ARC TRAINING CENTRE IN COGNITIVE COMPUTING FOR MEDICAL TECHNOLOGIES

THE UNIVERSITY OF MELBOURNE
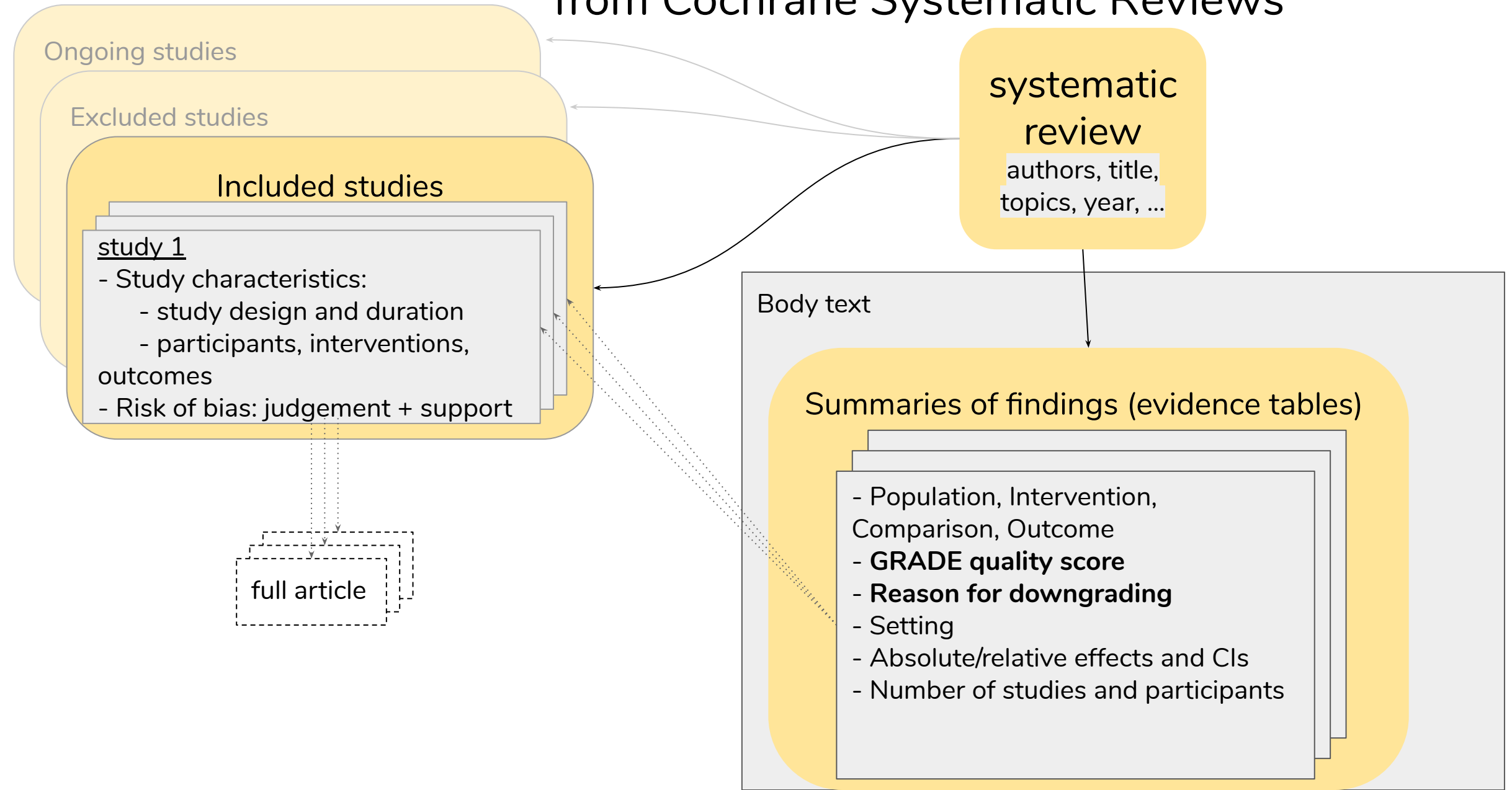
IBM

Assessing the quality of medical evidence

Simon Šuster

with Tim Baldwin, Antonio Jimeno Yepes, Jey Han Lau,

David Martinez Iraola, Yulia Otmakhova, Karin Verspoor

Stream 4

09/10/2020

Constructing a quality assessment dataset
from Cochrane Systematic Reviews

Validation of extraction procedures

- The scraped data was parsed with a combination of rules and regular expressions
- Verification of selected extracted fields against manual annotations:
    - fields: <u>GRADE score</u>, <u>categorisation of reasons for downgrading</u>, <u>number of studies</u> and <u>number of participants</u>
    - data from Cochrane reviews on anaesthesia (Conway et al. 2017)
    - own manual work
- > 0.9 accuracy for all tested categories, conservative extraction

Entire collection:
with GRADE labels (~30k)

Smaller collection:
+ valid reasons for downgrading
(~13.8k)
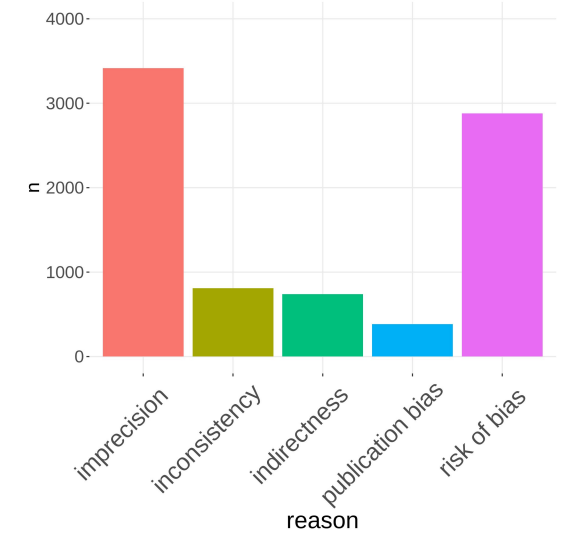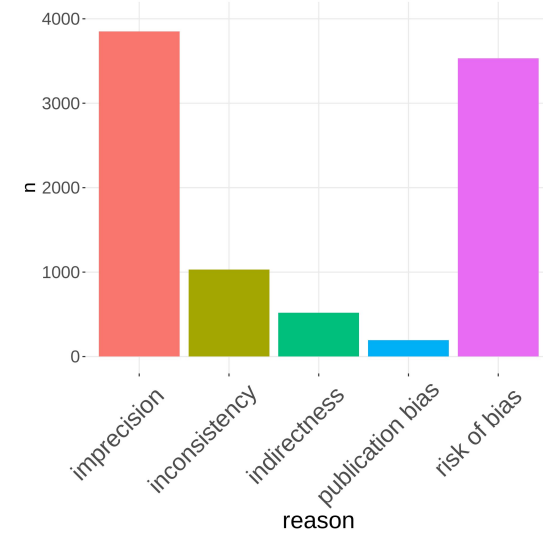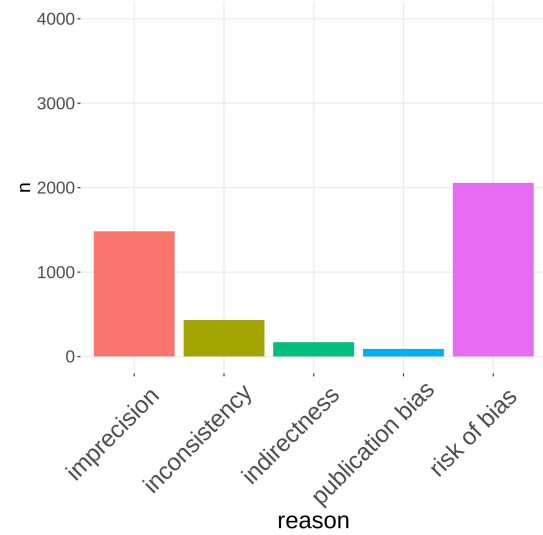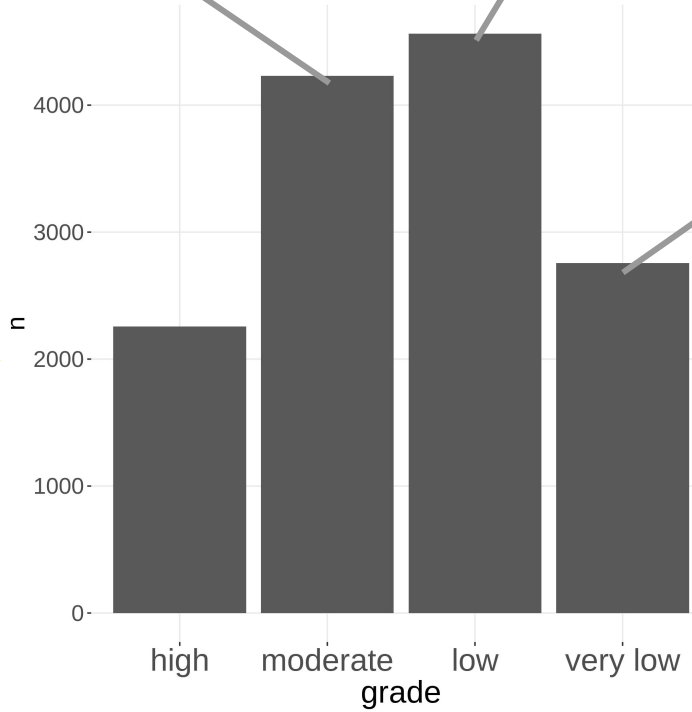
Evidence is of high quality only ⅙ of the time

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

Evidence synthesis and quality assessment

Systematic reviewing seeks to collect, summarise and **appraise** all empirical evidence that fits pre-specified eligibility criteria.

- Assuming already summarised evidence, to what extent can quality appraisal be done automatically?

- Is the task more difficult for specific question types, outcomes, medical specialties?

- What level of NL understanding is needed?

- Can we use structured data as a substitute for manual annotations?

Simon Suster, Stream 4

Constructing a quality assessment dataset from Cochrane Systematic Reviews

Ongoing studies

Excluded studies

Included studies

study 1
- Study characteristics:
    - study design and duration
    - participants, interventions, outcomes
- Risk of bias: judgement + support

full article

systematic review
authors, title, topics, year, ...

"Mediterranean-style diet for the primary and secondary prevention of cardiovascular disease"

Body text

Summaries of findings (evide

- Population, Intervention, Comparison, Outcome
- **GRADE score**
- **Reason for downgrading**
- Setting
- Absolute/relative effects and CIs
- Number of studies and participants

"In adults without cardiovascular disease, does Mediterranean diet (compared to no dietary intervention) help reduce the risk of myocardial infarction?"

low

imprecision, risk of bias