1

Introduction and Motivation

Kaye Husbands Fealing, Julia I. Lane, John L. King, and
Stanley R. Johnson

1.1 Overview

In the United States, improving the safety of the food supply has become a national priority, and food safety research has been identified as central to achieving that goal. Yet, little is known about answers to key questions, such as: What research is already being done in the field? How many researchers are active in food safety research? What are the characteristics of those researchers? How do federal research funding patterns affect current workforce development and future research capacity? What are the reciprocal influences between food safety issues and federally funded research? In short, what are the key ways in which federal investment in food safety research funding will affect the research pipeline?

Of course, these questions are not unique to food safety research, but this type of research is particularly interesting because of the diversity of scientific fields and funding sources (including agricultural, health, and veterinary) and the diversity of economic actors involved in agriculture, food production, storage, and the movement of food safety risks across domestic and international jurisdictions. Further, a continually evolving dynamic relationship exists between private-sector agriculture (including food production interests) and public-sector food safety research. To a large degree, these are scientifically complementary, with each entity exerting influence in the policy arena.

In addition, the importance of the field is undeniable. The Centers for Disease Control and Prevention (CDC) estimates that more than 48 million individuals in the United States alone – one in every six – will get sick from a foodborne illness. Many of these foodborne illnesses will pass unacknowledged as generalized discomfort. Many will be more severe, resulting in lost time from work. Others will result in permanent disabilities or even death. The CDC estimates that 128,000 cases of foodborne illness will require medical treatment and 3,000 individuals will die every year. The literature on the economic burden of foodborne illness is estimated as up to $77 billion annually (*1*). The US Department of Agriculture (USDA) estimates that just 15 pathogens account for more than $15 billion of economic burden from treatment, lost work, morbidity, and mortality, and this does not include other nonpathogenic sources of food safety risk such as food contaminants. Moreover, food safety is an issue of international scope: The total impact of foodborne illnesses is orders of magnitude higher than the effects in the United States alone, with incidence and impact higher in other countries and especially so throughout the developing world. In response to this, important policy changes have taken place in the field of food safety. Most significant is the legislation – the Food Safety Modernization Act of 2011 – which contains provisions designed to enhance the coordination of food safety research. Implementation of the act will affect long-standing research programs at federal laboratories, universities, hospitals, and other research institutions. Appendix 2.2 in Chapter 2 reviews the laws and regulations in the food safety industry.

A 2012 report published by the President's Council of Advisors on Science and Technology (PCAST) called for "creation of a new innovation ecosystem for agriculture that leverages the best from different parts of the broad US science and technology enterprise." In that report, PCAST recommended an annual increase in "investment" in agricultural research of $700 million, with suggested allocations to new graduate and postdoctoral fellowships ($180 million), new competitively funded research at the USDA's Agriculture and Food Research Institute (AFRI; $235 million), basic research at National Science Foundation (NSF; $130 million), and new private-public institutes ($150 million). The PCAST report is illuminating for two additional reasons. It notes that (1) mechanisms are needed for distributing funds to earn their highest return and (2) returns are not merely economic but also include the increase in human capital (or talent) developed as research is conducted. There is also an important role for food safety research because the sheer ubiquity of food consumption poses risks and creates opportunities for food safety science to reduce those risks.

Yet simply investing in research is not sufficient. The PCAST report highlighted lingering questions about the "appropriate allocation of research funds and whether they could be better spent on research challenges that are not a strong focus of the private sector" (p. 36). Of course, the lack of information about the impact of research is not confined to the field of food safety. In a speech titled "Why Policy Implementation Needs a

Science of Science Policy," John H. Marburger III voiced frustration that policymakers were not asking the right questions nor were they provided with sufficient evidence to formulate effective science policy. "How much should a nation spend on science? What kind of science? How much from private versus public sectors? Does demand for funding by potential science performers imply a shortage of funding or a surfeit of performers?" Marburger called for a new "science of science policy" to emerge, offering compelling guidance for policy decisions (*2*). He also noted that if there were to be better management of the national science and technology enterprise, then the practice of science policy must be professionalized.

Fortunately, that new "science of science" policy has emerged and is what forms the basis of the work in this book. That science is based on integrating new data on all steps of the research process, from the funding inputs to the outputs and consequences, by taking advantage of data from the federal statistical system. These new data, which are called the UMETRICS data (*3*), are what the authors build on in this book. This approach builds up from data at the level of individuals who conduct research – data that have broader economic and social impacts. The data provide answers to questions about the results of federal funding in the agricultural sector, particularly establishing mechanisms for assessing the impacts in food safety sectors. These are some of the key questions that must be answered for effective use of public resources to achieve food safety goals.

In sum, the work in this book seeks to answer some of the most important questions that are necessary to improve public policy about food safety research. This book describes new data and techniques that will enable, for the first time, a detailed examination of the outcomes of federally funded research in the agricultural sector generally and scientific outputs and outcomes related to food safety in particular. As such, this book provides a novel template that the science of science policy community can use to assess the impact and value of research that extends to other scientific fields. Of course, as with any research, much work remains to be done to characterize the full complexity of the impact of scientific research – and this book provides the first steps along a new pathway to do so.

1.2 Science of Science Policy: The Research Framing for This Book

A 2009 Pew Research Center Survey found that almost three-quarters of Americans agreed that government spending on basic scientific research, as well as on engineering and technology, "usually pays off in the long run." The same survey also found that roughly 60 percent of Americans said that "government investment in research is essential for scientific progress," while almost one-third said that "private investment will ensure that enough scientific progress is made, even without government investment." That year, private-sector firms and government agencies spent roughly 3 percent of total output in the United States on research and development (R&D). Federal expenditure on R&D was $133 billion, with about 25 percent of that spent on basic research. Almost half of the nondefense R&D budget went to basic research. Arguably, these expenditures advanced science, which in turn affected social outcomes, such as national security, health outcomes, food safety and security, energy and natural resource use, transportation, communication, and education.

However, estimates of the impact of science, technology, and innovation on society (from both the government and private sector) are typically based on multipliers and other proximate values. The calculation of economic returns, such as financial earnings from patent licenses, commercialized products, and spinoff companies, have typically been one-off approaches to assessing the benefit streams of expenditures on science. The calculation of scientific returns has often been based on counting the papers generated by researchers – a field known as *bibliometrics*. However, these measures suffer from several flaws. First, they do not strictly identify the outputs generated by any specific stream of funding. Second, the gross measures ignore the obvious necessary comparison: What is the additional output from these expenditures beyond what would have occurred given the status quo? Furthermore, these measures of outputs from research activities do not go far enough to measure the social impacts of research. The public wants to know how much their tax dollars contribute to improvements (or retrenchments) in social well-being. Assessing the public value of science and technology, therefore, is a critically important activity, because without such assessments, the collective citizenry would not be able to grasp the return on their "investments" in the scientific enterprise (*4*).

The lack of data on the impact of science expenditures has been a major impediment for some time for an informed decision-making process among both policymakers and legislators alike. Indeed, the 2008 White House Science of Science Policy Interagency Task Group undertook a literature review to determine the state of the science to date. The Task Group circulated a questionnaire to federal agencies to ascertain what methods are currently being used for programmatic investment decision making, as well as to ask what tools and resources federal agencies need that are currently unavailable. The Task Group found the following:

- A well-developed body of social science knowledge exists that could be readily applied to the study of science and innovation.
- Although many federal agencies have their own communities of practice, the collection and analysis of data about the science and scientific communities they support is heterogeneous and unsystematic.
- Agencies are using very different models, data, and tools to understand their investments in science and technology.
- The data infrastructure is inadequate for decision making. (*5*)

Historically, most of the estimates that were used for estimating the impact of science expenditures came from the Bureau of Economic Analysis's RIMS II model, which was derived from a decades-old input–output model of spending flows (and mostly uses national coefficients for industry or locally specific application). This approach also functionally equates the impact of science to the impact of building a football stadium or an airport: The impact is derived from the demand side and depends on the amount of spending on bricks and mortar and workers (*6*).

There are several challenges to building a better data infrastructure. The first is that the US scientific data infrastructure is oriented toward program administration rather than empirical analysis. The result is that the agencies primarily responsible for funding science operate in different data silos, with different identifiers, different reporting structures, and different sets of metrics (*5*). The second is that the focus of data collection is on awards, which are not the appropriate unit of behavioral analysis. Awards are the intervention of interest, and the activities of the scientists who receive the awards are what need to be followed. In other words, awards are temporal, but knowledge generation and resulting innovation require studying the activities of the objects of continuous analysis: scientists and their scientific interaction with other scientists. A third reason is that the current data infrastructure does not allow science expenditures to be coupled with scientific and economic outcomes. In particular, Grants.gov provides a unified portal to find and apply for federal government grants, but goes no further. Research.gov and Science.gov provide information about R&D results associated with specific grants, and a consortium of federal agencies provides R&D summaries (www.osti.gov/fedrnd). Another obvious challenge is the fact that the reporting system is manual (with obvious quality implications) and relies on principal investigators to make reports during the active period of the award – even though the impacts of science expenditures are often unknown until many years after the award has ended. Finally, despite the fact that funding agencies believe that their impact includes both workforce and social impacts, there is no systematic tracking of the students supported by federal funds. A previous effort to collect R&D information on federal awards, RADIUS, was discontinued in 2006.

The need to do better is compelling. There are continuing demands for evidence-based decision making on the part of research agencies, and agencies are eager to find methods that more accurately measure outcomes and impacts of their outlays (*7*). Traditional estimates are useful for comparative analyses, provided that counterfactuals are properly stated and measurable.

1.3 The Contribution of This Book

The goal of this book is to build a better understanding of how returns to research are generated, focusing mainly on data-intensive methodologies. As Daniel Kahneman has noted, the first big breakthrough in our understanding of the mechanism of association was an improvement in a method of measurement (*8*). The authors believe that this work will provide a new pathway for informing the link between research expenditures and research outcomes by building data at the most granular level possible: the project level.

This book builds on a vast literature on productivity growth and the social rate of return on expenditures on R&D in agriculture based on macro- or industry-level data. A number of important articles made great strides in using these more aggregated data to assess the economic returns to government expenditures: benefit-cost, risk and multiplier analyses, as well as econometric methods used to calculate multifactor productivity indexes (*9–14*). However, the resultant estimates of returns to research expenditures that use those techniques vary widely, particularly given the broad range of assumptions used to model relationships within the system. These measures are also highly aggregated, and they are most accurate for very near-term outputs from R&D expenditures. It is very difficult to quantify the longer-term impacts or spillover effects, at least partly because of the meso- or macro-level of the data analyzed.

This book's contribution provides a more granular approach. It exploits project-level data at a detailed temporal level to begin to describe *what* is funded, *who* is doing the research, and *what* the results are. At the heart of this methodology is the innovative UMETRICS approach of tracing research funding, which makes use of new computational tools to tie together disparate datasets (*15*). Chapter 3 spells out the approach in more detail; the approach uses natural language processing to describe (1) what research is being done, using proposal and award text to identify the research topics in a portfolio. Administrative records at universities and funding agencies describe (2) who is doing the research on federally supported grants on food safety and with whom. This is possible because of data drawn directly from payroll records, which also have the occupational classifications of each individual employed – including graduate students. This enables a characterization of the variety of occupational categories directly supported by agency funding. Finally, in response to the question of (3) what are the results, this research creates analytical links between researchers funded to do work on food safety and US Census Bureau data on earnings and employment outcomes. This approach represents a marked departure from the bulk of work seeking to quantify the results of research insofar as it focuses on the outcomes of the people who are involved in research projects as opposed to bibliometric method (i.e., counting the publications written).

Several methodological contributions go beyond the application of computational science to characterizing food safety research. One is that the approach focuses on the activities of not just principal investigators, but also the postdocs, graduate students, and undergraduate students working on food safety, as well as those working in related fields such as microbiology, zoology, epidemiology, and chemistry. This permits the construction of comparison groups. Another is the ability to construct direct measures of the way in which research funding supports research teams – this is particularly important given that science is increasingly being done by teams. A third contribution is the matches to outside datasets, which enable the capture of an important subset of the activities of researchers after the receipt of research funding – such as their PhD dissertations and their placement and outcomes. These sources are used to describe what results the funding has generated. Chapter 3 of this book describes the conceptual framework and data infrastructure used to assess the results of investments in food safety research.

1.4 Audience for the Book

There are multiple audiences for this study, both general and specific. First is the public. Federal research spending costs every man, woman, and child in the United States more than $200 a year. The returns to that spending are neither well documented nor well understood. This book shows how to trace the public value of investments in basic and applied research, with a particular focus on an area of great public interest – food safety. The second audience consists of funding agencies. The framework developed here should lead to a better understanding of the pathways to impact resulting from the investment of money in research. Third, university administrators can build on the data infrastructure at their own institutions to better understand the structure of research activities at their institutions. Fourth, researchers who work on science and innovation policy issues will benefit from the data infrastructure that has been created in the process of doing this study. The administrative records linked to Census data and to dissertation and patent databases should provide a fertile field for research in multiple areas.

In the specific area of food safety, policymakers in agricultural, science, and technology policy agencies should be able to benefit from the ways in which this study traces economic impact. The work provides new insights into the nature of food safety research, the composition of the existing and future workforce, and the pathways whereby food safety researchers connect to the larger economy.

1.5 The Plan of the Book

Chapters 1–3 introduce the conceptual premise of this book. Chapter 2 presents information about the nature of the food safety system in the United States as it is currently organized and regulated, which is quite complex, fragmented, and prone to obsolescence based on unanticipated events. It also provides a synthesis of the results of a workshop in which stakeholders from across the food safety research and food production chain provided input, and participants in that workshop produced two white papers. Chapter 3 describes the conceptual and empirical framework used for food safety research throughout the book.

Chapters 4 and 5 provide an in-depth discussion of new analytical and empirical techniques for describing research. Chapter 4 describes the fundamental step of identifying publicly funded food safety research from open records using computational techniques. Chapter 5 describes the structure of research funding in the sample of research institutions for which data exist, as well as the effects of different assumptions about food safety definitions on the scope of the research field.

The focus of this book then turns to an analysis of food safety research on the researchers and the research teams carrying it out. Chapter 6 begins by focusing on the individual researchers. It describes the way in which the data can be used to characterize who is doing food safety research, then matches these data to Census Bureau data to characterize the demographics of the food safety research workforce. Chapter 6 also describes how it is possible to use these new data to construct a control group of individuals that can be used as a comparison for investments in food safety research. Of course, since science is increasingly done in teams, one can also use the data to describe the structure of teams and their links to other areas of research; that is the focus of Chapter 7.

The book then turns to documenting the results of food safety research, using both traditional and nontraditional frameworks. Chapter 8 focuses on early career outcomes of graduate students and postdoctoral scholars who participate in federal research awards as part of their training. This analysis allows for employment and earnings effects of federal funding to be determined, compared with carefully constructed comparison groups. Chapters 9 and 10 examine patent and publication activity. While it is understood that patents are not a critical vehicle for the transfer of new knowledge in the food safety innovation ecosystem, the analysis in Chapter 9 does address the following questions: (1) What has happened to the pace and direction of patenting in the food safety sector? (2) What are the characteristics of US and foreign firms that are most active in food safety patenting? (3) What are the geographical and sectoral distributions of food safety patents? Chapter 10 follows with an analysis of scientific papers, which are an important source of policy governance. The methods employed in that chapter use new computational approaches designed to address two major weaknesses of traditional bibliometric analysis: (1) the limited coverage (and bias) of analyzed literature, due to the limitations of existing databases that tend to include a specific set of journals and subjects (interesting to their primary readership) and (2) the high cost of running a large-scale qualitative analysis of retrieved publications. Chapter 11 provides both a conclusion and a look forward to a future research agenda.

References

[1] R. L. Scharff, Economic Burden from Health Losses Due to Foodborne Illness in the United States. *J. Food Prot.* **75**, 123–131 (2012).

[2] J. H. Marburger, Wanted: Better Benchmarks. *Science (80-.)* **308**, 1087 (2005).

[3] J. Lane, J. Owen-Smith, R. Rosen, B. Weinberg, New Linked Data on Science Investments, the Scientific Workforce and the Economic and Scientific Results of Science. *Res. Pol.* **44** (9), 1659–1671 (2015).

[4] K. Husbands Fealing, "Public Value of Science and Technology." Humphreys School of Public Affairs, University of Minnesota, Working paper (2012).

[5] National Science and Technology Council, "The Science of Science Policy: A Federal Research Roadmap" (National Science and Technology Council, Science of Science Policy Interagency Task Group, Washington, DC, 2008).

[6] J. Lane, Assessing the Impact of Science Funding. *Science (80-.)* **324**, 1273–1275 (2009).

[7] K. Husbands Fealing, J. Lane, J. Marburger, S. Shipp, *The Handbook of Science of Science Policy* (Stanford University Press, 2011).

[8] D. Kahneman, *Thinking Fast and Slow* (Farrar, Straus and Giroux, 2011).

[9] Z. Griliches, Research Cost and Social Returns: Hybrid Corn and Related Innovations. *J. Polit. Econ*. **66**, 419–431 (1958).

[10] Z. Griliches, Productivity, R&D, and the Data Constraint. *Am. Econ. Rev*. **84**, 347–374 (1994).

[11] D. W. Jorgenson, F. M. Gollop, Productivity Growth in US Agriculture: A Postwar Perspective. *Am. J. Agric. Econ*. **74**, 745–750 (1992).

[12] J. M. Alston, P. G. Pardey, Attribution and Other Problems in Assessing the Returns to Agricultural R&D. *Agric. Econ*. **25**, 141–152 (2001).

[13] J. M. Alston, M. A. Andersen, J. S. James, P. G. Pardey, The Economic Returns to U.S. Public Agricultural Research. *Am. J. Agric. Econ*. **93**, 1257–1277 (2011).

[14] J. Mullen, Productivity Growth and the Returns from Public Investment in R&D in Australian Broadacre Agriculture. *Aust. J. Agric. Resour. Econ*. **51**, 359–384 (2007).

[15] I. Foster, R. Ghani, R. S. Jarmin, F. Kreuter, J. I. Lane, *Big Data and Social Science: A Practical Guide to Methods and Tools* (Taylor & Francis Group, 2016).