THE UNIVERSITY OF MELBOURNE

ARC TRAINING CENTRE IN COGNITIVE COMPUTING FOR MEDICAL TECHNOLOGIES

IBM

Creating a dataset for automated quality assessment of medical evidence

Simon Šuster

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

Stream 4

22/5/2020

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?

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

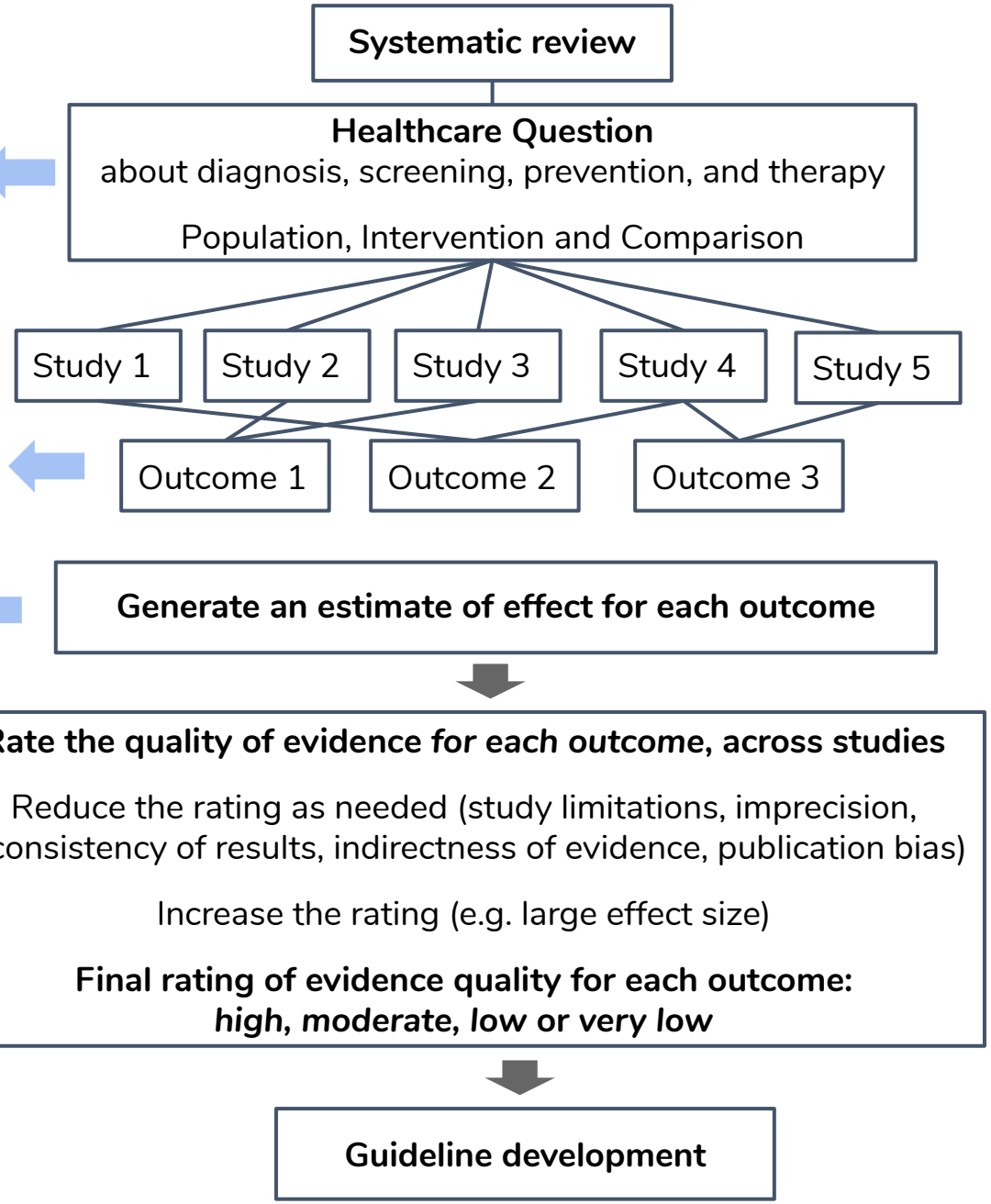
CVD mortality
stroke
myocardial infarction
total cholesterol change
...

Myocardial infarction as outcome:
Risk: 12 per 1000 (Intervention)
16 per 1000 (Control)
...

GRADE: ⊕⊕⊖⊖ (low)

Downgraded by one level for imprecision. Confidence interval is wide enough to include both an important increase or decrease in the outcome.

Downgraded by one level for risk of bias. The only included study was the PREDIMED trial retracted due to methodological issues with randomisation [...]



Modified from Guyatt et al. 2011

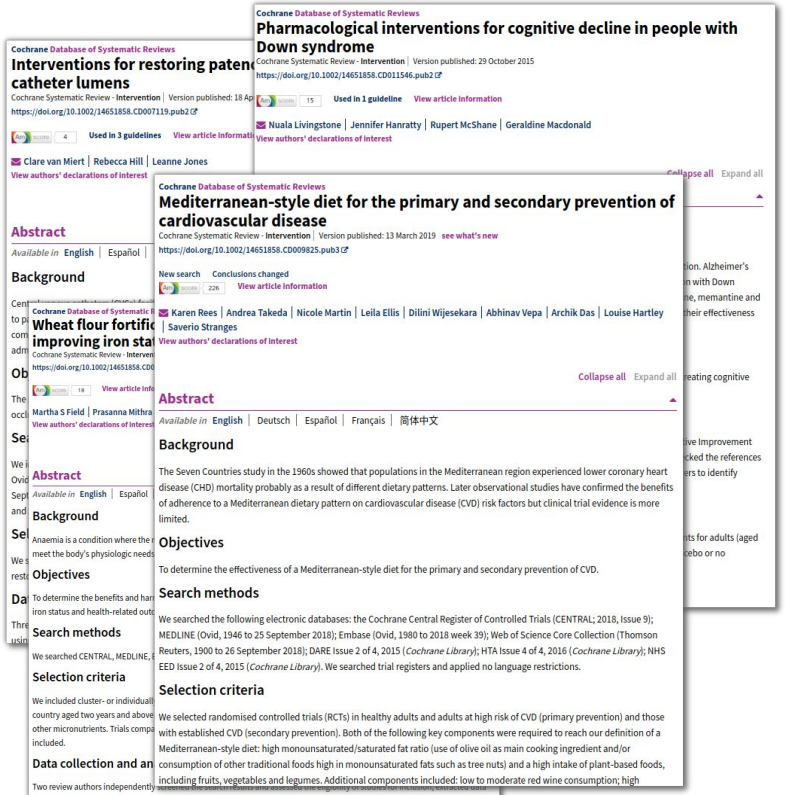
Critical appraisal in systematic reviewing using GRADE

(Grading of Recommendations, Assessment, Development and Evaluation)

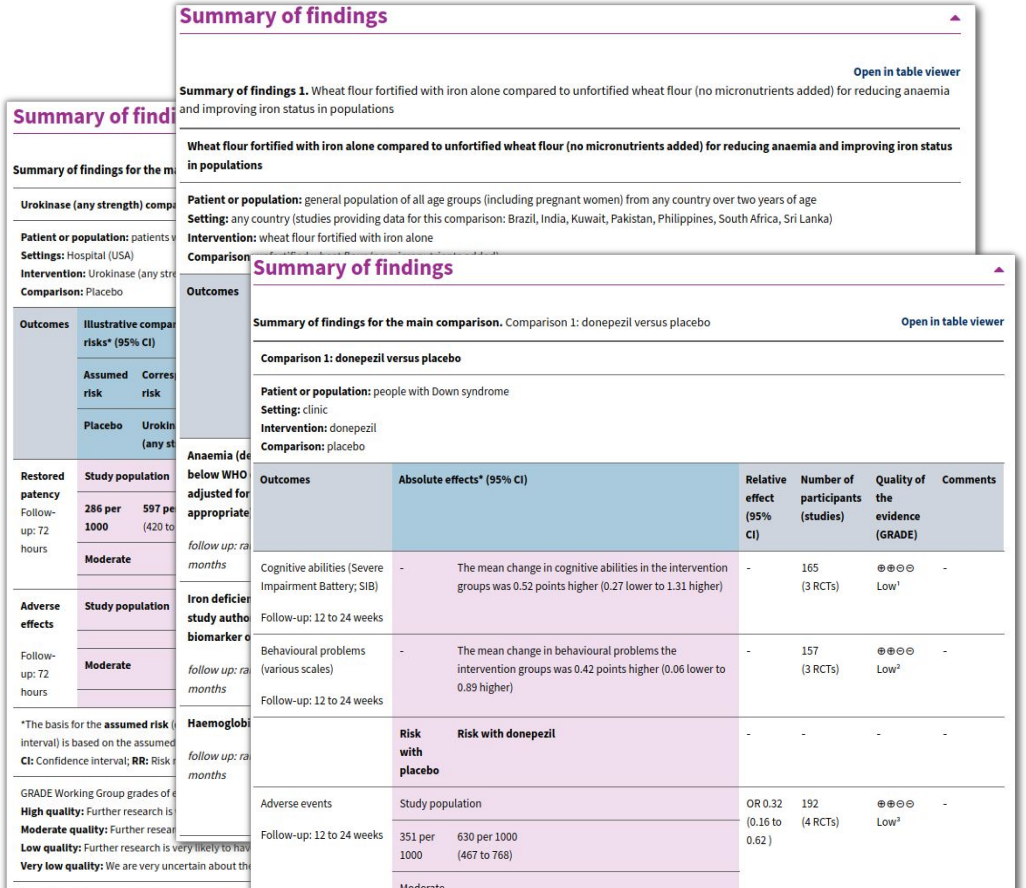
Cochrane Database of Systematic Reviews (CDSR)

- comparatively high methodological and reporting quality
- adopts the GRADE framework
- a huge human effort of reasonable consistency

~8,000
systematic reviews



~3,000 tabular summaries
of findings



~27,000
GRADEd
outcomes
with
justification

Current status

- Scraped and parsed all of CDSR into JSON files
- Quality checks and external validation of extracted data (ongoing)
- Modelling (not yet started)

Clinical question as PICO elements
Statistical data
Characteristics of included studies
Reviewers' interpretation
Implications for practice/research
Free-text summaries
…

GRADE score
Reasons for score adjustment