# Automated support for systematic reviews: dream or reality ?

**Workshop contributors:**

- Jeremy Wyatt (Wessex Institute, Southampton): Workshop aims & scope; overview of the potential role of automated tools to support the SR process
- James Thomas (EPPI Centre, UCL): How well do current and emerging tools perform ?
- Elaine Williams (NETSCC, Southampton): Can study publishers such as the NIHR Journals Library provide machine readable protocols and study results ?
- Geoff Frampton, (SHTAC Southampton): That's all very well, but how might these tools help me ?
- You: discussion on training needs, likely niche areas of use, user requirements, criteria for adoption etc.
- JW: Closing remarks & next steps

UNIVERSITY OF Southampton

# Workshop aims & Scope

**Aims:**

- To help reviewers understand the current and potential role of automation in supporting the SR process
- To help those working on automated tools to better understand the review process and reviewers' needs
- To explore the implications of automated support tools for reviewers

**Scope:** tools that go beyond simple data management

**Outputs**: report & recommendations for partners; journal article / manifesto; other ?

# Overview of SR automation

Jeremy Wyatt
Professor of Digital Healthcare & Director,
Wessex Institute, University of Southampton
j.c.wyatt@soton.ac.uk

# Overview

- Do we have a problem with SRs ?
- Why is this happening ?
- Where *might* technology be able to help ?
- Insights from Rogers & Gartner
- Some key questions to ask

# Crequit's question: Do SRs include relevant evidence?

**Methods**:

- Identified 29 SRs (13 since 2013) on 47 treatments for non-small cell lung cancer
- Compared with 6 cumulative network meta analyses 2009-2015 of 77 RCTs (pub 2000-Nov 2014) on same treatments (54 comparisons, 29000 pts)

**Results:**

- SRs in best year covered 55% of RCTS, 70% of patients, 60% of treatments, 62% of comparisons
- Persisted when they excluded RCTs on drugs that failed Ph2 studies, were pub. as abstracts or after the last SR
- Median interval from last SR search to publication: 9m (IQR 5-13m)
- Only 21% of SRs reported duplicate study selection & extraction, comprehensive search of lit + industry sources

**Conclusions:** "*SRs of a given condition provide a fragmented, out of date panorama of the evidence…. This waste of research might be reduced by cumulative network meta analysis*". Crequit et al, BMC Medicine 2016

UNIVERSITY OF Southampton

# Crequit's live cumulative network meta-analysis

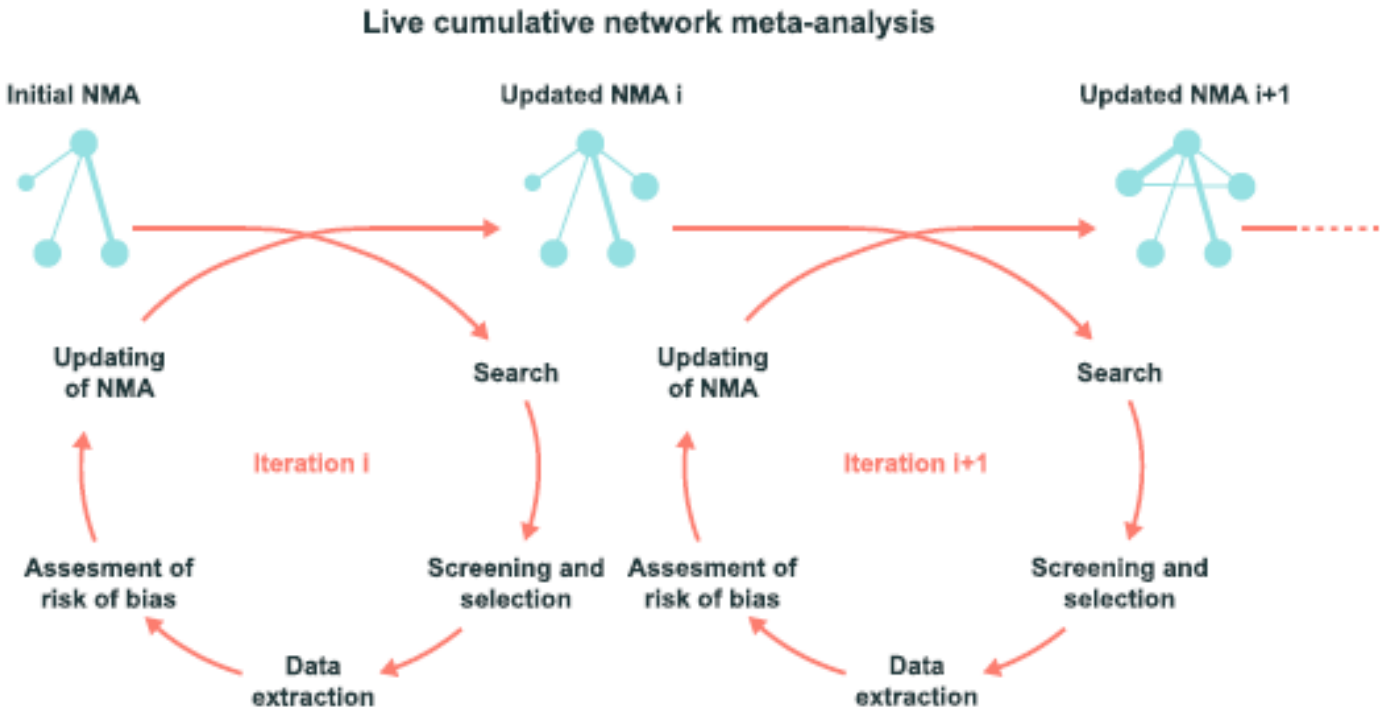



**Fig. 5** A new approach to synthesize evidence: live cumulative network meta-analysis. Starting from an initial NMA, a research community would regularly (e.g., every 3 months), search for, screen, and select trials with new results and, if any, extract data, assess the risk of bias, and update the NMA. NMA: network meta-analysis

UNIVERSITY OF Southampton

# Some possible reasons for these problems

**Supply side challenges:**

- The tsunami of new trials: 40,000 pa. (ie. > 100 / day) [PT = clinical trial, publication year = 2014]
- Trials published only as abstracts: 20% in Crequit 2016
- Inadequate RCT reports eg. intervention descriptions (TIDIER checklist)
- Wider range of interventions & measures, inadequate lexicon & indexing processes

**SR process issues:**

- Increasingly complex review processes following growing evidence of SR biases and shortcomings
- Shortage of SR funding and skilled review staff
- Reluctance of some J to publish SR updates
- Insistence of some reviewers to use gold standard methods even when time & resources are short
- Failure to exploit new technology (Elliott 2014, Tsafnat 2014) – or new tech that doesn't tackle the real problems ?

# Some barriers to review excellence

| Stage | Barrier | Potential solution |
|---|---|---|
| Searching | Too many studies | Cl… Queries, PubMed "Studies …his" ? |
| | Missing studies | CRG study registers<br>Full text searches ?<br>Natural language understanding ?<br>Machine translation ? |
| Critical appraisal | M…, poor quality …dies | Duplicate assessment<br>Robot Reviewer ? |
| Data extraction | Incorrect data | Duplicate extraction<br>XML structured study reports |
| Data synthesis | Ignoring heterogeneity | Check $I^{2,}$ investigate via sensitivity analysis etc. |
| | Other ? | |

We need studies – evidence – to inform this !

UNIVERSITY OF Southampton

# Emerging tools to consider

Search, screening & updating:

- Query expansion
- Machine translation
- NLU for full text searches
- ML to build RCT database

Critical appraisal:

- Robot Reviewer etc.

Data extraction:

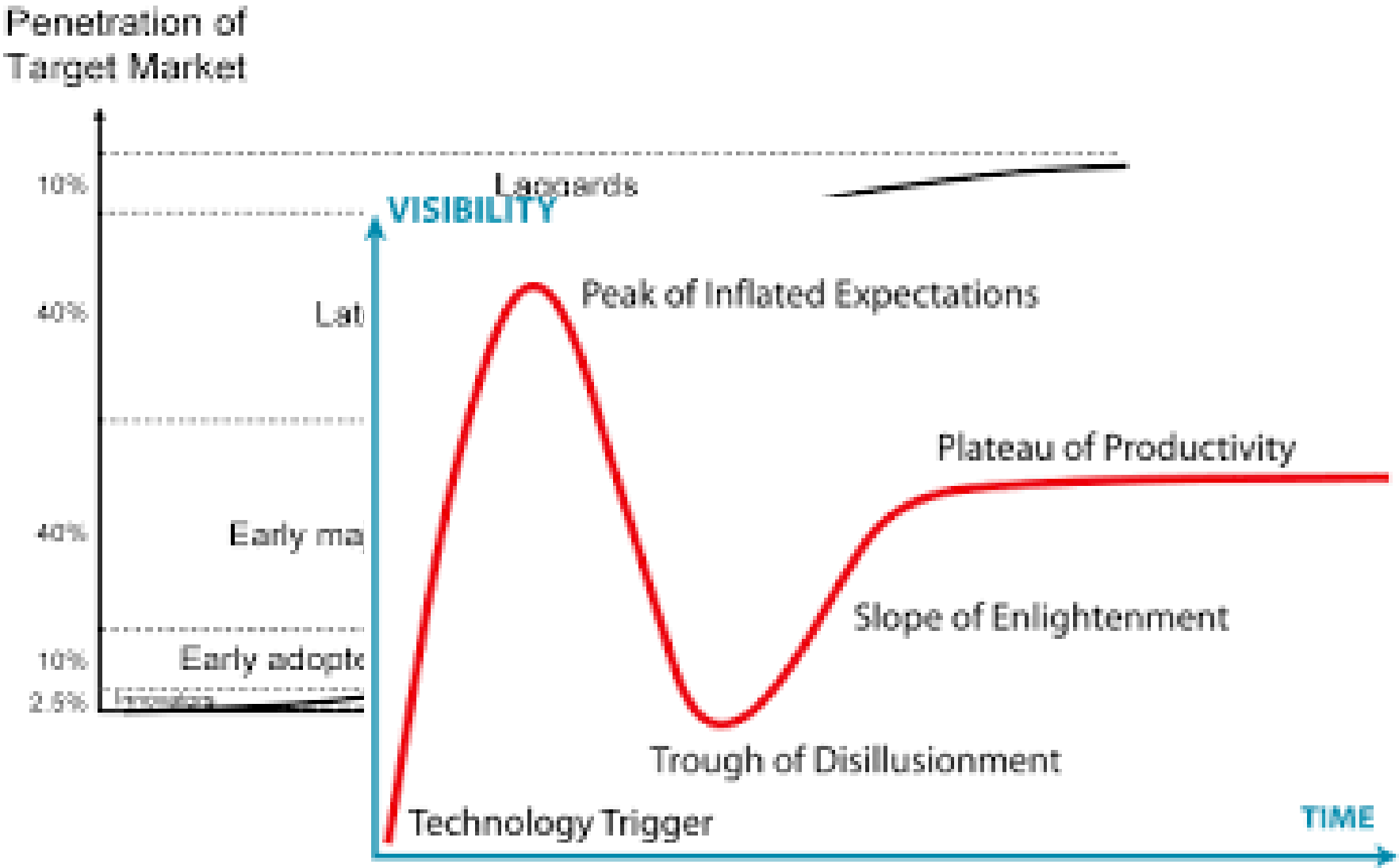
- Machine translation
- XML-structured study reports (methods & data)
- Natural language understanding for automated data extraction

Synthesis and conclusions:

- Automated synthesis tools
- Automated summaries
- Graphical summaries / data graphics

All stages: support for crowd sourcing

# Where are we on the Rogers curve and Gartner Hype cycle ?



# Some questions

1. What are the **real** reviewing problems & challenges that reviewers need help with ?
2. How easy to use, fast and accurate **are** these automated tools now ?
3. How fast & accurate would these tools **need** to be to help us ?
4. How to link up tool developers with **typical** reviewers, to ensure that the resulting tools are usable and useful ?
5. What are the potential **implications** of these tools:
   - Will we need training in these tools ?
   - Will we see de-skilling of reviewers ?
   - Will they hasten moves towards structured methods & results sections in study reports (Ida Sim's Trial Bank) ?
6. Should we even start from here, or is now the time to re-engineer the whole knowledge chain

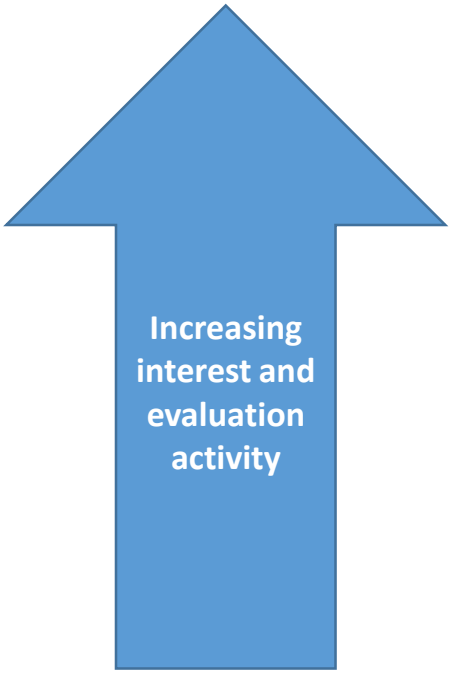
UNIVERSITY OF Southampton

# How well do current and emerging tools perform?

James Thomas, EPPI Centre, UCL

# Tools can perform different functions

- Search screening and updating
  - Screening of citations
  - ‘Mapping’ research activity
  - Database creation / curation
- Critical appraisal
- Data extraction
- Synthesis and conclusions



# 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%
- Automation can help in three areas, with increasing ‘risk’ to obtaining 100% recall:

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.

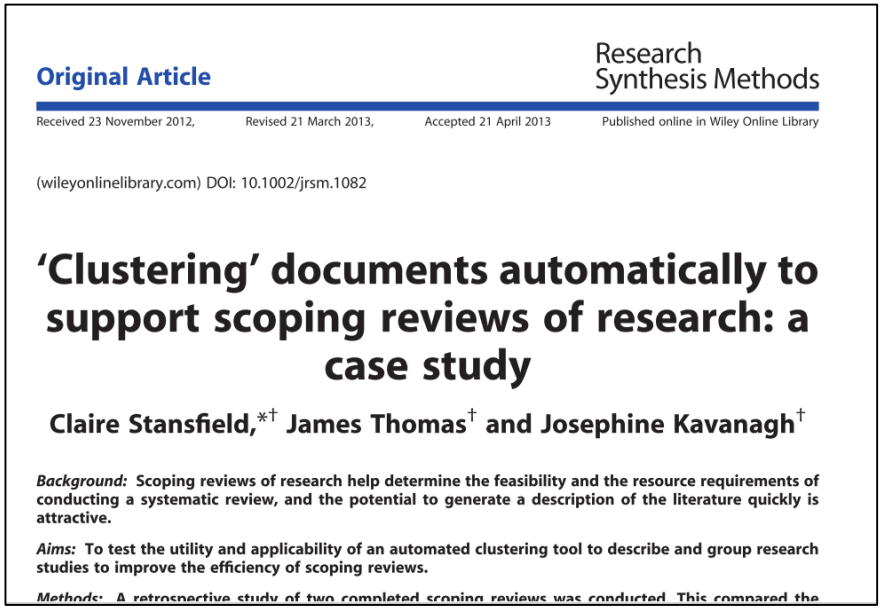
**Methods:** Five research questions led our review: what is the state of the evidence base; how has workload reduction been evaluated; what are the purposes of semi-automation and how effective are they; how have key contextual problems of applying text mining to the systematic review field been addressed;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

- 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

UNIVERSITY OF Southampton

# Mapping research activity

- It is possible to apply ‘keywords’ to text automatically, without needing to ‘teach’ the machine beforehand
- This relies on ‘clustering’ technology – which groups studies which use similar combinations of words
- Very few evaluations
  - Can be promising, especially when time is short
  - But users have no control on the terms actually used



Original Article

Research Synthesis Methods

Received 23 November 2012, Revised 21 March 2013, Accepted 21 April 2013 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1082

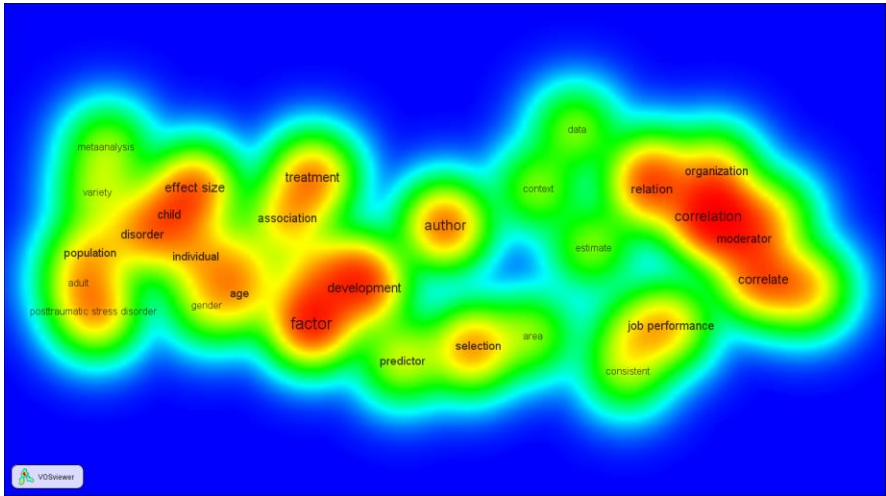
**‘Clustering’ documents automatically to support scoping reviews of research: a case study**

**Claire Stansfield,*† James Thomas† and Josephine Kavanagh†**

***Background:*** **Scoping reviews of research help determine the feasibility and the resource requirements of conducting a systematic review, and the potential to generate a description of the literature quickly is attractive.**

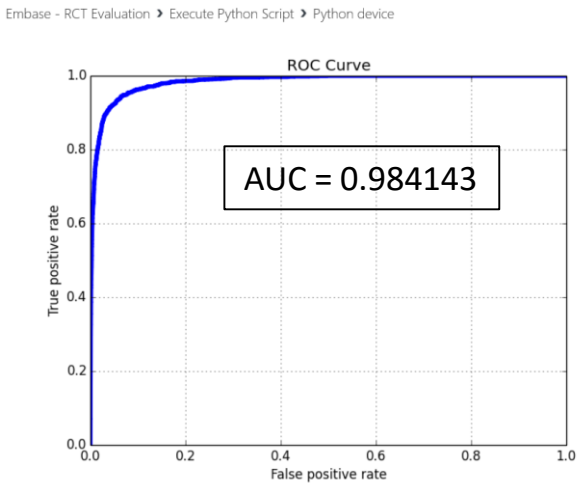
***Aims:*** **To test the utility and applicability of an automated clustering tool to describe and group research studies to improve the efficiency of scoping reviews.**

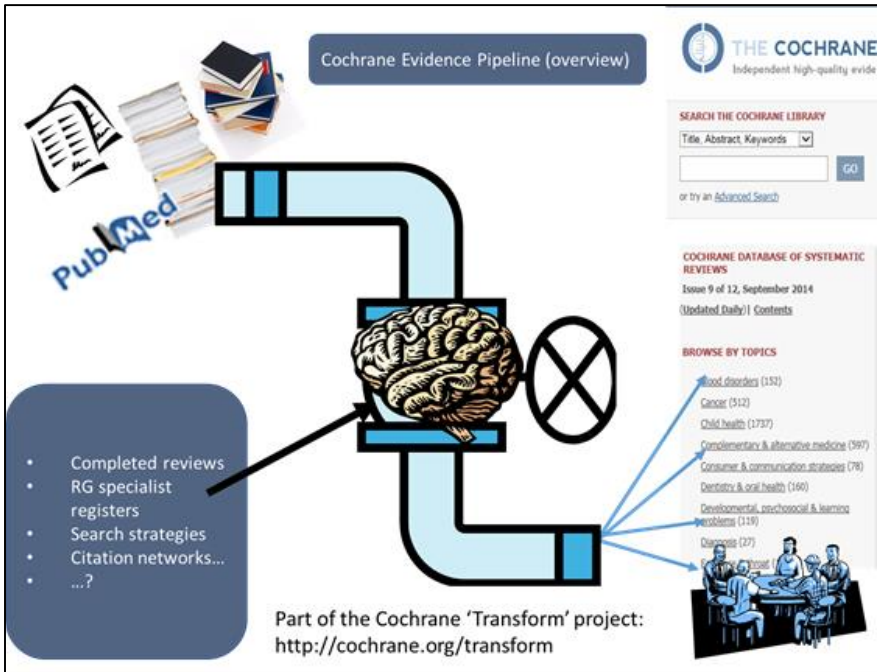
***Methods:*** **A retrospective study of two completed scoping reviews was conducted. This compared the**



# Database creation / curation

- If training data are available, it is possible to build a classification tool which can determine whether a given citation is within the scope of a database or not
- For simple categorisations – such as whether something is an RCT or not – performance is impressive
- The more data the better



# Risk of Bias assessment

- Emerging area; e.g.
  - RobotReviewer
  - Millard, Flach and Higgins
- Tools can accomplish two purposes:
  - Identify relevant text in the document
  - Automatically assess risk of bias
- Can perform very well on some dimensions of RoB

Int. J. Epidemiol. Advance Access published December 8, 2015

IJE

*International Journal of Epidemiology*, 2015, 1–12
doi: 10.1093/ije/dyv306
Original article

Original article

## Machine learning to assist risk-of-bias assessments in systematic reviews

**Louise A.C. Millard,[1,2,3]* Peter A. Flach[1,3] and Julian P.T. Higgins[1,2]**

[1]MRC Integrative Epidemiology Unit, [2]School of Social and Community Medicine and [3]Intelligent Systems Laboratory, University of Bristol, Bristol, UK

*Corresponding author. School of Social and Community Medicine, Oakfield House, Oakfield Grove, BS8 2BN. E-mail: louise.millard@bristol.ac.uk

Accepted 23 October 2015

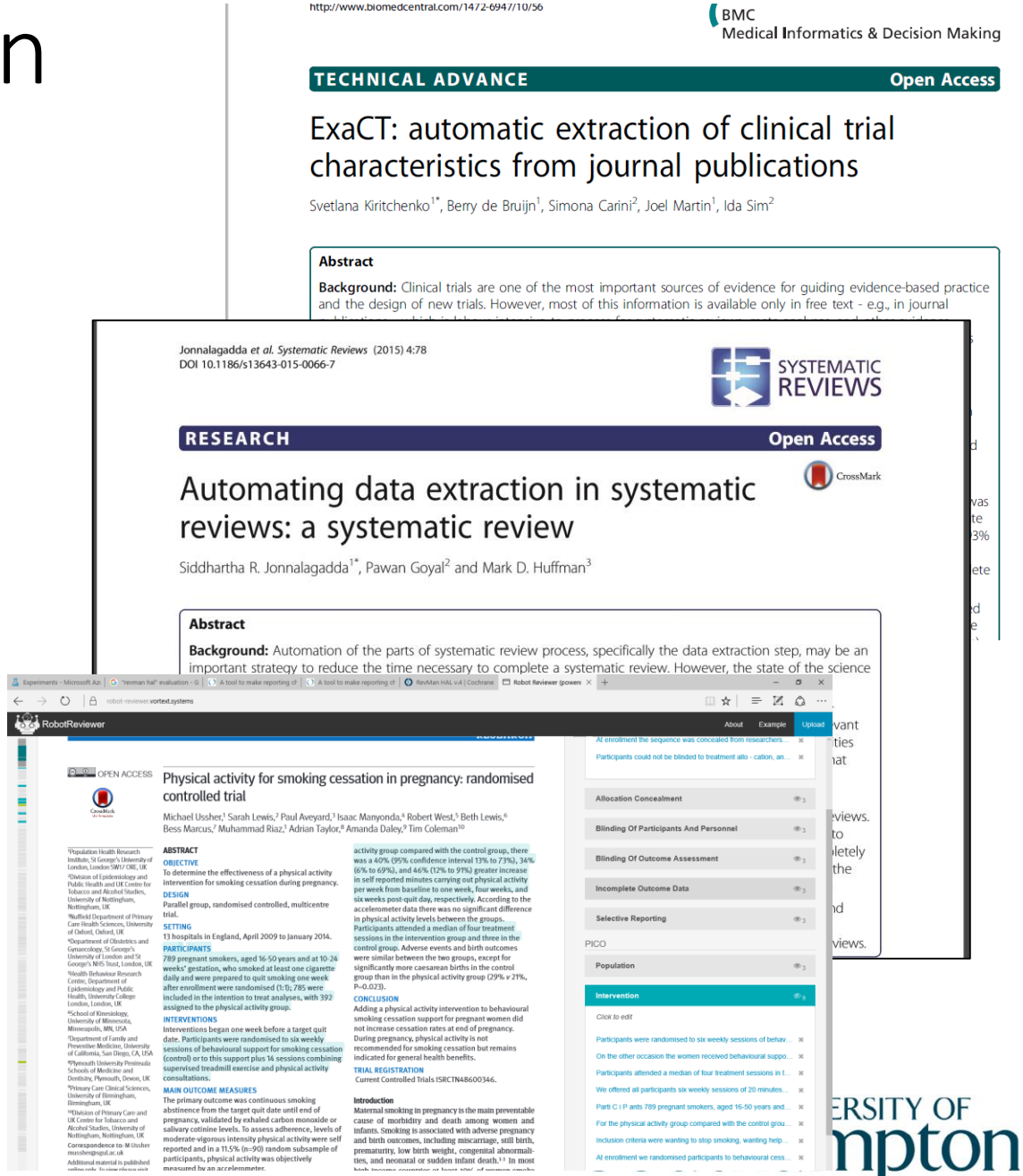
**Abstract**

**Background:** Risk-of-bias assessments are now a standard component of systematic



# Data extraction

- RobotReviewer can identify phrases relating to study PICO characteristics
- ExaCT extracts trial characteristics (e.g. eligibility criteria)
- Systematic review found that no unified framework yet exists
- More evaluative work is needed on larger datasets



http://www.biomedcentral.com/1472-6947/10/56

BMC Medical Informatics & Decision Making

TECHNICAL ADVANCE Open Access

ExaCT: automatic extraction of clinical trial characteristics from journal publications

Svetlana Kiritchenko[1*], Berry de Bruijn[1], Simona Carini[2], Joel Martin[1], Ida Sim[2]

Abstract

Background: Clinical trials are one of the most important sources of evidence for guiding evidence-based practice and the design of new trials. However, most of this information is available only in free text - e.g., in journal

Jonnalagadda et al. Systematic Reviews (2015) 4:78
DOI 10.1186/s13643-015-0066-7

SYSTEMATIC REVIEWS

RESEARCH Open Access

Automating data extraction in systematic reviews: a systematic review

Siddhartha R. Jonnalagadda[1*], Pawan Goyal[2] and Mark D. Huffman[3]

Abstract

Background: Automation of the parts of systematic review process, specifically the data extraction step, may be an important strategy to reduce the time necessary to complete a systematic review. However, the state of the science

RobotReviewer

OPEN ACCESS

Physical activity for smoking cessation in pregnancy: randomised controlled trial

Michael Ussher,[1] Sarah Lewis,[2] Paul Aveyard,[3] Isaac Manyonda,[4] Robert West,[5] Beth Lewis,[6] Bess Marcus,[7] Muhammad Riaz,[3] Adrian Taylor,[8] Amanda Daley,[9] Tim Coleman[10]

Allocation Concealment

Blinding Of Participants And Personnel

Blinding Of Outcome Assessment

Incomplete Outcome Data

Selective Reporting

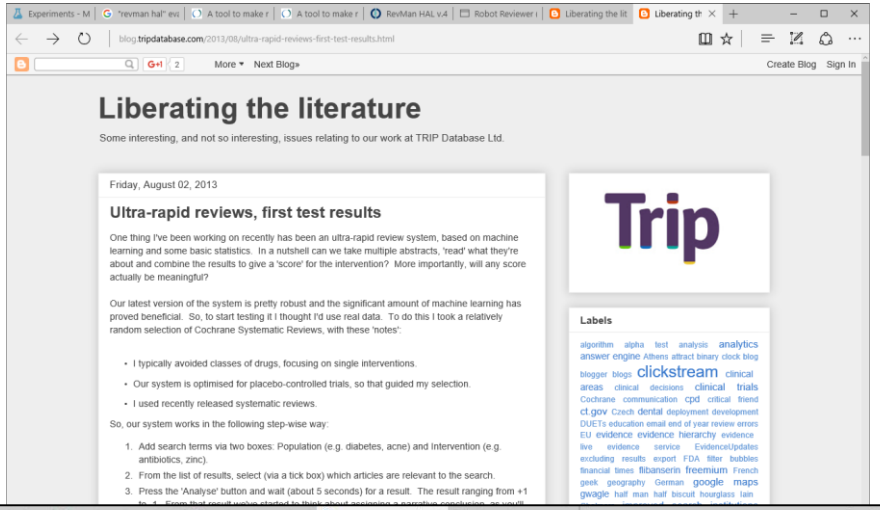
PICO

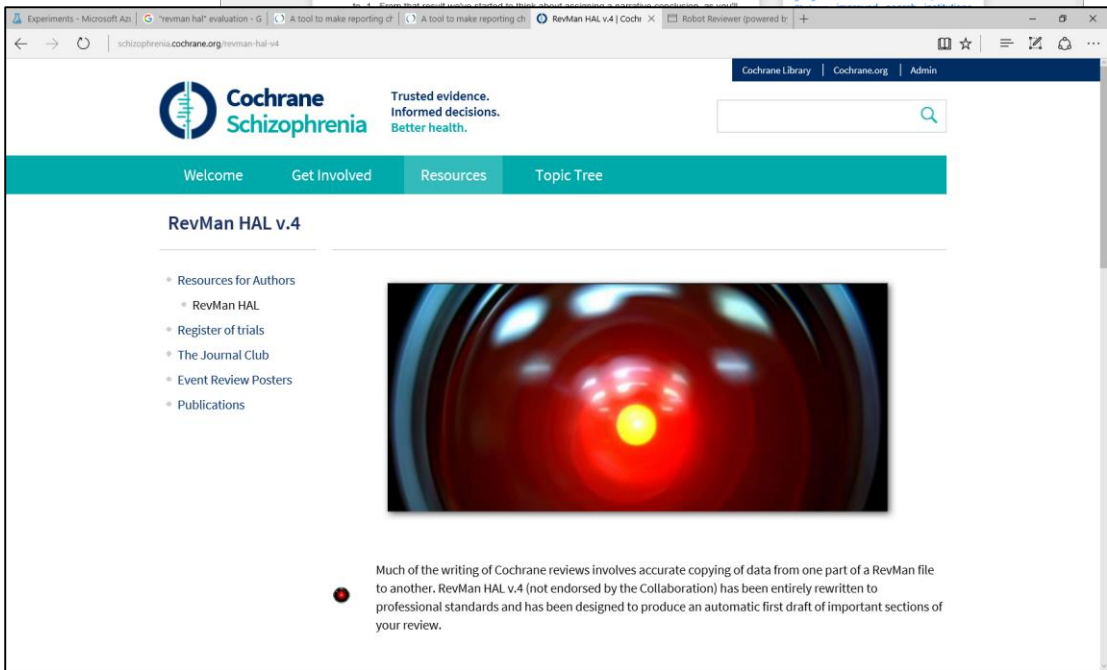
Population

Intervention

# Synthesis and conclusions

- Summarisation and synthesis of text is an active area for development in computer science
- Many hurdles to overcome before this technology can be used routinely
- Some systems automate parts of the process



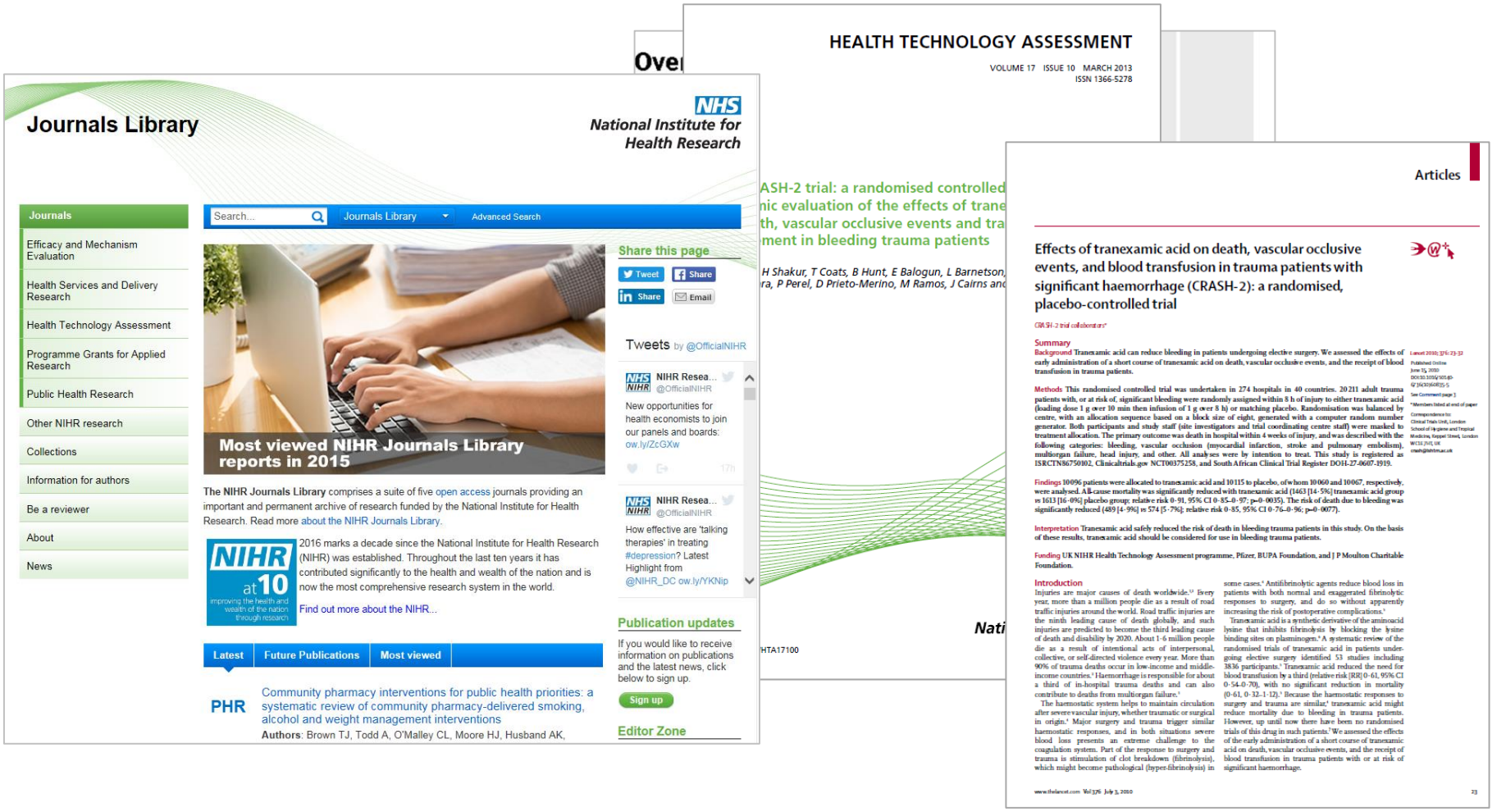
*Automated support for systematic reviewers: dream or reality?*

# Can publishers provide machine readable protocols and study results?

Cochrane UK & Ireland Symposium 2016

Elaine Williams, Director of Research Delivery and Impact,
NIHR Evaluation, Trials and Studies Coordinating Centre

# Publishing today



# NIHR Journals Library

- 5 open access journals – only health research funder with own journal series
- Builds on *Health Technology Assessment* journal
- Full reporting and permanent archive of research and other project information, **after project completion**
- Over 1,000 issues published - **£309m** research funding (November 2015)
- Academic primary audience
- HTA widely referenced in NICE Clinical Guidelines[1]

[1]Turner S, Bhurke S, Cook A. *Impact of NIHR HTA Programme funded research on NICE clinical guidelines: a retrospective cohort.* Health Research Policy and Systems (2015) 13:37.
http://www.health-policy-systems.com/content/13/1/37

# Features of NIHR journals

| | |
|---|---|
| Full description of research methods | ✓ |
| Full reporting of results - positive, neutral and negative | ✓ |
| Peer-reviewed and copy edited | ✓ |
| Reporting of patient and public involvement | ✓ |
| Published in an online open access journal | ✓ |

FIGURE 1 Consolidated Standards of Reporting Trials (CONSORT) flow diagram for all trial participants. a, Based on clinically intention

TABLE 22 Cost impact analysis of managing BSIs occurring with standard compared with antibiotic CVCs, with best- and worst-case scenarios for all PICUs in 2012[a] and hypothetical scenarios for a typical PICU with 350 admissions per year

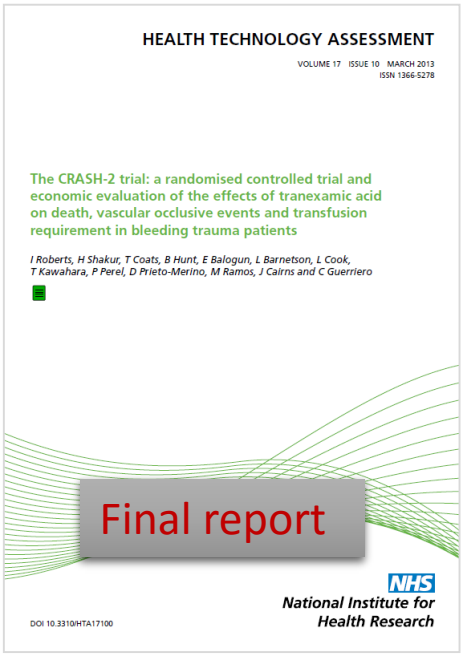
| Scenario | BSI rate using standard CVCs | Rate ratio | BSI rate using antibiotic CVCs | BSI with standard CVCs | BSI with antibiotic CVCs | BSI averted cost–impact | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Cost per BSI: –£2801 | Cost per BSI: £10,975 | Cost per BSI: £24,751 |
| Base case | 4.58 | 0.40 | 1.83 | 385.9 | 154.4 | 231.6 | 231.6 | 231.6 |
| | | | | | | –648,606 | **2,541,397** | 5,731,401 |
| Worst case | 4.42 | 0.97 | 4.29 | 372.5 | 361.3 | 11.2 | 11.2 | 11.2 |
| | | | | | | –31,297 | 122,631 | 276,559 |
| Best case | 4.74 | 0.17 | 0.81 | 399.4 | 67.9 | 331.5 | 331.5 | 331.5 |
| | | | | | | –928,583 | 3,638,415 | 8,205,414 |
| | 1 | 0.40 | 0.40 | 3.4 | 1.3 | 2.0 | 2.0 | 2.0 |
| | | | | | | –5,645 | 22,119 | 49,884 |
| | 2 | 0.40 | 0.80 | 6.7 | 2.7 | 4.0 | 4.0 | 4.0 |
| | | | | | | –11,290 | 44,238 | 99,767 |
| | 3 | 0.40 | 1.20 | 10.1 | 4.0 | 6.0 | 6.0 | 6.0 |
| | | | | | | –16,936 | 66,358 | 149,651 |
| | 4 | 0.40 | 1.60 | 13.4 | 5.4 | 8.1 | 8.1 | 8.1 |
| | | | | | | –22,581 | 88,477 | 199,534 |
| | 5 | 0.40 | 2.00 | 16.8 | 6.7 | 10.1 | 10.1 | 10.1 |

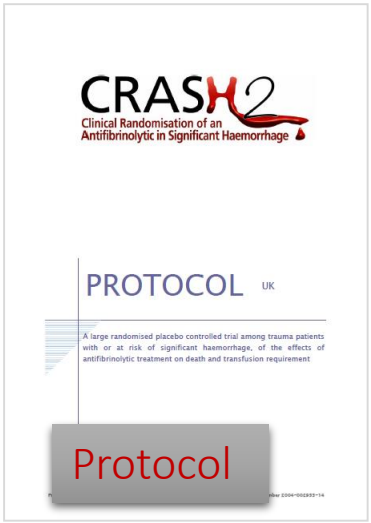
Harron K, Mok Q, Dwan K, Ridyard CH, Moitt T, Millar M, *et al.*CATheter Infections in CHildren (CATCH): a randomised controlled trial and economic evaluation comparing impregnated and standard central venous catheters in children. *Health Technol Assess* 2016;20(18)

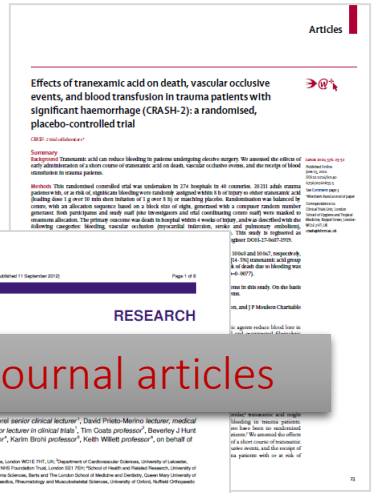
This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
▼<article xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchem
 type="research-article" dtd-version="3.0">
  ▼<front>
    ▼<journal-meta>
       <journal-id journal-id-type="publisher-id">hta</journal-id>
      ▼<journal-title-group>
         <journal-title>Health Technology Assessment</journal-title>
       </journal-title-group>
       <issn pub-type="ppub">1366-5278</issn>
       <issn pub-type="epub">2046-4924</issn>
      ▼<publisher>
         <publisher-name>National Institute for Health Research</publisher-name>
       </publisher>
     </journal-meta>
    ▼<article-meta>
       <article-id pub-id-type="publisher-id">08/13/47</article-id>
       <article-id pub-id-type="doi">10.3310/hta20180</article-id>
      ▼<title-group>
        ▼<article-title>
           CATheter Infections in CHildren (CATCH): a randomised controlled trial and economic evaluation comparing impregnated and standard ce
         </article-title>
         <alt-title alt-title-type="short">CATheter Infections in CHildren (CATCH)</alt-title>
       </title-group>
      ▼<contrib-group>
        ▼<contrib contrib-type="author">
          ▼<name name-style="western">
             <surname>Harron</surname>
             <given-names>Katie</given-names>
           </name>
           <xref ref-type="aff" rid="aff1">1</xref>
         </contrib>
        ▼<contrib contrib-type="author">
          ▼<name name-style="western">
             <surname>Mok</surname>
             <given-names>Quen</given-names>
           </name>
           <xref ref-type="aff" rid="aff2">2</xref>
         </contrib>
        ▼<contrib contrib-type="author">
```

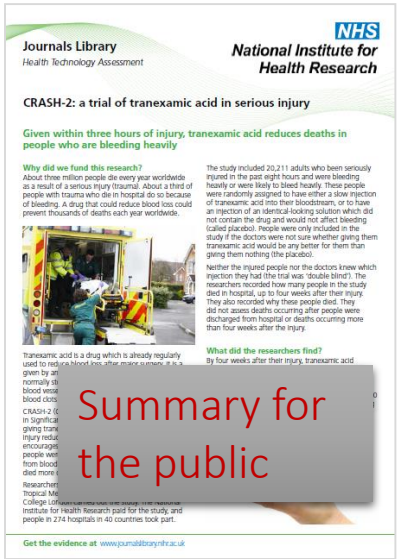
# Open access to more than the final report







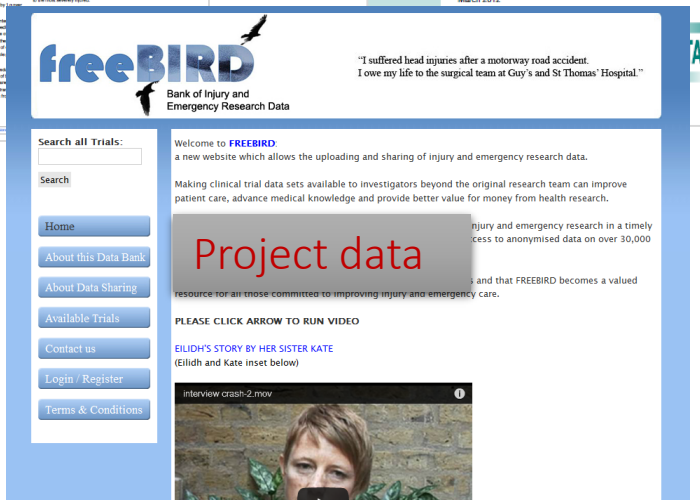



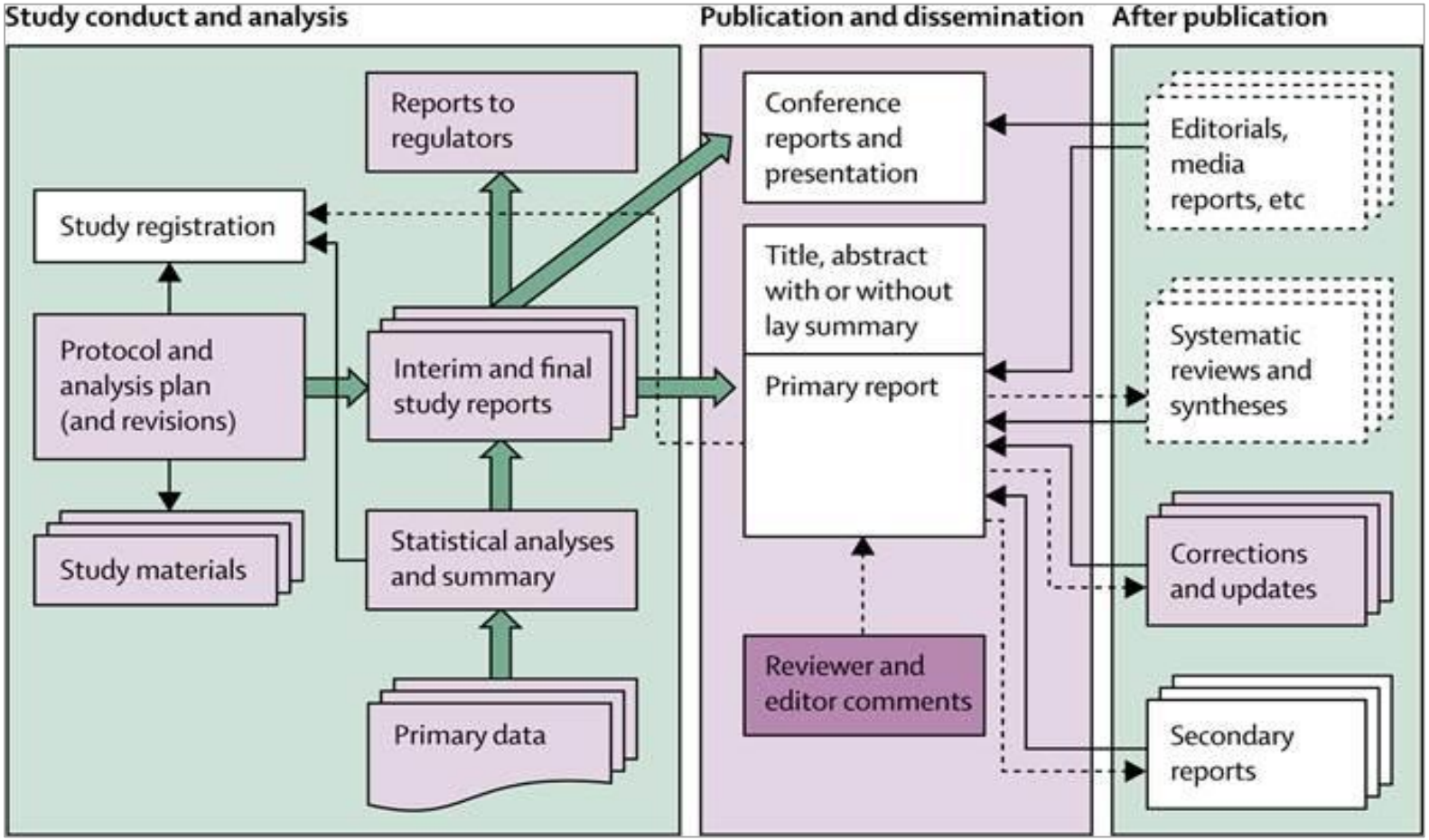
Table of contents | Abstract | Plain English summary | Scientific summary | Chapters
Important information | References | Appendices | Glossary | List of abbreviations

- Scientific Summary (PDF) (171.3 KB)
- Plain English Summary (PDF) (144.2 KB)
- Full Report (PDF) (19.9 MB)
- References (from this publication as a .ris file) (38.8 KB)
- Citation (for this publication as a .ris file) (595 Bytes)
- Protocol (PDF) (875.9 KB)

# The landscape is developing

- Greater focus on ‘avoidable waste’
- Open Access
- Dissemination and implementation
- Demonstrating impact
- Technology (eg XML)
- Data sharing

# Move to enhanced linking




*P Glasziou, Lancet* 2014; 383: 267–76

# Supporting systematic reviewers

*"Evidence-based medicine stipulates that all relevant evidence be used to make clinical decisions regardless of the implied resource demands"*[1]

- Quality in > Quality Out
- Reporting guidelines (*EQUATOR*) and associated tools (*eg Penelope*)
- Full text XML to support data mining
- Enhanced tagging
- References (.ris format)
- Access to data
- other?

[1] Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB: Evidence based Medicine: How to Teach and Practice EBM. Edinburgh: Churchill Livingstone; 2000.

# Automation of systematic reviews: the reviewer's viewpoint

*(…that's all very well, but how do these tools help me?)*

**Geoff Frampton**

*Southampton Health Technology Assessments Centre (SHTAC)*

**http://www.southampton.ac.uk/shtac**

# SHTAC: Who are we and what do we do?

- **A team of systematic reviewers and health economists**
- **We conduct systematic reviews (and maps) on a wide variety of health and social sciences topics (e.g. for NIHR, Cochrane Collaboration, WHO)**
- **We also critically appraise systematic reviews and economic analyses conducted by other parties, e.g. companies submitting evidence to NICE**

# Do we use automation for systematic reviews (SR) ?

- Depends on how “automation” is defined
- Yes, in bibliographic searching
  - running search strategies in databases or search engines
  - importing search results into reference management software
- Yes, within reference management software
  - identification of duplicate references
  - acquiring full-text documents
  - rule-based sorting (e.g. grouping) of references
- Not (yet) for other steps of systematic reviews (or maps)

# Our experiences

- **Bibliographic searching**

✓ – **Automation saves effort in searching and retrieving references**

**BUT…**

✗ – **Search functionality is not consistent across databases**

✗ – **Manual translation of search strategies is necessary for some databases**

✗ – **Reference import or download options are sometimes limited by quantity or completeness**

# Our experiences

- **Reference management software**

✓ – **Automation saves effort in organising references**

**BUT…**

✗ – **A proportion of references is often incomplete or incorrect**

✗ – **Duplicates are often missed**

✗ – **Full text documents are not always available or accessible**

# Where else in SR could automation help us?

- **Eligibility screening**
  - Especially if thousands of titles & abstracts require screening ✓

  **BUT…**

  - Might compromise recall (up to 5%?) ✗
  - Which tool(s) should we use?
  - Would automation replace one human reviewer?
  - Suitable for full-text screening?
  - Quality assurance process (reviewer agreement)?

# Where else in SR could automation help us?

- **Guide for data extraction?**
  - **Help reviewers to identify where relevant data are located in a report (but risk of over-reliance?)**
- **Guide for planning/formatting?**
  - **Auto-filling of relevant data fields in Protocol or Review report**
  - **Prompting for human input to ensure standardisation**

# Discussion points

- **Automation unlikely to be applicable to all *steps* of SR**
  - **Some steps require human judgement**
  - **SR need human inputs (e.g. stakeholder advisors to guide clinical interpretation and problem-spotting)**

# Discussion points

- **Automation unlikely to be applicable to all *steps* of SR**
  - Some steps require human judgement
  - SR need human inputs (e.g. stakeholder advisors to guide clinical interpretation and problem-spotting)
- **Automation unlikely to be applicable to all *types* of SR**
  - For some SR (e.g. complex interventions) even human reviewers find it challenging to locate and select evidence

**… automation could be valuable on a case-by-case basis**

**… may guide human reviewers on some SR steps**

# Wish list: what would we as reviewers like to see?

- **More efficient automation of searching and reference retrieval**

  – **Improved capability to interrogate multiple databases and search engines with the same search strategy**

  – **Improved quantity and completeness of references that can be imported into reference management software**

  – **Improved compatibility of databases and search engines with reference management software**

# Wish list: what would we as reviewers like to see?

- **More efficient reference management**
  - A tool to validate and update all references in a library to ensure completeness and accuracy (to also improve de-duplication)
- **Guidance on tools for automated eligibility screening**
  - Which tools are available?
  - Where to find them?
  - How to use them?
    - … training requirements for the operator?
    - … time and resources for machine learning processes?
  - Critical evaluation of strengths and weaknesses