Lancaster Environment Centre | Lancaster University

Comments on the draft BPA hazard assessment protocol

14 September 2017 | Four Points by Sheraton, Brussels

Paul Whaley | Lancaster Environment Centre

p.whaley@lancaster.ac.uk

About me

- Research into methods for conducting high quality systematic reviews and maps at Lancaster Environment Centre, UK
- Associate Editor for Systematic Reviews at *Environment International* (IF 7.088)
- Co-Chair of evolving working group on publishing standards for the Evidence-Based Toxicology Collaboration
- Promoting best practices in SR methods, e.g. involved with WHO Chemical Risk Assessment Network, NGOs, EFSA

“Science is supposed to be cumulative, but scientists only rarely cumulate evidence scientifically.”

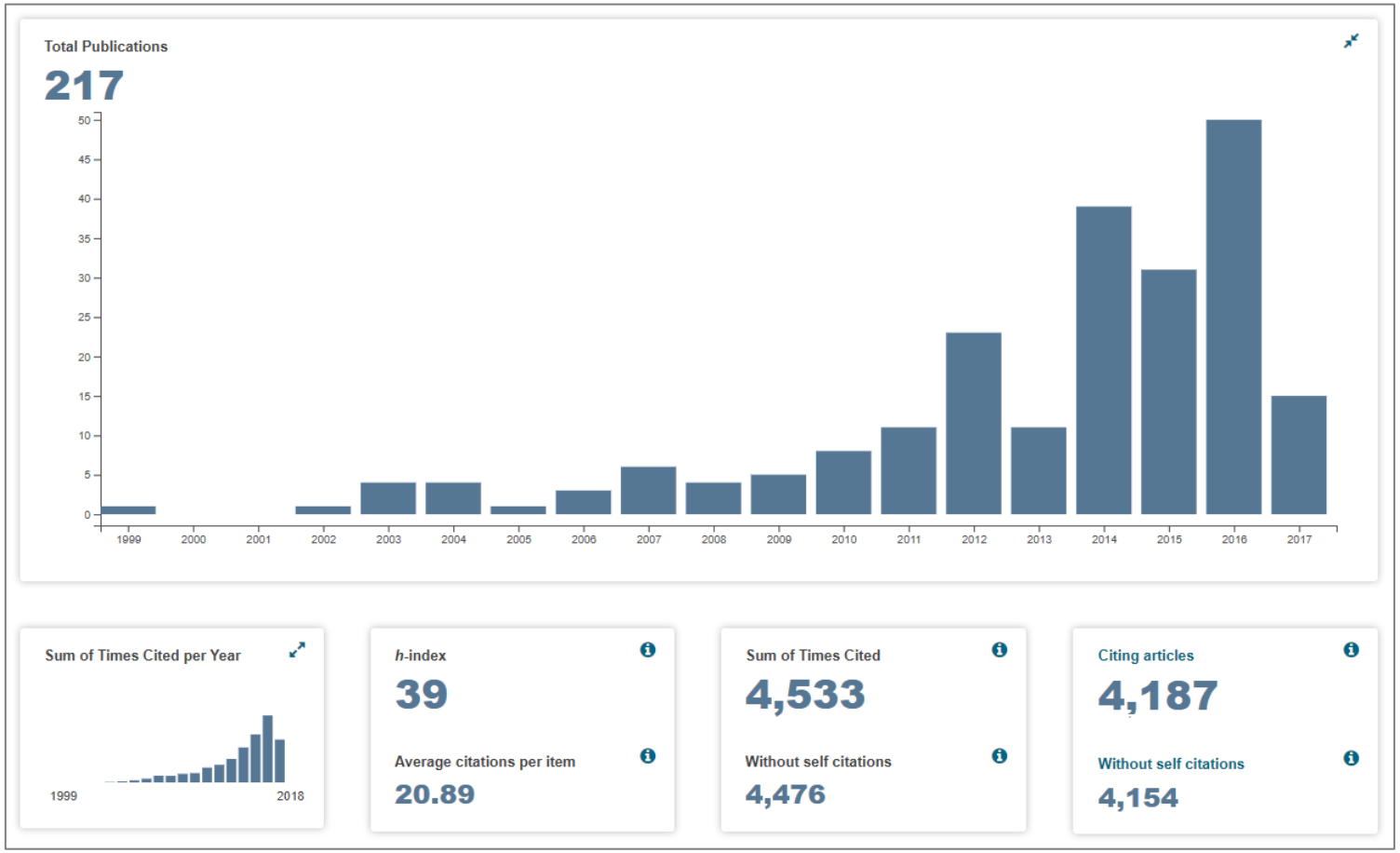
- Mulrow (1987): only 1 article of 50 in the top 4 medical journals had used scientific methods to identify, appraise and synthesise information
- Chalmers and Mosteller (1992): textbook advice on treatment of heart attacks lagged findings of clinical trials by 10 years

Chalmers, Hedges & Cooper (2002)

Systematic review methods

- Apply scientific method to the problem of summarising evidence
  - Follow a pre-specified protocol
  - All relevant evidence is found and included
  - Appraise all the evidence for risk of bias
  - Appropriate statistical and qualitative techniques to generate summary results
  - Transparent, comprehensive documentation of all decisions
- Same standard for reviewing evidence as for generating it

SRs increasingly common in EH research



Papers indexed in Web of Science (WoS) with the term “Systematic Review” in the publication title, filtered for “Toxicology” as topic, excluding topic of “Pharmacology Pharmacy”. Search date: 16 August 2017

The problem with the prestige of SR methods

- 8989 PubMed records tagged by 2004 as “systematic review” yet actual number of stringently-defined SRs was ~2500 (Moher et al. 2007)
- Most published SRs have major flaws in conduct and reporting (Page et al. 2016)
- ~3% of manuscripts are “decent and clinically useful” (Ioannidis 2016)
- Our own pilot data shows serious omissions in reporting of 19 of 25 SRs published in the top environmental health journals through 2014-2015

Why this matters

Flawed reviews will be mistaken for gold-standard research, resulting in:

- environmental health challenges being misidentified
- subsequent policy being based on incorrect interpretations of the available evidence
- ultimately the value of systematic review in decision-making being undermined

Comments on the protocol

The sort of feedback I would give as a systematic review editor

Summary

- Protocol is timely and best practice
- Ambition to use systematic methods is challenging and laudable
- Some major issues
  - » Use of results of previous assessments conducted with different methods
  - » Exclusion of studies on basis of design (cross-sectional and single dose)
  - » Unclear relevance assessment
  - » Unorthodox two-step assessment of study quality
  - » Unclear approach to assessing weight of evidence
  - » Absence of a plan for meta-analysis
- Has the protocol fallen between two stools?
- As an editor, would expect a lot of work before publication

The use of previous assessments (line 187)

- Previous assessments could have produced different results if they had been conducted according to systematic methods
- It will introduce a bias to the starting point to unquestioningly use the previous conclusions
- **Fix**: Start fresh, assessing all evidence with the same, robust methods

Cut-off dates for search (265)

- Sensitive search methods are very good to see
- Cut-off date of after previous reviews only makes sense if previous searches were fully comprehensive, AND all previous evidence is included in the new review
- **Fix**: Include all studies regardless of date of publication

Exclusion of cross-sectional and single dose studies (384)

- Cross-sectional and single-dose studies have obvious limitations
- Inclusion of and assessment of these studies in combinations with the rest of the evidence can overcome some of these limitations
- To have limitations is not to be of zero value; enough studies of these types could contribute to a signal on BPA toxicity
- We cannot know in advance if this is the case; therefore *a priori* exclusion of this evidence is not justified
- **Fix**: Do not exclude studies on basis of study design; include everything which is relevant

Assessing evidence for relevance (422)

- Not clear about the role of the apparent two-step relevance appraisal process in the hazard assessment
- At this stage, studies occupy a spectrum of relevance: not binary, not captured by yes/unclear/no
- The binary judgement happens at the initial inclusion phase, during screening of the literature where relevance is unambiguous
- Position on that spectrum determines weight they are accorded in the final analysis ("Do we believe these results are representative of what is happening in humans?")
- Exclusion at this stage is inappropriate, should be weighted instead
- **Fix**: Maybe just needs clearer explanation?

Internal validity assessment (446)

- Not clear how “quality” is differentiated from “risk of bias”; what quality constructs are not in the risk of bias assessment?
- Why is quality assessment a two-step process?
- Organising studies into tiers of validity appears to combine non-comparable quality constructs into a single judgement, which could lead to incorrect weighting of study limitations
- Not clear what it means to “integrate” the results of the NTP risk of bias assessment approach into single judgement of reliability
- **Fix**: Simplify by using standard one-step risk of bias approach

Weight of evidence approach (628)

- This is where the protocol begins to feel less like documentation of how decisions will be made, than it does a description of when subjective, opinion-driven processes will be used
- Not necessary: GRADE methodology already provides a framework for weighing the strength of the evidence when determining confidence in a summary result
- **Fix**: Clearer description of the factors to be used in weighing evidence; structure using GRADE-like approach

Lack of plan for meta-analysis

- Meta-analysis can be very powerful for providing pooled estimates of effect and is normally the foundation of a compelling systematic review, as it gives the hard numbers in a summary result
- Can analyse the result for heterogeneity, precision, conduct subgroup analysis with the best-quality studies, etc.
- Can transform understanding of the strength of the evidence
- Conditions under which meta-analyses should be conducted are part of the protocol of any systematic review aiming at a quantitative result; yet the plan here seems to be not to do one

Falling between two stools?

- Doesn’t seem to be consensus in the review team about the use of systematic methods
- Rather, there is a combination of traditional, narrative techniques and systematic approaches
- Problem is, being systematic is like being pregnant: you can’t really be just a bit systematic
- Not at all clear how the unconventional, non-traditional approaches to e.g. quality assessment increase rather than detract from the validity of the overall assessment