



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.

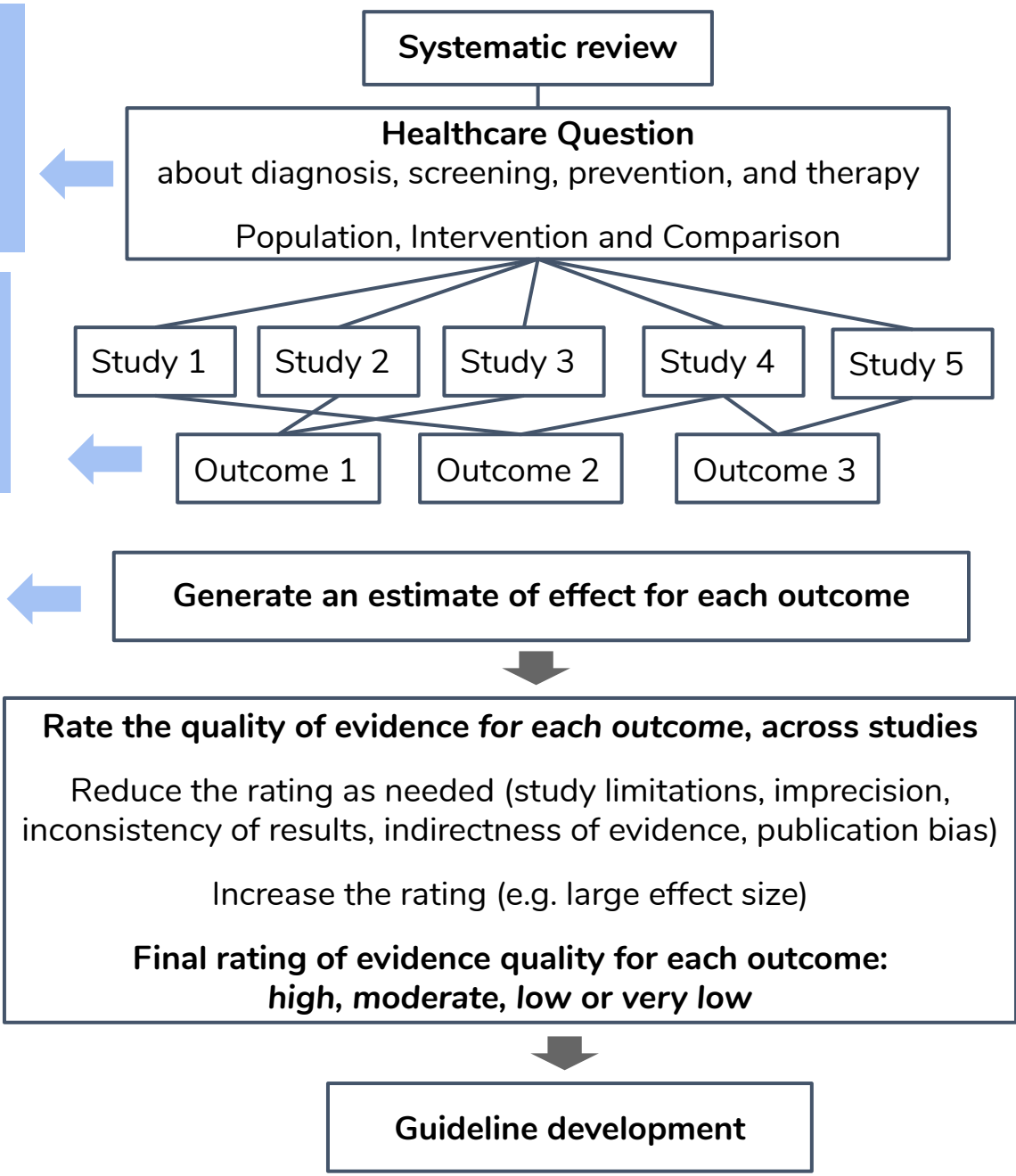
- Assuming already summarised evidence, to what extent can quality appraisal be done automatically?
- Is the task more difficult for specific question types, outcomes, medical specialties?
- What level of NL understanding is needed?
- Can we use structured data as a substitute for manual annotations?

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 [...]



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

The screenshot displays several Cochrane Database of Systematic Reviews (CDSR) abstracts. Visible titles include:

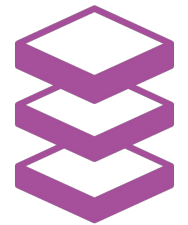
- Interventions for restoring patency of catheter lumens
- Pharmacological interventions for cognitive decline in people with Down syndrome
- Mediterranean-style diet for the primary and secondary prevention of cardiovascular disease
- Wheat flour fortified with iron alone compared to unfortified wheat flour for reducing anaemia and improving iron status in populations

 Each abstract shows the title, authors, publication date, and a brief summary of the review's background and objectives.



The screenshot shows a detailed 'Summary of findings' table. The table is organized into sections for different comparisons and outcomes. Key elements include:

- Comparison 1: donepezil versus placebo**
- Outcomes:** Cognitive abilities (Severe Impairment Battery; SIB), Behavioural problems (various scales), and Adverse events.
- Relative effect (95% CI):** -0.52 for cognitive abilities, -0.42 for behavioural problems, and 0.32 for adverse events.
- Number of participants (studies):** 165 (3 RCTs) for cognitive abilities, 157 (3 RCTs) for behavioural problems, and 192 (4 RCTs) for adverse events.
- Quality of the evidence (GRADE):** Low¹ for cognitive abilities and behavioural problems, and Low¹ for adverse events.
- Comments:** 'The mean change in cognitive abilities in the intervention groups was 0.52 points higher (0.27 lower to 1.31 higher)'.



Current status

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

