

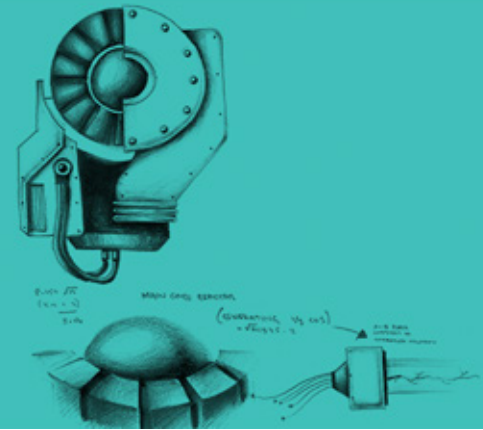
BIBLIOMETRIC ANALYSIS OF NZ RESEARCH PERFORMANCE

An Executive Summary of Motu Working Paper 18-xx

Adam Jaffe and Kate Preston

Motu Economic and Public Policy Research

adam.jaffe@motu.org.nz and kate.preston@motu.org.nz



SUMMARY HAIKU

Bibliometrics.
Look at New Zealand science.
Good but not stellar.

INTRODUCTION

Bibliometrics is the use of data on research publications as indicators of the research output of researchers, research institutions or geographic/political areas. Whenever a metric or statistic is constructed, choices have to be made as to which data to use and what procedures to follow in constructing the metric. The use of bibliometrics to evaluate research endeavours is relatively new and has not been standardised. There is an academic literature on the issues in constructing bibliometric measures, but there is no widely accepted international agreement as to which issues of measurement are most important. Disagreement also prevails over whether it is appropriate to use confidence intervals on bibliometric data, which can be considered a population rather than a sample.

Our research involved stress-testing bibliometric measures of New Zealand (NZ) publication output across disciplines and over time to elucidate the consequences of different choices. Though the focus of this report is on NZ's performance as a nation, many of the insights will also be relevant to the bibliometric assessment of individual researchers, departments, and institutions. In addition, we discuss the arguments for and against using confidence intervals for the bibliometric measures and present bootstrapped confidence intervals around some of our results.

The measures we used to evaluate the quantity and impact of NZ's research are described in Table 1 below. They were calculated from Scopus Custom Data 2002-2015, a large database of abstracts and citations of peer-reviewed literature.

Table 1: Bibliometric Measures of NZ's Research Performance

Quantity of Research	
Total Publications	The number of publications with a NZ affiliation
Impact of Research	
Average MNCS	Normalised citations per NZ publication, interpreted as above the world average if greater than 1.
Total MNCS	Combines both the amount of research output (publications) and its average impact (citations per publication).
Fraction in top percentiles	The fraction of NZ publications in the top percentiles of the citation distribution of similar publications.

As there are multiple choices of calculation mode to make, it is not practical to explore all possible combinations of choices. Our approach was to explore each choice in isolation relative to a set of baseline results.

The issues that we explored include:

- Types of publications to include
- Normalisation of citations by publication
- Comparing total output and impact to a benchmark
- Attributing internationally co-authored publications to countries
- Normalising citations of publications in multiple fields
- The effect of the integer nature of citations
- Measuring uncertainty
- An alternative assignment of publications to fields
- Consolidation of fields

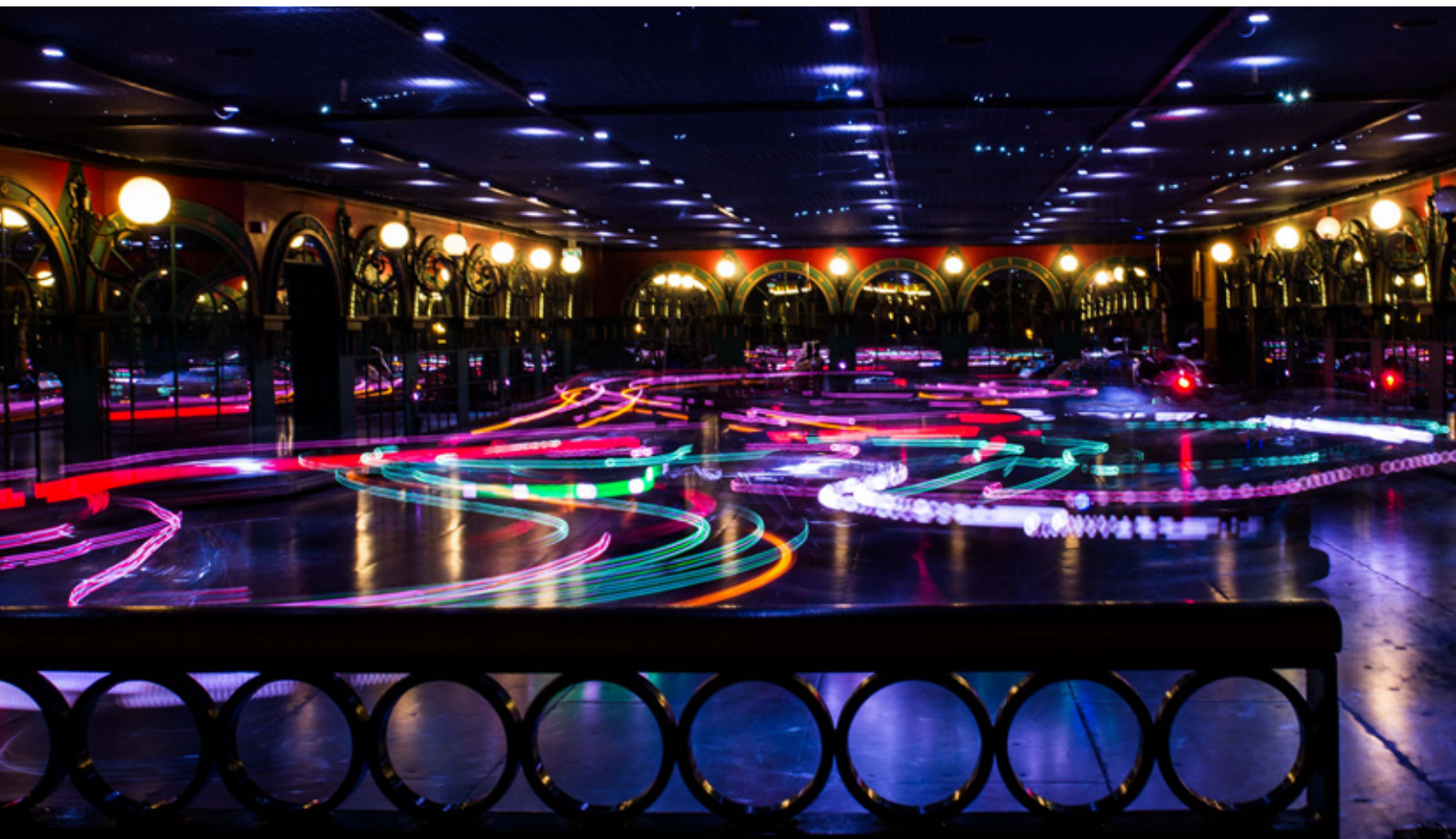
It is important to emphasise that in most cases there is not a ‘right’ or ‘wrong’ way to conduct the analysis. Rather, there are frequently equally valid alternative approaches, or approaches that inherently answer slightly different underlying questions. Therefore, except in a few cases where certain methods lead to inherently misleading results, our approach is to describe the consequences of different approaches, rather than identifying a correct approach.

KEY FINDINGS

The types of publications included in bibliometric analysis and how publication type and fields are accounted for in normalising citations are shown to have relatively small impacts on results for NZ’s bibliometric performance. Though small, one should be cautious that these impacts will be stronger for some fields than others.

Many NZ papers have both NZ and foreign authors. MBIE currently counts such papers as full NZ output. An alternative is to treat each of such papers as a fraction of a publication, with the fraction equal to the fraction of authors who are based in NZ. We find that NZ’s performance in terms of average MNCS across subjects tends to be worse when using fractional counting instead of full counting of authors, implying that the internationally co-authored papers are on average more highly cited than purely domestic papers.

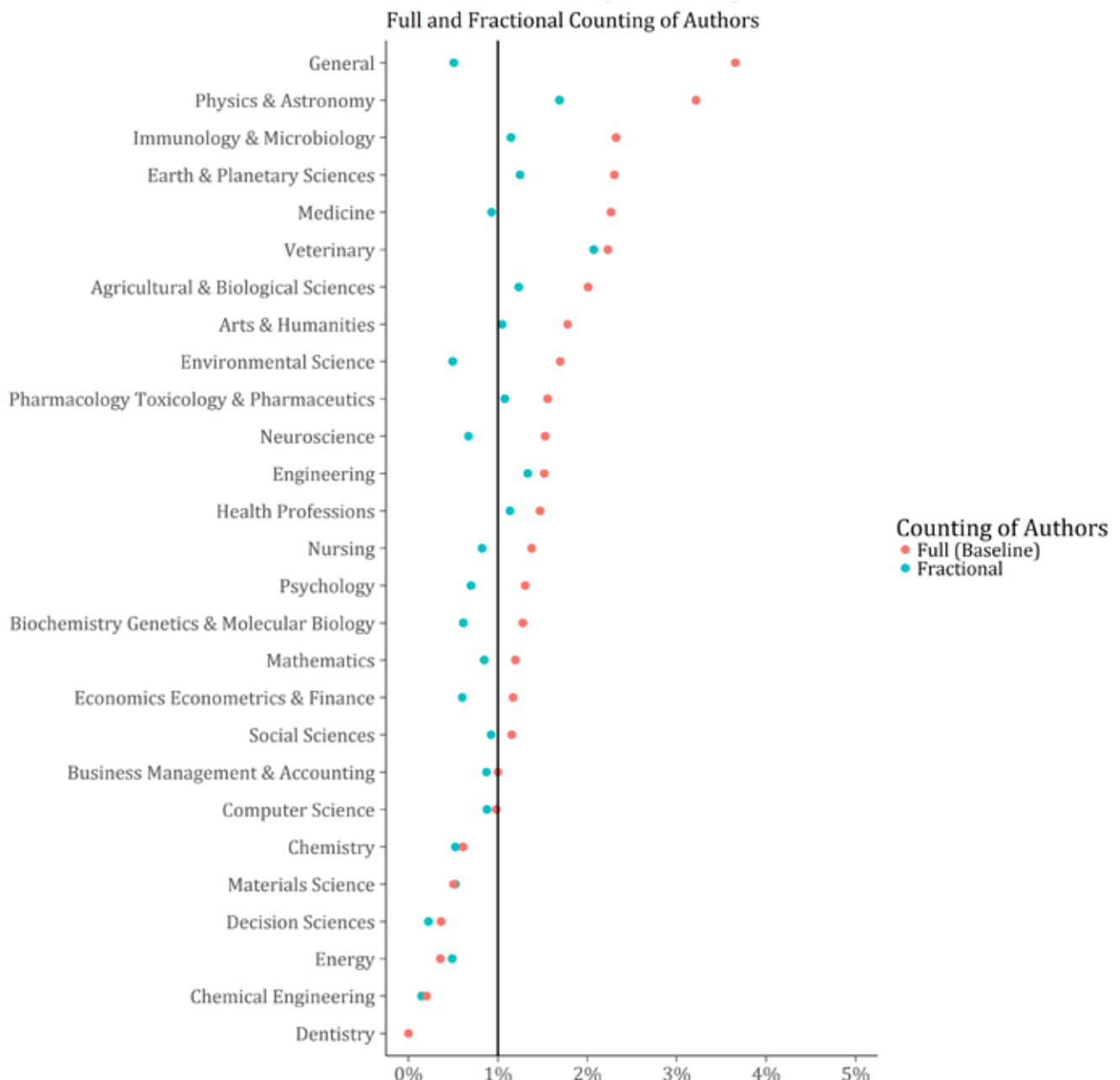
Whether internationally co-authored papers are viewed as a full or partial NZ output also has significant consequences for how the share of publications in top percentiles can be interpreted. In principle, the fraction of NZ publications



in top percentiles can be compared directly to the world at large so that, for example, having more than 1% of NZ publications in the top 1% of the world's citation distribution indicates that NZ has contributed a disproportionately high share of top publications. However, with full counting of authors, a paper with both NZ and U.S. authors that is in the top 1% is counted as a top-1% paper for both countries, meaning around the world there will be more than 1% of all papers in the 'top 1%'.

We show that when fractional counting is used, avoiding the double-counting issue, NZ's performance across disciplines on these percentile measures is much worse than it appeared with full counting.

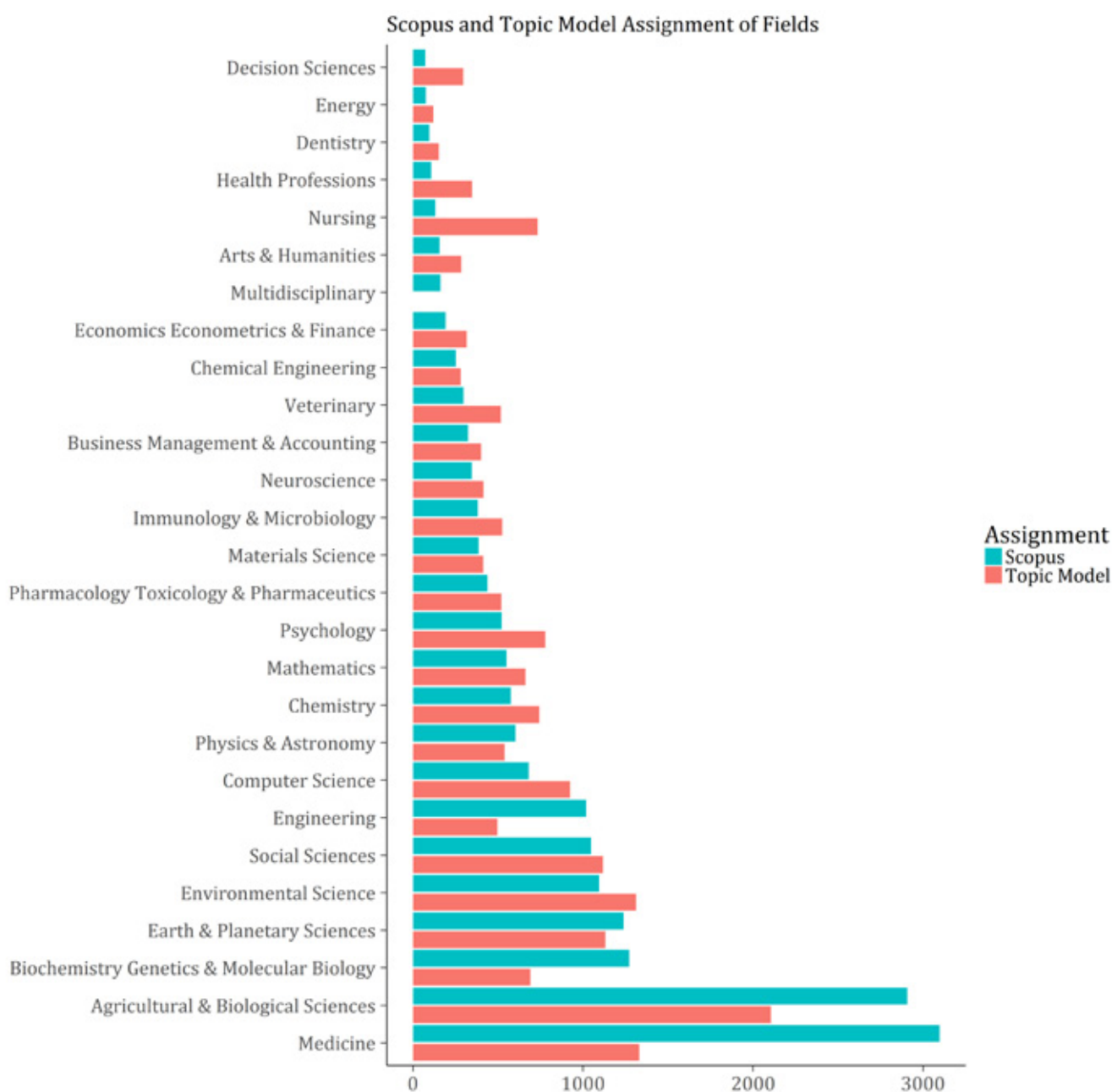
Figure 1: Fraction of NZ Papers in Top 1% in 2012-2014

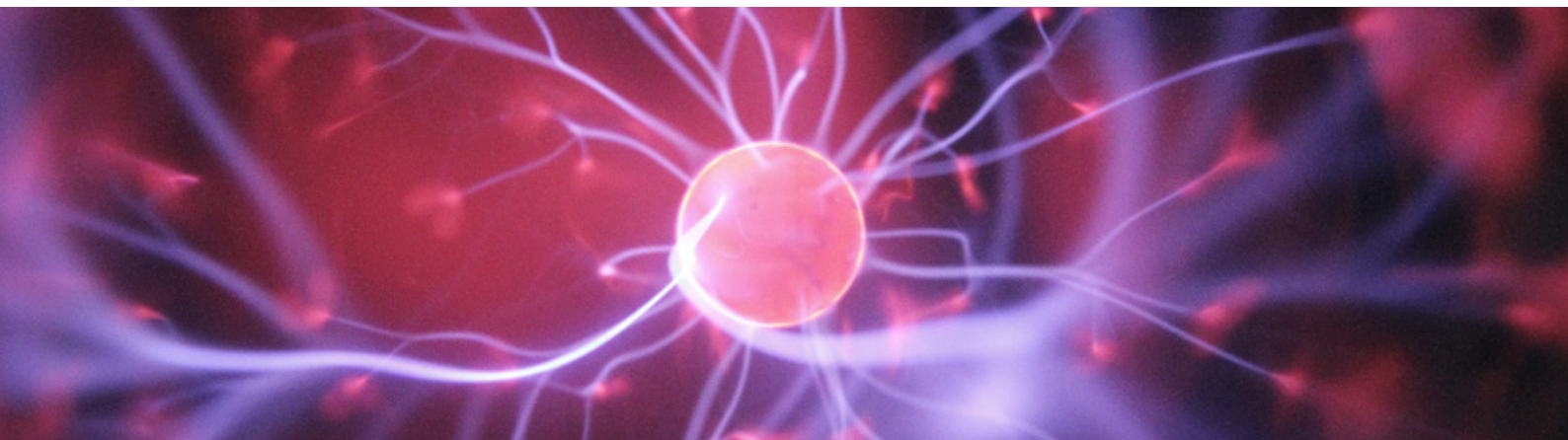


There is another problem in the calculation of fraction of publications in top percentiles that can distort the interpretation of results. Because citation data are discrete and many publications may have the same number of citations, it is possible, for example, that more than 1% of publications have at least as many citations as the threshold for being in the top 1% of the distribution. When calculating the fraction of publications above a certain threshold of the citation distribution, fractional weighting should be applied to publications with the same number of citations as the threshold so that the global fraction of publications in the top $x\%$ is exactly $x\%$. We find that leaving out this adjustment leads to positively biased results for the share of NZ's publications that fall in the top percentiles, particularly in recent years when citation distributions tend to be more condensed.

Publications in the Scopus dataset are linked to disciplines according to the main fields covered by the journal in which they were published. Where publications are in journals that span multiple fields, we do not know to what extent the publication itself covers either of those fields. As an alternative, we explore the use of topic modelling to assign publications to fields using the words in their abstracts. The classifications produced differ somewhat from those based on Scopus journal assignments, but we do not have any external reference to tell us which is better.

Figure 2: NZ's Total Publications 2002-2004





While bibliometric measures are not based on samples in a statistical sense, they are nonetheless potentially sensitive to how the population corresponding to a particular question of interest is defined. We proposed using bootstrapped confidence intervals to characterize this sensitivity and demonstrated this for some of our results, but this is a topic that merits further work. We observe that, for fractions in the top percentiles, uncertainty of results increases the closer the threshold considered is to zero.

CONCLUSION

Because methodological choices underlying bibliometric analysis affect both statistics and their interpretation, we must stress the importance of making these choices transparent. The MBIE 2016 Science and Innovation System Performance Report leaves readers in the dark about how publications in multiple fields or with multiple authors are dealt with. Those readers who are concerned with such issues will find it unclear what conclusions can be drawn from the results, while those who are not could easily draw misinformed conclusions.

In terms of which metrics should be studied, it would not be wise to draw conclusions about NZ's research performance based on any of the metrics we use in isolation. In most disciplines, NZ research is above average, by which we mean that the average MNCS is greater than unity and the proportion of publications with above median citations is greater than 50%. But when we focus on the upper 10% or upper 1% of the citation distributions, NZ's share in most fields is below the world share when properly calculated. This suggests that NZ has a healthy proportion of good researchers but a disproportionately low concentration of international star researchers.

In future work, we will explore the career trajectories of individual NZ scientists, and their relationships with each other and international scientists, to try to understand the dynamics that underlie the distribution of outcomes.

It is important to emphasize that the measures of bibliometrics that we have employed look only at the outputs of the research process. What we would really like to understand is how these outputs relate to research inputs such as researchers, laboratory equipment and supplies. Indeed, in the absence of data on research inputs, none of the bibliometric measures in this report can be used to ascertain in which fields NZ researchers or NZ research investments are more effective or more efficient than others. The launching of the National Research Information System (NRIS) will create a basic data structure for research inputs. Linking those data to bibliometrics should create the potential to begin to answer these questions.

READ THE FULL VERSION OF THE WORKING PAPER AT

WWW.MOTU.ORG.NZ OR CALL US ON 04 939 4250

Motu Economic and Public Policy Research is an independent research institute operating as a charitable trust. It is the top-ranked economics organisation in New Zealand and in the top ten global economic think tanks, according to the Research Papers in Economics (RePEc) website, which ranks all economists and economic research organisations in the world based on the quantity and quality of their research publications.