

How can we find relevant research more quickly in systematic reviews?

James Thomas and Alison-O'Mara-Eves

Obesity, eating disorders, and the media:
An interdisciplinary workshop

Oxford, 8th November 2011

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



Five streams of work:

- Childhood Studies
- Evaluation of Social Interventions
- Sexual Health, Reproduction and Social Exclusion
- Evidence for Policy and Practice Information and Co-ordinating Centre
- Perspectives, Participation and Research

<http://eppi.ioe.ac.uk/>

Evidence for Policy and Practice Information and Co-ordinating (EPPI) Centre



Conducting reviews since 1993
In health promotion,
education, social care, crime,
transport, work and pensions,
international development

Support and tools for review groups:
Education (25 groups, 70+ reviews),
criminology, employment,
speech and language, social care, development

EPPI-Reviewer software

Formal links
with Cochrane
and Campbell
Collaborations

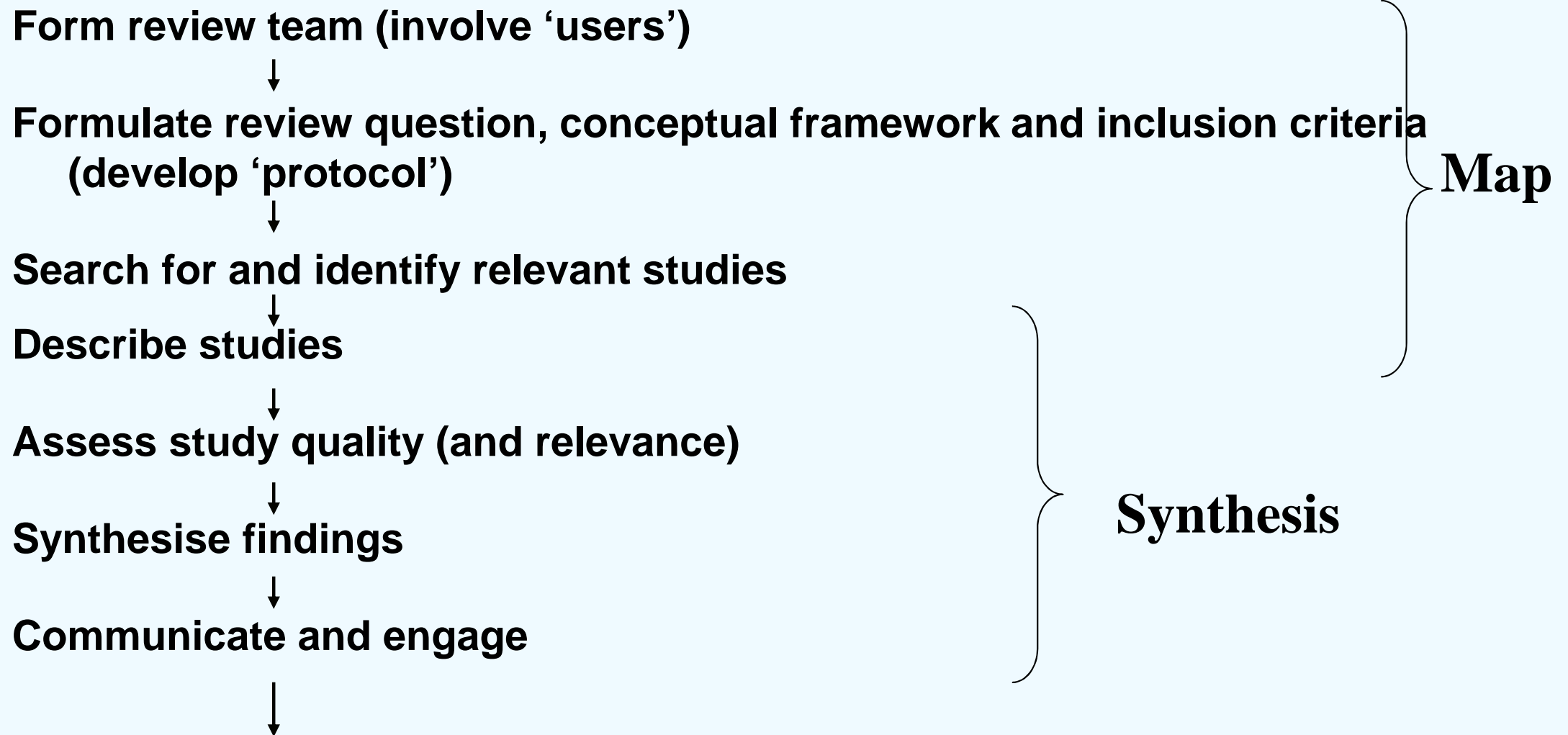
Methodological work, e.g.
Methods for Research Synthesis Node of
the ESRC National Centre
for Research Methods



On-line libraries
of research evidence

Short courses and
Masters course
in evidence for
public policy and practice

The common stages of a systematic review



Recent systematic reviews (all available at <http://eppi.ioe.ac.uk/>)

- Communities that cook: a systematic review of the effectiveness and appropriateness of interventions to introduce adults to home cooking (at peer review)
- Young people's views relating to obesity and body size, shape and weight (forthcoming)
- Caird J, Kavanagh J, Oliver K, Oliver S, O'Mara A, Stansfield C, Thomas J (2011) Childhood obesity and educational attainment: a systematic review. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Rees R, Oliver K, Woodman J, Thomas J (2009) Children's views about obesity, body size, shape and weight: a systematic review. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London.
- Woodman J, Lorenc T, Harden A, Oakley A (2008) Social and environmental interventions to reduce childhood obesity: a systematic map of reviews. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Aicken C, Arai L, Roberts H (2008) Schemes to promote healthy weight among obese and overweight children in England. Report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Finding relevant studies is getting harder...

- There are more papers being published
- More journals being established
- More databases created
- More tools to aid searching (e.g. citation networks)
- More methodological development

But we need to search widely...

- Searching only the large databases (e.g. PubMed) misses important research
- Large databases often have a North-American bias
- Smaller databases often have less powerful search engines, but contain highly relevant studies

The importance of searching a range of sources

How are decisions made about the entry of people aged 65+ to care services?

	SSCI	Medline	CINAHL	Caredata
Unique articles retrieved	237	182	27	16
Unique relevant articles	116	73	24	15

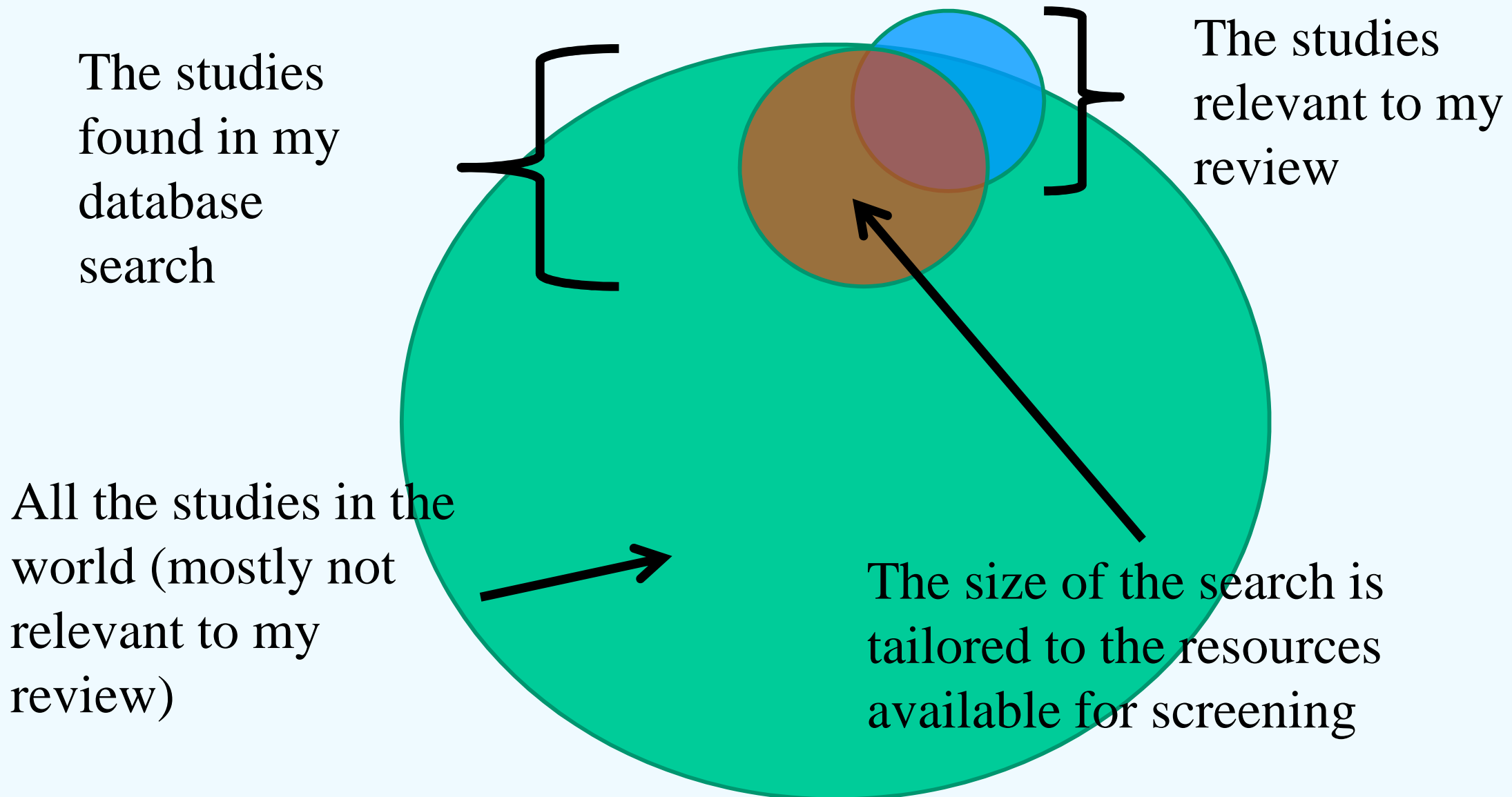
Taylor B, Dempster M and Donnelly M (2003) Hidden Gems: systematically Searching Electronic Databases for Research Publications for Social Work and Social Care. C J Social Work, 33:423-429.

The importance of using a variety of search terms

	Detail of search strategy	Total number of citations	Number of relevant studies	Sensitivity %
1	31 terms	1048	72	100
2	11 terms	669	64	89
3	7 terms	385	47	65

*Adapted from: Harden A, Peersman G, Oliver S, Oakley A (1999) Identifying primary research on electronic databases to inform decision-making in health promotion: the case of sexual health promotion. *Health Education Journal* 58: 290–301.

The traditional approach (not to scale)

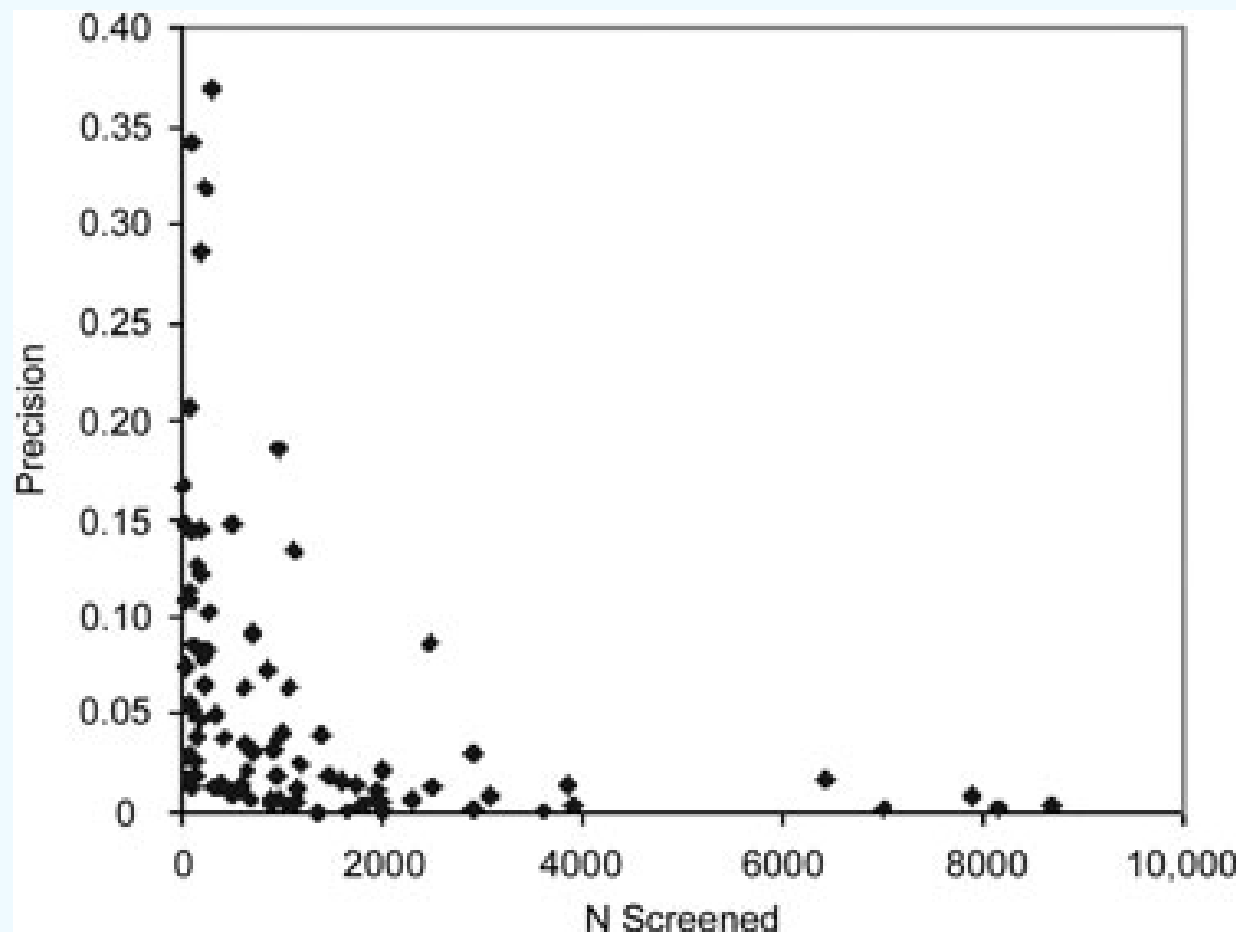


Precision and screening volume

Relationship between screening volume and precision of the search

Retrieving lower numbers of studies to screen means:

- Less screening (!);
- and
- Greater precision

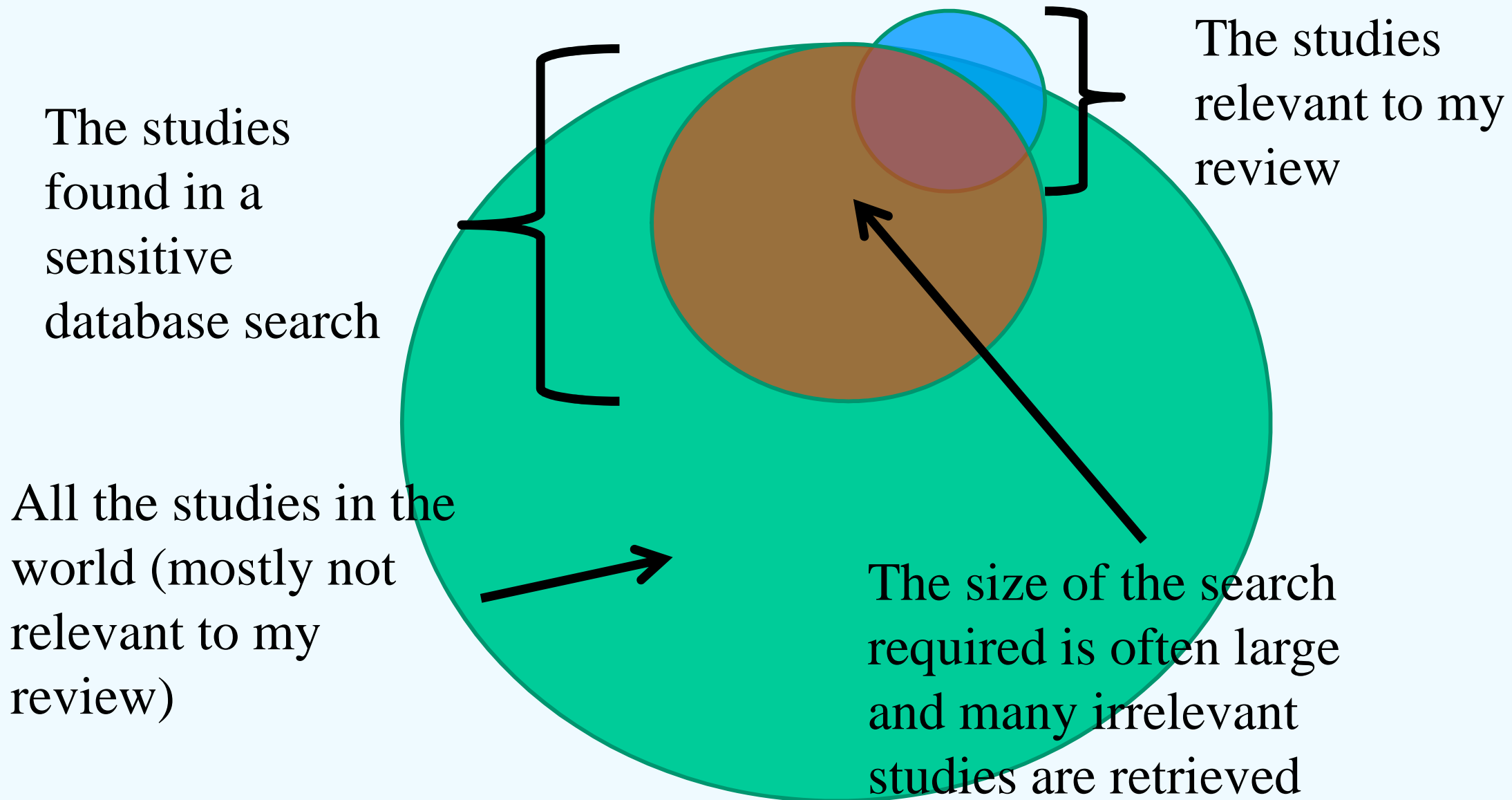


Sampson M, Tetzlaff J, Urquhart C (2011) Precision of healthcare systematic review searches in a cross-sectional sample. *Research Synthesis Methods* 2 (2): 119-125

But we need to search widely...

- Systematic reviews commissioned to inform policymaking are often broad in scope
- Searching to inform policymaking in a UK context often means searching multiple sources

The impact on a sensitive search



What can we do?

- Get bigger teams and / or longer timescales; or
- Find a better way of identifying relevant studies

The screening process

The screenshot shows the EPPI-Reviewer4 (V.4.2.1.1) interface. The main window displays the details for 'Item 14 in current list'. The title is 'Smoking habits in secondary school students.' The author(s) are 'Damas C ; Saleiro S ; Marinho A ; Fernandes G ;'. The month is 'January' and the year is '2009'. The 'Included?' checkbox is checked. The abstract text is visible below. On the left, there is a 'Screening Criteria' panel with various checkboxes, including 'T&A: Include Quantitative 98+ prevalence of sources' which is checked. At the top, there are navigation buttons: 'Next', 'Cancel', and 'Find on web'. A green callout box on the right contains three numbered steps: 1. Read title & abstract, 2. Click include / exclude, and 3. Click 'next' and move on to the next reference (repeat...). Arrows from the callout box point to the 'Next' button and the abstract area.

1. Read title & abstract
2. Click include / exclude
3. Click 'next' and move on to the next reference (repeat...)

The screening process

- Traditional screening methods require that every reference be screened manually
 - Hence the size of the pool of studies to be screened is limited according to the resources available
- This assumes that the relevant studies are randomly distributed throughout all the retrieved studies

Moving the goalposts

- Text mining changes the distribution of relevant studies
- The 'next' study has a higher chance of being relevant



Two technologies

- Term recognition (extraction)
 - ‘screening prioritisation’
- Automatic classification
- (also automatic clustering, but that’s another story)

Term recognition

TerMine (C-value) analysis

Found **235** terms in 4.71 seconds - all terms ([in table](#)) ([in text](#)) - threshold:

Background.

There are high levels of concern about **childhood obesity**, with **obese children** be... Children's attitudes to... and beliefs about... their bodies have also raised concern... this issue can inform...

This **systematic re...** shape or weight a...

Methods.

We conducted **ser...** of 1997 which rep... their methods. Inc...

Findings were syn... findings from close... also consulted you...

Results.

We included 28 st... abuse and isolatio... characteristics to o... their shape.

The National Centre for Text Mining's TerMine tool analyses the text from the relevant titles & abstracts and identifies their most significant terms. (Other methods include standard TF-IDF)

Rank	Term	Score
1	body size	64.875
2	overweight child	17
3	young people	9
4	body shape	8
5	interpretive synthesis	7
6	aggregative synthesis	6
6	overweight people	6
8	additional file	5
8	low quality	5
8	other section	5
11	public health	4.8
12	uk child	4
12	systematic review	4
12	physical activity	4
15	datum collection method	3.169925
15	body size matter	3.169925
17	obese child	3
17	size-related abuse	3
17	young child	3
17	acceptable body	3
21	nationwide public health research programme	2.321928
22	public health research	2.169925
23	healthy size	2
23	childhood obesity	2
23	thin body size ideal	2

Terms are combined in a weighted search

- 'ISABOUT (
 - "art therapy" weight(1), "physical activity" weight(0.952), "sport participation" weight(0.802), "mental health" weight(0.747), "control group" weight(0.702), "young people" weight(0.602), "music therapy" weight(0.592), "web site" weight(0.551), "tai chi" weight(0.531), "youth sport" weight(0.516), "sport participant" weight(0.401), "odd ratio" weight(0.401), "volunteer youth sport coach" weight(0.401), "greek traditional dance" weight(0.397), "public library" weight(0.351), "community theatre" weight(0.351), "martial art" weight(0.351), "team sport" weight(0.351), "lifetime incidence" weight(0.351), "athletic identity" weight(0.351), "easy tai chi" weight(0.347), "body image dissatisfaction" weight(0.347), "visual art" weight(0.334), "receptive language skill" weight(0.318), "body image" weight(0.317), "music education" weight(0.301), "rett syndrome" weight(0.301), "therapeutic art-making" weight(0.301), "arts-related predictor" weight(0.301), "alcohol intoxication" weight(0.301), "job satisfaction" weight(0.301), "community-based art classroom" weight(0.268), "experimental group" weight(0.251), "rugby player" weight(0.251), "blood pressure" weight(0.251), "fracture risk" weight(0.251), "therapeutic theatre" weight(0.251), "socioeconomic status" weight(0.251), "cultural expenditure" weight(0.251), "aggressive behavior" weight(0.251), "physical ability" weight(0.251), "recreational sport" weight(0.251), "yoga group" weight(0.251), "physical exercise" weight(0.251), "yogic practice" weight(0.251), "practical implication" weight(0.251), "creative art" weight(0.245), "box office revenue" weight(0.238), "diastolic blood pressure" weight(0.238), "confirmatory factor analysis" weight(0.238), "lumbar bone density" weight(0.238), "senior activity center" weight(0.238), "public cultural expenditure" weight(0.238), "oxford medial unicompartmental knee arthroplasty" weight(0.233), "library service" weight(0.201), "stigma consciousness" weight(0.201), "soccer world cup final" weight(0.201), "athletic identity measurement scale" weight(0.201), "football player" weight(0.201), "entertainment-education radio soap opera" weight(0.201), "physical disability" weight(0.201), "music activity" weight(0.201), "silver yoga" weight(0.201), "lumbar bone mineral density" weight(0.201), "art classroom" weight(0.201), "young child" weight(0.201), "sleep quality" weight(0.201), "dependent variable" weight(0.201), "massachusetts avenue 7th floor" weight(0.201), "state anxiety" weight(0.201), "social identity" weight(0.201), "teaching method" weight(0.201), "explanatory power" weight(0.201), "positive emotion" weight(0.201), "health status" weight(0.201), "research project" weight(0.201), "psychologic profile" weight(0.201), "leisure activity" weight(0.201), "united state" weight(0.201), "insomniac prisoner" weight(0.201), "poor sport behavior" weight(0.188), "scottish highland game" weight(0.188), "physical fitness activity" weight(0.188), "parent-created sport climate" weight(0.188), "highland game" weight(0.175), "bone density" weight(0.175), "age group" weight(0.175), "world cup" weight(0.163), "exploratory factor analysis" weight(0.159), "adolescent sex offender" weight(0.159), "mental health indicator" weight(0.159), "spectator decision-making inventory" weight(0.159), "feminist performance activism" weight(0.159), "contemporary anti-war movement" weight(0.159), "radio communication project" weight(0.159), "no big deal" weight(0.159), "long-term care facility" weight(0.159), "chestnut street suite" weight(0.159), "black football player" weight(0.159), "active music therapy" weight(0.159), "publisher sage publications." weight(0.159), "spinal cord injury" weight(0.159), "art therapy group" weight(0.159), "visual art programme" weight(0.159), "35-year-old age group" weight(0.159), "radial bone density" weight(0.159), "poor sportspersonship behavior" weight(0.159), "soft tissue injury" weight(0.159)
- (20) •)')

The result

- The result is an ordered list of titles and abstracts
 - With those that are *most similar* to the ones already marked as ‘include’ at the top
- The person screening continues to screen as usual, but behind the scenes the titles and abstracts remaining are re-ordered regularly (e.g. every 25 items)

Automatic classification

- Appears to the reviewer to be much the same as the other method
- Behind the scenes, a *support vector machine* is automatically classifying every remaining reference as being included or excluded
- Potentially a more refined approach -



Additional refinements / issues

- Target ‘boundary’ items – not those that can be easily classified as include or exclude
- We need to protect against *hasty generalisation*: i.e. ‘teaching’ the machine to look for certain subsets of studies
 - Use another tool to act as a check on the classifier such as reviewer-identified terms* or automatic clustering
- * Wallace BC, Small K, Brodley CE, Trikalinos TA (2010) Active Learning for Biomedical Citation Screening. ACM SIGKDD Conference on knowledge discovery and data mining, Washington.

Does it work?

- More research is needed
- Simulation studies using data from completed reviews refine appropriate technologies
- Active use of these tools in reviews helps to develop methods
 - E.g. ‘Baseline inclusion rate’

But does it work?

- Preliminary evaluations are promising
- E.g. in a review about young people's access to tobacco:
 - 36,000 titles and abstracts to screen
 - Baseline inclusion rate: 1.81%
 - i.e. about 652 items
 - 656 items identified as being relevant after screening 9,100 items (7.16%)

Conclusions (1)

- Text mining enabled us to identify the expected number of relevant studies with only 25% of the usual manual work, saving time.
- Prioritised screening allows the full-text document retrieval process to begin sooner, which can help prevent disruptions to workflow caused by delays in accessing copies of documents .
- One possible limitation is that it is impossible to know whether everything that was relevant has been found – short of reading all 36,000 titles and abstracts.
- Wallace et al also found that classification offered large potential savings in the screening burden.

Conclusions (2)

- Further evaluative work is needed before we are able to be more definitive.
- This method is highly promising and may save significant time and money, enabling research to be made available to policy and practice in a more timely way than can be achieved currently.
- Text mining shifts the emphasis of identification from the searching stage to screening. The bespoke nature of text mining tools allows greater control over the reasons for potentially missing relevant studies than can be achieved by narrowing the search process.

Further information

- Thomas J, McNaught J, Ananiadou S (2011) Applications of text mining within systematic reviews. *Research Synthesis Methods*. 2(1): 1-14.
- Ananiadou S, Okazaki N, Procter R, Rea B, Sasaki Y, Thomas J (2009) Supporting Systematic Reviews using Text Mining. *Social Science Computer Review*. 27(4)

Thank you

SSRU website: <http://www.ioe.ac.uk/ssru/>
SSRU's EPPI website: <http://eppi.ioe.ac.uk>

j.thomas@ioe.ac.uk



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