Institute of Education

UCL

Human and artificial intelligence: new technologies and processes to find studies for systematic reviews

EAHIL 2019

James Thomas[1], Anna Noel-Storr[2] and Claire Stansfield[1]

[1]EPPI-Centre, University College London
[2]Radcliffe Department of Medicine, University of Oxford

Acknowledgements & declaration of interest

- Many people… including: Sergio Graziosi, Jeff Brunton, Alison O'Mara-Eves, Ian Shemilt, Claire Stansfield (EPPI-Centre / EPPI-Reviewer and text mining / automation / information science); Chris Mavergames and Cochrane IKMD team; Julian Elliott and others on Cochrane Transform project; Iain Marshall (Kings College); Byron Wallace (Northeastern University); the Digital Services Team at the National Institute for Health & Care Excellence (NICE); Cochrane Crowd

- I am employed by University College London; receive funding from Cochrane and the funders below for this and related work; co-lead of Project Transform; lead EPPI-Reviewer software development.

- Parts of this work funded by: Cochrane, JISC, Medical Research Council (UK), National Health & Medical Research Council (Australia), Wellcome Trust, Bill & Melinda Gates Foundation, Robert Wood Johnson Foundation. All views expressed are my own, and not necessarily those of these funders.

- ('Creative commons' photos used for illustrations)

Aims and objectives

- AIM: outline the potential for using AI/ machine learning to make systematic reviewing HTAs more efficient
- OBJECTIVES:
  - How some of these technologies – especially machine learning - works
  - Demonstrate / discuss some current tools
  - Discuss future directions of travel

Outline

- Introduction and background (James 15 mins)
- Practical sessions:
  - Developing search strategies (Claire 30 mins)
  - Machine learning (James 30 mins)
  - BREAK
  - A bit more machine learning / trying tools out (15 mins)
  - Crowdsourcing + S4M + surveillance (Anna 30 mins)
- Discussion / practical (All 30 mins)

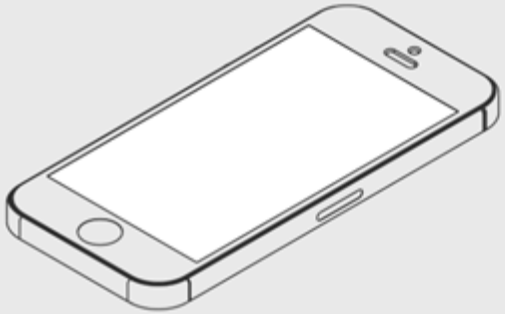
- http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3677

Discussion

https://www.mentimeter.com/app

Go to **www.menti.com** and use the code 24 45 92

What methods and processes will need to be developed to use these tools?



1 Grab your phone

2 Go to www.menti.com

3 Enter the code 80 60 84 and vote!

Context: systematic reviews and HTAs

- Demanding context
- Need to be correct
- Need to be seen to be correct
- Demand very high recall (over precision)
- At odds with much information retrieval work

Automation in systematic reviews HTAs – what can be done?

- Study identification:
  - Citation screening
  - RCT classifier
- Mapping research activity
- Data extraction
  - Risk of Bias assessment
  - Other study characteristics
  - Extraction of statistical data
- (Synthesis and conclusions)

Looking to the future…

The future is already here — it's just not very evenly distributed. William Gibson

- A lot of behind the scenes development has been taking place
- Niche tools and technologies will become more mainstream
- The future will continue to be unevenly distributed

Assisting search development

Purpose: to explore linkages or words in text or controlled vocabulary

Applications:

- Increase precision
- Increase sensitivity
- Aid translation across databases
- “Objective” search strategies
- Integrated search and screen systems

Sample of citations

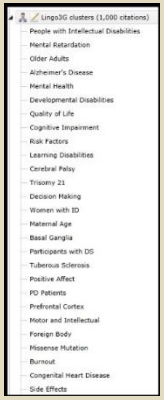
```
DP  - 2013
TI  - The exposure assessment in curren
      acceptability of real-time data c
      users.
PG  - 594
LID - 10
AB  - Ob
      fea
      of
      of
      wer
      dru
```

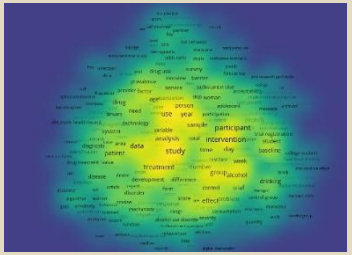
```
IP  - 2
DP  - 2015 Apr
TI  - Comparison of a smartphone app for alco
      intervention plus bibliotherapy: A pilo
PG  - 3
LID - 1
AB  - OB
      sm
      cu
      in
      mo
```

```
IP  - 2
DP  - 2014
TI  - Results of a pilot test of a
      for alcohol use disorders: u
PG  - 168-75
LID - 10.1080/08897077.2013.821437
AB  - BACKGROUND: This paper provi
      early-stage effectiveness an
      that provides a stand-alone,
```

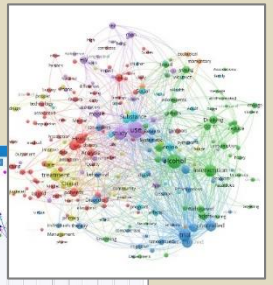
Citation elements (title, abstract, controlled vocabulary, body of text, etc)

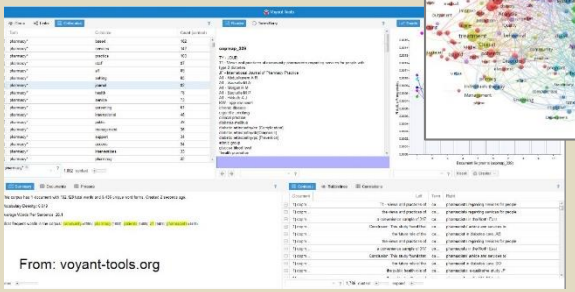
Output





Term	Score
ID	76.61
people	44.67
resident	39.81
adult	38.26
woman	37.08
older people	35.73
menopause	32.58
participant	29.33
client	28.07
life	27.35
older adult	26.21
death	25.76
older person	25.76
age	23.62
active ageing	19.76
retirement	19.55
person	18.72
population	18.14
end-of-life care	17.95
future planning	17.12
ageing adult	15.8





1

Sample of citations

Citation elements (title, abstract, controlled vocabulary, body of text, etc)

2

Analysis

Text analysis

Counting words, phrases or nearby terms

- Generic tools
- Database specific (PubMed) tools

Term extraction and automatic clustering

- TD*IDF

Statistical relations between terms

- Termine
- Automatic clustering

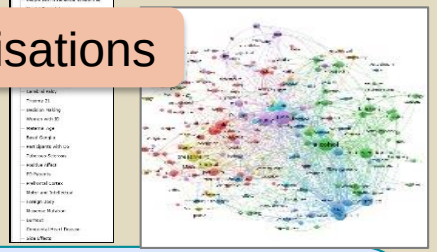
Statistical and linguistic analysis

3

Output

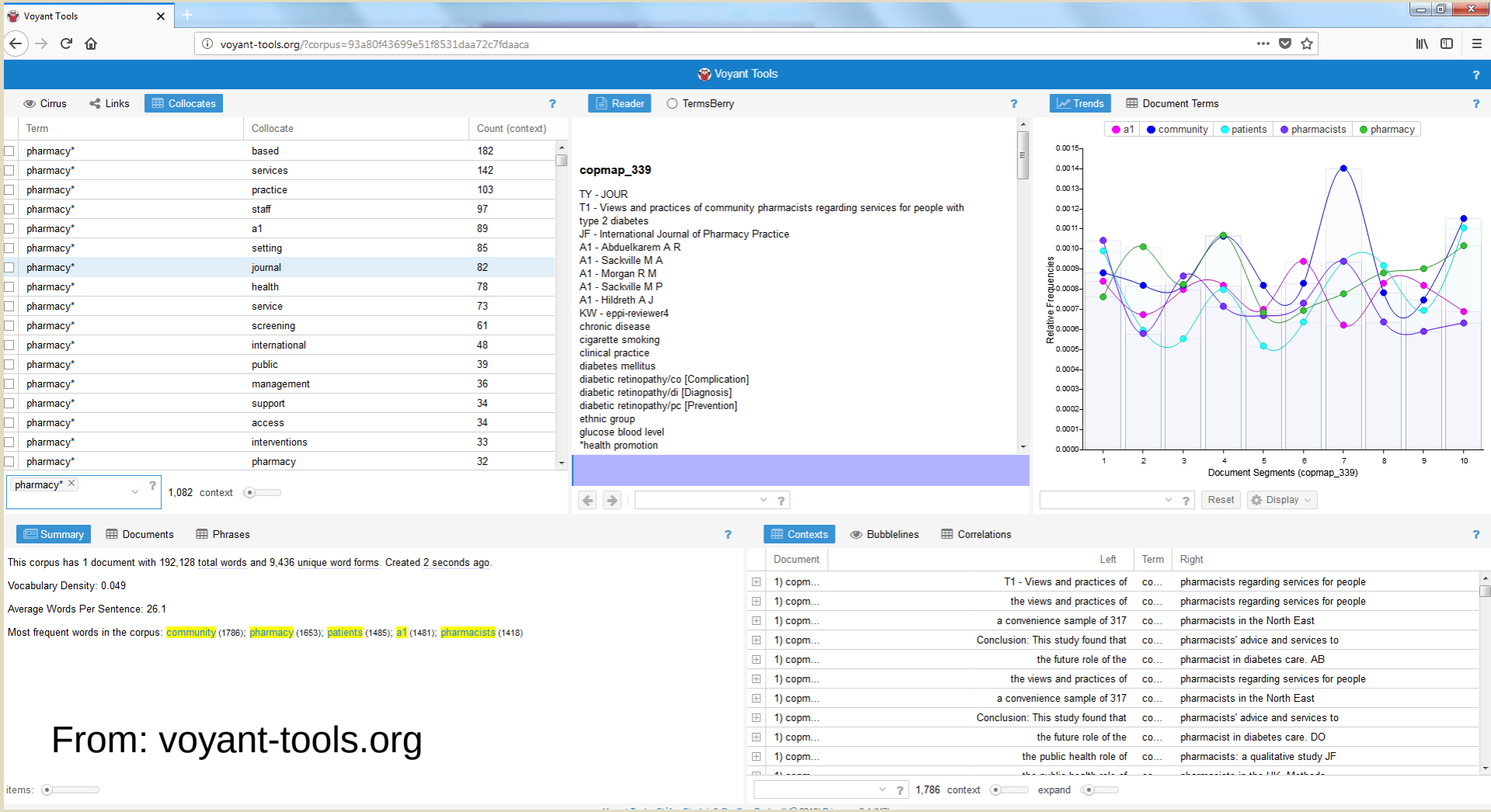
Word or phrase lists

Visualisations



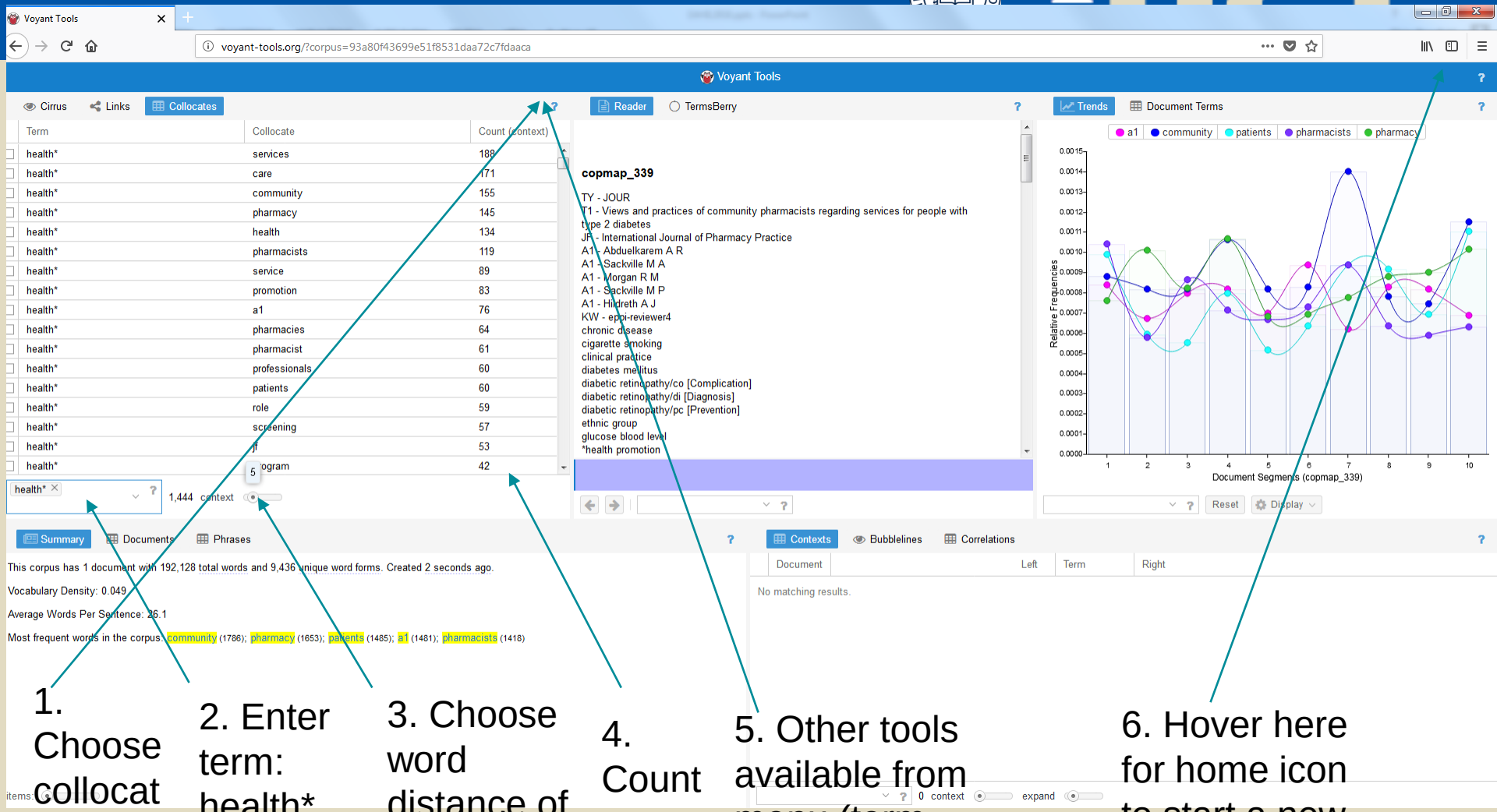
Humans assess relevance and impact to search.

Institute of Education



From: voyant-tools.org

Institute of Education



1. Choose collocates tool

2. Enter term: health*

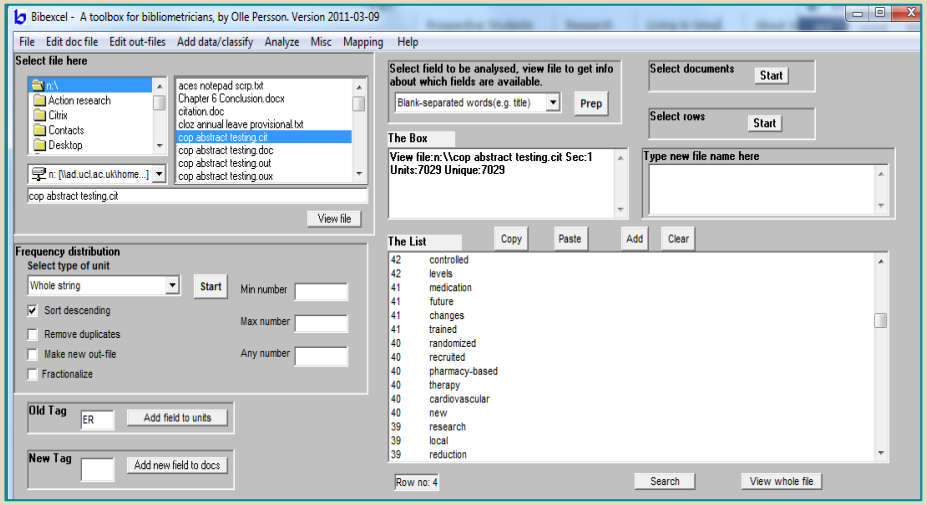
3. Choose word distance of collocates

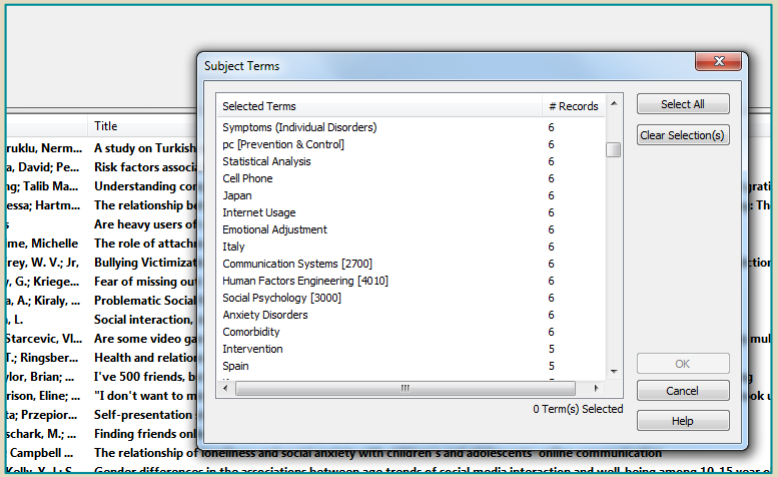
4. Count

5. Other tools available from menu (term grid, Cirrus word clouds etc.

6. Hover here for home icon to start a new analysis

Other tools that have useful functionality include for text analysis…





Using Bibexcel to count the number of abstracts a word occurs in

Using Endnote's Subject Bibliography to generate a list of keywords

Word-frequency analysis results

Quick-search...

Word	Unique	Title	Abstract	Keywords	Points
pharmacy	85	23	68	7	
care	88	36	67	59	
study	63	6	60	1	
from	60	8	56	0	
patients	56	13			
health	97	17			
pharmacists	60	24			

Systematic Review accelerator – Word-frequency analysis

Displays word/phrase occurrence per abstract

Sign-up required

TerMine (C-value) analysis

Found **8224** terms in 75.18 seconds - all terms (in table) (in text) - threshold: 0 Apply

TY - JOUR. T1 - Views and practices of community pharmacists regarding services for people with type 2 diabetes. JF - International Jou
KW - eppi-reviewer4. chronic disease. cigarette smoking. clinical practice. diabetes mellitus. diabetic retinopathy/co [Complication]. diab
service. home care. human. lifestyle. moslem. *non insulin dependent diabetes mellitus. patient care. patient counseling. patient monitori
N2 - Objective : To describe the views and practices of community pharmacists regarding services for people with type 2 diabetes Metho
The 26-item questionnaire covered the setting of the pharmacy , dispensing medication , and the pharmacist 's role in the primary preven
More than 80 % of respondents reported that they saw patients with diabetes `` very often '' or `` often '' when they collected their prescri
medication and gave information to help them have a better understanding of their disease More than 90 % of the pharmacists believed t
10 percent of the respondents reported that they `` often '' or `` very often '' promoted regular eye examinations Home blood glucose mor
Conclusion : This study found that community pharmacists ' advice and services to people with type 2 diabetes fell short of the standards
profession and with stakeholders about the future role of the community pharmacist in diabetes care. AB - Objective : To describe the vie
questionnaire survey of a convenience sample of 317 community pharmacists in the North East of England The 26-item questionnaire co
detecting undiagnosed diabetes and prevention of complications Key findings : There was a 51 % response rate More than 80 % of respo
quarters reported that they `` never/rarely '' or only `` sometimes '' advised patients what to expect from their medication and gave informa
is important , but the majority were `` never/rarely '' or only `` sometimes '' involved in its promotion Around 10 percent of the respondents
by 50 % of the respondents The majority reported that they checked prescriptions for drug interactions Conclusion : This study found that
the National Service Framework for Diabetes Our findings can be used to promote discussion in the profession and with stakeholders ab
SP - 161. EP - 168. CY -. SN - 0961-7671. U1 - 32847778. U2 - 136708. N1 -. ER -. . TY - RPRT. T1 - Findings of a survey of needle exc
Survey results. Needles for injection. Drug abuse. Drugs of abuse. Hepatitis C. Questionnaires. Data collection. Risk assessment. Nation
- This survey was instigated in response to the 2004 DH 'Hepatitis C Action Plan for England ' It examines the nature and extent of provis
commissioning and planning of needle exchange services , and it assesses the levels and quality of data collection The survey comprise
exchange scheme co-ordinators It uncovered a 'mixed economy'of needle exchange facilities present in the majority of drug action teams
activity and a lack of uniformity between services The level of data on needle exchange throughput and activity was poor , raising concer
- This survey was instigated in response to the 2004 DH 'Hepatitis C Action Plan for England ' It examines the nature and extent of provis
commissioning and planning of needle exchange services , and it assesses the levels and quality of data collection The survey comprise
exchange scheme co-ordinators It uncovered a 'mixed economy'of needle exchange facilities present in the majority of drug action teams
activity and a lack of uniformity between services The level of data on needle exchange throughput and activity was poor , raising concer
-. IS - Research Briefing : 17. CY - UK. UR - http : //www.nta.nhs.uk/publications/docs/RB17_ned_xch.pdf. PB - NHS National Treatment
T1 - Strategies enhancing the public health role of community pharmacists : a qualitative study. JF - Journal of Pharmaceutical Health Se
N2 - Objectives : This study interviewed healthcare professionals to identify strategies enhancing the public health role of community pha
'HD Call Recorder for Skype ' The qualitative data software package NVivo (version 10) was used for the storage , retrieval and analysis
strategies to enhance the public health role of community pharmacists in the UK They included empowerment through education and aw
social media in practice , the use of independent pharmacist practitioners (IPPs) , teaching communication methods to students and pha
and changing the undergraduate pharmacy curriculum to increase its public health content In terms of benefits , enhancing the public hea
between healthcare professionals , enhance the knowledge base of practitioners , reduce negative perceptions about pharmacists and bi

Text view: applying Termine to 338 studies of public health interventions in community pharmacies

From NacTeM http://www.nactem.ac.uk/software/termine/cgi-bin/termine_cvalue.cgi

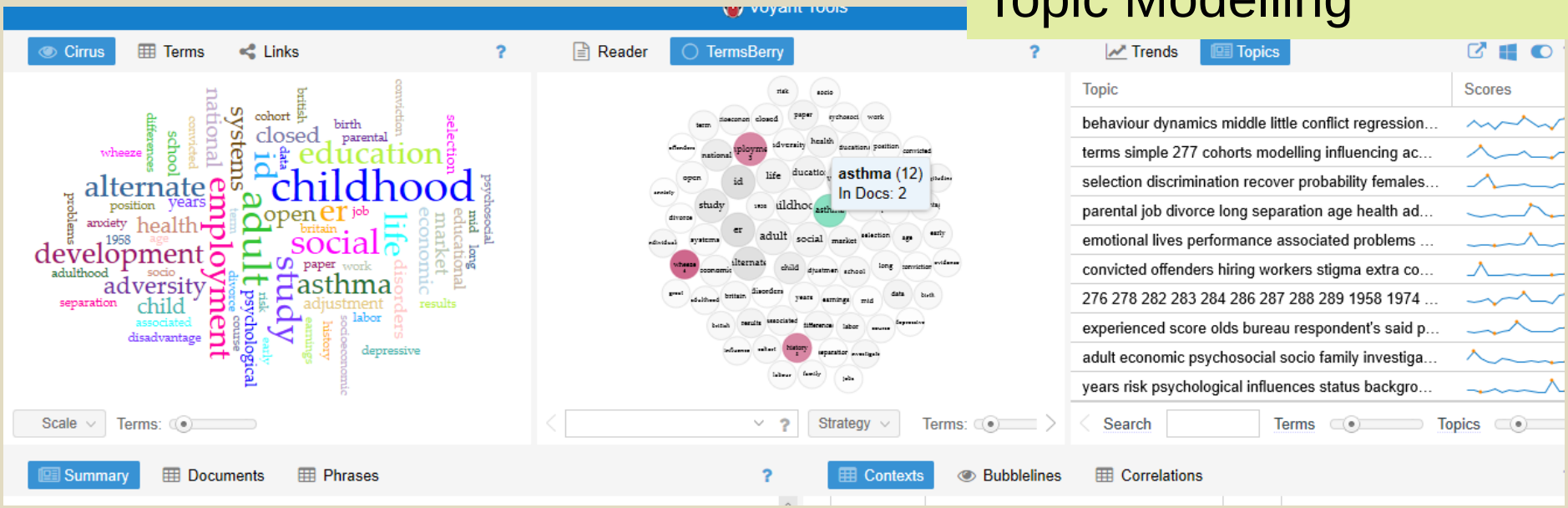
Rank	Term	Score
1	community pharmacy	1033.88501
2	community pharmacist	451.192322
3	public health	232.711411
4	blood pressure	175
5	risk factor	147.822144
6	primary care	138.600006
7	health service	122.838188
8	main outcome	117.029854
9	main outcome measures	113.789383
10	needle exchange	110.720993
11	drug user	100.849159
12	health care	99.918594
13	pharmacy service	96
14	intervention group	89.111115
15	public health service	88.340805
16	cardiovascular disease	82.647057
17	usual care	79.789474
18	health promotion	72.078949
19	control group	71.555557
20	pharmacy practice	71.099998
21	weight management	70.578125
22	body mass index	69.73835
23	cardiovascular risk	66.903847
24	vaccination rate	66.117645
25	international journal	62.5
26	pharmacy staff	62.421051
27	weight loss	61.708332
28	drug therapy	61
29	risk assessment	60.314285
30	hiv testing	57.882355
31	blood glucose	57.468086

Table view: Applying Termine to 338 studies of public health interventions in community pharmacies

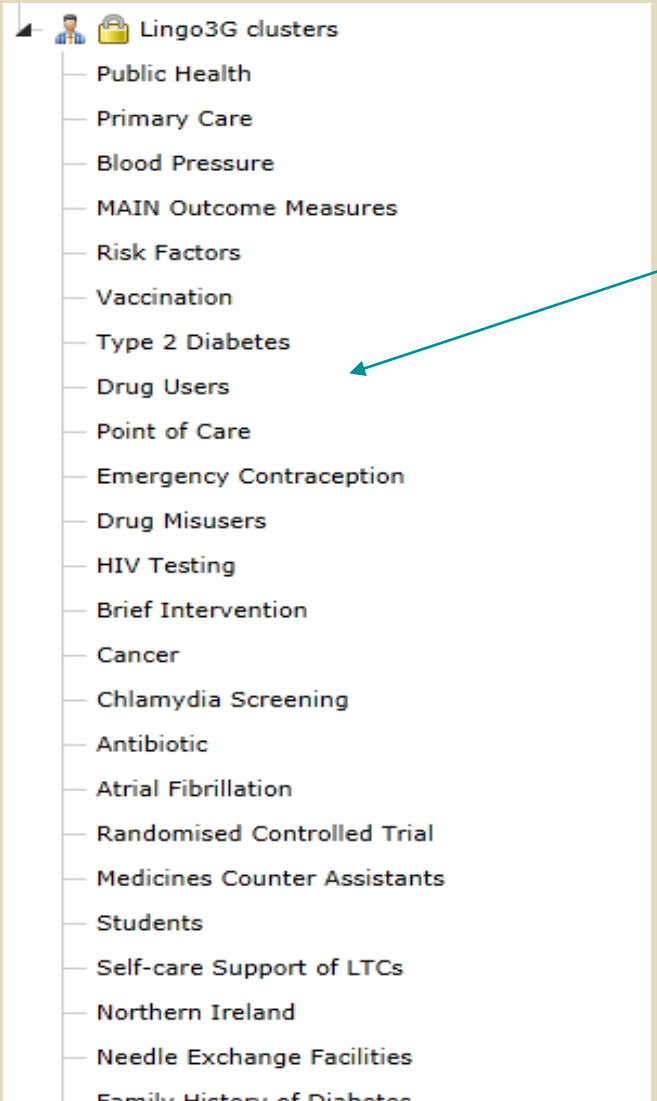
From NacTeM http://www.nactem.ac.uk/software/termine/cgi-bin/termine_cvalue.cgi

Clustering and visualisation

Cirrus word clouds
Terms Berry
Topic Modelling



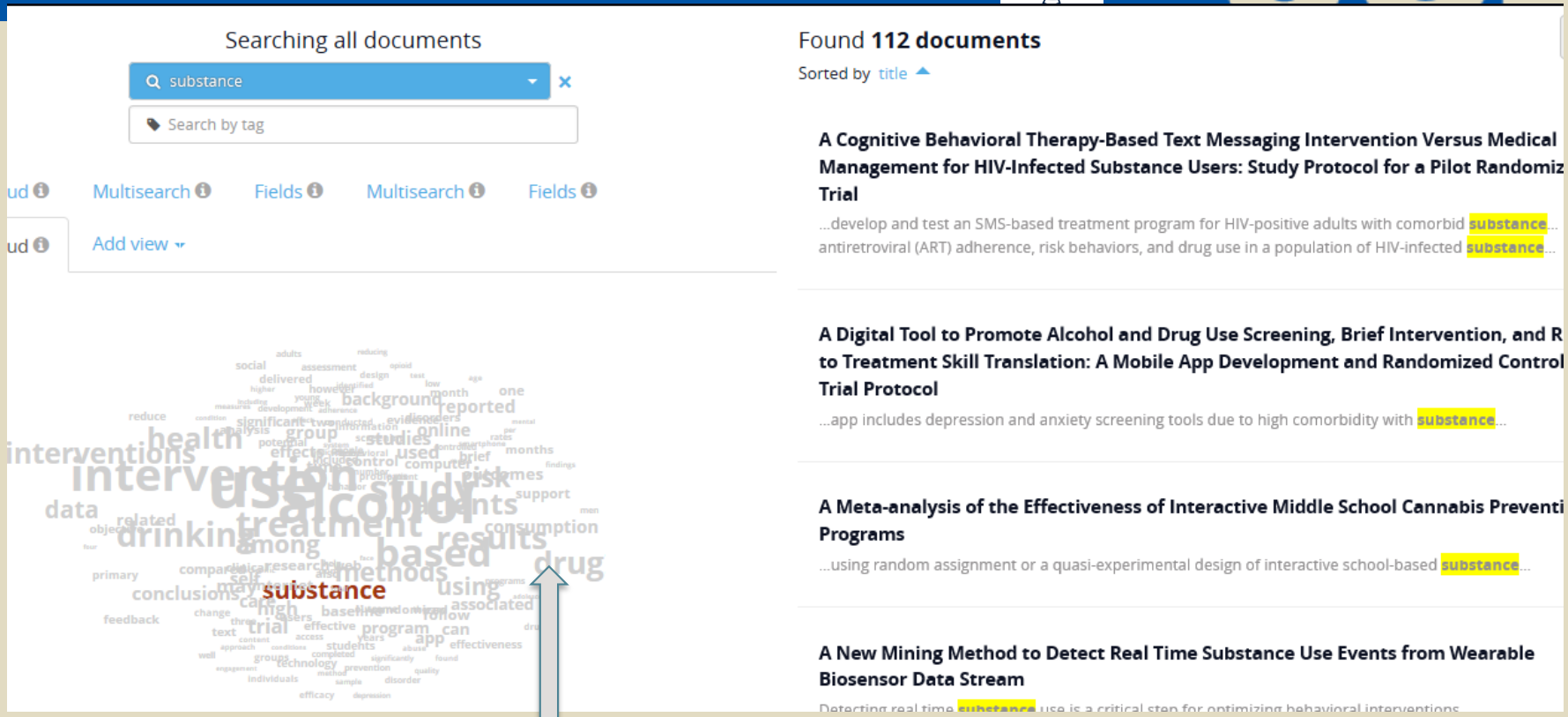
Clustering and visualisation



Lingo3G groups sets of citations and assigns labels

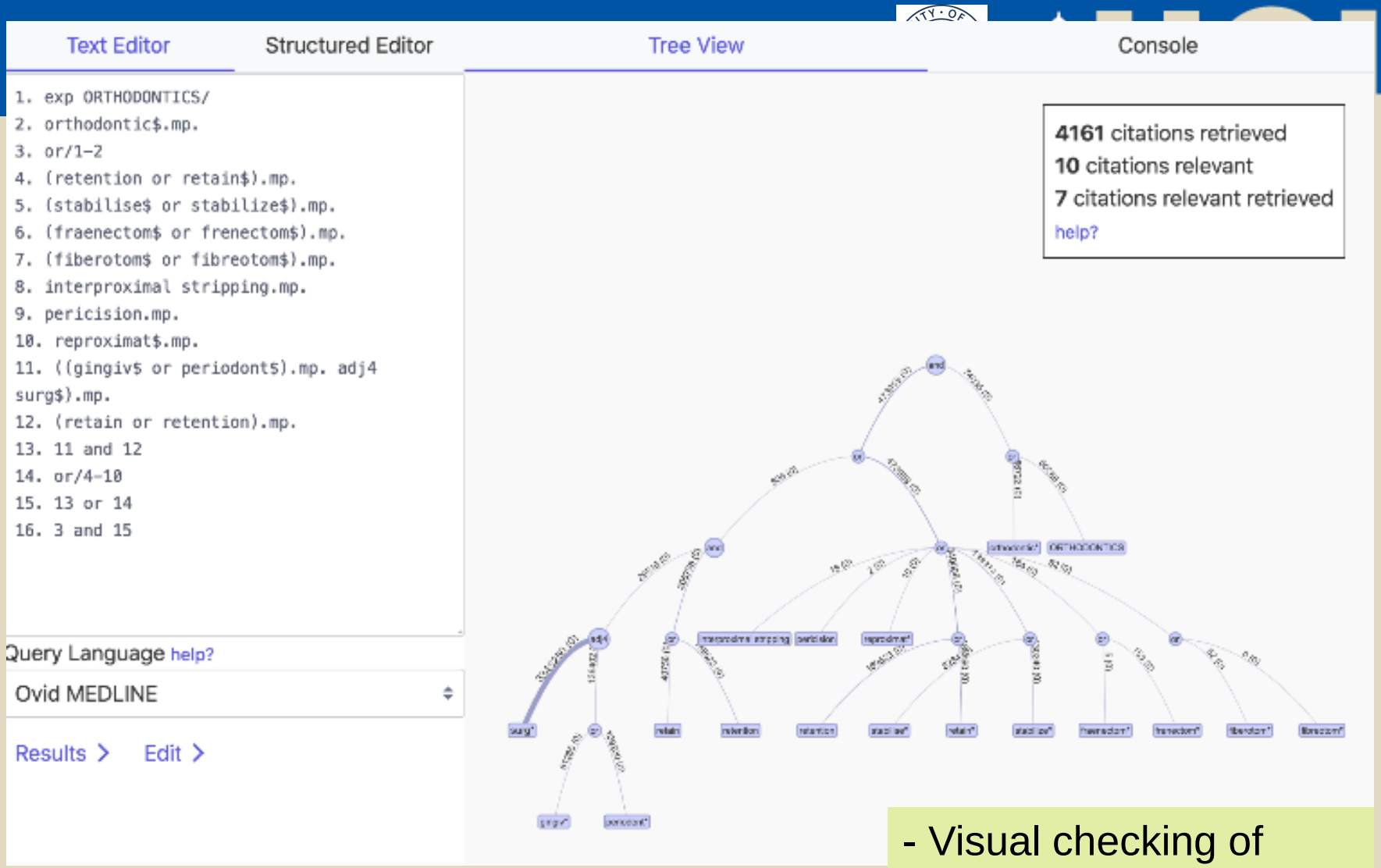
Citations within each cluster can be browsed

Using Lingo3G to map 338 studies of public health interventions in community pharmacies, (Interface: EPPI-Reviewer 4)



Overviewdocs

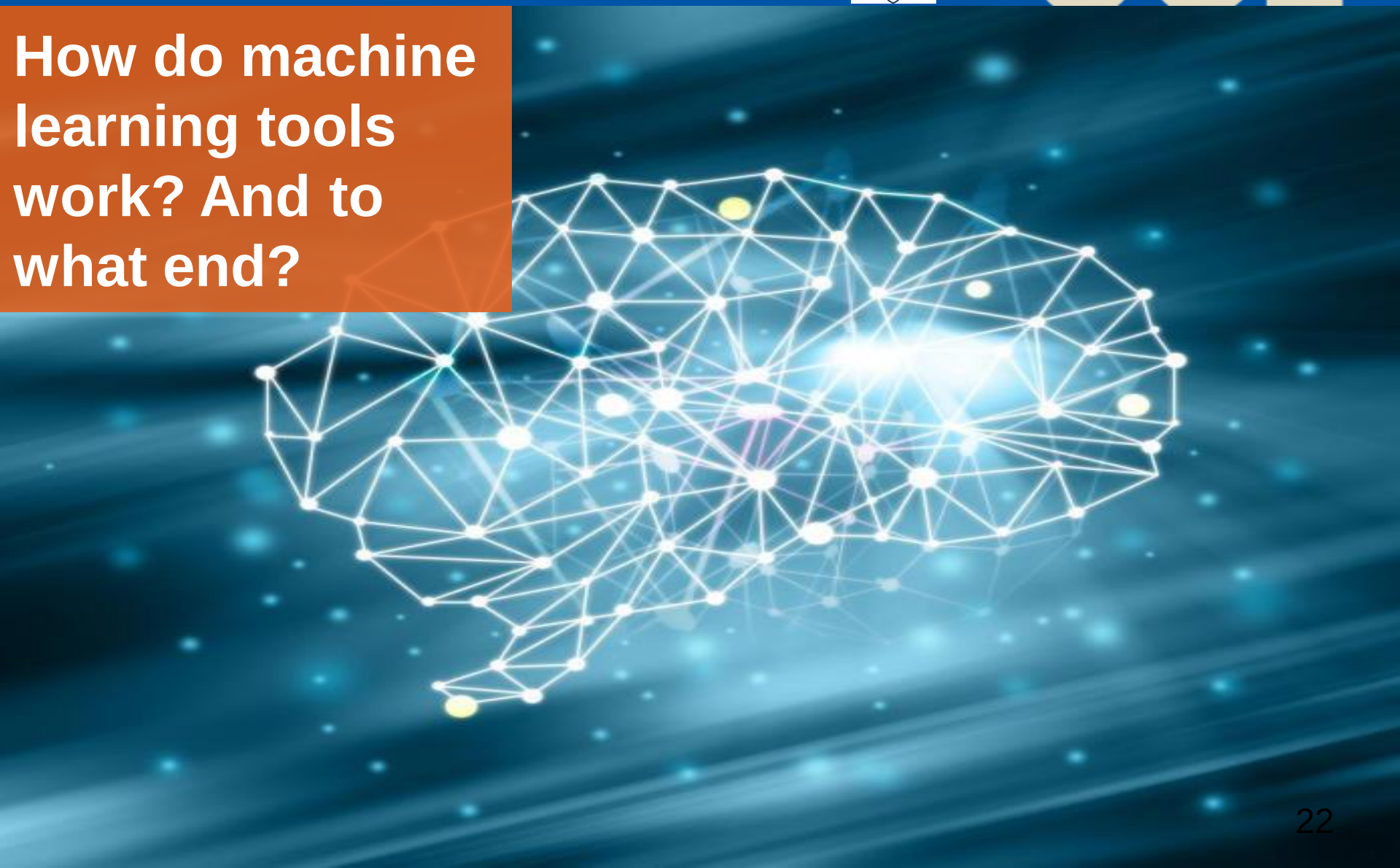
Interactive word cloud linked to browsing citations
Examples – import csv files
Sign-up and storage of files



Search refiner
Sign up required

- Visual checking of search elements,
- Comparison with test citations

How do machine learning tools work? And to what end?



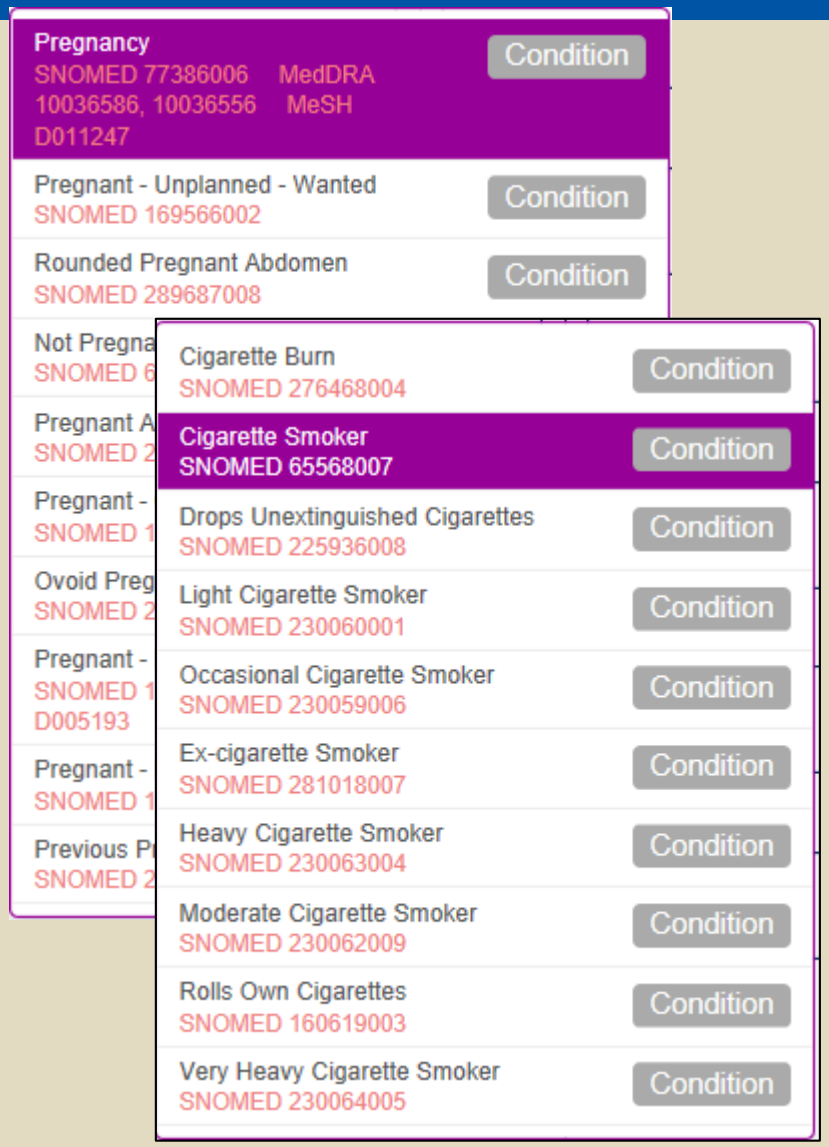
Three machine learning / automation paradigms

- Rules-based approaches
  - (strictly speaking, not machine learning)
- Unsupervised approaches
- Supervised approaches

- Covering in terms of technology not purpose, so we can consider their strengths and weaknesses more easily

Rules-based approaches

- As you might guess… a set of rules is constructed by humans and given to the machine
- For example
  - Look up a simple set of words
  - Use of synonyms
  - If a given phrase is present, apply a given code



PARTICIPANTS:
789 pregnant smokers, aged 16-50 years and at 10-24 weeks' gestation, who smoked at least one cigarette daily and were prepared to quit smoking one week after enrollment

PARTICIPANTS:
Women between 10 and 24 weeks' gestation smoking five or more cigarettes a day before pregnancy and one or more during pregnancy.

PARTICIPANTS:
Women between 12 and 24 weeks' gestation who smoked ≥ 10 cigarettes a day before and ≥ 5 during pregnancy, with an exhaled carbon monoxide (CO) reading of ≥ 8 parts per million (p.p.m.).

Rules can be accurate… but fragile

- If you stick within the rules, you get the anticipated results
- If you stray outside – even a little bit – the rule can fail altogether
- No grey area – it works, or completely fails

Rules are not fashionable!

Unsupervised approaches

- The machine is given no rules…
- And simply identifies patterns in the data
- E.g.
  - Relationships between words
    - https://projector.tensorflow.org/
  - Clustering documents
    - LDAVis
    - Carrot2 Search

Institute of Education

UCL

https://projector.tensorflow.org

Embedding Projector

DATA

5 tensors found

Word2Vec 10K

Label by: word

Color by: No color map

Edit by: word

Tag selection as

Load | Publish | Download | Label

Sphereize data

Checkpoint: Demo datasets

Metadata: oss_data/word2vec_10000_200d_labels.tsv

UMAP | T-SNE | PCA | CUSTOM

X: Component #1

Y: Component #2

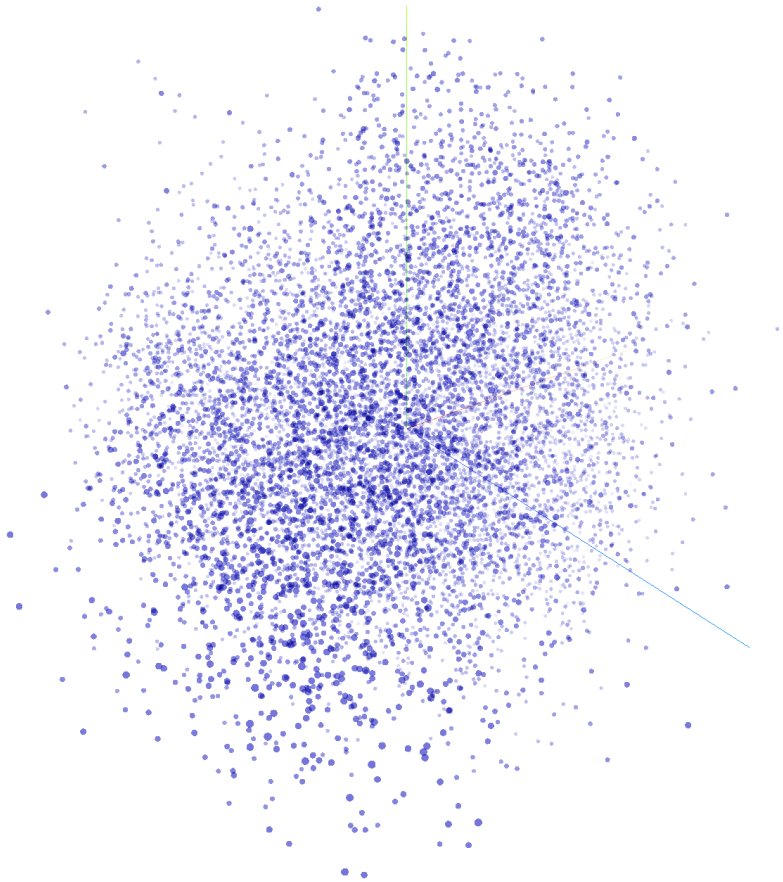
Z: Component #3

PCA is approximate.

Total variance described: 8.5%.

Points: 10000 | Dimension: 200

Reset zoom to fit all points



Show All Data | Isolate 101 points | Clear selection

Search .* by: word

BOOKMARKS (0)

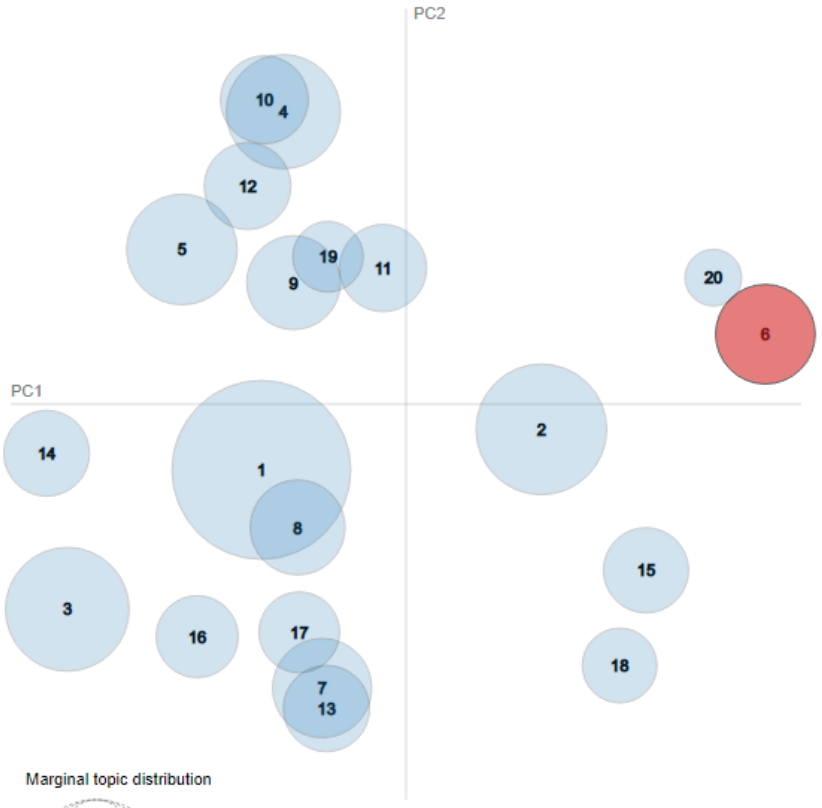
Selected Topic: 6 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric:(2)

λ = 0.63

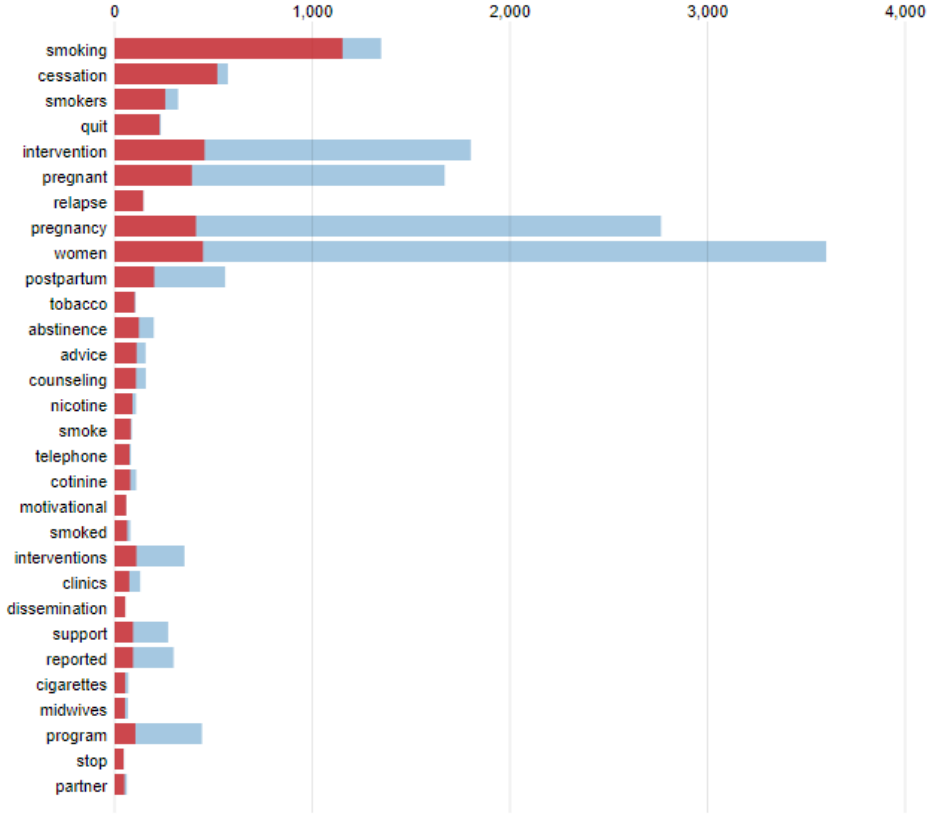
0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Marginal topic distribution

Top-30 Most Relevant Terms for Topic 6 (5% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Institute of Education

smoking - Carrot2 Clustering En

Not secure | search.carrot2.org/stable/search?query=smoking&results=100&source=pubmed&algorithm=lingo&view=circles&skin=fanc...

About | Search feeds | Download | Carrot Search | Contact

eTools Web Search | Wiki | PubMed | PUT

smoking Search More options

Folders | Circles | FoamTree

Top **100** results of about **268158** for **smoking**

1 Prospective, multi-centric benchmark study assessing delirium: prevalence, incidence and its correlates in hospitalized elderly Lebanese patients.

With the increase in the proportion of elderly Lebanese patients, little is known about delirium's prevalence, incidence and correlated factors. ... To identify the prevalence, incidence and factors associated with overall and incident delirium in hospitalized elderly Lebanese patients.

http://www.ncbi.nlm.nih.gov/pubmed/31203529

2 Familial cancer of unknown primary.

Cancer of unknown primary site (CUP) is a deadly disease diagnosed through metastases at various organs without primary tumor identification. Despite the major molecular and technological advances, the carcinogenesis of CUP remains enigmatic which hampers adequate study design of treatments leading to survival improvement. To date, the pathogenesis of CUP is still debatable with one hypothesis considering CUP is simply a group of metastatic tumors with unidentified primaries and another considering it a distinct entity with specific genetic and phenotypic aberrations. Familial CUP seems to favor the first hypothesis due to common genetic predisposition factors between known primaries and CUP. Two clinical implications may be withdrawn from the pathogenesis of familial clustering of CUP. The detailed family history and environmental risk factors may orient towards the primary tumor identification. In cases of familial, **smoking** avoidance and adherence to general population guidelines for cancer screening would be strongly encouraged.

http://www.ncbi.nlm.nih.gov/pubmed/31203526

3 Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI DIRECT Consortium.

Here, we describe the characteristics of the Innovative Medicines Initiative (IMI) Diabetes Research on Patient Stratification (DIRECT) epidemiological cohorts at baseline and follow-up examinations (18, 36 and 48 months of follow-up).

http://www.ncbi.nlm.nih.gov/pubmed/31203377

4 Health behaviours and mental and physical health status in older adults with a history of homelessness: a cross-sectional population-based study in England.

This study compared (1) levels of engagement in lifestyle risk behaviours and (2) mental and physical health status in individuals who have previously been homeless to those of individuals who have not.

http://www.ncbi.nlm.nih.gov/pubmed/31203244

5 Combined effects of lung function, blood gases and kidney function on the exacerbation risk in stable COPD: Results from the COSYCONET cohort.

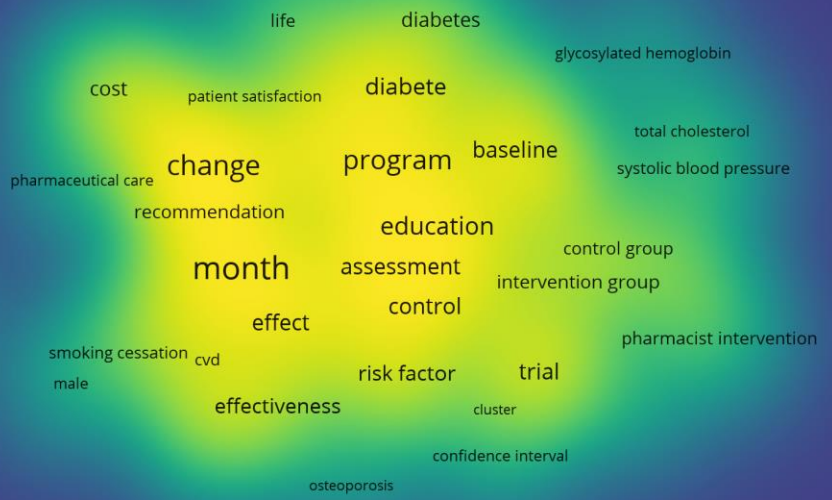
Alterations of acid-base metabolism are an important outcome predictor in acute exacerbations of COPD, whereas sufficient metabolic compensation and adequate renal function are associated with decreased mortality. In stable COPD there is, however, only limited information on the combined role of acid-base balance, blood gases, renal and respiratory function on exacerbation risk grading.

http://www.ncbi.nlm.nih.gov/pubmed/31203096

6 "Don't smoke in public, you look like trash": An exploratory study about women's experiences of **smoking**-related

Query: **smoking** -- Source: **PubMed** (100 results, 0 ms) -- Clusterer: **Lingo**

Map of research of public health interventions in community pharmacies N=338 - titles/abstracts (minimum occurrence of term =10



VOSviewer

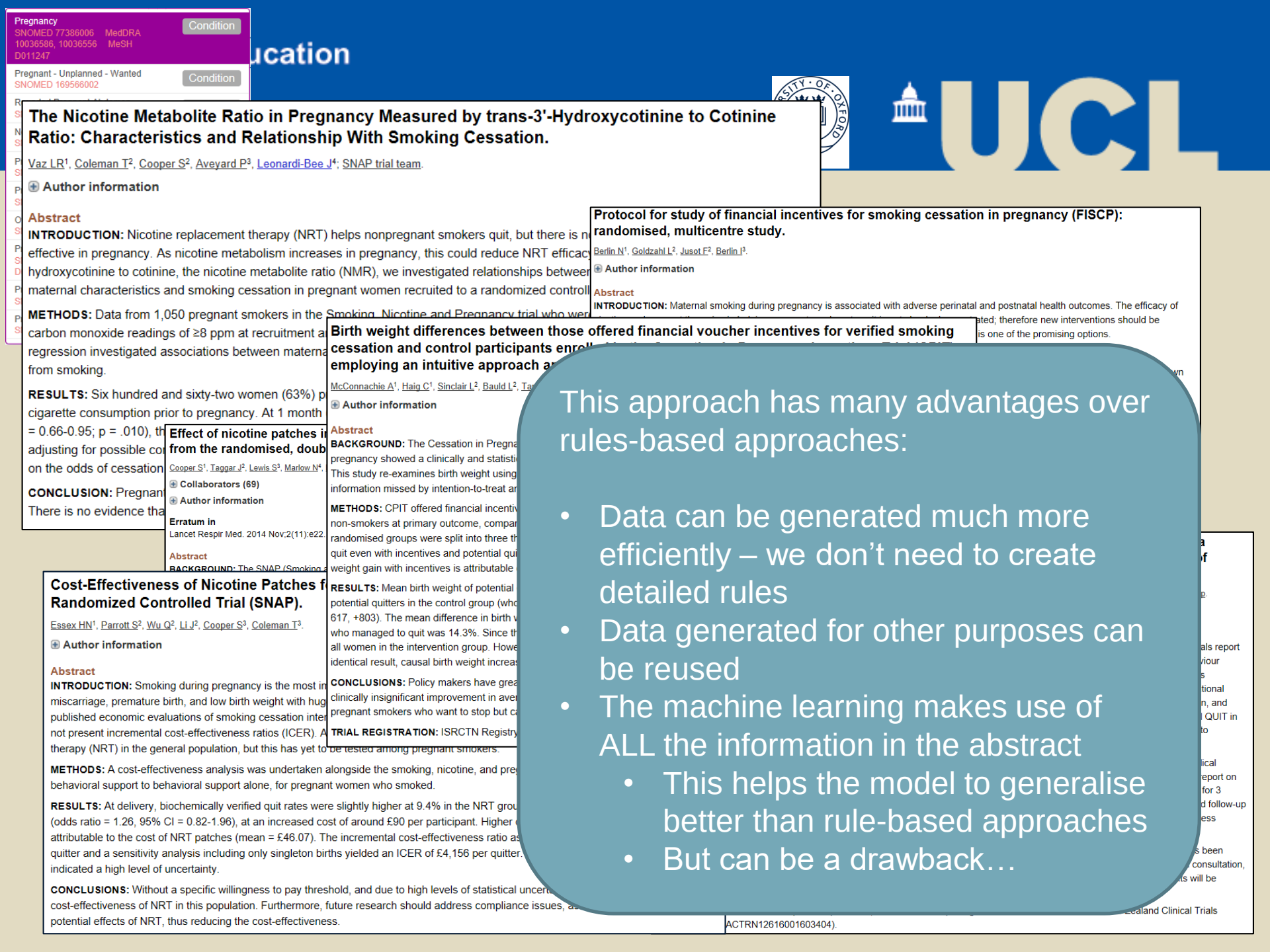
Unsupervised approaches lack control

- Very powerful – can reveal relationships in the data which are not necessarily obvious
- Very efficient – data often need no preparation

- But… you don't get to tell the machine which classifications to make

Supervised approaches

- Humans prepare ‘training’ data – containing data + labels which describe the desired classification
- E.g.
  - Image recognition
  - Text classification


Pregnancy
SNOMED 77386006 MedDRA 10036586, 10036556 MeSH D011247
Condition
Pregnant - Unplanned - Wanted
SNOMED 169566002
Condition
ucation
UCL
The Nicotine Metabolite Ratio in Pregnancy Measured by trans-3'-Hydroxycotinine to Cotinine Ratio: Characteristics and Relationship With Smoking Cessation.
Vaz LR[1], Coleman T[2], Cooper S[2], Aveyard P[3], Leonardi-Bee J[4]; SNAP trial team.
Author information
Abstract
INTRODUCTION: Nicotine replacement therapy (NRT) helps nonpregnant smokers quit, but there is n
effective in pregnancy. As nicotine metabolism increases in pregnancy, this could reduce NRT efficac
hydroxycotinine to cotinine, the nicotine metabolite ratio (NMR), we investigated relationships betwee
maternal characteristics and smoking cessation in pregnant women recruited to a randomized controll
METHODS: Data from 1,050 pregnant smokers in the Smoking, Nicotine and Pregnancy trial who wer
carbon monoxide readings of ≥8 ppm at recruitment a
regression investigated associations between materna
from smoking.
RESULTS: Six hundred and sixty-two women (63%) p
cigarette consumption prior to pregnancy. At 1 month
= 0.66-0.95; p = .010), th
adjusting for possible co
on the odds of cessation
CONCLUSION: Pregnan
There is no evidence tha
Protocol for study of financial incentives for smoking cessation in pregnancy (FISCP): randomised, multicentre study.
Berlin N[1], Goldzahl L[2], Jusot F[2], Berlin I[3].
Author information
Abstract
INTRODUCTION: Maternal smoking during pregnancy is associated with adverse perinatal and postnatal health outcomes. The efficacy of
ated; therefore new interventions should be
is one of the promising options.
Birth weight differences between those offered financial voucher incentives for verified smoking cessation and control participants enrol
employing an intuitive approach a
McConnachie A[1], Haig C[1], Sinclair L[2], Bauld L[2], Ta
Author information
Abstract
BACKGROUND: The Cessation in Pregna
pregnancy showed a clinically and statisti
This study re-examines birth weight using
information missed by intention-to-treat an
METHODS: CPIT offered financial incentiv
non-smokers at primary outcome, compar
randomised groups were split into three th
quit even with incentives and potential qui
weight gain with incentives is attributable
RESULTS: Mean birth weight of potential
potential quitters in the control group (who
617, +803). The mean difference in birth w
who managed to quit was 14.3%. Since th
all women in the intervention group. Howe
identical result, causal birth weight increas
CONCLUSIONS: Policy makers have grea
clinically insignificant improvement in aver
pregnant smokers who want to stop but ca
TRIAL REGISTRATION: ISRCTN Registry
Effect of nicotine patches i
from the randomised, doub
Cooper S[1], Taggar J[2], Lewis S[3], Marlow N[4],
Collaborators (69)
Author information
Erratum in
Lancet Respir Med. 2014 Nov;2(11):e22.
Abstract
BACKGROUND: The SNAP (Smoking
Cost-Effectiveness of Nicotine Patches f
Randomized Controlled Trial (SNAP).
Essex HN[1], Parrott S[2], Wu Q[2], Li J[2], Cooper S[3], Coleman T[3].
Author information
Abstract
INTRODUCTION: Smoking during pregnancy is the most in
miscarriage, premature birth, and low birth weight with hug
published economic evaluations of smoking cessation inter
not present incremental cost-effectiveness ratios (ICER). A
therapy (NRT) in the general population, but this has yet to be tested among pregnant smokers.
METHODS: A cost-effectiveness analysis was undertaken alongside the smoking, nicotine, and pre
behavioral support to behavioral support alone, for pregnant women who smoked.
RESULTS: At delivery, biochemically verified quit rates were slightly higher at 9.4% in the NRT grou
(odds ratio = 1.26, 95% CI = 0.82-1.96), at an increased cost of around £90 per participant. Higher
attributable to the cost of NRT patches (mean = £46.07). The incremental cost-effectiveness ratio as
quitter and a sensitivity analysis including only singleton births yielded an ICER of £4,156 per quitter.
indicated a high level of uncertainty.
CONCLUSIONS: Without a specific willingness to pay threshold, and due to high levels of statistical uncert
cost-effectiveness of NRT in this population. Furthermore, future research should address compliance issues,
potential effects of NRT, thus reducing the cost-effectiveness.
ACTRN12616001603404).
This approach has many advantages over rules-based approaches:
Data can be generated much more efficiently – we don't need to create detailed rules
Data generated for other purposes can be reused
The machine learning makes use of ALL the information in the abstract
This helps the model to generalise better than rule-based approaches
But can be a drawback…

Good supervision is required…

- Very dependent on quality and coverage of training data
- Performance very dependent on context
- For example…

Birth weight differences between those offered financial voucher incentives for verified smoking cessation and control participants enrolled in the Cessation in Pregnancy Incentives Trial (CPIT), employing an intuitive approach and a Complier Average Causal Effects (CACE) analysis.

McConnachie A[1], Haig C[1], Sinclair L[2], Bauld L[2], Tappin DM[3].

Author information

Abstract

BACKGROUND: The Cessation in Pregnancy Incentives Trial (CPIT), which offered financial incentives for smoking cessation during pregnancy showed a clinically and statistically significant in[illegible]d. This study re-examines birth weight using an intuitive an[illegible] information missed by intention-to-treat analysis.

METHODS: CPIT offered financial incentives up to £400 non-smokers at primary outcome, compared to 25 (8.7% randomised groups were split into three theoretical sub- quit even with incentives and potential quitters - require weight gain with incentives is attributable only to potenti

RESULTS: Mean birth weight of potential quitters in the potential quitters in the control group (who did not quit) 617, +803). The mean difference in birth weight betwee who managed to quit was 14.3%. Since the intervention all women in the intervention group. However, "compliar identical result, causal birth weight increase 21 g ÷ 0.14

CONCLUSIONS: Policy makers have great difficulty givi clinically insignificant improvement in average birth weig pregnant smokers who want to stop but cannot achieve s

TRIAL REGISTRATION: ISRCTN Registry, ISRCTN87508788

This means that:

- ALL of the text in the document can be used to 'learn' the classifications
- This increases the model's resilience to minor variations in wording that would break a rules-based system
- The disadvantage is that if you wanted to classify e.g. smoking cessation among young people, you'd need to ensure that the training data also covered young people – or performance would drop

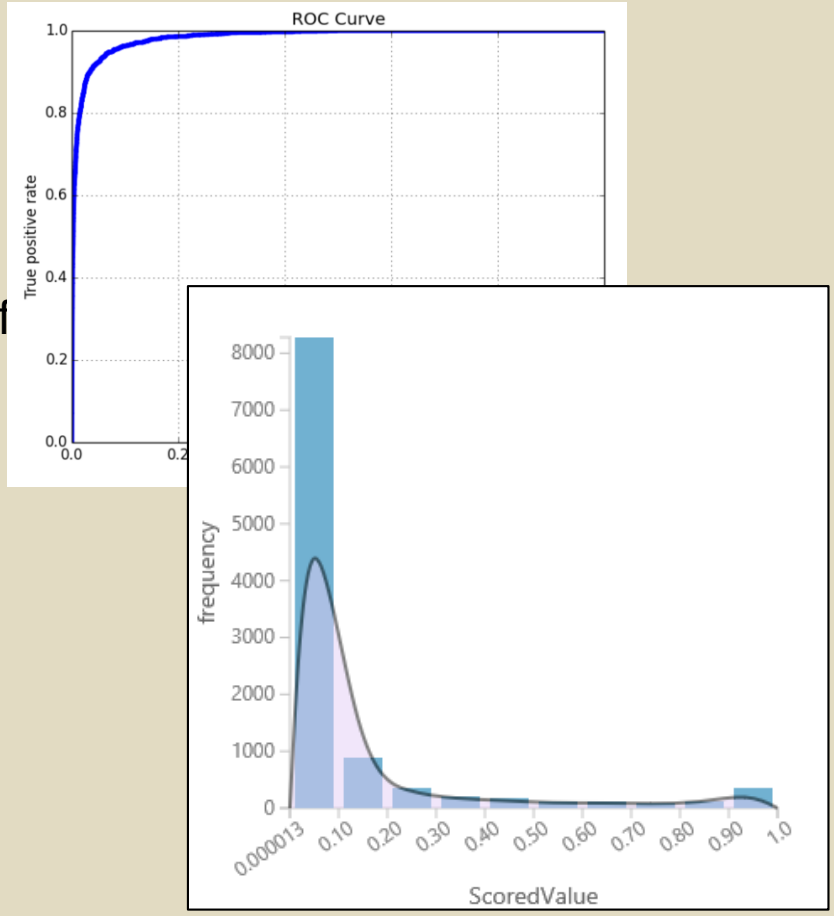
Where can ‘supervision’ come from?

- (‘supervision’: aka ‘training data’)

- Training data always comes from people. It includes:
  - data generated for other purposes (e.g. historic review data)
  - Specifically created data
  - Crowdsourced data

Example of supervised machine learning: the RCT Classifier

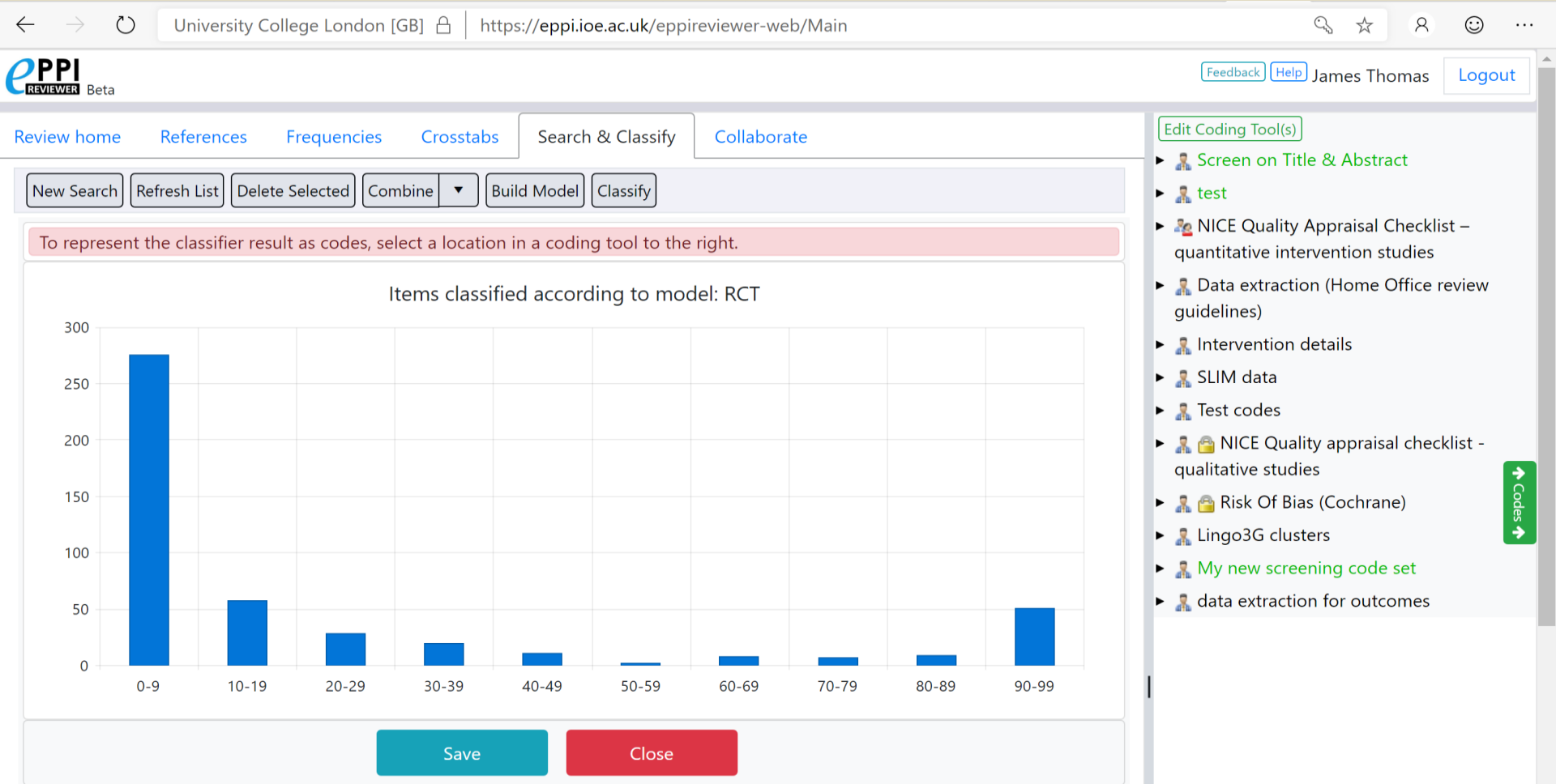
- An RCT classifier was built using more than 280,000 records from Cochrane Crowd
- It has been calibrated to achieve a recall of 99% on the McMaster ‘Hedges’ dataset
- It’s very accurate!
- But not all supervised learning can be so accurate…
- It is ‘simply’ applying single classification (RCT / not RCT) and is built on lots of high quality data



Institute of Education

UCL

Classification demo



Important questions to ask of any supervised machine learning system

- Where did the training data come from?
- Were there sufficient training data to build robust models?
- Were the training data biased in some way?
- How similar are the training data to my use scenario?

Tools

- Klasifiki [https://er5-alpha.ucl.ac.uk/klasifiki]
  - username: guestK password: WorkShop9
- Citation screening (within reviews)
  - Abstrackr
  - EPPI-Reviewer
  - Rayyan
  - Swift ActiveScreener

Ranking search results

- Different technologies operating behind the scenes
- E.g.
  - PubMed ‘similar articles’
  - Microsoft Academic ‘related articles’
  - Medline Ranker ‘ranked’ articles

In summary

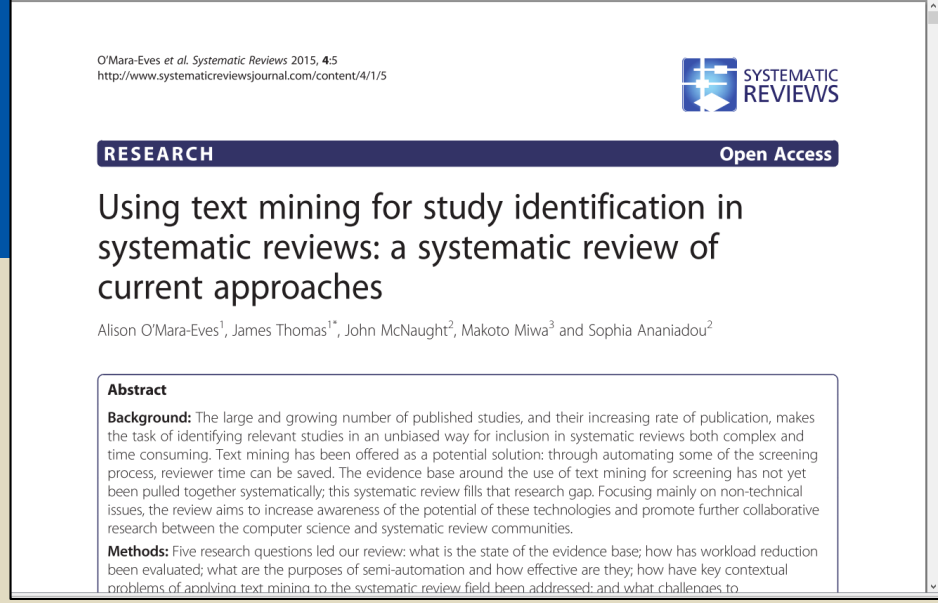
Rule-based	Unsupervised	Supervised
• Not fashionable • Potentially powerful • Very demanding in time • Rules can be fragile	• Very little time effort required to create rules or training data • No control over classifications	• Can utilise lots of training data which can be generated efficiently • Makes use of data created for other purposes • Does not break as easily as rule-based approaches • Can predict specific classification terms (unlike unsupervised)

Supporting citation screening

(if there's time)

Citation screening

- Has received most R&D attention
- Diverse evidence base; difficult to compare evaluations
- ‘semi-automated’ approaches are the most common
- Possible reductions in workload in excess of 30% (and up to 97%)



O'Mara-Eves *et al. Systematic Reviews* 2015, **4**:5
http://www.systematicreviewsjournal.com/content/4/1/5

SYSTEMATIC REVIEWS

RESEARCH Open Access

Using text mining for study identification in systematic reviews: a systematic review of current approaches

Alison O'Mara-Eves[1], James Thomas[1*], John McNaught[2], Makoto Miwa[3] and Sophia Ananiadou[2]

Abstract

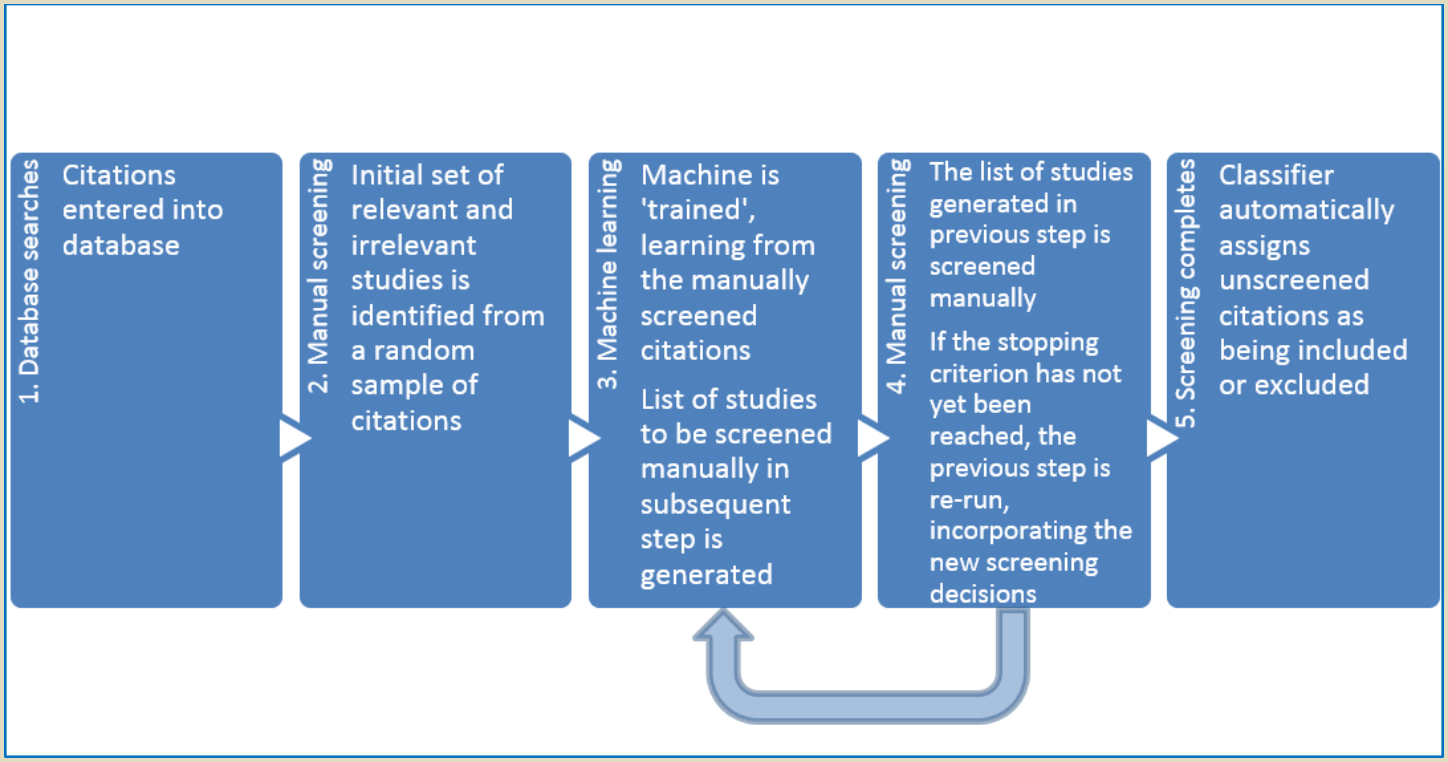
Background: The large and growing number of published studies, and their increasing rate of publication, makes the task of identifying relevant studies in an unbiased way for inclusion in systematic reviews both complex and time consuming. Text mining has been offered as a potential solution: through automating some of the screening process, reviewer time can be saved. The evidence base around the use of text mining for screening has not yet been pulled together systematically; this systematic review fills that research gap. Focusing mainly on non-technical issues, the review aims to increase awareness of the potential of these technologies and promote further collaborative research between the computer science and systematic review communities.

Methods: Five research questions led our review: what is the state of the evidence base; how has workload reduction been evaluated; what are the purposes of semi-automation and how effective are they; how have key contextual problems of applying text mining to the systematic review field been addressed; and what challenges to

Summary of conclusions

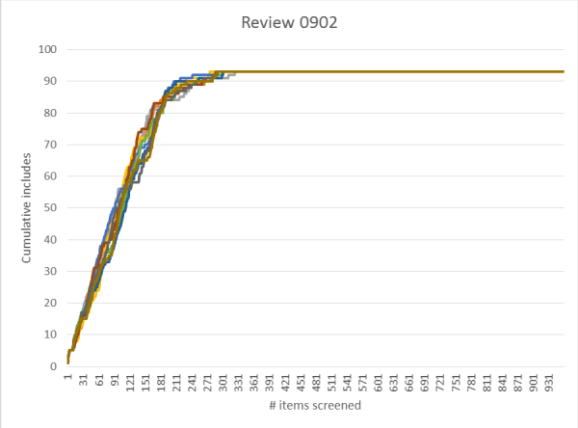
- Screening prioritisation
  - ‘safe to use’
- Machine as a ‘second screener’
  - Use with care
- Automatic study exclusion
  - Highly promising in many areas, but performance varies significantly depending on the domain of literature being screened

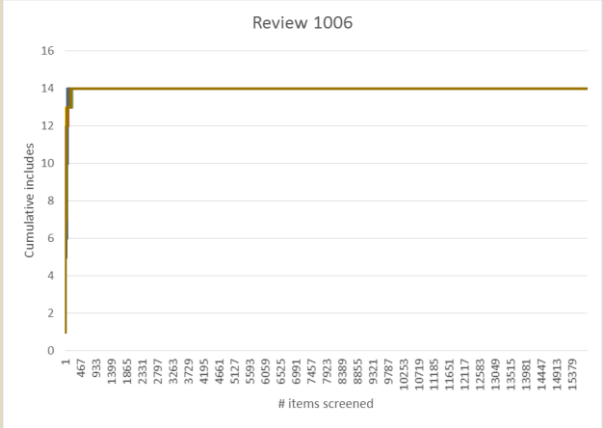
How the machine learns…

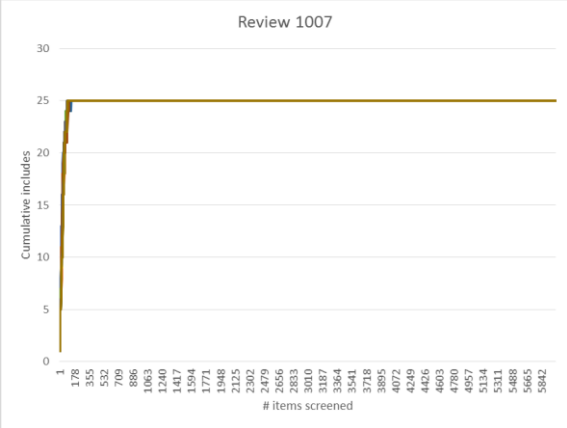


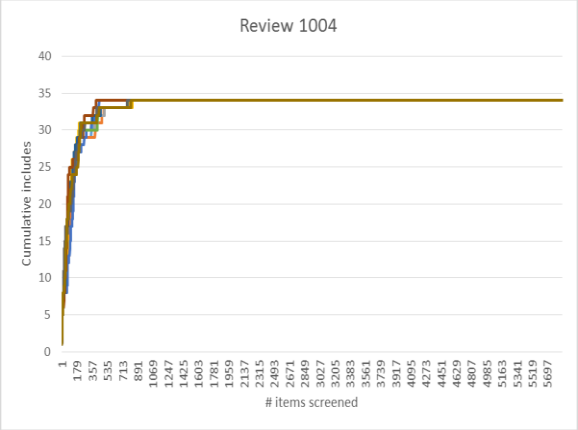
And it can work quite well…

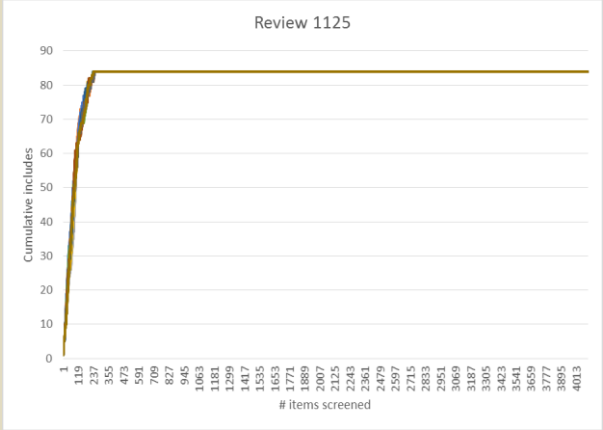
Does it work? e.g. reviews from Cochrane Heart Group

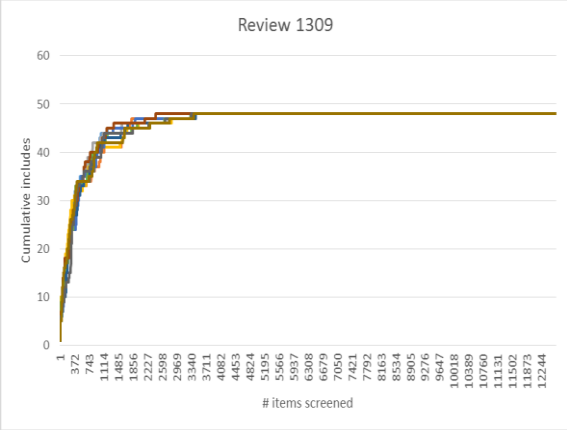












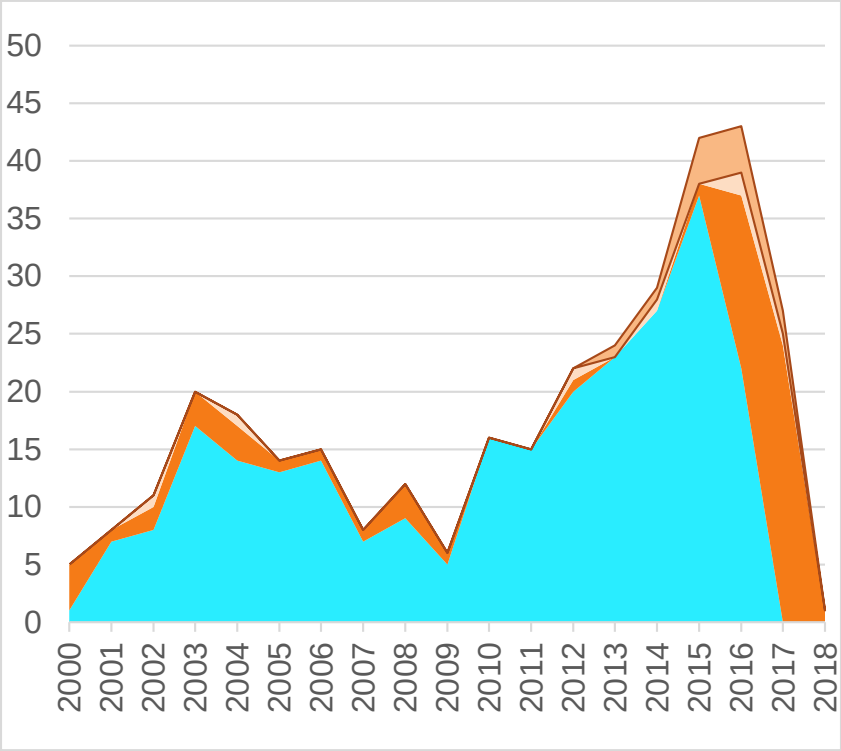
Custom-build classifiers for update searches

Community Pharmacy map of public health interventions

21,555 citations from update search.

Classifier used for 61% reduction in screening titles and abstracts.

- 8,449 title and abstracts for 62 includes
- 1,788 titles only for 7 includes
- Website searches for 12 includes



Challenges

- Uncertainty – when to stop screening
- Managing the screening process

Less useful if

- Vocabulary changes over time
- Eligibility criteria is expanded

Studies in map N=336

- Original searches (n=225)
- Update: relevance-ranked (n=62)
- Update: title-only citations (n=7)
- Update: website searches (n=12)

Using crowdsourcing in health evidence synthesis: Cochrane Crowd

Anna Noel-Storr

18 June 2019

Trusted evidence.
Informed decisions.
Better health.

Cochrane

Crowdsourcing

The practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers

Crowdsourcing

Crowd creation
e.g. Threadless

4 types of crowdsourcing:
crowd creation

Crowdsourcing

Crowd creation
e.g. Threadless

Crowd wisdom
e.g. The SIM exchange

4 types of crowdsourcing:
crowd creation, crowd wisdom

Cochrane

Crowdsourcing

Crowd creation
e.g. Threadless

Crowd wisdom
e.g. The SIM exchange

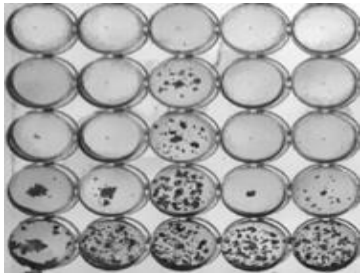
Crowd funding
e.g. Kickstarter

4 types of crowdsourcing:
crowd creation, crowd wisdom, crowd funding

Cochrane

Crowdsourcing

Crowd creation
e.g. Threadless



Crowd voting
e.g. Zooniverse

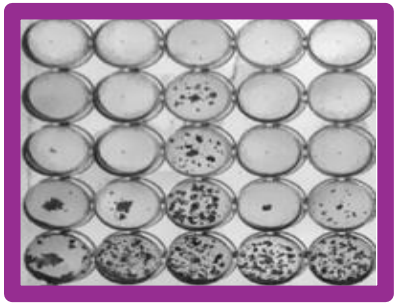
Crowd wisdom
e.g. The SIM exchange

Crowd funding
e.g. Kickstarter

4 types of crowdsourcing:
crowd creation, crowd wisdom, crowd funding, crowd voting

Crowdsourcing

Crowd creation
e.g. Threadless



Crowd voting
e.g. Zooniverse

Crowd wisdom
e.g. The SIM exchange

Crowd funding
e.g. Kickstarter

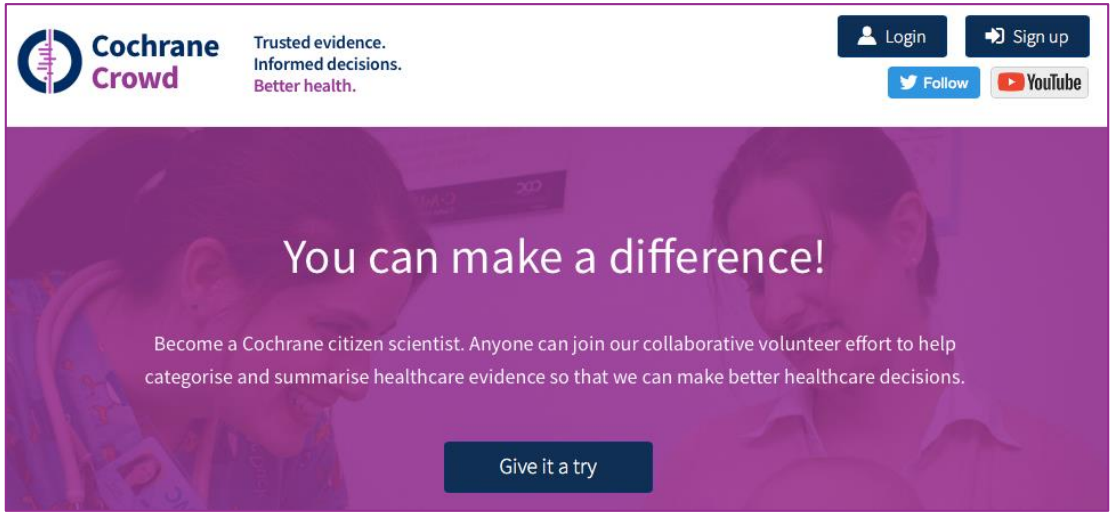
4 types of crowdsourcing:
crowd creation, crowd wisdom, crowd funding, crowd voting

What is Cochrane Crowd?

A platform that hosts **tasks** that help produce high quality health evidence using **crowdsourcing**

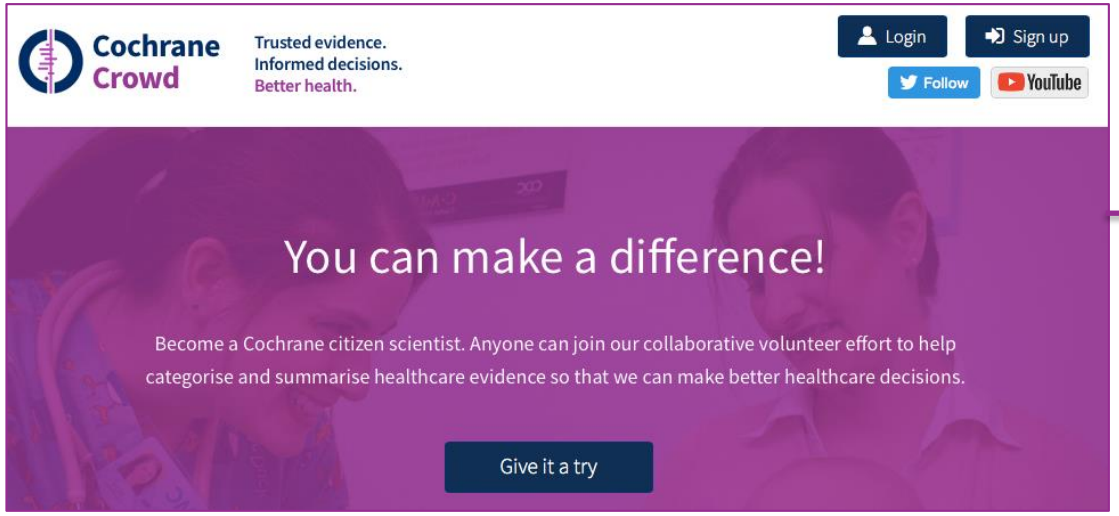
crowd.cochrane.org

Cochrane Crowd



Cochrane Crowd: http://crowd.cochrane.org

Cochrane Crowd



Contributors complete classification tasks

Cochrane Crowd: http://crowd.cochrane.org

Cochrane Crowd

Identifying health evidence

Describing health evidence

Cochrane Crowd: http://crowd.cochrane.org

Classifying or categorising

Effects of combined administration of weight reducing diet and spirulina platensis on anthropometric measures and glycemic markers in obese and overweight subjects: a randomized, double-blinded, placebo-controlled clinical trial [622989046]

Introduction: This study was conducted to determine the possible effects of combined administration of a weight reducing diet and spirulina supplements on the anthropometric measures and glycemic markers in obese and overweight subjects. Materials and Methods: Fifty-two subjects (25<=BMI<40) were enrolled and randomly allocated to receive spirulina platensis (SP) (2 grams per day) or placebo tablets. Anthropometric measures and glycemic markers, including fasting blood glucose, serum insulin and Homeostatic Model Assessment-Insulin resistance (HOMA-IR) were assessed before and after the study; participants also received a weight reduction diet during 12 weeks of intervention. Results: Thirty-eight individuals who completed the intervention, included 38 participants (7 males) with an average age of 40 years and BMI of 32.9 kg/m2. Reduction of body weight and body mass index (BMI) were significantly higher in the SP group, compared to the placebo group (-3.22+/-1.97 kg vs.-1.45+/-1.86 kg, P=0.008 and-1.23+/-0.79 kg/m2 vs.-0.63+/-0.68 kg/m2, P=0.01). Body fat was also considerably reduced in the SP group, compared to the placebo group (-2.28+/-1.74 vs.-1.22+/-1.55, P=0.01). In addition, HOMA-IR reduced marginally, significant in the SP group (P=0.05). Conclusion:

Is the record describing a randomised controlled trial (RCT)?

- Yes
- No
- Unsure

Ensuring accuracy

Each task is supported by brief, interactive training

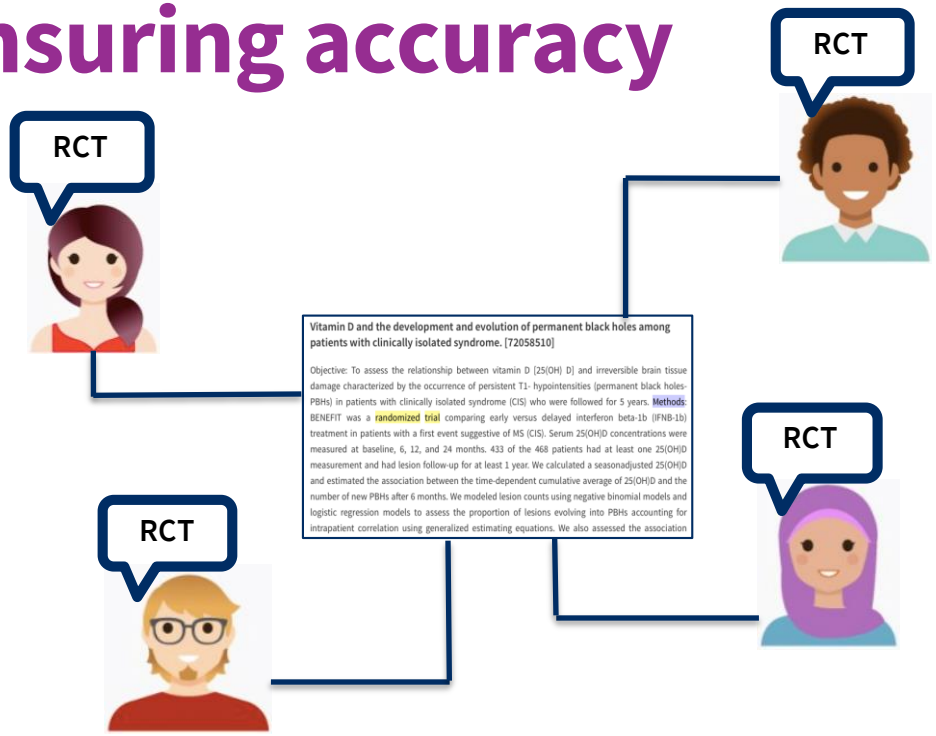
Ensuring accuracy

Each task is supported by brief, interactive training

The training is made up of practice records and commentary

This helps to ensure individual accuracy

Ensuring accuracy



No record is just looked at once. Most records need 4 agreements for it to either be deemed an RCT or not.

This helps to ensure collective accuracy

Ensuring accuracy

RCT Identification

	Info specialist and methodologist	
Cochrane Crowd	TP 457	FP 58
	FN 4	TN 5522

Sensitivity: 99.1% Specificity: 99%

DTA Identification

	Info specialist and methodologist	
Cochrane Crowd	TP 100	FP 22
	FN 20	TN 978

Sensitivity: 83% Specificity: 97%

CTgov Identification

	Info specialist	
Cochrane Crowd	TP 8191	FP 77
	FN 17	TN 5823

Sensitivity: 99.7% Specificity: 98.6%

With the right agreement algorithm in place very high collective accuracy is possible.

Implementation

At the **individual review level** with a workflow called Screen4Me

Implementation: Screen4Me

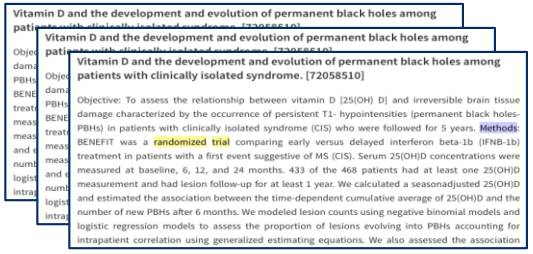
A results screening workflow that uses…

- Known assessments
- RCT classifier
- Cochrane Crowd

Screen4Me: what is it?

Known assessments

The records that have already been through Cochrane Crowd

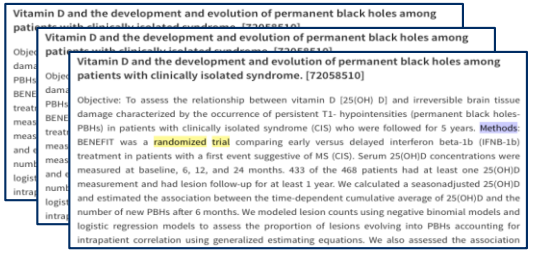


To date, over 500,000 bibliographic records have been through Cochrane Crowd

Screen4Me: what is it?

Known assessments

The records that have already been through Cochrane Crowd

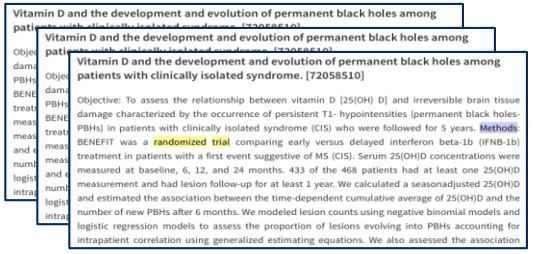


42,500 RCTs

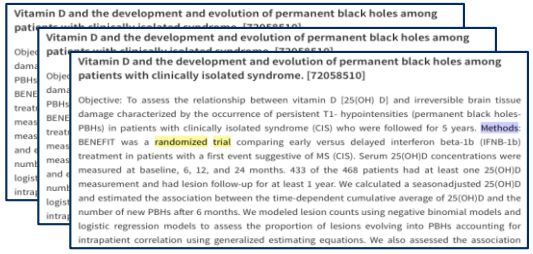
Screen4Me: what is it?

Known assessments

The records that have already been through Cochrane Crowd

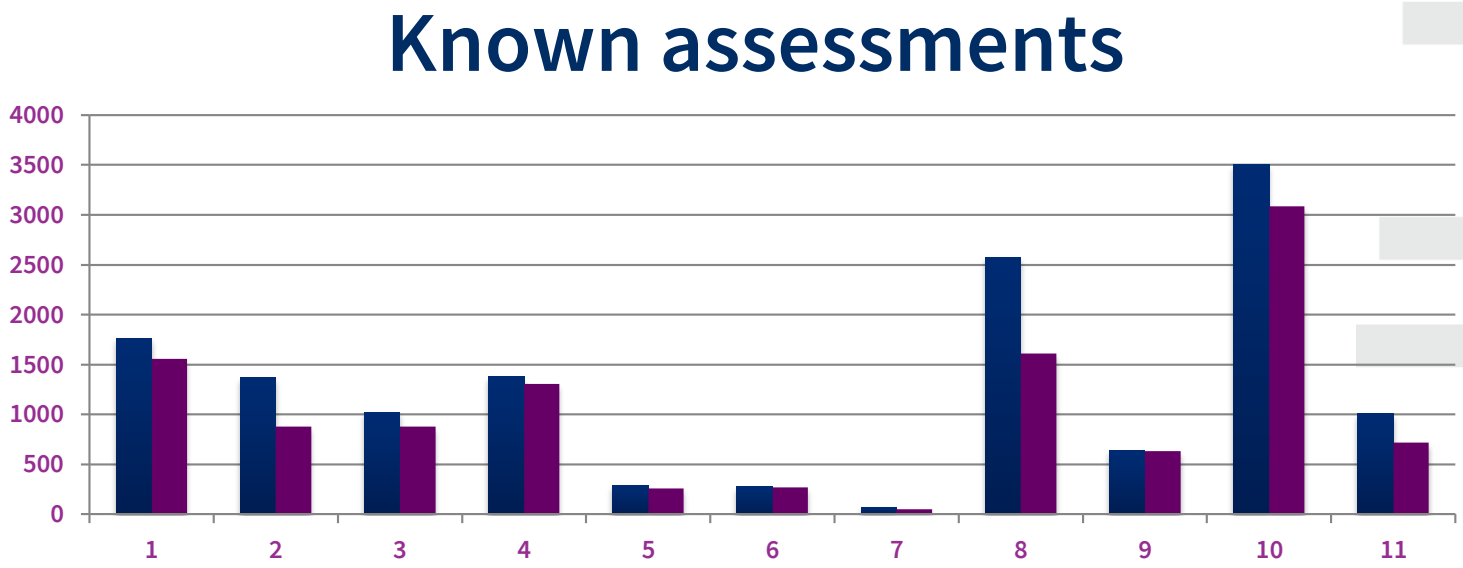


42,500 RCTs



465,000 Rejects

Screen4Me: what is it?



In a month's worth of updated reviews, the results identified in Embase searches for those updates had already been screened by the Crowd: 62%-98%

Screen4Me: what is it?

Known assessments

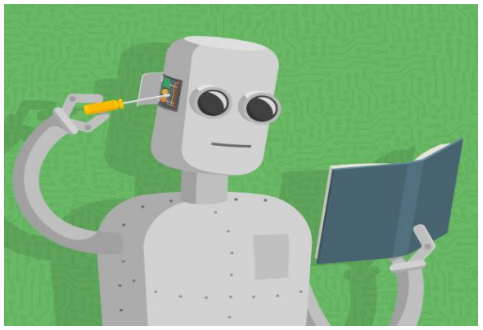
Making better use of known assessments could really help to reduce some duplication of effort.

Over 90% of Cochrane reviews currently include only RCTs.

Screen4Me: what is it?

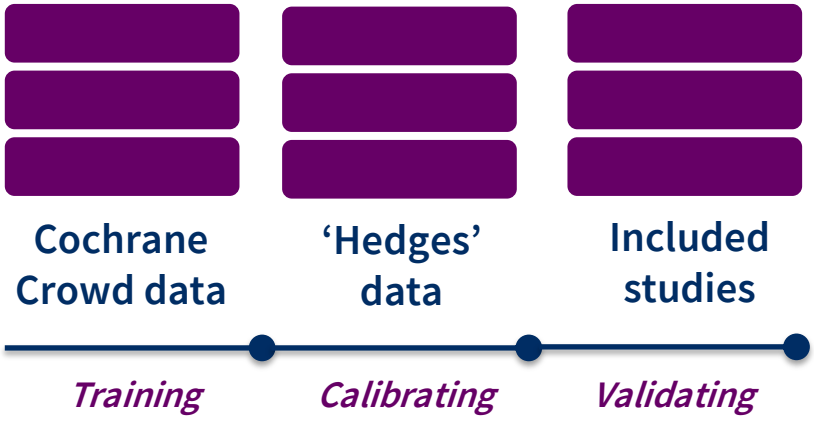
RCT classifier

The classifier assigns records a likelihood score of it describing an RCT



The classifier was built using data (known assessments) generated by Cochrane Crowd.

Screen4Me: what is it?



Three main datasets have been involved in building and testing the RCT classifier.
The machine threshold will be set to achieve a 99% recall as requested by IRMG.

Screen4Me: what is it?

Cochrane Crowd

Online platform:
crowd.cochrane.org

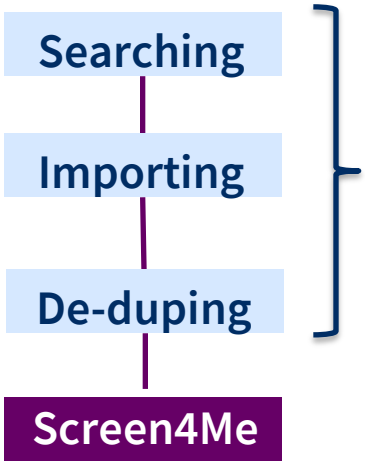
13,500 sign-ups

Screen4Me: what is it?

Brining the three components together and enabling review teams to access them for their specific reviews

- Known assessments
- RCT classifier
- Cochrane Crowd

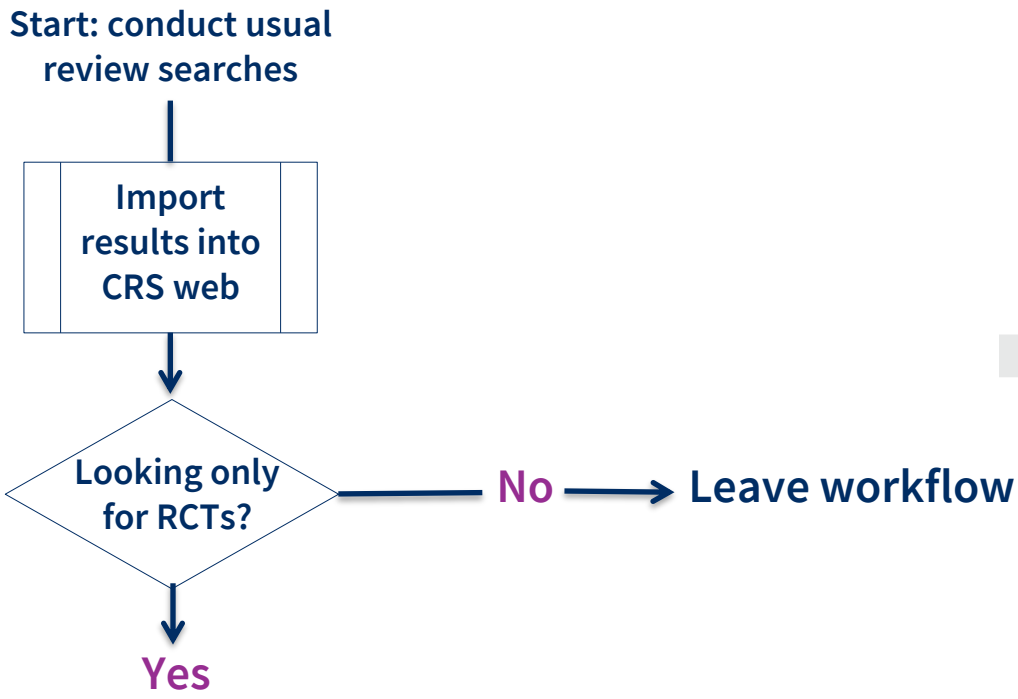
Screen4Me: how will it work?



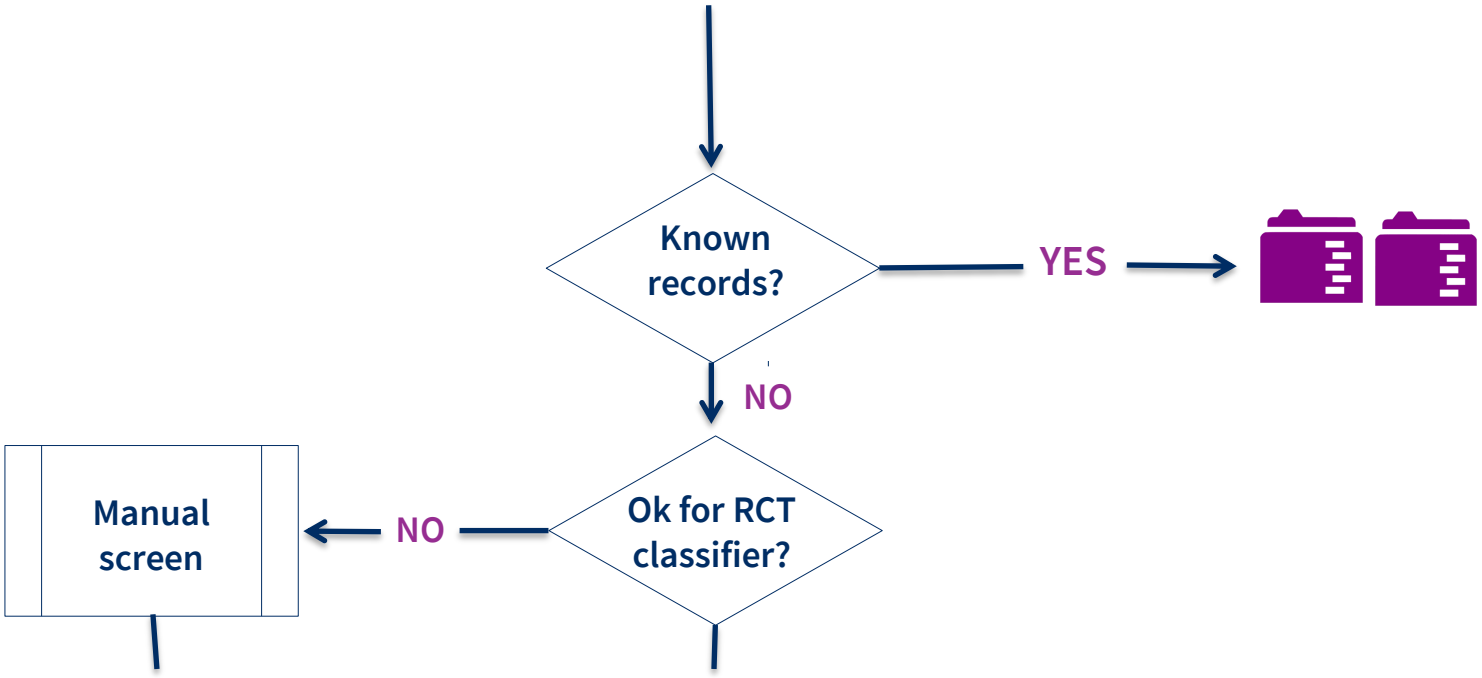
These processes done just as they usually are

Screen4Me is operated from within the CRS web, through folders.

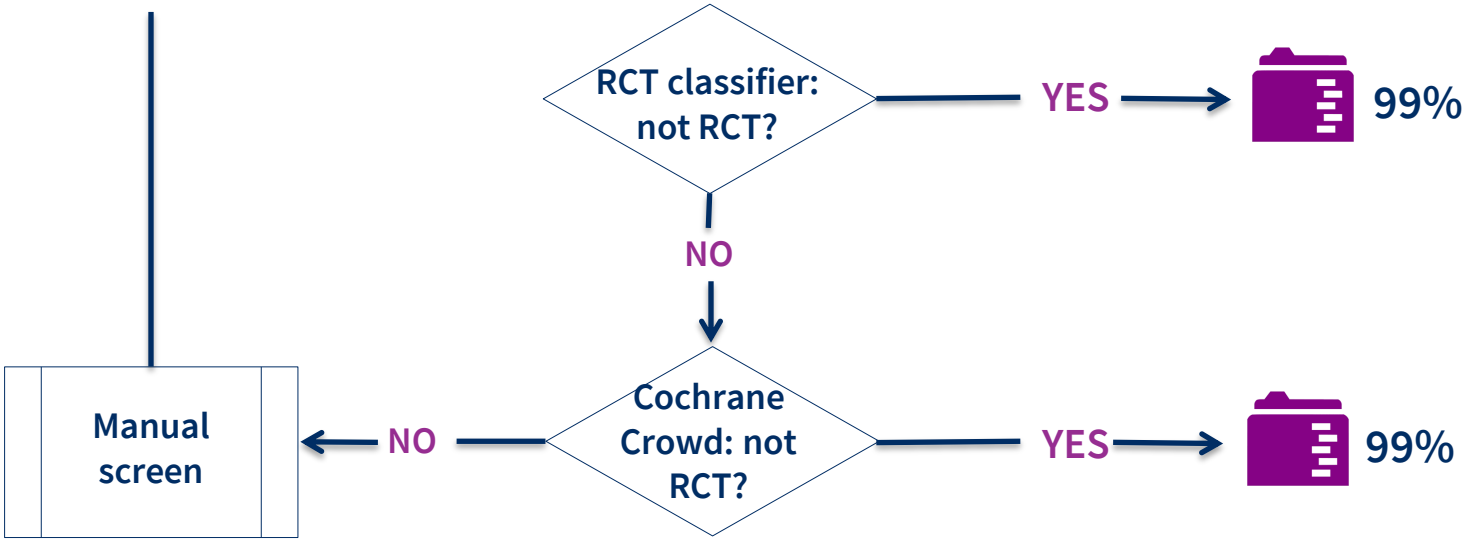
Screen4Me: workflow



Screen4Me: workflow *cont.*

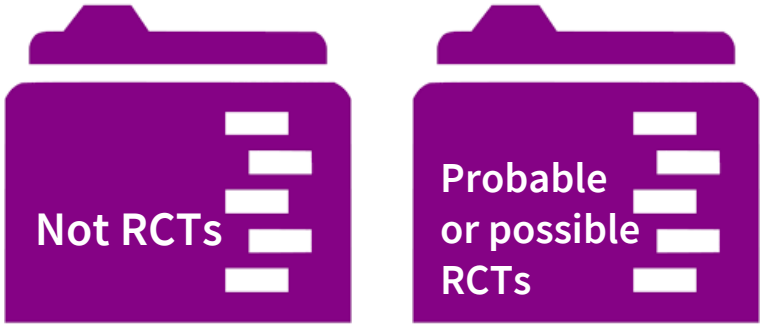


Screen4Me: workflow *cont.*



Cochrane

Screen4Me



Estimated **reduction in records** for author teams to have to screen: **50-85%**

Screen4Me: usage

Used 80 times

150,000 records have gone through Screen4Me

Between 40-70% reduction in the records for authors to screen

Screen4Me has been live since April 2019

Implementation

At the **individual review level** with a workflow called Screen4Me

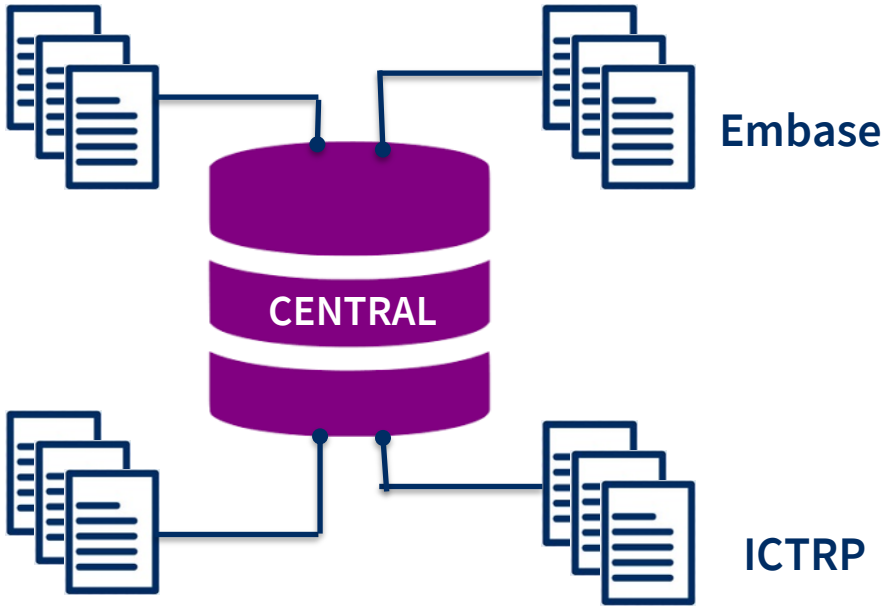
Implementation

At the **individual review level** with a workflow called Screen4Me

At a **'meta' level** through Cochrane's Centralised Search Service

Centralised Search Service

The centralised search service aims to make CENTRAL as comprehensive as possible for RCT evidence



Enabling better surveillance

Moves us closer to a surveillance approach making it easier to find the evidence with far greater specificity but without compromising sensitivity

Cochrane

Time to find some RCTs!

1. Go to Cochrane Crowd: http://crowd.cochrane.org
2. Sign-up or login
3. Go to the task called: RCT ID
4. Click on the: Training records button, then the Start basic training

Cochrane

How many records can we screen in 5 minutes?

Ready. Steady. GO!

Super Screeners

Beatrice
Joshua David
Genevieve C.
Martina
Blanche
Daniela
Samuel
Inga
Alexandre
Joëlle
Caro
Elisabeth
Maria
Lisa
Oliver
Martina
Eric
Nina
Alice
Therese

Speedy Screeners

Isabelle
Kaitlin
Zahra
Wichor
Nicole
Shelley
Leonard
Janice Yu Chen
Guus
Sandy
Pernille Skou
Glyneva
Fiona Joan Laird
Heather K
Pablo
Igor
David
Marli
Doris
Latifa

Cochrane

Thank you!

Trusted evidence.
Informed decisions.
Better health.

Which new approach(es) are you most likely to try out for yourself?

What are your concerns?

What do you think are the potential benefits?

What methods and processes will need to be developed to use these tools?

Institute of Education

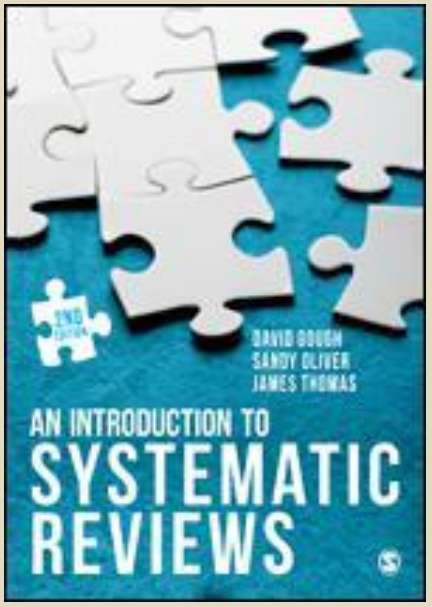
UCL

Thank you

EPPI-Centre website: http://eppi.ioe.ac.uk

Email
j.thomas@ucl.ac.uk
anna.noel-storr@rdm.ox.ac.uk
c.stansfield@ucl.ac.uk

EPPI CENTRE



The EPPI-Centre is part of the Social Science Research Unit at the UCL Institute of Education, University College London

EPPI-Centre
Social Science Research Unit
Institute of Education
University of London
18 Woburn Square
London WC1H 0NR

Tel +44 (0)20 7612 6397
Fax +44 (0)20 7612 6400
Email eppi@ioe.ac.uk
Web eppi.ioe.ac.uk/