Cochrane Crowd:
Leveraging human and machine capability for living evidence syntheses

NIHR ESU Living Evidence Syntheses Working Group Meeting
15 January 2026

Anna Noel-Storr

Trusted evidence.
Informed decisions.
Better health.

Cochrane

What will I cover?

- **Why** build Cochrane Crowd?
- **What** is Cochrane Crowd and how does it work?
- **Supporting** living evidence syntheses
- **Vision/next steps** for the future

Why build Cochrane Crowd?

We are struggling to keep pace with the amount of research produced

Why build Cochrane Crowd?

“Pinpointing needles in giant haystacks” Shemilt 2017

We have a huge **specificity problem** in information retrieval for evidence synthesis.

Questions are becoming more complex.

It’s hard to find what we need without retrieving large quantities of noise.

Why build Cochrane Crowd?

Unfindable and un-FAIR

Siloed working and siloed data resulting in significant duplication of effort and research waste. Often starting every question from scratch.

Why build Cochrane Crowd?

- ***Improve discoverability of primary research***
  - *To know what has been done, and where the gaps are*
- ***Speed up 'the first third' of review production***
  - *From question formulation to identifying includable studies*
- ***Better enable the reuse of data/metadata***
  - *Stop starting from scratch every time and support efforts towards living evidence*
- ***Provide meaningful opportunities for contribution***
  - *Enable anyone with an interest to be able to join the collective effort*

What is Cochrane Crowd and how does is work?

Howe 2006: coined the term crowdsourcing

> *"the act of a company or an institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call"*

Crowdsourcing

Estelles-Arolas 2012: produced a more exhaustive definition, having identified multiple differing definitions. They identified eight key characteristics:

(i) There is a clearly defined crowd
(ii) There exists a task with a clear goal
(iii) The recompense received by the crowd is clear
(iv) The crowdsourcer is clearly identified
(v) The compensation to be received by the crowdsourcer is clearly defined
(vi) It is an online assigned process of participative type
(vii) It uses an open call of variable extent
(viii) It uses the internet

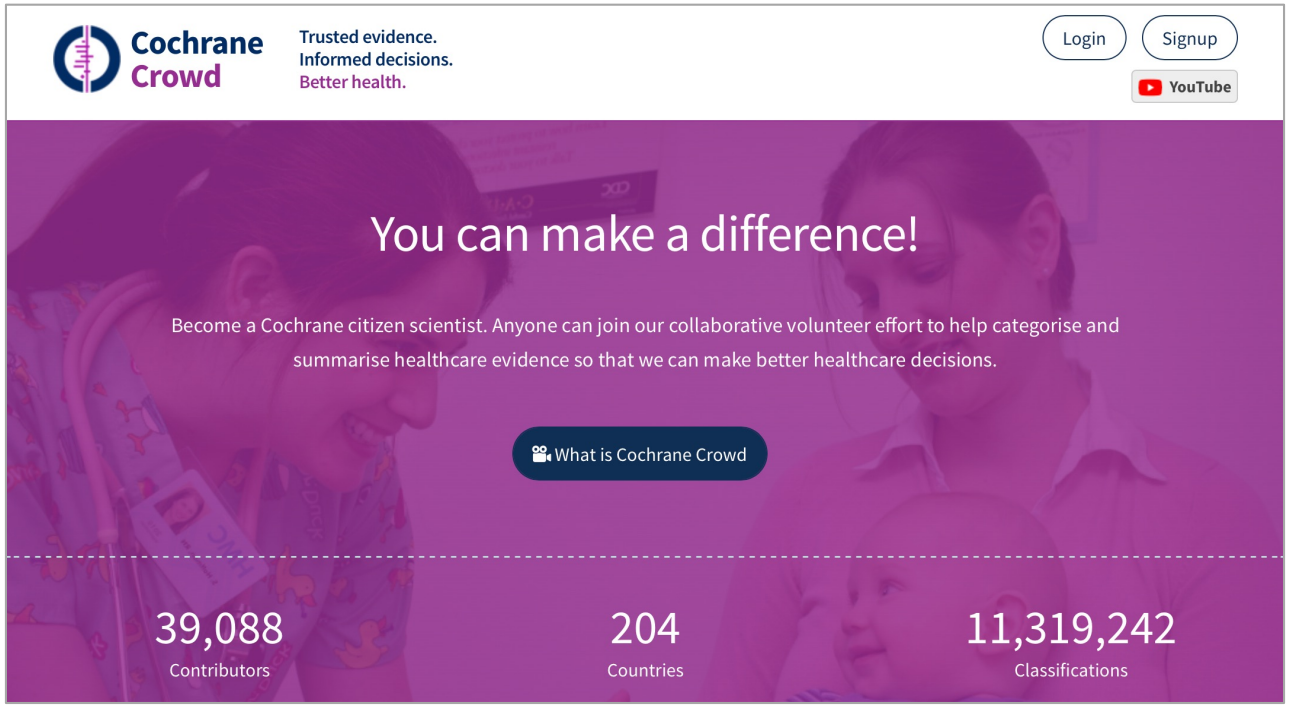
And there are different types of crowdsourcing. We use a type often called…

Crowdsourcing type: microtasking

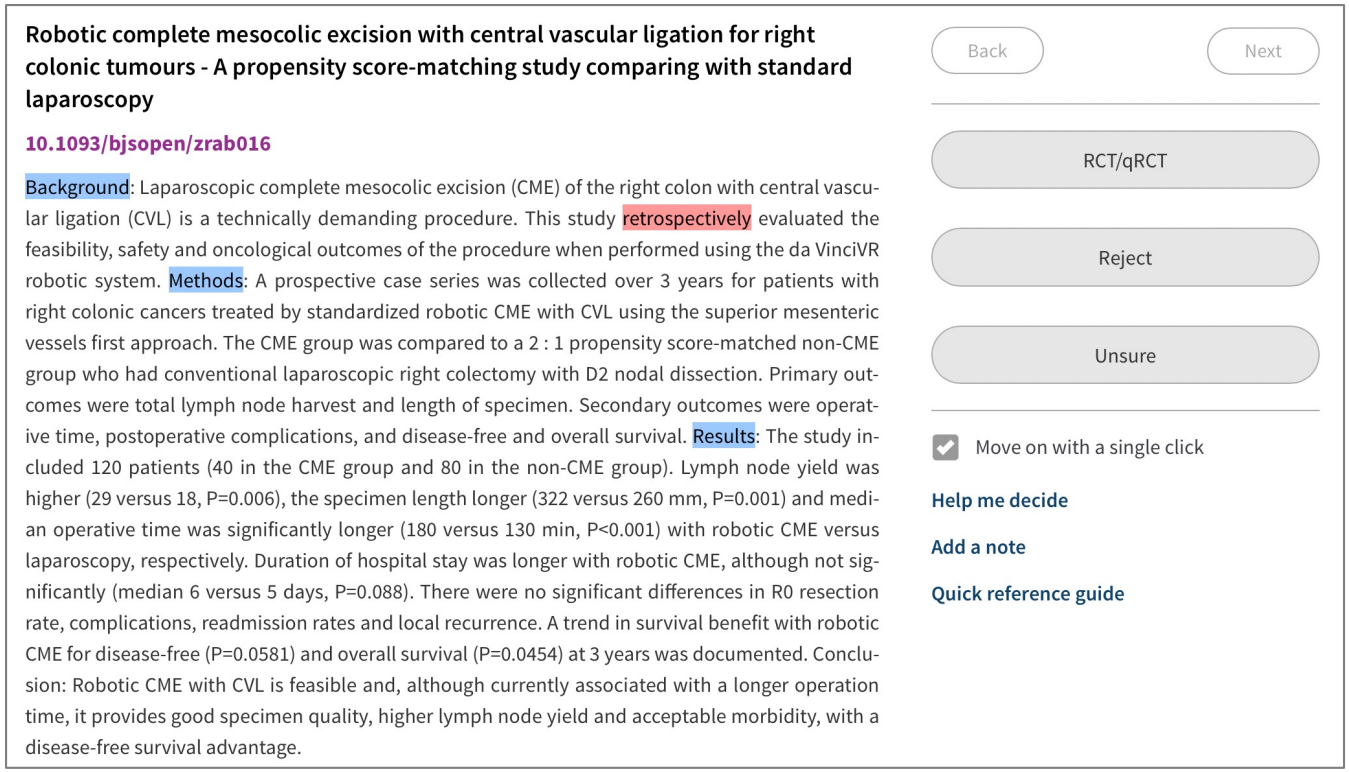
Breaking down a large corpus of data into smaller units and distributing those units to a large online crowd

"The distribution of small parts of a problem"

Launched on Valentines Day



https://crowd.cochrane.org

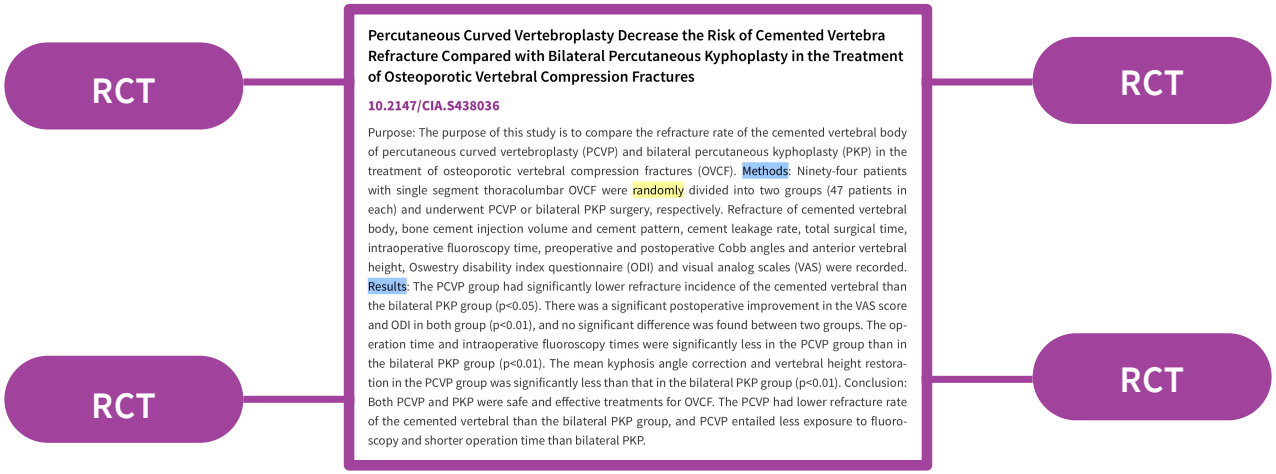


RCT Identification

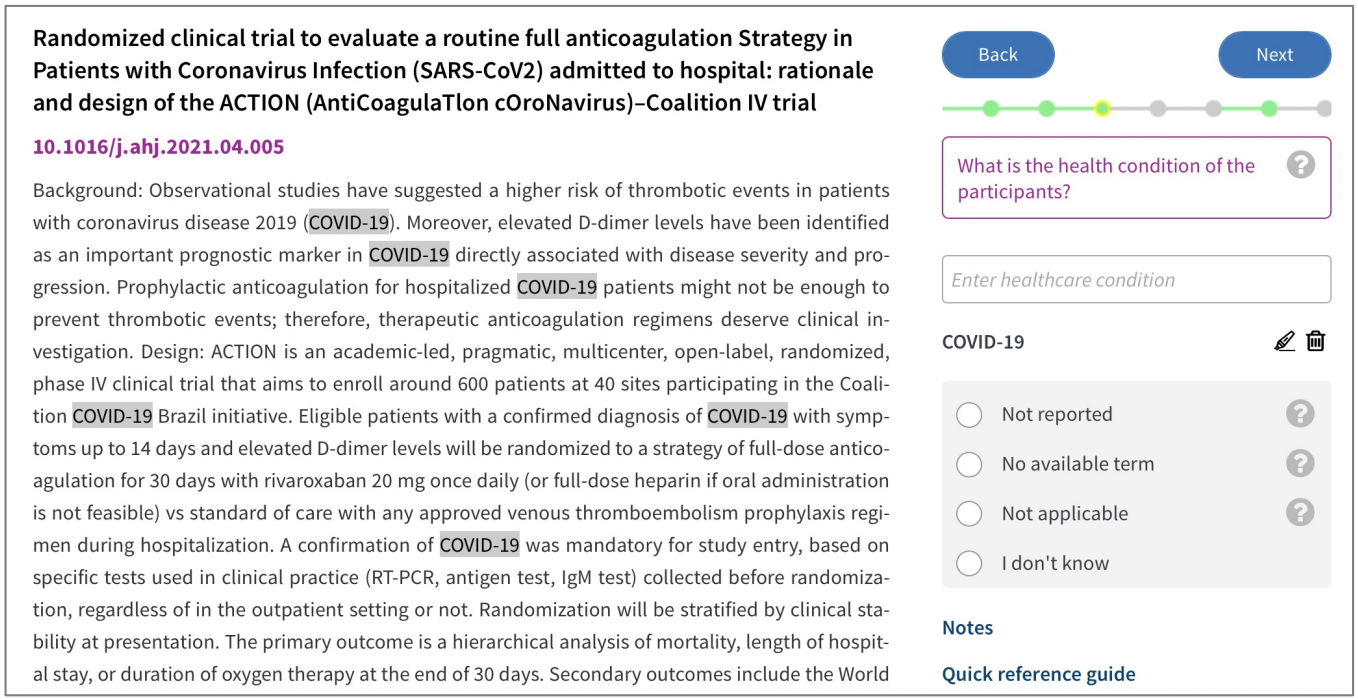
A “mainstream” task on Cochrane Crowd.
Our first task. Crowd have identified thousands reports of RCTs. Records not indexed as RCT.

Efficiency and Accuracy

The Crowd model is both efficient and accurate



The agreement algorithms mean that the final classification given to a record is for most microtasks 99% correct



PICO Extract

First hybrid task using machine classifier suggestions integrated in the task

Track your activity and progress

How many tasks you have contributed to…

My Crowd activities summary

View summary

Here's a summary of your Cochrane Crowd activity to date. For more detail and download options, click View summary.

You have been signed up since **February 2014**

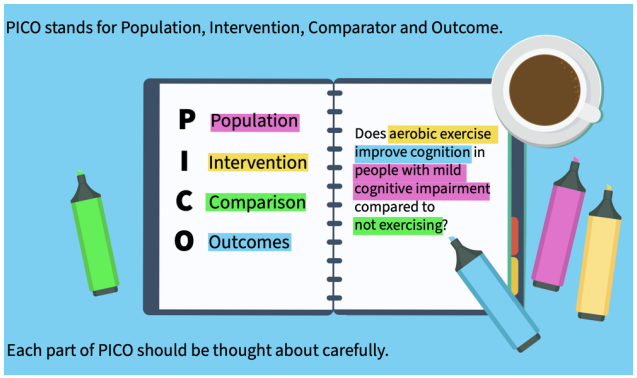
You have contributed to **79 tasks**

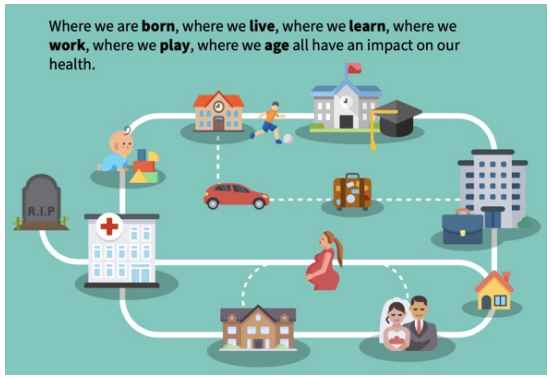
You have Earned **22 badges**

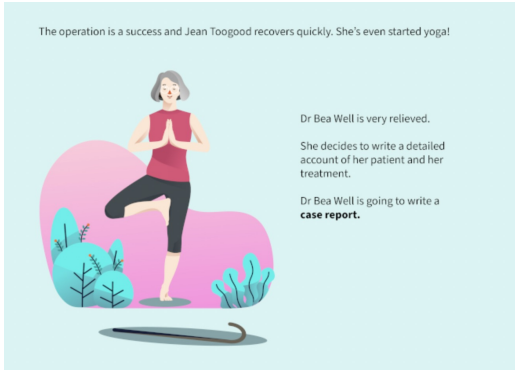
You have participated in **148 challenges**

Supporting beginners

The microtasks are supported by microlearning



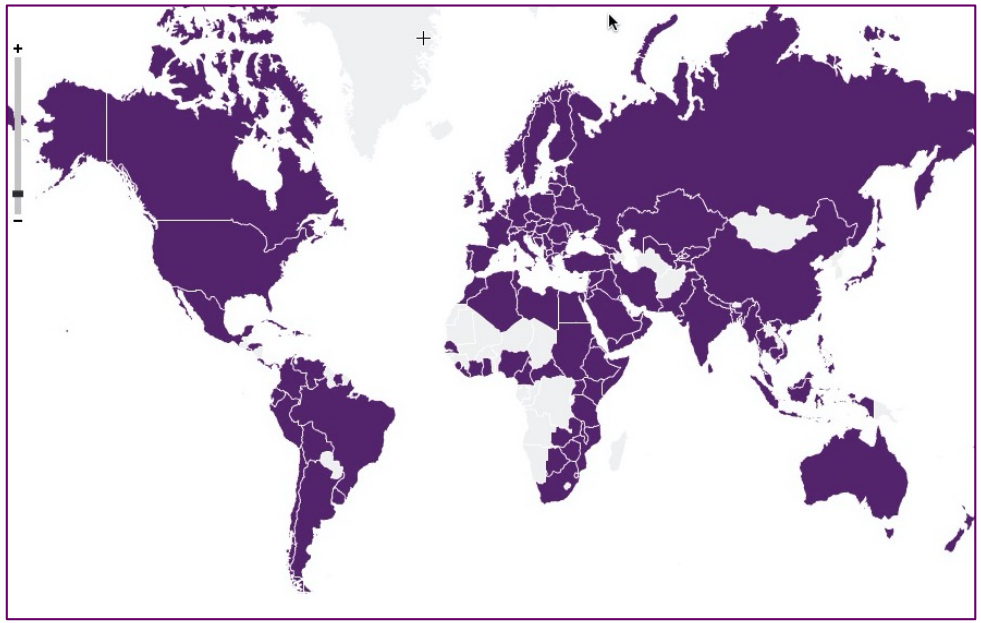




Bite-sized learning modules that complement the tasks:
Key concepts in health research, understanding study designs and introducing health equity

Crowd characteristics

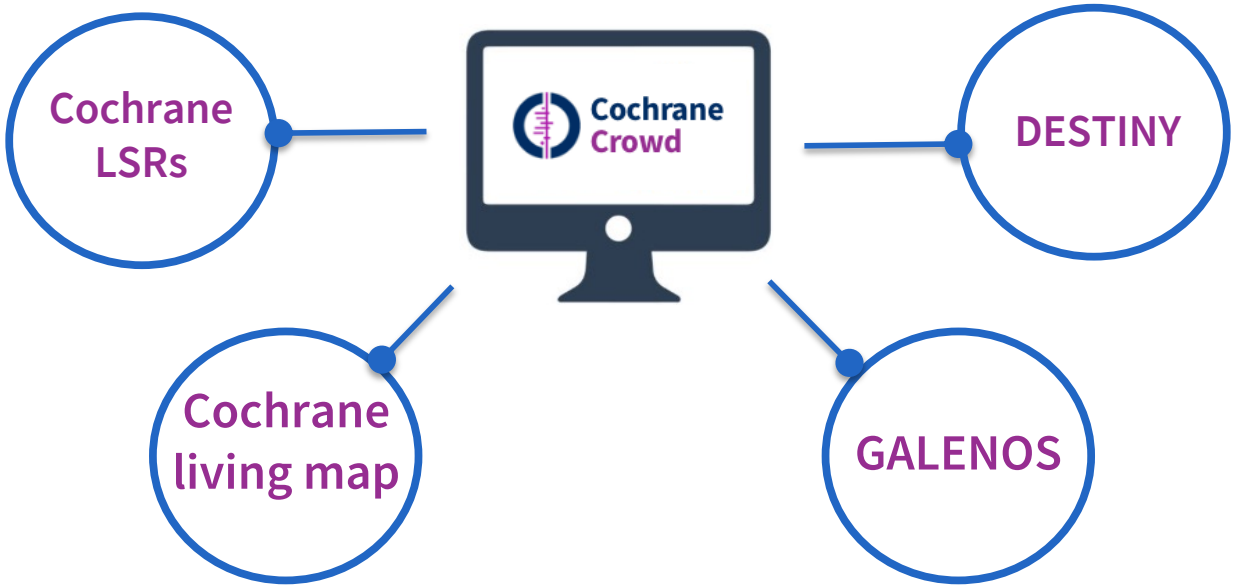
Over 39,000 people have joined Cochrane Crowd



- 46% educated to post-graduate level
- 19% don’t have a degree
- 24% completely new to health research
- 33% had no or little idea of SRs
- 20% involved in review production
- 41% student in health-related area
- 32% aged 17-24 years

Supporting living evidence synthesis

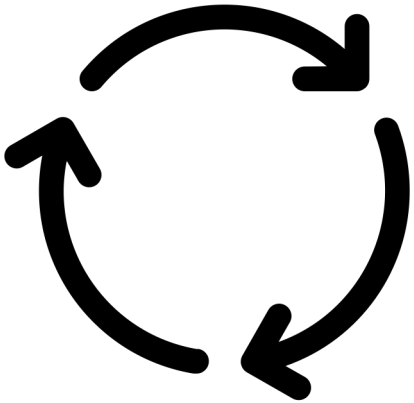
The Crowd have been involved in various living evidence syntheses



Supporting living evidence synthesis

Two Cochrane public health reviews which use Screen4Me

Cochrane living systematic reviews



Helping specific reviews via a workflow that uses crowd and machine called Screen4Me

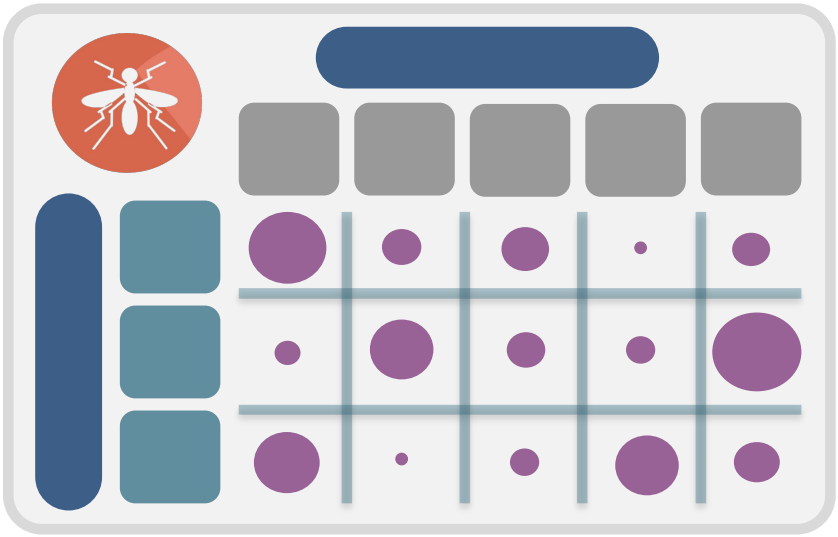
Over 200 reviews have used Screen4Me. Mean workload reduction in results screening: 72%

Every two months, the search results are loaded into the Screen4Me workflow. They go to the classifier first, then anything left goes to Crowd

Supporting living evidence synthesis

A living evidence and gap map on dengue prevention

Cochrane living gap map

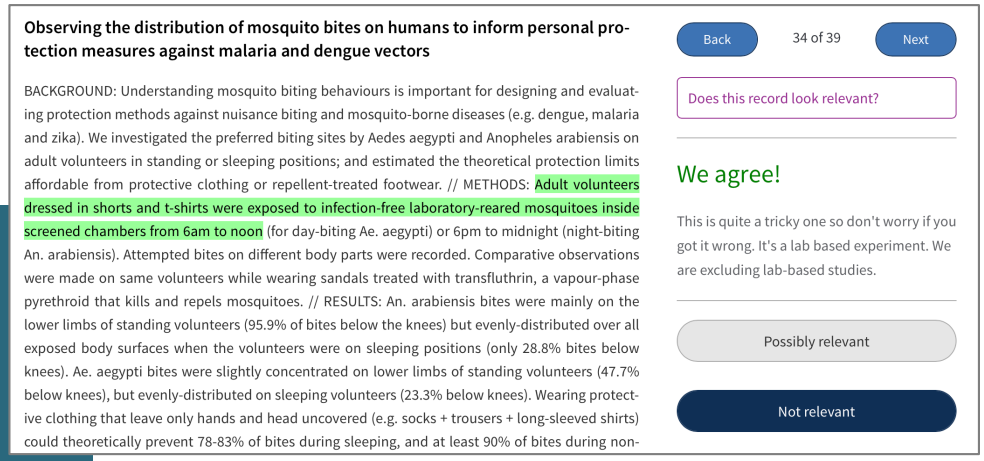


Not just supporting Cochrane reviews but also a Cochrane living evidence and gap map: dengue prevention.

Created a bespoke Crowd task to provide initial eligibility screen.

Supporting living evidence synthesis

A living evidence and gap map on dengue prevention

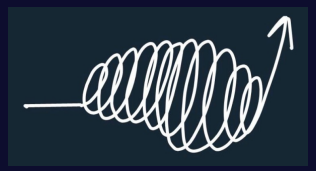


Once we'd built the task, we tested it on a gold standard dataset made up of 2000 records screened by the core author team. Crowd had to achieve over 97% recall for us to go live with the task. The Crowd achieved 99% recall.

Supporting living evidence synthesis

Four non-Cochrane living systematic reviews on the topic of depression, anxiety and psychosis

The GALENOS project



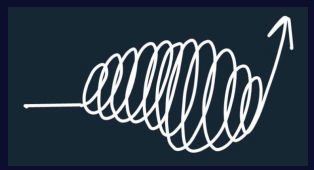
Screening for 4 living systematic reviews. Most were non-RCT based.

Tested out three different **screening algorithms/approaches…**

Supporting living evidence synthesis

Four non-Cochrane living systematic reviews on the topic of depression, anxiety and psychosis

The GALENOS project



3-agreements (3 consecutive agreements needed; disagreements go to 1 resolver

4-agreements (4 consecutive agreements needed; disagreements go to 1 resolver

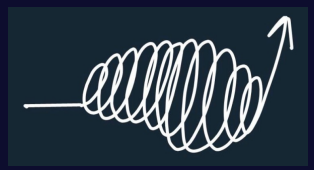
3-agreements (3 consecutive agreements needed; disagreements go to 2 resolver

4-agreements (4 consecutive agreements needed; disagreements go to 1 resolver = one member of author team

Supporting living evidence synthesis

Four non-Cochrane living systematic reviews on the topic of depression, anxiety and psychosis

The GALENOS project



3-agreements (3 consecutive agreements needed; disagreements go to 1 resolver)

4-agreements (4 consecutive agreements needed; disagreements go to 1 resolver)

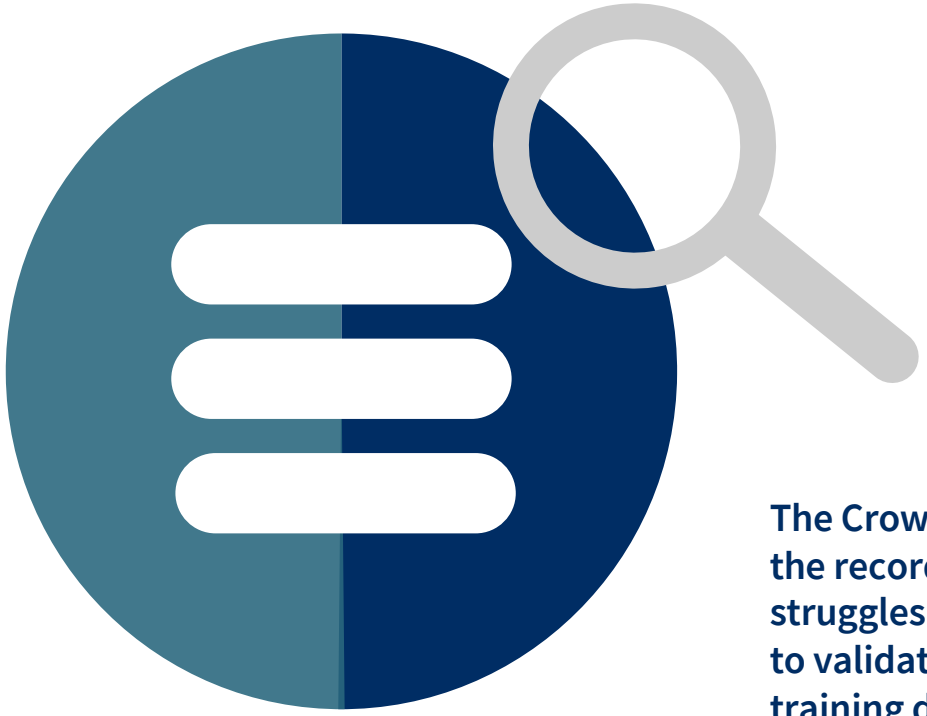
3-agreements (3 consecutive agreements needed; disagreements go to 2 resolver)

4-agreements (4 consecutive agreements needed; disagreements go to 1 resolver = one member of author team

Supporting living evidence synthesis – what next?

Creation of comprehensive repository of climate and health studies

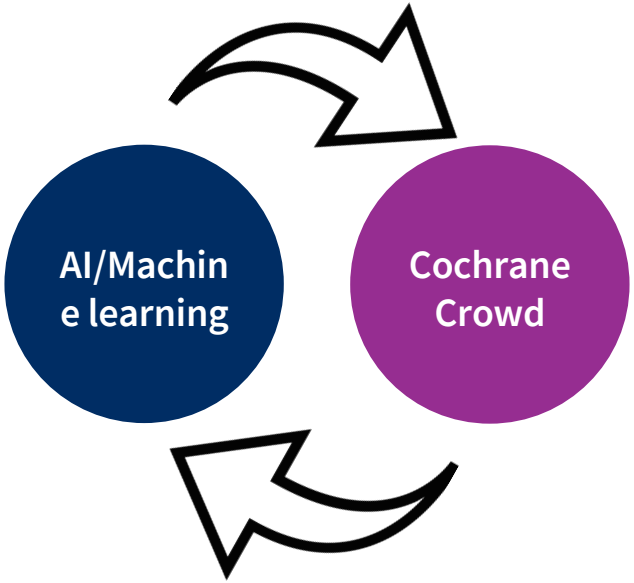
DESTINY
Living meta-repository on climate and health



The Crowd will help to screen the records the machine struggles with as well as help to validate models or create training data for new models

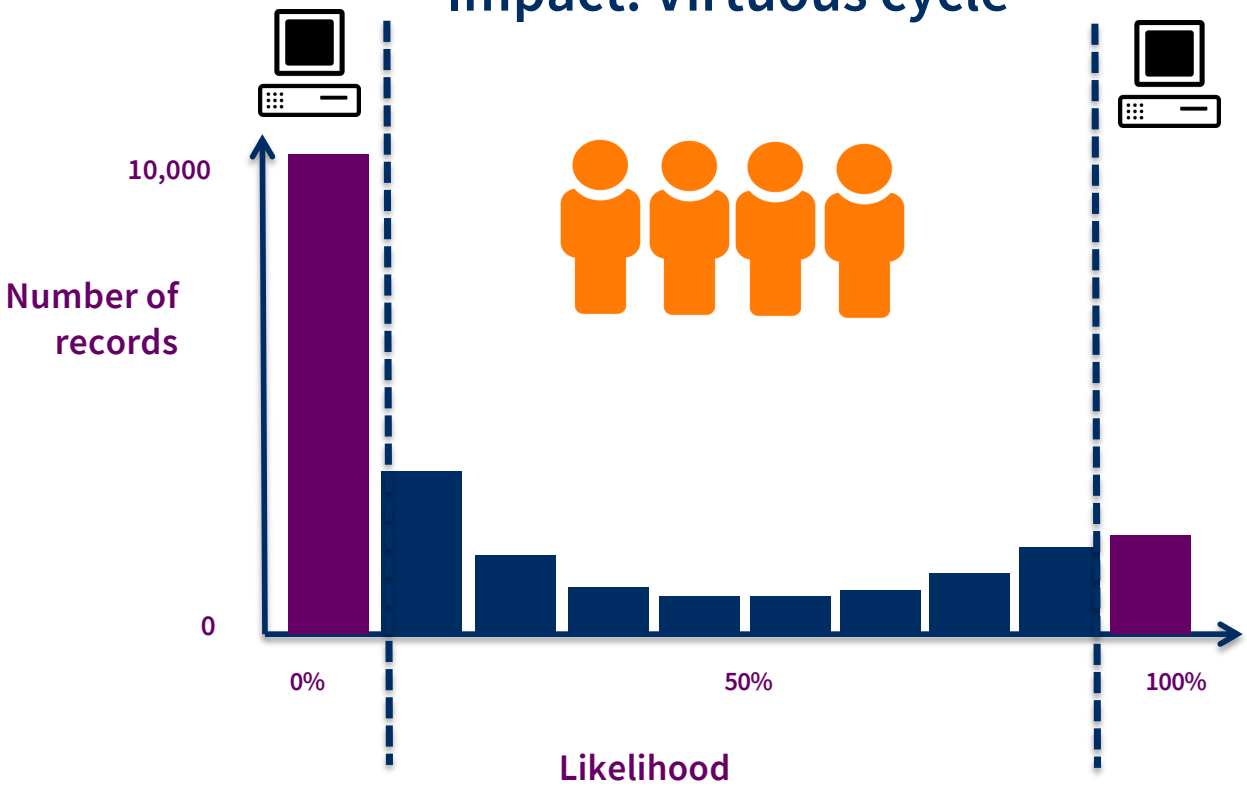
Impact: Virtuous cycle

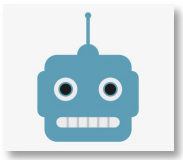
Interaction between human and machine capability



As the Crowd generates more data, it is fed to the machine who continues to learn and is in turn able to perform more and more of the task

Impact: Virtuous cycle



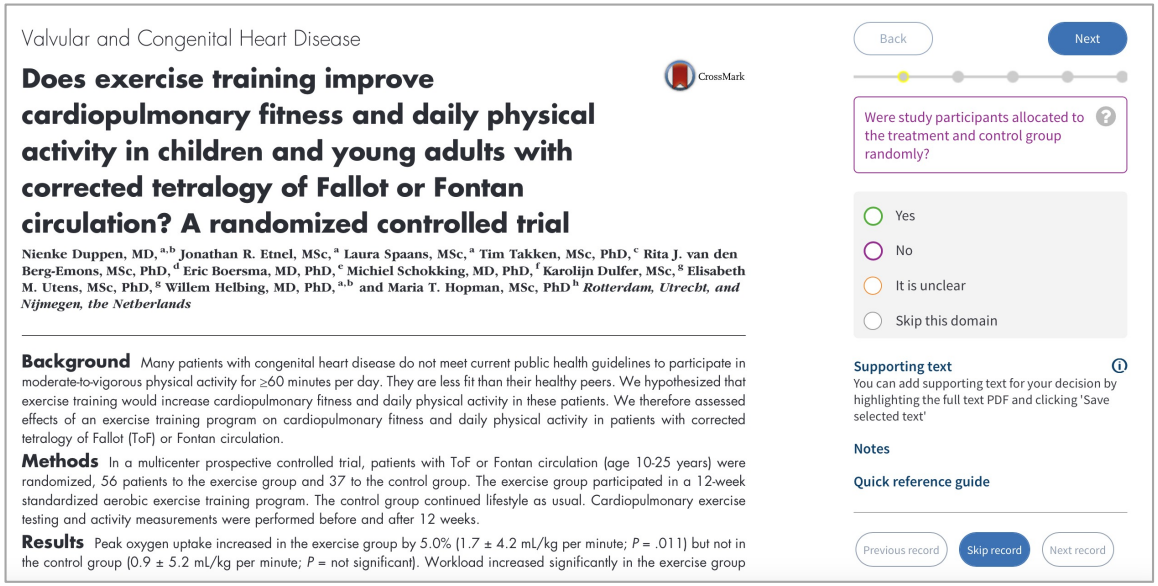


As the classifier improves, the proportion of records that need to go to the Crowd will decrease…

Future

Expand Cochrane Crowd into other parts of the evidence production process

More Crowd tasks



Prototype Risk of Bias task

Cochrane

Thank you!

anoel-storr@Cochrane.org

Trusted evidence.
Informed decisions.
Better health.