



NTP

National Toxicology Program

Draft OHAT Approach Part 1 Preparing the Topic Through Assessing the Quality of Individual Studies

Abee Boyles, Ph.D.

Division of the National Toxicology Program,
Office of Health Assessment and Translation (OHAT)

Web-Based Informational Meeting
April 23, 2013 12:00 - 4:00PM EDT



Step 1. Prepare the Topic

- **Scope and focus the topic** to answer specific questions
 - Consult with appropriate experts to focus objective and questions
- **Develop draft protocol** through iterative process to detail project-specific procedures used throughout the evaluation
 - Literature search, selection of relevant studies
 - Determination of primary, secondary, and grouping of outcomes
 - Data extraction, assessment of risk of bias, evaluation of confidence
- **Obtain input on draft protocol** from experts (technical advisors, BSC) and solicit public comment
 - Draft protocols illustrate the application of this framework
 - Protocols available at: <http://ntp.niehs.nih.gov/go/38673>
- **Document** project-specific procedures before proceeding with the evaluation



Importance of the Protocol

- **The protocol contains enough details so that the process and the procedures could be reconstructed**
 - For example
 - The literature search strategy is presented in enough detail so that it could be replicated
 - State which outcomes are considered primary and secondary
 - Criteria for assessing individual study quality are established stating what defines high and low risk of bias for each question on study design or performance
- **Revisions to the protocol**
 - It is recognized that valid reasons for modifying a protocol may occur during an evaluation
 - Revisions are permitted and they are documented and justified

Primary and Secondary Outcomes

Example – PFOA/PFOS Exposure and Immunotoxicity

Humans	Animals*	<i>In vitro</i> Assays
<i>Primary outcomes</i>	<i>Primary outcomes</i>	<i>Primary outcomes</i>
Immune-related diseases and measures of immune function <ul style="list-style-type: none">• <i>Immunosuppression</i>• <i>Sensitization and allergic response</i>• <i>Autoimmunity</i>	Disease resistance assay or measures of immune function <ul style="list-style-type: none">• <i>Disease resistance assays</i>• <i>Immune function assays following <u>in vivo</u> exposure to the test substance</i>	<i>Immune function assays following <u>in vitro</u> exposure to the test substance</i>
<i>Secondary outcomes</i>	<i>Secondary outcomes</i>	<i>Secondary outcomes</i>
<i>Immunostimulation**</i> <i>Observational immune endpoints</i>	<i>Observational immune endpoints following <u>in vivo</u> exposure to test substance</i>	<i>Observational immune endpoints following <u>in vitro</u> exposure to the test substance</i>

Primary and Secondary Outcomes

Example – PFOA/PFOS Exposure and Immunotoxicity

More detail and examples provided in the protocol


Humans	Animals*	<i>In vitro</i> Assays
<i>Primary outcomes</i>	<i>Primary outcomes</i>	<i>Primary outcomes</i>
<p>Immune-related diseases and measures of immune function</p> <p><i>Immunosuppression</i> (e.g., otitis, infections, or decreased vaccine antibody response);</p> <p><i>Sensitization and allergic response</i> (e.g., atopic dermatitis or asthma);</p> <p><i>Autoimmunity</i> (e.g., thyroiditis or systemic lupus erythematosus)</p>	<p>Disease resistance assay or measures of immune function</p> <p><i>Disease resistance assays</i> (e.g., host resistance to influenza A or trichinella, changes in incidence or progression in animal models of autoimmune disease)</p> <p><i>Immune function assays following in vivo exposure to the test substance</i> (e.g., antibody response [T-cell dependent IgM antibody response (TDAR)], natural killer cell [NK] activity, delayed-type hypersensitivity [DTH] response, phagocytosis by monocytes, local lymph-node assay [LLNA])</p>	<p><i>Immune function assays following in vitro exposure to the test substance</i> (e.g., natural killer cell [NK] activity, phagocytosis or bacterial killing by monocytes, proliferation following anti-CD3 antibody stimulation of spleen cells or lymphocytes)</p>
<i>Secondary outcomes</i>	<i>Secondary outcomes</i>	<i>Secondary outcomes</i>
<p><i>Immunostimulation**</i> (e.g., unintended stimulation of humoral immune function)</p> <p><i>Observational immune endpoints</i> (e.g., lymphocyte counts, lymphocyte proliferation, cytokine levels, serum antibody levels, or serum autoantibody levels)</p>	<p><i>Observational immune endpoints</i> (e.g., lymphoid organ weight, lymphocyte counts or subpopulations, lymphocyte proliferation, cytokine production, serum antibody levels, serum or tissue autoantibody levels, or histopathological changes in immune organs)</p>	<p><i>Observational immune endpoints following in vitro exposure to the test substance</i> (e.g., general mitogen-stimulated lymphocyte proliferation, cytokine production)</p>

Primary and Secondary Outcomes

Example – BPA Exposure and Obesity

More detail and examples provided in the protocol

Humans	Animals	Supporting Evidence
<i>Primary outcomes</i> overweight, obesity measures, or measures of adiposity (e.g., BMI, waist circumference, fat composition, skin-fold thickness)	<i>Primary outcomes</i> adiposity (e.g., fat mass, percent fat)	<i>Phenotypic or “apical” outcomes</i> e.g., adipogenic endpoints such as adipocyte number, adipocyte differentiation, or adipocyte lipid accumulation
<i>Secondary outcomes</i> adipokines, ghrelin, leptin, adiponectin, resistin, feeding behavior	<i>Secondary outcomes</i> adipokines, ghrelin, leptin, adiponectin, resistin, feeding behavior, body weight	<i>Pathway and cellular endpoints</i> e.g., <i>ex vivo</i> , cellular, genomic, or mode of action outcomes reported in eligible animal or human studies; cellular, genomic, or mode of action outcomes reported in <i>in vitro</i> studies of adipocytes; interactions with key receptors involved in regulating adipogenesis, e.g., peroxisome proliferator-activated receptors (PPAR), <i>retinoid X receptor (RXR)</i> , <i>liver X receptor (LXR)</i> , or glucocorticoid receptor (GR) in any <i>in vitro</i> model or high throughput screening system



Step 2: Search for and Select Studies for Inclusion

- **Perform comprehensive literature search**

SCOPUS™



PubMed

National Library of Medicine
NLM

- Follow search strategy specified in protocol
 - Including search dates, frequency of updates, etc.
 - Use of unpublished studies – e.g., peer review of critical studies

- **Screen studies for inclusion**

- Two reviewers evaluate each study at the title/abstract level
- Follow procedures defined in protocol to
 - Select relevant studies based on pre-defined criteria
 - Resolve conflicts between reviewers
 - Document reasons for exclusion
 - Complete full-text review



Web-based Reference Management

- DistillerSR[®] Systematic Review Software
- Project management and workflow
- Tracks which studies were included or why excluded

The screenshot shows the DistillerSR web interface. The browser address bar indicates the URL is <https://systematic-review.ca/Reports/ArticlePr>. The page title is "DistillerSR". The main content area displays the project name "Excess Folic Acid" and the user "kris.thayer". A navigation menu includes "Review", "Datarama", "Reports", "References", "Forms", "Manage Levels", "Users", "Project", and "Logout". A table below the menu shows the status of studies at different review levels. Two red arrows point from the bullet points in the list above to the "Included" and "Excluded" columns of the table.

	Unreviewed	Some Reviews	Included	Excluded	Conflict	Fully Reviewed
Level 1 - Title/Abstract Screening	12255	434	2692	7365	228	10285
Level 2 - PDF Screen	1178	165	1205	144	0	1349
Level 3 - Folic Acid Data Extraction Human Studies	96	0	5	0	0	7
Level 3 - Risk of Bias	12	0	5	0	0	5
Level 3 - Meta Analysis	26	0	0	0	0	3

Exclusion Report



DistillerSR

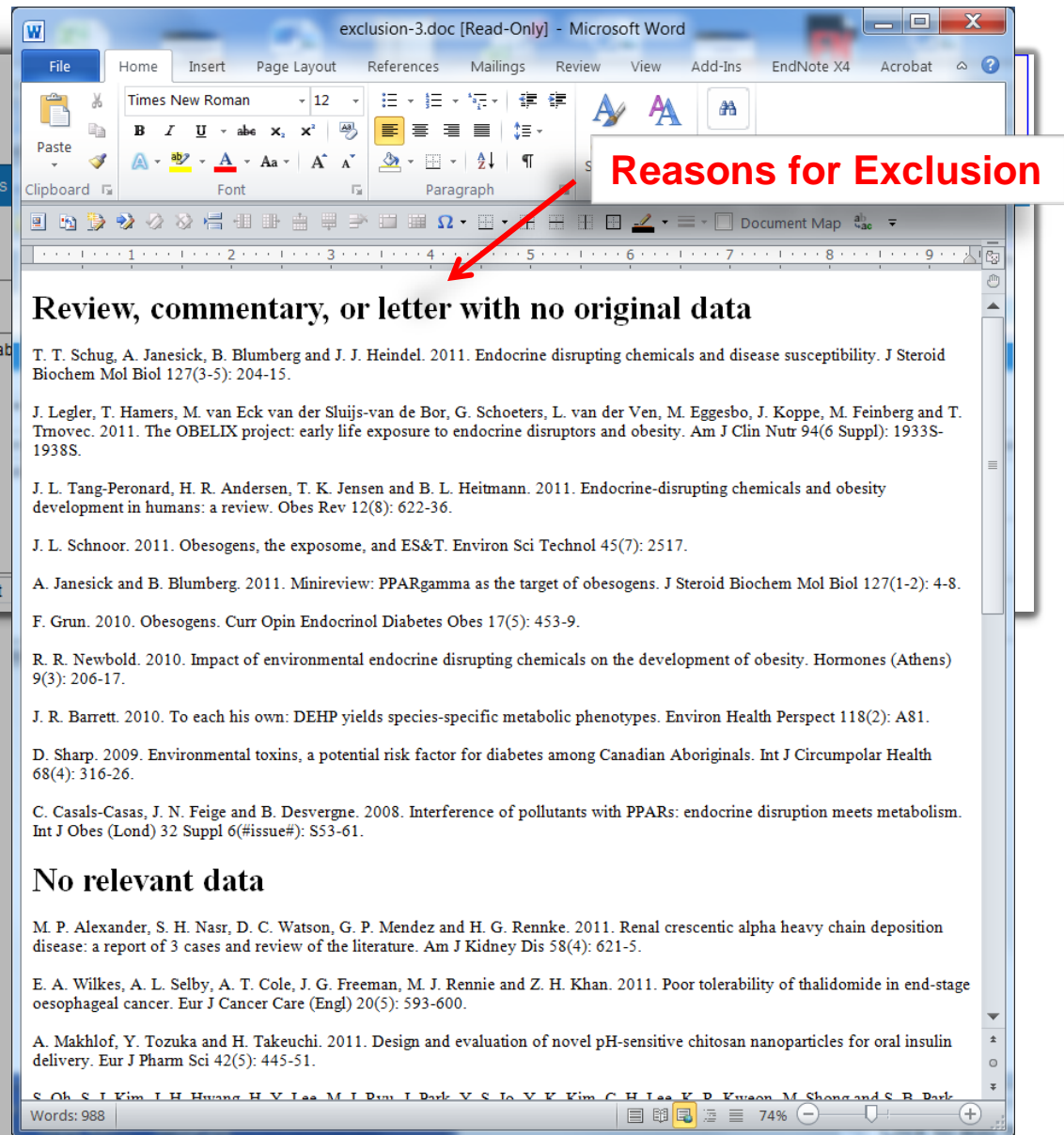
Review | Datarama | Reports | References | Forms

Select a Level: All Levels

Level	Exclusions [?]	Form	Question
Level 1	35	Abst Scre	Do the title or ab conta

Download Exclusion Document with format No Custom Format

Reports menu items: Article Status, Exclusions, User Metrics, Kappa, Statistical, User Workload



exclusion-3.doc [Read-Only] - Microsoft Word

File | Home | Insert | Page Layout | References | Mailings | Review | View | Add-Ins | EndNote X4 | Acrobat

Times New Roman 12

Clipboard | Font | Paragraph

Reasons for Exclusion

Review, commentary, or letter with no original data

T. T. Schug, A. Janesick, B. Blumberg and J. J. Heindel. 2011. Endocrine disrupting chemicals and disease susceptibility. *J Steroid Biochem Mol Biol* 127(3-5): 204-15.

J. Legler, T. Hamers, M. van Eck van der Sluijs-van de Bor, G. Schoeters, L. van der Ven, M. Eggesbo, J. Koppe, M. Feinberg and T. Tmovec. 2011. The OBELIX project: early life exposure to endocrine disruptors and obesity. *Am J Clin Nutr* 94(6 Suppl): 1933S-1938S.

J. L. Tang-Peronard, H. R. Andersen, T. K. Jensen and B. L. Heitmann. 2011. Endocrine-disrupting chemicals and obesity development in humans: a review. *Obes Rev* 12(8): 622-36.

J. L. Schnoor. 2011. Obesogens, the exposome, and ES&T. *Environ Sci Technol* 45(7): 2517.

A. Janesick and B. Blumberg. 2011. Minireview: PPARgamma as the target of obesogens. *J Steroid Biochem Mol Biol* 127(1-2): 4-8.

F. Grun. 2010. Obesogens. *Curr Opin Endocrinol Diabetes Obes* 17(5): 453-9.

R. R. Newbold. 2010. Impact of environmental endocrine disrupting chemicals on the development of obesity. *Hormones (Athens)* 9(3): 206-17.

J. R. Barrett. 2010. To each his own: DEHP yields species-specific metabolic phenotypes. *Environ Health Perspect* 118(2): A81.

D. Sharp. 2009. Environmental toxins, a potential risk factor for diabetes among Canadian Aboriginals. *Int J Circumpolar Health* 68(4): 316-26.

C. Casals-Casas, J. N. Feige and B. Desvergne. 2008. Interference of pollutants with PPARs: endocrine disruption meets metabolism. *Int J Obes (Lond)* 32 Suppl 6(issue#): S53-61.

No relevant data

M. P. Alexander, S. H. Nasr, D. C. Watson, G. P. Mendez and H. G. Rennke. 2011. Renal crescentic alpha heavy chain deposition disease: a report of 3 cases and review of the literature. *Am J Kidney Dis* 58(4): 621-5.

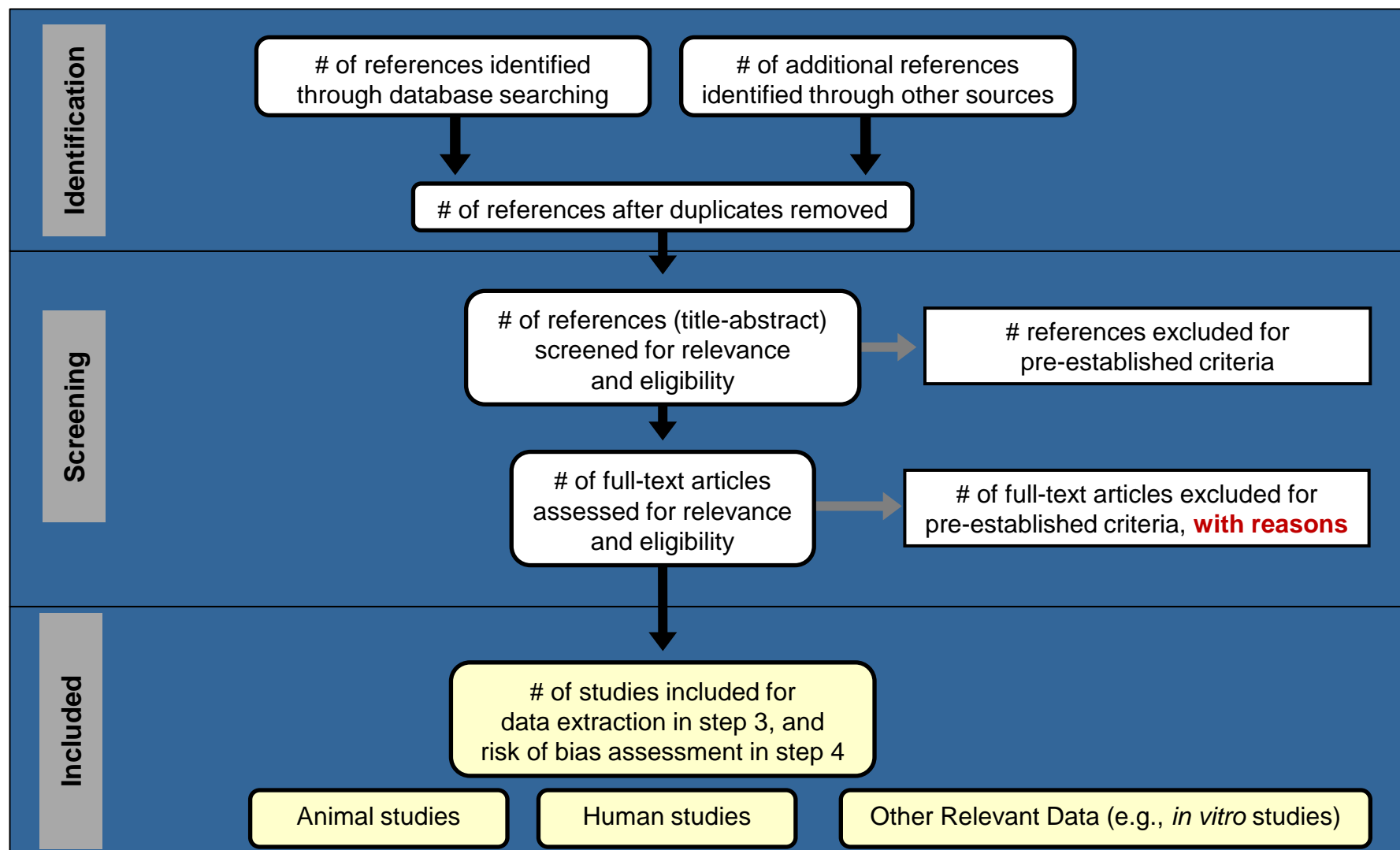
E. A. Wilkes, A. L. Selby, A. T. Cole, J. G. Freeman, M. J. Rennie and Z. H. Khan. 2011. Poor tolerability of thalidomide in end-stage oesophageal cancer. *Eur J Cancer Care (Engl)* 20(5): 593-600.

A. Makhlof, Y. Tozuka and H. Takeuchi. 2011. Design and evaluation of novel pH-sensitive chitosan nanoparticles for oral insulin delivery. *Eur J Pharm Sci* 42(5): 445-51.

S. Oh, S. J. Kim, J. H. Hwang, H. Y. Lee, M. I. Park, J. Park, Y. S. Jo, Y. K. Kim, C. H. Lee, K. P. Kwon, M. Shong and S. B. Park

Words: 988 | 74%

From Literature Search to Study Selection



Step 3: Extract Data from Studies

- **Extract data**

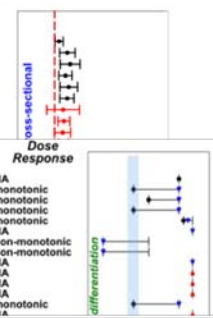
- Individual study information collected systematically
- There are separate template data extraction forms for human, animal, and *in vitro* studies
- Follow procedures defined in protocol for
 - Data extraction by a member of the evaluation team
 - Data extraction forms would be customized for each evaluation
 - Quality assurance of data



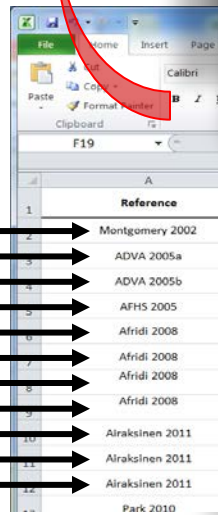
Data Extraction Applications

- Data extraction is transparent and consistent
- Files can be disseminated for data mining
- Variety of outputs without re-entering data
 - Figures (Meta Data Viewer)
 - Appendix tables (Mail Merge)

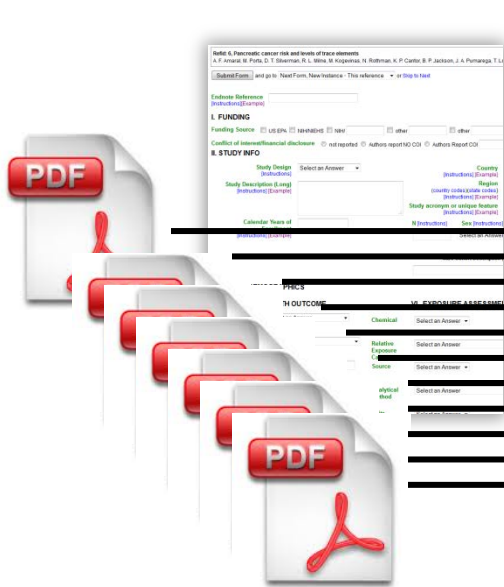
Reference	Study Description (n)	Statistic adjOR (95% CI)	Exposure
Gorg et al. 2009	Europe (6 countries) CESAR, 9-12y, (8,926)	1.26 (1.03, 1.55)	yes/no
Toschke et al. 2002	Germany (Bavaria) 5-6.9y (1997), (8,365)	1.92 (1.29, 2.86)	yes/no
Toschke et al. 2003	Germany (Bavaria) 5-6y (2001/2002), (4,974)	2.22 (1.33, 3.69)	yes/no
Toschke et al. 2007	Germany (Bavaria) 5-9y (2001/2002), (5,472)	1.75 (1.25, 2.43)	yes/no
Von Kries 2002	Germany (Bavaria) 5-6.9y, (6,483)	2.06 (1.31, 3.23)	yes/no
Von Kries 2008	Germany (Bavaria) 5-6.9y, (5,899)	1.9 (1.3, 2.7)	yes/no
Ino et al. 2011	Japan (Kumagaya) 9-10y, (2,508)	1.55 (0.67, 3.57) [crude prevOR]	yes/no
Kondry et al. 2011	UK (Merseyside) 5-11y, (3,038)	1.61 (1.19, 2.18)	yes/no
Raum et al. 2011	Germany (Aachen) 6y, (1,954)	1.51 (0.99, 2.28)	yes/no



Reference & Study Design	Study population	Outcome & Diagnostic	Chemical	Risk Estimate (95% CI)	Exposure Comparison	Adjustment factors
(Chang et al. 2010) cross-sectional	Taiwan[NR] near PCP factory, (1,234(1,478)) N Analysis (Total N): 1,234(1,478) Cases or prevalence: NR Exposed cases: NR	HOMA β-cell ≥75th percentile	PCDD/PCDF	1.3 (95%CI: 0.9,1.8) adjOR	>20.6 vs ≤20.5 pg WHO98-TEQDF/g lipid (serum)	age, sex, BMI, smoking, weight control, physical activity, and family history of diabetes
(Chang et al. 2010) cross-sectional	Taiwan[NR] near PCP factory, (1,234(1,478)) N Analysis (Total N): 1,234(1,478) Cases or prevalence: 178(14%) Exposed cases: NR	HOMA-IR HOMA-IR≥75th percentile	PCDD/PCDF	1.7 (95%CI: 1.2,2.4) adjOR	>20.5 vs ≤20.5 pg WHO98-TEQDF/g lipid (serum)	age, sex, BMI, smoking, weight control, physical activity, and family history of diabetes
(Chang et al. 2011) cross-sectional	Taiwan[NR] near PCP factory, (333(1,449)) N Analysis (Total N): 333(1,449) Cases or prevalence: 355 Exposed cases: 28(1.9%)	HOMA β-cell ≥75th percentile	PCDD/PCDF	1.45 (95%CI: 0.62,3.42) adjOR	≥30.3(T3) vs <15.9(T1) pg WHO98-TEQDF/g lipid (serum)	age, gender, smoking, physical activity, waist circumference, systolic blood pressure, diastolic blood pressure, and a family history of diabetes
(Chang et al. 2011) cross-sectional	Taiwan[NR] near PCP factory, (357(1,449)) N Analysis (Total N): 357(1,449) Cases or prevalence: 355 Exposed cases: 28(1.9%)	HOMA-IR HOMA-IR≥75th percentile	PCDD/PCDF	4.89 (95%CI: 2.09,12.2) adjOR	15.9-<30.3(T2) vs <15.9(T1) pg WHO98-TEQDF/g lipid (serum)	age, gender, smoking, physical activity, waist circumference, systolic blood pressure, diastolic blood pressure, and a family history of diabetes

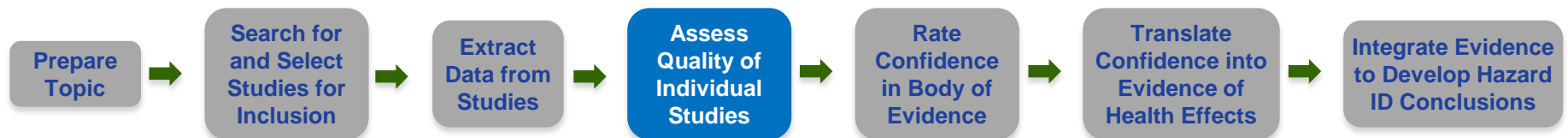


Reference	Study Design	Study Description	N Analysis (Total)	Health Outcome
Montgomery 2002	Pros	UK[nat'l] NCDS, 33y, (8,926)	4917	obese
ADVA 2005a	Retro	Australia[nat'l] Viet. vets, (8,365)	6,166 deaths (59,179)	diabetes
ADVA 2005b	Retro	Australia[nat'l] Viet. vets, (4,974)	1,052 deaths (59,179)	diabetes
AFHS 2005	Pros	US[AFHS] ORH 2002 exam cycle, (8,365)	776(1950)	diabetes
Afridi 2008	CS	Pakistan[Hyderabad] non-smokers, (5,899)	225	diabetes
Afridi 2008	CS	Pakistan[Hyderabad] smokers, (5,899)	209	diabetes
Afridi 2008	CS	Pakistan[Hyderabad] non-smokers, (5,899)	225	diabetes
Afridi 2008	CS	Pakistan[Hyderabad] non-smokers, (5,899)	225	diabetes
Airaksinen 2011	CS	Finland[Helsinki] 57-70y, (2,508)	1988	T2D
Airaksinen 2011	CS	Finland[Helsinki] 57-70y, (2,508)	398(1988)	T2D
Airaksinen 2011	CS	Finland[Helsinki] 57-70y, (2,508)	1988	T2D
Park 2010	CC	S. Korea[Ullim] 240y, (1,954)	70(100)	met.synd.



Step 4: Assess the Quality of Individual Studies

- **Study quality or risk of bias**
 - How credible are the study findings?
- **State of the art for assessing risk of bias**
 - Single summary scores for “study quality” are strongly discouraged
 - Reporting quality checklists are of limited utility (mix bias and reporting)
- **Existing methods**
 - Established risk of bias tools for randomized controlled trials
 - No existing consensus on how to assess risk of bias for
 - Observational human studies, or
 - Animal studies
 - *In vitro* studies



Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies

METHODOLOGICAL EVALUATION OF OBSERVATIONAL RESEARCH (MEVORECH)—Observational Studies of Risk Factors of Chronic Diseases

Please define in the protocol specific to your research quality components:

1. Define and justify target population
2. Define and justify population subgroups if applicable, race, gender, other
3. Response rate, justify acceptable response rate, and rate that can be defined as a major flaw of the study in the total sample and in race, gender, and other subgroups if applicable.
4. Exclusion rate from the analysis - define in the protocol ranges specific for your research and rate that can be defined as a major flaw of the study in the total sample and in race, gender, and other subgroups if applicable.
5. Source of measure outcomes. Define and justify minor flaws specific for the nature of the condition:

Sources

Suggested minor flaws

Self reported (collected for the study) _____

Proxy reported (collected for the study) _____ Minor flaw

Objectively measured with diagnostic methods for the purpose of the study (independent on health care) _____

Measured by interviewers for the study _____

Observed during clinical exam for the purpose of the study _____

Obtained from medical records (mining of the data collected for health care purposes) _____ Minor flaw

Obtained from administrative database (mining of the data collected for health care purposes) _____ Minor flaw

Obtained from registries or administrative databases (collected for epidemiologic evaluation independent of health care) _____

Other (specify) _____

1. Reference period is specific for the nature
2. Severity (degree of severity) is applicable
3. Frequency of the importance of the outcomes
4. Gold standard to measure outcomes
5. Reliability of the exposure

Methods Guide for Comparative Effectiveness Reviews

Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions

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



The Cochrane Handbook

13.5.2.3 Tools for assessing methodological quality or risk of bias in non-random

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	
Performance bias	Did the strategy for recruiting participants into the study differ across study groups?	x	x	x	x	x
	Does the design or analysis critically account for important confounding and modifying variables through matching, stratification, multivariable analysis, or other approaches?	x	x	x	x	x
Attrition 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	x
Detection 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
	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		
Reporting bias	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
Reporting bias	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	x
Reporting bias	Were the potential outcomes prespecified by the researchers? Are all prespecified outcomes reported?	x	x	x	x	x
	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 similar as to be matched for the exposure of interest.

Collaborative Approach to Meta Analysis and Systematic Review of Animal Data from Experimental Studies

C·A·M·A·R·A·D·E·S

Review of Animal Data from Experimental Studies

Welcome

The CAMARADES collaboration provides a supporting framework for groups involved in the systematic review and meta-analysis of data from animal studies in experimental stroke.

Our interests range from identifying potential sources of bias in animal work, developing recommendations for improvements in the design and reporting of animal studies, developing the meta-analysis methodology, the better to apply it to animal studies, through to the selection of candidate stroke drugs for clinical trial.

CAMARADES aims to provide a central focus for data sharing, to act as a repository for those wishing to carry out such reviews, to provide a web-based stratified meta-analysis bioinformatics engine (under development), and to act as a repository for completed reviews.

While the CAMARADES data set is curated from Edinburgh it is mirrored at the National Stroke Research Institute in Melbourne Australia and is the shared property of all those contributing data.

Some CAMARADES members from Edinburgh, Sweden and Melbourne ran the Edinburgh Marathon Relay on 31st May to raise funds for Macmillan Cancer Support. We finished in 30h 28 minutes, 42 out of 500 were female.

The 2nd International Symposium on Systematic Reviews in Laboratory Animal Science will take place in Edinburgh, 8-9th March 2013.

CAMARADES

CHALLENGES: Predicting the future: Study successfully detects Alzheimer's disease decades before symptoms first manifest

CAMARADES: Migration talk on the epidemiology of animal research at World Health Summit #WHStalk available at bit.ly/1D9v7Tt

Twitter: @camarades

authors are expected to use for assessing outcome reporting and other potential judgement is formulated by assessing judgement of bias. The tool was not developed for bias assessments for NRS.

studies (see Box 13.1.a and Section 13.5.2.3) should carry out and more details will be provided in the next section on the purposes of illustrating the heterogeneity of bias assessments for NRS.

randomized studies of interventions have been created, and were reviewed systematically by Deeks et al (Deeks 2003) in their review they located 182 tools, likely useful for systematic reviews as they force the reviewer to be systematic in their study assessments and attempt to ensure that quality judgements are made in a degree of adjustment as they neglected to elicit detailed information about how study participants were allocated to groups, which in terms of the risk of bias is a major concern. In common with some tools for assessing the quality of randomized trials, some did not distinguish items relating to the most useful tools identified in this review are the Downs and Black instrument and the Newcastle-Ottawa Scale (Downs 1998, Wells 2008).

methodological systematic review (MacLachlan 2006). The reviewers found that some of the 29 items were difficult to apply to case-control studies, that the tool was time consuming to use. The Newcastle-Ottawa Scale, which has been used in NRS/MG workshops to illustrate issues in data extraction from primary studies, however, the items may still need to be customized to the review question of interest. Review authors also need to be aware of differences in Newcastle-Ottawa Scale uses the term 'selection bias' to describe what others may call 'applicability' or 'generalizability'.

Review authors may also find it helpful to consider items included in reporting statements for randomized trials (see 2007) in order to highlight gaps in reporting (and execution) in NRS (Reeves 2004, Reeves 2007).

Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies

Consideration of 5 traditional risk of bias domains

Study design determines which questions apply

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	
Performance bias	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
Attrition 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	
Detection 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
	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
Reporting bias	Were confounding variables assessed using valid and reliable measures, implemented consistently across all study participants?		x	x	x	x
	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.

Methods Guide for Comparative Effectiveness Reviews
Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions

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



Adaptation of Existing Study Quality Methods

- Although there are a variety of risk of bias methods for human studies, animal tools are generally reporting quality checklists (e.g., ToxRTool)
- The recent AHRQ method guide* was particularly useful as a model because it covers RCTs and a range of human observational studies
- The clarity group scale for answering risk of bias questions was also useful (definitely low, probably low, probably high, to definitely high)

The screenshot displays the Clarity website interface. At the top, the Clarity logo is on the left, and the tagline "CLINICAL ADVANCES THROUGH RESEARCH AND INFORMATION TRANSLATION" is on the right. Below the tagline is a navigation menu with links for home, background, team members, research, education, press releases, contact, and members only. The main content area features two overlapping tool pages. The top page is titled "Tool to Assess Risk of Bias in Randomized Controlled Trials" and includes a question: "1. Was the allocation sequence generated and concealed?" with response options: "Definitely yes (low risk of bias)", "Probably yes", "Probably no", and "Definitely no (high risk of bias)". The bottom page is titled "Tool to Assess Risk of Bias in Cohort Studies" and includes a question: "1. Was selection of exposed and non-exposed cohorts drawn from the same population?" with response options: "Definitely yes (low risk of bias)", "Probably yes", "Probably no", and "Definitely no (high risk of bias)". A red arrow points from the text in the first bullet point to the "Definitely yes (low risk of bias)" option in the Cohort Studies tool.

Step 4: Assess the Quality of Individual Studies



- **Study quality or risk of bias**

- Judge whether the design and conduct of individual studies compromise credibility in the link between exposure and outcome
- Evaluation is endpoint/outcome specific
- Use predefined set of questions adapted from AHRQ to address **both human studies and animal toxicology studies**

- Study design determines which questions are applicable
- Answers equate to risk of bias rating for each question/criteria

++	Definitely Low risk of bias
+	Probably Low risk of bias
-	Probably High risk of bias
--	Definitely High risk of bias



General Risk of Bias Answer Format



++ Definitely Low:

- Direct evidence of low risk of bias practices
(Protocol includes specific examples of relevant low risk of bias practices)

+ Probably Low:

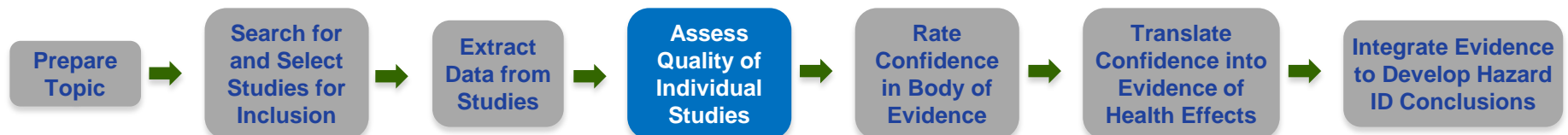
- Indirect evidence of low risk of bias practices
- OR deviations from low bias practices would not appreciably bias results

- Probably High:

- Indirect evidence of high risk of bias practices
- OR there is insufficient information provided

-- Definitely High:

- Direct evidence of high risk of bias practices



Selective Reporting Bias

Example Question – Appendix 2 of Protocols



Were all measured outcomes reported?

++ Definitely Low:

- There is **direct** evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction.

Explicit guidance is provided for each answer to determine the risk of bias rating

Selective Reporting Bias

Example – Appendix 2 of Protocols (continued)



- Were all measured outcomes reported?

++ Definitely Low:

- There is **direct** evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported. This would include outcomes reported with sufficient detail to be included in meta-analysis or fully tabulated during data extraction.

+ Probably Low:

- There is **indirect** evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have been reported
- **OR** analyses that had not been planned at the outset of the study (i.e., retrospective unplanned subgroup analyses) are clearly indicated as such and it is deemed that the omitted analyses were not appropriate and selective reporting would not appreciably bias results. This would include outcomes reported with insufficient detail such as only reporting that results were statistically significant (or not).

Selective Reporting Bias



Example – Appendix 2 of Protocols (continued)

- **Were all measured outcomes reported?**

- **Probably High:**

- There is **indirect** evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported
- **OR** there is insufficient information provided about selective outcome reporting.

- **Definitely High:**

- There is **direct** evidence that all of the study's measured outcomes (primary and secondary) outlined in the protocol, methods, abstract, and/or introduction (that are relevant for the evaluation) have not been reported. In addition to not reporting outcomes, this would include reporting outcomes based on composite score without individual outcome components or outcomes reported using measurements, analysis methods or subsets of the data (e.g., subscales) that were not pre-specified or reporting outcomes not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect).

Risk of Bias Rating Individual Animal Studies

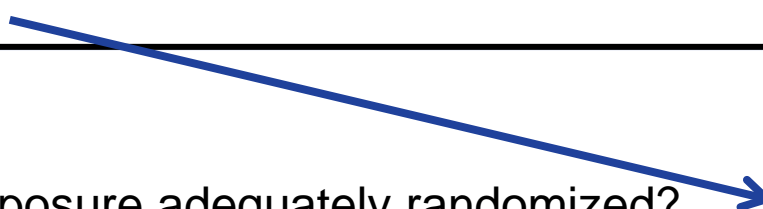


- ++ Definitely Low risk of bias
- + Probably Low risk of bias
- - Probably High risk of bias
- -- Definitely High risk of bias

○ Not applicable due to study design

Draft OHAT Risk of Bias Questions

	Andy et al., 2010	Bucher et al., 1999	Wolfe et al., 2000	Boyles et al., 2011	Thayer et al., 2008
Was administered dose or exposure adequately randomized?	+				
Was allocation to the study groups adequately concealed?	-				



Risk of Bias Rating Individual Animal Studies



- ++ Definitely Low risk of bias
- + Probably Low risk of bias
- - Probably High risk of bias
- -- Definitely High risk of bias

Draft OHAT Risk of Bias Questions

○ Not applicable due to study design

		Andy et al., 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?	+	++	+	-	+

Using Risk of Bias Data



- **Ability to categorize or “tier” studies based on risk of bias**
 - Can clearly show risk of bias for individual factors or “tier” by all factors
 - Can define “key” risk of bias questions on a project-specific basis
- **Enhance transparency** – risk of bias released as part of evaluation
- **Can stratify or restrict confidence rating conclusions**
 - Stratified analysis with high risk of bias studies included to assess impact
 - Use studies with lower risk of bias (1st and 2nd tier)

Example of Tiers for Risk of Bias



Category	Guidance	Key Criteria		
1 st tier	<p>“definitely low” or “probably low” risk of bias for key criteria AND “definitely low” or “probably low” risk of bias for ≥50% of other</p> <p style="text-align: center;"> ++ + + + + + - - </p>	+	++	+
2 nd tier	Study does not meet criteria for 1 st or 2 nd tier			
3 rd tier	<p>“definitely high” or “probably high” risk of bias for key criteria AND “definitely high” or “probably high” risk of bias for ≥50% of other</p> <p style="text-align: center;"> ++ + + - - - - -- </p>	-	-	--

- **Observational Studies (most human) – 3 key criteria**
 - Can we be confident in the exposure characterization?
 - Can we be confident in the outcome assessment?
 - Does the study design or analysis account for important confounding and modifying variables?
- **Experimental Studies (most animal) – 1 key criteria**
 - Can we be confident in the outcome assessment?

Questions?