SPRU - SCIENCE POLICY
RESEARCH UNIT

To what extent does biomedical research change direction? Comparing research for diseases of the poor and research for diseases of the rich

Presentation for DISCUS seminar series
Josie Coburn, 24th July 2021

UNIVERSITY OF SUSSEX | BUSINESS SCHOOL

EFMD EQUIS ACCREDITED

ASSOCIATION OF MBAS ACCREDITED

The problem

Perennial science policy debate

- Science is inherently uncertain and cannot be directed (Bush, 1945; Meyers 2007) vs.
- Science is a public good and should address societal needs (Polanyi, 1962; Sarewitz, 1996)

Misalignments and inequalities in disease-related research

- Research funding ≠ Disease burden (e.g. Ràfols and Yegros, 2017)
- 'Diseases of the poor' research funding < 'Diseases of the rich' research funding

Increase in targeted funding

- Including to diseases of the poor (Chapman et al., 2017, 2018)
- But neglected diseases are still underfunded, and

But…

Research is given directions irrespectively of intentions

- Institutions (in the sociological sense: norms, values, etc.) direct research outcomes

Research changes direction relatively often

- From grants and funding institutes to publications, clinical trials, patents, and drugs (Sampat, 2015)

Serendipity = unexpected beneficial discoveries (Bush, 1945; Meyers, 2007)

- But serendipity in research not yet well understood (Sampat, 2015; Yaqub, 2018)

Research questions

How and why does publicly-funded biomedical research change direction or remain on target, comparing research for diseases of the poor and research for diseases of the rich?

Today's question: How does health research change direction?

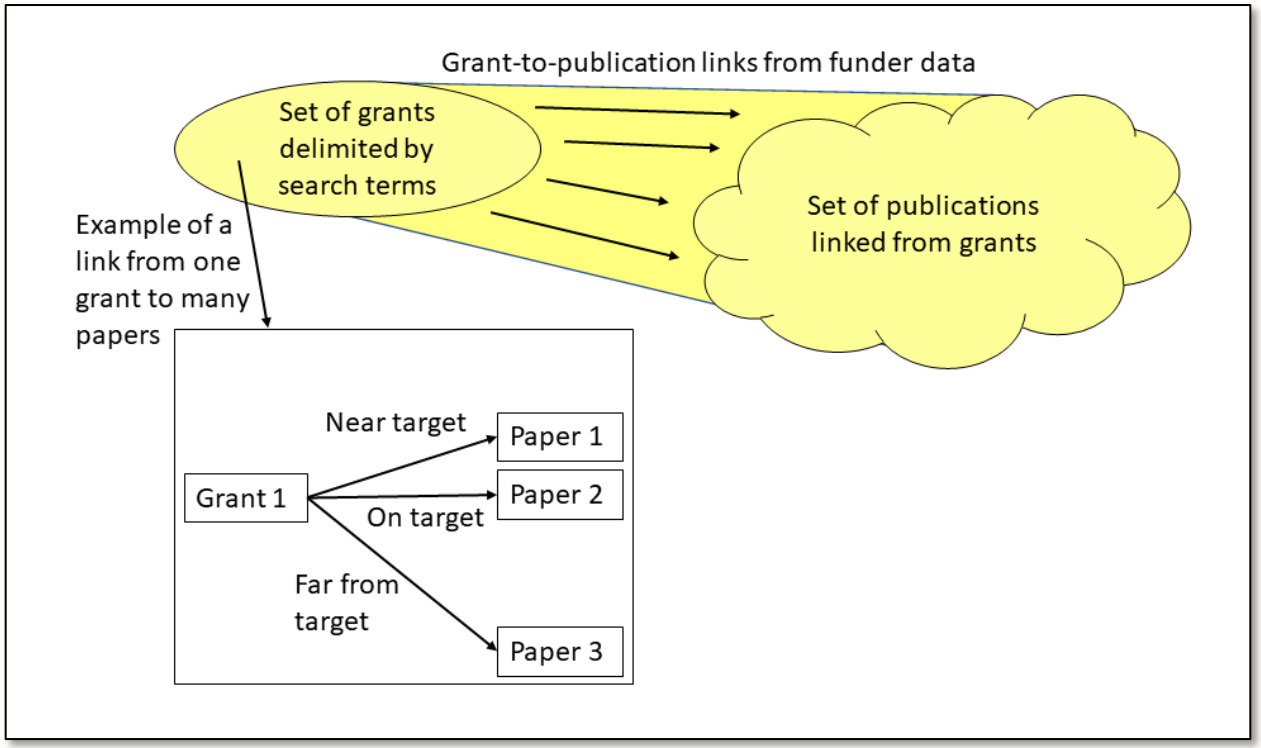
1. To what extent does research targeting a particular disease (grants) produce progress for the disease being targeted (publications)?
2. To what extent does progress in disease-related research (publications) rely on research targeted to that disease (grants)?
3. What similarities and differences are there comparing research for diseases of the poor and research for diseases of the rich?

Future questions:

- Why does research change direction?
- To what extent can research changing direction be called serendipity?
- What are the implications for science policy?

Measuring how research changes direction

- Map the relationships between grants and publications for a selection of diseases
- Identify the extent to which research changes direction
  - Link grants to publications (funder data) and publications to grants (acknowledgements)
  - To what extent do the topics match?



Selecting diseases to compare

- These diseases vary by disease burden, geographical spread, research funding levels and causes (e.g. pathogens, vectors, genetic and environmental factors)

Disease	Chagas disease	Malaria	Ischaemic heart disease	Breast cancer
Deaths	7,853	619,827	8,930,369	611,625
Total DALYs worldwide (DALYs per 100,000 inhabitants)	2.73	484.24	1,809.28	178.55
- **LMIC**	4.65	830.46	2,635.85	224.85
- **HIC**	0.29	2.87	2,579.20	440.94
DALYs ratio (DALYS per person LIC+MIC) / (DALYS per person HIC) (Røttingen et al 2013)	1,869.10	185.10	1.18	0.45
Updated DALYs ratio (Yegros-Yegros et al 2020)	15.81	289.36	1.02	0.51
Disease type (Yegros-Yegros et al 2020)	2	3	1b	1a
R&D Funding (RCDC US NIH for 2019, $ million)	11	196	421	709
Grant funding proposals (2004-2018)	769	5,800	7,387	24,659
Publications (2009-2018)	4,994	22,778	82,149	93,820

Data sources

- There are lots of great data sources out there
- Many now have APIs
- And you can also do web-scraping to gather more / different data

Grant proposal data sources	Publication data sources
Digital Science Dimensions (database of grants, publications, clinical trials, patents and policy documents), NIH Research Portfolio Online Reporting Tools (RePORTer) for US biomedical research data, Bill and Melinda Gates Foundation, US National Science Foundation, Gateway to Research for UK data, Wellcome Trust, EU Cordis for EU data, and many other public funding bodies such as Canadian Institutes of Health Research.	Clarivate Analytics Web of Science, Elsevier Scopus, Digital Science Dimensions, PubMed, Centre for Agriculture and Bioscience International (CABI), and 1science for malaria-related publications

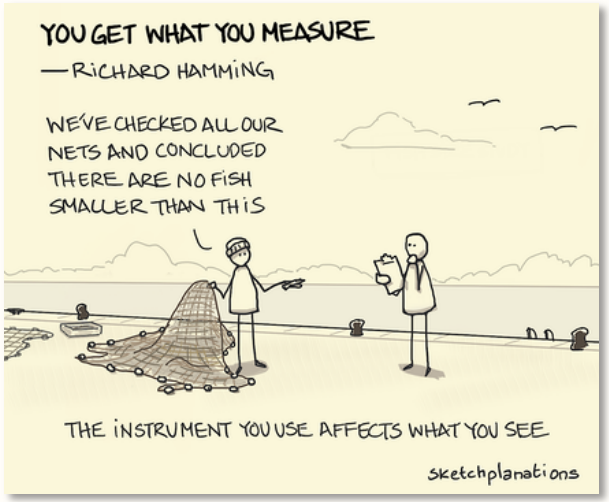
- I have collected grant data and publication data for the 4 diseases
- Grants were delineated using search terms
- Publications were delineated using Medical Subject Headings (MeSH)
- I have linked them using funder data and publication acknowledgements
- I have used an NIH tool (NIH National Library of Medicine Medical Text Indexer to assign MeSH to grants based on titles and abstracts
- Publications are also categorised by MeSH terms in PubMed

But beware of data biases

Don't just look under the lamppost

Pay attention to delineation

Don't compare apples and oranges



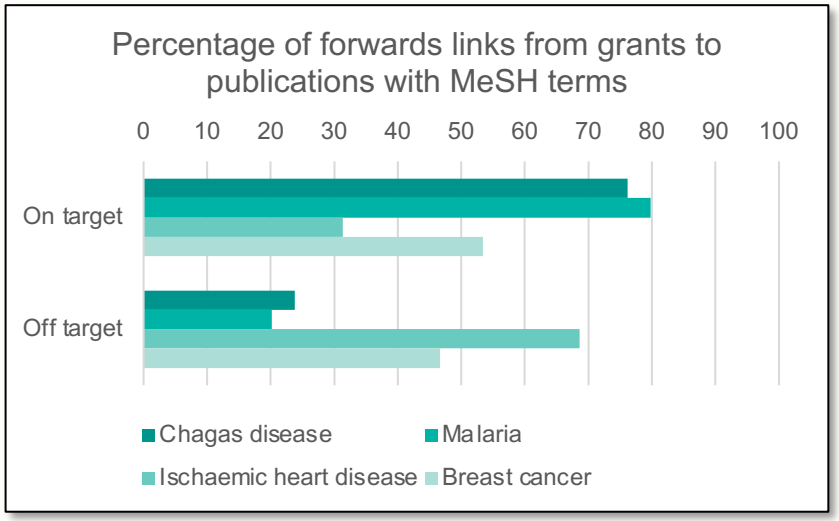
You get what you measure and **measuring has consequences**

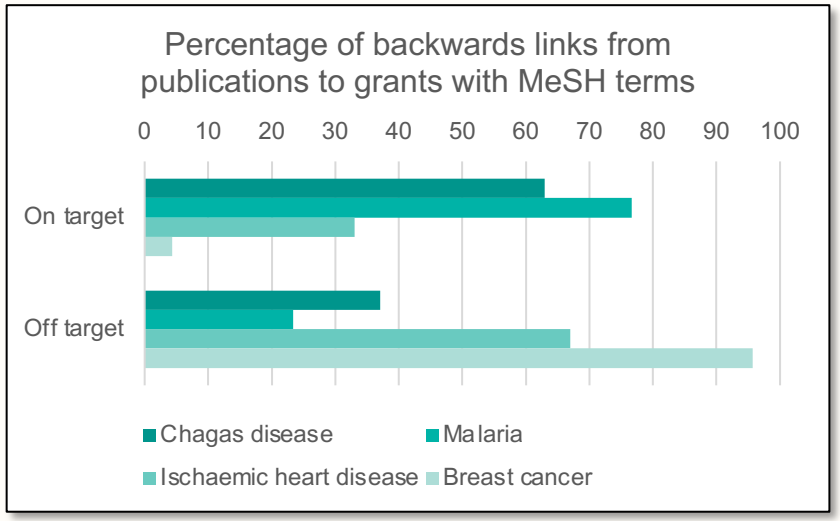
One solution: use different lenses

Use different lenses

- Theoretical
- Methodological
- Analytical

Method and results: Binary measure

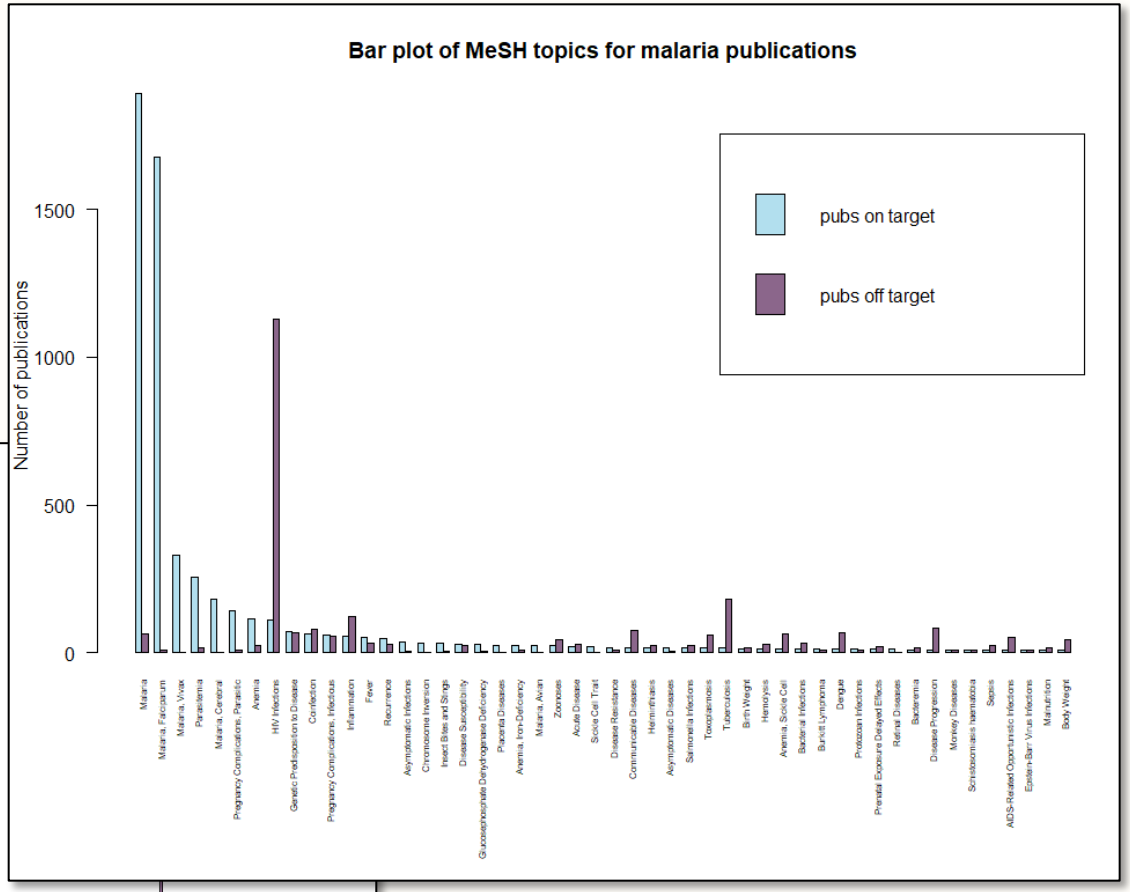


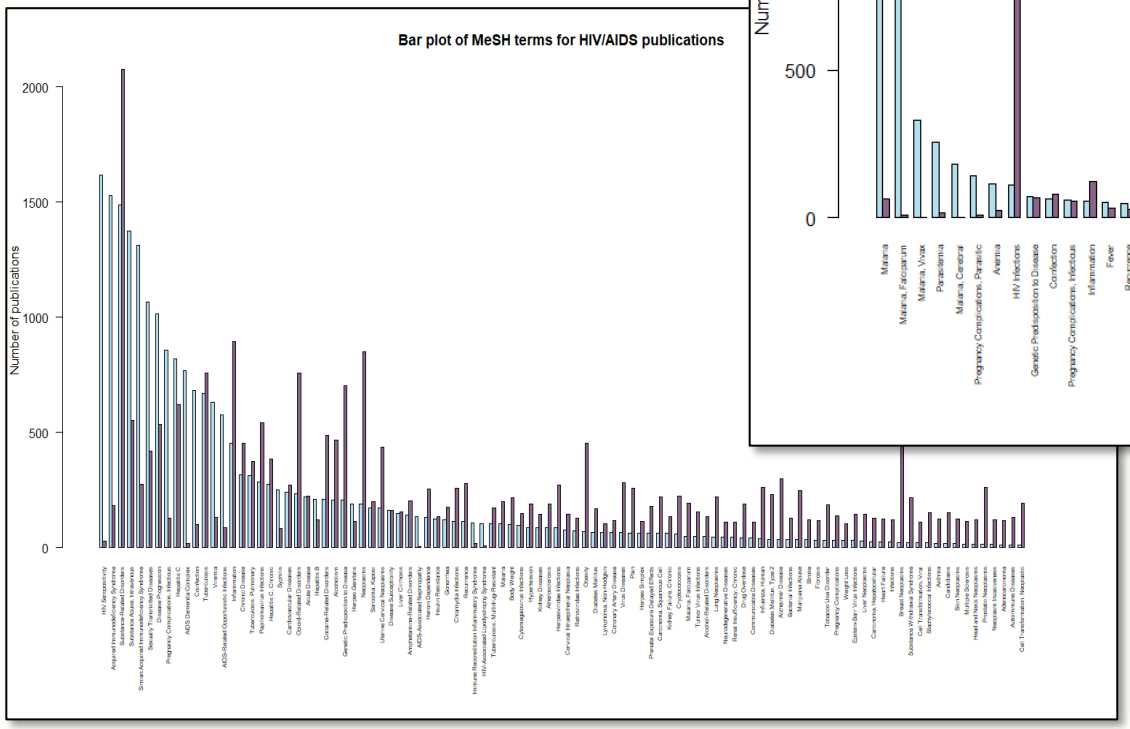


- These charts are based on a **binary measure**: – do the Medical Subject Headings (MeSH terms) match or not? (e.g. Malaria, Malaria falciparum, Malaria vivax)
- There are proportionally **less changes in direction in research for diseases of the poor** (malaria and Chagas disease) **than in research for diseases of the rich** (breast cancer and ischaemic heart disease) in both directions
- A **higher proportion of off target malaria research 'donates' to publications** on other diseases whereas a **higher proportion of off target breast cancer research 'free-rides' on grants** from other diseases

Method and results: MeSH term bar plots

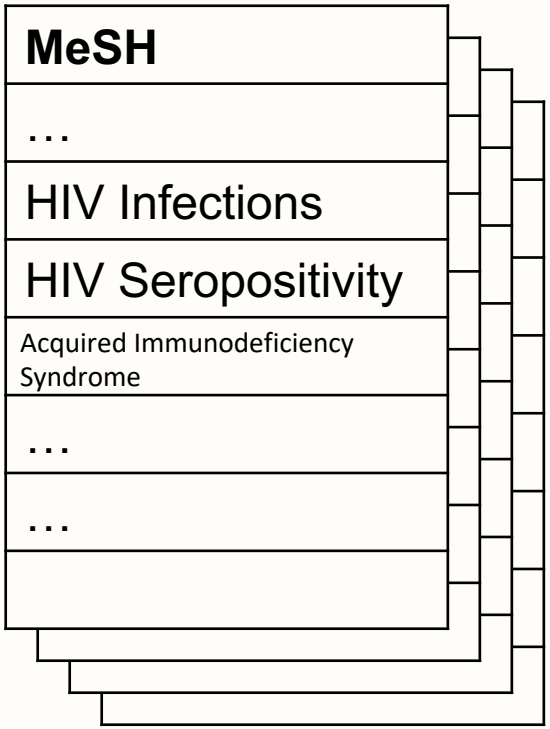
- These charts show that the MeSH terms associated with the **on target** publications are different to the MeSH terms associated with the **off target** publications





- Publications were split into **on target** and **off target** based on **MeSH terms**

Method and results: Cosine similarity



On target reference vector made from 50 most frequent MeSH in on target publications

$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

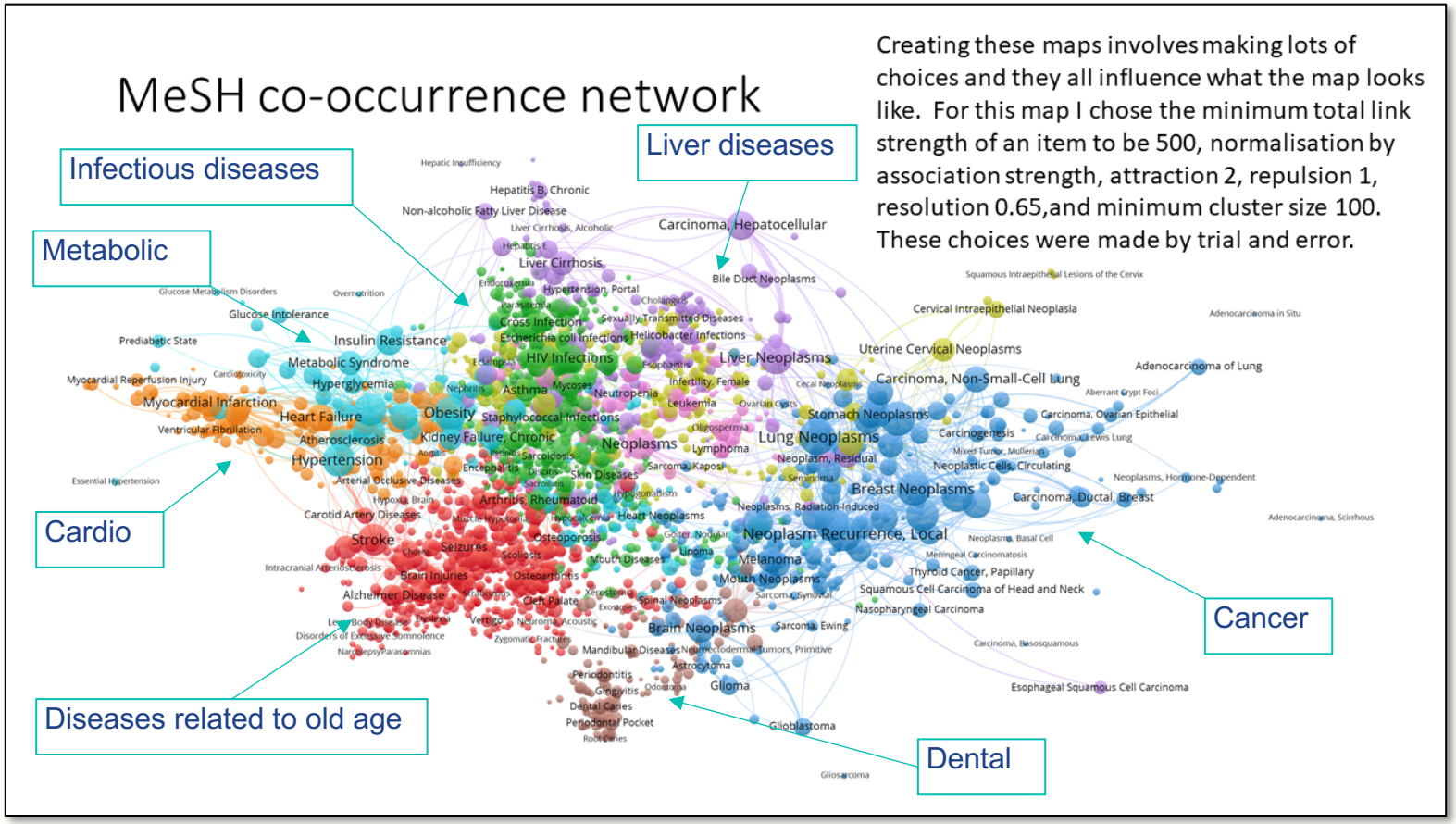
MeSH
…
HIV Infections
Tuberculosis
Diabetes
…
…

Off target publication vector

For Terry Pratchett fans out there – for HIV **the answer is 0.42!**

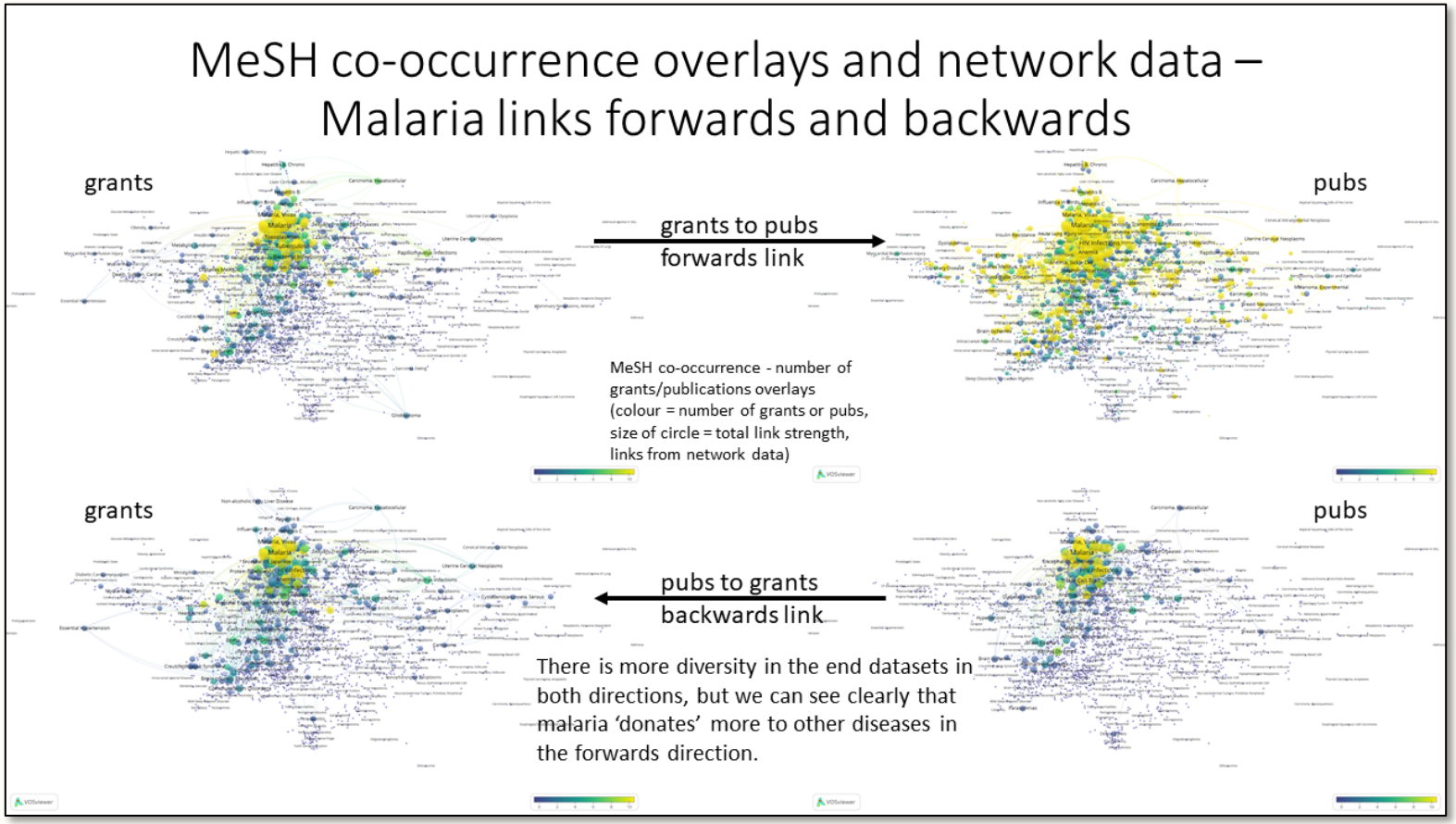
But we can also measure **how off target each publication is** by comparing its MeSH to the MeSH in the reference vector using **cosine similarity**

Method and results: Mapping MeSH co-occurrence



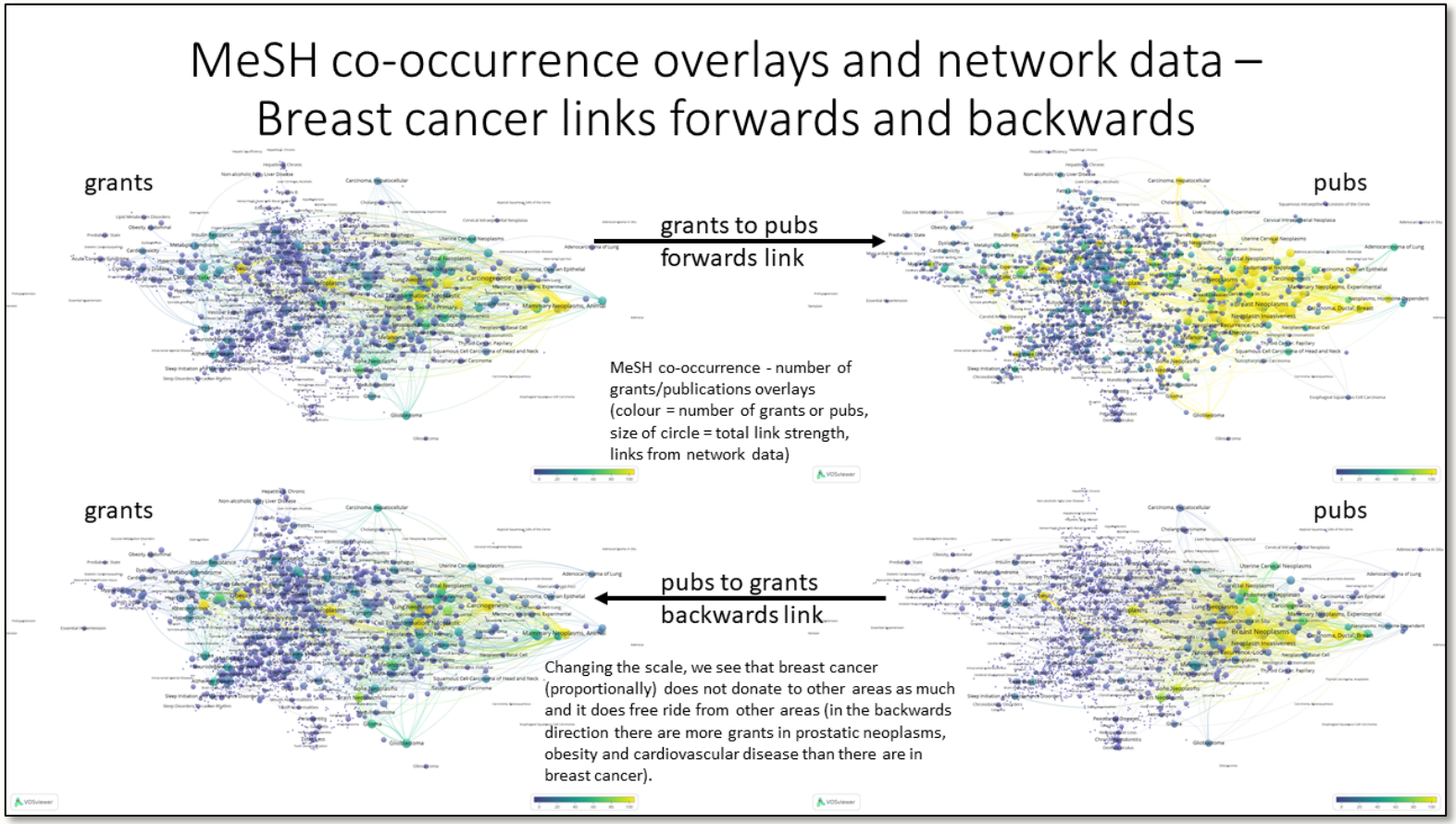
- Map based on all MeSH terms associated with all publications from PubMed and their co-occurrence within each publication
- Colour = cluster, size of circle = total link strength

Method and results: Overlay mapping - malaria



- Colour = number of publications
- Size of circle = total link strength

Method and results: Overlay mapping – breast cancer



- Colour = number of publications
- Size of circle = total link strength

Method and results: Knowledge flow

Disease	Flow forwards	Flow forwards normalised by number of forwards links (donate)	Flow backwards	Flow backwards normalised by number of backwards links (free-ride)	Ratio of forwards to backwards (norm flow forwards / norm flow backwards) (donate / free-ride)
Chagas	187.72	0.25	51.68	0.21	1.19
Malaria	2,041.63	0.21	280.37	0.11	1.91
Ischaemic heart disease	6,552.42	0.39	2071.05	0.47	0.83
Breast cancer	13,005.75	0.49	4520.75	0.64	0.77

- We have a category of concern, disease **alpha**. Then we look for the spread/**knowledge flow (kf)** of the **alpha grants** across categories **as we move to publications**. If disease **alpha** was only in category **alpha (j)**, then the spread would be given by the vector **flow (f_{ji})** of grants in category j, publishing in category i, weighted by the distance from j to i, written as follows:
- ***$kf_j = \sum$ i of flows f_{ji} from j to i times the distance between j and i (d_{ji})***
- In order to have a comparative measure, we normalize the flows so that their sum is 1.
- Now, it turns out that there are co-occurrences – which means that flow does not start from only a single j category, but from other categories as well: a distribution $\mathbf{p_i}$ of the grant across categories (diseases). Now we can look at the spread from each of these categories (where before it was only j) as follows:
- ***$kf = \sum j\ p_j \sum$ i of flows f_{ji} from j to i times the distance between j and i (d_{ji})***

Observations/lenses/open questions

- **Key findings so far**
- For all diseases, in the forwards direction (grants to pubs), publications have more diverse MeSH associated with them than grants. In the backwards direction (pubs to grants), grants have more diverse MeSH associated with them than publications. Thus, **research does change direction**
- For all diseases, in the backwards direction, the difference between the patterns for publications and grants is less than in the forwards direction – this appears to show that **all diseases 'donate' more than they 'free ride'**
- A **higher proportion of off target malaria research 'donates' to publications** on other diseases whereas a **higher proportion of off target breast cancer research 'free-rides' on grants** from other diseases
- **Why? And with what implications? What would you want to ask researchers and research funders in interviews?**

- **On using different lenses**
- I know how many **data-related and methodological choices** I have made along the way, and I'd like to use **different lenses** to examine **how the answers we get depend on these choices**
- **What do you think about that?**
- **How would you do it systematically?**

Thank you!

- **Any questions?**