Unlocking the Cochrane Data Vault

Deirdre Beecher  Senior Metadata Specialist
Chris Mavergames  Head of Informatics and Technology (IT) Services/Chief Information Officer

Conflicts of interest

Chris Mavergames and Deirdre Beecher are salaried employees of Cochrane.
Workshop goals

• Understand how Cochrane is developing PICO (Patient-Intervention-Comparison-Outcomes) metadata to describe and enhance access to Cochrane systematic reviews and their included studies

• Explore approaches and scenarios to utilize PICO metadata in literature searching
Summary

• Presentation (15 mins)
  • Background to Cochrane Linked Data Project
  • PICO annotation
  • Ongoing development of the search tool
• Exercise - using PICOfinder search tool (prototype) (30 Mins)
• Discussion (25 mins)
• Close (5 mins)
Linked Data

- Enrich our content and data with metadata using controlled vocabularies (SNOMED CT etc.)
- Construct knowledge models and structures (ontologies) that will allow re-use of this metadata (annotations) for both downstream (dissemination) and upstream (production) use
- Become more interoperable with other projects, products, datasets, and systems (for example, NICE guidelines)
- Improve production and dissemination of evidence
Making existing research data “FAIR”

Findable
Accessible
Interoperable
Reusable
How do we prepare the data?

**Digoxin for preventing or treating neonatal respiratory distress syndrome** CD001080

<table>
<thead>
<tr>
<th>Population</th>
<th>Interventions</th>
<th>Comparisons</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preterm Infant (Less than 37 weeks)</td>
<td>Pharmacological</td>
<td>No active treatment</td>
<td>Mortality, Physiological or clinical, Death</td>
</tr>
<tr>
<td>Neonatal Respiratory Distress Syndrome</td>
<td>Digoxin</td>
<td>Placebo</td>
<td>Respiratory Distress, Bronchopulmonary Dysplasia</td>
</tr>
<tr>
<td>Birth to 1 mo, Male and Female</td>
<td></td>
<td></td>
<td>Chronic Lung Disease, Intraventricular Hemorrhage</td>
</tr>
</tbody>
</table>
What is PICO annotation?

- **Patient, Population or Problem**
  - What are the characteristics of the patient or population?
  - What is the condition or disease of interest?
- **Intervention**
  - Drug therapy, surgery, education etc
- **Comparison**
  - Usual care, placebo, different drug etc
- **Outcome**
  - Quality of life, change in clinical status, morbidity, adverse events
What data are we annotating?

- Present
  - Cochrane systematic reviews (inclusion criteria) published from 2015
- Included studies – only done for some topics
- Future
  - Meta-analyses – automation/design change to annotation widget in progress
Dynamic study discovery via Evidence Pipeline

Goal: to create a dynamic discovery system for the upfront identification and description of human studies.

Evidence Pipeline
Finding and classifying relevant research through human and machine effort
Machine Curation (PICO)

Using machine learning to identify and filter evidence prospectively

In [7]: r = requests.post("http://104.41.231.151:5000/annotate", json=json.dumps({"source": 'cochrane-review', 'task': 'pico', 'data': {'cdno': 'CD006064', 'characteristics': {'participants': 'All pregnant women attending antenatal care at least once.', 'outcomes': 'The primary outcome measure is the rate of breastfeeding initiation in all pregnant women after birth (as defined by trial authors). Secondary outcomes include: \n', 'interventions': 'Breast examination, for any purpose, conducted at least once during an antenatal care visit, compared with \'usual\' care (that is, that which does not include antenatal breast examination). "}, 'annotator-id': 'unique annotator ID'}}))

In [8]: r.text
Using crowdsourcing to perform complex annotations as a series of micro-tasks

Early inhaled steroid use in extremely low birthweight infants: A randomised controlled trial. [201631]

Objective We hypothesised that a prophylactic inhaled steroid would prevent the progression of bronchopulmonary dysplasia (BPD) in extremely low birthweight infants (ELBWIs). Design This study was a multicentre, randomised, double-blinded, placebo-controlled trial. Setting This investigation was conducted in 12 level III neonatal intensive care units (NICUs). Patients A total of 211 ELBWIs requiring ventilator support were enrolled. Intervention Starting within 24 h of birth and continuing until 6 weeks of age or extubation, two doses of 50 µg fluticasone propionate (FP) or placebo were administered every 24 h. Main outcome measurement The primary outcome measure used to indicate the morbidity of severe BPD incidence was death or oxygen dependence at discharge from the NICU. The secondary measures were neurodevelopmental impairments (NDIs) at 18 months of postmenstrual age and 3 years of age. We performed subgroup analyses based on gestational week (GW) and the presence of chorioamnionitis (CAM). Results Infants were randomised into the FP (n=107) or placebo (n=104) groups. No significant differences were detected between the FP and placebo groups with respect to either the frequency of death or the oxygen dependence at discharge or NDIs. In subgroup analyses, the frequencies of death and oxygen dependence at discharge were significantly decreased in the FP group for infants born at 24-26 GWs and for infants with CAM, regardless of the GW at birth. Conclusions Inhaled steroids have no effect on the prevention of severe BPD or long-term NDI but might decrease the severity of BPD for ELBWIs with a risk factor. Trial registration number UMIN-CTR C000000405. Copyright Â© 2016 BMJ Publishing Group Ltd & Royal College of Paediatrics and Child Health.
Who is doing the work?

- Cochrane Review Group Staff
- Cochrane Information Specialists
- Health Sciences Information Specialists
- Healthcare Professionals
- PICO annotation Central Support Team
Progress to date – review annotations

- Complete: 3533
- In Progress: 878
Progress to date – study annotations

- Complete: 6191
- In Progress: 2122
The Cochrane Vocabulary

- Using already developed ontologies
- Creating terms where not available in these ontologies
- Relabelling some terms but keeping the hierarchy imported from the ontology
- Mapping/Merging terms to make it easier to select a term
Terms added to our vocabulary since 2017

- Population/Condition: 255
- Intervention/Comparison: 731
- Outcome: 473
Intervention classification

• 15 item classification of interventions from Davey et al (2011)
• No active treatment used for Comparisons (sham, placebo, etc.)

‘The COMET (Core Outcome Measures in Effectiveness Trials) Initiative brings together people interested in the development and application of agreed standardised sets of outcomes, known as ‘core outcome sets’ (COS)’
How do we access this data?

- For now we use the PICOfinder
- This is ONLY a prototype
- Proof of concept for the principle of PICO searching
- For retrieval testing metadata in progress
How the PICOfinder prototype has helped to develop searching tools

- Usefulness of PICO search
- Functionality needed for possible future PICO search platform
- How PICO search might be used in conjunction with Cochrane Library search
- How PICO search might be used in conjunction with search on other platforms
PICO search – present & future development

- A PICO metadata search widget integrated into the Cochrane Library
- At the point of launch, search to be used with Cochrane reviews’ metadata
- Suitable for search across different content type e.g. included studies
- Cochrane Colloquium
  - Presentation of metadata in Cochrane Library Search Results
  - Find reviews with similar PICOs
  - New PICO search tab within Advanced search
So we would like to …

- Search Cochrane Reviews (or included studies) with greater speed, precision, sensitivity
- And complement traditional searching on the Cochrane Library
Acknowledgements

Anna Last – Medical Terminology Manager
Lorne Becker – Product Owner/Clinical Expert
Carol Friesen – Metadata Specialist, Cochrane
Robin Featherstone – Information Specialist, Editorial Methods Dept. Cochrane
Cochrane Review Group Staff
Cochrane Information Specialists
Volunteers who are annotating or doing quality assurance (QA)
Exercise

PICOfinder prototype

https://data.cochrane.org/pico-finder2
Questions – provided on handouts

• How did you construct your search?
• How many reviews did you find?
• Would you be confident that you have retrieved all reviews for this topic?
• In what situations would you use this tool? i.e. teaching, assisting other researchers, as a clinical librarian etc.
• How long did it take to construct your search?
• What did you like about PICOfinder?
• What didn’t you like about PICOfinder?
• To whom would you recommend PICOfinder?
• How easy was it to search PICOfinder? (on a scale of 1-5, 1=very easy, 5=very hard)
Discussion

• How could PICO metadata change the way you search for health evidence?
• What avenues would you like Cochrane to explore with PICO metadata?
Contacts

Deirdre Beecher
Senior Metadata Specialist
dbeecher@cochrane.org

Chris Mavergames
Head of Informatics & Technology Services/Chief Information Officer
cmavergames@cochrane.org