



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.

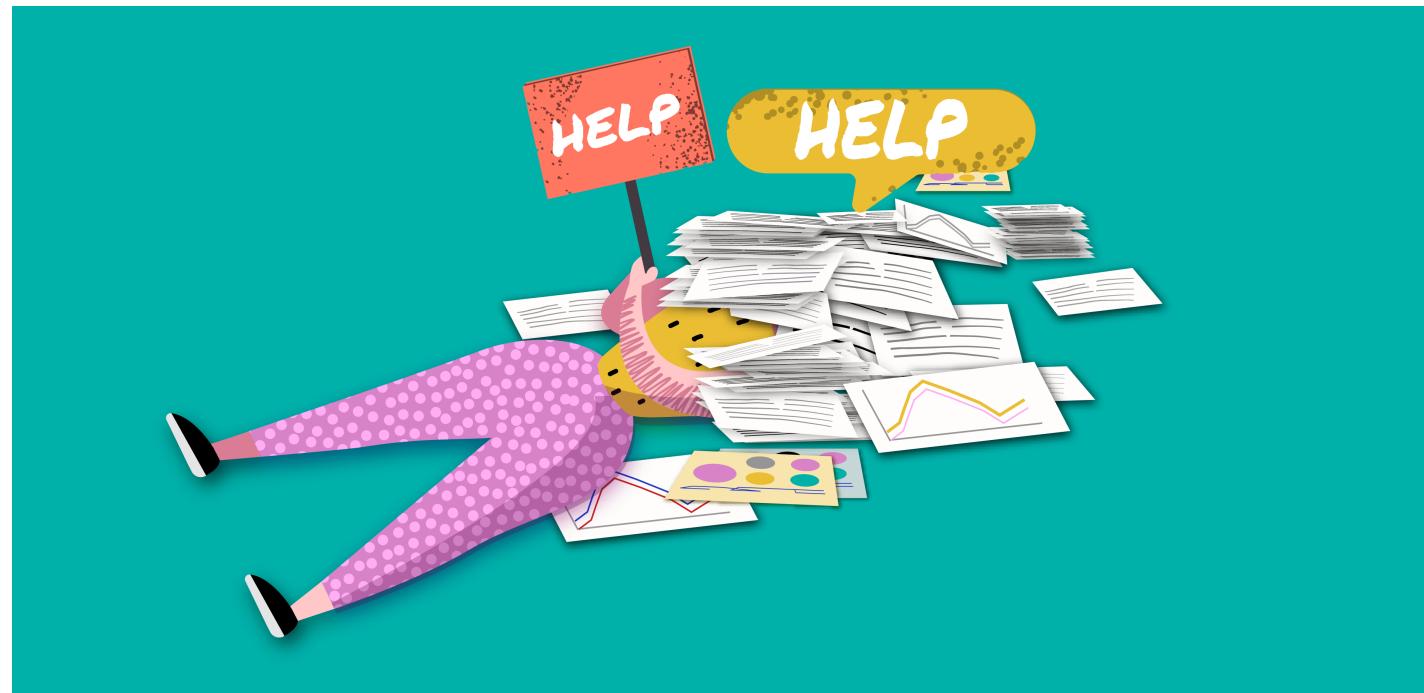


What will I cover?

- **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 it 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”

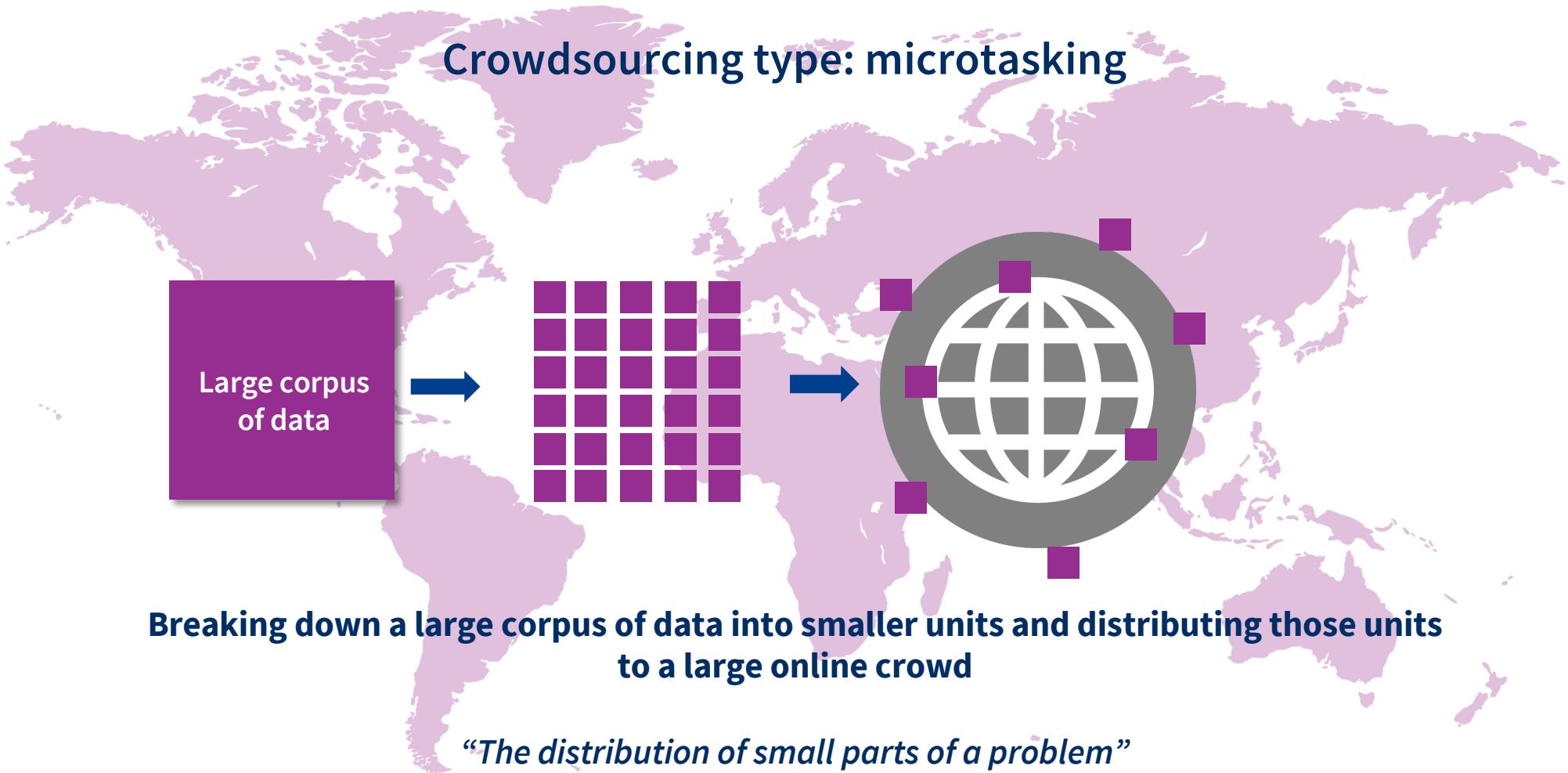


Crowdsourcing

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...





Launched on Valentines Day

The screenshot shows the homepage of the Cochrane Crowd website. At the top, there is a purple header with the Cochrane Crowd logo, the tagline "Trusted evidence. Informed decisions. Better health.", and buttons for "Login", "Signup", and "YouTube". The main content area features a background image of two smiling healthcare professionals (a doctor and a nurse) with a baby. The text "You can make a difference!" is prominently displayed in the center. Below this, a subtext reads: "Become 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." A "What is Cochrane Crowd" button is located in the lower-left quadrant of the main image. At the bottom, there is a summary of the platform's statistics: "39,088 Contributors", "204 Countries", and "11,319,242 Classifications".

Trusted evidence.
Informed decisions.
Better health.

Login Signup YouTube

You can make a difference!

Become 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.

What is Cochrane Crowd

39,088 Contributors

204 Countries

11,319,242 Classifications

<https://crowd.cochrane.org>

Robotic complete mesocolic excision with central vascular ligation for right colonic tumours - A propensity score-matching study comparing with standard laparoscopy

10.1093/bjsopen/zrab016

Background: Laparoscopic complete mesocolic excision (CME) of the right colon with central vascular ligation (CVL) is a technically demanding procedure. This study **retrospectively** evaluated the feasibility, safety and oncological outcomes of the procedure when performed using the da VinciVR robotic system. **Methods:** A prospective case series was collected over 3 years for patients with right colonic cancers treated by standardized robotic CME with CVL using the superior mesenteric vessels first approach. The CME group was compared to a 2 : 1 propensity score-matched non-CME group who had conventional laparoscopic right colectomy with D2 nodal dissection. Primary outcomes were total lymph node harvest and length of specimen. Secondary outcomes were operative time, postoperative complications, and disease-free and overall survival. **Results:** The study included 120 patients (40 in the CME group and 80 in the non-CME group). Lymph node yield was higher (29 versus 18, $P=0.006$), the specimen length longer (322 versus 260 mm, $P=0.001$) and median operative time was significantly longer (180 versus 130 min, $P<0.001$) with robotic CME versus laparoscopy, respectively. Duration of hospital stay was longer with robotic CME, although not significantly (median 6 versus 5 days, $P=0.088$). There were no significant differences in R0 resection rate, complications, readmission rates and local recurrence. A trend in survival benefit with robotic CME for disease-free ($P=0.0581$) and overall survival ($P=0.0454$) at 3 years was documented. **Conclusion:** Robotic CME with CVL is feasible and, although currently associated with a longer operation time, it provides good specimen quality, higher lymph node yield and acceptable morbidity, with a disease-free survival advantage.

Back Next

RCT/qRCT

Reject

Unsure

Move on with a single click

[Help me decide](#)

[Add a note](#)

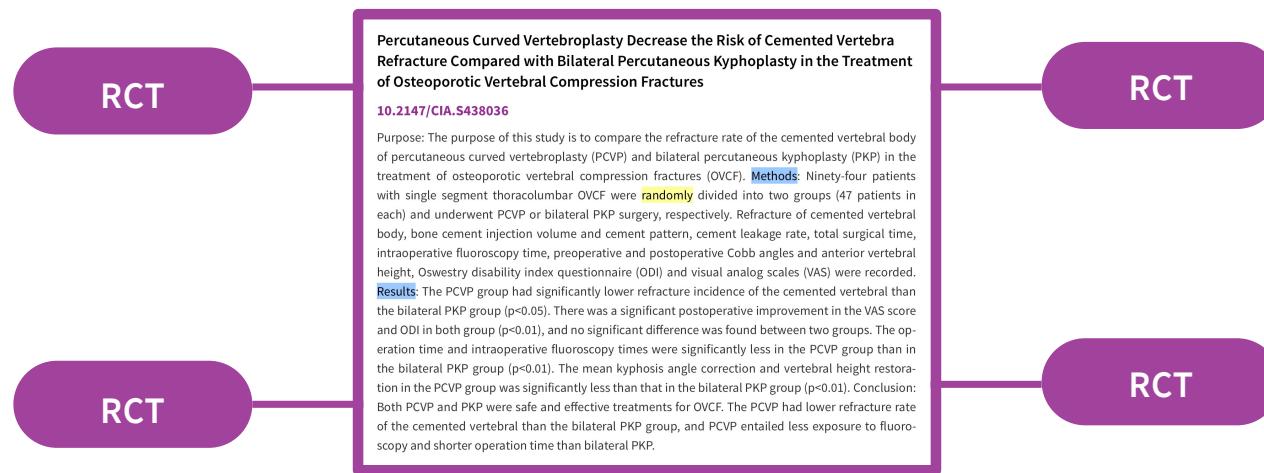
[Quick reference guide](#)

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

Randomized clinical trial to evaluate a routine full anticoagulation Strategy in Patients with Coronavirus Infection (SARS-CoV2) admitted to hospital: rationale and design of the ACTION (AntiCoagulaTlon cOroNavirus)-Coalition IV trial

10.1016/j.ahj.2021.04.005

Background: Observational studies have suggested a higher risk of thrombotic events in patients with coronavirus disease 2019 (**COVID-19**). Moreover, elevated D-dimer levels have been identified as an important prognostic marker in **COVID-19** directly associated with disease severity and progression. Prophylactic anticoagulation for hospitalized **COVID-19** patients might not be enough to prevent thrombotic events; therefore, therapeutic anticoagulation regimens deserve clinical investigation. Design: ACTION is an academic-led, pragmatic, multicenter, open-label, randomized, phase IV clinical trial that aims to enroll around 600 patients at 40 sites participating in the Coalition **COVID-19** Brazil initiative. Eligible patients with a confirmed diagnosis of **COVID-19** with symptoms up to 14 days and elevated D-dimer levels will be randomized to a strategy of full-dose anticoagulation for 30 days with rivaroxaban 20 mg once daily (or full-dose heparin if oral administration is not feasible) vs standard of care with any approved venous thromboembolism prophylaxis regimen during hospitalization. A confirmation of **COVID-19** was mandatory for study entry, based on specific tests used in clinical practice (RT-PCR, antigen test, IgM test) collected before randomization, regardless of in the outpatient setting or not. Randomization will be stratified by clinical stability at presentation. The primary outcome is a hierarchical analysis of mortality, length of hospital stay, or duration of oxygen therapy at the end of 30 days. Secondary outcomes include the World

What is the health condition of the participants?

Enter healthcare condition

COVID-19

- Not reported
- No available term
- Not applicable
- I don't know

Notes

Quick reference guide

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**



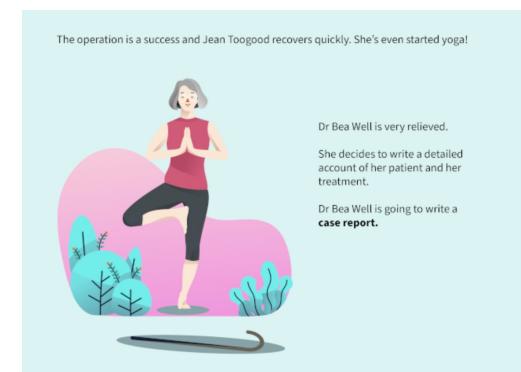
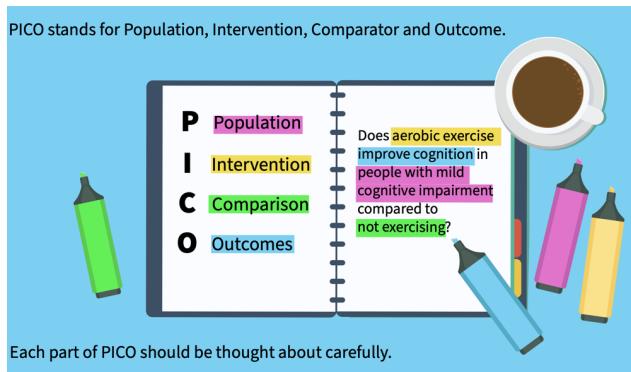
You have Earned **22 badges**



You have participated in **148 challenges**

Supporting beginners

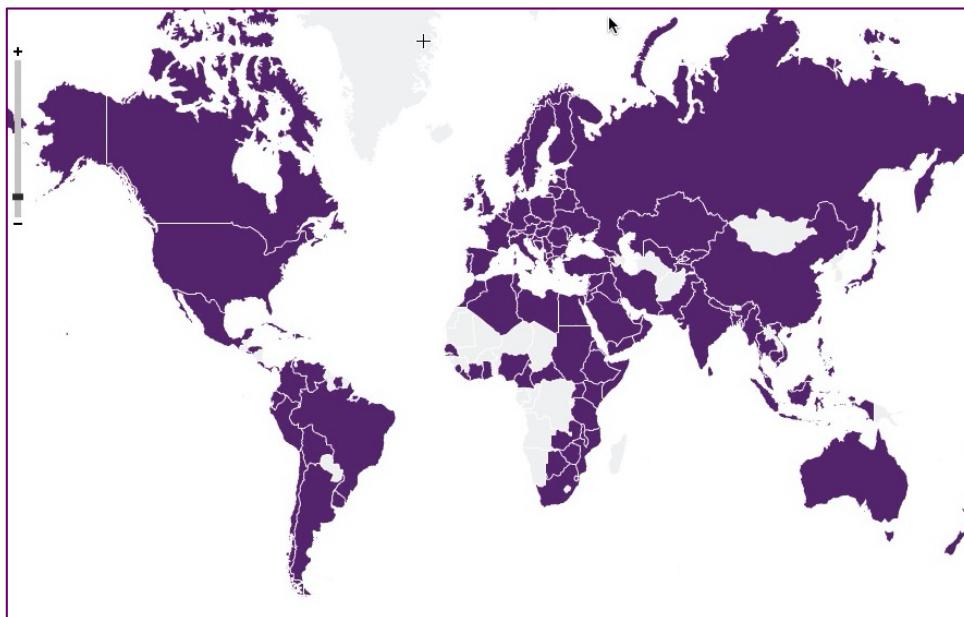
The microtasks are supported by microlearning



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

Crowd characteristics

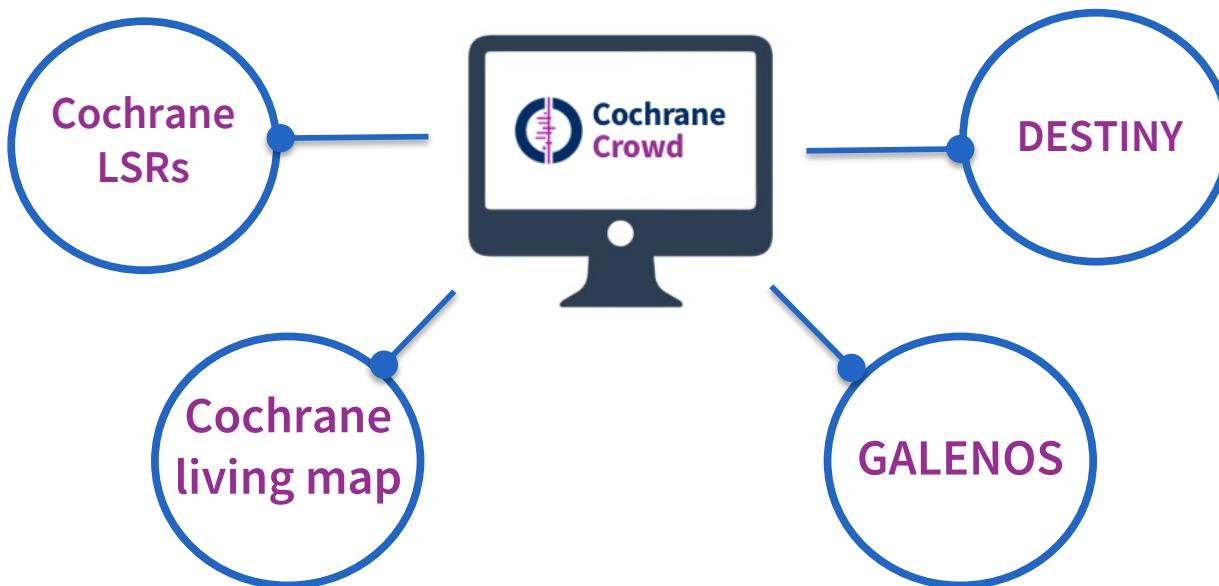
Over 39,000 people have joined Cochrane Crowd



- 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

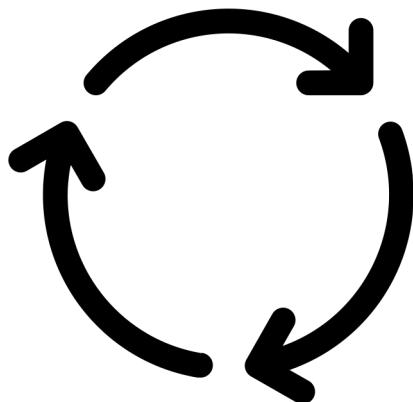
The Crowd have been involved in various living evidence syntheses



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

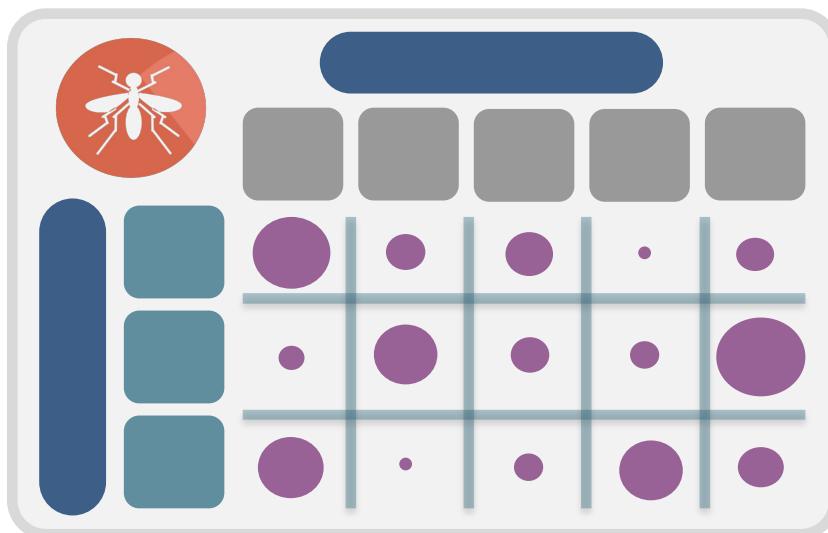
Over 200 reviews have used Screen4Me. Mean workload reduction in results screening: 72%

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

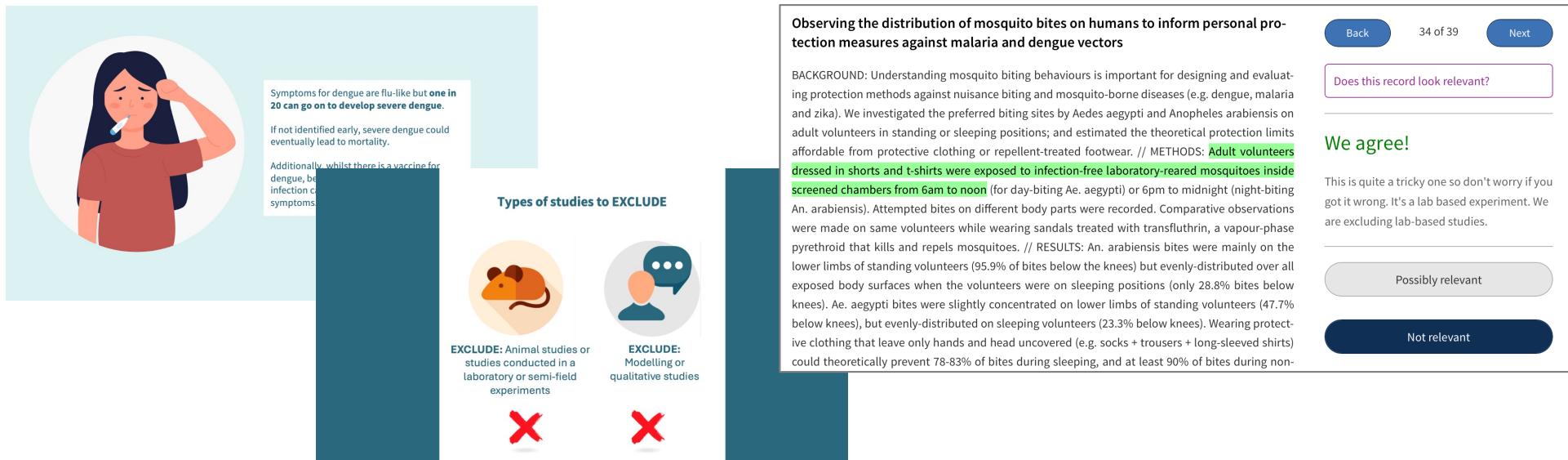


Not just supporting
Cochrane reviews but
also a Cochrane living
evidence and gap map:
dengue prevention.

Created a bespoke
Crowd task to provide
initial eligibility screen.

Supporting living evidence synthesis

A living evidence and gap map on dengue prevention



Symptoms for dengue are flu-like but **one in 20 can go on to develop severe dengue.**
If not identified early, severe dengue could eventually lead to mortality.

Additionally, whilst there is a vaccine for dengue, because of the long time between infection and symptoms, it is not effective.

Types of studies to EXCLUDE

- EXCLUDE:** Animal studies or studies conducted in a laboratory or semi-field experiments
- EXCLUDE:** Modelling or qualitative studies

Observing the distribution of mosquito bites on humans to inform personal protection measures against malaria and dengue vectors

BACKGROUND: Understanding mosquito biting behaviours is important for designing and evaluating protection methods against nuisance biting and mosquito-borne diseases (e.g. dengue, malaria and zika). We investigated the preferred biting sites by Aedes aegypti and Anopheles arabiensis on adult volunteers in standing or sleeping positions; and estimated the theoretical protection limits affordable from protective clothing or repellent-treated footwear. // **METHODS:** **Adult volunteers dressed in shorts and t-shirts were exposed to infection-free laboratory-reared mosquitoes inside screened chambers from 6am to noon** (for day-biting Ae. aegypti) or 6pm to midnight (night-biting An. arabiensis). Attempted bites on different body parts were recorded. Comparative observations were made on same volunteers while wearing sandals treated with transfluthrin, a vapour-phase pyrethroid that kills and repels mosquitoes. // **RESULTS:** An. arabiensis bites were mainly on the lower limbs of standing volunteers (95.9% of bites below the knees) but evenly-distributed over all exposed body surfaces when the volunteers were on sleeping positions (only 28.8% bites below knees). Ae. aegypti bites were slightly concentrated on lower limbs of standing volunteers (47.7% below knees), but evenly-distributed on sleeping volunteers (23.3% below knees). Wearing protective clothing that leave only hands and head uncovered (e.g. socks + trousers + long-sleeved shirts) could theoretically prevent 78-83% of bites during sleeping, and at least 90% of bites during non-sleeping.

Does this record look relevant?

We agree!

This is quite a tricky one so don't worry if you got it wrong. It's a lab based experiment. We are excluding lab-based studies.

Possibly relevant

Not relevant

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)

The GALENOS project



Supporting living evidence synthesis

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

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

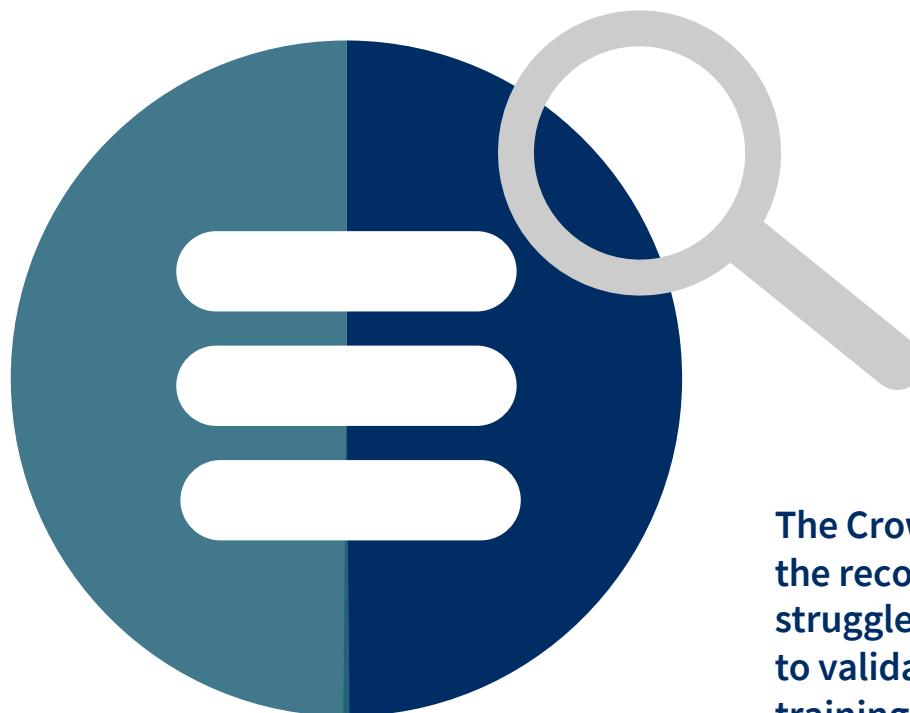
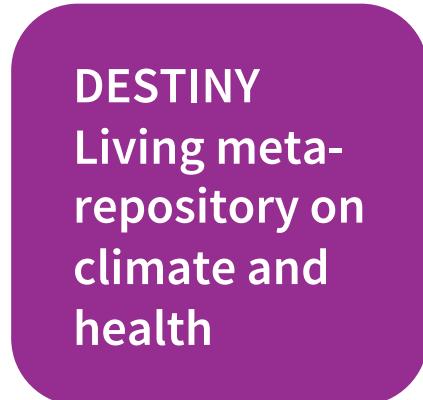
4-agreements (4 consecutive agreements needed; disagreements go to 1 resolver)

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?

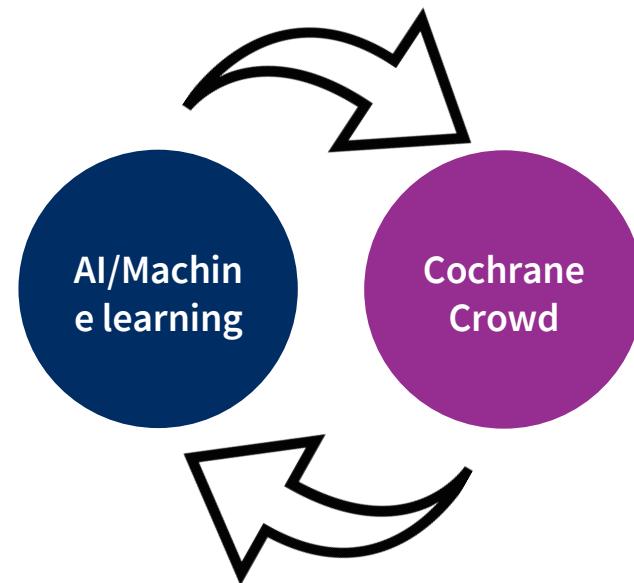
Creation of comprehensive repository of climate and health studies



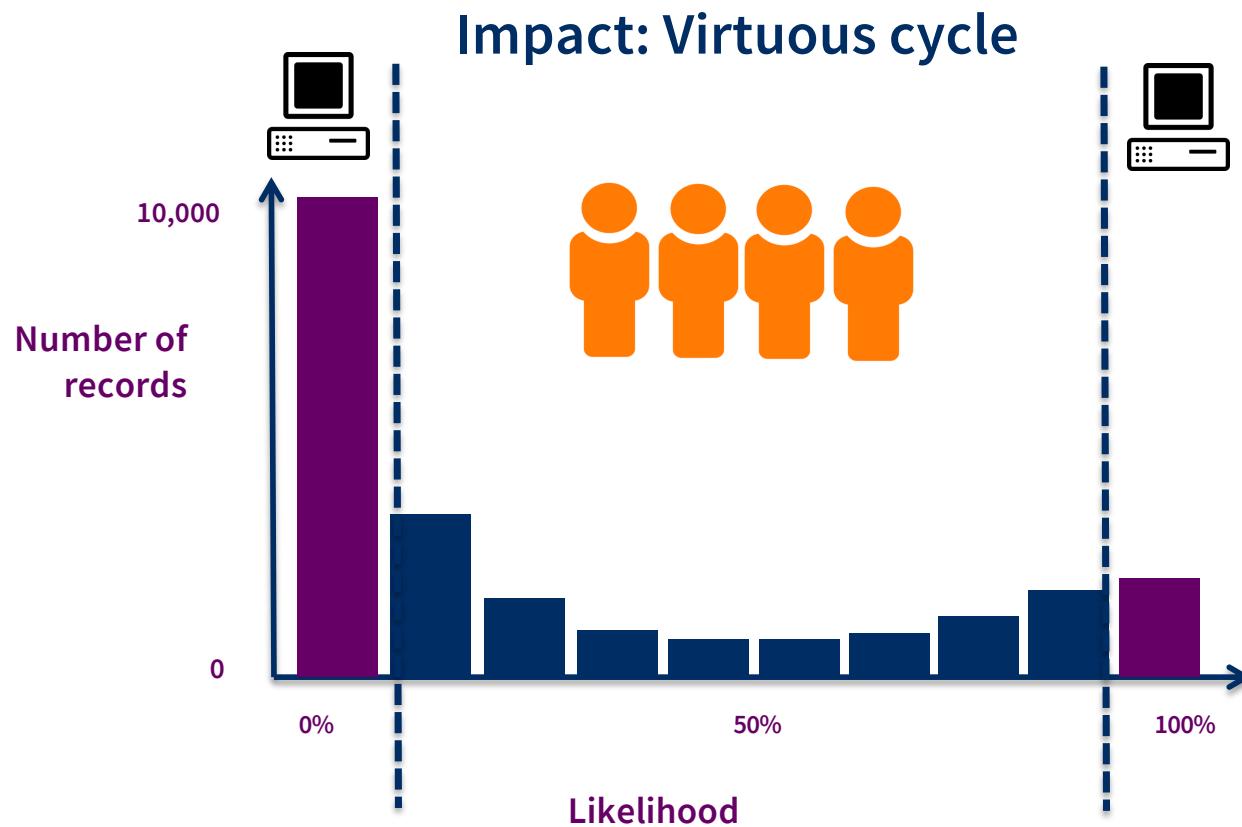
The Crowd will help to screen
the records the machine
struggles with as well as help
to validate models or create
training data for new models

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



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

More Crowd tasks

Future

Expand Cochrane Crowd into other parts of the evidence production process

Valvular and Congenital Heart Disease

Does exercise training improve cardiopulmonary fitness and daily physical activity in children and young adults with corrected tetralogy of Fallot or Fontan circulation? A randomized controlled trial

Nienke Duppen, MD,^{a,b} Jonathan R. Etnel, MSc,^a Laura Spaans, MSc,^a Tim Takken, MSc, PhD,^c Rita J. van den Berg-Emons, MSc, PhD,^d Eric Boersma, MD, PhD,^a Michiel Schokking, MD, PhD,^a Karolijn Dulfer, MSc,^a Elisabeth M. Utens, MSc, PhD,^a Willem Helbing, MD, PhD,^{a,b} and Maria T. Hopman, MSc, PhD^b Rotterdam, Utrecht, and Nijmegen, the Netherlands

Background Many patients with congenital heart disease do not meet current public health guidelines to participate in moderate-to-vigorous physical activity for ≥60 minutes per day. They are less fit than their healthy peers. We hypothesized that exercise training would increase cardiopulmonary fitness and daily physical activity in these patients. We therefore assessed effects of an exercise training program on cardiopulmonary fitness and daily physical activity in patients with corrected tetralogy of Fallot (ToF) or Fontan circulation.

Methods In a multicenter prospective controlled trial, patients with ToF or Fontan circulation (age 10-25 years) were randomized, 56 patients to the exercise group and 37 to the control group. The exercise group participated in a 12-week standardized aerobic exercise training program. The control group continued lifestyle as usual. Cardiopulmonary exercise testing and activity measurements were performed before and after 12 weeks.

Results Peak oxygen uptake increased in the exercise group by 5.0% ($1.7 \pm 4.2 \text{ mL/kg per minute}$; $P = .011$) but not in the control group ($0.9 \pm 5.2 \text{ mL/kg per minute}$; $P = \text{not significant}$). Workload increased significantly in the exercise group

 CrossMark

Were study participants allocated to the treatment and control group randomly?

Yes

No

It is unclear

Skip this domain

Supporting text  You can add supporting text for your decision by highlighting the full text PDF and clicking 'Save selected text'

Notes

Quick reference guide

Previous record  Next record

Prototype Risk of Bias task



Thank you!

anoel-storr@Cochrane.org

Trusted evidence.
Informed decisions.
Better health.

