

What's Missing?

Diseases Correlation Ignored by the Medical Research

<https://blownhither.github.io/>

Diseases are correlated - some you can easily think of, some you cannot! Arthritis is recently found to increase the risk of diabetes. Such relations can be too obscure to be found by tuition, even for researchers and medical practitioners.

- Disease/habit index covariance coefficient**

We analyze clinical correlation of diseases/habits with covariance coefficient

e.g. $X = \text{bottles of alcoholic drinks}$, $Y = \text{diagnosed with diabetes}$

Data Source	Size	Attributes
BRFFS The Behavioral Risk Factor Surveillance System	450K	Drinking, diabetes, sleep...

- Keyword co-occurrence in medical research / media coverage**

We analyze coverage of correlated diseases/habits in papers by with portion of papers with both keywords in title.

$$\frac{\#("Alcohol" \cap "Diabetes")}{\#("Alcohol" \cup "Diabetes")}$$

Data Source	Size	Attributes
PubMed medical research	29M	Paper title, abstract, ...
Google News media coverage	...	

- Potential “missing” researches**

$\approx$ highly-related disease/habits that are rarely studied on!

- Next step**

- Grow in size and diversity beyond the current 450k-row BRFFS data
- Leverage NLP to extract semantic-level correlation in Google News and PubMed
- More sophisticated metric for measuring discrepancy

Diseases v.s. Life Habits

correlations ignored in the research

