Integrated Computational Approaches in Systematic Evidence Mapping: Bridging Human Expertise and Machine Learning for Advanced Toxicological Assessments

I.A. Lea, J. Holt, E. Bunnage, M. Long, R. Chew, S. Bell, S. Edwards, R. Sayre, S.M.Vliet, S.G. Lynn, and M.J. Kristan

Monday, March 17, 2025
9:15 AM – 11:45 AM
3409/J540
Computational Toxicology I
Convention Center
W Hall A2

Abstract:

Background and Purpose: As the volume and complexity of scientific literature grow, the need for more sophisticated systematic evidence mapping (SEM) methodologies and associated tools has become increasingly apparent. Traditional SEM approaches, often constrained by predefined criteria and labor-intensive processes, may fail to capture the full breadth of relevant data, limiting their utility in emerging fields such as toxicology and environmental health. This abstract synthesizes findings from exploration of innovative methods to enhance SEM through use of computational tools (e.g., visualizations, machine learning (ML), and large language models (LLMs)). The overarching goal is to enhance the accuracy, efficiency, and applicability of SEM, providing a robust framework for future systematic reviews which can support regulatory decision-making.

Methods: The integration of computational tools to the SEM process was explored in a series of experiments utilizing data from a thyroid focused systematic evidence map. First, an evaluation of the accuracy of article labeling was conducted on 773 articles to identify features of abstract text that predict when evaluation of full-text is not necessary to determine under what conditions abstract-only labeling is a viable option. Second, the value of using generalized LLMs, specifically GPT-4, to categorize 636 full-text journal articles evaluated as part of the thyroid evidence map was assessed in a series of case studies. The LLM predictions were compared to the human curated labels applied to these articles to assess the model's proficiency. The results of these experiments were applied in the development of an updated SEM workflow.

Two approaches were used to develop the updated SEM workflow, both leveraged label information generated by human reviewers or computational tools. The first approach looked at how to group or cluster papers based on the title and abstract containing similar themes and concepts. This resulted in development of the LitMapper tool that uses dynamic visuals to explore the results of a literature search. The second approach sought to use explicitly defined labels generated as part of the evidence mapping process, as well as label context, to create dynamic cooccurrence networks. The tool developed from this approach, LitConnector, connects articles so that relationships between concepts could be explored. The tools, LitMapper and LitConnector, developed as part of the updated SEM workflow will be presented and examples of their use provided.

Results: The abstract-versus-full-text coding experiment showed that under certain conditions, abstract-only labeling could be a viable alternative, with explainable ML highlighting specific features that predict when full-text review is essential. In total, 45.6% (21/46) of inventory labels had recall < 0.5. A recall of < 0.5 means that more than half of the papers deemed relevant for a category based on the full text are not

being identified as relevant when only the title/abstract data is considered. Some labels were promising for abstract labeling, including reference type - primary (recall = 0.99, precision = 0.98, N=707), chemical - yes (recall = 0.98, precision = 0.93, N=696), and mechanism - thyroid membrane transporters, (recall = 0.94; precision = 0.97, N=242). The case study using the LLM application revealed that these models could achieve near-human proficiency in categorizing literature, particularly in identification of study type with balanced accuracy ranging 0.94-1 when a non-human organismal group was identified, effectively reducing the manual burden on researchers. Several categories of labels showed strong performance including mechanism labels, as well as several organism group labels, particularly rodent (balanced accuracy = 0.97; MCC = 0.92).LitMapper was used to explore biological concepts contained in a literature set focused on thyroid hormone (TH) serum binding proteins. Two major article clusters were identified centered on either thyroid or heart. Further exploration into the cardiac sub-set identified 'amyloidosis' as one of the differentiating concepts that linked the TH serum distributor protein, transthyretin, to the brain and heart through the development of transthyretin amyloid fibrils. Using this information allowed a refinement of the SEM approach for these proteins. LitConnector can be used to explore and visualize relationships between concepts identified during article review. Using labels applied by human reviewers, the amount of literature describing TH membrane transport or serum distributor proteins during development were assessed. This showed, that while data for non-adult life stages were available, the number of articles was limited and may not provide a comprehensive evaluation. This type of information can support data-driven decisions and has the potential to allow for refinements of the scope of a project.

Conclusions: This work illustrates a cohesive, integrated approach to advancing SEM through the strategic application of computational tools that can also be deployed modularly. By leveraging case studies and iterative refinement, the updated workflow addresses key limitations of traditional SEM, offering a more robust, efficient, and scalable solution for systematic mapping of complex scientific literature. This integrated approach accelerates the process and enhances the accuracy and contextual relevance of SEMs. It also provides a practical framework that has the potential for use in future downstream applications, such as systematic reviews to support chemical risk assessments or developing Integrated Approaches to Testing and Assessment and benchmarking future SEM methodological improvements.