Neurotoxicology and Teratology 33 (2011) 354–359

Contents lists available at ScienceDirect

Neurotoxicology and Teratology

journal homepage: www.elsevier.com/locate/neutera

A proposal to facilitate weight-of-evidence assessments: Harmonization of Neurodevelopmental Environmental Epidemiology Studies (HONEES)

Eric Youngstrom [a], Lauren Kenworthy [b], Paul H. Lipkin [c], Michael Goodman [d], Katherine Squibb [e], Donald R. Mattison [f], Laura Gutermuth Anthony [g], Susan L. Makris [h], Ambuja S. Bale [h], Kathleen C. Raffaele [h], Judy S. LaKind [i,j,k,*]

[a] Departments of Psychology and Psychiatry, University of North Carolina at Chapel Hill/Davie Hall, CB 3270, University of North Carolina, Chapel Hill, NC 27599, USA
[b] Children's National Medical Center, Center for Autism Spectrum Disorders, Departments of Pediatrics, Neurology and Psychiatry, George Washington University School of Medicine, 15245 Shady Grove Road, Suite 350, Rockville, MD 20850, USA
[c] Center for Development and Learning, The Kennedy Krieger Institute, Department of Pediatrics, The Johns Hopkins University School of Medicine, 707 North Broadway Baltimore, MD 21205, USA
[d] Department of Epidemiology, Emory University School of Public Health/1518 Clifton Rd., Atlanta, GA 30322, USA
[e] Department of Medicine, University of Maryland School of Medicine/11 South Paca Street, 2nd Floor Baltimore, MD 21201, USA
[f] Eunice Kennedy Shriver National Institute of Child Health and Human Development National Institutes of Health, Department of Health and Human Services, 6100 Executive Blvd, Room 7B03N, Bethesda MD 20892–7510, USA
[g] Children's National Medical Center, Center for Autism Spectrum Disorders, Departments of Pediatrics and Psychiatry, George Washington University School of Medicine, 15245 Shady Grove Road, Suite 350, Rockville, MD 20850, USA
[h] U.S. Environmental Protection Agency, National Center for Environmental Assessment, 1200 Pennsylvania Ave., NW, Washington, DC 20460, USA
[i] LaKind Associates, LLC, 106 Oakdale Avenue, Catonsville, MD 21228, USA
[j] Department of Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD 21202, USA
[k] Department of Pediatrics, Penn State College of Medicine, Hershey, PA 17033, USA

ARTICLE INFO

Article history:
Received 20 October 2010
Received in revised form 13 January 2011
Accepted 30 January 2011
Available online 18 February 2011

Keywords:
Neurodevelopmental function tests
Interstudy consistency
Environmental epidemiology
Weight-of-evidence assessments
Risk assessment
Environmental chemicals

ABSTRACT

The ability to conduct weight-of-evidence assessments to inform the evaluation of potential environmental neurotoxicants is limited by lack of comparability of study methods, data analysis, and reporting. There is a need to establish consensus guidelines for conducting, analyzing, and reporting neurodevelopmental environmental epidemiologic studies, while recognizing that consistency is likewise needed for epidemiology studies examining other health outcomes. This paper proposes a set of considerations to be used by the scientific community at-large as a tool for systematically evaluating the quality of proposed and/or published studies in terms of their value for weight-of-evidence assessments. Particular emphasis is placed on evaluating factors influencing the risk of incorrect conclusions at the level of study findings. The proposed considerations are the first step in what must be a larger consensus-based process and can serve to catalyze such a discussion. Achieving consensus in these types of endeavors is difficult; however, opportunities exist for further interdisciplinary discussion, collaboration, and research that will help realize this goal. Broad acceptance and application of such an approach can facilitate the expanded use of environmental epidemiology studies of potential neurodevelopmental toxicants in the protection of public health, and specifically children's health.

1. Introduction

Pediatric neurodevelopmental disorders have generated substantial attention from the scientific community, the public and media. This high visibility is due in part to an observed increased prevalence of autism spectrum disorders (Rice et al., 2010) and possibly attention deficit hyperactivity disorder (ADHD) (Pastor and Reuben, 2008). Many chemicals are known to be human neurotoxicants, at least following acute exposure to adults (Grandjean and Landrigan, 2006). In addition, associations between specific environmental chemicals and neuropsychiatric disorders have been reported (e.g., depression and air pollutants (Szyszkowicz et al., 2009); anxiety and PCB 153 (Plusquellec et al., 2010); ADHD and polyfluoroalkyl chemicals

Abbreviations: ADHD, attention deficit hyperactivity disorder; CONSORT, Consolidated Standards of Reporting of Trials; HONEES, Harmonization of Neurodevelopmental Environmental Epidemiology Studies; IRIS, Integrated Risk Information System; PCBs, polychlorinated biphenyls; QUADAS, Quality Assessment of Diagnostic Accuracy Studies; US EPA, United States Environmental Protection Agency; STARD, Standards for Reporting of Diagnostic Accuracy; STROBE, STrengthening the Reporting of OBservational studies in Epidemiology.

* Corresponding author at: LaKind Associates, LLC, 106 Oakdale Avenue, Catonsville, MD 21228, USA. Tel./fax: +1 410 788 8639.
E-mail address: lakindassoc@comcast.net (J.S. LaKind).

0892-0362/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ntt.2011.01.004

Table 1
IRIS chemicals with toxicological reviews.[a]

No neurotoxicity data in animals or humans		
Acrolein	Methylene diphenyl diisocyanate (monomeric and polymeric MDI)	Thallium chloride
Bentazon (Basagran)	Mirex[b]	Thallium oxide
Beryllium and compounds[b]	Naphthalene[b]	Thallium selenite
1,3-Butadiene	Phosgene	1,2,3-Trichloropropane
Chromium(III), insoluble salts	Propionaldehyde	2,2,4-Trimethylpentane
Chromium(VI)	Thallium carbonate	Zinc and compounds
Ethylene glycol monobutyl ether (EGBE) (2-Butoxyethanol)[b]		
No neurotoxicity data in animals; some data in adult humans (mostly positive); no data in children		
Chloroform	Hexachlorocyclopentadiene (HCCPD)	Thallium nitrate
Animal DNT[c] data (all positive); no adult animal data; no adult or child human data		
	1,2-Dibromoethane	Perchlorate and Perchlorate salts[b]
Chlorite (sodium salt)	Diesel engine exhaust	2,2′,4,4′-Tetrabromodiphenyl ether (BDE-47)
2,2′,3,3′,4,4′,5,5′,6,6′-Decabromodiphenyl ether (BDE-209)	2,2′,4,4′,5,5′-Hexabromodiphenyl ether (BDE-153)	
Adult animal data; some DNT data, no human data; results in animals are mixed		
Acetonitrile	Cyclohexane	2,2′,4,4′,5-Pentabromodiphenyl ether (BDE-99)
Bromobenzene	Dibutyl phthalate[b]	Phenol
Cerium oxide and cerium compounds	1,1-Dichloroethylene (1,1-DCE)	Platinum[b]
Cis- and *trans*-1,2-Dichloroethylene[a]	1,3-Dichloropropene	Quinoline
Cumene	Ethyl *tert*-butyl ether (ETBE)[b]	Tributyltin oxide (TBTO)
	2-Methylnaphthalene	1,3,5-Trinitrobenzene
Neurotoxicity testing in adult animals and data in adult humans (all are positive); a few assessments of DNT in animals; no data in children		
Acetone	Chlordecone (Kepone)	Nitrobenzene
Acrylamide[b]	*n*-Hexane	1,1,2,2-Tetrachloroethane[b]
Barium and compounds	2-Hexanone	Tetrahydrofuran[b]
Benzene	Hydrogen cyanide[b]	Thallium(I), soluble salts
Carbon tetrachloride[b]	Methanol[b]	Thallium acetate
Chlordane (technical)	Methyl chloride	Thallium(I) sulfate
Chloroprene	Methyl ethyl ketone (MEK)	Toluene
Dichloroacetic acid	Methyl isobutyl ketone (MIBK)	1,1,1-Trichloroethane
Dichlorobenzenes[b]	Methyl methacrylate	Vinyl chloride
1,4-Dioxane[b]	Pentachlorophenol[b]	Xylenes

Data in children		
Neurodevelopmental outcomes observed (congenital anomalies)	**Case reports in children identified effects on nervous system**	**Neurodevelopmental testing was conducted in children**
Chlorine dioxide	Boron and compounds	Methyl mercury
Trichloroacetic acid[b]	Bromate	Perchloroethylene[b]
	Chloral hydrate	Trichloroethylene[b]
	Hydrogen sulfide	

[a] To facilitate consistency of comparison across chemicals, this evaluation includes only those substances for which an IRIS Toxicological Review was available on the IRIS website (U.S. EPA, 2010a); it is not a comprehensive list of neurotoxic environmental chemicals. Thus, a number of important chemical assessments that include neurodevelopmental testing in children (e.g., lead, PCBs, some pesticides) are not included in the list.
[b] External review draft.
[c] DNT = developmental neurotoxicity.

(Hoffman et al., 2010); ADHD and PCBs and lead (Eubig et al., 2010); autism and pesticides (Roberts et al., 2007)).

In spite of the growing concerns about perturbations in neurodevelopment that may be associated with environmental exposures, few environmental chemicals are regulated on the basis of neurodevelopmental outcomes. A review of the U.S. Environmental Protection Agency (US EPA) Integrated Risk Information System (IRIS) database illustrated the limited impact of studies assessing neurodevelopmental endpoints in children have had on the characterization of hazards resulting from early life stage exposures to environmental toxicants. The IRIS database is a publicly available peer-reviewed source of toxicological information maintained by the US EPA National Center for Environmental Assessment (NCEA) for over 550 environmental chemicals (U.S. EPA, 2010a). For those chemical assessments that were issued in 1997 or later (a total of 85, listed in Table 1), detailed hazard and dose response information is provided in a "Toxicological Review," thereby allowing an examination of the use and influence of neurobehavioral data in oral and inhalation reference value derivation. Out of the 85 assessments examined, 19 had no neurotoxicity data in either animals or humans, 24 had animal but not human neurotoxicity data, and 33 had adult human neurotoxicity data (30 of which were supported by animal data). Only 9 of the 85 IRIS toxicological databases that were examined included any information on neurotoxicity in children, and of those only 3 reported

data from actual neurodevelopmental testing (the others being 2 chemicals with reports of congenital anomalies and 4 chemicals with case reports in children that suggested effects on the nervous system). The chemicals for which neurodevelopmental testing was conducted were methyl mercury, perchloroethylene, and trichloroethylene. While these data were in all cases considered qualitatively important in the weight-of-evidence evaluation of the chemical toxicity profile, methyl mercury was the only final assessment for which neurodevelopmental testing data in children were actually used as the quantitative basis for a reference value.

For these particular examples, the use of the data from neurobehavioral testing in children was limited for reference value derivation by a number of considerations related to study conduct or data reporting. These included issues such as adequacy of sample size, inconsistent testing methodologies, questions about the selection or implementation of testing procedures, inadequate consideration of confounding factors, uncertainties regarding exposure characterization, reproducibility of study findings, and inconsistencies due to timing of exposure or life stage of assessment (U.S. EPA, 2010a). For two of the draft assessments (perchloroethylene and trichloroethylene) that included neurodevelopmental data in children, other toxicological outcomes were observed at lower exposure levels and thus were selected to derive reference values (U.S. EPA, 2008, 2010b). It is important, however, to recognize the possibility that the full range of neurobehavioral outcomes at the lower end of the dose–response array may not have been fully characterized, thus limiting the weight-of-evidence evaluation of the data, and/or selection of the point of departure.

There are several factors that limit consideration of neurodevelopment in weight-of-evidence assessments. First, neurobehavioral testing in children is not mandated by any regulatory agency. Second, studies in animals may not fully capture the range of adverse outcomes in neurological development that might be observed in children. In fact, it has been noted that developmental neurotoxicity might not be evident in routine toxicology tests (Grandjean and Landrigan, 2006; Raffaele et al., 2010). Third, use of epidemiology data in weight-of-evidence assessments has been hindered by differences in methods used in conducting neurodevelopmental epidemiologic studies as well as differences in data analysis and reporting (Goodman et al., 2010; Youngstrom et al., 2010).

To address these and related issues, a symposium on “Advancing Neurodevelopmental Evaluation in Children: An Interdisciplinary Approach” was held on June 29, 2010 at the joint 34th annual Neurobehavioral Teratology Society, 50th annual Teratology Society, and 23rd annual Organization of Teratology Information Specialists meetings, in Louisville, KY. The objective of the symposium was to present recent multidisciplinary research centered on advancing epidemiological evaluation of children for the detection of neurobehavioral effects potentially associated with exposure to environmental chemicals and the application of these data to human health risk assessment. The symposium presentations included a critical review of commonly used neurodevelopmental tests in epidemiological studies of associations between environmental chemical exposure and adverse health effects, and tests that should be considered for future studies (Youngstrom et al., 2010; Kenworthy et al., 2010; Anthony et al., 2010); a critical review of tests used in environmental epidemiology studies of PCBs (Goodman, 2010; Goodman et al., 2010); discussion of a 2009 interdisciplinary Needs Assessment Workshop which resulted in innovative thinking about paths forward for neurodevelopmental function testing in environmental epidemiology (LaKind et al., 2010); and considerations for selection, administration, and interpretation of neurodevelopmental tests in studies of environmental chemicals. The symposium provided a forum for further discussion on advancing the science of environmental neurodevelopmental/epidemiological research.

A crucial finding of the analysis of epidemiology literature on PCBs and neurodevelopment was that our ability to conduct weight-of-evidence assessments, even with several studies, is limited when study methods, data analysis, and reporting lack comparability (Goodman et al., 2010). Goodman et al. (2010) called for establishing consensus standards for the conduct, analysis, and reporting of neurodevelopmental environmental epidemiologic studies with the recognition that consistency is needed for epidemiology studies examining other outcomes, as well. Youngstrom et al. (2010), in their assessment of use of neurodevelopmental function tests used in environmental epidemiology, noted that “the field of environmental epidemiology may be nearing a stage where a formal set of reporting guidelines could be developed to help the design of future studies, as has been done with clinical trials, studies of diagnostic assessment tools, and medical epidemiological studies.”

We attempted to address these issues by proposing a set of considerations to be used by the scientific community at-large as a tool for systematically evaluating the quality of proposed and/or published studies in terms of their value for weight-of-evidence assessments. We specifically propose considerations and quality criteria for epidemiological research in the area of neurodevelopmental disorders and exposure to chemicals. Particular emphasis is placed on evaluating the sensitivity and specificity of the measures, as these characteristics directly affect the risk of false positive and false negative decisions at the level of the individual, thus influencing the risk of incorrect conclusions at the level of study findings.

We recognize that the considerations proposed here, which we refer to as HONEES (Harmonization of Neurodevelopmental Environmental Epidemiology Studies), are the first steps in what must be a larger consensus-based process. We hope that this serves to catalyze such a discussion. Ultimately, the end product – a clearly articulated set of standards and quality criteria – is intended to help inform the design of proposed studies and to assist funding agencies in deciding which proposals are most likely to advance the field. Perhaps most importantly, use of this type of guidance would enhance the ability of scientists to utilize results from such studies as the National Children's Study (http://www.nationalchildrensstudy.gov/Pages/default.aspx) for weight-of-evidence assessments in support of chemical regulation for developmental neurotoxicants.

2. Methods

The proposed considerations are modeled on the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) guidelines (von Elm et al., 2008) (http://www.strobe-statement.org/). They also incorporate elements from the STARD criteria (Standards for Reporting of Diagnostic Accuracy) for assessment studies (Bossuyt et al., 2003), the Evidence Based Medicine criteria for the validity of a study of harm (Straus et al., 2005), and prior recommendations from the fields of toxicology and psychology (Youngstrom et al., 2010). These considerations were gathered, reviewed for relevance, and then collated into a more relevant set of recommendations focused on epidemiological studies related to neurodevelopment and chemical exposures.

For the purposes of reporting and evaluating studies of the neurodevelopmental effects associated with toxicant exposure, the STROBE guidelines provide an excellent starting point. STROBE includes both general recommendations and specific criteria that are relevant to longitudinal cohort studies, case control studies, and cross-sectional studies — the three major designs used to study neurodevelopmental effects of chemical exposure. For present purposes, the major contribution of the STROBE guidelines was to increase the focus on study factors that might change assessment results at the study level, such as variables that could influence the sensitivity of the assessment to exposure effects.

After comparing the tools available in the literature for evaluating epidemiological studies (STROBE), studies of diagnostic efficiency (STARD), and recent “evidence based assessment” reviews in

Table 2
Proposed Guidance for Evaluating Neurodevelopmental Environmental Epidemiology Studies: Harmonization of Neurodevelopmental Environmental Epidemiology Studies (HONEES).

Item	Yes	No	Unclear
Sampling and participants			
1. * Were participant selection criteria clearly described?			
2. # Were there clearly defined groups of participants, similar in all important ways other than exposure to the chemical? (e.g. IQ scores, SES, age) Alternately, are sufficient details reported to allow stratification, covariate or propensity score correction, or other appropriate adjustment?			
3.* Were the participants representative of the population to whom results would be generalized in practice?[a]			
4. * Were withdrawals from the study explained? (e.g., flow diagram, or other accounting)			
Assessment procedures			
5. # Were exposures and clinical outcomes measured in the same ways in both groups? Or did the methods allow potential introduction of bias for one group? (i.e., was the assessment of outcomes either objective or blinded to exposure?)[b]			
6. # Was the follow-up of study participants sufficiently long for the outcome to occur?[c]			
7. # Did the follow-up have an acceptable level of participant retention to avoid bias? Or was sufficient information gathered to estimate and adjust for potential bias (e.g., propensity scoring)?[c]			
8. Did the design avoid potential "work up" bias or differential selection of subjects for assessment between the exposed versus unexposed groups? Did the whole sample (or a random selection of the sample) receive the same assessment protocol?			
9. * Did participants receive the same assessments regardless of degree of exposure to toxicant?			
10. * Was the method for exposure measurement described in enough detail to permit replication or application to new cases?			
11. * Is the assessment tool likely to correctly measure or classify the target construct? (adequate diagnostic sensitivity and accuracy may be particularly important for neurodevelopmental studies)			
12. * Was the neurodevelopmental assessment procedure appropriate?			
a. * Were methods described in sufficient detail to permit replication?			
b.* Was the test appropriate for the ages at which it was used?			
c.* Did the protocol avoid differential burden or fatigue effects across groups that might invalidate results?			
d. * Are there normative data for comparison, outside of the study of environmental chemicals?[d] Alternately, is there evidence of a dose-response gradient within the same study?			
e. * Was the administration valid, or were there major departures from standardization? (e.g., poor training, used outside of age norms, translated version without supporting psychometric data. Study must demonstrate evidence of good training, administration, and scoring and checks for continued administration validity across the length of the study by the review of video tapes or other similar measures)			
13. * Was exposure status determined without knowledge of the results of the neurodevelopmental assessment?			
14. * Were the neurodevelopmental assessment results interpreted without knowledge of the exposure status? (e.g., blinding of assessment administration and scoring staff; computerized administration)			
15. * Was the contextual and supporting information used to interpret the test similar in the research protocol versus standard clinical practice? (e.g., if an assessment tool requires probing or collateral information for clinical scoring, were these additional steps also gathered as part of the research protocol?)[e]			
16. * Were uninterpretable/intermediate test results reported? (e.g., treatment of missing data, and reporting of borderline or midrange scores versus only extreme groups)			
Interpretation and causal inference			
17. # Do the results of the study fulfill some of the methodological tests for inferring causation? (Note that it is not necessary for any single study to meet all of these criteria; the more that are met, the greater converging evidence in support of causal interpretations).			
a. # Is it clear that the exposure preceded the onset of the outcome?			
b. # Is there a dose–response gradient?			
c. # Is there any positive evidence from other "dechallenge–rechallenge" studies?			
d. # Is the association consistent from study to study?			
e. # Does the association make biological sense?			

*Adapted from Whiting et al., 2003.
Adapted from Straus et al., 2005.
[a]Although internal validity of the design is paramount for drawing valid inferences from any particular study, the intent of this recommendation is to encourage reflection on whether the results generalize from the sample to the intended population. Genetic polymorphisms and diet are examples of factors that could moderate the resilience or susceptibility of different groups to the effect of toxicants.
[b]If the measurement in question is truly objective, such as performed by a computer, then it will be impartial. However, this is rarely the case with neurocognitive evaluations. If the assessment involves human interaction with the subject or any degree of subjectivity in scoring, then blinding or some similar methodological safeguard would be important to protect impartiality. Either objective assessment or blinding would be sufficient.
[c]If the design was cross-sectional, then the time period between evaluation and prior exposure should have been long enough to allow toxicant effects to manifest.
[d]Because studies of potential harm with human subjects will almost always involve quasi-experimental designs, there are well-established potential threats to validity (Campbell and Stanley, 1963). Comparing the control group to established norms for measures can indicate whether there are secular trends influencing results, such as the Flynn Effect that has been observed with measures of cognitive ability (Flynn, 1999).
[e]For example, one of the most widely used behavior checklists includes several items that are supposed to be probed or queried before they are scored (Achenbach and Rescorla, 2003), but this step is often ignored when the checklist is used in research settings, resulting in exaggerated estimates of pathology (Drotar et al., 1995).

psychology (Hunsley and Mash, 2005; 2008), we used the *Qu*ality *A*ssessment of *D*iagnostic *A*ccuracy *S*tudies (QUADAS) as the basis for a coding tool (Whiting et al., 2003). The original QUADAS was derived from the 25 criteria in the STARD Guidelines for assessment studies, and the content captured the major study design and reporting elements that are common to both epidemiological (e.g., STROBE) and diagnostic (e.g., STARD) criteria, while also including a stronger emphasis on exposure and outcome assessment and discriminating accurately between groups. The QUADAS has been used in multiple reviews and meta-analyses, accumulating a substantial evidence base supporting its use (Hollingworth et al., 2006; Whiting et al., 2006).

The QUADAS distilled the 25 recommendations from the STARD into 14 key criteria that could be coded from reports of studies in order to evaluate the quality of the study and reporting, and the potential for bias to have influenced findings. Two of these 14 criteria did not appear relevant to studies of environmental chemical exposure. One criterion (QUADAS #4) pertained to the length of time between the screening and reference standard test. The intention

of Criterion #4 was to capture whether the screening and reference tests were given close enough together that the person's diagnostic status was unlikely to change in between test sessions. Treating "diagnostic change over time" as a source of bias does not appear relevant for studies of environmental chemical exposure both because the central issue is not agreement between two tests and because the diagnostic status may change over time as a result of the toxicant exposure. In fact, long-term outcomes may be a central question of interest, not a source of bias in the study design (e.g., a study may examine the relationship between a toxicant level in cord blood at birth and the children's later functioning at age 10). Thus we omitted this criterion from the revised coding tool. Similarly, a second criterion (QUADAS #7) focused on issues relevant to diagnostic screening and not to toxicant studies and was therefore omitted.

The Evidence Based Medicine (EBM) guidelines for evaluating the quality of evidence for studies of harm (Straus et al., 2005) also provided a helpful set of questions and principles for evaluating studies of effects of chemical exposure. The final augmented set (Table 2) includes several additional criteria adapted from the EBM guidelines, as denoted in the item list. These include factors such as whether the subjects were similar in all respects except for chemical exposure, or whether the neurodevelopmental outcomes were assessed using the same methods in both the exposed and non-exposed groups (avoiding potential "work-up" bias).

3. Results

The product of this effort is a set of proposed considerations and recommendations (HONEES) for a process of selecting a test, evaluating its validity, and presenting findings. The main areas for evaluation of the study are Sampling and Participants, Assessment Procedures, and Interpretation and Causal Inference. In the case of the proposed considerations in Table 2, coding would be accomplished by assigning a numerical value to a "yes" or "no" response (with the response of "unclear" leading to further evaluation of the study). The emphasis on criteria and process is consistent with the philosophy of Evidence Based Medicine and we selected the parameters in Table 2 because they have been refined over more than a decade of interdisciplinary review, they focus on evaluating and quantifying the potential risk of harm, and they include study characteristics that influence estimation of risks. New evidence continues to evolve, and existing tests will continually be revised or replaced. Thus a canonical list is ultimately a less useful product as it will be static and will grow ever more dated as the field progresses. Principles and processes, on the other hand, equip the audience to seek new data, evaluate it, and decide when to change assessment practices. The need for recommendations such as those given in Table 2 was highlighted in a review of the epidemiologic literature on PCBs and neurodevelopment (Goodman et al., 2010).

4. Conclusions

The ultimate goal of conducting neurodevelopmental environmental epidemiology studies is to improve public health by providing information necessary to determine whether and how to limit exposures to specific environmental chemicals. However, even multiple, well-conducted observational studies may not provide regulatory agencies with information needed for decision-making if the methods used (including methods for measuring and expressing exposure and methods for examining associations) are not sufficiently consistent across studies. This has resulted in calls for establishing consensus standards for the conduct, analysis, and reporting of epidemiologic studies (Goodman et al., 2010; Goodman, 2010; Bellinger, 2009; Moher et al., 2010).

The proposed considerations in Table 2 are offered as a means of catalyzing the conversation needed to build a set of consensus standards that can be used by the scientific community to encourage the type of inter-study harmonization necessary for improved weight-of-evidence assessments. These criteria can be applied prospectively, in the design of new studies, and retrospectively, in the evaluation of existing studies. Prospectively, these considerations are aspirational and offer a roadmap to research designs that will enhance the validity of findings and avoid pitfalls that pock the domain of neurocognitive assessment. Applied retrospectively to the existing literature, they offer a method for systematically evaluating studies, and possibly loosely ranking them. There is no expectation that any study would necessarily achieve all criteria, nor are the criteria all equally important, though all warrant consideration. Ultimately, for such standards to be of value, acceptance by interested parties, through round-table or consensus conferences involving key scientific and policy leaders, will be necessary. We are aware of the difficulty in achieving consensus in these types of endeavors, but opportunities exist for further interdisciplinary discussion, collaboration, and research that will help realize this goal. Broad acceptance and application of such an approach can facilitate the expanded use of environmental epidemiology studies of potential neurodevelopmental toxicants in the protection of public health, and specifically children's health.

Conflicts of interest statement

The research was supported by a grant from Cefic-Long-range Research Initiative (Cefic-LRI). DRM, SLM, AB, and KCR did not receive grant support from Cefic-LRI. Cefic-LRI was not involved in the design, collection, management, analysis, or interpretation of the data; or in the preparation or approval of the manuscript. Dr. LaKind consults to both government and industry. Dr. Lipkin was an Advisory Board member (2008–2009) for the Autistic Disorder Registrational Trials (Bristol-Myers Squibb) and from 2007–present has been a consultant to InfaCare Pharmaceutical Corporation (Clinical Trial Protocol — Use of Stannsoporfin in Neonatal Jaundice). Dr. Youngstrom has received travel support from Bristol Meyers Squibb. DRM was supported in part by the Intramural Research Program at NICHD for this work and has no financial conflict of interest. The other authors report no conflict of interest.

Acknowledgments

The NBTS symposium was funded through EPA contract number EP10H000282. The research was supported by a grant from Cefic-Long-range Research Initiative (LRI) (DRM, SLM, ASB, and KCR did not receive support from Cefic). The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, the National Institutes of Health, or Cefic-LRI. Cefic-LRI was not involved in the design, collection, management, analysis, or interpretation of the data; or in the preparation or approval of the manuscript. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Achenbach TM, Rescorla LA. Manual for the ASEBA adult forms & profiles. Burlington, VT: University of Vermont; 2003.

Anthony LG, Youngstrom E, Kenworthy L, LaKind JS, Goodman M, Squibb K, et al. Threats to study validity: the Flynn Effect, examiner drift, confounders, lost in translation, and other important considerations. Neurotoxicol Teratol 2010;32 (NBTS36):506.

Bellinger DC. Interpreting epidemiologic studies of developmental neurotoxicity: conceptual and analytic issues. Neurotoxicol Teratol 2009;31:267–74.

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Standards for reporting of diagnostic accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003;326:41–4.

Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin; 1963.

Drotar D, Stein REK, Perrin EC. Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. Special Issue: Methodological issues in clinical child psychology research. J Clin Child Psychol 1995;24:184–92.
Eubig PA, Aguiar A, Schantz SL. Lead and PCBs as risk factors for attention deficit hyperactivity disorder. Environ Health Perspect 2010. doi:10.1289/ehp.0901852.
Flynn JR. Searching for justice: the discovery of IQ gains over time. Am Psychol 1999;54: 5-20.
Goodman M. Weight-of-the-evidence assessment in neurodevelopmental epidemiology: a plea for consistency. Neurotoxicol Teratol 2010;32(NBTS37):507.
Goodman M, Squibb K, Youngstrom E, Anthony LG, Kenworthy L, Lipkin PH, et al. Using systematic reviews and meta-analyses to support regulatory decision making for neurotoxicants: lessons learned from a case study of PCBs. Environ Health Perspect 2010;118:727–34.
Grandjean P, Landrigan PJ. Developmental neurotoxicity of industrial chemicals. Lancet 2006;368:2167–78.
Hoffman K, Webster TF, Weisskopf MG, Weinberg J, Vieira VM. Exposure to polyfluoroalkyl chemicals and attention deficit hyperactivity disorder in U.S. children aged 12–15 Years. Environ Health Perspect 2010 Dec;118(12):1762–7.
Hollingworth W, Medina L, Lenkinski RE, Shibata DK, Bernal B, Zurakowski D, et al. Interrater reliability in assessing quality of diagnostic accuracy studies using the QUADAS tool. A preliminary assessment. Acad Radiol 2006;13:803–10.
Hunsley J, Mash EJ. Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. Psychol Assess 2005;17: 251–5.
Hunsley J, Mash EJ. Developing criteria for evidence-based assessment: an introduction to assessments that work. In: Hunsley J, Mash EJ, editors. A guide to assessments that work. New York: Oxford University Press; 2008. p. 3-16.
Kenworthy L, Anthony LG, Goodman M, LaKind JS, Lipkin PH, Mattison D, et al. Getting the biggest bang for your buck: choosing neurodevelopmental tests that maximize power. Neurotoxicol Teratol 2010;32(NBTS35):506.
LaKind JS, Youngstrom E, Goodman M, Squibb K, Lipkin PH, Anthony LG, et al. A multidisciplinary approach to advancing the science of neurodevelopmental testing in cohorts of infants and young children. Neurotoxicol Teratol 2010;32 (NBTS 34):505.
Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010;7(2):e1000217.
Pastor PN, Reuben CA. Diagnosed attention deficit hyperactivity disorder and learning disability: United States, 2004–2006. National Center for Health Statistics. Vital Health Stat 2008;10(237):1-14. http://www.cdc.gov/nchs/data/series/sr_10/Sr10_237.pdf.
Plusquellec P, Muckle G, Dewailly E, Ayotte P, Bégin G, Desrosiers C, et al. The relation of environmental contaminants exposure to behavioral indicators in Inuit preschoolers in Arctic Quebec. Neurotoxicology 2010;31:17–25.
Raffaele K, Rowland J, May B, Makris S, Schumacher K, Scarano L. The use of developmental neurotoxicity data in pesticide risk assessments. Neurotoxicol Teratol 2010;32:563–72.
Rice C, Nicholas J, Baio J, Pettygrove S, Lee LC, Braun KVN, et al. Changes in autism spectrum disorder prevalence in 4 areas of the United States. Disabil Health J 2010;3:186–201.
Roberts EM, English PB, Grether JK, Windham GC, Somberg L, Wolff C. Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. Environ Health Perspect 2007;115: 1482–9.
Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 3rd ed. New York: Churchill Livingstone; 2005.
Szyszkowicz M, Rowe BH, Colman I. Air pollution and daily emergency department visits for depression. Int J Occup Med Environ Health 2009;22:355–62.
U.S. EPA. IRIS toxicological review of tetrachloroethylene (Perchlorethylene), external review draft. Washington, DC: U.S. Environmental Protection Agency, National Center for Environmental Assessment; 2008. http://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=192423.
U.S. EPA. Integrated risk information system (IRIS). Washington, DC: U.S. Environmental Protection Agency, National Center for Environmental Assessment; 2010a. http:/www.epa.gov/iris.
U.S. EPA. IRIS toxicological review of trichloroethylene, external review draft. Washington, DC: U.S. Environmental Protection Agency, National Center for Environmental Assessment; 2010b. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=215006.
von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344–9.
Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006;6:9.
Youngstrom E, LaKind JS, Kenworthy L, Lipkin PH, Goodman M, Squibb K, et al. Advancing the selection of neurodevelopmental measures in epidemiological studies of environmental chemical exposure and health effects. Int J Environ Res Public Health 2010;7:229–68.