Advanced Information Retrieval on the Edge
Living Systematic Reviews

Robin Featherstone
*Information Specialist, Alberta Research Centre for Health Evidence (ARCHE), Cochrane Child Health, Canada*

Steve McDonald
*Information Specialist, Cochrane Australia, Monash University, Australia*

Anna Noel-Storr
*Information Specialist, Radcliffe Department of Medicine, University of Oxford, UK*

James Thomas
*Associate Director, Evidence for Policy and Practice Information and Coordinating (EPPI)-Centre, UK*
AGENDA

8:40-9:00  PRESENTATION 1: Introduction to living systematic reviews (LSRs)
9:00-9:15  PRESENTATION 2: LSR searching
9:15-9:25  GROUP ACTIVITY 1: LSR planning exercise
9:25-9:40  PRESENTATION 3: LSR search: assessment and maintenance
9:40-10:00  GROUP ACTIVITY 2: Data analysis
10:00-10:30  BREAK
10:30-11:00  PRESENTATION 4: LSR enablers
11:00-11:30  GROUP ACTIVITY 3: Discussion and presentations
Disclosures

• We are all members of the Living Systematic Review network
• We are all involved in either developing and/or evaluating the methods and tools to support LSR search methods
• James Thomas is responsible for the RCT Classifier
• Anna Noel-Storr is responsible for Cochrane Crowd
Presentation 1

Introduction to Living Systematic Reviews

Steve McDonald
Information Specialist, Cochrane Australia
Monash University, Australia
The review currency challenge

- Burgeoning volume of primary evidence
- Reviews take long time to produce (mean 20 months)
- Can be out of date on publication
- Updating a review not dissimilar from starting again
- Even with explicit updating commitment; difficult to maintain up-to-date review portfolio
Breaking the health evidence trade-off

- New updating model: to achieve greater currency and retain accuracy, with similar workload
What is a living systematic review?

We define an LSR as a **systematic review** that is continually updated, incorporating relevant new evidence as it becomes available. In practice, this means **continual surveillance** for new research evidence through ongoing or frequent searches and the **inclusion of relevant new information** into the review in a **timely manner** so that the findings of the systematic review remain current.

Elliott et al. 2014 *PLoS Med*
When to consider an LSR

- The review question is a priority for decision making
- Certainty in the existing evidence is low or very low
- There is likely to be new research evidence
- For the LSR pilots: sufficient capacity and resources within the author and editorial teams to maintain an LSR
Features of Cochrane LSR approach

• Applies to any review type (e.g. RCTs, qualitative)
• Retains core systematic review methods
• LSR-specific methods pre-specified in protocol
• Explicit and a priori commitment to frequent search and updating
• Starts with a standard ‘baseline’ review

• Monthly Searches
• Tell reader ‘what’s happening’ monthly
• Republish the review, informed by the evidence
Process and LSR-specific methods

- **Searching**
  - Specify frequency of all sources (databases, trial registers, etc.)

- **Screening**
  - Should match search frequency; may use ‘enablers’ (Classifiers, Crowd)

- **Data synthesis**
  - Incorporate new evidence immediately OR use decision rules to delay

- **Other**
  - Methods and searches reviewed over time
LSR production model

1. Monthly database & trial register searches (Review group)
2. 3-6 monthly searches ‘other sources’ (Authors)
3. Apply RCT Classifier (Review group)
4. Screening of potential RCTs (Authors)
5. Screening ‘low probability’ RCTs (Cochrane Crowd)
6. Via CRG
7. Study = YES
8. Study = NO

One month
Study = NO
- Alert the reader (amend UCS) (Review Group)
- (Authors)

Study = YES
- Alert the reader (amend UCS) (Review Group)
- Data extraction, risk of bias, synthesis (Authors)
- IMMEDIATE OR DELAYED INCORPORATION
Example Cochrane LSR

- Baseline review Sep 2017: 50 trials
- 1st LSR published Jan 2018: 5 new
- 2nd LSR published May 2018: 8 new
- 3rd LSR expected Jul 2018: 3 new

What's new

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>26 March</td>
<td>Amended</td>
<td>This is a Living Systematic Review. Searches are run and screened monthly. Searches run up to 25 January 2018 identified 10 new studies and 4 ongoing studies. These studies are currently being incorporated in the review (expected publication in May 2018). In addition, the team continues with monthly screening (last search date 25 February 2018) and has found an additional 2 new studies to be included in a future update.</td>
</tr>
</tbody>
</table>
Evaluation of pilot LSRs

Aims

• Explore feasibility, and implications for contributors, processes and workflows
• Identify opportunities to improve the LSR model(s)

Participants

• Author teams, editorial staff, peer reviewers
• Cochrane and non-Cochrane

Data collection

• 3 x interviews and monthly surveys

Results

• Available September 2018
Resources

Series published Sep 2017

1. Introduction – why, what, when, how
2. Combining human and machine effort
3. Statistical methods for updating M-As
4. Living guideline recommendations

Cochrane Living Systematic Reviews

Interim guidance for pilots (Version 0.3, 21 April 2017)

cochrane.org/lsr
Presentation 2
LSR Searching

Robin Featherstone
*Information Specialist, Alberta Research Centre for Health Evidence (ARCHE), Cochrane Child Health*
LSR Challenges for Information Specialists (IS)

- Translating bespoke strategies into LSR production models
- Accounting for publication bias
- Assessing and revising strategies as needed
- Maintaining strategies over time
LSR Searching Goal

Maximize efficiency while ensuring quality
Adapting Bespoke Strategies

Considerations

- How well did the original search perform?
- Can it be replicated?
- Can it be automated (in full or in part)?

What revisions to the search are needed:

- To reduce workload?
- To ensure precision?
# Search Reporting – Why Standards Matter

<table>
<thead>
<tr>
<th>Topic</th>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food environment</td>
<td>Food environment OR nutrition environment OR retail food OR neighbourhood OR neighborhood OR environment OR food desert* OR food swamp OR food availability OR food cost OR food affordability OR food pric* OR food quality</td>
</tr>
<tr>
<td>Retail food outlets</td>
<td>Supermarket OR grocery store OR convenience store OR corner store OR dollar store OR fast food OR restaurant OR food store OR bodega OR tienda</td>
</tr>
<tr>
<td>Dietary intake</td>
<td>Food OR fruit OR vegetable OR diet* OR nutrition OR processed food</td>
</tr>
<tr>
<td>Weight- and health-related outcomes</td>
<td>Obes* OR overweight OR BMI OR body mass index OR waist circumference OR anthropometric OR health OR cardiovascular OR cancer OR diabetes OR hypertension OR disease OR illness</td>
</tr>
<tr>
<td>Socio-economic status</td>
<td>Income OR disparity OR equity OR inequity OR inequality OR disadvantage OR poverty OR depriv* OR marginaliz*</td>
</tr>
<tr>
<td>Canada</td>
<td>Canada OR Canadian OR British Columbia OR Alberta OR Saskatchewan OR Manitoba OR Ontario OR Quebec OR Nova Scotia OR New Brunswick OR Prince Edward Island OR Newfoundland OR Yukon OR Northwest Territories OR Nunavut</td>
</tr>
</tbody>
</table>

* A Boolean search function indicating truncation, allowing multiple forms of a given word (e.g., depriv* identifies deprived, deprivation.)
Automated Alerts

Considerations

• Do search sources support auto-alerts?
• Will auto-alerts match planned update frequency?
• How to apply date limits?
• How to identify and remove duplicates?
• Who will manage results?
Objectives of LSR Auto-alerts

Retrieve precise, ready-to-screen, unique search results on predictable intervals
Facilitated LSR Searching

Complementary search approaches

• Simplified search strategies
• Cited/citing references
• Handsearching

Grey literature

• Trial registers
• Agency reports
• Industry websites
• Funded research databases
Facilitated Searches in LSR Production Models

Considerations

• Which grey literature sources are likely to yield new unique evidence?
• How to integrate results with auto-alerts?
• Should frequency be the same as auto-alerts (in full or in part)?
• Who should conduct?
Objectives of LSR Facilitated Searching

Complement (not replicate) auto-alerts
Group Activity 1
Search planning exercise
Discussion questions for small groups

1. Is the topic appropriate for a living systematic review?
2. How will you ensure new evidence is retrieved on a monthly basis?
Discussion questions for small groups

1. Is the topic appropriate for a living systematic review?

2. How will you ensure new evidence is retrieved on a monthly basis?

Group A
What influence does marijuana legislation have on attitudes and behaviours towards tobacco smoking?

Group B
For patients with Non-Small Cell Lung Cancer (NSCLC), what is the cost-effectiveness and economic impact of second-line treatment with nivolumab, pembrolizumab and atezolizumab, with and without the use of PD-L1 (Programmed death-ligand 1) testing, for patient selection?
Presentation 3
LSR search: Assessment and Maintenance

Robin Featherstone
Information Specialist, Alberta Research Centre for Health Evidence (ARCHE), Cochrane Child Health
LSR Challenges for Information Specialists (IS)

- Translating bespoke strategies into LSR production models
- Accounting for publication bias
- Assessing and revising strategies as needed
- Maintaining strategies over time
Potential Pitfalls of LSR Searching

I never make the same mistake twice. I strive for five or six times.

© Gabriel UTAJ 04-09-07
Assessing LSR Searches

Considerations

- Which assessment methods to use?
- How to maximize knowledge gained from the review process?
Case Study: Retrospective Search Assessment

- Data from a 2018 update of a 2016 systematic review
- Number of included studies = 139

<table>
<thead>
<tr>
<th>Database</th>
<th>Records retrieved</th>
<th>Total included studies retrieved</th>
<th>Precision</th>
<th>NNTR (Number Needed to Read)</th>
<th>Unique studies retrieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDLINE</td>
<td>4037</td>
<td>118</td>
<td>0.029</td>
<td>34</td>
<td>3</td>
</tr>
<tr>
<td>Embase</td>
<td>2623</td>
<td>121</td>
<td>0.046</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>970</td>
<td>128</td>
<td>0.132</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>CINAHL</td>
<td>254</td>
<td>34</td>
<td>0.134</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>
Maintenance Required

Considerations

• Is the topic question maturing over time?
• What new indexing terms (e.g., MeSH) are available?
• What database enhancement will impact auto-alerts?
• Which new evidence sources are available?
• How often should the strategy be examined?
Objectives of LSR Search Assessment and Maintenance

Improve performance and enhance strategies on a recurring basis

Created by Yu Luck from Noun Project
Group Activity 2
Data analysis
Discussion questions for small groups

1. What do the precision, NNTR and unique studies data tell you about the original search?

2. How can we enhance our search based on these data?
COFFEE BREAK
Presentation 4
LSR Enablers

James Thomas
Associate Director, Evidence for Policy and Practice
Information and Coordinating (EPPI)-Centre, UK

Anna Noel-Storr
Information Specialist, Radcliffe Department of Medicine,
University of Oxford, UK
Enablers in a nutshell…

Involves taking research curation outside the confines of individual reviews

Uses three core ‘technologies’:

1. Human effort in Cochrane Crowd
2. Machine Learning
3. Re-use of data
Systematic review ‘pain points’

- Search: 0 hours
- Citation screening*: 120 hours
- Importing full-text articles: 10 hours
- Full-text review: 20 hours
- Data extraction~: 80 hours
- Risk of bias assessment: 10 hours
Cochrane Crowd

You can make a difference!
Becomes a Cochrane citizen scientist. Anyone can join our collaborative volunteer effort to help categorise and summarise healthcare evidence so that we can make better healthcare decisions.

Give it a try

A platform for crowdsourced **micro-tasks**
that helps produce high quality health evidence
Reviews are complex but they are made up of a number of rule-driven, systematic tasks. These kinds of tasks can be re-formed as ‘microtasks’.
The microtask: is it an RCT?

Restricted versus continued standard caloric intake during the management of refeeding syndrome in critically ill adults: A randomised, parallel-group, multicentre, single-blind controlled trial. [2015522581]

Background: Equipoise exists regarding the benefits of restricting caloric intake during electrolyte replacement for refeeding syndrome, with half of intensive care specialists choosing to continue normal caloric intake. We aimed to assess whether energy restriction affects the duration of critical illness, and other measures of morbidity, compared with standard care. Methods: We did a randomised, multicentre, single-blind clinical trial in 13 hospital intensive care units (ICUs) in Australia (11 sites) and New Zealand (two sites). Adult critically ill patients who developed refeeding syndrome within 72 h of commencing nutritional support in the ICU were enrolled and allocated to receive continued standard nutritional support or protocolised caloric restriction. 1:1 computer-based randomisation was done in blocks of variable size, stratified by enrolment serum phosphate concentration (>0.32 mmol/L vs <0.32 mmol/L) and body-mass index (BMI; >18 kg/m² vs <18 kg/m²). The primary outcome was the number of days alive after ICU discharge, with

Cochrane Citizen Scientists can see a title and an abstract and have to decide whether they think the record is describing a randomised trial
More microtasks

Identifying

- Is it an RCT? YES
- Is it a DTA? YES

Describing

Available
- Planned / in beta

P I C O

P E C O
Results: collective accuracy

<table>
<thead>
<tr>
<th>Info specialist and methodologist</th>
<th>True positives 457</th>
<th>False positives 58</th>
</tr>
</thead>
<tbody>
<tr>
<td>False negatives 4</td>
<td>True negatives 5522</td>
<td></td>
</tr>
</tbody>
</table>

**Sensitivity:** 99.1%  **Specificity:** 99.0%

Natural sample; blinded to crowd decision; dual independent screeners as reference standard.
Results to date

- **75,000** trials found
- **2 million** classifications
- **9000** contributors
- **99%** accuracy
Text mining

Deriving high-quality information from text

Machine learning

Models that learn from data to make predictions or decisions
Machine classifiers

Abstract

A random allocation study was performed to compare the clinical value of monitoring of apparently normal pregnancies for fetal well-being. Patients were excluded. In the group counting fetal movements, 577 women recorded assessments had an average of five determinations of serum total IgG respectively. Women having hormone assessments were seen more frequently for management did not differ. Women making fetal movement counts were noted to be less likely to have a C-section.

Notes

Abstract

RCTs

non-RCTs
RCT Classifier

- First iteration is implemented in CRS Web and EPPI-Reviewer
- Trained on 400,000 classifications by the Crowd
- Provides a score for each citation (0-100)
- Recall of 99.8% at 10% threshold

- 25,000 records
- 15,655 very unlikely to be RCT
- **99.9% correct**
Deployed in Cochrane Register of Studies
Cochrane, the Crowd and the Machine

The story so far

- 2 million classifications
- 500,000+ citations
- 75,000 RCTs

Since January 2017

Machine reduces the amount for the Crowd to screen by 50-70%

Crowd identifies around 500 difficult to identify reports of RCTs

A perfect partnership
Machine/Crowd synergy

- Previous model
- Crowd 2016
- Crowd & Machine 2017

Number of records

- 300K
- 250K
- 200K
- 150K
- 100K
- 50K

WS12: Living Systematic Reviews
HTAi 2018, Vancouver
Efficiency: four completed pilots

<table>
<thead>
<tr>
<th>N</th>
<th>RCT</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3635</td>
<td>872</td>
<td>76% reduction</td>
</tr>
<tr>
<td>4913</td>
<td>831</td>
<td>83% reduction</td>
</tr>
<tr>
<td>1200</td>
<td>370</td>
<td>69% reduction</td>
</tr>
<tr>
<td>3424</td>
<td>1446</td>
<td>58% reduction</td>
</tr>
</tbody>
</table>

1 day 3 days 4.5 hrs 5 days
Re-using existing data

What evidence do we have so far?

11 review updates published in Sept 2016 that were looking for RCTs and had searched Embase using a methodological filter
Crowd has already screened between 62% to 98% (mean 83%) of records identified from Embase searches done for review updates.
‘Screen 4 Me’ workflow

A new service which is just being rolled out for Cochrane authors

1. Upload search results
2. Non-RCTs removed using:
   a) Data reuse
   b) Machine learning
   c) Crowdsourcing
3. Remaining records returned to authors

Offers huge efficiencies for these reviews
‘Screen 4 Me’ workflow

Start: conduct usual review searches

Are these records already known NOT to be RCTs?

No

Are these records very unlikely to be RCTs?

No

Are these records RCTs according to Cochrane Crowd?

No

End (Manual screening of remainder)

Yes

Existing data
Use of enablers by LSR pilots

Pilot 1 – reviews on anticoagulation in cancer patients
• Monthly searches of Cochrane, MEDLINE and Embase
• ~50-130 citations p/month sent to RCT Classifier
  • Citations scoring 10-99 (~20-50 p/month) >>>> AUTHORS
  • Citation scoring 0-9 (~30-80 p/month) >>>> CROWD
• 1 new RCT; no RCTs from among citations sent to Crowd

Pilot 2 – fruit and veg review
• Monthly searches of Cochrane, MEDLINE and Embase
• 2,600 citations since Sep 2017 sent to RCT Classifier
  • Citations scoring 10-99 (1260, ~50%) >>>> AUTHORS
  • Citation scoring 0-9 (1340) >>>> CROWD
• 16 new RCTs; 1 RCT from among citations sent to Crowd
What are the statistical features of the study? Full text retrieval and data extraction

What are the characteristics of the study? eg RCT, DTA...
A probability is assigned

What is the study design? Population, Intervention and Outcomes?

Finding and classifying relevant research through human and machine effort

http://community.cochrane.org/tools/project-coordination-and-support/transform
A Living Systematic Review is a systematic review which is continually updated, incorporating relevant new evidence as it becomes available.
Group Activity 3
Discussion and presentations
Discussion questions

- Finalize and present your plan for a “living” search

Questions

1. What is your topic? Is this a good candidate for an LSR?
2. How will you ensure new evidence is identified each month?
3. How will you ensure search performance over time?
4. What enablers could assist your LSR production model?
Summary and Recommendations

• LSRs are an emerging approach to evidence syntheses

• LSR searches combine automated and facilitated strategies

• LSRs provide opportunities for continual improvement of search performance

• LSR production models embrace enabling technology
References & Recommended Resources


Cochrane Community. Living Systematic Reviews. 2018. Available from: http://community.cochrane.org/review-production/production-resources/living-systematic-reviews


Acknowledgements

• Annie Synott (Living Systematic Review Network)
• Jo Brooker, Justin Clark, Jo Platt and Charlene Bridges (LSR pilots)

Funding

• Cochrane (Project Transform)
• Australian National Health and Medical Research Council

More Info

• cochrane.org/lsr
• Slides available under ‘Other LSR Resources’