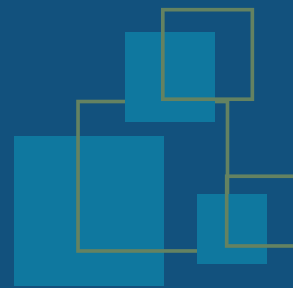


WORK HEALTH AND SAFETY RESEARCH HORIZON SCAN, SCOPING REVIEW AND EVIDENCE GAP MAPPING PROJECT



Scoping literature review methodology

A report prepared for Safe Work Australia



About this project

This project was commissioned by Safe Work Australia, and it maps the existing landscape of work health and safety and workers' compensation research across the five Safe Work Australia *Research and Evaluation Strategy* initial priority areas to provide a data driven understanding of the current evidence base, gaps, and emerging research areas.

This report was developed with guidance from an Expert Working Group of experts from across work health and safety research and related areas. The Academy of the Social Sciences in Australia and the Australian Academy of Technological Sciences and Engineering gratefully acknowledges the Expert Working Group for their contributions.

Expert Working Group Members

Professor Peng Bi FASSA

Professor Alex Collie

Professor Maureen Dollard FASSA

Professor Andrew Neal FASSA

Professor Sharon Parker FASSA

Professor Michelle Tuckey

Professor Mark Wooden FASSA

Report prepared by Dr Harry Rolf, Dr Honae Cuffe, Nicola Smillie and Olivia Allen

© Safe Work Australia 2026

This report was commissioned by Safe Work Australia and prepared by the Academy of the Social Sciences in Australia. With the exception of third-party material and where otherwise noted, this work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

Date of publication

April 2026

Publisher

Academy of the Social Sciences in Australia

3/95 Northbourne Ave

Turner ACT 2612

Australia

www.socialsciences.org.au

Citation

Academy of the Social Sciences in Australia. (2026). Work health and safety research horizon scan, scoping review and evidence gap mapping project: Scoping literature review methodology. <https://doi.org/10.60651/84Y4-MY06>

Table of contents

Overview	3
Protocol	4
<i>Research questions</i>	4
<i>Inclusion and exclusion criteria</i>	4
<i>The use of generative AI</i>	5
<i>Quality assurance</i>	6
Search Strategy	7
<i>Stage 1: Retrieval</i>	7
<i>Stage 2: Screening and enrichment</i>	7
<i>Stage 3: Processing and consolidation</i>	8
<i>Stage 4: Validation and refinement</i>	8
Analysis and results	8
Outputs	9
References	10
Appendix 1 Search query	11

Overview

This paper outlines the methodology and approach for the scoping literature review component of Safe Work Australia's (SWA's) Horizon Scan and Evidence Gap Mapping project.

Scoping literature reviews are an established approach for synthesising the breadth and characteristics of research evidence. They are valuable when the objective is to capture the diversity of studies and perspectives in a field, rather than to evaluate the quality or effect size of specific interventions. By systematically identifying, classifying, and charting published and grey literature, scoping literature reviews provide a structured overview of the available evidence while revealing areas where knowledge gaps or research saturation exist.

The Horizon Scan and Evidence Gap Mapping project is intended to establish the current Australian work health and safety (WHS) research landscape and identify emerging trends, risks and opportunities relevant to the future of Australian work and workplaces. The purpose of this scoping literature review is to systematically identify WHS literature that is relevant to the five research priority areas in SWA's Research and Evaluation Strategy and that is either authored by Australian researchers or directly concerns the Australian context in the last two decades. This is achieved through the application of a structured taxonomy of WHS concepts and terms that operationalises the conceptual boundaries of each priority area. The taxonomy is used to guide literature searches in Dimensions.ai, enabling comprehensive retrieval of relevant Australian, international, and cross-disciplinary research.

In this project, the scoping literature review complements the horizon scan. Where the horizon scan adopts a broad, forward-looking perspective, focusing on weak signals and drivers of change that cut across priority areas, the scoping literature review takes a focused, retrospective view, systematically mapping the existing evidence base within each of the five priority areas. The two methods are linked through the search term taxonomy, which structures the scope of the search and evolves iteratively.

The scoping literature review methodology and approach presented here are underpinned by established frameworks for evidence synthesis that emphasise transparency, reproducibility, and iteration. In keeping with these principles, the process combines a structured protocol with a degree of flexibility to accommodate the iterative refinement of the taxonomy and the inclusion of diverse evidence sources. The results will be presented as enarrative summaries and bibliometric datasets that align with the taxonomy's categories, enabling straightforward comparison with the horizon scan outputs and the interpretation of evidence trends, strengths and gaps.

The scoping literature review methodology and approach is organised into the following parts:

- **A protocol** that establishes the boundaries of the review and outlines the approach to quality assurance;
- **A search strategy** designed to capture research publications using Dimensions.ai; and
- **An approach to the analysis** of screened results, mapping them against the taxonomy and describing the characteristics of records using bibliometrics and quantitative analysis.

Protocol

The protocol defines the scope of a review, it covers the guiding research questions, inclusion and exclusion criteria, quality assurance and use of generative AI.

Research questions

The review is guided by a primary research question and three subsequent questions:

What is the extent and distribution of identifiable Australian work health and safety (WHS) and workers' compensation research within each taxonomy for SWA's five priority areas?

- What are the characteristics of Australian WHS and workers' compensation research in a priority area? (e.g. publication characteristics such as types, years published, organisations, authors etc. and topics defined by combinations of taxonomy terms that publications are tagged with).
- To what extent are evidence gaps visible within a priority area?
- To what extent is the current evidence base able to address emerging weak signal and drivers of change identified through the horizon scan for a priority area?

These questions inform the identification, classification, and analysis of literature.

Inclusion and exclusion criteria

The scope of each review is defined by the relevant taxonomy which sets the conceptual boundaries for the inclusion of identifiable literature. All literature must relate to WHS within the Australian context, whether through study setting, institutional affiliation, or explicit reference to Australian policy, regulation, or workforce conditions. Consistent with the horizon scan's scope, the following criteria apply:

- **Inclusion:** Research publications that address WHS-relevant issues within at least one taxonomy category; materials published in English from 2010 onward; and publications that concern Australian contexts or have Australian author affiliation.
- **Exclusion:** Materials that are not related to WHS and/or workers' compensation, not relevant to Australia, not written in English, or fall outside the taxonomy's conceptual scope.

WHS is a broad, multi-disciplinary research domain that spans health sciences, psychology, engineering, law, organisational behaviour, public policy and practice-based research. Each field applies different standards for methodological rigour and publication, and evidence is produced across a range of scholarly and professional outlets from peer-reviewed journals to technical reports, conference papers, books and regulatory or practitioner publications. For this reason, the review does not apply quality hierarchies such as prioritising "top" journals over conferences or books and chapters when considering the inclusion or exclusion of publications.^{4,5}

Quality in the scoping literature review is managed through transparent documentation of inclusion decisions, expert validation, and the inclusion of sources that demonstrate clear relevance, credibility, and contribution to WHS knowledge and practice. This approach is consistent with scoping review methodology, which seeks to map the breadth of evidence rather than assess study design quality, and supports SWA's *Research and Evaluation Strategy* principle of fostering collaboration across the research–practice ecosystem to strengthen the evidence base.

We acknowledge that inherent limitations apply to any approach dealing with this breadth of research. The global and interdisciplinary nature of WHS research means that relevant studies may not explicitly reference “work health and safety” or be Australian authored or about the Australian context, they may be indexed under other disciplines or may not be recently published. These limitations are balanced in our method by broadly searching all available sources — Australian, international, and cross-discipline — for relevant literature, using Dimensions.ai's AI-enhanced full-text search and semantic discovery capabilities to improve keyword and phrase matching and thereby maximise coverage.

The use of generative AI

Under supervision, generative AI will be used to assist with data processing, taxonomy tagging, summarisation of literature and term discovery in each scoping literature review. Prompt-engineered screening workflows have been found to achieve high sensitivity for ultimately included studies in reviews while reducing workload.⁶ Large Language Models (LLMs) like ChatGPT are capable of abstractive summarisation by identifying key entities, relationships, and thematic content in text, a process already recognised in AI-assisted systematic reviews.^{7,8}

Generative AI will be used in the following ways:

- **Data processing:** Automated deduplication and metadata standardisation of retrieved records.
- **Taxonomy tagging:** Application of natural language processing models to classify records according to taxonomy categories based on textual similarity and contextual meaning.
- **Summarisation:** Generation of summaries such as abstracts or keywords for records lacking metadata.
- **Term discovery:** Identification of significant recurring or co-occurring terms not already identified that may inform the taxonomy or the horizon scan.

The use of generative AI and AI-assisted processes are subject to human verification and when used the model will be asked to provide an internal measure of confidence about the results generated using log probability.^{9,10} This internal confidence score allows the researcher to assess the model's confidence in its outputs. Outputs with a low confidence score will be reviewed by the researcher, and a statistically significant sample of outputs with confidence scores over the threshold will also be reviewed.

To enhance the interpretability and reliability of generative AI outputs, we will also adopt a confidence–diversity calibration framework that complements the use of confidence scores.¹¹ The method recalibrates LLM probability scores into context-sensitive confidence intervals rather than relying on a single fixed threshold.

In this framework, records are divided into three confidence tiers:

- High-confidence outputs (above 85 %) are automatically accepted and logged for random validation.
- Intermediate-confidence outputs (typically between 70 – 85 %) are deferred for targeted human review.
- Low-confidence outputs (below 70 %) are flagged for exclusion or re-evaluation.

This tiered calibration allows the model to explicitly recognise uncertainty and route ambiguous classifications to human reviewers, improving both efficiency and accuracy in large-scale literature screening. Recent evaluations have shown that generative AI tools do face limitations when used for literature searching and have variable error rates in screening. Independent benchmarking also shows that a model’s self-assessed confidence can be miscalibrated, so supervised, auditable use under human supervision is essential.^{12, 13} For this reason, human verification and confidence testing forms part of the quality assurance approach when using AI during every part of this scoping literature review.

Quality assurance

To ensure the integrity, accuracy and presentiveness of results we will apply three levels of quality assurance during a scoping literature review: internal researcher review, expert oversight, and analytical validation.

At the researcher level, internal cross-checking is conducted at each stage of the review with searches, screening and coding decisions recorded, and reviewed by at least two members of the Academy of the Social Sciences in Australia (the Academy) project team. AI-assisted classifications will be reviewed using confidence scores and a random sample of results, with low-confidence or ambiguous results reviewed to confirm accuracy, and the results of these checks recorded to provide an audit trail.

Expert oversight is provided by SWA and the Expert Working Group (EWG). The EWG will be provided with the opportunity to review interim outputs at key stages including search, screening and analysis. Their role is to confirm that the review sufficiently reflects the domain knowledge of WHS and workers’ compensation and that findings and insights correspond to their understandings of the field.

At the analytical level, validation will be performed using descriptive and comparative statistics to assess the completeness and representativeness of search results. These analyses will examine the distribution of retrieved records by taxonomy category, source type, and organisational affiliation. They will highlight the coverage of terms, key organisations, and significant researchers within each priority area, enabling interrogation of results by SWA, the EWG and project stakeholders.

The process of quality assurance will not attempt to benchmark the results against the global WHS body of knowledge but seeks instead to ensure that the dataset includes the principal Australian and international organisations, authors, and publication venues known to shape WHS and workers' compensation research and policy. Where notable gaps are identified, such as the absence of a recognised stakeholder or source, iterative adjustments to search or analysis can be performed.

Search Strategy

The search strategy is designed to identify, retrieve, and consolidate research evidence relevant to the SWA Research and Evaluation Strategy priority areas using Dimensions.ai, selected for its capacity to integrate scholarly publications, funded research projects, policy documents, and patents within a unified metadata structure, enabling broad yet precise coverage of the WHS research landscape. The database's advanced search schema supports complex Boolean queries and field-level filtering, allowing the taxonomies to be directly operationalised in the search and retrieval process.¹⁴

The search process follows an iterative four-stage design. It continues iteratively until the results achieve conceptual saturation, meaning that further query adjustments no longer yield substantively new terms, sources, or stakeholders. All search iterations, parameters, and versioned datasets are recorded in the project's audit trail to ensure transparency and reproducibility.

Stage 1: Retrieval

The first stage involves constructing a series of standardised Boolean search queries to capture the breadth of research in a priority area. Primary and related terms from the taxonomy are converted into Boolean strings ("trauma exposure" OR "fatigue and burnout"), linked with domain anchors ("work health and safety" OR "occupational health and safety") and restricted by the filter (Australia OR Australian). Filters will also be used to place temporal limits that confine results to the period beginning in 2010.

Each query has core and optional modifiers that can be applied to ensure that optimum search coverage is achieved. Search queries will be run separately within Dimensions.ai, and metadata records are retrieved across all available research publications identified. Query use and results are logged in line with the quality assurance portal to form an audit trail, see Appendix 1 for the standardised query template.

Stage 2: Screening and enrichment

Following retrieval, records undergo a combined process of semi-automated screening and enrichment using generative AI to create a curated subset of relevant records.

Generative AI is used to assist the researcher by screening the relevance of each record against inclusion and exclusion criteria. The model is asked to classify records using titles and abstracts. Summary statistics describing the total number of records retrieved, screened, excluded, and retained will be captured and reported in PRISMA-ScR format, ensuring transparency and replicability.^{15,16}

In parallel with screening, generative AI will be applied to enhance the metadata of some retained records. Where metadata such as abstracts and keywords are missing or incomplete, the model will be asked to produce metadata to facilitate consistent interpretation and analysis. The model will also be used to tag records with relevant taxonomy terms using rule-based keyword matching and semantic classification to provide additional metadata.¹⁷

This integrated approach allows the review to filter and structure the dataset simultaneously, reducing repetition between screening and data preparation. All model outputs, including confidence scores and tags, will be recorded to provide an audit trail and are subject to the quality assurance procedures outlined in the protocol.

Stage 3: Processing and consolidation

The third stage involves processing and consolidation of the retrieved results. Screened search results from Dimensions.ai are exported and merged into a single dataset using a unified schema and stored in a spreadsheet. During this step, duplicate records are identified and removed, and metadata fields are normalised. This processing and consolidation ensure that the dataset is standardised and ready for subsequent analysis.

Stage 4: Validation and refinement

The final stage is validation and refinement. The consolidated dataset is examined descriptively to evaluate its coverage and alignment with the taxonomy. Descriptive statistics, such as publication and term frequencies and distributions will be produced to identify any areas of underrepresentation. These analyses help determine whether the search strategy adequately captures the breadth of WHS and workers' compensation research or whether additional refinement is required.

Analysis and results

Analysis will transform search results into a coherent representation of the WHS evidence base. The quantity, distribution and characteristics of research will be described to reveal areas where evidence is sparse or absent within the priority areas and set against the weak signals and drivers of change emerging from the case studies.

In keeping with the purpose of a scoping literature review, the analysis will be descriptive rather than evaluative, focusing on mapping patterns of research activity rather than assessing study quality, design or outcomes. The approach draws on established guidance for scoping literature reviews and evidence gap mapping that recommends combining quantitative summary statistics with visual representations of evidence coverage to convey complex findings clearly and accessibly.

Analysis begins with descriptive statistical profiling of the dataset. Frequency counts and cross-tabulations will be produced for each taxonomy category including vulnerable contexts and workers' compensation. These summaries will reveal how research effort is distributed across the five priority areas and allow identification of heavily studied and under-represented topics.

Temporal analysis will chart publication trends over time, showing the growth or decline in particular topics. Distributions will be examined across document types such as publications, grants, patents, and policy outputs to understand research activity within the broader WHS knowledge base.

Bibliometric and network analysis will be used to depict the configuration of WHS knowledge. Network analyses (using co-occurrence) will be conducted to visualise relationships among key entities such as authors, organisations, and taxonomy terms. Co-occurrence analysis will identify which taxonomy terms appear most frequently together, revealing clusters of research activity and intersections among the priority areas. Organisational and geographic mappings will highlight the institutional and spatial Dimensions of WHS and workers' compensation research, showing which organisations or geographies are most active within each domain.

Outputs

Two outputs will be produced through the scoping literature review for each priority area: a narrative report and bibliometric dataset.

A narrative report will integrate the quantitative findings of analysis with interpretive commentary that situates evidence patterns within the broader WHS and workers' compensation research context. Each report will summarise the overall volume and distribution of research, highlight key trends and gaps, and discuss the implications in the context of SWA's research strategy. The reports will be written to complement the horizon scan case studies by providing a retrospective, evidence-based view of the same domains that the scan examines prospectively.

A comprehensive bibliometric dataset, comprising the cleaned, coded, and categorised records identified through the search and retrieval process will be provided for each priority area. This dataset forms the empirical foundation of the review and can be used for secondary analyses, monitoring trends in WHS research over time, and supporting ongoing updates to the taxonomy. The dataset provides a reproducible resource that can be used to verify and reproduce the scoping literature reviews findings.

References

- ¹ Munn, Zachary, Danielle C. Stern, Edoardo Aromataris, Tiffany J. Lockwood, and Fiona M. Tufanaru. "Systematic Review or Scoping Review? Guidance for Authors When Choosing between a Systematic or Scoping Review Approach." *BMC Medical Research Methodology* 18, no. 1 (2018): 143.
- ² Tricco, Andrea C., Erin Lillie, Patricia Zarin, et al. "A Scoping Review on the Conduct and Reporting of Scoping Reviews." *BMC Medical Research Methodology* 16, no. 1 (2016): 15.
- ³ Scherbakov, Dmitry, et al. "The Emergence of Large Language Models as Tools in Literature Reviews: A LLM-Assisted Systematic Review." *Journal of the American Medical Informatics Association* (2025): ePub ahead of print.
- ⁴ Sorensen, Gloria, Rachael B. Henning, and L. Casey Chosewood. "The Future of Research on Work, Safety, Health and Well-Being: A Guiding Framework." *Social Science & Medicine* 269 (2021): 113593. <https://doi.org/10.1016/j.socscimed.2020.113593>.
- ⁵ Banks, David, Letitia Davis, Marissa Baker, and John Howard. "Scoping Total Worker Health®: An Expanded View of Worker Safety, Health, and Well-Being." *Journal of Occupational and Environmental Medicine* 67, no. 7 (July 2025): e403–e412. <https://doi.org/10.1097/JOM.0000000000003185>.
- ⁶ Homiar, Amir, et al. "Development and Evaluation of Prompts for a Large Language Model to Screen Titles and Abstracts in a Living Systematic Review." *BMJ Mental Health* (2025): ePub ahead of print. <https://doi.org/10.1136/bmjment-2025-301762>.
- ⁷ Enrico M. Bucci et al., "Artificial Intelligence in Systematic Reviews: Promises and Pitfalls," *Systematic Reviews* 12, no. 1 (2023): 45.
- ⁸ O'Mara-Eves, Alison, et al. "Using Text Mining for Study Identification in Systematic Reviews: A Systematic Review of Current Approaches." *Systematic Reviews* 4 (2015): 5.
- ⁹ Gautam Chutani. "Unlocking LLM Confidence Through Logprobs." *Medium*, February 18, 2024. <https://gautam75.medium.com/unlocking-llm-confidence-through-logprobs-54b26ed1b48a>
- ¹⁰ Vatvenger. "Confidence Unlocked: A Method to Measure Certainty in LLM Outputs." *Medium*. Accessed October 14, 2025. <https://medium.com/@vatvenger/confidence-unlocked-a-method-to-measure-certainty-in-llm-outputs-1d921a4ca43c>.
- ¹¹ Zhilong Zhao and Yindi Liu, *A Confidence–Diversity Framework for Calibrating AI Judgement in Accessible Qualitative Coding Tasks* (Guangzhou: School of Journalism and Communication, South China University of Technology, 2025). <https://doi.org/10.48550/arXiv.2508.02029>.
- ¹² Clark, James, et al. "Generative Artificial Intelligence Use in Evidence Synthesis: A Systematic Review." *Research Synthesis Methods* (2025): ePub ahead of print. <https://doi.org/10.1017/rsm.2025.16>
- ¹³ Omar, Mohamed, et al. "Benchmarking the Confidence of Large Language Models in Answering Clinical Questions." Preprint, 2024.
- ¹⁴ Gusenbauer, Michael, and Sarah Gauster. "How to Search for Literature in Systematic Reviews and Meta-Analyses: A Comprehensive Step-by-Step Guide." *Technological Forecasting & Social Change* (2025): ePub ahead of print.
- ¹⁵ Marshall, Iain J., and Byron C. Wallace. "Toward Systematic Review Automation: A Practical Guide to Using Machine Learning Tools in Research Synthesis." *Systematic Reviews* 8 (2019): 163.
- ¹⁶ PRISMA. "PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation." Accessed October 13, 2025. <https://www.prisma-statement.org/scoping>.
- ¹⁷ Polak, M.P., Morgan, D., "Extracting accurate materials data from research papers with conversational language models and prompt engineering," *Nature Communications* no. 15, (2024).

Appendix 1 Search query

The standardised query (Table 1) has a set of core parameter and optional modifiers that can be applied to optimise search.¹⁹ The core parameters include relevant WHS taxonomy terms, anchor terms (“work health and safety”, “occupational health and safety”, WHS, OHS), an Australian context filter (“Australia” Or “Australian”), and the timeframe of 2010–2025. Optional modifiers include term specific to hazards and risks, vulnerable contexts, workers’ compensation drawn from supplementary tables accompanying the taxonomy, and key authors identified by SWA or the EWG.

- **Hazard/risk** terms to broaden coverage to specific exposures or health outcomes.
- **Vulnerable context** terms to identify risks in high-risk worker groups or industries.
- **Workers’ compensation** terms to ensure recovery and system-design issues are systematically captured.
- **Authors** such as experts in specific domains identified by the EWG or during search for inclusion.

Query Part	Purpose	Modifier	Fields
("Primary term" OR "Related term 1" OR "Related term 2" ...)	Focus the query on relevant taxonomy terms (from one of the five SWA priority areas).	Core	Title, Abstract, Full text
AND ("work health and safety" OR "occupational health and safety" OR WHS OR OHS)	Ensure results fall within the domain of WHS research.	Core	Title, Abstract, Full text
AND (Australia OR Australian)	Restrict to Australian context or authorship.	Core	Affiliated organisation or author country, Title, Abstract
AND (2010 OR 2011 ... OR 2025)	Limit to the timeframe 2010–2025.	Core	Date published
AND ("hazard*" OR "risk*" OR "exposure" OR "fatigue" OR "bullying")	Integrate hazards and risks (cross-cutting category).	Optional	Title, Abstract, Full text
AND ("vulnerable work" OR "migrant worker*" OR "gig work" OR "SME" OR "small business" OR "FIFO" OR "older worker*" OR "young worker*")	Include vulnerable contexts systematically.	Optional	Title, Abstract, Full text
AND ("workers’ compensation" OR "return to work" OR "injury management" OR "rehabilitation" OR "insurance scheme*" OR "claims")	Ensure compensation implications are considered.	Optional	Title, Abstract, Full text
AND (“Author” OR “Author”)	Enable searching the literature for identified experts during search.	Optional	Title, Abstract, Full text

Table 1: Standardised search query template

¹⁹ Gusenbauer, Michael, and Sarah Gauster. “How to Search for Literature in Systematic Reviews and Meta-Analyses: A Comprehensive Step-by-Step Guide.” *Technological Forecasting & Social Change* (2025): ePub ahead of print.