A Framework for Evaluating the Beneficial Impacts of Publicly Funded Research

Adam B. Jaffe

Motu Economic and Public Policy Research
adam.jaffe@motu.org.nz

Motu

MOTU NOTE #15
JANUARY 2015

EXECUTIVE SUMMARY

Many public entities are interested in empirical estimation of the overall impact, in terms of benefits to citizens, of publicly funded research. These include increased incomes, better health, cleaner environment, enhancement of social and cultural values, and any other benefit that could be an objective of public policy. The ultimate objective of this analysis is to develop reliable measures of such impacts in order to inform decisions about the level of public resources to devote to research, and to evaluate the relative effectiveness of different modes and mechanisms of research support. Because different types of impacts are fundamentally non-commensurable, it is not possible to derive a single composite metric of all research impacts that would be useful for decision purposes. Further, for some important impacts there may be no meaningful quantitative measure, or there may only exist metrics that are illustrative or indicative of the impact in some approximate way. . Any evaluation that excluded those impacts that cannot be directly quantified would be biased. For these reasons, the approach of this paper is to identify a set of metrics and indicators that broadly covers the major categories of impacts. In each case, metrics are proposed that come as close as possible to direct quantification of the impact of interest, but where indirect or approximate indicators are the only ones available for potentially important impacts, they are included.

The report addresses a number of cross-cutting issues that arise in any attempt to evaluate the long-run impacts of research. These include the need to distinguish between true objectives and intermediate outcomes, the need to distinguish between the impact of a particular investment and benefits that would have occurred anyway (the "treatment effect"), the long and uncertainty delay between research and its impacts, irreducible uncertainty about impacts, and the question of whether an attempt should be made to measure all impacts in monetary terms. Although these problems cannot be truly solved, the report suggests approaches for dealing with them.

Previous frameworks for research evaluation have been proposed by various parties around the world. Some common features of these published frameworks are discussed.

A possible framework is proposed that uses eleven identified dimensions grouped into five categories:

Economic

1. New or improved products or services
2. Reduced operating cost or reduced commercial risk
3. Increased wages or improved job opportunities

Environmental

4. Reduced pollution or other anthropogenic environmental impact

Public policy

5. Improvement of public policy or of the delivery of public services

Capability

6. Enhancement of the scientific and technological capabilities of the work force

Social

7. Improved morbidity and mortality, or reduction in the cost of maintaining health

8. Increased knowledge and interest in science

9. Reduction in real or perceived communal risk

10. Enhancement of NZ international reputation, or contribution to sustainable development in other countries

11. Enhancement of social, cultural or community values

More detail on the coverage of each of these dimensions is presented in Table One (p. 16).

Metrics are then proposed for each of these dimensions that provide either direct measures of the impact, proxy or indicator measures, or measures of intermediate objectives that might be indicative of possible eventual impacts. Illustrative examples are provided of the kinds of impacts within each dimension that might be generated by publicly *funded* research.

Conducting a quantitative assessment of a specific body of public research would require consideration of each of the dimensions and their possible metrics, and determination of the extent to which the necessary data exist or could be collected, either for specific programmes or overall. The framework was consciously constructed without regard to the cost or feasibility of measuring the postulated dimensions, so it is expected that full implementation will not be possible for any given programme. By having such a complete framework, however, any partial assessment that is conducted can be presented to decisionmakers with appropriate context as to the potentially important dimensions of impact that are not being addressed.

Acknowledgements

This research was supported by the New Zealand Ministry of Business, Innovation and Employment; GNS Science; ESR, Inc and BRANZ. I have benefitted from discussions with Richard Capie, Peter Cressey, Des Darby, Ewan Delany, Rob Lake, and numerous researchers and staff at each of the funding institutions. The opinions and positions expressed herein are my own and have not been accepted or rejected by the funding organizations, each of which is considering what steps to take with respect to evaluation of its own research impacts.

1. Introduction

Many public entities are interested in empirical estimation of the overall impact, in terms of benefits to citizens, of publicly funded research. These include increased incomes, better health, cleaner environment, enhancement of social and cultural values, and any other benefit that could be an objective of public policy. The ultimate objective of this analysis is to develop reliable measures of such impacts in order to inform decisions about the level of public resources to devote to research, and to evaluate the relative effectiveness of different modes and mechanisms of research support. Because different types of impacts are fundamentally non-commensurable, it is not possible to derive a single composite metric of all research impacts that would be useful for decision purposes. Further, for some important impacts there may be no meaningful quantitative measure, or there may only exist metrics that are illustrative or indicative of the impact in some approximate way. . Any evaluation that excluded those impacts that cannot be directly quantified would be biased. For these reasons, the approach of this paper is to identify a set of metrics and indicators that broadly covers the major categories of impacts. In each case, metrics are proposed that come as close as possible to direct quantification of the impact of interest, but where indirect or approximate indicators are the only ones available for potentially important impacts, they are included.

The next section of the report discusses several cross-cutting issues that arise with respect to a number of different categories of impacts, and proposes approaches to handling these issues. The following section reviews similar or analogous evaluation framework efforts undertaken by others, and identifies some common themes. Section IV describes the proposed framework and possible metrics to implement it. Section V concludes the paper.

2. Cross-cutting issues

2.1. Objectives, means-to-an-end, and indicators

The pathway from research results to ultimate impacts is often complex, and the steps along the way difficult to identify. As a result, it is sometimes useful to identify intermediate outcomes that are not beneficial impacts in themselves, but may be indicative of a higher likelihood of such outcomes eventually occurring. For example, the licensing of a research result by a firm, or investment by a firm in development of a product from public research, are not economic or social impacts in themselves. Impacts only occur when a new product or process is sold or used, generating income and consumer benefits. But licensing and private investment are means to those ends, and their occurring may logically be seen as indicators that the true impacts (increased income and consumer benefit) are more likely than if such private involvement were not occurring. It is important, however, to distinguish between measuring a true impact and measuring an indicator or intermediate step toward such impacts. Otherwise, policy might begin to see the indicator as the objective, and see success in increasing the level of the indicator even in circumstances where it is not likely to be connected to the ultimate desired impacts.[1]

1 In some formulations, a further distinction is drawn, between "outputs" of research such as published papers and granted patents, and "outcomes" of research such as a new product or a new/revised public policy (Jaffe, 1998; Science Foundation Ireland, 2013). Both "outputs" and "outcomes" are intermediate steps on the way to impacts, which are desired consequences for citizens.

2.2. Isolating the "treatment effect"

The goal of identifying the impacts of public research can be thought of in analogy with evaluating the effectiveness of a drug or medical device. In evaluating a drug, it is not enough to know if the people given the drug get better; we need to know whether they get better more often or faster than people who are not given the drug. This difference in outcomes between those given a drug and those who are not is called the "treatment effect." Analogously, if jobs and incomes are created commercializing the outcome of a public research program, we want to know whether and to what extent those impacts are greater than what would have occurred without the public investment.

Measuring the "treatment effect" from public research investments is difficult, because it is often hard to know what would have occurred in the "but for" world in which the public investment was not made. In some cases, it is not clear even conceptually what is the right way to think about the "but for" world. Suppose it were the case that if a given research project were not funded by the sponsoring agency, the results probably would still have been obtained because the research would have been funded by another government ministry. One could say that the agency funding was not necessary, but if each ministry took the same approach then none would fund it and the impact would be lost. More generally, it may be hard to analyse the diverse pathways by which a given impact might be realised. As a result, there is no general solution to the problem of isolating the treatment effect. But it is an issue that should always be part of the analysis of the apparent impacts of any program (European Science Foundation, 2012a). Observed impacts should be credited to a project or programme only on the basis of evidence that they would not otherwise have occurred.

2.3. Time lags

The amount of time that passes between successful research and ultimate impacts can be long and is always uncertain. This means that looking for impacts at any point in time subsequent to completion of a research project always risks missing impacts that have yet to occur. Indeed, the macro evidence is that the average lag between public research and measurable impacts in terms of economic productivity is on the order of decades (Adams, 1990). This suggests that research that *does* create significant impacts will typically be undervalued in its initial aftermath, but of course it does not imply that all research should be assumed to eventually generate impacts. A partial solution to this problem is the identification of intermediate outputs that may be indicative of the likelihood of eventual value, as discussed above. More generally, the problem of long and uncertain lags feeds into the problem of uncertainty.

2.4. Uncertainty

In any evaluation of research impacts, there will always be numerous ways in which the nature and magnitude of the eventual impacts remains uncertain (Lane, 2009). The standard economic approach to this uncertainty is to estimate probabilities of different outcomes, and then calculate "expected impacts" as a probability-weighted average. That is, if there is a 50% chance of a $100 impact and a 50% chance of a $10 impact the expected impact is $55 a(.5 times 100 plus .5 times 10). With respect to research impacts, however, it is difficult even to estimate probabilities for the possible outcomes. Hence a qualitative rather than quantitative approach to characterizing the uncertainty around the outcomes becomes appropriate. For this purpose, it is useful to identify different sources of uncertainty about ultimate impacts, because these different types of uncertainty may have different implications for policy:

- Scientific uncertainty: Depending on how mature research is at the point that evaluation is undertaken, there may remain uncertainty regarding the scientific outcomes, i.e. to what extent the research questions will be answered or proposed technical solutions will work.
- Market uncertainty: for a given technical problem solution, it may be unknown how the costs of production and demand for a product will translate into market success.
- Financial/entrepreneurial uncertainty: even if a product is a potential market success, it is possible that financial or other barriers might prevent its being developed (although distinguishing this possibility from underlying market uncertainty may be difficult)
- Regulatory uncertainty: new products may depend on policy or regulatory settings for their success, in which case their ultimate value will depend on policy decisions.
- Policy "uptake" uncertainty: if the potential impact of research is an improvement in public decisionmaking (e.g. better building codes or environmental regulations), then realisation of the impact may depend on policymakers' willingness to utilise the research.
- An evaluation of impacts for a project or programme should identify the sources and describe qualitatively the magnitude of the uncertainty surrounding the measured impacts. This characterisation of the uncertainty associated with the estimated impact should then be conveyed to decisionmakers along with the estimates.

2.5. Monetisation of impact metrics

As noted above, many of the potential impacts of research are non-commensurable. For some impact types, however, methods do exist for estimating a dollar-equivalent value for non-monetary measures of impact. For example, both improvements in human health and protection or enhancement of environmental amenities can be valued in dollars if one is willing to make the necessary assumptions and collect the necessary data on preferences. There are two basic difficulties with undertaking such monetisation. First, the valuation methods require data such that the valuation becomes a moderately complex research project in itself, and the results are sensitive to the methods used and difficult to generalise and reproduce. Second, even with these efforts, there will remain some impact categories that cannot be monetised. Given that at the end of the day the overall valuation will need to include some benefits that cannot be monetised, it is not clear that replacing some of the non-monetary valuations with inherently imprecise dollar valuations will improve decisionmaking. For the purpose of this report, no attempt will be made to monetise those impacts that are not naturally measured in dollars. Investigation of the possibilities for partial monetisation of some impact categories could be considered in future work.

3. Other frameworks

The problem of measuring impacts of public research is not new; both MBIE and other agencies have considered it in the past. This section provides a brief overview of some of the most relevant previous studies.

Science Foundation Ireland

SFI published a statement on Research Impact in January 2013. It proposes a framework in which all research proposals must contain an "impact statement," defined as "a clear articulation of the potential impact of a proposed research project or programme and actions proposed to help realise this impact." It suggests the following types of impacts (and provides examples for each):

- Economic and commercial impacts, for which the beneficiaries are firms;
- Societal impacts, for which the beneficiaries are individuals, groups, or organisations for whose quality of life, knowledge, behaviours or creative practices have benefitted;
- Impacts on public policy and services, for which the beneficiaries are government, NGOs, charities or other public sector organisations;
- Health impacts, for which the beneficiaries include individuals whose health outcomes have been improved or quality of life enhanced through application of enhanced healthcare;
- Environmental Impacts, where the key beneficiaries are the natural environment or the built environment together with people who benefit as a result; and
- Impacts on professional services. Where beneficiaries include organisations or individuals involved in the development of and delivery of professional services.

SFI emphasises that "impacts" are distinct from what they call research "outputs" such as patents or papers. However, they include other intermediate outcomes such as co-investment by firms and generation of a spin-out company as impacts. The SFI effort to specify the framework for research impact statements is ongoing.

Tekes, The Finnish Funding Agency for Technology and Innovation

Tekes published "A framework for evaluating the societal impact of Research and Innovation" in 2011. They identified four "areas" of impact:

- Economy and economic renewal: "addresses the economic impacts of research and innovation;"

- Environment: "addresses the impacts of research and innovation in the face of environmental challenges such as climate change and resource depletion;"
- Well-being: "consists of impacts of research and innovation on the objective and subjective factors of well-being, such as health and quality of life, working life, and the living environment;" and
- Skills and culture: "includes the impacts of research and innovation activities on the accumulation of knowledge, skills, expert networks, culture and creativity."

The report notes that an alternative formulation would put all of these areas under "well being," and also notes that however defined the areas are "partially overlapping and much interconnected." Although focused on impacts, the Tekes report also discusses indicators for inputs to the research process and levels of research activity.

European Science Foundation Member Organisation Forum on Evaluation of Publicly Funded Research

In 2012, the ESF held a forum on Evaluation of Publicly Funded Research. Its Working Group on Impact Assessment identified the following categories of research impact (in addition to scientific and technological impact):

- Economic impact: contribution to the sale price of products, a firm's costs and revenues (micro level), and economic returns either through economic growth or productivity growth (macro level).
- Social impact: contribution to community welfare, quality of life, behaviour, practices and activities of people and groups.
- Political impact: contribution to how policy makers act and how policies are constructed and to political stability.
- Environmental impact: contribution to the management of the environment, for example, natural resources, environmental pollution, climate and meteorology.
- Health impact: contribution to public health, life expectancy, prevention of illnesses and quality of life.
- Cultural impact: contribution to understanding of ideas and reality, values and beliefs.
- Training impacts: contribution to curricula, pedagogical tools, qualifications (European Science Foundation, 2012b).

The Forum did not address specific examples or metrics.

Hall and Scobie; 2013 MBIE Biological Industries R&D Evaluation

MBIE SSI published an evaluation of the Biological Industries R&D funding in September 2013 (Pells, 2013). This report focused on an examination of the strategy of the fund, and its portfolio management. For the purpose of identifying the beneficial impact of such funding, it relies on previous work, particularly (Hall and Scobie, 2006) for evidence that publicly funded research in this area generates significant positive impact in terms of increased productivity in the agricultural sector. These studies use statistical techniques that examine the association between the level of public spending and the rate of agricultural productivity improvement to infer the impact of the former on the latter. Thus, in relation to the objectives of this report, these studies look at a specific potential impact (increased productivity), and infer the magnitude of that impact from the macro statistical association between funding and productivity. They can be thought of as "top down" rather than "bottom up" estimates of impact, because they do not look at the impacts of specific projects or programmes, but rather infer the impact from the macro correlation between public expenditures and performance.

NZIER Report on Fresh Water Research

NZIER carried out an evaluation in 2012 of the impacts of MBIE fresh water research. They developed and applied a methodology based on systematic tabulation of judgments of research impacts based on self-reporting in research project reports. Using specific impacts claimed by the research teams in these reports, they assigned a numerical significance score for each impact based on the nature of the claimed impact and the degree of quantification of that impact. They then categorised these claimed impacts as economic, cultural, social or environmental. An overall measure of improvement in "well-being" flowing from each project was constructed by weighting the assigned significance scores using the importance assigned to each of these four categories by different groups, specifically business, Iwi/Māori, NGOs/Community and Government.

What is significant about this methodology for the purpose of this report is the idea that the nature of economic, cultural, social and environmental impacts may differ or may be perceived to differ for different stakeholder groups. This is an important insight and may be important for detailed evaluation of a specific research area such as fresh water. For the higher-level analysis desired here, it introduces a level of complexity that is problematic. In the analysis of the next section the interests of different stakeholders are accommodated by including specific dimensions of impact that relate to key stakeholders (e.g. enhancement of Iwi/Māori interests), rather than by tracking each stakeholder group's take on every dimension. This seems to be a reasonable compromise between the need to allow for the interests of different groups and the need to keep the framework manageable.

Ministry of Economic Development Evaluation of Public R&D Assistance to Firms

In 2011 MED published two reports evaluating the impact on the performance of firms of public R&D support programs. These reports look at a specific category of potential impacts—changes in the behaviour and performance of firms—and consider various means of measuring those impacts. These analyses are quite relevant to subsequent implementation of the framework developed herein, because they address carefully the question of isolating the treatment effect of government expenditure. But they do not speak to the current task of establishing an overall framework for identifying and quantifying the distinct dimensions of public research impacts.

Dare to Measure: Evaluation Designs for Industrial Policy in the Netherlands

This study considers different methods for solving the "treatment effect" problem discussed above, and applies them to several industrial research and innovation programmes in the Netherlands. It does not address impact evaluation *per se*, taking the stated goal of each programme as its potential impact and focusing on the methodology of identifying that impact. Hence it's usefulness is similar to the previous MED reports on R&D programs, in that it does not address the definition of impact dimensions, but could be very useful with respect to studies based on this framework to measure the impacts of each programme.

Deloitte Access Economics for the Australian National Health and Medical Research Council

This study purports to calculate the economic rate of return on investments made in health-related research in Australia. Unlike the other studies discussed, it does not make any attempt to characterise or identify the actual impacts of the funded research. It simply takes existing projections of future improvements in health in Australia, and makes a series of assumptions (what fraction of projected improvements result from research; what fraction of the research-based improvements derive from Australian research) to calculate a dollar value that it imputes to Australia's health research investment. It cannot be considered a framework for research impact evaluation. It is a methodology for assigning a dollar value to a postulated level of research impact.

4. Proposed framework

4.1. Considerations in framework design

As the above survey of others' approaches demonstrates, there is no "correct" way to capture and characterise the impacts of publicly funded research. Any chosen framework represents a set of subjective choices regarding how the framework best represents reality while being workable for decision purposes. Before proposing a specific framework, it is useful to identify some of these choices.

<u>Tradeoff involving breadth, precision and manageability:</u> In defining the set of impact dimensions to be considered, there is an inherent tension among:

- breadth of coverage—inclusion of impacts that span economic, social, cultural environmental objectives;
- precision—specification of impacts in a way that precisely captures the nuance and particularity of impacts in each specific research area; and
- manageability—keeping the number of impact dimensions small enough that an overall picture of research impact emerges in a way that is useful for decision purposes.

In general, it is not hard to achieve any two of these three objectives. One could have a very broad and very precise framework, but doing so would require an unmanageably large number of dimensions. One could have a very precisely defined set of impact dimensions, and keep the number manageable, if one were willing to focus narrowly on (for example) a specific set of economic impacts. Or one can have broad coverage with a manageable set of dimensions, but only at the expense of precision with respect to some differences in the nature of impacts across different research areas. The optimal balance among these objectives depends on the purposes for which the analysis will be used. Roughly speaking, this report leans towards defining a set of impact dimensions that is broad and manageable in number, at the inevitable expense of some degree of precision with respect to the nuances of impacts across different research areas.

International comparability: Any one country inevitably has only limited data from which to explore empirically the relative effectiveness of different modes or mechanisms for public research investment. Therefore, it would be desirable to be able to exploit other countries' experience in making policy decisions. This will be easier to do to the extent that each country's impact evaluation framework can be mapped or translated to others' systems. Such mapping or translation is facilitated by similarity of structure, and by clear definition.

Groupings or categories: Conceptually, an impact framework could consist of an enumerated set of dimensions of potential research impact, without any grouping or assignment of these dimensions to named categories. In practice, all of the previously described frameworks do group impact dimensions into broad categories such as economic, social, cultural, etc. Such grouping is useful for decisionmaking, and also facilitates identification of types of impacts that may be missed or captured only inadequately.

Measurability: As noted above, some dimensions of impact are inherently easier to measure or quantify than others. It is natural in constructing an impact evaluation framework to concentrate on impact dimensions that are measurable. But such an approach inherently biases any evaluation based on the framework. Such a bias can potentially lead to bad decisionmaking, and also undermines the legitimacy of the fundamental project of systematic evaluation. Therefore, this report presents a framework for characterizing impacts that includes all dimensions that seem important, even if they can be measured only approximately or imperfectly.

Inclusion of intermediate outcomes or objectives: As noted above, there is some degree of ambiguity as to the distinction between "impacts" and "outcomes." The approach taken here is to limit the definition of impact dimensions to potential research consequences that are, in and of themselves, of value to citizens. Potential research consequences that are desirable because they may ultimately lead to beneficial impacts for citizens, or which may be indirectly indicative that such benefits will occur, will be considered in the context of potential metrics or indicators for each impact dimension, but will not be treated as impacts themselves. As an example, higher incomes are an end in themselves, but formation of spin-out firms or co-investment by firms in the outcomes of public research are not valuable in and of themselves, and hence will be considered as metrics or indicators but not as impacts.

4.2. Proposed dimensions of impact

The goal in specifying categories of impact is to facilitate the usefulness of the framework for decision purposes, to maintain conceptual clarity so that the framework can be meaningfully mapped onto other frameworks that might be used by others, and to keep the number of categories relatively small. The following five categories are proposed:

- Economic impacts are benefits enjoyed by individual citizens in the form of higher incomes or consumption of higher-quality goods and services.
- Environmental impacts are improvements or avoided harm to the natural environment.
- Public policy impacts are the facilitation of better decisionmaking with respect to public policies, laws and regulations, or more effective government operations.
- Capability impacts are the improvement in the scientific and technical capabilities of the work force.
- Social and cultural impacts are all benefits enjoyed by citizens other than those defined in the other categories, including health and safety,[2] maintenance or enhancement of heritage, and cultural or national enjoyment.

As noted, this categorisation is inherently somewhat arbitrary. So long as the specific impact dimensions are well-defined, they could be easily categorised in different ways.

The proposed dimensions of research impact are presented in Table One (p. 16). In identifying the multiple dimensions, there is no intention that the various dimensions are of equal importance, or are comparable to each other in terms of the breadth of impacts they encompass. There is an unavoidable degree of arbitrariness as to whether related potential impacts should be grouped under one dimension, or identified as a separate dimension. For example, "increased knowledge and interest in science" could have been included within the more general "enhancement of social, cultural and community values" rather than identified as a dimension of its own. Conversely, "cost reduction" and "risk reduction" could have been broken into two separate categories rather than being grouped together. The particular aggregations and disaggregations shown are based on judgment regarding which impacts will likely be viewed similarly by decisionmakers, but it could be modified if more detail is desired in a particular area, or if a particular dimension doesn't seem to have much activity.

Note that a given outcome of the research process can have impact along multiple dimensions. For example, research that generates a new technology for mitigating some environmental impact could potentially lead to reduced environmental impact (4), improved environmental or land use regulation (5) and a new product (1). Many research projects or programmes might have an education or training component that contribute to (3), in addition to whatever other impacts are generated along the other dimensions by the research outcomes.

2 One could think of health as something that people consume, and hence categorise health benefits as economic. But some health benefits of research (e.g. diminished earthquake hazard) do not come from better health care, and health care itself is delivered outside markets to a significant degree, so it seems preferable to consider improved health and safety as a social and cultural impact rather than an economic one.

Table Two (p. 17) gives illustrative examples of hypothetical or actual impacts that might be generated across the different dimensions for each of the Vote Science and Innovation contestable funds. Obviously, each fund is likely to have impacts that are more concentrated along some dimension than other funds, and some funds might never generate an impact along some dimensions. But the eleven dimensions should collectively span the space of all potential impacts of all of the funds.

4.3. Impact metrics

In choosing metrics for each dimension of impact, it is useful to consider what attributes make a given metric most useful and reliable. Desirable attributes include:

- <u>Conceptual validity:</u> the metric is connected to the underlying impact concept by a meaningful and valid model
- <u>Statistical validity:</u> the is metric highly correlated with the underlying impact
- <u>Consistency over time:</u> the metric's relationship to the underlying impact concept is stable over time
- <u>Consistency over context:</u> the metric's relationship to the underlying impact concept is the same in different contexts (e.g. different scientific fields or different countries)
- <u>Aggregability:</u> the metric can be aggregated across projects, programmes, agencies, etc., so that the impact of the aggregate is equal to the sum of the measured impacts of the components

Undesirable attributes include:

- <u>Manipulability:</u> the level of the metric can be increased without increasing the underlying impact concept through conscious manipulation
- <u>Cost:</u> what resources are necessary to collect the data or information necessary to construct the metric

In most cases tradeoffs will have to be made in choosing metrics, or weaknesses of specific metrics will have to be considered as they are used for different decisionmaking purposes. For example, in constructing an overall assessment of the impact of public research, it is essential that the metrics be aggregable across projects, but this may simply not be possible for some dimensions of impact. This inability to aggregate will then have to be addressed in constructing an overall assessment.

Table Two (p. 17) proposes possible metrics for each of the dimensions. In order to facilitate thinking about the information presented by different metrics, they are classified into three types. For some dimensions, there is a "direct measure," meaning a metric that captures conceptually the manner in which citizens' well-being is enhanced along this dimension by research outcomes. For example, if new products or services are sold, economic theory suggests that the revenue generated by the sales is a measure of the social value of the new goods or services being sold. If environmental impact is reduced, quantitative measurement of the extent of reduction of specific pollutants measures the social benefit thereby created.

There will be many cases, however, where such direct measures are not available. For some dimensions of impact, such as enhancement of scientific capability, it is not clear that a direct measure exists, even in principle. And even for dimensions such as new products for which such a measure may exist in principle, there will likely be many specific research projects or programmes for which the direct measure is not available, either because of data limitations, or because the passage of time has been inadequate to determine whether a new product will ultimately be forthcoming and how successful it might be. For this reason, it is useful to consider additional types of metrics. The second column presents "proxy" or "indicator" metrics. These are metrics that do not truly capture the benefit to citizens, but are likely to be indicative of such a benefit. That is, a list or count of new products created doesn't measure the benefit, because many new products with limited sales do not represent as large a benefit as a single new product that is wildly successful. But it is nonetheless potentially useful as an indicator, however imprecise, of ultimate benefit. Similarly, expert evaluations or measures of knowledge network connectedness are inherently imperfect but nonetheless indicative of impact dimensions such as science capability and communal risk.

Finally, the third column presents metrics that are potentially meaningful as measures of intermediate outcomes, i.e. steps on the path to realisation of the ultimate impact. These differ from the proxies or indicators in the second column in that they measure *something* accurately, but that something is not really of value, in and of itself, to citizens. That is, at least in principle we can measure private sector development investment with reasonable accuracy, so it is not a "proxy" in the same sense as an expert assessment is only a proxy for the true degree of risk reduction. And having increased private sector investment is not something of value, so it is not a desired research impact. But it is relevant because it suggests that new products or efficiency improvements (which *are* valuable) may materialise down the line. Because these ultimate impacts may not be observable (particularly early in the process), it is useful to measure progress on these intermediate outcomes as an indirect indicator of eventual impact. Of course, intermediate success does not guarantee ultimate success, so these measures are also subject to error.

Referring back to the desired attributes of metrics introduced at the beginning of this section, there is considerable uncertainty has to how good many of these metrics would be. In particular, there is no way to know their statistical validity without further research (and even with such research, statistical validity could only be established for those dimensions for which a valid direct measure exists). There are, however, some clear limitations. In particular, except for the economic impacts, most of the metrics suffer from poor aggregability. This is partially inherent in the non-commensurable nature of the non-economic impacts: no matter how good your measurements, it is not clear conceptually how one would add together a positive impact on cultural heritage and a positive impact on social cohesion to calculate an overall impact on social/cultural values. In other cases it is more of a measurement issue. One might imagine calculating an overall impact on aggregate community risks, but if different projects' impact on that dimension are measured using subjective expert assessments it is not clear how to add those together. These considerations will need to inform decisions about how to implement the framework and how to present the results.

5. Conclusion

This paper presents a framework designed to be used as the basis for a bottom-up assessment of the overall impact of publicly funded research. The next step towards such an assessment is to consider each of the dimensions and their possible metrics, and determine the extent to which the necessary data exist or could be collected, either for specific programmes or overall. The framework was consciously constructed without regard to the cost or feasibility of measuring the postulated dimensions, so it is expected that full implementation will not be possible. By having such a complete framework, however, any partial assessment that is conducted can be presented to decisionmakers with appropriate context as to the potentially important dimensions of impact that are not being addressed.

REFERENCES

Adams, James D. 1990. "Fundamental Stocks of Knowledge and Productivity Growth." *Journal of Political Economy* 98:4 pp. 673-702.

Deloitte Access Economics. 2011. "Returns on NHMRC funded Research and Development," Australian Society for Medical Research, 17 October 2011.

European Science Foundation. 2012a. "Evaluation in Research and Research Funding Organisations: European Practices." http://www.esf.org/coordinating-research/mo-fora/evaluation-of-publicly-funded-research.html

European Science Foundation. 2012b. "The Challenges of Impact Assessment." http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/MO_FORA/MOFORUM_Eval_PFR__II_/Publications/WG2_new.pdf&t=1389068701&hash=39d2cdb57845ded4abcdfc50c3282546db323daf

Impact Evaluation Expert Working Group. 2012. "Dare to Measure: Evaluation designs for industrial policy in The Netherlands." Impact Evaluation Expert Working Group, the Netherlands.

Hall, Julia and Grant Scobie. 2006. "The Role of T&D in Productivity Growth: The Case of Agriculture in New Zealand: 1927 to 2001," *Working Paper 06/01*, The Treasury, Wellington.

Jaffe, Adam B. 1998. "Measurement Issues." Lewis Branscomb, ed., *Investing in Innovation*. Cambridge: MIT Press.

Lane, Julia. 2009. "Assessing the Impact of Science Funding," *Science* 324:5932 pp. 1273–1275.

Ministry of Economic Development Evaluation Team. 2011. "Innovation Policies and Funding in New Zealand: How Effective Are They? A Survey of the Evidence from Recent Evaluations, Research Papers, Statistical Studies and Policy Documents." Ministry of Economic Development, February 2011.

Ministry of Economic Development Evaluation Team. 2011. "Evaluation of the

Impacts of Cross-vote Government Assistance on Firm Performance: Stage 2, Impacts of Direct Financial Support for R&D," Ministry of Economic Development, April 2011.

NZIER. 2012. "Evaluating Research: An Experimental Method and Application to Freshwater Research." Report to the Ministry of Business, Innovation and Employment, Wellington.

Pells, Sharon. 2013. "Evaluation of Biological Industries R&D—Final Report," *MBIE SSI File Reference ES10008*, Ministry of Business, Innovation and Employment, Wellington.

Science Foundation Ireland. 2013. "SFI Research Impact Jan 2013". Available online at http://www.sfi.ie/assets/files/downloads/Funding/Impact/SFI_Research_Impact_2013.pdf.

Tekes, The Finnish Funding Agency for Technology and Innovation. 2011. "Better Results, More Value: A Framework for Analysing the Societal Impact of Research and Innovation," *Tekes Review 288/2011*, Tekes, Helsinki.

TABLE ONE
DIMENSIONS OF PUBLIC RESEARCH IMPACT

Economic

1. New or improved products or services

   Includes specific innovations sold to the market, and knowledge infrastructure that facilitates innovation or makes existing products better or more valuable (e.g. new methods of testing that improve reliability or maintenance of databases that can be mined for innovations)

2. Reduced operating cost or reduced commercial risk

   Includes specific efficiency improvements, and knowledge infrastructure that facilitates reduction of cost or risk (e.g. generically applicable production methods or maintenance of databases that lower risk of commercial activities)

3. Increased wages or improved job opportunities

   Includes creation of new or better jobs, and enhancement of human capital

Environmental

4. Reduced pollution or other anthropogenic environmental impact

   Technologies, methods or knowledge that reduce environmental impacts, lower the cost of achieving existing impact targets or goals, or facilitate enhancement of environmental goals

Public policy

5. Improvement of public policy or of the delivery of public services

   Technologies, methods or knowledge that facilitates better public decionmaking or more effective or efficient delivery of public services

Capability

6. Enhancement of the scientific and technological capabilities of the work force

   Increase in the number of skilled people or in the skill level of the existing force

Social

7. Improved morbidity and mortality, or reduction in the cost of maintaining health

   Technologies, methods or knowledge that allow people to live longer and/or healthier lives, or reduce the cost of maintaining existing health states, either through reduction of health risks or improvement of health care

8. Increased knowledge and interest in science

   Generation of excitement and knowledge among the non-expert community about science and the scientific dimensions of public issues

9. Reduction in real or perceived communal risk

   Technologies, methods or knowledge that reduce risks or threats to the existence, stability and cohesion of communities, or which mitigate fear of those risks in a positive way

10. Enhancement of international reputation, or contribution to sustainable development

    Diplomatic benefit or national pride/satisfaction from being perceived as an international leader in specific research fields, or from creation of technologies, methods or knowledge that contribute to the improvement of well-being in less developed countries

11. Enhancement of social, cultural or community values

    Technologies, methods or knowledge that facilitate maintenance or enhancement of heritage values, social cohesion, cultural and aesthetic values, or other communal values beyond economic and environmental

TABLE TWO

Category	Impact dimension	Direct Measure	Proxy or indicator	Intermediate outcome
Economic	1. New or improved products or services	additional revenue	enumeration of new products and processes enumeration of commercialization/licensing discussions	private sector development investment licensing to private firms; achievement of license agreement milestones new firm creation venture capital investment in new firms exports
	2. Reduced operating cost or commercial risk	cost reduction or productivity improvement		increased investment (plant and equipment)
	3. Increased wages or improved job opportunities	wages in new positions		person-years of training at various levels employment of research project staff in industry employment at newly created firms
Environmental	4. Reduced pollution or other anthropogenic environmental impact	reduction in emissions or other environmental impact (tons; percent of total emissions)		
Public policy	5. Improvement of public policy or of the delivery of public services	issuance or implementation of policy or practice incorporating research results		workshops or other delivery of policy, programmatic or operational advice to governmental body collaboration between researchers and public employees or politicians

Continued page 18

Category	Impact dimension	Direct Measure	Proxy or indicator	Intermediate outcome
Capability	6. Enhancement of the scientific and technological capabilities of the NZ work force		increased connectness or collaboration measures in innovation network	person-years of training at various levels international collaborations attraction of scientists and enginners from overseas to locate in NZ
Social	7. Improved morbidity and mortality, or reduction in the cost of maintaining health.	increase in quality-adjusted life years		adoption of new technology or practice in health care
	8. Increased knowledge and interest in science			time spent in interactions with public development and use of educational materials
	9. Reduction in real or perceived communal risk		expert assessment of communal risk reduction survey results regarding public risk perceptions	
	10. Enhancement of NZ international reputation, or contribution to sustainable development outside NZ	increased income or measured well-being of overshore beneficiaries	expert assessment of reputational impacts international rankings	
	11. Enhancement of social, cultural or community values		expert assessment of values impacts	