WORKING PAPER

Time Lags

A Feasibility Study

FLAVIA TSANG, JONATHAN GRANT,
STEVEN WOODING

RAND EUROPE

Executive summary

This memo describes the feasibility of using data from Alberta Innovate Health Solutions (AIHS) to study time lags between research expenditure and the eventual health benefits. Our key finding is that existing data do not form a sufficient basis for an empirical study of time lags. A large scale data matching and/or collection exercise would be required.

The correct unit of analysis to examine time lags is the grant

Because examining time lags necessarily involves looking the time delay between different activities and events, we need to analyse events with relatively short time spans – hence we need to focus on grants and not individual careers.

The ISHTAR data set does not capture the necessary data to examine lags without external data sources

We found that the ISHTAR data base is limited:

- research publications – probably the main research output - are not captured
- patent information is effectively absent – it is only available for 4 years out of the 30 years covered in the database. Even this data does not appear to be complete or consistently recorded.
- medical subject headings (MeSH) in the database are linked to individuals, not to grants

Identifying links between research events (grants, papers, and patents) is a vital step towards studying time lags, as the links allow us to trace the development of research ideas. Unfortunately, these links are virtually absent from the ISHTAR database – only a small number of grant-patent links are recorded.

Additional data on research outputs/outcomes is available in external databases, but is hard to link to the grants in the ISHTAR database

External databases such as ISI web of knowledge and the United States Patent and Trademark Office (USPTO) can provide paper-to-paper, patent-to-paper and patent-to-patent linkages. However, the ISHTAR database does not provide the information to easily link papers and patents identified in these databases to specific grants.

Citation databases contain information in linking some papers to funders through 'funding acknowledgements' however in the case of AIHS these acknowledgements do not normally

mention specific grant numbers. Because ISHTAR database provides MeSH headings for researchers, and not grants, this information cannot be used to help link papers to grants.

Recommendations for prospective data collection

We recommend the AIHS do the following:

- AIHS should ensure that grant recipients cite grant IDs in funding acknowledgements
- AlHS should also collect information about what papers arose from each grant directly from grant recipients
- AIHS should collect information on patents that arise from grants directly from grant recipients

Recommendations for retrospective data improvement

AIHS should explore methods of improving their retrospective data set linking grants to papers, patents and other impacts. It would also be helpful if grants could be linked to follow on grants. Two possible approaches are:

- Web based survey of previous grant recipients – possibly providing a list of their papers/patents/grants and asking them to link them.
- Use of text/data mining techniques to link grants, papers and patents – this would need to be tested against matching by researchers.

Sources of data that might allow downstream impacts to be added to the dataset should also be investigated. These might include citations on clinical guidelines and information about patent licensing and exploitation.

Development of visualisations was hampered by lack of data

Because of data limitations, we were forced to carry out the proof of concept analysis on only one researcher, and make a number of strong assumptions to allow us to link data at the grant level. Our key conclusion is that in order to effectively explore visualisation techniques that would be appropriate, and informative, for use with a large dataset, we need a large dataset to test them on. For small datasets it is possible to produce layouts by hand; however, this would not be feasible for large datasets.

However by working with our small data set we concluded that we would be likely to need a mathematical algorithm which optimizes the graphical layout by varying the vertical positions of the research events while keeping the horizontal positions constrained to a timeline. So far we have not been able to identify an existing software with such an algorithm.