



June 11, 2013

To: Office of Health Assessment and Translation, National Toxicology Program

**RE: National Toxicology Program draft OHAT approach for systematic review and evidence integration for literature-based health assessments.**

We and other scientists at the University of California, San Francisco Program on Reproductive Health and the Environment (PRHE), read with great interest the National Toxicology Program's (NTP) draft OHAT approach and the accompanying case-specific protocols (bisphenol A exposure and obesity, and PFOA & PFOS exposure and immunotoxicity). Thank you for the opportunity to provide input on NTP's important efforts.

On behalf of PRHE scientists we write to say that we believe NTP is undertaking a groundbreaking approach for the environmental health field that follows in the footsteps of the highly successful endeavor to use transparent and systematic approaches in evidence-based medicine. The success of research groups such as COCHRANE in evidence-based medicine clearly shows that this is the way forward for evidence-based environmental health research. Moreover, beginning in 2009, a similar approach has been developed by UCSF and its collaborators in the Navigation Guide; (Woodruff & Sutton, 2011) and applying the Navigation Guide to our first case study on PFOA demonstrated that NTP's proposed approach is both feasible and of great potential value to public health. (Johnson et al, 2013; Koustas et al, 2013) We know that the NTP's approach will be of enormous benefit in speeding the translation of environmental health research into improved health outcomes.

Below we provide comments and suggestions on ways in which we believe the methods, clarity, and direction of the protocols could be improved. We have also enclosed the draft protocol with some suggested text edits.

Overall, beyond our strong endorsement of NTP's general direction, we would like to *strongly support* the following specific aspects of NTP's proposals:

(1) NTP states the exposed population will have input on what question is even asked *before* the assessment is conducted – we believe that framing a meaningful question is essential to the value of research synthesis methods in environmental health;

(2) NTP proposes to make the data used in its analyses publicly available --- we believe this is a huge and welcome step towards transparency in environmental health research synthesis;

(3) NTP states, "*a single, well conducted study may provide evidence of toxicity or a health effect.*" We strongly agree with this health protective evidentiary bar and note that is consistent with current best practices in environmental health research synthesis; and

- (4) NTP states they will not require mechanistic or other types of relevant data for hazard identification. While, when available, mechanistic or related data can help support decision-making, we strongly support this approach as we believe the absence of these data should not preclude decision-making as such delays are not neutral but permit potentially harmful exposure to persist.

We believe NTP's protocol could be strengthened in the following key areas:

- (1) We recommend that NTP should not require prior peer-review of articles in order for a study to be included in its systematic review. We believe this step is not needed because NTP will be making an objective and transparent assessment of risk of bias for every included study by virtue of doing the systematic review. We believe peer review in this situation is inserting unnecessary "expert opinion" into the process, at a minimum is redundant, and of serious concern is that having to include an extra peer-review step for a study to be included in the systematic review can also cause unwarranted delays in decision-making with potentially adverse impacts on public health;
- (2) We recommend that NTP include conflict of interest as a risk of bias domain as standard practice. There is ample evidence in the clinical literature to support funding source as a potential source of bias (Lundh et al, 2012);
- (3) There is a lack of clarity among the protocols as to whether NTP will use the ToxRTool to assess the internal validity of *in vitro* studies. We believe the ToxRTool is not appropriate for this use but rather it is a tool primarily for evaluating reporting. Reporting quality is not considered an appropriate metric to assess risk of bias (Higgins and Green 2011; Viswanathan et al. 2012). As such, NTP should not use ToxRTool for evaluating internal validity as it is not an appropriate tool.
- (4) The bar NTP sets for conducting a meta-analysis may be too high. We note that it has been demonstrated in the preclinical literature that including studies that are somewhat heterogenous in meta-analysis can sometimes demonstrate a consistent exposure effect across these studies (i.e. homogeneity), and in doing so it can increase overall confidence in the body of evidence. See ([www.camarades.info](http://www.camarades.info));
- (5) We note that the BPA and PFOS/PFOA protocols include a step for contacting authors for missing data which we strongly support; however we think that the draft OHAT approach should also contain this information as explicit standard NTP practice;
- (6) We strongly disagree with a step for downgrading evidence in Step 7. It will be difficult to define what NTP means by "*strong*" *opposition for biological plausibility*" that has not already been accounted for in NTP's risk of bias evaluation and other earlier rating of the confidence of the body of evidence.
- (7) In the last paragraph of the Draft OHAT approach NTP states, "*As appropriate, the NTP also discusses information about outcomes from evidence streams not used in reaching a final hazard identification conclusion...*" It is unclear what this means but it is concerning that additional data would be brought in at this point - the benefit of a systematic review could be lost if it devolves into an expert narrative review at the final moment of its conclusion. Anything relevant should be in the protocol and PECO question - the rest is not relevant to the hazard statement.

Our specific comments are as follows:

## Introduction

- How are exposure levels among the population integrated into the decision? In the Navigation Guide that would be in Step 4 – grading recommendations for prevention. As risk management is not in the purview of NTP, and in a systematic review the hazard is defined independent of the population exposure level, do you mean to say that is how NTP would choose a substance for review? We recommend that NTP’s guidance for determining “level of concern” conclusions be developed and provided *a priori* in the protocols.
- *Re: “The systematic review format provides increased transparency”.* What is this increased transparency compared to? It might be beneficial to very briefly state the method (or lack thereof) that NTP’s new approach is replacing.
- Near the start of this introductory paragraph it might help to include a couple of sentences on why systematic review and meta-analysis are useful. The necessity of a systematic approach is far more obvious, but in our experience there is far less understanding of the value of the meta-analysis/data synthesis component of a systematic review, so making it clear at the start why this is being done and how it is useful might help to clarify this for the public.
- *Re: “The method for data synthesis includes steps...”.* We think there are two things missing here: first, “assessment of study quality” is a little ambiguous. It might be clearer to say “assessment of study reporting and risk of bias”. Second, in any meta-analysis it is important to take into account sources of heterogeneity; therefore it would be helpful to include another step: “explore the impact of study design characteristics and potential sources of bias on the estimates of [health effects/outcome/effect size]”. Additionally, prior to this it would be useful to explain why this is being done, for example “evidence from different streams of evidence (human, in vivo and in vitro) all have value in determining the effects of environmental chemicals on human health outcomes. Therefore our approach is to assess confidence within each individual stream of evidence using data-synthesis methods and then to integrate across evidence streams to reach hazard identification conclusions (Figure 1)”.

### Step 1: Prepare Topic

- We would note that a strength of systematic reviews in the clinical arena is that there is the capacity to develop “individual-relevant” questions - this is echoed in the National Academy of Sciences publication *Science and Decisions* in terms of scoping (National Research Council, 2009) - this is a very important feature of NTP’s proposal - that the exposed population can have input on what question is even asked *before* the assessment is conducted.
- There are inconsistencies when labeling sections under eligibility criteria as well as differences in defining these criteria between the PFOS/PFOA and BPA case studies, without any explanation. For example, the eligibility criteria for the human/animal participants are very different between the two protocols, and it is not clear why these were selected differently.
- In the 6 point list, there is some information potentially missing:
  - Step 3: It is not clear what “grouping” means here. Perhaps “identification” is a better term to use?
  - Step 4: *Re: “data extraction elements”.* This needs a bit more elaboration. Perhaps “the data to be extracted from each relevant study”?
  - Step 6: It is not clear whether you mean to take into account all of the streams of evidence separately here. In which case it would make more sense to say “evaluation of confidence in each body of evidence (human, in vitro, in vivo)....”

### Step 2: Search for and Select Studies for Inclusion

- Mention that the comprehensive literature search will include carefully selected search terms to maximize the sensitivity and specificity of the search.
- In the NTP protocol the term “snowballing” refers only to scanning the references of each included article; we recommend also searching for other articles that cite the included articles as this targets more current studies.
- *Re: “If a study that may be critical to the evaluation..”* Who/what determines whether a study is “critical” - do you mean one that meets the inclusion criteria?
- *Re “If a study that may be critical to the evaluation has not been peer reviewed, the NTP will have it peer reviewed.”* What is the need for or purpose of this peer-review step? NTP will not be relying on any of the authors conclusions but will be making an independent assessment of risk of bias of the study by virtue of doing the systematic review - isn't peer review in this situation “expert opinion” and at a minimum redundant? Of serious concern is that having to include this step can also cause unwarranted delays - what is the purpose of setting such a high bar for inclusion in the review? If a study meets the inclusion criteria the review itself will reveal what the data do and do not say in a transparent manner.
- *Re: “protocol establishes criteria for including or excluding references”.* It is clearer to say that the criteria are based on the PECO statement - that is how the inclusion/exclusion criteria are established - and it also points to why a well-formulated PECO statement is critical - it is the driving force behind all the subsequent steps.
- The sentence “If major limitations in a specific study type or design for addressing the question are known in advance (e.g., unreliable methods to assess exposure or health outcome), a basis for excluding those studies may be described a priori in the protocol.” We recommend care be taken when *a priori* excluding certain types of studies from the body of evidence. This opens up avenues to bias in the inclusion/exclusion of studies, and seems unnecessary because the “major limitations” of studies would be evaluated and revealed during the review (risk of bias).
- *Re: “the protocol also outlines the specific plans for...”* Again, we would relate whatever measures you want to include to your PECO statement - things that do not meet the standards will fall out of the inclusion criteria.
- *Re: reference screening,* you will need a referee if the 2 people cannot agree. Also describe the process for settling disagreements; similar to what is outlined for screening studies for eligibility.

### **Step 3: Extract Data from Studies**

- We recommend that as standard practice NTP utilize two independent data extractors for all the data as a worthwhile quality assurance/control procedure.
- Wherever possible, we recommend utilizing drop-down menus for multiple-choice answers for as many of the data fields as possible. This will minimize user input errors, as well as encourage *a priori* decisions about the data.
- We recommend that NTP add more detail for contacting study authors for the data and make this a standard practice in the OHAT method. In our application of the Navigation Guide method we have documented that a challenge to conducting risk of bias assessments and quantitative analyses is that all the necessary data were not reported in the published studies. Our efforts to contact study authors were critical to our ability to conduct the review. We anticipate that contacting study authors will be a necessary step for conducting systematic reviews until such time that steps are undertaken--by journals, funding agencies and through study registries---to standardize optimal reporting. Our findings of our case study of applying the Navigation Guide underscore the urgency of calls for improved access to the data needed to conduct scientifically robust reviews of environmental health science (Goldman and Silbergeld 2013), and the importance to environmental health of nascent efforts in the pre-clinical arena to develop improved experimental animal study design and reporting (Landis et al. 2012; van der Worp and Macleod 2011; Vesterinen et al. 2011).

- Public availability to the data is excellent and a huge step towards transparency - NTP should be highly commended for this breakthrough in government science.

#### **Step 4: Assess the Quality of Individual Studies**

- Study quality is not the same as risk of bias and in our experience this can be very confusing to people -we recommend saying that individual studies will be assessed for risk of bias and explain what that is - and stress that it should not be confused with quality or reporting - you can have a very high quality study that is biased - Cochrane points out you can have multiple studies that say the same thing but due to bias it is not a true finding.
- *Re: final paragraph:* All individual studies need to be assessed for risk of bias - we recommend saying that risk of bias tools for mechanistic and *in vitro* studies are important and have yet to be developed - in these cases applicable risk of bias tools will be adapted and stated a priori in the protocol.
- There is a large body of empirical evidence in the clinical sciences that supports that conflict of interest, e.g. funding sources, should be evaluated as a risk of bias and included in the risk of bias questions in protocols. We highly recommend that NTP include conflict of interest as a risk of bias domain as a standard practice.
- The BPA protocol says “we will not use ToxRTool to assess the internal validity of *in vitro* studies,” but no such statement is in the PFOS/PFOA protocol which makes it seem like NTP is open to using ToxR for assessing *in vitro* risk of bias. This is not appropriate because the ToxRTool is not a tool for risk of bias it is a tool for evaluating reporting.

#### **Step 5: Rate the Confidence in the Body of Evidence**

- Maybe use “body of evidence” throughout this section as it is occasionally confusing whether individual studies are being evaluated versus when a body of evidence is being evaluated.
- *Re: first sentence;* It is not clear whether this is referencing each stream of evidence or the entire body of evidence (human, *in vivo* and *in vitro* combined). Additional clarification is needed.
- *Re: “When other relevant data are necessary to address the question of the evaluation...”* It may be clearer to say “for each body of evidence included in the evaluation, the specific definitions used to rate the confidence in the body of evidence will be defined *a priori* in the protocol” - as written it might be misinterpreted as some other data rather than a body of evidence.
- *Re: second paragraph* “Conclusions developed in the subsequent steps of the approach are based on the evidence with the highest confidence.” We understand this to mean, and strongly agree, that you would not be “diluting” highest confidence evidence with lower confidence evidence. That is consistent with methods of evidence-based medicine. However, at the same time, there may be circumstances when conclusions will be informed by evidence for which there is lower confidence. For instance, another less-confident body of evidence may support the higher confidence body of evidence and may contribute to the conclusion. We recommend that the intent of what you appear to be saying be more clearly explicated.
- *Re: “A single, well conducted study may provide evidence of toxicity or a health effect.”* We strongly agree with this important point that is consistent with current best practices in environmental health research synthesis. For example, USEPA’s minimum criteria for animal data for a reproductive or developmental hazard is data demonstrating an adverse reproductive effect in a single appropriate, well-executed study in a single test species. ( U.S. EPA, 1996; U.S. EPA, 1991)

- *Re: "if a sufficient body of very similar studies is available..."* This might require some careful consideration to eliminate subjectivity. Firstly, how will you determine that studies are sufficiently similar? There is limited value in an overall estimate of effect if the entire body of included evidence is similar, but all of poor study quality and design. Some scientific judgment is required, but in general it is acceptable to conduct a meta-analysis on studies which are not very similar. Reasonable scientific judgment could be to say "we will perform a separate meta-analysis for all streams of evidence on only our primary outcome measures". I.e. obtain an overall estimate of effect for fetal weight in humans; or gestational age in animals. However, in combining different studies, it is important to consider heterogeneity and explore its potential sources (e.g. using multivariate meta-regression). It is a way of demonstrating empirically the impact of study design characteristics and potential sources of bias which may increase confidence in the overall conclusions, if for example you can show that there is still an effect in higher quality studies which have study designs which are deemed scientifically appropriate to answer the question (e.g. for animal studies was the time of exposure relevant to humans?). Another reason to avoid excluding studies which are dissimilar is that if you can demonstrate a consistent exposure effect across these studies (i.e. homogeneity), overall confidence may be increased. Examples of this type of analysis in the preclinical domain can be found in the work by Macleod et al ([www.camarades.info](http://www.camarades.info)).
- *Re: "Finally, confidence conclusions are developed across multiple outcomes for those outcomes..."*: It is not clear what you mean by this - GRADE recommendations are to consider each outcome separately.
- *Re "judgments are "inherently" subjective";* It may be more accurately described as there is judgment involved in establishing the level of proof you will be requiring for the ultimate confidence rating - the judgment is in where and how to set the bar for the evidence - and people may disagree about that, depending on whether it is a public health/precautionary decision or some other criteria. But that is not "inherent" - it is simply that judgment will be applied in a transparent manner.
- *Re section on page 3 "An initial confidence rating is determined by the ability of the study design to address causality as reflected in the confidence that exposure preceded and was associated with the outcome".* "Causality" as an evidence bar is overstating what the evidence in environmental health can provide - we recommend you stick with "association" as you explain later in the document that causality is implied by an association.
- *Re: "observational animal studies".* Maybe explain that these are wildlife studies and the importance of these observations to human health.
- *Re: "if the decision to downgrade is borderline for two properties..."* It might be clearer to say that you are not adding the up and down grades but looking at what are the key drivers of the strengths or weaknesses of the body of evidence and up and down grading accordingly.
- *Re: "There is empirical evidence that studies with negative results (no association)..."* Perhaps mention that this is in clinical sciences and that the direction of publication bias may be the opposite in environmental health.
- *Re: section "Upgrade confidence rating";* maybe specify that upgrading only refers to observational studies, not experimental studies (i.e., studies which meet the 4 criteria for starting at "high" quality).
- *Re: "Large magnitude of effect".* We recommend that whenever possible you specify in the protocol on a case by case basis what you will consider a "large magnitude of effect." In our experience, this will make the evaluation of the data much smoother.

- *Re: "Dose-response"*. As with large magnitude of effect, whenever possible, we recommend that you specify in the protocol on a case by case basis what you will consider to be a demonstrated dose response suitable for upgrading.
- *Re: "consistency"*. Do you mean consistency for the body of evidence being rated - i.e., human or animal, etc. ? It might be helpful to state that as it could be misconstrued that you are looking across animal and human studies here.
- *Re: section "Combine confidence conclusions for all study types and multiple outcomes"*. Here the approach is to only base conclusions on the evidence with the highest confidence; however we would recommend evaluating whether there are differences between evidence of high and low confidence.
- Contrary to our reading of the previous section, it seems here that the method to upgrade and downgrade evidence is based on the summation of the individual factors. We would not recommend this step being strictly quantitative, rather, as above, these factors should be more informative for upgrading or downgrading the evidence by one or two levels based on the key drivers of the strengths or weaknesses of the body of evidence. In addition, summation of the individual factors may lead to double counting. Moreover, the quantitative process itself is not very clear. The process for translating individual studies' scores for directness into whether and by how much to downgrade the confidence rating is not discussed, nor is the process for evaluating publication bias.

#### **Step 6: Translate Confidence Ratings into Level of Evidence for Health Effect**

- *Re: "evidence of no effect"*. We suggest adding that the conclusion is only for that specific health outcome, allowing for situations where there may be evidence of no health effect for "X" outcome but that does not preclude that there may be an effect for another separate health outcome.

#### **Step 7: Integrate Evidence to Develop Hazard Identification Conclusions**

- The step of combining from the human and non-human evidence streams is a valuable exercise. It is a very simple and logical method. It may be challenging, at least to begin with, simply due to the wide range of different *in vitro* and mechanistic studies and working out what is directly or indirectly relevant and equally, what is not relevant or simply bad science. However it is a very worthwhile venture to try to figure out these steps as there is undoubtedly a lot of value in some of these studies. Working out their relevance to humans also has a number of very important cost-reduction and ethical implications, i.e. could an *in vitro* study replace certain animal models? However we have strong reservations about the consideration of using *in vitro* studies to downgrade hazard ID conclusions in the presence of human data that otherwise indicate a certain level of evidence.
- *Re: "...with consideration of special situations related to exposure information that may apply..."* We are not clear what that means - since NTP states explicitly elsewhere in the document that exposure is considered *after* hazard ID.
- *Re: "Note that mechanistic data or another type of other relevant data is not required to reach a final hazard identification conclusion"*. We strongly support this - i.e., that you do not need mechanistic or other type of relevant data to make a hazard ID.
- *Re: "If other relevant data provide strong opposition for biological plausibility of the relationship between exposure..."* We strongly disagree with a step for downgrading at this point - It will be difficult to define what is "*strong*" *opposition for biological plausibility* that has not already been accounted for in the risk of bias evaluation and other earlier rating of the confidence of the body of evidence - this additional step allows for a double counting down - if you have moderate data after

all the previous steps that should not be reduced by a downgrade here --- the purpose of the upgrade at this point is to account for additional *supporting* evidence, not additional contradictory evidence- downgrading here leaves open raising the evidence bar rather than making an informed decision based on the available evidence --- while this step appears on a superficial level to be "balanced" with the opportunity for an upgrade it really creates an imbalance as it allows you to downgrade what has already been determined to be moderate data.

- *Re: "As appropriate, the NTP also discusses information about outcomes from evidence streams not used in reaching a final hazard identification conclusion..."* It is unclear what this means but concerning that additional data would be brought in at this point - the benefit of a systematic review could be lost if it devolves into an expert narrative review at the final moment of its conclusion. Anything relevant should be in the protocol and PECO question - the rest is not relevant to the hazard statement. The inclusion and exclusion criteria are explicit - so the reasons why these data were not used should be in the protocol - it does not matter whether they are or are not supportive - it only matters whether they are pertinent to the study question.

#### Figure 1:

- Step 5: Not assigning a particular confidence rating based on particular study design (RCT, cohort, case-control, etc.) is a good idea. However, the system of plus signs (body of evidence receiving 4 pluses if it has all 4 "key features") looks as if you are adding up the confidence ratings, and giving equal value to each feature. But if a body of evidence does not include "exposure prior to outcome," then it would seem that it should be at best "low confidence," and not be possible to rate as equal to a body of evidence that is only missing the feature of "controlled exposure," i.e. a group of observational studies.
- Steps 5 + 6: It is not clear why "high" confidence in Step 5 has 4 plus signs but "high" confidence in Step 6 has only 3 plus signs. Are the plus signs really necessary, or can you just use the words "high, moderate, low?"

#### Appendix A:

- *Re: important confounding;* what criteria will be used to determine which potential confounders are the important ones on which to judge risk of bias?
- *Re: performance bias;* "adjust or control for other exposures" sounds like confounding too. For epidemiology studies, it seems difficult to assess whether a study adequately adjusted for "other exposures." In reality, most probably don't because we don't know what they are.
- *Re: detection bias;* By saying that the time window for exposure must be "prior to the outcome," this could rate a study that for example was on birth weight and measured maternal serum levels of a contaminant a few days postpartum with higher risk of bias than a study that measured levels a year before birth. It should be clear that the exposure measurement can represent an exposure for a different time period than the one where it is actually measured if there is data to support that a proxy measure can be used. Also, how will you determine if an exposure characterization is "reliable?" Will there be specific criteria?

#### Appendix B

- It appears that a body of evidence that relied on "ecologic" studies might be downgraded at the individual risk of bias level (for exposure characterization) and at the confidence in the body of evidence level (not having individual outcome data). Would this "double dinging" for attributes of study design be allowed, because the downgrading relates to exposure and outcome separately?

**Other comments:**

We think it is also important to consider study reporting, specifically factors which are independent from risk of bias, but relevant to an overall assessment of study quality. For example, in the preclinical domain it is unfortunately quite common for studies to fail to report the number of animals per group, attrition rates, the variance or what the measure of variance is (e.g. SD or SEM; see Landis et al, 2012). Other aspects of study design may be important, but not reported, e.g. the route of administration, the method used to stain histological samples etc. The reporting of these can probably be inferred once the data extraction has been completed, but it could be helpful to actually document study reporting and highlight if any of these are areas of concern.

In summary, PRHE's scientists strongly support NTP's commendable efforts, welcome collaboration with NTP scientists on improving methods of systematic and transparent review in environmental health and wish the NTP great success in its endeavor. We hope that our detailed comments will help clarify and sharpen NTP's documents and the application of its method.

Sincerely,

Tracey J. Woodruff, PhD, MPH

Professor and Director,

University of California, San Francisco, Program on Reproductive Health and the Environment

Hanna M Vesterinen, PhD.

Postdoctoral Research Fellow

University of California, San Francisco, Program on Reproductive Health and the Environment

## References:

Goldman, LR. & Silbergeld, E. 2013. Access to chemical data used in regulatory decision making. *Environmental Health Perspective*. 121(4): A111-2.

Higgins, JPT. & Green, S. 2011. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).

Kousta EL, *et al.* 2013. The Navigation Guide as an Evidence-Based Medicine Methodology to Evaluate Human Health Effects of Environmental Chemicals: Perfluorooctanoic Acid (PFOA) and Fetal Growth. Under review at *Environmental Health Perspectives*.

Landis, SC., et al. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 490: 187-191.

Lundh A, Sismondo S, Lexchin J, Busuioac OA, Bero L. Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews* 2012; 12:MR000033. [dx.doi.org/10.1002/14651858.MR000033.pub2](https://doi.org/10.1002/14651858.MR000033.pub2)

Johnson, et al. 2013. The Navigation Guide - Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth. Under review at *Environmental Health Perspectives*.

National Research Council. 2009. *Science and Decisions: Advancing Risk Assessment*. The National Academies Press, Washington, D.C.

U.S. EPA. Guidelines for Reproductive Toxicity Risk Assessment. U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC, 630/R-96/009, 1996.

U.S. EPA, Risk Assessment Forum. Guidelines for Developmental Toxicity Risk Assessment. Washington, DC, EPA/600/FR-91/001, 1991.

van der Worp, HB. & Macleod, MR. Preclinical studies of human disease: time to take methodological quality seriously. *J Mol Cell Cardiol*. 51(4): 449-450.

Vesterinen, HM., et al. 2011. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the *Journal of Cerebral Blood Flow and Metabolism*. 31(4): 1064-1072.

Viswanathan, M., Golin, CE., Jones, CD., et al. 2012. Closing the quality gap: revisiting the state of science (Volume 4: Medication adherence interventions: comparative effectiveness). Rockville (MD): Agency for Healthcare Research and Quality.

Woodruff TJ, Sutton P. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Aff (Millwood)* 30:931-937.