Implementing machine learning methods in literature searching and screening to identify and categorize mechanistic evidence for the Integrated Science Assessments

Jennifer L. Nichols, Ryan Jones, Michael J. Stewart, and Steven J. Dutton

National Center for Environmental Assessment, US EPA, RTP, Durham, NC, 27711

Jennifer L. Nichols | nichols.jennifer@epa.gov | 919-541-0708

Background

The National Ambient Air Quality Standards (NAAQS) for criteria air pollutants are established by the EPA as mandated by the Clean Air Act to protect public health and welfare. A critical component in reviewing the NAAQS is the development of the Integrated Science Assessments that assess the state of the science informing the relationship between ambient exposures and a range of health and welfare effects. This requires that EPA scientists identify, evaluate, and integrate a broad range of scientific evidence, including mechanistic evidence that is an important aspect to understanding the nature of the relationship between exposure and effect. Because of the vast amount of evidence pertinent to ISAs from both observational and experimental studies, identifying and evaluating the relevant evidence in a comprehensive and efficient manner is a challenge. In the current review of the Ozone NAAQS, recently initiated under an accelerated timeline by directive from the EPA Administrator, methods and approaches for ISA development have been streamlined and modernized to further adopt systematic review methodologies, including for the identification and categorization of mechanistic evidence. As presented in this case study, a combination of machine learning and automated approaches in literature searching through the Health and Environmental Research Online database and the adoption of the SWIFT-ActiveScreener tool, have resulted in a substantially more efficient process to identify and categorize evidence. Ultimately, this streamlined workflow provides an easily adaptable process to effectively search and screen for mechanistic evidence that is a critical component to drawing conclusions in scientific assessments supporting key Agency policy decisions.

Literature Workflow for the Ozone ISA

Keyword search:
- Ozone or o3, 2011/01/01-2018/03/31
- PubMed, Web of Science, TOXLINE
- n = 30,797

Excluded
- Not peer reviewed, n = 883
- Not English language, n = 329
- Retracted, n = 7

References for Automated Topic Classification, n = 29,519
- Experimental seeds, n = 364
- Epidemiology seeds, n = 425

Not Topic Classified, n = 24,497

Automated Topic Classified References
- Experimental seeds, n = 3,298
- Epidemiology seeds, n = 1,724

Title/abstract Exclude
- Experimental n = 701
- Epidemiology, n = 343
- Not screened, n = 2,620

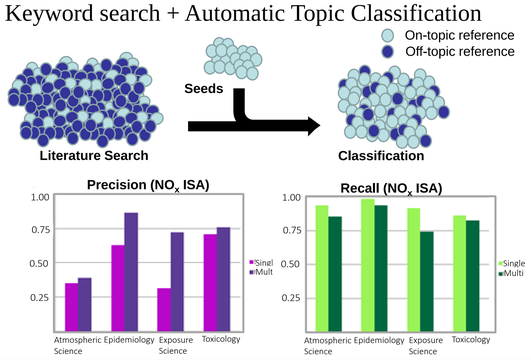
Title/abstract Include
- Experimental, n = 541
- Epidemiology, n = 820

Categorization of References
- Experimental: Health outcome, Exposure duration, Exposure concentration
- Epidemiology: Health outcome, Exposure duration, Location

U.S. Environmental Protection Agency
Office of Research and Development

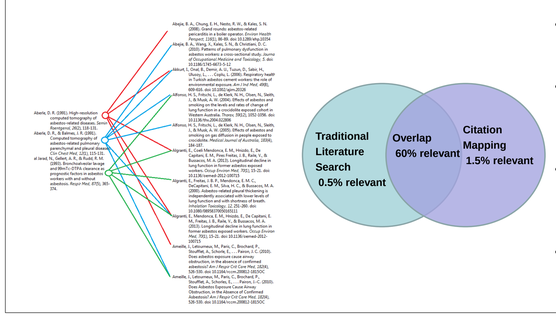
Novel Methods for Literature Identification: Topic Classification and Citation Mapping

Keyword search + Automatic Topic Classification



- Seed items from specific disciplines are used to train an algorithm
- This algorithm is applied to a set of search results and predicts the relevant discipline of each item
- Method is invaluable for large assessments where results must be distributed to specialized teams
- Results from analyses of precision and recall from application in the 2016 Oxides of Nitrogen ISA demonstrate the effectiveness of the automatic topic classification across scientific disciplines.

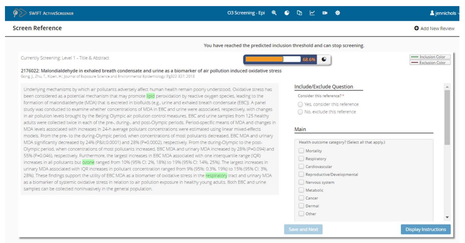
Citation Mapping

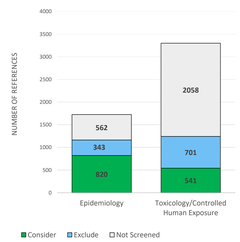


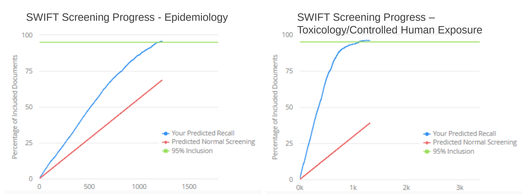
- Starts from a list of known relevant references
- Aggregates expert judgments of relevance, then ranks results by probability of relevance
- Not bound by metadata -- able to return relevant items that would have been missed by keyword searches
- Citations found in both the keyword search and citation mapping have a much higher probability of relevance.
- Citation mapping used to identify ecological references for Ozone ISA, but not health references.

Implementation of SWIFT-ActiveScreener for Title/Abstract Screening

- SWIFT-ActiveScreener (Sciome): title/abstract screening with machine learning in real-time based on exclusion/inclusion decisions
- Allows optional screening questions
- Predicts 95% threshold, comparable to human judgment
- Reduced number of references needed to undergo title/abstract screening
- Added efficiency from SWIFT-AS depends on specificity of the references obtained from the literature search.







Categorization of Evidence and Application of Evidence Maps

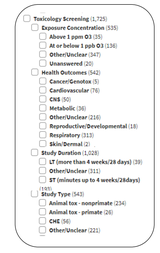
Experimental

Screening Questions

Title/Abstract Screening – Experimental (Animal Toxicology and Controlled Human Exposure

Seed references: n =
Screening questions:
1. Consider this reference (full-text screen)?
   - Yes, consider this reference
   - No, exclude this reference
2. Health outcome category?
   - Mortality
   - Respiratory
   - Cardiovascular
   - Reproductive/Developmental
   - Nervous System
   - Metabolic
   - Cancer
   - Dermal
   - Other
   - Unclear
3. Exposure duration?
   - Short-term (up to 28 days)
   - Long-term (more than 28 days)
   - Unclear
4. Study type?
   - Controlled human exposure
   - Animal tox - primate
   - Animal tox – nonprimate
   - Unclear

Tags in HERO Database



Evidence Map

Exposure	Study Type	Health Outcome Category: CV	Resp	Repro/ Dev	Metabolic	Nervous System	Cancer	Dermal	Other	Total
Short-term	CHE	16	40	0	1	4	2	0	9	72
	Tox - Nonprimate	25	9	7	17	23	0	0	17	185
	Tox-Primate	5	8	4	0	0	1	1	0	19
	Unclear	1	8	0	0	0	0	0	0	9
Long-term	CHE	0	0	0	0	0	0	0	0	0
	Tox - Nonprimate	6	23	1	5	12	0	0	7	52
	Tox-Primate	1	5	2	0	0	0	0	0	8
	Unclear	2	0	0	0	1	0	0	0	3
Unclear	CHE	0	0	0	0	0	0	0	0	0
	Tox - Nonprimate	7	62	1	7	9	1	0	23	112
	Tox-Primate	2	15	5	0	0	0	0	2	24
	Unclear	14	49	1	4	7	1	0	44	120
	Total	79	304	23	34	56	5	1	103	604

Refine Scope for ISA

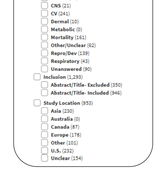
Epidemiology

Screening Questions

Title/Abstract Screening – Epidemiology

Seed references: n = 66
Screening questions:
1. Consider this reference (full-text screen)?
   - Yes, consider this reference
   - No, exclude this reference
2. Health outcome category?
   - Mortality
   - Respiratory
   - Cardiovascular
   - Reproductive/Developmental
   - Nervous System
   - Metabolic
   - Cancer
   - Dermal
   - Other
   - Unclear
3. Exposure duration?
   - Short-term (up to 28 days)
   - Long-term (more than 28 days)
   - Unclear
4. Study locations?
   - U.S.
   - Canada
   - Europe
   - Australia
   - Asia
   - Other
   - Unclear

Tags in HERO Database



Evidence Map

Exposure	Location	Health Outcome Category: Mortality	CV	Resp	Repro/ Dev	Metabolic	Nervous System	Cancer	Dermal	Other	Total
Short-term	US	36	36	38	11	2	0	0	5	11	133
	Canada	10	14	12	1	0	0	0	3	7	48
	Europe	33	17	17	6	2	5	0	1	5	126
	Asia	40	13	18	5	3	8	1	1	10	179
	Other/Unclear	21	8	15	5	1	1	0	0	2	148
Long-term	US	16	8	12	46	5	4	5	0	7	128
	Canada	5	1	4	3	1	0	1	0	1	16
	Europe	5	5	12	16	6	1	2	0	1	49
	Asia	10	17	19	25	4	4	4	1	6	90
	Other/Unclear	4	14	17	40	0	4	1	0	4	88
Unclear	US	3	1	1	1	1	1	1	1	1	11
	Canada	0	0	1	0	0	0	0	0	0	1
	Europe	0	3	4	2	1	0	0	0	2	12
	Asia	0	5	2	2	0	0	0	0	2	11
	Other/Unclear	0	1	2	2	0	4	0	1	1	11
	Total	183	141	174	163	27	32	15	12	60	1013

Refine Scope for ISA

Summary

- The evidence base for ISAs spans many disciplines and is vast. References identified from the literature search for recent ISAs range from approximately 50,000-500,000.
- Automatic topic classification and citation mapping are novel literature identification methods that are modules within the HERO Database. The approaches have successfully reduced the number of off-topic references that disciplinary teams of scientists need to screen for ISAs.
- Recently, SWIFT-AS was implemented in the Ozone ISA development process, adding another layer of machine learning capabilities and effectively reducing the number of references to be screened at the title/abstract level.
- SWIFT-AS optional screening questions were utilized to refine the categorization of references during title/abstract screening.
- For mechanistic evidence, experimental and epidemiologic studies were primarily binned by health outcome category and exposure duration to facilitate a more strategic and systematic consideration of the evidence by the assessment team.
- In future applications, this streamlined workflow could easily be adapted to allow for further refinement and categorization of mechanistic evidence based on assessment needs.