THE UNIVERSITY OF MELBOURNE

ARC TRAINING CENTRE IN COGNITIVE COMPUTING FOR MEDICAL TECHNOLOGIES

IBM

Stream 4: Synthesising, appraising and exploring medical evidence

Simon Šuster

with Tim Baldwin, Antonio Jimeno Yepes, Jey Han Lau, David Martinez Iraola, Yulia Otmakhova, Karin Verspoor

5 November 2020

Evidence-based decision support

clinical scenario

suggest most relevant guideline (IR & NLP)

- relate patient characteristics to inclusion/exclusion criteria of guidelines
- use summaries or direct clinical questions

clinical situation

guidelines
(recommendations for specific questions)

systematic reviews

quality appraisal

summarisation

IR & NLP

relevant studies
(RCTs, observational...)

exploration and selection

medical literature

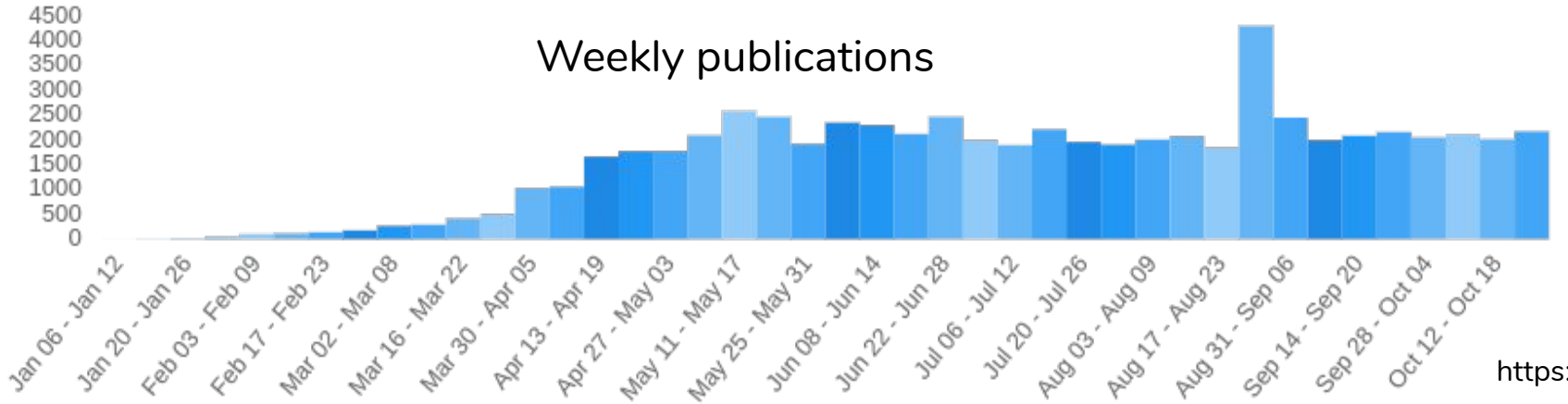
clinical question

in **PICO** format:
Population/Problem
Intervention
Comparison
Outcome

Exploring scientific evidence

Verspoor, K., Šuster, S., Otmakhova, Y., Mendis, S., Zhai, Z., Fang, B., Lau, J.H., Baldwin, T., Yepes, A.J. and Martinez, D., 2020. COVID-SEE: Scientific Evidence Explorer for COVID-19 related research. arXiv preprint arXiv:2008.07880.

Otmakhova, Y., Verspoor, K., Baldwin, T. and Suster, S., 2020. Improved Topic Representations of Medical Documents to Assist COVID-19 Literature Exploration. EMNLP 2020 Workshop NLP-COVID.



https://www.ncbi.nlm.nih.gov/research/coronavirus/

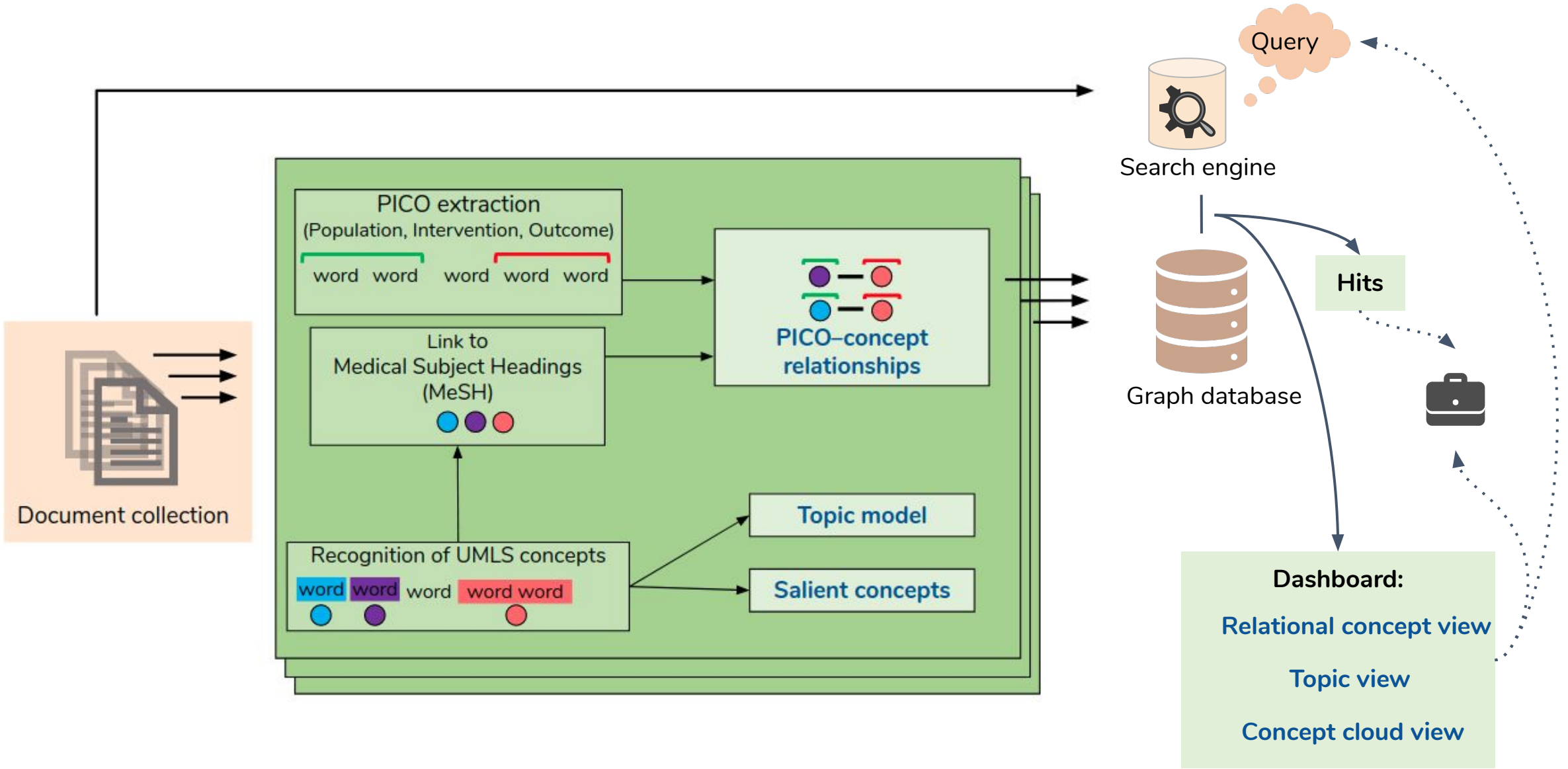
Medical literature around COVID-19 is the major hub of research activity!

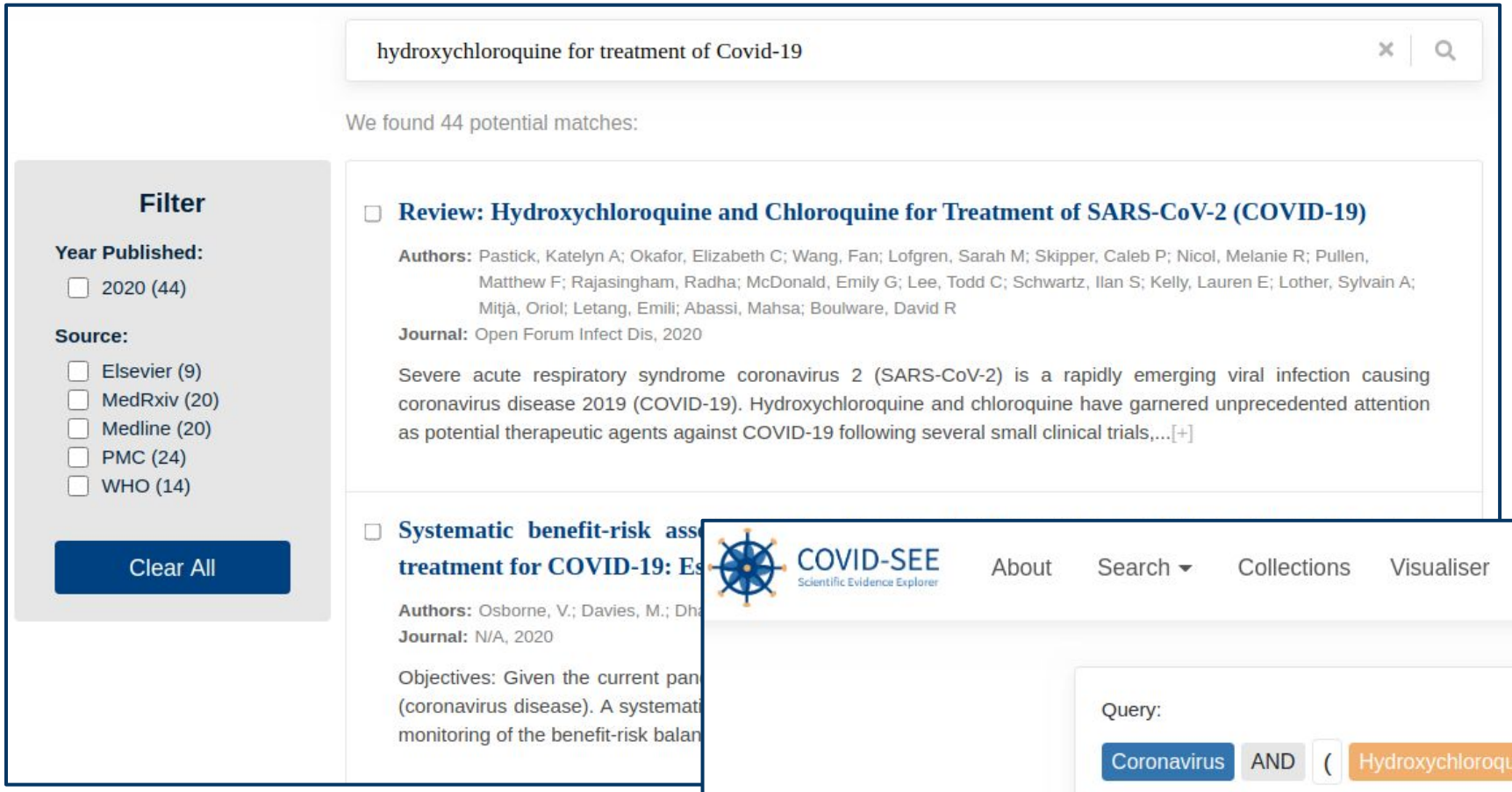
	COVID-SEE (ours)	*SciSight*	*DOC Search*	*COVID-19 Navigator*	*LitCOVID*	*SemViz*	*WellAI*	*COVID Intel. Search*	*Le Bras et al. 2020*	*COVID-19 LOVE*	*Trialstreamer*
Search:											
NL/IR	✓	✗	✓	✗	✓	✗	✗	✓	✗	✗	✓
Concepts	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PICO	✓	✓	✓	✗	✗	✗	✗	✗	✗	✓	✓
Visualisation:											
Concepts	✓	?	?	✗	✗	✓	✗	✗	✗	✗	?
Relations	✓	✓	✗	✗	✗	✓	✗	✗	✗	✗	✗
Topics	✓	✗	✗	✗	✗	✗	✗	✗	?	✗	✗

Our goals

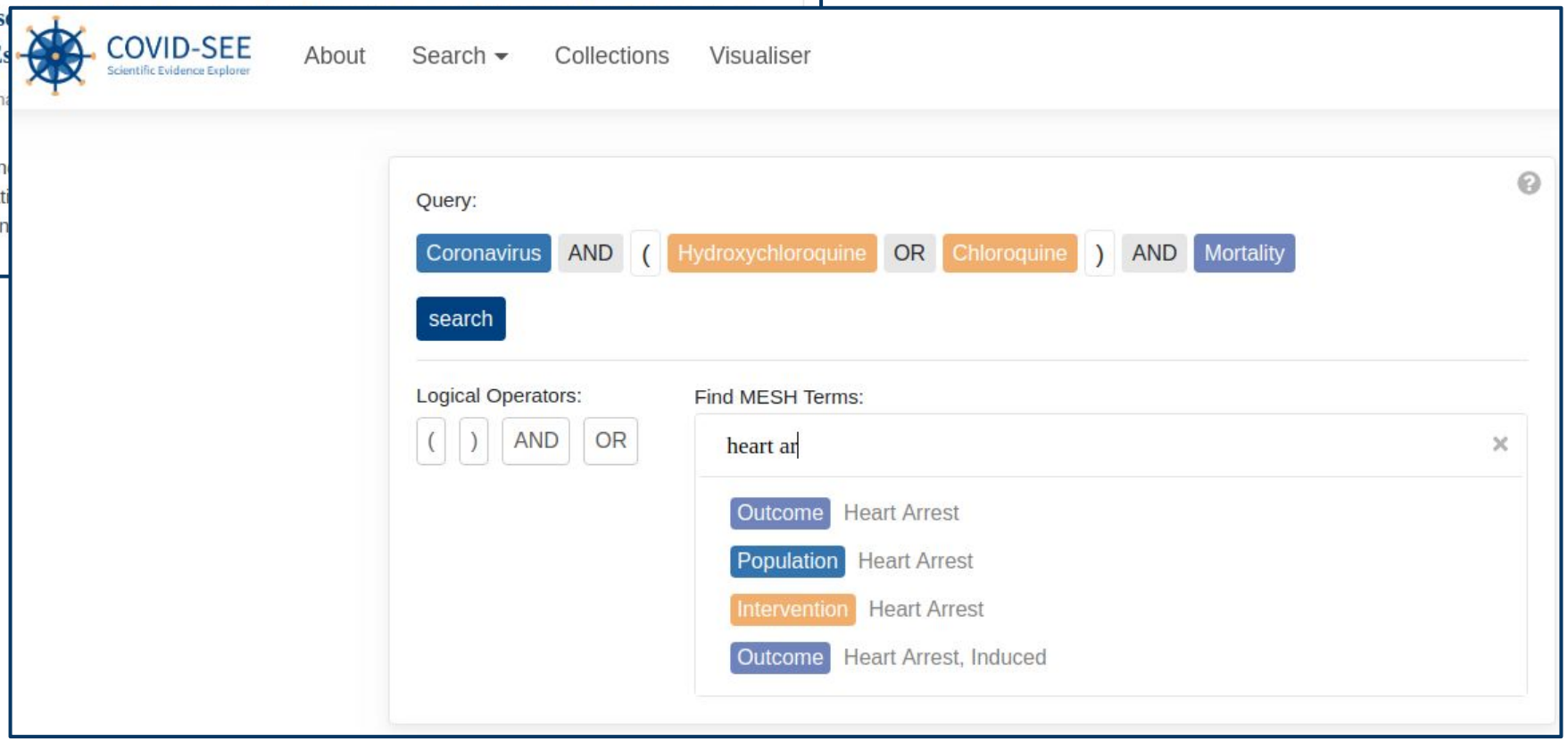
- Design a system to navigate the COVID-19-related literature effectively
- Use NLP methods, including concept recognition, detection of PICO elements and topic modeling, to structure the key information
- Offer visual overviews of document content to facilitate information discovery

System overview

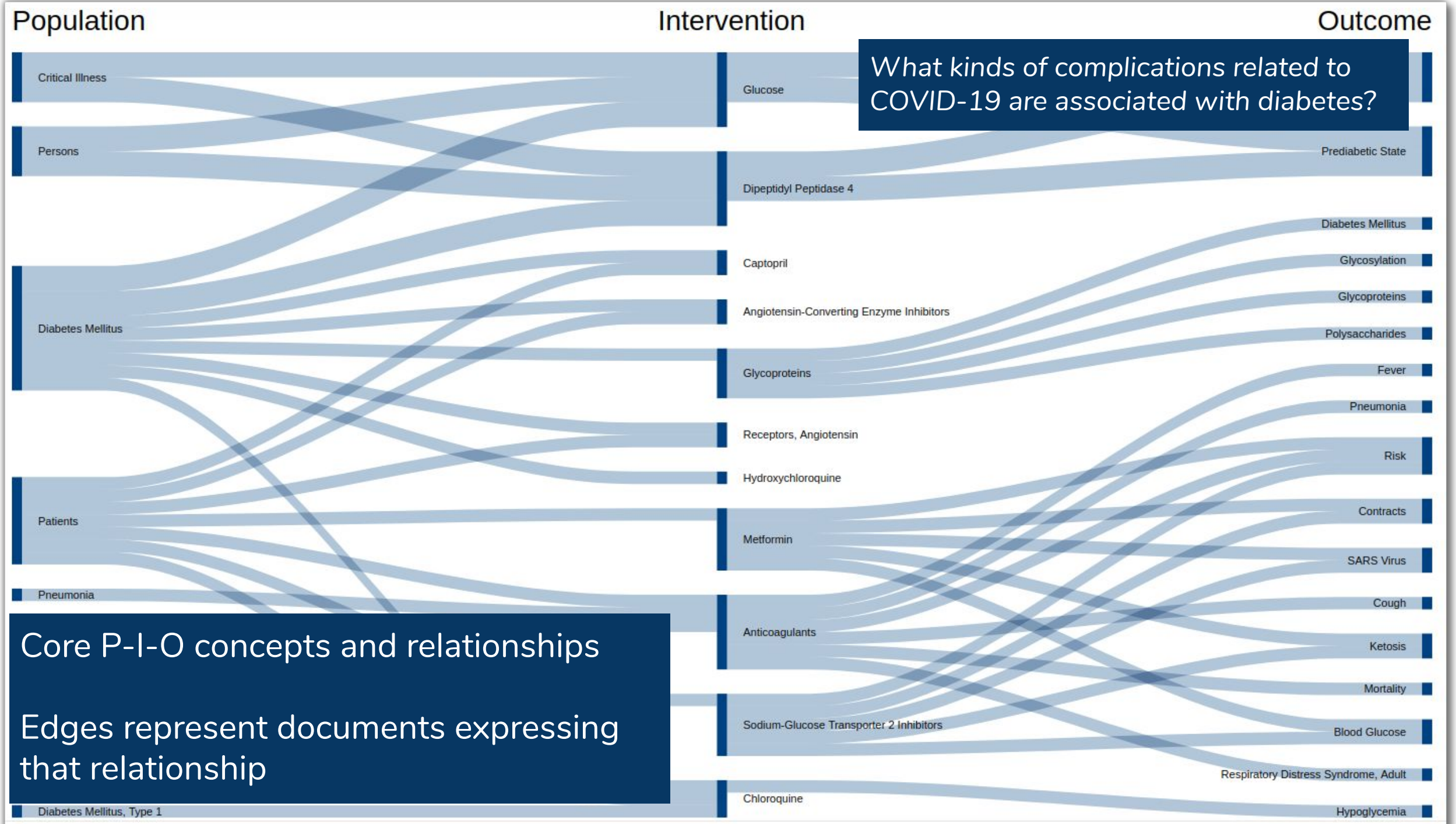


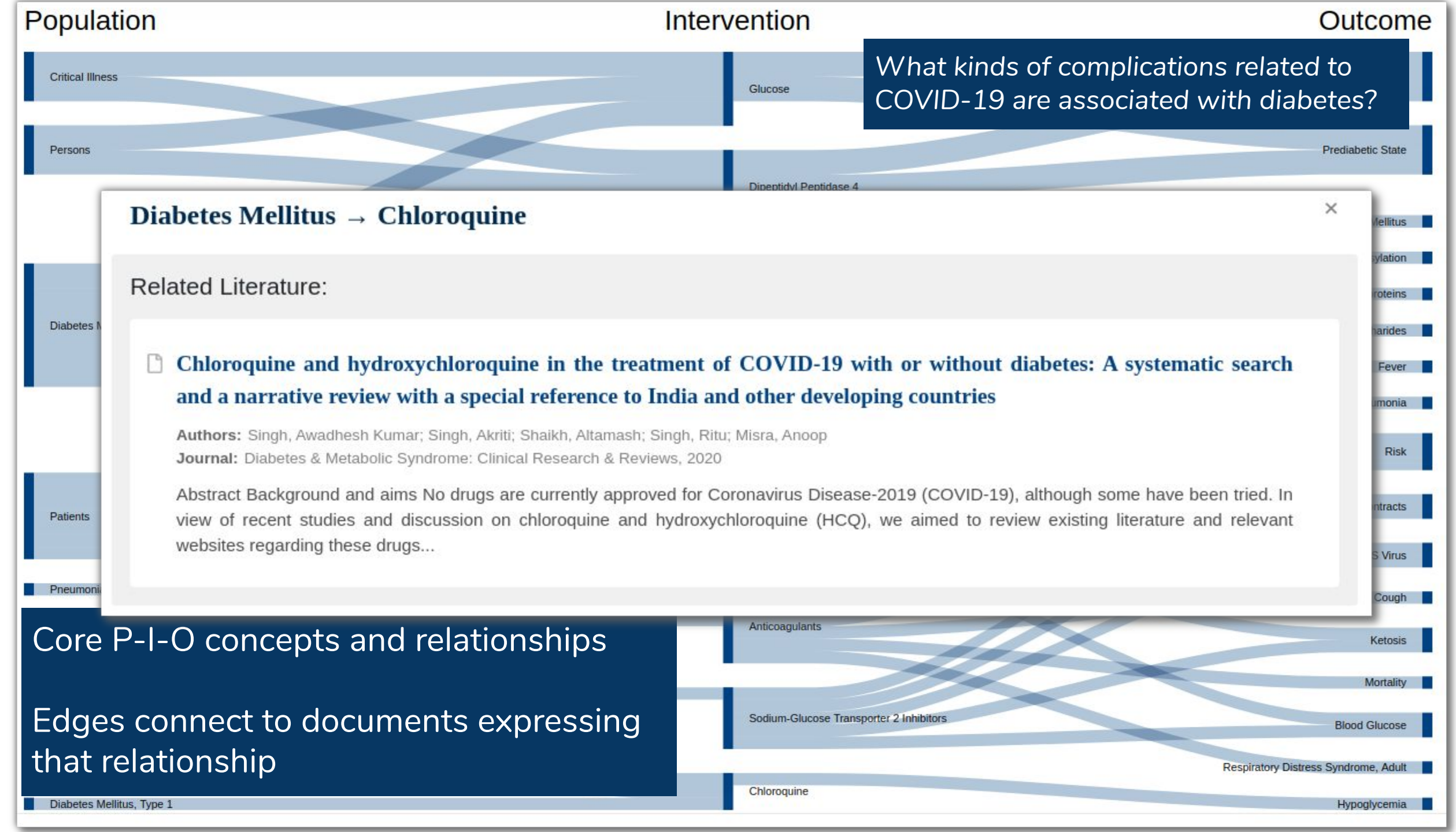


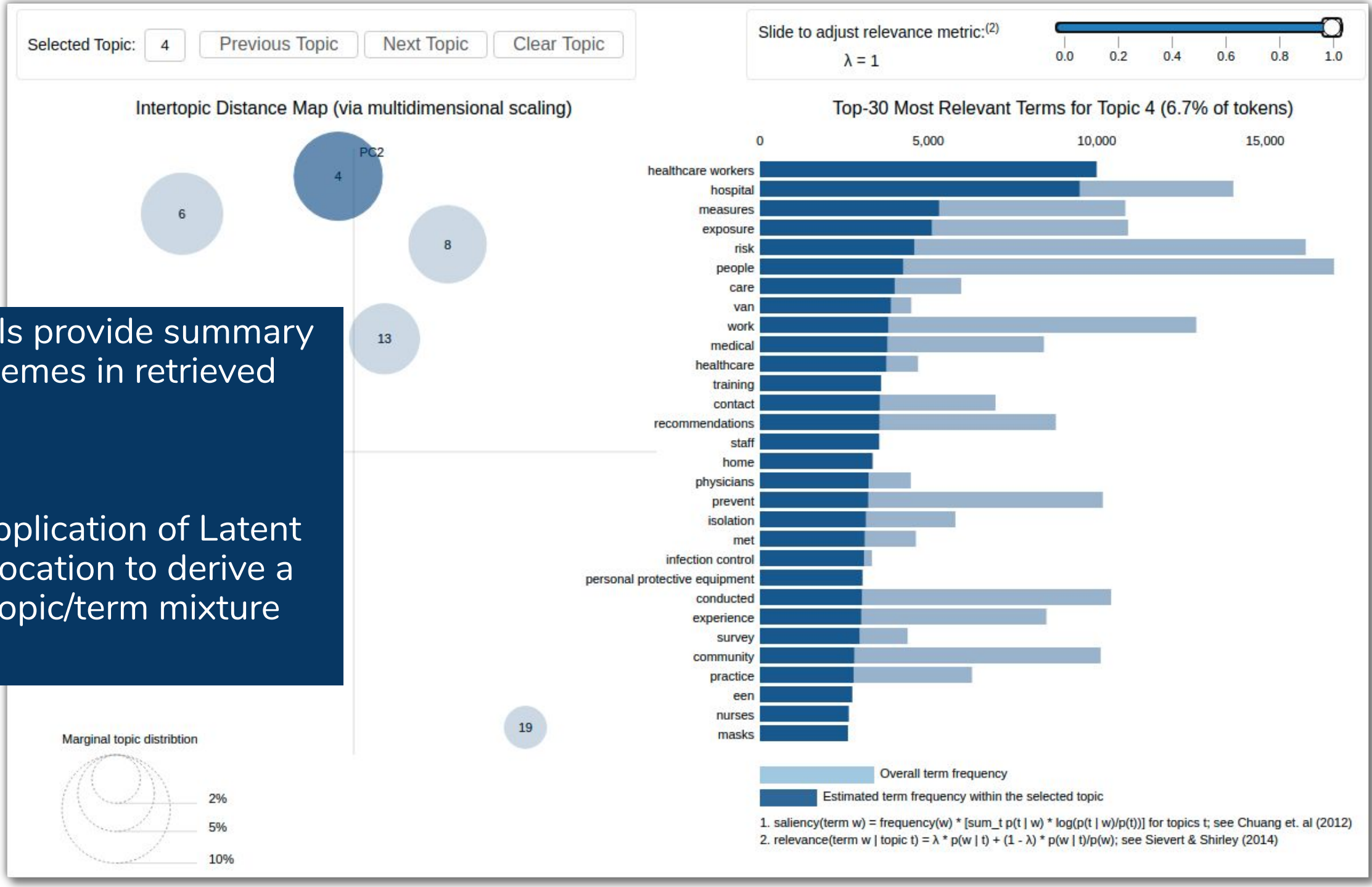
Simple search



Structured search


Population
Intervention
Outcome
What kinds of complications related to COVID-19 are associated with diabetes?
Critical Illness
Persons
Diabetes Mellitus
Patients
Pneumonia
Diabetes Mellitus, Type 1
Glucose
Dipeptidyl Peptidase 4
Captopril
Angiotensin-Converting Enzyme Inhibitors
Glycoproteins
Receptors, Angiotensin
Hydroxychloroquine
Metformin
Anticoagulants
Sodium-Glucose Transporter 2 Inhibitors
Chloroquine
Prediabetic State
Diabetes Mellitus
Glycosylation
Glycoproteins
Polysaccharides
Fever
Pneumonia
Risk
Contracts
SARS Virus
Cough
Ketosis
Mortality
Blood Glucose
Respiratory Distress Syndrome, Adult
Hypoglycemia
Core P-I-O concepts and relationships
Edges represent documents expressing that relationship


Population
Intervention
Outcome
What kinds of complications related to COVID-19 are associated with diabetes?
Critical Illness
Persons
Glucose
Dipeptidyl Peptidase 4
Prediabetic State
Diabetes Mellitus → Chloroquine
Related Literature:
Chloroquine and hydroxychloroquine in the treatment of COVID-19 with or without diabetes: A systematic search and a narrative review with a special reference to India and other developing countries
Authors: Singh, Awadhesh Kumar; Singh, Akriti; Shaikh, Altamash; Singh, Ritu; Misra, Anoop
Journal: Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020
Abstract Background and aims No drugs are currently approved for Coronavirus Disease-2019 (COVID-19), although some have been tried. In view of recent studies and discussion on chloroquine and hydroxychloroquine (HCQ), we aimed to review existing literature and relevant websites regarding these drugs...
Patients
Pneumoni
Anticoagulants
Sodium-Glucose Transporter 2 Inhibitors
Chloroquine
Diabetes Mellitus, Type 1
Fever
Risk
Cough
Ketosis
Mortality
Blood Glucose
Respiratory Distress Syndrome, Adult
Hypoglycemia
Core P-I-O concepts and relationships
Edges connect to documents expressing that relationship



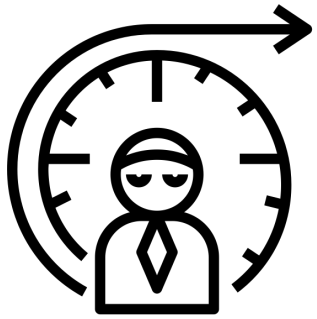
Topic models provide summary of salient themes in retrieved collection

Based on application of Latent Dirichlet Allocation to derive a document/topic/term mixture model

Concept clouds allow for quick overview of salient concepts in a document

Based on UMLS concepts

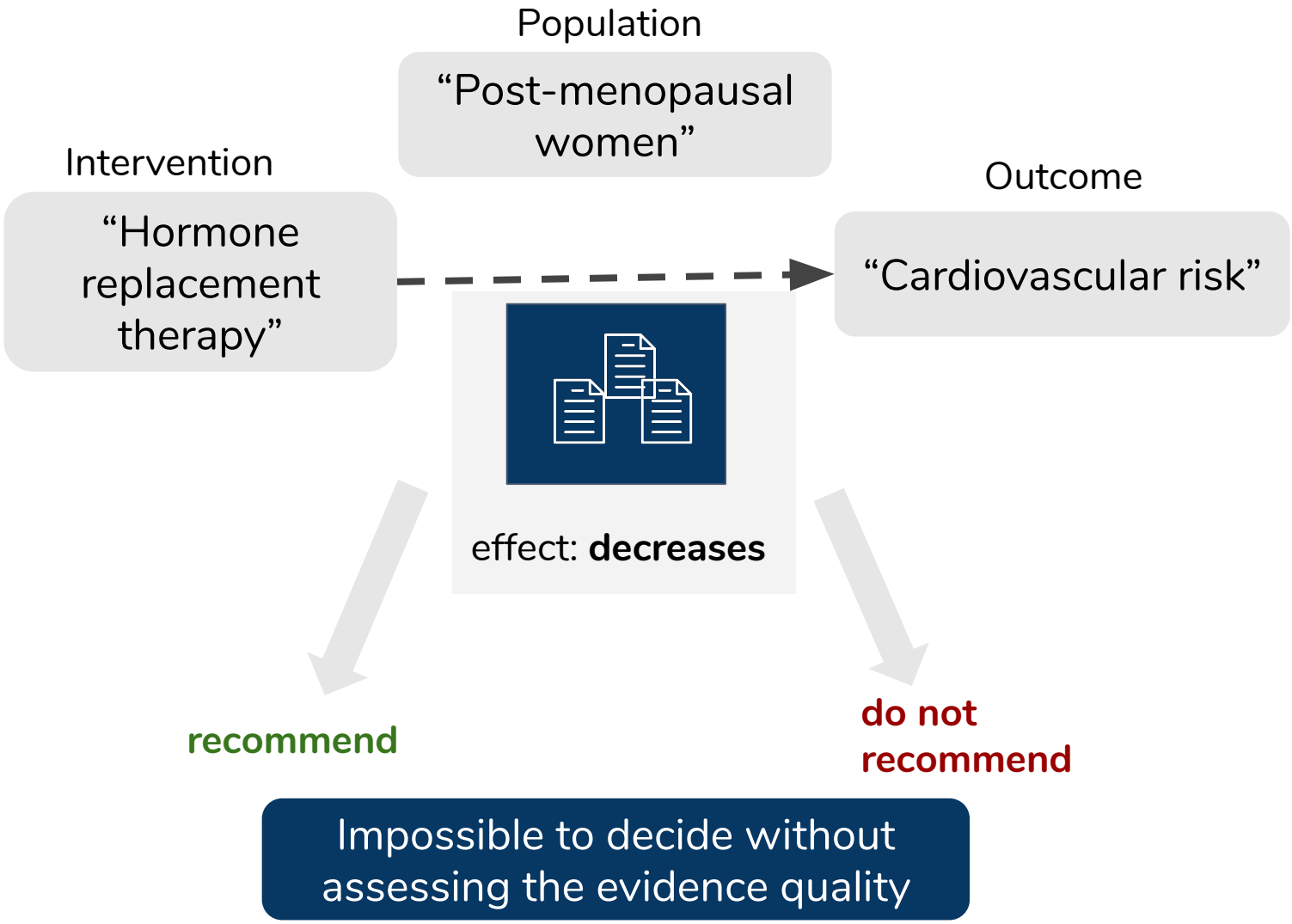
Weighting of terms based on relative distinctiveness



- Refine structured search and PICO visualisation
- Recommend articles from similar topic distributions
- Evaluate with an existing information-retrieval test collection
- Carry out a user study to better understand exploration and design

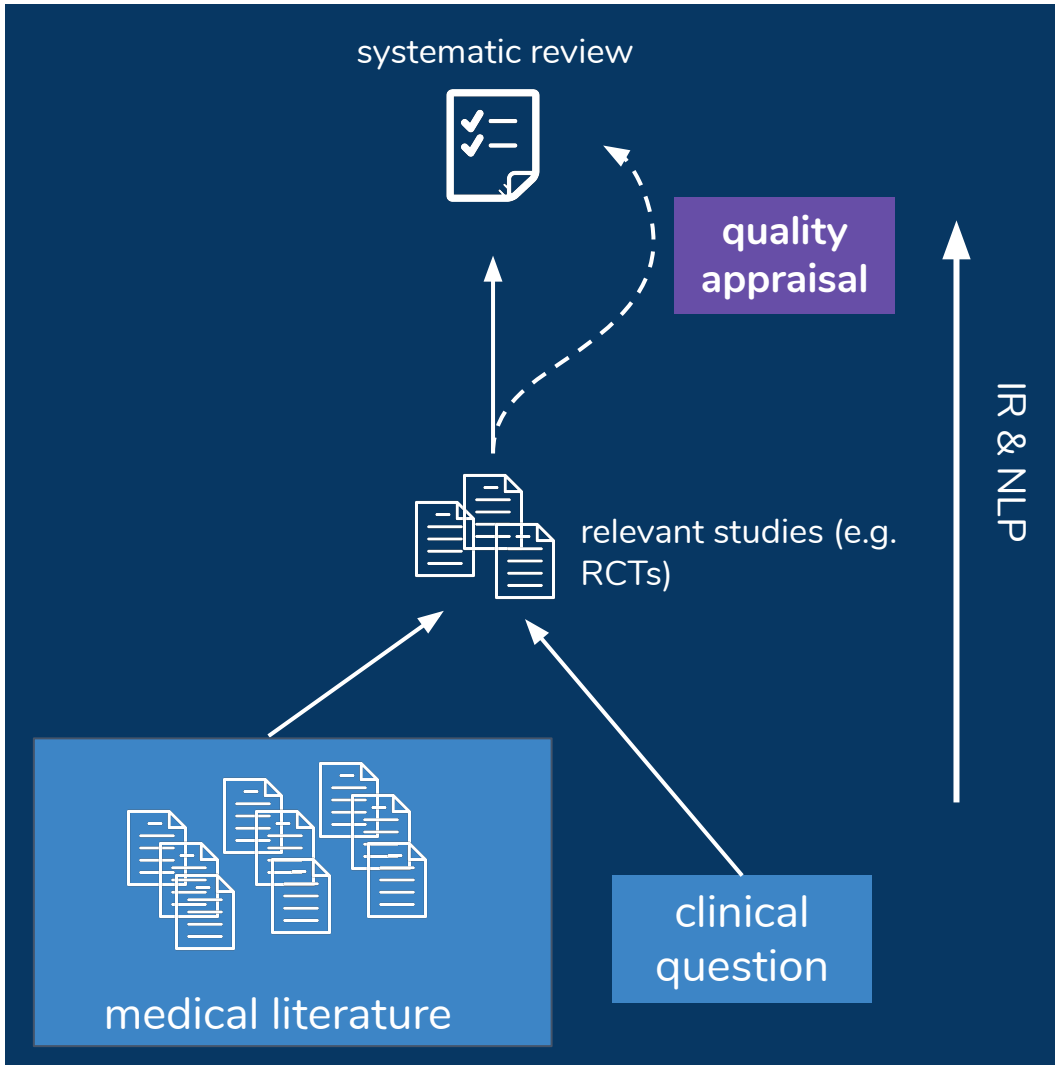
Automated quality appraisal of medical evidence

Quality assessment in evidence synthesis



Inspired by Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*, 336(7650), 924-926.

Quality assessment in evidence synthesis



What confidence should we have in the estimate (based on flaws in trial design, conduct, analysis or reporting)?

Systematic reviews are often out of date:

- Extremely labour intensive (typically take 1-2 years)
- High and increasing publication rate of primary studies

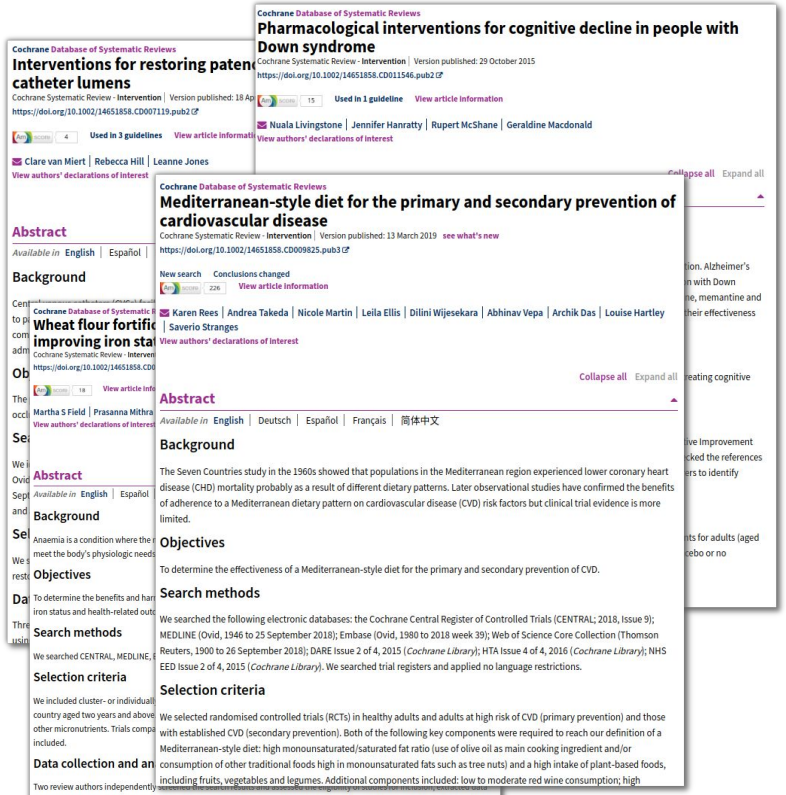
Can we expedite the reviewing process with ML?

- So far, strong focus on search and screening
- In quality appraisal, work limited to risk-of-bias detection for individual studies, e.g. RobotReviewer

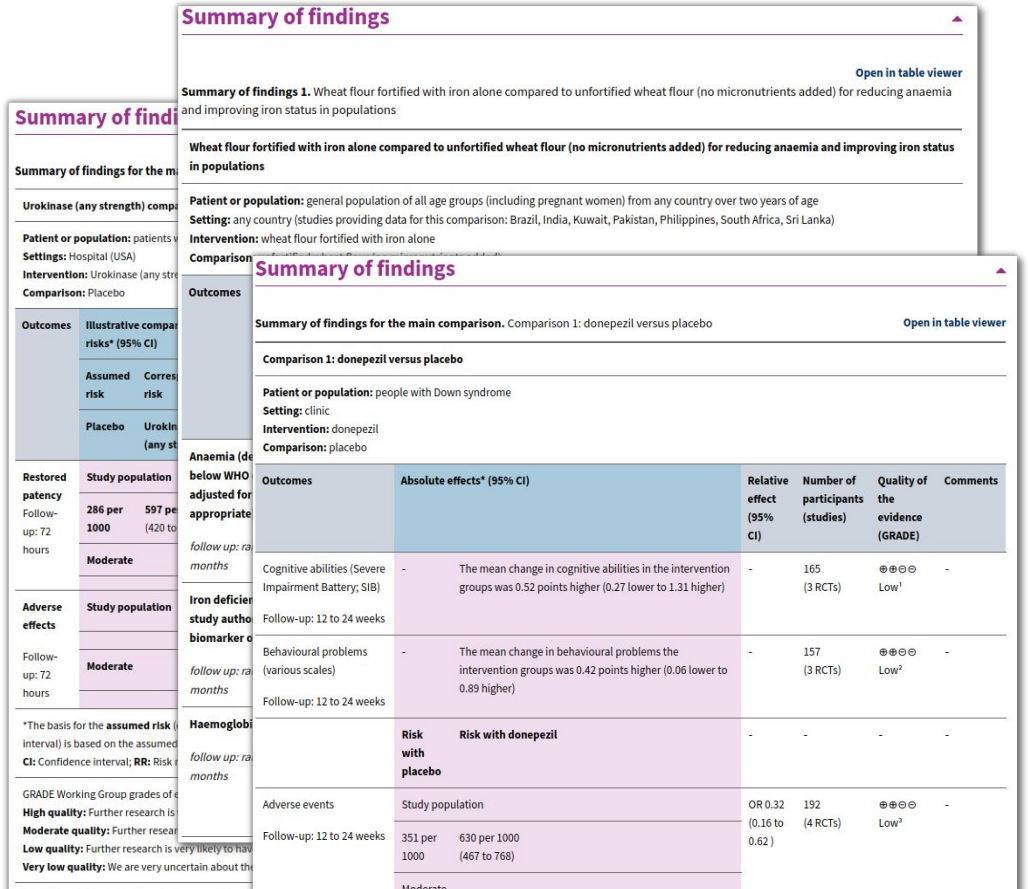
Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ open, 7(2), e012545.

Constructing a dataset for automated quality appraisal

~8,000
systematic reviews

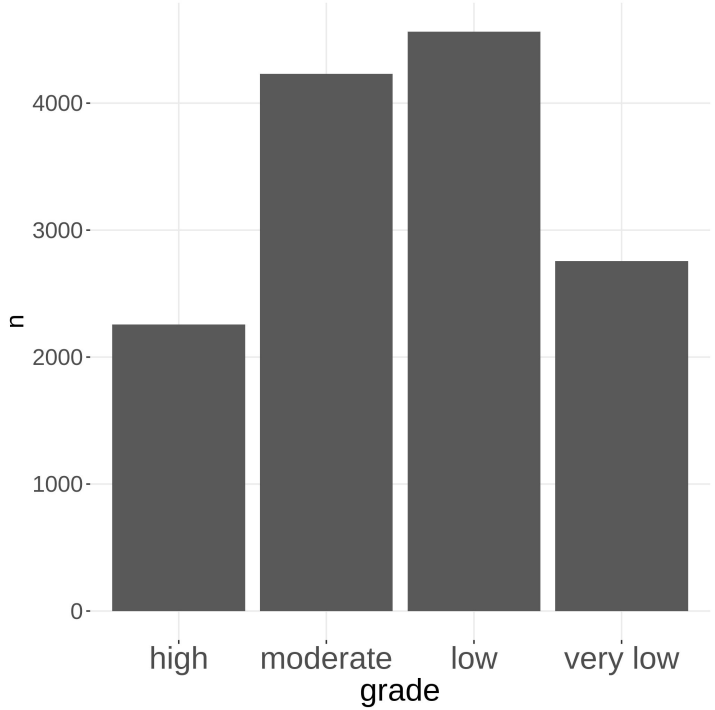


extract data related to quality appraisal from summaries of findings



14,000
outcomes rated for quality (with justification)

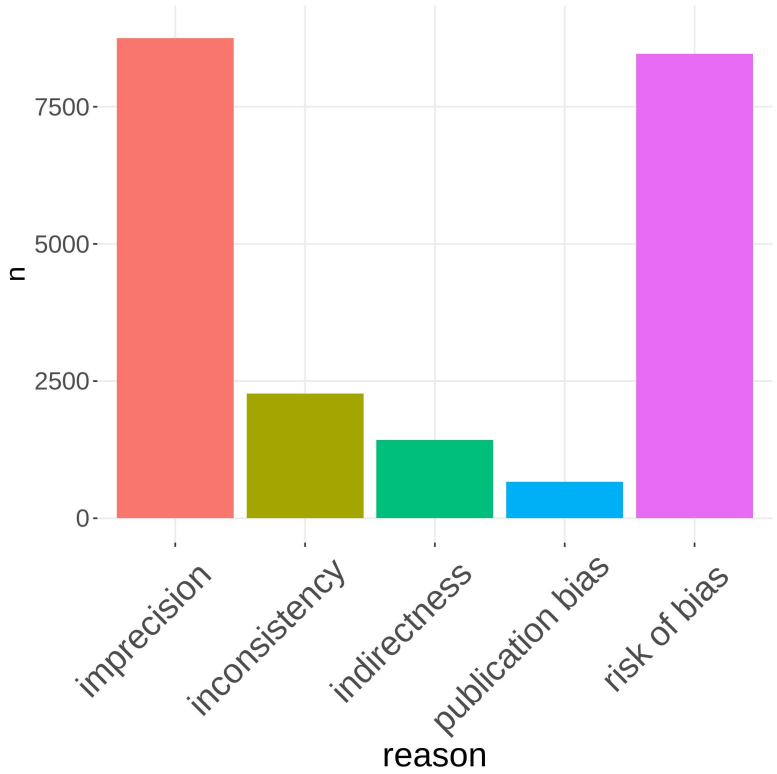
Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework

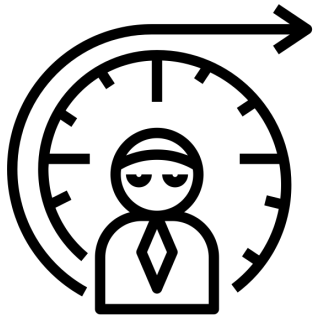


Medical evidence is of high quality only ⅙ of the time

Half of the time, evidence is of (very) low quality

When medical evidence is downgraded, it is because of...



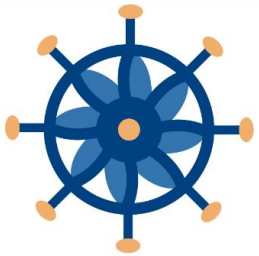


- Neural models for text classification/regression with extension to support numerical and categorical feature types (e.g. fine-tuning pre-trained language models for scientific texts)
- Provide explainable quality ratings: pointers to study flaws when downgrading
- Semi-automatically derived quality ratings for supervision: How subjective are they?

Summary

Evidence exploration

- System for navigating COVID-19-related publications
- Leveraging IR and NLP techniques for detecting concepts, PICO categories and topics



COVID-SEE
Scientific Evidence Explorer

Quality assessment of medical evidence

- New dataset from Cochrane Systematic Reviews
- A broad-scope quality rating task based on GRADE