CREATE
National Center for Risk and Economic Analysis of Terrorism Events
University of Southern California

S&T Analysis and Management of Innovation Activity III (STAMINA III)

Appendix C: Literature Review of R&D Indicators of Success (IoS) Related to Transition

December 23, 2022

DHS Science and Technology Directorate (S&T)
Office of Science and Engineering (OSE)
Technology Scouting and Transition Division (TST)

Basic Ordering Agreement (BOA):
HSHQDC-17-A-B0004/70RSAT21FR0000143

Detlof von Winterfeldt, Principal Investigator (PI), Isaac Maya, Co-PI

Project Team: Jennifer Sosenko and Kaitlyn Werden

Center for Risk and Economic Analysis of Threats and Emergencies (CREATE)
University of Southern California (USC)
1150 S. Olive St., 17th Floor
Los Angeles, CA 90015

This research was supported in whole or in part with Federal funds from the United States Department of Homeland Security (DHS) through the Center for Risk and Economic Analysis of Threats and Emergencies (CREATE) at the University of Southern California (USC) under Basic Ordering Agreement (BOA), HSHQDC-17-A-B0004/70RSAT21FR0000143. This report does not necessarily reflect the views or policies of DHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. Furthermore, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of DHS, or USC, or CREATE.

S&T Analysis and Management of Innovation Activity III (STAMINA III) Annual Report

Appendix C: Literature Review of R&D Indicators of Success (IoS) Related to Transition

Contents

1. Foundations for the Development of Indicators of Success of R&D Transition Uptake Benefit Assessment

This section provides an introduction into the broad range of definitions and interpretations associated with the terms used in describing R&D Indicators of Success (IoS) specific to transition uptake benefit assessment. The objective is to establish a base for common understanding and unify interpretation, thus providing a practical perspective useful to S&T. The goal is to inform the S&T process in its development of IoS for R&D transition project uptake benefit assessment.

1.1. INTRAC 2015: Outputs, Outcomes and Impact

This article (https://www.intrac.org/wpcms/wp-content/uploads/2016/06/Monitoring-and-Evaluation-Series-Outcomes-Outputs-and-Impact-7.pdf) aims to clarify the process of input through output, outcomes, and eventually value while defining these terms and eliminating some confusion among them. Inputs, activities, outputs, outcomes and impact are terms that are used to describe changes at different levels from the delivery of goods and services to long-term, sustainable change in people's lives. Whilst the terminology is in common use, there is great inconsistency in how the terms are interpreted. Figure 1-1, from the article, shows the definitions developed by OECD DAC in 2002.

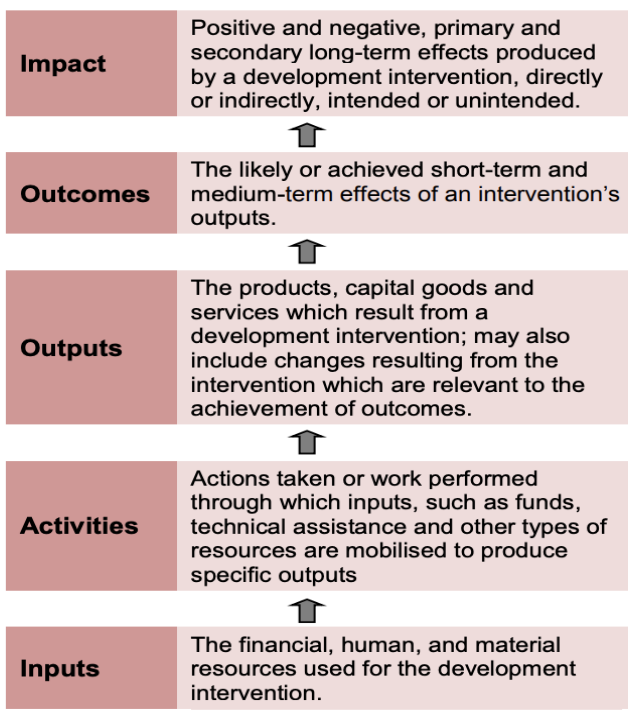

Figure 1-1. Definitions of Inputs, Activities, Outputs, Outcomes and Impact originally developed by OECD DAC in 2002 (OECD, 2010)[1].

Tying these terms to Results and Objectives leads to Figure 1-2, from the INTRAC 2015 article cited.

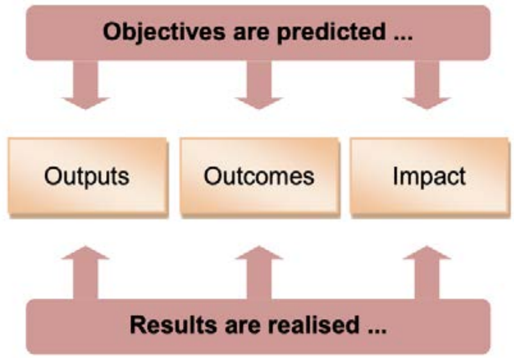

Figure 1-2. Connecting Objectives & Results to Outputs, Outcomes & Impact (from the INTRAC 2015 article cited).

1.2. 2021 Output, Outcomes, Impact and KPIs

The goal of this article (https://jeffgothelf.com/blog/output-outcomes-impact-and-kpis/) is to build a shared understanding of and a consistency with vocabulary. It encourages movement away from outputs and towards outcomes, referring to a 2019 book by Joshua Seiden, Outcomes over Output: Why customer behavior is the key metric for business success[1]. Quoting from the article,

- **"Output** – This is the stuff that we make. These are our features, products, content, services, policies, programs, initiatives and projects. At the end of each work day, output is what you have produced. It is what our users and customers buy, use and consume from us. Examples of output could be a new video streaming app, a faster way to scan my receipts for my expense reports, a new vacation policy at your company or a diversity and inclusion initiative at the office.
- **"Outcome** – Outcome is a measurable change in human behavior we see when we give the output to our users and customers. Outcome answers the question, "What are people doing differently now that we have delivered the output?" Outcomes are not features. They are metrics. For example, an outcome is not, "we shipped the app." Instead, an outcome is, "50% of our audience has upgraded to the new app." Outcomes tell us when we've delivered something of value (or not). They measure, empirically, whether we've made the user experience better for our target audience. If the outcomes we expected aren't met, we missed something with the output. There was a design flaw, a business model mistake, a poor rollout of the initiative, etc. Outcomes are the measures of success of our output.
- **"Impacts** – Impact metrics are the high level measures of the health of your business. These are often the metrics you will find on an executive dashboard. Sales, revenue, profit margin, customer satisfaction, churn are just a few examples you can label as impact metrics. Impact

[1] From the book's cover: In the old days, when we made physical products, setting project goals wasn't that hard. But in today's service- and software-driven world, "done" is less obvious. When is Amazon done? When is Google done? Or Facebook? In reality, services powered by digital systems are never done. So then how do we give teams a goal that they can work on? Mostly, we simply ask teams to build features—but features are the wrong way to go. We often build features that create no value. Instead, we need to give teams an outcome to achieve. Using outcomes creates focus and alignment. It eliminates needless work. And it puts the customer at the center of everything you do. Setting goals as outcomes sounds simple, but it can be hard to do in practice. This book is a practical guide to using outcomes to guide the work of your team.

metrics are lagging indicators. In other words, they look backwards to things that have already happened. Outcomes are leading indicators. They are tactical metrics that give us a sense of how our impact metrics will do in the future. In large organizations, c-level executives and business unit leaders worry about impact metrics while product-level teams focus mostly on outcomes.

- **"Key Performance Indicators (KPIs)** – As a term, KPI is far more common than any of the other ones in this article. I regularly get asked how outcomes, impacts and Objectives and Key Results (OKRs) integrate with KPI's, if they do at all. My take is this: KPI's are impact metrics. More often than not the metrics you'll find on a "KPI dashboard" are the same metrics you would call impact metrics — high level measures of the health of the business."

1.3. intuitix: Best KPI for Research and Development (R&D)

This article (https://intuitix.co/best-kpi-for-research-and-development-rd/) states that the two most clear metrics to assess in R&D are inputs and outputs. Input can usually be defined as the time or cost spent on R&D. Output on the other hand can be more complex, must summarize what the department produces - the results of successful projects, and might include tangible outputs, such as,

- Time savings
- Monetary savings
- Safety improvements
- Product improvements
- Service improvements

R&D output however should also consider the less tangible outputs, such as,

- Learning achieved throughout the process
- Transferable skills gained
- Capital assets acquired
- Valuable business relationships formed

Input and output provide a solid base to communicate a high level overview of performance, but to gather more detailed insights, can use Quantity, Quality and Time,

- Quantity is simply the amount of output. This may be the amount of,
    - R&D projects that succeed
    - Projects that are undertaken
    - A specific measure within each project (i.e cost)
- Quality looks more at the standard of output. This might include,
    - Amount of successful projects divided by unsuccessful projects
    - Average market readiness or Technology Readiness Level of projects
    - Variance between project concept and delivery
- Time is the speed of output. This can include average times for,
    - A project to be delivered
    - A project to break even
    - A project to pass through a certain phase

Use these metrics to explore strengths and weaknesses, for example, an R&D department produces a large amount of projects, but none of them are successful. In this instance, let's say they mostly fail to adhere to the brief. The department can deduce that they must improve on the quality of their output, rather than the quantity.

2. Survey of R&D Indicators of Success (IoS)

The literature on R&D indicators of performance is extensive. This section presents references focusing on articles applicable to S&T's R&D transition uptake assessment.

2.1. GAO 1997: Measuring Performance, Strengths and Limitations of Research Indicators

In studying complex processes or concepts such as innovation, it is not always possible to measure them directly. As a result, researchers turn to the use of "indicators." Indicators point to or illustrate the process or concept in question but do not directly measure it.[2] This GAO reference also points out that R&D spending data provide an indication of how much research is being performed but do not provide a measure of the impacts of that spending, and in fact states that R&D spending is <u>Not</u> a good indicator of R&D results, and decisionmakers in the public and private sectors typically have chosen to measure outcomes using a variety of proxies, including

1. **Return on Investment (ROI)** – measuring the sales and profits resulting from investments in R&D, but there are many issues with this metric, such as the lag between the R&D and the industry effect, and the difficulty in tracing outcomes to specific inputs.
2. **Patent Counts** –measuring technical change and inventive input and output over time, but with intrinsic drawbacks, such as inconsistency across industries regarding propensity to patent inventions, and patent statistics that do not distinguish between major and minor improvements, and the limited applicability of patent evaluation to government R&D.
3. **Bibliometrics** – the study of published data, bibliometrics counts citations in an attempt to address questions of productivity, influence, and the transfer of knowledge. However, this metric is most applicable in areas such as basic research where the results are more often published than protected by firms, and its usefulness as a measure of research results remains somewhat controversial. Issues include weight by publication or other quality measure, the frequency of citation, interdisciplinary comparisons of results, and its limited use in government.
4. **Peer Review** – the National Science and Technology Council concluded that quantitative measures alone cannot supplant the essential element of expert judgment/peer review. However, peer review generally depends on criteria that are inherently difficult to measure, rely on subjective judgment that is vulnerable to bias, depends on the expertise of the reviewer, and expensive.

The GAO reference ends with a caveat that quantitative metrics, such as tasks completed, patents awarded, articles published, etc., measure <u>Outputs</u> and not <u>Outcomes</u>.

2.2. NAP 1997: Measuring Outputs and Outcomes of Innovation

This workshop[3] recognized that innovation output may be indicated by statistics on patents, papers, prizes, invention disclosures, and degrees awarded, while outcome proxies include patent and paper citations, expert evaluations, innovation counts, new product sales, measured productivity growth, and benefit/cost or rate-of-return estimates. The workshop explored the following in more detail,

- Patent Counts and Patent Citations as indicators of Innovation Output
    - Strengths

    - Are available to the public,
    - Contain considerable detail, and
    - Are available over time, enabling time series analysis.
  - Limitations
    - Some important technologies are not patentable, such as software that is protected by copyright.
    - Not all inventions are patented. Firms protect the returns to their investment in other ways such as through secrecy, lead-time advantages, and marketing.
    - Firms patent for different reasons, not all of them related directly to commercial exploitation, for example, to protect an invention from imitation, to block competitors from patenting or pursuing a line of research, or to evaluate the productivity of their R&D activities. This adds considerable uncertainty to the interpretation of patent count data.
    - Patents represent only the practical application of ideas, not more general advances in knowledge.
    - Patents represent inventions, not activities and investments to commercialize new technology.
    - Patents have widely varying commercial value and therefore significance with respect to innovation.
- New Measures of Innovative Output
  - Financial Market Valuations of R&D Investments - For example, researchers have used financial reports to link R&D investments to changes in company stock prices
  - Innovation Counts
  - Bibliometrics
  - Technometrics - Measuring and comparing the dimensions of technical performance of a product or production process
- Outputs and Outcomes Occurring as Spillovers – Benefits accruing to other firms and industries without compensation to the innovating firm. To understand the nature and extent of innovation spillovers, input, output, and outcome data need to be matched at different levels of aggregation.

2.3. NIH 2010: STAR METRICS: New Way to Measure the Impact of Federally Funded Research

The Science and Technology in America's Reinvestment – Measuring the EffecT of Research on Innovation, Competitiveness and Science (STAR METRICS) initiative promised to monitor the impact of federal science investments on employment, knowledge generation, and health outcomes. The initiative—Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science, or STAR METRICS—was a multi-agency venture led by the National Institutes of Health, the National Science Foundation (NSF), and the White House Office of Science and Technology Policy (OSTP).

The expectation had been that "STAR METRICS will yield a rigorous, transparent review of how our science investments are performing," said NIH Director, Francis S. Collins, M.D., Ph.D. "In the short term, we'll know the impact on jobs. In the long term, we'll be able to measure patents, publications, citations, and business start-ups." Plans called for data for the program to come from research institutions that volunteer to participate and the federal agencies that fund them. Information was to have beeen gathered from the universities in a highly automated way, with minimal or no burden for the scientists and the university administration.

There were to have been two-phases to the program. The first phase was to use university administrative records to calculate the employment impact of federal science spending through the American Recovery and Reinvestment Act and agencies' existing budgets. The second phase was to have measured the impact of science investment in four key areas,

- ***Economic growth*** will be measured through indicators such as patents and business start-ups.
- ***Workforce outcomes*** will be measured by student mobility into the workforce and employment markers.
- ***Scientific knowledge*** will be measured through publications and citations.
- ***Social outcomes*** will be measured by long-term health and environmental impact of funding

The program also aimed to capture research activities and outputs in the form of a distributed database, making it possible to automate many administrative tasks such as creating biosketches, progress reports, final reports, and tenure reviews. However, though the STAR METRIC initiative initially generated significant participation by research institutions, generating working groups and a Steering Committee (https://thefdp.org/default/committees/star-metrics-working-group/,) it was discontinued without fanfare and without producing significant results.

2.4. Harvard Business Review 2012: The Trillion-Dollar R&D Fix

This article discusses a new metric to measure R&D productivity in the business sector that economists have created called a research quotient, which defines the relationship between a firm’s inputs and outputs. From the article’s summary,

> How does a company know what kind of return it’s getting from R&D? Is it better at R&D than the competition? How much should it be spending and what can it do to improve the effectiveness of those investments? Existing measures of R&D effectiveness don’t answer any of those questions.
>
> In this article Anne Marie Knott, a professor at Washington University’s Olin School of Management, presents a new metric for R&D productivity: RQ, short for research quotient. It allows managers not only to estimate the effectiveness of R&D investment relative to the competition but also to see how changes in R&D expenditure affect the bottom line and, most important, a company’s market value.
>
> According to her research, if the top 20 firms traded on U.S. exchanges had increased their 2010 R&D investment using the RQ method, the collective increase in market cap would have been a whopping $1 trillion. The longer-run benefits are potentially even greater because the measure allows companies to more closely link changes in R&D strategy, practices, and processes to profitability and value.
>
> This is a story that has played out before: Thirty years ago, W. Edwards Deming’s quality metrics inspired the TQM movement, which revolutionized the way companies manufacture. RQ can do the same for R&D, the author believes—and the payoff will be even greater.

Existing measures of R&D effectiveness (for example, amount of spending or number of patents) don’t answer those questions or reliably predict market value. Year after year, consulting giants point out that R&D spending does not correlate with market value or growth. Calculating RQ doesn’t involve fancy new math. Economists have been calculating capital and labor productivity for years—that is, determining the marginal value of increasing either one. R&D productivity can be determined using the same method, although few, if any, analysts or academics have done so at the level of individual companies. Essentially, the equation defines the relationship between a firm's inputs (what it spends) and its output (its revenues). The formula typically considers two costs, capital and labor, as shown in Figure 2-1.

The Theory Behind RQ

$$Y = K^{\alpha} L^{\beta}$$

OUTPUT (REVENUES) | CAPITAL | LABOR

Figure 2-1. The Theory Behind RQ (from the article.)

Good measures have three properties: universality, uniformity, and reliability. Uniformity means the measure is interpreted the same way in all contexts; universality means it applies to all relevant entities (in this case, firms); and reliability means that its predictions confirm what theory says should happen. The easiest way to explain why these properties are important is to show why another measure often used to gauge R&D effectiveness—patent counts—fails because it lacks them. RQ exhibits all three properties. RQ is estimated entirely from standard financial data, so it can be calculated for any firm doing R&D. And because RQ is a ratio, its interpretation is uniform across firms regardless of currency. Most importantly, RQ is reliable.

2.5. NIH 2014: The Usefulness and Limitations of Metrics in Measuring the Returns on Publicly Funded Research

This book chapter[4] reviews existing metrics, the uses and limitations of a commonly used input indicator and a commonly used output indicator; the challenges of data collection to inform measurement tools, with a focus on the STAR METRICS Program; the limitations of existing metrics; and the need to move beyond current indicators.

It cites a report[5] that explores methods (e.g., data mining, visualization, site visits, economic analyses), metrics, and indicators, providing examples,

- *Bibliometrics* are a quantitative measure of the quantity, dissemination, and content of research publications. They reveal the volume of outputs from the research system, and can shed light on pathways of knowledge transfer and the linkages among various scientific fields. However, the use of citations as a measure of quality or impact varies among fields and individual scientists, making this a difficult metric to apply across the research system.
- *Case studies* are useful in capturing the complex and varied inputs that influenced a particular output. This method is a valuable way to reveal the context of a discovery, but case studies often provide examples and generalizable information rather than definitively linking research to a particular output or outcome.
- *Economic analysis* can be used to understand the relationship between costs and benefits, compare possible outcomes among a range of alternative strategies, and reveal the cost-efficiency of an approach. This method is useful whenever it is possible and appropriate to assign a monetary value to both costs and benefits.

- *Logic models* provide a visual interpretation of the trajectory through which inputs contribute to a particular output, and can be useful for planning, monitoring, and evaluating research programs. The limitation of logic models is that a trajectory can change in unexpected ways. Moreover, these models tend to disregard the counterfactual, or the most likely scenario had the research program not existed.
- *Peer review* is a method based on the idea that experts in a field are best suited to determining the quality of work in that field. Some have criticized peer review for discouraging the funding of high-risk research or radically new research approaches.
- *Statistical analysis* is a valuable, albeit time-consuming way to identify patterns in existing datasets. This method depends greatly on access to and the quality of existing data.

This review of metrics for research inputs and outputs revealed them to be lacking, including

- Input Indicator: Research and Development (R&D) as a Percentage of Gross Domestic Product (GDP)
- Outcome Indicator: Return On Investment

The review further concluded that metrics used to assess any one aspect of the system of research in isolation without a strong understanding of the larger picture may prove misleading. A key challenge is to establish the most appropriate set of metrics for achieving the goal of the data collection effort. The review included an excellent summary of research impact frameworks being used or under development in the US and other nations, included herein at Table 2-1.

Table 2-1. Research Impact Frameworks Used by the United States and Other Nations (Guthrie 2013)

Framework	Origin and Rationale	Scope	Measurement	Application to Date	Analysis	Wider Applicability
Research Excellence Framework, UK	Evolved from its predecessor, the RAE, and the RQF. Intended to be low burden, but pressure from researchers led to changes. Includes wider societal impact.	Assessment at subject level on three elements: quality of research outputs, impact of research (not academic) and vitality of environment.	Assessment by subject peer review panel of list of outputs, impact statement and case studies, and statement on research environment.	Piloted 2009. First round of assessment 2014; results will determine funding allocation.	Burden not reduced, but adds wider impact to evaluation. Originally metrics based, but this was dropped as too unpopular.	Suitable for similar cross institutional assessment of performance. High burden in institutions, arguably expensive. Best for significant funding allocation uses.
STAR METRICS, U.S.	Key aim to minimize burden on academics. Helps to meet U.S. federal accountability requirements.	Two levels: Level 1, number of jobs supported; Level 2, range of research funded researcher interactions and wider impacts.	Data mining approach, automated. At present, only gathers jobs data. Methodologies for Level 2 still being developed.	Level 1 rolled out to 80 universities. Level 2 still under development. Voluntary participation so full coverage unlikely.	Feedback generally positive, but feasibility of Level 2 not proven.	Potentially very wide depending on success of Level 2. There has been international interest, e.g., from Japan, EC.
Excellence in Research for Australia, Australia	Perceived need to include assessment of quality in block funding allocation (previously volume only). Advocacy purpose to demonstrate quality of Australian research.	Assesses quality, volume, application of research (impact), and measures of esteem for all Australian universities at disciplinary level.	Indicator approach; uses those appropriate at disciplinary level. Dashboard provided for review by expert panel.	First round in 2010, broadly successful. Next round 2012, with minor changes. Intended for funding allocation, but not used for this as yet.	Broadly positive reception. Meets aims, and burden not too great. Limitation is the availability of appropriate indicators.	Should be widely applicable; criticism limited in Australian context. Implementation appears to have been fairly straightforward.
Canadian Academy of Health Sciences Payback Framework, Canada	Draws on well-established 'payback' framework. Aims to improve comparability across a disparate health research system, Covers wide range of impacts.	Five categories; advancing knowledge; capacity building; informing policies and product development; health and health sector benefits; broader economic benefits.	Specific indicators for each category. Logic model has four research 'pillars': biomedical, clinical; health services; social cultural, environmental and population health.	Used by public funders; predominantly CIHR (federal funder), but there has also been some uptake by regional organizations (e.g., Alberta Innovates).	Strengths: generalizable within health sector; can handle unexpected outcomes. But understanding needed at funder level—may limit uptake. Early stages hard to	Breadth, depth, and flexibility mean framework should be widely applicable. However, it only provides a guide and needs significant work to tailor to specific circumstances.
National Institute of Health Research Dashboard, UK	Aim is to develop a small but balanced set of indicators to support strategic decision making, with regular monitoring of performance.	Data collected quarterly at programme level on inputs, processes, outputs and outcomes for three elements: financial, internal process and user satisfaction.	Programme specific data can be pooled to provide a system level dashboard; 15 indicators selected, matching core aims, collected quarterly.	Launched July 2011 NIHR-wide, with data to be provided by the four coordinating centres, analyzed and aggregated centrally.	Designed to fit strategic objectives, so in that sense likely to be effective. However, only just launched, so detailed analysis premature.	Should be applicable to other national health research funders. Performance indicators selected can be tailored to assessment needs.

2.6. McKinsey 2015: Brightening the black box of R&D

McKinsey & Company has taken a new approach to measuring R&D outcomes by creating a formula that seemingly uses a single metric to measure outcomes. The formula, visually displayed in Figure 2-2, takes a novel approach to measuring R&D outcomes: multiplying a project's total gross contribution by its rate of maturation and then dividing the result by the project's R&D cost. The formula demonstrates several virtues. First, it's a single metric rather than a collection of them. Second, it aims to measure what R&D contributes within the sphere of what R&D can actually influence. Finally, by measuring productivity both at the project level and across the entire R&D organization (the latter through simple aggregation), it endeavors to speak to the whole company, from the boardroom all the way to the cubicle. These three elements—total gross contribution, rate of maturity, and cost of R&D—come together in a formula to quantify R&D's overall performance and shed light on separate aspects of productivity. This, in turn, facilitates more confident managerial interventions to improve them. However, measuring productivity is just a starting point—it won't change R&D's efficiency on its own. The formula must be integrated into existing management processes or lead to the creation of new ones.

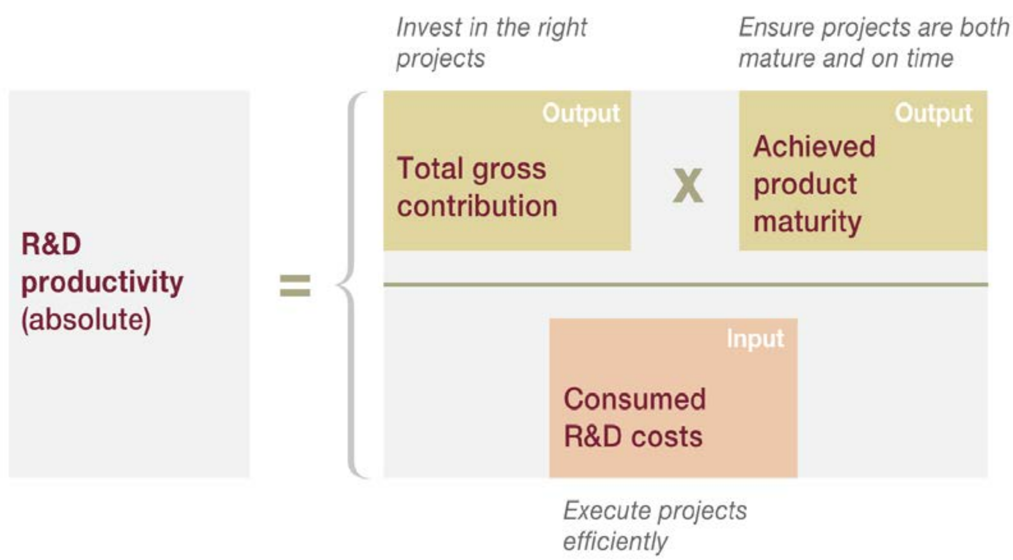

Figure 2-2. A simple formula provides companies with a single measure to assess the productivity of the R&D function (from McKinsey article.

2.7. Mehta 2015: How to Measure R&D Effectiveness?

This article (https://inspird.com/2015/04/19/how-to-measure-rd-effectiveness/) presents the traditional types of R&D metrics,

1. R&D Investments / Expenses
    - Total R&D Headcount
    - Total R&D Expense
    - R&D Expense as % of Revenue
    - R&D Expense increase/decrease from prior year
    - R&D Expense compared with peers/industry average
    - Etc.
2. Project execution status

- Performance relative to plans (costs and schedule)
- Concept to Market Time
- Number of Projects in the Pipeline
- Etc.

3. Historic results-based Metrics
   - Fraction of Revenues from New Products
   - Number of Patents generated
   - Number of Papers Published
   - Customer satisfaction with new products
   - Etc.
4. ROI-based metrics
   - Return on Innovation Investment
   - Target NPV for each new product
   - • Etc.

The article concludes that the problem with most of these metrics is that they are not actionable, as it is hard for managers to take concrete actions using these metrics to improve R&D effectiveness.

2.8. Gough 2016: Measuring the impact of R&D spending

This article (measuring the output and impact of science, technology and innovation) reviews the OECD's perspective on measuring the output and impact of science, technology and innovation, and its call for new indicators that capture the movements of scientists and the flow of knowledge, and better metrics to assess innovation. Key points include,

- Science policy in much of the industrialized world is largely underpinned by the belief that research and development investment will yield economic growth. The major indicator used to measure the intensity of this outlay is a country's gross domestic expenditure on research and development (GERD), as a percentage of its GDP. The OECD average is 2.38%, which includes government, university and private investment. The European Union has a target of 3% by 2020.
- There have been a number of rigorous studies showing direct relationship between spending on research and scientific output
  - Countries that spent more on R&D, and had more universities and indexed journals, produced more high-quality research publications across the sciences and social sciences.
  - Public investment in R&D increases private sector investment in science and attracts foreign investment.
  - But a link between R&D spending and the translation of that research into commercial outcomes is less clear.
- Australia Case Study showed key problem with this approach,
  - Producing high-quality research publications – it was 12th in the Nature Index 2015 Global ranking – but struggles to reap commercial and wider socio-economic rewards.
  - Ranked 11th in research inputs, but 73rd in innovation efficiency ratio, which measures how much output or innovation success a country has had relative to its inputs, such as knowledge creation and number of researchers.
  - Part of the problem lies in the transfer of knowledge from campuses to companies. In 2013, Australia ranked 29th out of 30 OECD countries on industry collaboration with universities and public research organizations.

- Another major challenge with measuring the scientific output from R&D investments is timing: "the measurable outputs of the scientific process don't appear immediately after funding, teasing out the causal effect of any specific R&D investment is difficult."
  - Research into the renewable energy sector in the US found that patent applications began three years after funding and continued for up to 15 years.
  - In medicine, the lag from initial public R&D investment to the development of new drugs can exceed two decades.
- Another challenge is selection bias in funding schemes. People who receive grants from the government publish papers afterwards, but it is uncertain whether the government grants helped them. "Or did they get the grant because they are star researchers who likely would have published anyway?"
- The definition of scientific output is another issue for the scientific community. Traditionally, funding agencies and governments have looked at what can be counted, or reported anecdotally: papers, citations, patents, and success stories. Some have suggested measures that more accurately account for research training, human networks, idea transferring, and even research failures need to be developed to account for the true value of the scientific process.

2.9. Ståhle 2016: R&D Measurement Methods and Models

This work, R&D Measurement Methods and Models,[6] developed a measurement tool for evaluation of efficiency, productivity and impact of R&D projects, forming a ground for further development of the R&D measurement methodology and practice. It states that the top five metrics used for R&D are the following: R&D spending as a % of sales, R&D headcount, Current year % sales due to new products released in the past N-years, # Patents filed/pending/awarded/rejected, and # of new products released. However, none of these really gives a good idea of how well the R&D function is performing, since the metrics give no information of real efficiency of R&D. Engineering utilization, productivity and throughput have also been defined as most important metrics for measuring R&D perform. In the end, it indicates that a universal set of R&D metrics does not exist.

Since R&D investments compete, it must prove its productivity to its stakeholders. Productivity refers to the relation between output and input. Inputs are the resources used in the production (e.g. labor, materials and services) and outputs are products, services or both. Productivity involves two dimensions, internal and external efficiency. Internal efficiency shows how well the work is done and exposes efficiency of project operations: e.g. methods, routines and functions. External efficiency shows what has been achieved. It exposes the results, i.e. financial and other kind of outputs and impacts. Internal efficiency is important for development of R&D functions, and external efficiency is important to motivate R&D investments within a company from strategic and shareholder perspectives. The relationships among these concepts is shown in Figure 2-3 (Figure 1 from this Section's reference.)

Figure 2-3. Dimensions of Project Efficiency (Figure 1 from the Section reference.)

Since efficiency is never dependent only on one variable, but instead based on a comprehensive process, or even the system as whole, R&D measurement must cover all the phases of the process and/or elements of the system,, such as that from Brown and Svenson,[7] which includes the following elements: input, process, output, implementation, and outcome (impact), as shown in Figure 2-4.

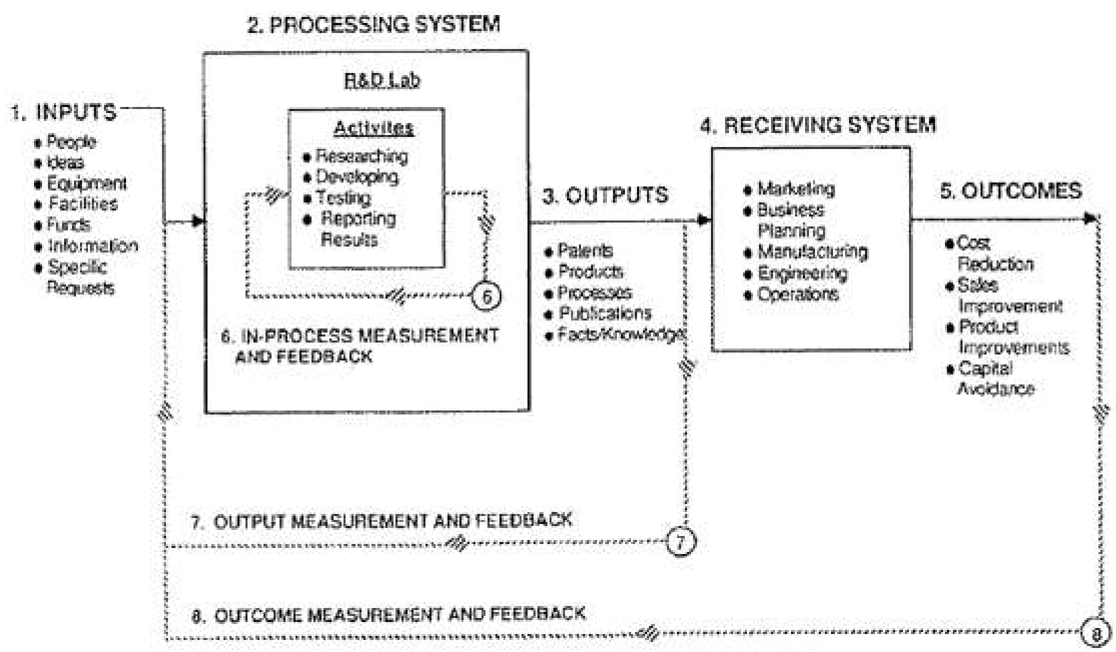

Figure 2-4. R&D System by Brown & Svenson (1988)

2.10. RAND 2017: Accelerating the Transition of Technologies Created through the U.S. Department of Defense Small Business Innovation Research Program

This study[8]of the success of transition of technologies generated through the DOD SBIR program used four indicators for the success of technology transition,

1. Action obligation of federal procurement contracts (FPC) – the money value of federal procurement contracts (measure for the effectiveness of the SBIR program)

2. Return on Investment (ROI)[2] – measure for the efficiency of the program
3. Jobs created (JOB) – the average change in the number of employees per year (measure for the economic performance of the program)
4. Patents filed (PAT) – the number of patent applications (measure for the technological performance of the program)

The first metric directly measures technology transition (in terms of the SBIR program's effectiveness) while others indirectly capture technology transition (in terms of the program's technological and economic impacts). In particular, the second (ROI), third (JOB), and fourth (PAT) metrics shed light on the efficiency, economic, and technological performance, respectively. Those metrics have been widely employed in the evaluation of R&D programs[9] and are also suggested by the SBIR/STTR Interagency Policy Committee.[10]

2.11. Zemlickienė 2022: Performance Measurement in R&D Projects: Relevance of Indicators Based on US and German Experts

This article[11] discusses indicators for R&D project performance measurement, especially relevant for measuring project performance where the project is perceived as an independent business unit. It selects indicators for measuring the performance of R&D projects and identifies and compares their relevance among US and German experts. After the experts' selection, based on using multi-criteria decision making (MCDM), the finalized list of indicators is presented in Table 2-2.

Table 2-2. R&D Project indicators from Reference 2

R&D Project Indicators
1. Profit due to R&D/Profits from new product sales
2. Project cost versus budget/Average development cost per product/Total cost of new product effort as % of revenue/R&D spending as a percentage of sales
3. Realized IRR/ROI/Net margin ROI
4. New product sales
5. Customer satisfaction with new products/market response/% of products succeeding in the market/New product acceptance rate
6. Market share gained due to R&D
7. Products completed speed/the length of time spent on each development activity in each NPD stage/actual project schedule versus planned
8. Product life cycle in the market
9. Technical quality of the project execution/maturity of the project

At the early stage of development, the performance of R&D activities is measured as the Technology Transfer Office (TTO) outcome, with the following metrics,

1. Number of invention disclosures

[2] Return on investment (ROI) used in this reference does not exactly correspond with one used in finance, instead following the new concept of ROI proposed by the SBIR/STTR Interagency Policy Committee (SSIPC): ROI = (Amount of federal procurement contracts – Amount of SBIR awards)/Amount of SBIR awards, from the perspective of not private sector but public sector.

2. Number of patent applications
3. Number of patents granted
4. Number of licenses signed
5. Number of start-ups formed
6. Research expenditure of university scientists
7. Expenditure of patenting activities
8. Operation expenditure
9. Number of new commercial products
10. Employment and productivity growth of start-up partners
11. Changes in stock prices of industrial partners, etc.

3. Data Sources for R&D Indicators

3.1. National Science Board (NSB)

The National Science Board (NSB) (https://www.nsf.gov/nsb/) provides the Science and Engineering Indicators data tool, presenting indicators and their trends in S&T education, employed workforce, R&D, patents and publications, and knowledge- and technology-intensive industries, down to the state level. The indicators are divided into six categories, of which three are relevant to R&D IoS, including,

- **Financial Research and Development Inputs (Figure 3-1)** – Financial indicators present the sources and level of funding for R&D. The indicators show how much R&D is being performed relative to the size of a state's economy and workforce. They also present the extent to which R&D is conducted by industrial and academic performers, covering two topics,
  - Level of R&D performed by businesses, universities, and other performers
  - Public-sector support for R&D activities

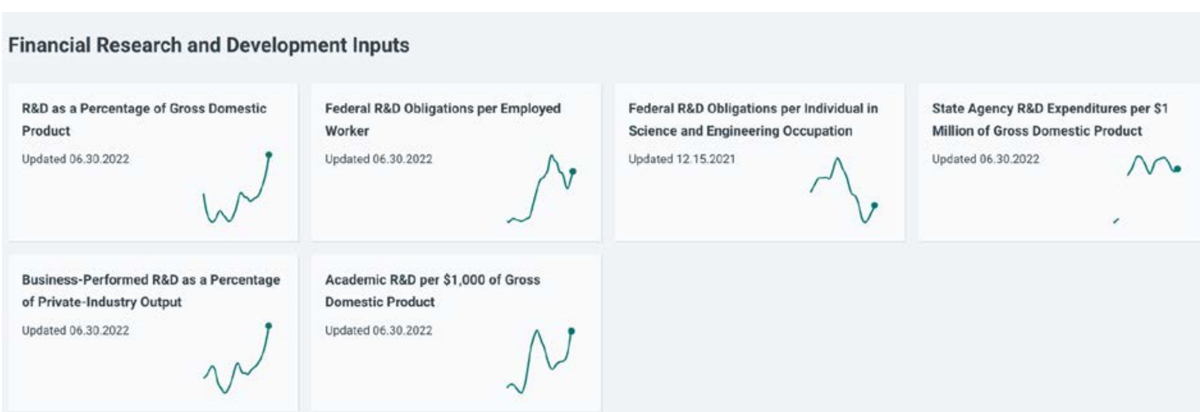

Figure 3-1. Financial Research and Development Inputs

- **Research and Development Outputs (Figure 3-2)** – These indicators show the number of new doctorates conferred, the publication of academic articles, and patent activity from the academic community and from all sources in the state. Outputs are measured relative to states' S&E workforces, academic workforces, and financial R&D inputs, covering two topics,
  - Human capital outputs

- Research-based outputs

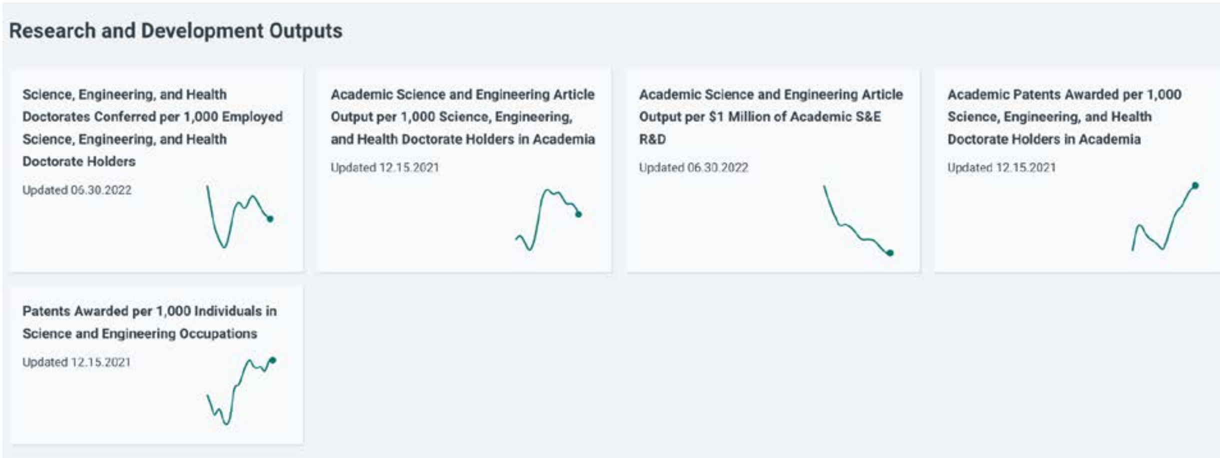

Figure 3-2. Financial Research and Development Inputs

- **Science and Technology in the Economy (Figure 3-3)** -- These indicators include venture capital activity, Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) awards, and business activity in industries with high employment in knowledge- and technology-intensive fields, covering two topics,
  - Business activity in industries with high employment in knowledge- and technology-intensive fields
  - Early-stage, high-risk capital investments

Figure 3-3. Financial Research and Development Inputs

3.2. Science & Technology Indicators of the Technology Observatory of the University of Alicante (OVTT UA)

The Technology Observatory of the University of Alicante (OVTT UA) (https://www.ovtt.org/en/guidelines/research-development-indicators/) offers sources of data for the following indicators as most relevant for science, research and innovation,

1. Context indicators, such as R&D policies and investment (public and private) or technology markets evolution, among others
2. Science and Technology indicators

3. Patent indicators
4. Innovation indicators
5. Bibliometric indicators
6. Indicators on R&D talent
7. Indicators on scientific culture and engagement
8. Indicators on university-industry cooperation
9. Indicators on entrepreneurship
10. Indicators on training and mobility

The main organizations dedicated to the production of data on science, technology and innovation indicators with international, regional and national scopes are listed on the reference website, and include,

- OECD
- UNESCO
- CEPAL
- RICYT
- IDB
- EUROSTAT
- OCTS-OEI

4. Summary of Indicators of Success (IoS) Characteristics

Bibliometrics, case studies, operational and economic metrics have been used most commonly as R&D output success metrics of R&D outcomes after commercialization, and may be able to be adapted as Indicators of Success (IoS) for R&D transition outcomes. Tables 4-1 and 4-2 summarize the findings of the literature review on these IoS. Other IoS, such as peer review, patent counts, job creation, etc., can also be adapted for S&T use, as described in the main body of this report.

Table 4-1. Bibliographic Methods for Measuring Outcomes

Bibliographic Method	Description
Publications or Citations	Count of publications or citations in peer-reviewed journals. Citation counts in particular are a traditional and important measure of impact. https://clarivate.libguides.com/authors/impact
Journal Impact Factor	A factor that determines the importance or rank of a journal by measuring the frequency that the average article in a journal has been cited that year. https://researchguides.uic.edu/if/impact
H-Index	The H-Index (Hirsch Index) measures productivity and scientific impact of a researcher or scientist. The H-index calculates the number of published papers that have each been cited at least h amount of times in other peoples' publications. An H-Index can be obtained by using the Web of Science, Scopus, or Google Scholar. https://researchguides.uic.edu/c.php?g=252299&p=1683205
ISI or Web of Science	Publication service that provides citation counts for articles indexed within it. It indexes over 12,000 journals worldwide. https://researchguides.uic.edu/c.php?g=252299&p=1683205

Scopus	Provides citation counts for articles indexed within it (indexes over 15,000 journals from over 4,000 publishers). https://researchguides.uic.edu/c.php?g=252299&p=1683205

Table 4-2. Summary of Indicators of Success (IoS) Characteristics

IoS Characteristic	Bibliometrics	Case Studies	Operational Indicators	Economic Indicators
Description	A set of mathematical and statistical methods used to measure the quantity and quality of knowledge products. Bibliometrics can be broken down into three categories (quantity, quality, and structural) to separately measure productivity, quality of output, and connections for more information on the publication.	A case study is a research approach used to create an in-depth and well-developed understanding of an issue in its real-life context. Many R&D impact evaluations qualify as case study research designs themselves. (Somewhat similar to operational indicators but only measures operations in a given setting)	Operational indicators encompass all operational improvements/changes due to an invention. Operational indicators vary widely.	ROI (Return on Investment) is a strong economic indicator that attempts to link R&D with monetary gain. ROI can provide a measurement of investment vs. profits. Econometrics is the application of statistical methods to economic data in order to give information on economic relationships. Input Indicator - Research and Development (R&D) as a percentage of Gross Domestic Product (GDP) Outcome Indicator - ROI
Strengths	- It is to quantify citation counts and papers published - It is neutral which allows for comparative assessment - Has shown promising results in measuring success over other indicators	- Many R&D impact evaluations already qualify as case studies, making it easy/timely to collect information on this indicator in these cases - Can allow for a more realistic and relevant measure for the specific use/operational context	- Indicators are specific to the invention and therefore typically more relevant - Chosen operational indicators are typically fairly easy to measure in terms of output	- ROI addresses one of the fundamental concerns about the numerical worth of R&D investments - Quantifiable and neutral so allows for comparative assessment

IoS Characteristic	Bibliometrics	Case Studies	Operational Indicators	Economic Indicators
Limitations	- Struggles to quantify the value of citations (ex: how much more/less is a journal paper "worth" compared to a conference paper?) - It is difficult to interpret the level of innovation and impact that has occurred as the result of a single knowledge product - Focusing on the quantitative aspect of bibliometrics rather than the qualitative aspects as well can inflate numbers quickly	- Can be difficult to generalize findings from one setting to definitively link research to a particular output or outcome - Detail and rigor of case studies can vary, making it difficult for comparative analysis amongst inventions/products	- Operational indicators can vary widely in which indicators are chosen which can make it difficult for comparability. However, amongst different inventions, if similar measurements are chosen, comparability is possible.	- Econometric studies typically underestimate the contribution of R&D, especially since it can be difficult to trace spillover effects - Complexity of the innovation process and its inherently long time frames pose obstacles to the calculation of ROI
Measuring Output	Output in this case would be the number of relevant publications produced by a source.	Output would be the immediate result of the case study (for example, a case study testing a new vehicle may have an output that involves the speed of the vehicle during the case study)		
Measuring Outcome	Outcome could be measured by the amount of reads or views a publication receives. This is more of a short-term measure and the next step would be to measure the actual impact of these reads/views through citation analysis (see below).	Outcome could be measured as the immediate improvement/change due to the output measured during the case study (in the example above, the outcome would be that the vehicle drove x times faster during the case study		

IoS Characteristic	Bibliometrics	Case Studies	Operational Indicators	Economic Indicators
		than the previous design did).		
Measuring Impact	Bibliometrics assess the quantitative impact of a given publication by measuring the amount of times a certain work is cited by other resources. - Journal Impact Factor: a way to measure impact that is calculated by averaging the number of citations per paper published in a journal during the two preceding years. - H-index: A way to measure impact of a researcher that is calculated as the published number of papers (h) each of which has been cited in other papers at least h times.	Impact would refer to the long-term changes that result from the invention. A case study may struggle to identify these long-term effects unless the case study itself lasts over a longer period of time or researchers pull from past, related case studies to estimate the impact of a similar product. (In the example mentioned previously, more of a long-term impact may be that all of the organization's vehicles are updated with the parts the new vehicle is made of.)		
Measuring Value	The value of a publication can be difficult to measure but can possibly be determined with more qualitative measurements such as peer review.	Value would be measured as how much a given product made progress towards its final goal (the final goal in the example case study used may be to catch more people). Some type of quantifiable metrics and		

IoS Characteristic	Bibliometrics	Case Studies	Operational Indicators	Economic Indicators
		equations would likely need to be used to develop a numerical value based on the outcome and impact of an invention.		
Examples	Number of publications or citations	Testing inventions in their appropriate context, interview case studies	Performance improvements, reduction of effort, speed/optimization of processes	R&D spending, dollars saved, lives saved, Cost-benefit analysis

5. References

1. OECD (2010). Glossary of Key Terms in Evaluations and Results Based Management. OECD, 2002, re-printed in 2010.
2. Measuring Performance, Strengths and Limitations of Research Indicators, GAO/RCED-97-91, March 1997, https://www.gao.gov/assets/rced-97-91.pdf.
3. National Academies of Sciences, Engineering, and Medicine. 1997. Industrial Research and Innovation Indicators: Report of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/5976.
4. Committee on Assessing the Value of Research in Advancing National Goals; Division of Behavioral and Social Sciences and Education; National Research Council; Celeste RF, Griswold A, Straf ML, editors. Furthering America's Research Enterprise. Washington (DC): National Academies Press (US); 2014 Oct 28. 4, The Usefulness and Limitations of Metrics in Measuring the Returns on Publicly Funded Research. Available from: https://www.ncbi.nlm.nih.gov/books/NBK253898/
5. Guthrie S, Wamae W, Diepeveen S, Wooding S, Grant J. Measuring Research: A Guide to Research Evaluation Frameworks and Tools. Santa Monica, CA: RAND Corporation; 2013. [December 2013]. http://www.rand.org/pubs/monographs/MG1217.html.
6. Pirjo Ståhle & Sten Ståhle, R&D Measurement Methods and Models, Fingrid Measure 1.0 / Phase 1, Review and analysis of R&D measurement approaches, Version 12.9.2016
7. Brown, W.B. and Gobeli, D. (1992) Observations on the Measurement of R&D Productivity: A Case Study. IEEE Transactions on Engineering Management, 39, 4, 325-331.Brown & Svenson (1988).
8. Ryu, Youngbok, Accelerating the Transition of Technologies Created through the U.S. Department of Defense Small Business Innovation Research Program. Santa Monica, CA: RAND Corporation, 2017, https://www.rand.org/pubs/rgs_dissertations/RGSD405.html.
9. Lott, Melissa C., "Measuring "Success" in R&D," Scientific American, 2011, https://blogs.scientificamerican.com/plugged-in/measuring-success-in-rd/
10. SBIR/STTR Interagency Policy Committee, The Small Business Innovation Research (SBIR) & Small Business Technology Transfer (STTR) Program Interagency Policy Committee Report to Congress: SBIR/STTR Standard Evaluation Framework, Small Business Administration.
11. Zemlickienė, Vaida, and Zenonas Turskis. “Performance Measurement in R&D Projects: Relevance of Indicators Based on US and German Experts.” Sustainability, vol. 14, no. 18, 2022, p. 11737., https://doi.org/10.3390/su141811737.