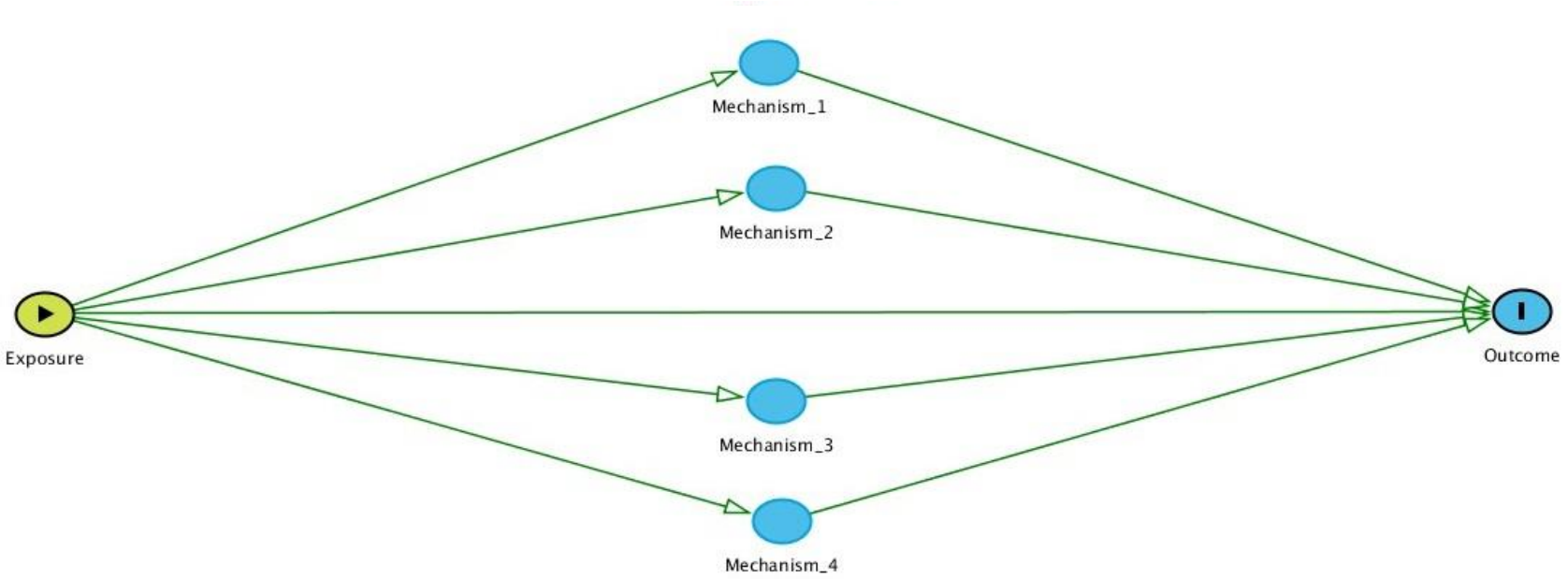
Text mining to identify potential mechanisms linking sedentary behaviour to disease outcomes

Brigid M. Lynch, PhD

Sedentary Behaviour Council
ISPAH satellite workshop
Cambridge
13 October 2018

Cancer Council Victoria

Understanding mechanisms

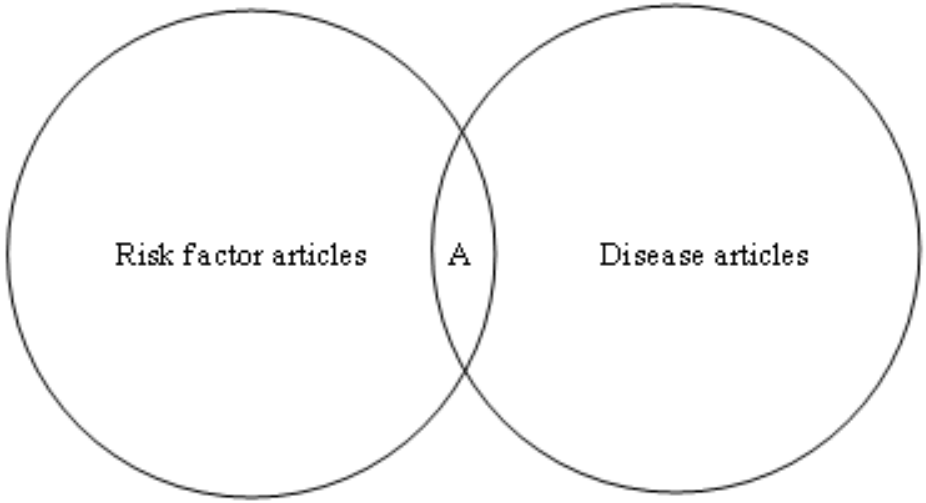


- Improve causal inference
- Easier/more effective to intervene on mediator
- Alter exposure in way to maximise mediation effect

Cancer Council Victoria

How to identify plausible mechanisms?

Search for common terms (potential mechanisms) in literature using PubMed®

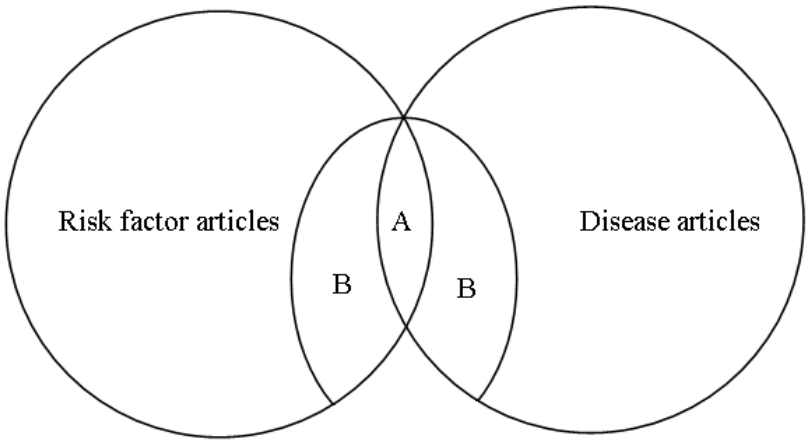


A overlapping terms identified by Boolean operator 'and'

Cancer Council Victoria

How to identify plausible mechanisms?

Search for common terms (potential mechanisms) in literature using text mining



A overlapping articles (with common terms)

B non-overlapping articles (with common terms) associated with both risk factor and disease

Cancer Council Victoria

Example - BDNF

- Protein expressed by hypothalamus
- Also excreted by skeletal muscle during exercise
- Exercise and cognitive function research (humans)
- Mouse models have shown upregulation of BDNF reduces tumour burden (various cancers)

Cancer Council Victoria

Text-mining tools

University of BRISTOL

MRC Integrative Epidemiology Unit

Current students Current staff Alumni

search

Integrative Cancer Epidemiology Programme

- Research themes
- **Research outputs**
  - ↳ Publications
  - ↳ Grants
  - ↳ **Data and analysis tools**
- People
- Collaborations
- Impact and engagement
- Contacts

Data and analysis tools

MR Base

MR Base enables online Mendelian randomization analysis using a comprehensive manually curated database of GWAS studies.

MELODI

MELODI is a literature mining platform to identify potential mechanistic pathways between exposures and disease outcomes.

"MR Base has transformed the ability of researchers world-wide to run rapid two-sample Mendelian randomization analyses, which can form the basis for in-depth analyses of many important questions regarding possible prevention or treatment of disease"

— PROF GEORGE DAVEY SMITH

TeMMPo

TeMMPo is a literature search tool to quantify the literature for specific disease mechanism.

LD Hub

LD Hub supports online LD score regression analysis using a comprehensive manually curated database of GWAS studies.

TeMMPo tool

- Text Mining for Mechanism Priorisation
- Uses Medical Subject Headings (MeSH) system
- Identifies co-ocurrence of MeSH headings in publications
  → Link ‘mechanism’ to exposure and/or outcome
- Targeted approach: specify *a priori* potential mechanism terms

https://www.temmpo.org.uk/

TeMMPo

TEXT MINING FOR MECHANISM PRIORITISATION

World Cancer Research Fund

University of BRISTOL

MRC Integrative Epidemiology Unit

Cancer Council Victoria

MELODI

International Journal of Epidemiology, 2018, 1–11
doi: 10.1093/ije/dyx251
Software Application Profile

Software Application Profile

MELODI: Mining Enriched Literature Objects to Derive Intermediates

Benjamin Elsworth,[1]* Karen Dawe,[1] Emma E Vincent,[1] Ryan Langdon,[1] Brigid M Lynch,[2–4] Richard M Martin,[1] Caroline Relton,[1] Julian P T Higgins[1] and Tom R Gaunt[1]

[1]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK, [2]Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, VIC, Australia, [3]Centre for Epidemiology and Biostatistics, University of Melbourne, Melbourne, VIC, Australia and [4]Physical Activity Laboratory, Baker Heart and Diabetes Institute, Melbourne, VIC, Australia

*Corresponding author. MRC Integrative Epidemiology Unit (IEU), Bristol Medical School: Population Health Science, Oakfield House, Oakfield Grove, University of Bristol, Bristol BS8 2BN, UK. E-mail: ben.elsworth@bristol.ac.uk

Editorial decision 2 November 2017; Accepted 3 January 2018

Cancer Council Victoria

MELODI

- A text mining platform designed to identify and prioritise intermediates between two datasets.
- Uses PubMed to create datasets (exposure and outcome).
- MELODI looks for overlapping
  - MeSH terms
  - Single SemMedDB concepts
  - SemMedDB ‘triples’ (subject – predicate – object) datasets.
- Enrichment step – compares frequency of terms within datasets to background frequency in PubMed.

Cancer Council Victoria

MELODI

- Enrichment using OR and two-tailed FET.
- P-values corrected for multiple testing using the Benjamini/Hochberg (non-negative) correction with a cutoff of $p<1e-5$.
- The results (corrected p-value and OR) used to filter results.
- Also uses frequency of predicate term (SemMedDB Triple) and minimum position in the MeSH hierarchy (MesH method).

Cancer Council Victoria

MELODI

- Visualisation of results via Sankey plot.
- SemMedDB Triples method also shows directed network diagram.
- These graphs are ‘live’ and change as filters applied.
- Once spurious intermediates removed and filters applied need to manually curate results.

Cancer Council Victoria

MELODI

Search results are retained if they represent a biological marker that could plausibly be associated with both the exposure and outcome. Two independent extractions!

Exclusion criteria for intermediates:

- synonyms, antonyms or similar terms for the exposure or outcome
- scientific methods, diagnostic tests, therapies (including drugs), anatomical or physiological nomenclature, comorbid conditions.

Cancer Council Victoria

Example application

Identifying known and novel mechanisms underpinning association between sedentary behaviour and breast cancer risk

Cancer Council Victoria

Strengths and limitations

- Strengths:
  - Sophisticated text-mining techniques
  - Mechanism discovery
  - Combining literature from different fields

- Limitations:
  - Co-occurrence in article → not necessarily association
  - Multiple intermediates in pathway:
    Not really possible to combine

Cancer Council Victoria

Acknowledgements

University of Bristol, UK

Sarah Lewis
Richard Martin
Tom Gaunt
Benjamin Elsworth
Luke Robles
Karen Dawe
Julian Higgins

Cancer Council Victoria

Funding

NBCF Career Development Fellowship

2015 – 2018

UICC Yamagiwa-Yoshida Memorial International Cancer Study Grant

Jan – Mar 2017

Thank you

brigid.lynch@cancervic.org.au

@drbrigidmlynch

http://www.bristol.ac.uk/integrative-epidemiology/faciliitiesresources/software/

Cancer Council Victoria

Further enquiry