Systematic Reviews, Machine Learning and the Liberation of Knowledge from Information in Environmental Health Research

Paul Whaley
Evidence Based Toxicology Collaboration Research Fellow
Lancaster Environment Centre, UK

Environmental Health Collaborative 2018 AI Summit, 19 October 2018
About me

- Researcher at Lancaster University and the Evidence-Based Toxicology Collaboration at Johns Hopkins Bloomberg School of Public Health
- Background in environmental health NGO advocacy and science communication, now working in chemical risk assessment, primarily around developing and advocating use of systematic methods in chemical risk research
- Associate Editor for Systematic Reviews at Environment International (IF 7.297)
- The “frameworks guy” in systematic review methods for environmental health research: systematic approaches to evidence surveillance and synthesis; critical appraisal tools; codes of practice; quality assurance and control
- Not a computer scientist, moving in the direction of machine learning anyway
What I’m going to talk about

• Winning the argument about using systematic review methods, but...
• ...neglecting to mention the data volume problem
• Machines should read scientific documents into graph databases
  • Even simple graph databases are pretty neat: e.g. data-driven AOPs
  • Large databases are really neat: chemical and disease signatures
• How non-computer-scientists can help machines learn to read
• Pay-off: we can capture the sum total state of human knowledge
Systematic review methods in chemical risk assessment
The data volume problem
Graph databases
What non-computer scientists can do
The pay-off
Reproducibility crisis in primary research

1,500 scientists lift the lid on reproducibility

Estimating the reproducibility of psychological science

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effort.

The Reproducibility Project: Cancer Biology is a collaboration between the Center for Open Science and Science Arena to independently replicate selected results from a national panel of researchers.
Reproducibility crisis in risk assessment?

Bisphenol-A
Bisphenol-A and impaired fertility

Wisniewski at al. 2015
Lassen et al. 2014
Tiwari & Vanage 2013
Peng et al. 2016
Rahman et al. 2017
Martínez-Peña et al. 2017
...no health concern for any age group from dietary exposure
EFSA 2015

...effects have been demonstrated for BPA [at] levels 10–10,000x lower than the current LOAEL of 50 mg/kg/day
Vandenberg et al. 2014

...a potential risk to the unborn children of exposed pregnant women [relating to] a change in the structure of the mammary gland
ANSES 2013

...a TDI for BPA has to be 0.7 μg/kg bw/day or lower to be sufficiently protective
National Food Institute, Denmark 2015
Same evidence, different conclusions

...no health concern for any age group from dietary exposure
EFSA 2015

...a TDI for BPA has to be 0.7 μg/kg bw/day or lower to be sufficiently protective
National Food Institute, Denmark 2013

...effects have been demonstrated for BPA [at] levels 10–10,000x lower than the current LOAEL of 50 mg/kg/day
Vandenberg et al. 2014
What can we do about this?

• Medicine has already seen this problem and come up with a solution
• We can borrow that
Infants born prematurely are at increased risk of life-threatening respiratory distress syndrome.

1970s: Could risk be reduced by giving a dose of steroids to women expecting to give birth prematurely?
Steroids and premature birth

- 1972: large trial shows steroids reduce mortality in premature birth
Steroids and premature birth

- 1977-1993: More trials, small and individually unconvincing, lots of differences between studies
- Divided expert opinion
Steroids and premature birth

- 1994: All studies aggregated in a systematic review, showing clear benefit
- 22 years wasted: we should already have known
- Unethical to conduct unnecessary studies
Success of systematic review in medicine

• Major role in resolving a lot of debates and uncertainty about what the evidence says about the effectiveness of medical treatments

• Systematic reviews are the most-cited type of research in the medical literature and form the evidence base for medical guidelines worldwide

• Can we apply SR methods and get the same benefits when assessing health risks posed by chemical substances?
What does it mean to be systematic?

• To use transparent, reproducible methods for reviewing what existing evidence says in answer to a research question, in a process which minimises the risk that the results of the review will be biased
Elevating the literature review to the status of a science
Three hallmarks of the scientific method

• Transparency: document everything (exhaustively)
• Reproducibility: two different research teams should be able to get the same results (or if not, at least be able to explain why not)
• Truthfulness: results should be unbiased (as free as possible from systematic error)
Three sources of bias in reviewing evidence

• Bias arising from flaws in the design, conduct, analysis and reporting of included studies being transmitted through to the results of a review (bias from limitations in the evidence)

• Bias due to systematic differences in results between the retrievable and irretrievable evidence (publication bias)

• Bias from conduct of the review itself, e.g.
  • Selective use of evidence (using part rather than all of the evidence base)
  • Selective interpretation of the evidence (seeing what you want / expect)
Systematic methods help prevent bias

• Pre-planned protocol defining review methodology
• Comprehensive search strategy
• Screening search results for relevance against objective criteria
• Comprehensive data extraction
• Critical appraisal of the included studies (risk of bias assessment)
• Valid qualitative and quantitative methods for synthesis
• Valid methods for interpreting confidence in results
Systematic methods compare favourably with expert-led approaches
Uptake of SR methods in chemical risk and environmental health research

- 2008: Arguably first mooted by Hartung and Hoffmann (EBTC)
- 2015: First journal Special Issue dedicated to SR methods in CRA
- 2016: First specialist SR editor at an environment health journal
- 2017: Next WHO/ILO Global Burden of Disease estimate to be based on 18 SRs, with pre-published protocols; EFSA bases a risk assessment (BPA) on SR methods for first time
- 2018: WHO protocols published; US EPA and GRADE Special Issues initiated; second editor
- SR described in EFSA, ECHA, NTP, EPA, TCEQ guidance. In legal text for identification of EDCs in EU. NGOs, agencies, industry and academia all support these methods.
Systematic review methods in chemical risk assessment

The data volume problem

Graph databases

What non-computer scientists can do

The pay-off
We have to review a lot of research

• It takes about 18 months to systematically review a few dozen studies
  • Planning
  • Searching and screening
  • Extraction of relevant data
  • Reporting
• Only small systematic reviews are feasible
  • Focus on a single exposure/outcome pair to keep data volume down
• Thousands of chemicals need assessing with systematic methods
• Data volume problem gets worse as in vitro testing is mainstreamed
The problem with focus

• Excludes relevant evidence e.g. BPA+1, which seems silly
  • It is relevant, just indirectly so: we know structural similarities between chemicals can inform risk estimates
  • But including BPA+1 increases evidence by orders of magnitude

• Risk management questions are rarely so focused
  • Need comprehensive view of evidence relating to potential health risks
  • Many health end-points, mixed exposures, etc.
Computers will have to read for us

- Managing the data volume problem by excluding it
- To maintain systematic standards and take advantage of all the data we already have and generate every day, we need to hand over the reading of documents to computers
Systematic review methods in chemical risk assessment
The data volume problem
Graph databases
What non-computer scientists can do
The pay-off
Data infrastructures for machine reading

• Currently, summarise studies as tables, or as tables with primary keys and relationships (relational DBs)

• Requires us to figure out how to represent all the information in a document in a relational scheme. Hard work.

• Also, small number of relationships; crudely summarised to get key queries working, but a poor imitation of real complexity in the data

• Since how things are related is something we are discovering all the time, relational databases are not a great engineering solution to storing data contained in scientific documents
There is a better way: semantic databases

- Most of what is going on in research can be represented or summarised in a finite set of subject-predicate-object triples
  - Rat group | IS DOSED WITH | BPA
  - BPA | CONCENTRATION IS | 5 mg/kgbw/d
  - Rat group | INVESTIGATED FOR | liver tumors
  - Liver tumors | SHOW | increase
- Graph databases are built direct from these triples
Adverse Outcome Pathways and graphs

- AOPs are increasingly important for predicting health outcomes from environmental exposures, connecting exposures and initiating events through to outcome via intermediate events.
Data-driven AOPs

• Networks of events which are incompletely presented or investigated in any given document

• A systematic, data-driven approach would be to read all literature about chemicals and events, and map the relations

• Not practical: Too big a job: thousands of events, hundreds of thousands of documents

• Machine learning makes the data-driven identification of AOPs possible, and at least positions the relevant data in situ even if it needs a human to analyse it
Eventually, chemical and disease signatures

- Lots of data atoms, and relations between them, creates a big, interrelated space in which chemicals start having signatures.
- Compare the signature of a PFAS to carcinogenicity; RA disappears as we know it, probabilistic assessment takes its place
Systematic review methods in chemical risk assessment
The data volume problem
Graph databases
What non-computer scientists can do
The pay-off
Prerequisites of effective, automated SR

• More supervised learning (annotated corpora), especially for tasks like recognising causal claims: requires domain expertise

• Stuff like changing publishing practices (what is it with tables?)
  • Remove needless impediments to making research machine-readable

• No point in automating the generation of biased results
  • Better reporting practices so we understand the quality of the input data (garbage in / garbage out: already a big problem in SRs, mega-problem when machines read studies into mega-databases
  • Validity of the methods we are automating, e.g. risk of bias, strength of evidence assessment in GRADE
Systematic review methods in chemical risk assessment
The data volume problem
Graph databases
What non-computer scientists can do
The pay-off
Humans are silly, part nine million

- We already have a database of all written knowledge, it just consists of millions of isolated documents unevenly distributed across millions of square kilometres of digital and physical space
- We query this database by looking for documents, reading them, trying to remember the bits that matter to us, and ignoring the bits that don’t
- Reports of these queries are more self-contained documents, which are added to the pile of documents which are unevenly distributed across physical and digital space, which someone has to read, etc. etc.
- Someone else with different information requirements then reads a bunch of these documents again, ignores and remembers different bits, produces a report which is another document, etc. and so forth
The pay-off

- ML and graphs allow a formal representation of state of human knowledge, not just a bunch of data points in isolated PDFs
- The machines take care of making data accessible; frees up humans to make best use of what we do know, and figure out how to find out what we don’t know. (e.g. documents around the data-driven AOP)
- As a systematic reviewer, it entirely puts me out of a job, of course. See you on the beach!
Thank you for listening!