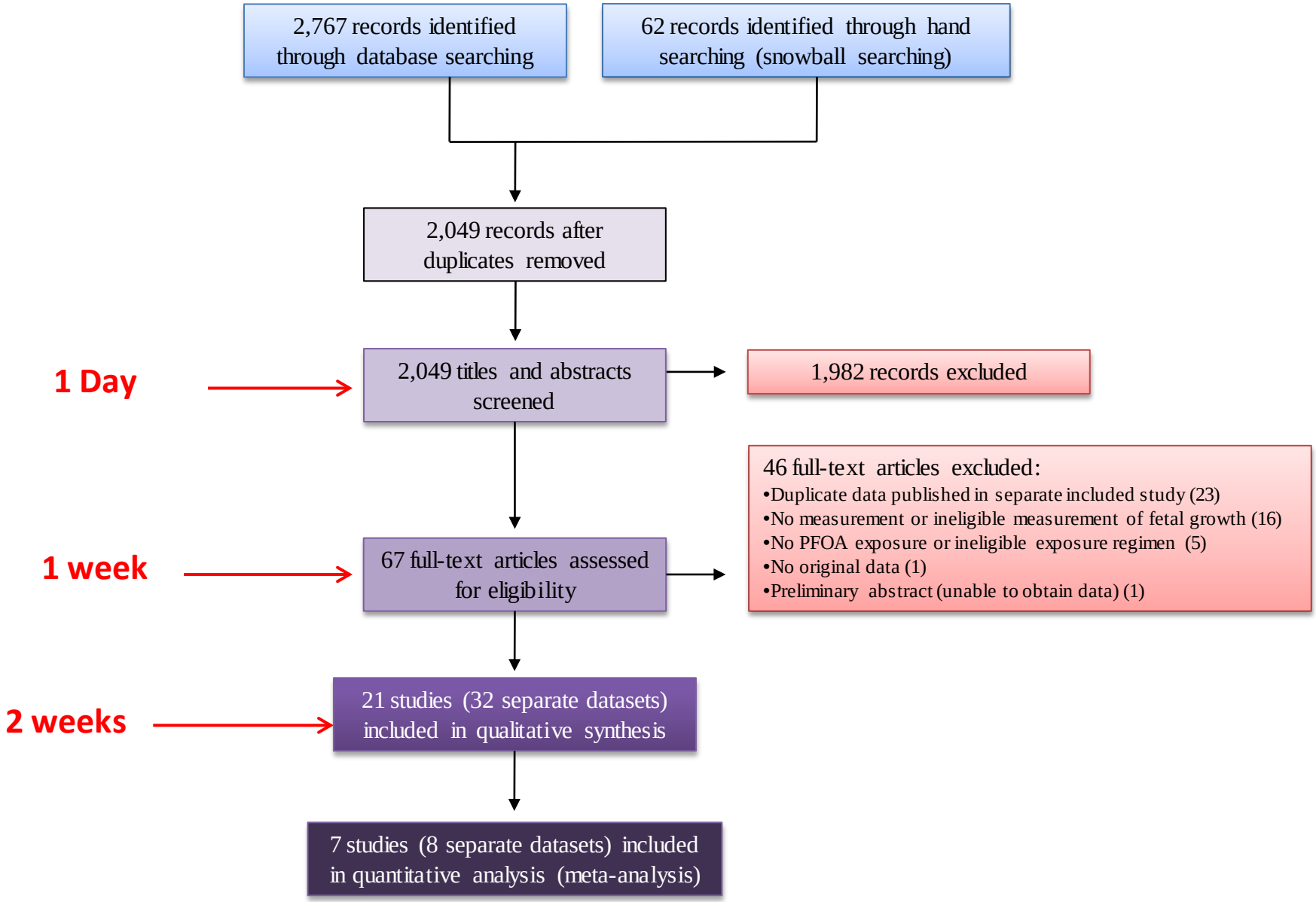
Outcome: Effects on fetal growth, birth weight, and/or other measures of size, such as length.

Step 2. Select the Evidence

- **Systematic Search**
  - Designed based on keywords from papers of interest
  - Reproducible
  - Inclusive of non-English papers and non-published sources (grey literature)

- **Study selection**
  - Compared to *a priori* defined criteria
  - Performed by 2 reviewers, subset confirmed by 3rd reviewer
  - Carefully tracked to maximize transparency

Non-human study selection process



Search Strategy Comparison for Non-Human Studies

Traditional Search

January 2011

PubMed = 140 studies
Web of Science = 10 studies
Handsearching Citations = 11 studies

Remove Duplicates

146

Screening

11

Systematic Search

February 2012

PubMed = 1462 studies
Web of Science = 1060 studies
Handsearching Citations = 62 studies
Tox Databases = 263 studies

Remove Duplicates

2049

Screening

21

NOTE: For 1/11 search, screened over 7000 articles (screened for each combination of terms)

Summary of Study Characteristics

Species


Mouse


Chicken


Rat


Fly


Salmon


Zebrafish

Route of Exposures

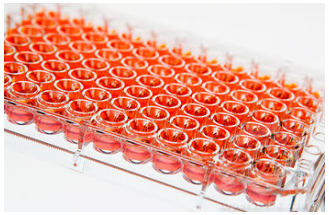

Gavage


Food


Drinking Water


Inhalation


Injection into Egg


Egg Immersion

Summary of Study Characteristics

Time point of Growth Measurement

At Birth

Near Term

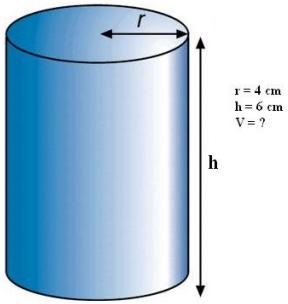
Not Stated

During larval development

Method of Growth Measurement

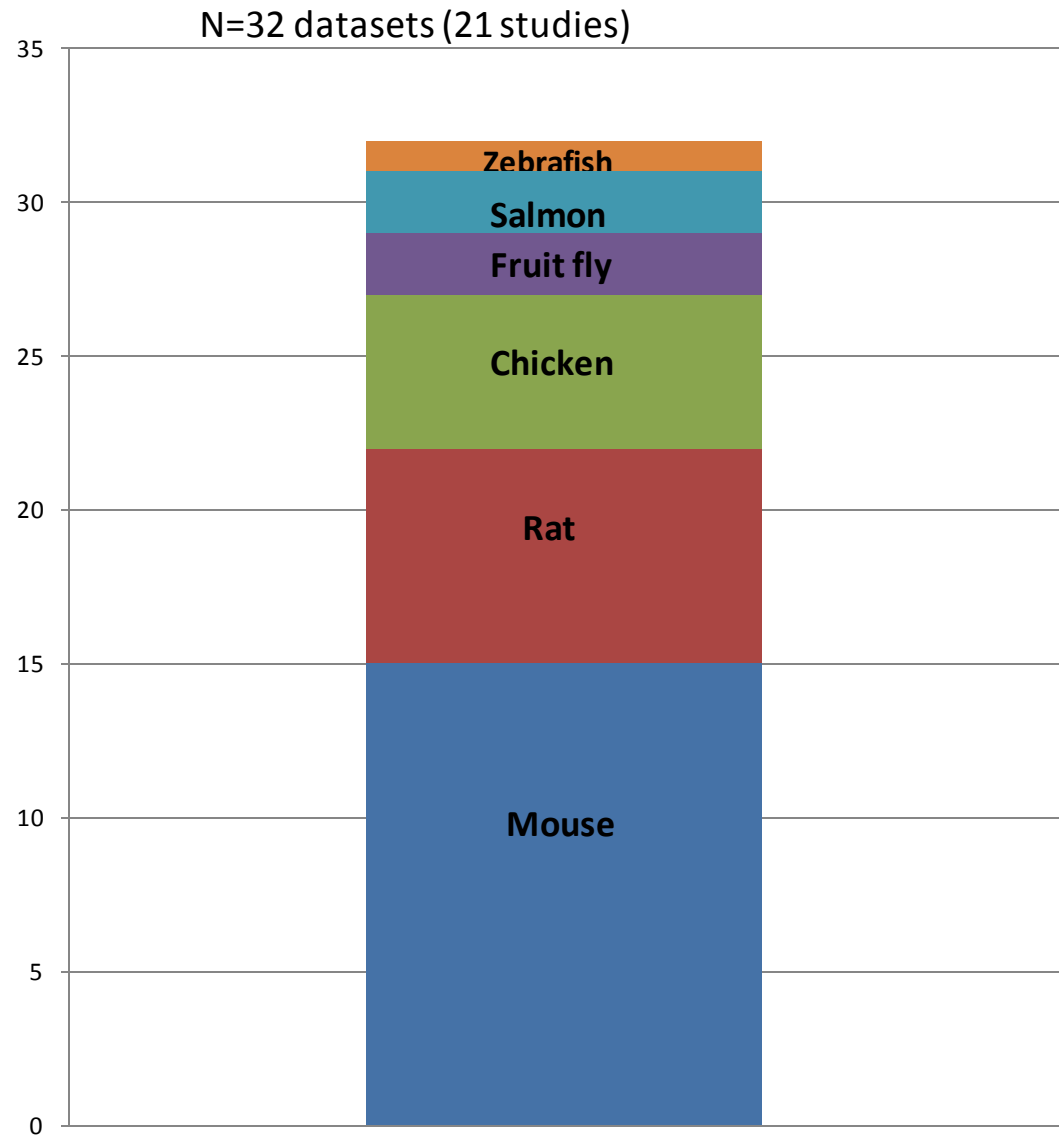
Weight

Length

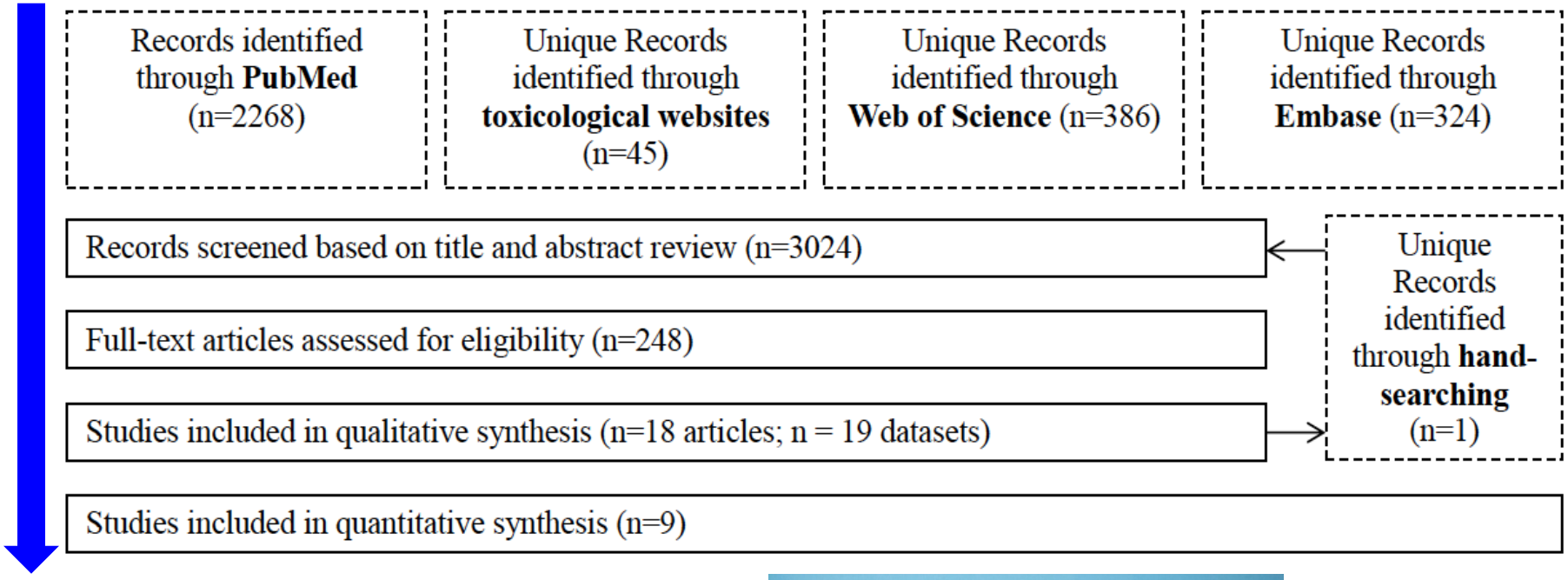


Larval Volume

Results of Non-Human-Non-Mammalian Evidence



Human study selection process



1 Day Title and Abstract Review
1 Week Full text review
2 Days Data Extraction
~ 2 Weeks - Total

Search Strategy Comparison for Human Studies

C8 Science Panel (Dec 2011)

Apelberg et al 2007
Fei et al 2007
Hamm et al 2010
Monroy et al 2008
Nolan et al 2009
Savitz et al 2012a
Savitz et al 2012b
Stein et al 2009
Washino et al 2009

The Navigation Guide search strategy was a more comprehensive method

Navigation Guide (2012)

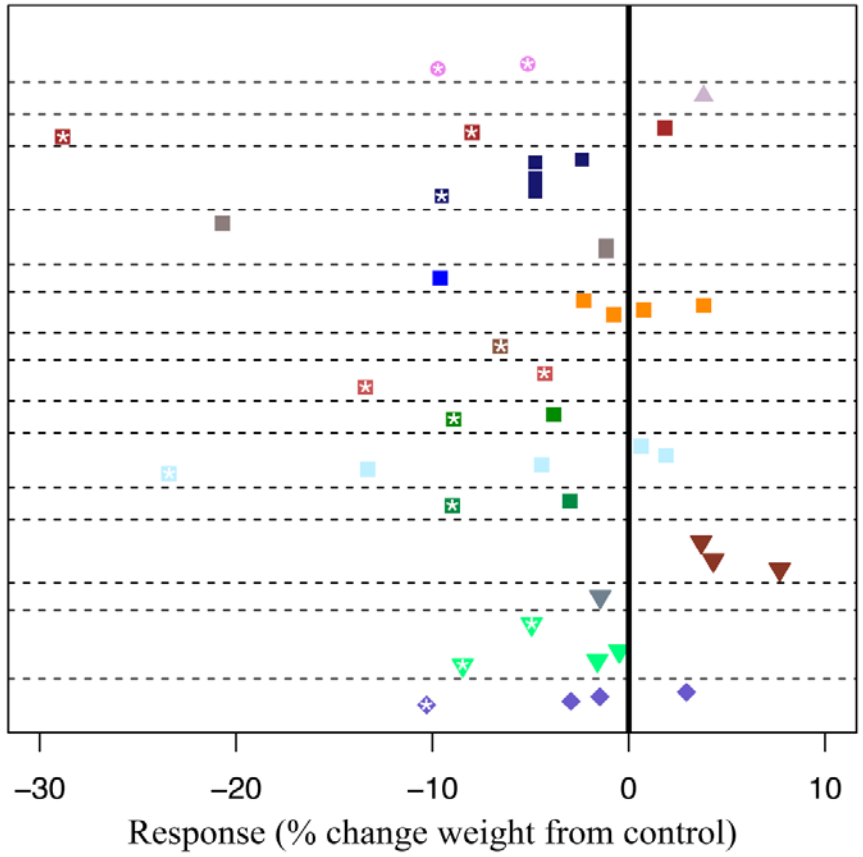
Apelberg et al 2007
Arbuckle 2012
Fei et al 2007
Fromme et al 2010
Halldorsson et al 2012
Hamm et al 2010
Kim S et al 2011
Kim S-K et al 2011
Monroy et al 2008
Nolan et al 2009
Savitz et al 2012a
Savitz et al 2012b
Stein et al 2009
Wang et al 2011 -> Chen et al 2012
Washino et al 2009
Whitworth et al 2012
Maisonet et al 2012

Step 3&4. Data Extraction & Analysis

- Data extracted by two reviewers to ensure accuracy
- Summary plots allow all data to be compared on the same scale
- Identify similarities/differences across studies

Study data: Pup mammalian weight

Study [study ID]	Species	Route of exposure	Maximum dose**
Hu 2010 [68]	Mouse	Drinking Water	1
Onishchenko 2011 [3610]	Mouse	Food	0.3
Yahia 2010 [103]	Mouse	Gavage	10
Hines 2009 [260]			5
Fenton 2009 [264]			5
White 2009 [312]			5
Abbott 2007 [528]			1
White 2007 [566]			5
Wolf 2007 [571b]#			20
Wolf 2007 [571a]#			5
Lau 2006 [635]			20
White 2011 [3862]			5
Hinderliter 2005 [711]	Rat	Gavage	30
Staples 1984 [1871]			100
York 2002 [5122]			30
Staples 1984 [1871]	Rat	Inhalation	25 mg/m^3



Doses in figure decrease as y-axis increases

***mg/kg BW/day unless otherwise specified*

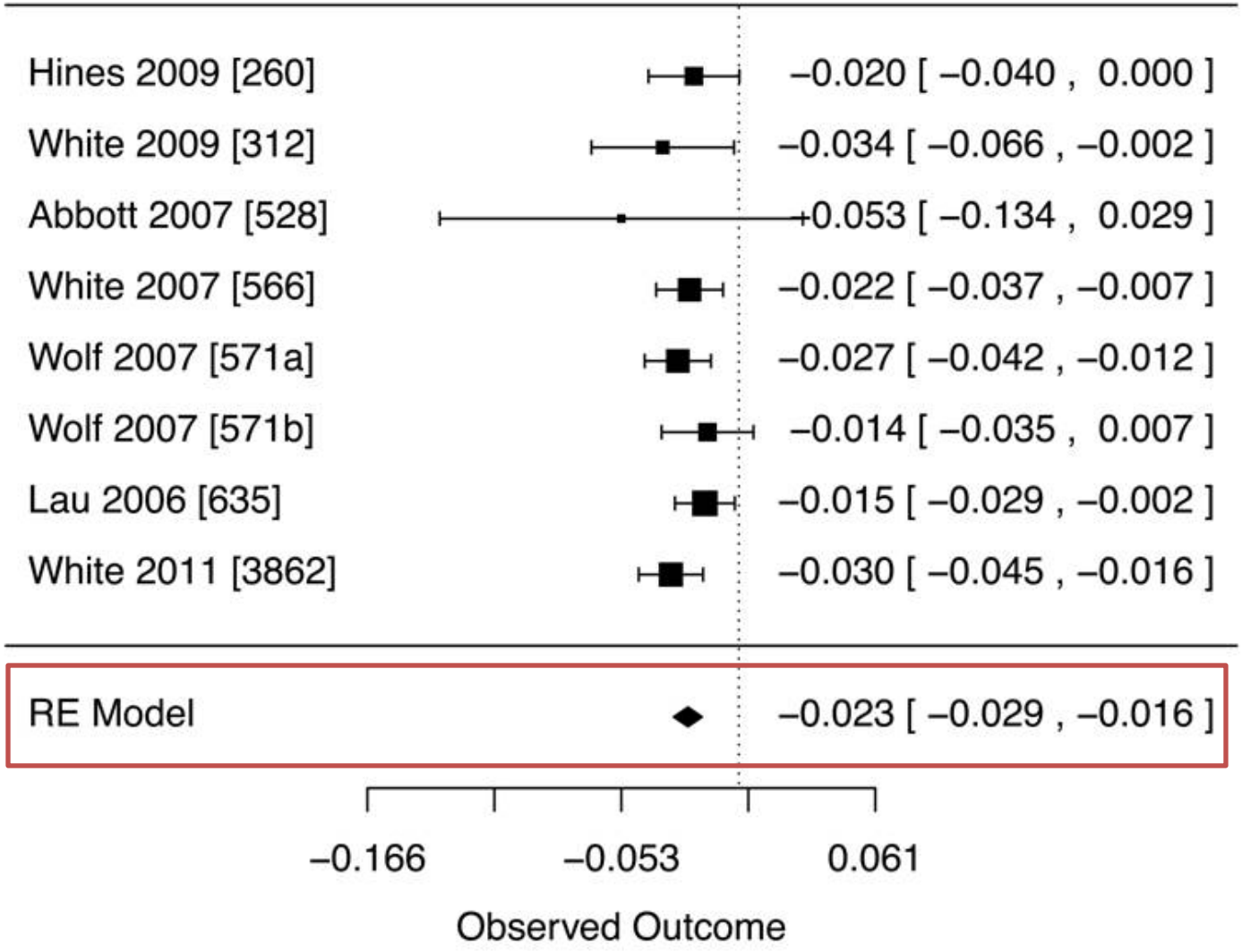
#Wolf study contributed two data sets—"a" exposed one group of animals from GD1-17 and "b" exposed a different group during a varied subset of days between GD1-17

Subset of studies for meta-analysis

Comparability across studies determined based on study characteristics:

- Animal model used:
  Mouse
- Developmental stage at measurement:
  Birth
- Outcome reported:
  Weight
- PFOA exposure:
  Oral Gavage (similar dose, frequency, timing, and duration)

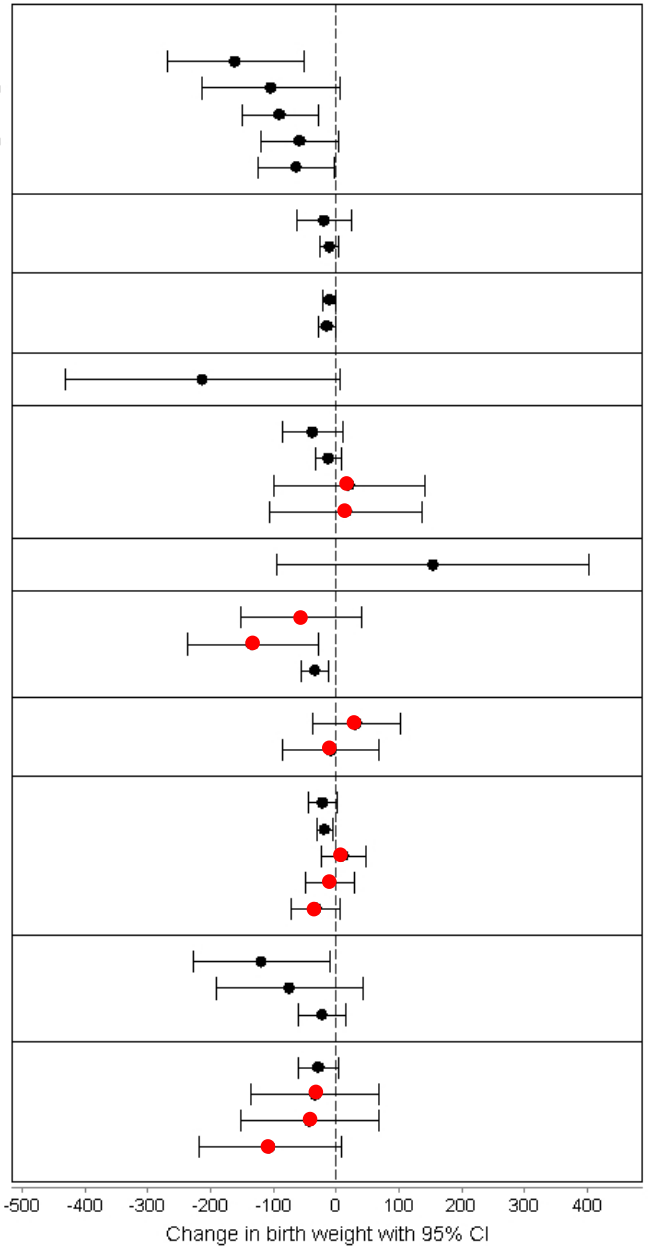
Meta-analysis results: Decrease in birth weight with increase in PFOA exposure



←Estimates a **0.023g decrease** in birthweight for every mg/kg/day increase in PFOA exposure

Summary of All Studies with Continuous Outcome of Birth Weight

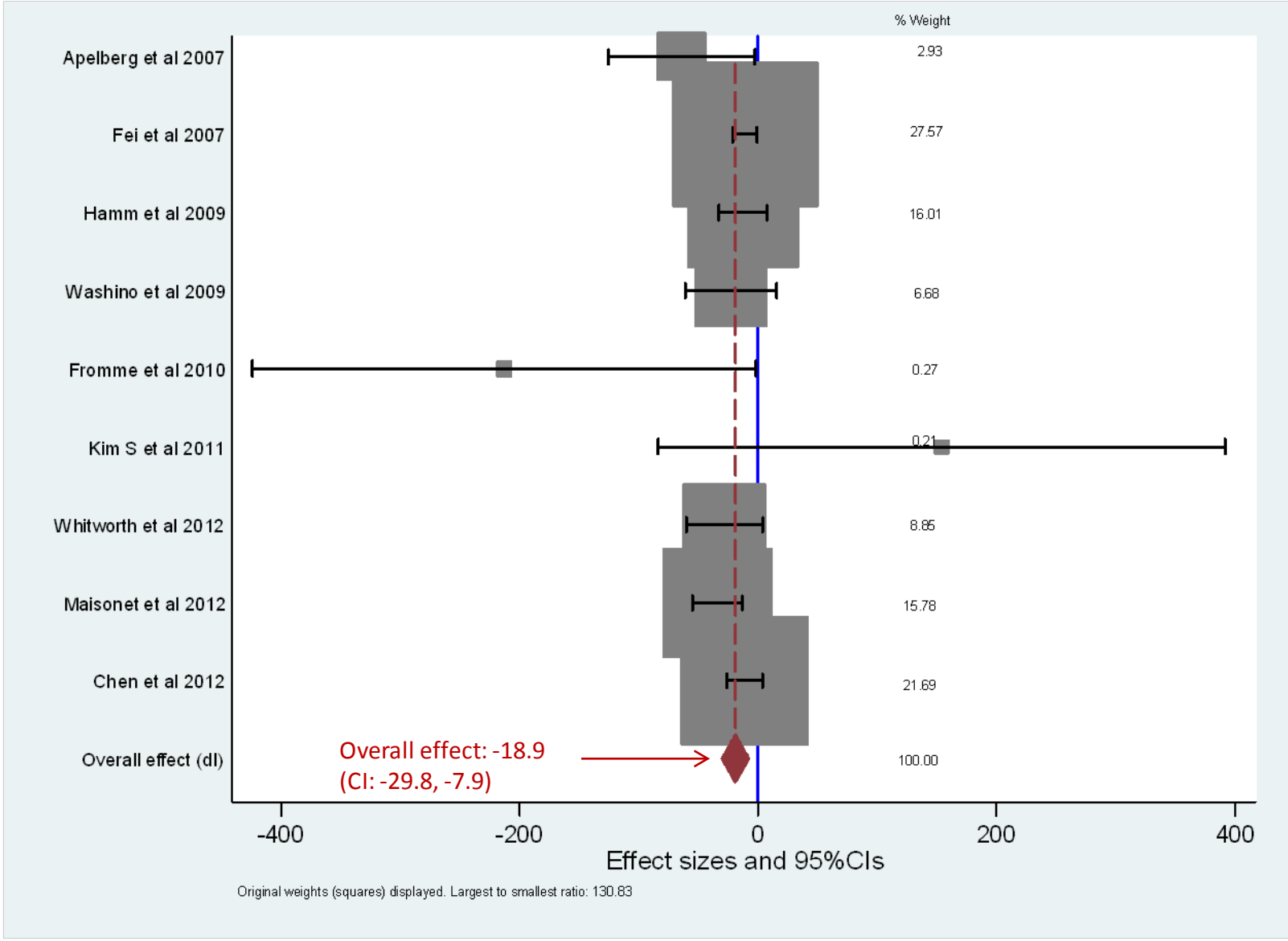
Study	PFOA increase	PFOA range (ng/mL)	Covariates
Apelberg et al 2007	ln ng/mL	0.3-7.1	ga
Apelberg et al 2007	ln ng/mL	0.3-7.1	ga, ma, bmi, race, par, smk, sex, ht, wtg, dia, hyp
Apelberg et al 2007	25th to 75th percentile	1.2-2.1	ga
Apelberg et al 2007	25th to 75th percentile	1.2-2.1	ga, ma, bmi, race, par, smk, sex, ht, wtg, dia, hyp
Apelberg et al 2007*	ng/mL	0.3-7.1	ga, ma
Chen et al 2012	ln ng/mL	geomean(stdev)=1.84(2.23)	ga, ma, bmi, par, cot, sex, edu, delmode
Chen et al 2012*	ng/mL	geomean(stdev)=1.84(2.23)	ga, ma
Fei et al 2007*	ng/mL	<LLOQ - 41.5	ga, ma, bmi, par, smk, sex, SES, gabd
Fei et al 2007	ng/mL	<LLOQ - 41.5	ga, ma, bmi, par, smk, sex, SES, gabd, PFOS
Fromme et al 2010*	ng/mL	0.50-4.20	none
Hamm et al 2010	ln ng/mL	<LOD - 18	ga, ma, race, grav, mwt, matht, smk, sex
Hamm et al 2010*	ng/mL	<LOD - 18	ga, ma, race, grav, mwt, matht, smk, sex
Hamm et al 2010	1st to 2nd tertile (ng/mL)	<LOD - <1.1 to 1.1-2.1	ga, ma, race, grav, mwt, matht, smk, sex
Hamm et al 2010	1st to 3rd tertile (ng/mL)	<LOD - <1.1 to >2.1 - 18	ga, ma, race, grav, mwt, matht, smk, sex
Kim S et al 2012*	ng/mL	0.4-3.23	ga, ma, par
Maisonet et al 2012	1st tertile to 2nd tertile	<3.1 to 3.1-4.4	ga, bmi, par, smk
Maisonet et al 2012	1st tertile to 3rd tertile	<3.1 to >4.4	ga, bmi, par, smk
Maisonet et al 2012*	ng/mL	1.0-16.4	ga, bmi, par, smk
Nolan et al 2009	low to mid exposure	na	ga, ga2, ga3, ma, race, sex, SES
Nolan et al 2009	low to high exposure	na	ga, ga2, ga3, ma, race, sex, SES
Savitz et al 2012 study II-b	25th to 75th IQR (lnPFOA)	1.92	ga, ma, par, edu, smk, exposyr, state
Savitz et al 2012 study II-b	100 ng/ml PFOA	100 ng/mL	ga, ma, par, edu, smk, exposyr, state
Savitz et al 2012 study II-b	1st/2nd quintile to 3rd quintile	3.9 - <8.9 to 8.9 - <19.6	ga, ma, par, edu, smk, exposyr, state
Savitz et al 2012 study II-b	1st/2nd quintile to 4th quintile	3.9 - <8.9 to 19.6 - 53.1	ga, ma, par, edu, smk, exposyr, state
Savitz et al 2012 study II-b	1st/2nd quintile to 5th quintile	3.9 - <8.9 to 53.1 -1897.0	ga, ma, par, edu, smk, exposyr, state
Washino et al 2009	log10PFOA	ND - 5.3	ga
Washino et al 2009	log10PFOA	ND - 5.3	ga, ma, bmi, race, par, smk, sex, edu, bsp
Washino et al 2009*	ng/mL	ND - 5.3	ma, ga
Whitworth et al 2012*	ng/mL	median(IQR)=2.2(1.6-3.0)	ga, ma, bmi, par
Whitworth et al 2012	first to second quartile	<1.65 to 1.65 - 2.24	ga, ma, bmi, par
Whitworth et al 2012	first to third quartile	<1.65 to 2.25 - 3.03	ga, ma, bmi, par
Whitworth et al 2012	first to fourth quartile	<1.65 to >3.03	ga, ma, bmi, par



* Estimate included in meta-analysis
● Data can be used to evaluate dose-response

ga=gestational age; ma=maternal age; bmi=body mass index; par=parity; smk=smoking status; sex=infant gender; ht=maternal height; wtg=maternal weight gain during pregnancy; dia=diabetes; hyp=hypertension; cot=serum cotinine; edu=maternal education level; delmode= delivery mode; SES=socioeconomic status; gabd=gestational age at blood draw; PFOS=serum perfluorooctane sulfonic acid; grav=gravidity; mwt=maternal prepregnancy weight; exposyr=year of exposure estimate; state=state of residence; bsp=blood sampling period

Meta-analysis for Birth Weight (n=9 studies)

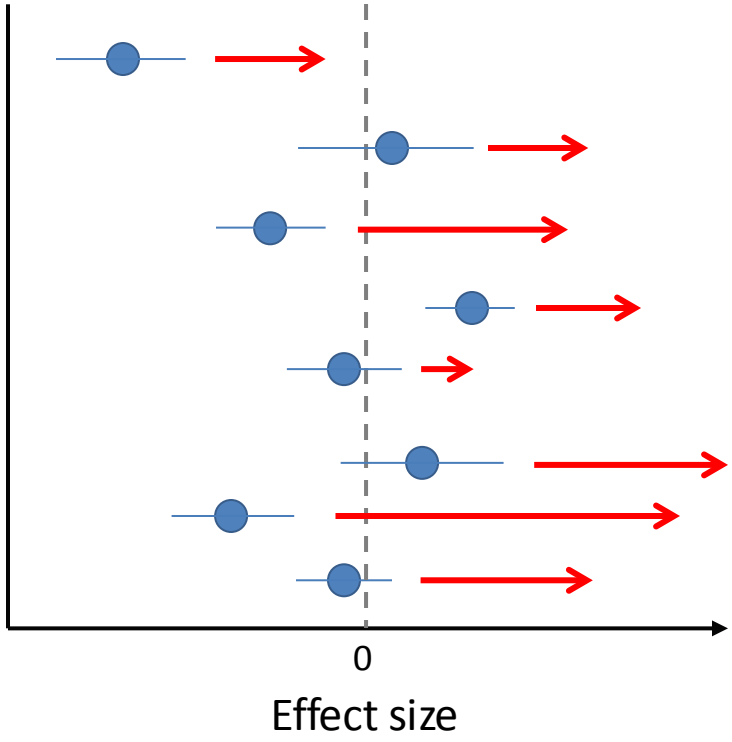


Effect: 18.9 gram reduction in birth weight per ng/mL serum PFOA increase

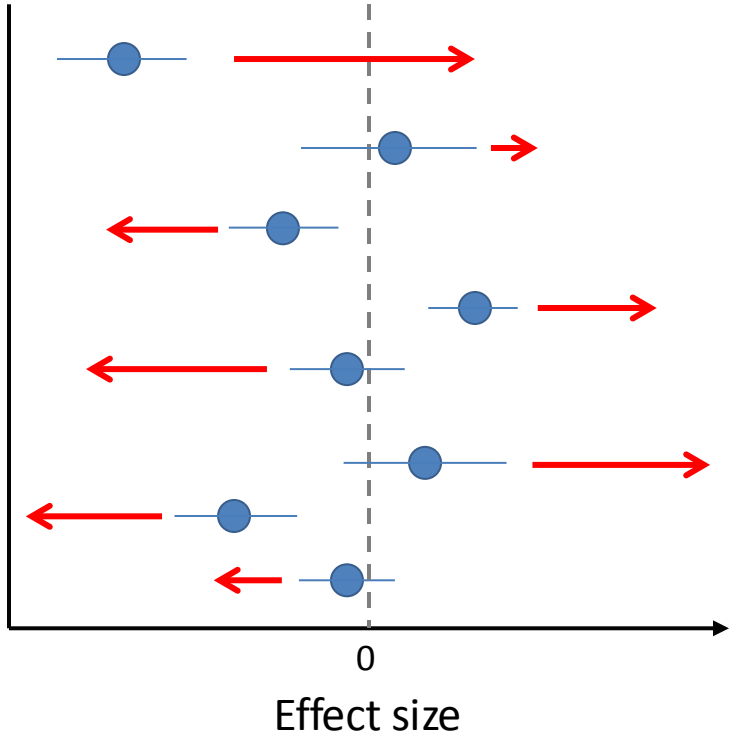
Results Step 5:
Rate the Quality and Strength of the Evidence

Risk of Bias vs Random Error

Bias

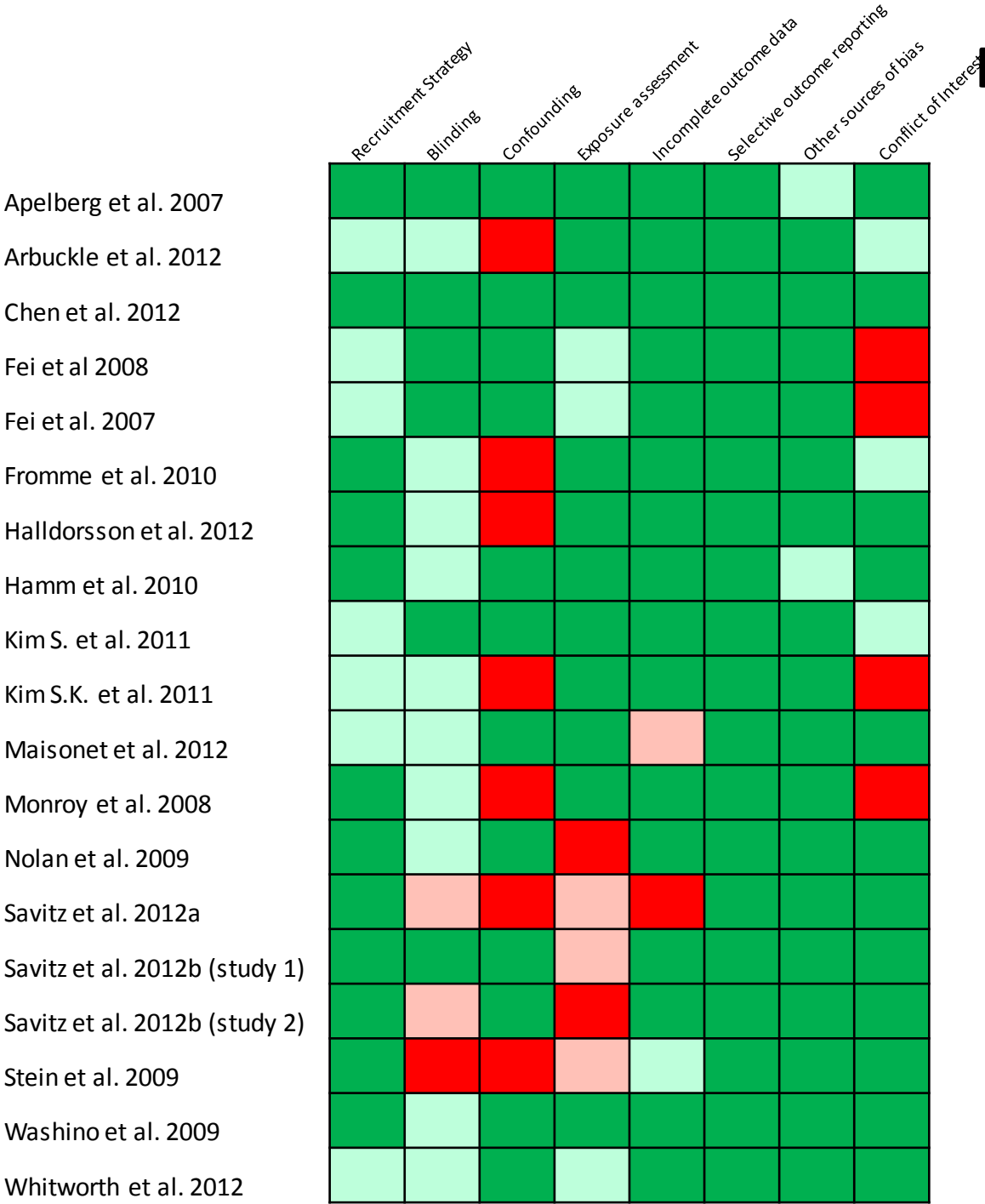


Random Error



Risk of Bias: Methodological characteristics of a study that can introduce a systematic error in the magnitude or direction of the results (Higgins and Green 2008).

Results: Risk of Bias Human Evidence N=19



For individual studies (N=19)

- Low risk
- Probably low risk
- Probably high risk
- High risk
- N/A

Rating Quality of Human Evidence

Moderate Quality

Human Evidence

Risk of Bias

Risk of bias is determined for ***each individual study***.

Domains

- Recruitment strategy
- Blinding
- Exposure assessment
- Confounding
- Incomplete outcome data
- Selective reporting
- Conflict of interest
- Other bias

Determinations
(for each risk of bias domain)

- Low risk
- Probably low risk
- Probably high risk
- High risk

Quality of Evidence

Quality is rated ***across all studies***. Human evidence begins as 'moderate quality' and may be downgraded (-1 or -2) or upgraded (+1 or +2) according to criteria.

Downgrade Criteria

- **Risk of bias across studies**
- Indirectness
- Inconsistency
- Imprecision
- Publication bias

Upgrade Criteria

- Large magnitude of effect
- Dose response
- All possible confounding would confirm negative result

Rating
(based on all quality criteria)

- High quality
- Moderate quality
- Low quality

Strength of Evidence

Strength is rated ***across all studies***. The final ratings represent the level of certainty of toxicity.

Considerations

- **Quality of body of evidence**
- Direction of effect
- Confidence in effect
- Other compelling attributes of the data that may influence certainty

Rating
(based on all strength considerations)

- Sufficient evidence
- Limited evidence
- Inadequate evidence
- Evidence of lack of toxicity

Factors that DECREASE Quality

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels.

1. **RISK OF BIAS**
   Study limitations - substantial risk of bias across most of body of evidence to downgrade

2. **INDIRECTNESS**
   Evidence was not directly comparable to the question of interest(i.e., population, exposure, comparator, outcome)

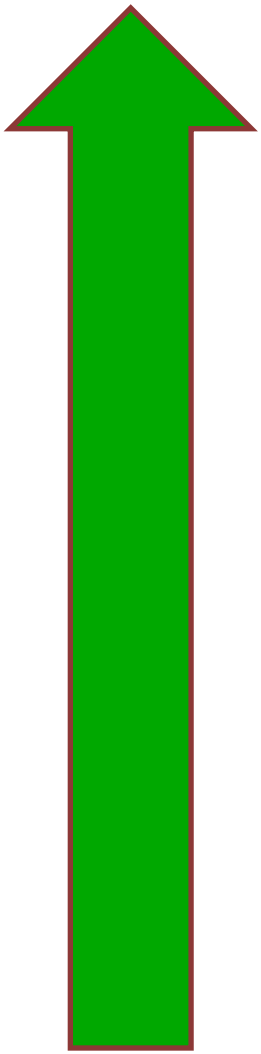
3. **INCONSISTENCY**
   Widely different estimates of effect (heterogeneity or variability in results)

4. **IMPRECISION**
   Studies had few participants and few events (wide confidence intervals)

5. **PUBLICATION BIAS**
   Studies missing from body of evidence, resulting in an *underestimate* of true effects from exposure

Factors that **INCREASE** Quality (Human only)

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels.



1. **LARGE MAGNITUDE OF EFFECT**
   Associations with relative risk greater than 2

2. **DOSE RESPONSE**
   Consistent dose response gradient in one or multiple studies, and/or dose response across studies

3. **CONFOUNDING MINIMIZES EFFECT**
   All possible residual confounders or biases would reduce demonstrated effect

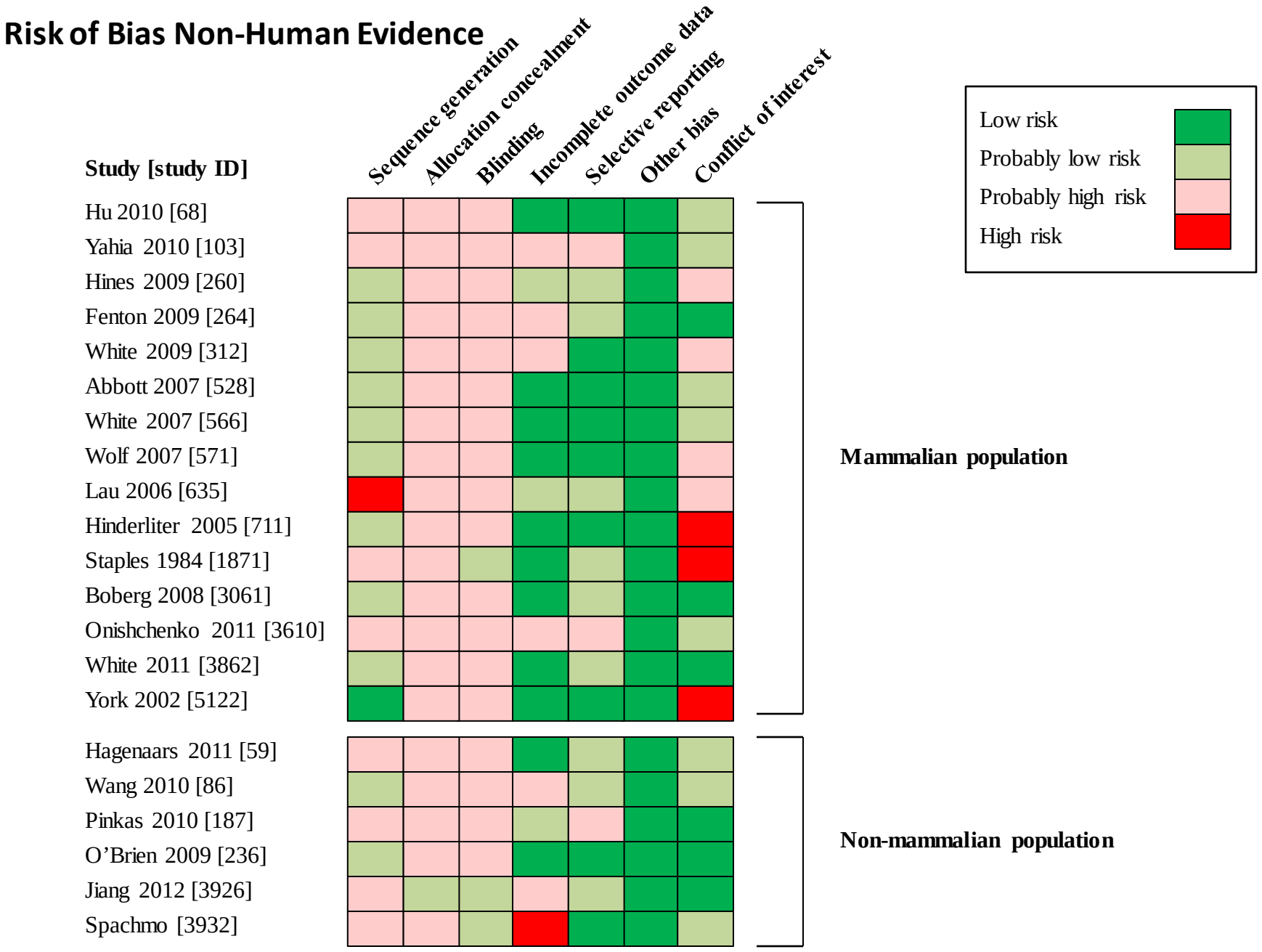
Quality of Human Evidence




0 Downgrade, 0 Upgrade = Moderate Quality

	Downgrade					Upgrade		
	Risk of Bias	Indirectness	Inconsistency	Imprecision	Publication Bias	Large Magnitude of Effect	Dose Response	Confounding Minimizes Effect
Final	0	0	0	0	0	0	0	0

Risk of Bias Non-Human Evidence



Step 3. Rate the Quality and Strength of the Evidence

Animal Evidence

Separate for Mammalian and Non-mammalian Populations

Risk of Bias

Risk of bias is determined for ***each individual study***.

Domains

- Sequence generation
- Allocation concealment
- Blinding
- Incomplete outcome data
- Selective reporting
- Conflict of interest
- Other bias

Determinations
(for each risk of bias domain)

- Low risk
- Probably low risk
- Probably high risk
- High risk

Quality of Evidence

Quality is rated ***across all studies***. Animal evidence begins as 'high quality' and may be downgraded (-1 or -2) according to criteria.

Criteria

- **Risk of bias across studies**
- Indirectness
- Inconsistency
- Imprecision
- Publication bias

Rating
(based on all quality criteria)

- High quality
- Moderate quality
- Low quality

Strength of Evidence

Strength is rated ***across all studies***. The final ratings represent the level of certainty of toxicity.

Considerations

- **Quality of body of evidence**
- Direction of effect
- Confidence in effect
- Other compelling attributes of the data that may influence certainty

Rating
(based on all strength considerations)

- Sufficient evidence
- Limited evidence
- Inadequate evidence
- Evidence of lack of toxicity

Rating Quality of Non-Human Experimental Studies

High Quality

Factors that DECREASE Quality

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels.

1. **RISK OF BIAS**
   Study limitations - substantial risk of bias across most of body of evidence to downgrade

2. **INDIRECTNESS**
   Evidence was not directly comparable to the question of interest(i.e., population, exposure, comparator, outcome)

3. **INCONSISTENCY**
   Widely different estimates of effect (heterogeneity or variability in results)

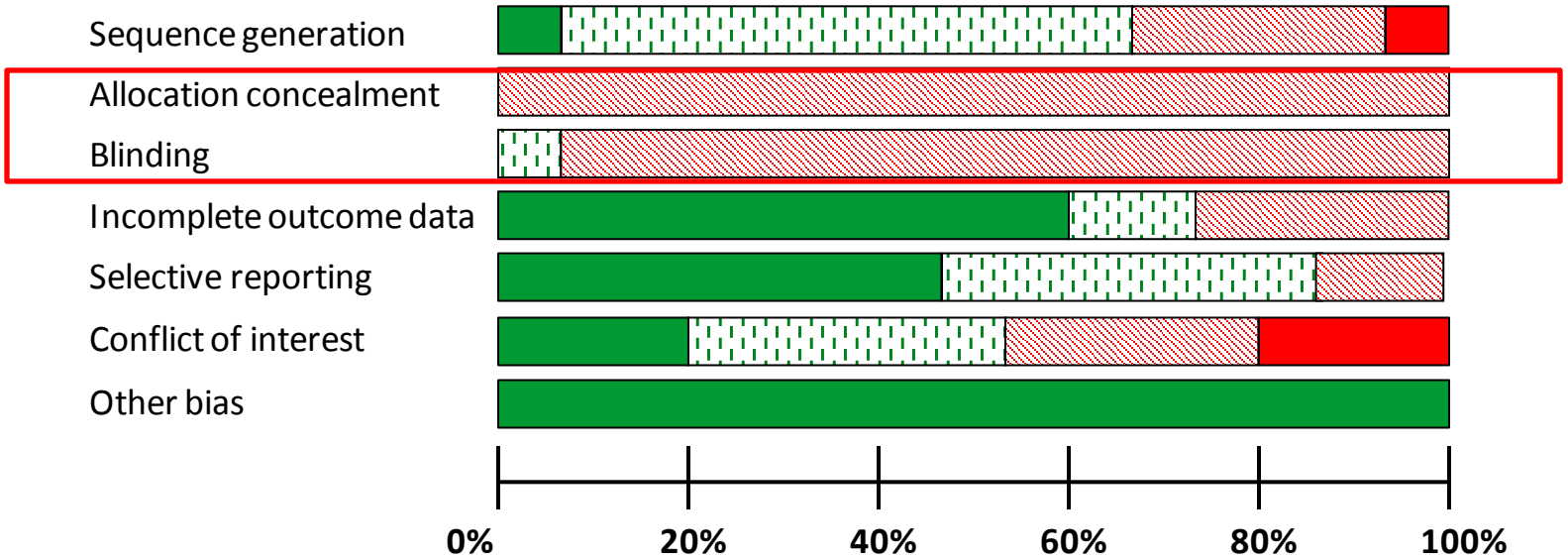
4. **IMPRECISION**
   Studies had few participants and few events (wide confidence intervals)

5. **PUBLICATION BIAS**
   Studies missing from body of evidence, resulting in an *underestimate* of true effects from exposure

Quality of Mammalian Evidence

-1 Downgrade = Moderate Quality

	Downgrade				
	Risk of Bias	Indirectness	Inconsistency	Imprecision	Publication Bias
Final	-1	0	0	0	0



Results Step 6:
Moving From Quality of Evidence to Strength of Evidence

Summary of Factors Considered

Risk of Bias of Individual Studies

- Recruitment strategy
- Blinding
- Confounding
- Exposure assessment
- Incomplete outcome
- Selective outcome
- Other ROB
- Confounding

Quality of Evidence

- ROB (overall studies)
- Indirectness of evidence
- Inconsistency
- Imprecision
- Publication bias
- Magnitude
- Residual Confounders
- Dose Response

- Quality Rating
- Direction of Effect
- Confidence in Effect
- Other compelling factors*

Strength of Evidence
Human Evidence = "Sufficient"

CRITERIA:

1. Quality of evidence: Moderate
2. What is the direction of effect? Decrease in fetal growth with PFOA exposure
3. What is the confidence in the effect? A new study would be unlikely to change the certainty in the direction of the effect
4. Are there other compelling attributes of the data that influence certainty?

Sufficient evidence of toxicity

The available evidence includes consistent results from well-designed, well-conducted studies and the conclusions are unlikely to be strongly affected by the results of future studies. A positive relationship was observed between exposure and outcome where chance, bias and confounding can be ruled out with reasonable confidence.

Strength of Evidence

Non-Human Mammalian Evidence = "Sufficient"

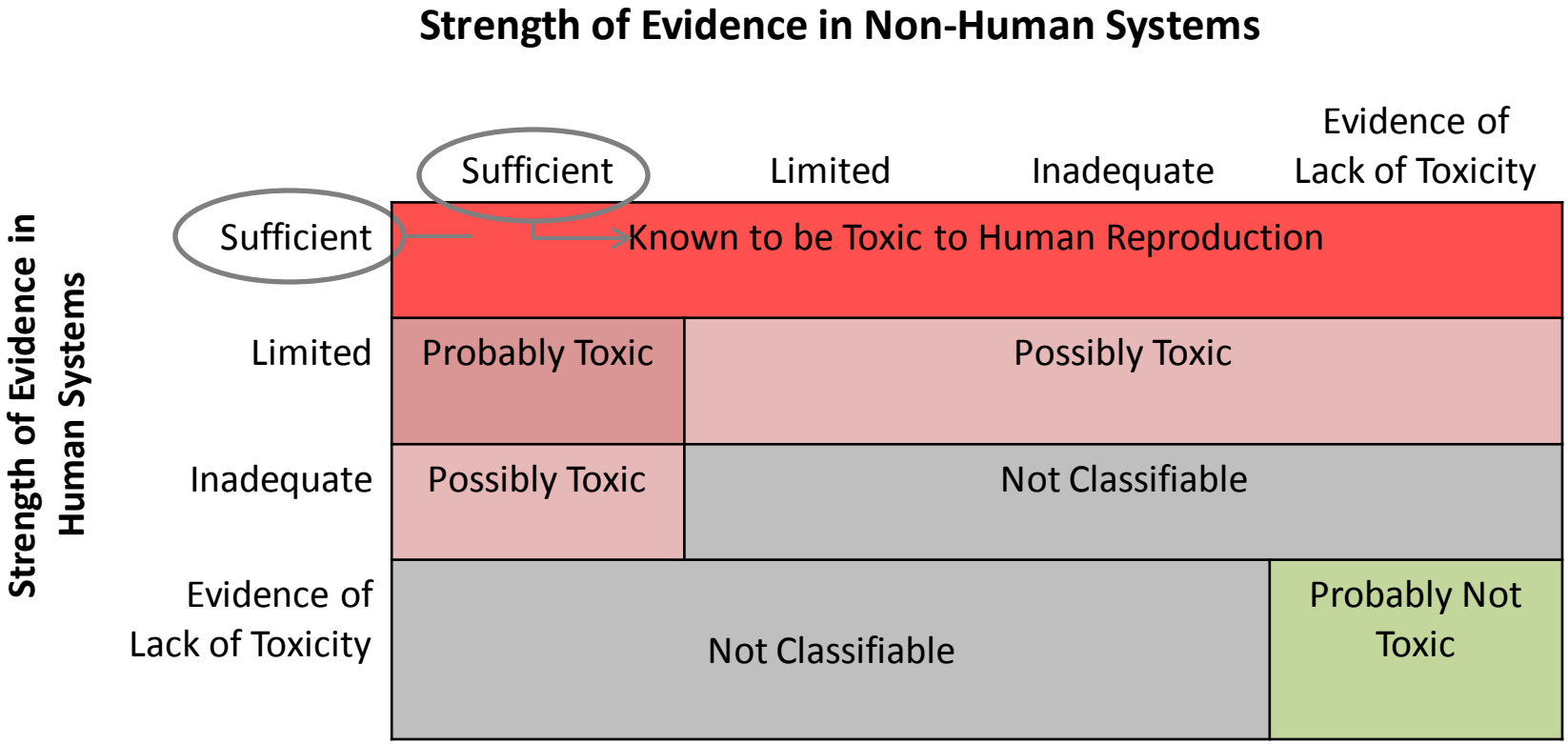
CRITERIA:

1. Quality of evidence: Moderate
2. What is the direction of effect? Decrease in fetal growth with PFOA exposure
3. What is the confidence in the effect? A new study would be unlikely to change the certainty in the direction of the effect
4. Are there other compelling attributes of the data that influence certainty?

Sufficient evidence of toxicity

Positive association has been established through multiple positive results or a single appropriate study in a single species.

Integrating the Streams of Evidence



***Conclusion*: Human exposure to PFOA is known to be toxic to human reproduction and development based on sufficient evidence of decreased fetal growth in both human and non-human mammalian species.**

Comparison of Narrative reviews and Navigation Guide/OHAT Approach

Reference	Specify study question	Specify inclusion/exclusion criteria	Conduct reproducible search	Assess Risk of Bias	Data analysis and/or meta-analyses	Summary of findings table (Health Effects)	Assess quality of body of evidence	Integrate evidence streams
Post et al 2012	Yes	No	No	No	No	Yes	No	No
Lindstrom et al 2011	Yes	No	No	No	No	No	No	No
Stahl et al 2011	Yes	No	No	No	No	Yes	No	No
White et al 2011	Yes	No	No	No	No	Yes	No	No
Steenland et al 2010	Yes	No	No	No	No	No	No	No
DeWitt et al 2009	Yes	No	No	No	No	No	No	No
Olsen et al 2009	Yes	No	No	No	No	Yes	No	No
Jensen and Leffers 2008	No	No	No	No	No	No	No	No
Lau et al 2007	Yes	No	No	No	No	No	No	No
Butenhoff et al 2004	Yes	Yes	No	No	Some data analysis (MOE, $LBMIC_{10}$)	Yes	No	No
Kennedy et al 2004	Yes	No	No	No	No	Yes	No	No
Lau et al 2004	Yes	No	Limited discussion of literature search	No	No	No	No	No
Hekster et al 2003	Yes	Some inclusion criteria described in cited report by same authors	Limited discussion of literature search	No	No	Yes	No	No
Kudo and Kawashima 2003	Yes	No	No	No	No	No	No	No
Navigation & OHAT/NTP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Strengths

- Permits action on available data
- Systematic and transparent
- Based on empirically-proven methods
- Capacity to evolve with change in evidence streams
- Can identify evidence gaps for future work
- Can support identification of safer alternatives
- Separates science from values and preferences

Limitations

- Analysis limited to available data
- Not every criterion developed a priori – some aspects of method developed simultaneously
- Novel parts of methodology need validation
- Further definition of moving from quality of evidence to strength of evidence
- Does not address non-scientific barriers to prevention-oriented action
- Need step 4

Future Directions

Step 4. Rate Strength of Recommendations

4. Grade Strength of Recommendation:

Strength of Evidence (from Step 3 above)

Exposure	Known to be Toxic	Probably Toxic	Possibly Toxic	Not Classifiable
High [1]	S	S	S	S
Medium [2]	S	S	D	D
Lower [3]	D	D	D	D

Is a Less Toxic Alternative Available?

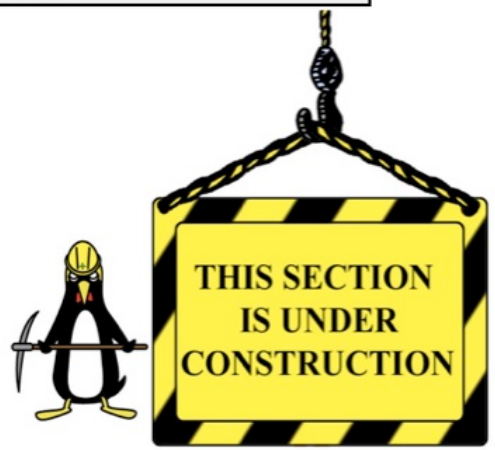
Patient Values and Preferences

Strong or Discretionary Recommendation

S = Strong Recommendation
- denotes "we recommend"

D = Discretionary Recommendation
- denotes "we suggest"

1. High Exposure =
   - Exposure at any level that occurs during critical or sensitive windows of development or during other periods of heightened vulnerability (i.e., nutritional deficiencies, chronic diesease/ immunosuppressed state, etc.);
   - Exposure at high level for any duration;
   - Exposure of moderate or low level for long (chronic) duration
2. Medium Exposure =
   - Exposure at moderate level for short or intermittent duration
3. Lower Exposure =
   - Exposure at low level for short or intermittent duration



Methodological Needs

- Criteria for moving from quality to strength of evidence
- Methods to include all potential types of evidence, i.e., assessing chickens, flies and *in vitro* data
- Improved methods of animal toxicity testing – high ROB may be prevalent for key domains
- Mechanistic data is considered under other considerations…. Further development needed
- Consider the nature and extent of consensus that is needed for a decision

Conclusion

- We can do it now!
- Comparable to OHAT/NTP approach
- Rigorous, systematic, transparent and doable
- Capacity to evolve with changes in evidence stream

Acknowledgements

Navigation Guide Funders (since 2009)

- USEPA
- US Environmental Protection Agency STAR (RD83467801)
- National Institute for Environmental Health Sciences (ES018135)
- Cal-EPA
- Fred Gellert Foundation
- Clarence Heller Foundation
- New York Community Trust
- Forsythia Foundation
- Passport Foundation
- Johnson Family Foundation
- Heinz Endowments
- Rose Foundation
- Kaiser Permanente
- Planned Parenthood Federation of America
- UCSF Phillip R Lee Institute for Health Policy Studies

Authors

Dylan Atchley
UCSF
Research Assistant

Paula Johnson
UCSF
Post-doctoral Fellow

Saunak Sen
UCSF
Associate Professor

Patrice Sutton
UCSF
Research Scientist

Tracey Woodruff
UCSF
Professor,
Director, PRHE

Daniel Axelrad
EPA
Environmental
Scientist

Erica Koustas
EPA
ORISE Post-doctoral
Fellow

Juleen Lam
EPA
ORISE Post-doctoral
Fellow

Karen Robinson
Johns Hopkins
Associate Professor

Thank You

UCSF
University of California
San Francisco

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

THE JOHNS HOPKINS UNIVERSITY BALTIMORE 1876
VERITAS VOS LIBERABIT

Application of Systematic Review Frameworks to Environmental Health

Colleen Lanier-Christensen

Research Fellow, Environmental Defense Fund

MPH Candidate, Columbia University Mailman School of Pubic Health

EDF
ENVIRONMENTAL DEFENSE FUND®
Finding the ways that work

Overview

- Background and goals
- NRC guidance/available frameworks
- Key components of systematic review and evidence integration
- Role of mechanistic data
- Priorities and next steps

IRIS program and systematic review

- Goal: High-quality, transparent, and timely scientific assessments based on available evidence
- How we get there: adopting transparent, objective, empirically validated systematic review methods

NRC recommendations

- Empirically based approaches are available

  "…models are available that have proved successful in practice. They have several common elements: transparent and explicitly documented methods, consistent and critical evaluation of all relevant literature, application of a standardized approach for grading the strength of evidence, and clear and consistent summative language."*

*NRC, Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde, 2011, 155.

Key elements of systematic review for multiple evidence streams

- Developing a protocol *a priori*
- Transparent and consistent method of data extraction/collection
  - Standard and clear procedures for missing data
- Criteria for assessing study risk of bias:
  - Must address internal validity – whether studies tell you something meaningful about health effects
  - Must be appropriate for specific evidence stream (e.g., animal, human, mechanistic, etc.)
  - Should be empirically tested to understand the impact on biasing effect estimates
- Characterization of quality and strength of evidence
- Guidance for integration of evidence

Mechanistic data

- Mechanistic data is rapidly becoming readily available
- At this time, significant limitations exist:
  - Lack of full knowledge of mechanism(s) of action (e.g., benzene, arsenic)
  - Presumption of a single or set of mechanisms:
    - Could exclude valuable, high-quality studies that illustrate less understood mechanisms
    - Inappropriately simplifies complex biological processes (multiple mechanisms may be involved)
- Given these limitations, mechanistic understanding
  - Should not be required for IRIS assessments
  - Should not serve as organizing framework for systematic review

A scientifically grounded approach to integrating evidence

- Should assume default that animal effects are relevant to humans, lacking sufficient evidence otherwise
  - Consistent with NRC recommendation in considering uncertainties
  - Basic principle of US EPA cancer risk assessment that site concordance across species is not required in hazard evaluation
- Data streams can and should be considered complementary

Priorities moving forward

- Importance of all evidence streams
  - Development and evaluation of tools to evaluate internal validity of animal and mechanistic studies
- Empirical evaluation of study elements
  - Criteria unique to each type of evidence: human, animal, and mechanistic
  - Criteria evaluated in human studies and warrant consideration in others:
    - Conflict of interest
    - Selective reporting

Next steps

- Leverage existing efforts to protect public health
  - Significant work has been done; we need to build on these existing, evaluated frameworks
  - Delays in scientifically sound IRIS assessments have real world consequences
- Keep the process moving based on available frameworks and evidence

Thank you!

Colleen Lanier-Christensen
ctc2129@columbia.edu
Research Fellow, Environmental Defense Fund
MPH Candidate, Columbia University Mailman School of Pubic Health

EDF
ENVIRONMENTAL DEFENSE FUND®
Finding the ways that work

FRAMEWORKS FOR SYNTHESIZING AND INTEGRATING EVIDENCE

Panel Discussion

Frameworks for Synthesizing and Integrating Evidence

1. Some frameworks consider human data and animal data jointly and some frameworks consider human data and animal data independently, and then integrate these results at the end. In what types of circumstance/scenario (e.g., type of data available, or primary study question), if any, would one approach be preferred?

Frameworks for Synthesizing and Integrating Evidence

2. The type of evidence available varies for different pollutants. How does the lack or uneven strength of one line of evidence (e.g., human data, mechanistic understanding) impact the weight of evidence and the ability to draw causal conclusions and evaluate hazard and dose-response relationships?

Frameworks for Synthesizing and Integrating Evidence

3. The availability of mode of action data can vary across chemicals. Where is the appropriate place in a framework for incorporating mode of action information?
4. How do you allow for flexibility and scientific judgment in developing a framework for integration? What aspects of a framework can be established a priori? What aspects will depend on the data and scenario/questions?