The Therapeutic Consequences of the War: World War II and the 20th-Century Expansion of Biomedicine*

Daniel P. Gross†
Duke University and NBER

Bhaven N. Sampat‡
Johns Hopkins University and NBER

August 2025

Abstract:

During World War II, the U.S. Committee on Medical Research (CMR) undertook an integrated, cross-sectoral effort to develop medical science and technology for war, representing the U.S. government's first substantial investment in medical research. Although it had mixed results during the war, using data on all CMR research contracts we show that it laid a foundation for the postwar takeoff of the U.S. biomedical innovation system. New and emerging research areas it supported experienced rapid growth in postwar science. It also fueled new postwar drug development, influenced medical practice, and shaped extramural research funding at the National Institutes of Health. These changes were accompanied by growing interdependence between science and technology and between firms, universities, and the federal government, reflecting biomedicine's increasingly systemic character. The evidence points to the long-run effects coordinated, integrative research policy can have in spurring the development of innovation systems and shifting technology sectors into a high-investment, high-growth development path.

JEL Classification: H51, H56, N42, N72, O14, O31, O32, O38

Keywords: Public R&D; Innovation systems; Biomedicine; Institutional development; World War II

*We thank Ashish Arora, Tim Bresnahan, Wes Cohen, Bob Cook-Deegan, Kyle Myers, and Paolo Surico for helpful conversations, as well as seminar audiences at Harvard Business School, Johns Hopkins University, UMass-Amherst, the University of Hawaii, University of Maryland, Washington University in St. Louis, Warwick Business School, and the Wharton School; conference audiences at the 2022 AEA annual meetings, the 2022 Taiwan Symposium on Innovation Economics and Entrepreneurship, the 2024 Conference on the Economics of Innovation in Memory of Zvi Griliches, and 2025 NBER Summer Institute; and discussants Pierre Azoulay, Kyle Myers, and Gary Pisano. We also thank Hayley Pallan, Senan Hogan-Hennessy, and Oum Lahade for research assistance during this project's early stages and Eric Gilliam for helpful discussions and related background research. Duke University's Fuqua School of Business, the Harvard Business School Division of Research and Faculty Development, and the National Bureau of Economic Research (Innovation Policy grant) provided financial support. This material is based upon work supported by the National Science Foundation under Grant No. 1951470. All errors are our own.

†Address: Duke University Fuqua School of Business, Durham, NC 27708, USA; email: daniel.gross@duke.edu.

‡Address: Johns Hopkins University, School of Government and Policy & Carey Business School; Washington DC 20001, USA; email: sampat@jhu.edu.

1 Introduction

Improvements in human health rank among the most transformative developments of the twentieth century: Nordhaus (2003) estimates that the 30-year increase in U.S. life expectancy over the century was as valuable as measured economic growth in all other sectors combined. Although the reasons for these gains have been debated, a crucial contributor is widely thought to be advances in medical science, technology, and practice (e.g., Murphy and Topel 2003, Cutler and Kadiyala 2003, Cutler et al. 2006). Many observers attribute these advances to the productivity of a highly developed U.S. biomedical R&D ecosystem (e.g., Laitner et al. 2024).

This ecosystem, however, is a post-World War II phenomenon. A century ago, U.S. universities were far less research intensive than they are today and had limited resources for research, medical or otherwise (Geiger 1986).[1] The pharmaceutical industry was small and concentrated, spent a smaller share of sales on R&D, and undertook little formal testing (Cockburn et al. 1999). The National Institutes of Health (NIH) was also small and intramural, interfacing little with the university or commercial sector. As Figure 1 shows, U.S. biomedicine was relatively stagnant until the 1940s. World War II then placed new demands on the country's medical research system, and between 1941 and 1945, the newly-created U.S. Office of Scientific Research and Development's (OSRD) Committee on Medical Research (CMR) enlisted universities, firms, and hospitals around the country in a large, collaborative effort to develop medical science and technology for war—an undertaking which was both the federal government's first extramural investment in biomedical research and one of its few pursuits (ever) of an actively-managed biomedical research policy. The end of the war in 1945 marked a turning point in annual scientific publications and drug-related patents (Panels A and B), which have been growing ever since.

[Figure 1 about here]

This evidence motivates the central question of this paper: what impact did World War II have on postwar U.S. biomedical innovation, and why? Using newly-collected archival data on all CMR research contracts, we show that the war effort triggered long-lasting growth in scientific research and drug development in subjects it supported, with effects persisting into the 1960s and beyond. Firms engaged in CMR-led drug development projects entered a two-decade period of prolific drug innovation, which was increasingly characterized by systematic approaches to drug discovery, closer links to manufacturing, and more connections to science. CMR also provided a foundation for the postwar NIH, which absorbed CMR's portfolio and much of its apparatus after it demobilized.

[1] As Geiger (1986), Urquiola (2020), and others have observed, U.S. universities began growing their research capacity in the 1920s and 1930s, but had limited capabilities by modern standards (and nearly none outside of a small number of universities) and limited research funding (even from leading foundations).

With academic science, the pharmaceutical industry, and NIH being the three pillars of the U.S. biomedical innovation system today, World War II was a turning point in its historical development. Adapting big-push theories of economic development to the study of innovation ecosystems, and evaluating it against the evidence, we argue that the World War II medical research effort established the key capabilities, inputs, and institutional relationships around which a more integrated biomedical innovation system could coalesce and start to grow.

Our first set of analyses examines science. To do so, we collect data on the universe of biomedical research publications between 1930 and 1970 and map both these and CMR contracts to Medical Subject Headings (the National Library of Medicine's controlled vocabulary for indexing research in the life sciences), as a measure of research space. Comparing the long-run growth trajectories of research subjects that were targets of CMR investment against those that were not, we find evidence of both continuity and change: less-developed (pre-1940) subjects that were a focus of CMR research grew substantially during and after the war, while the more established subjects it supported continued their pre-1940 growth trajectory, before tapering.

Our second set of analyses examines CMR's impacts on drug development in the postwar pharmaceutical industry. Using a list of new drugs introduced to the U.S. market between 1940 and 1975, we show that drug categories where CMR was active yielded significantly more new drugs in the 1950s and 1960s. Firms involved in the war effort—some of whom, like Pfizer, were not previously in the drug business—became particularly prolific drug developers for the next two decades. Many of these drugs developed as a result of new discoveries, processes, and capabilities which CMR directly supported—from semi-synthetic antibiotics to natural drugs which could be produced using techniques for industrial-scale fermentation borne of the war effort.

One practical effect of this investment was a reduction in military morbidity and mortality, particularly from infectious disease. But CMR-led wartime innovation also had the potential to influence civilian health. Though specific links to civilian disease burdens are difficult to establish econometrically, in our third set of analyses, we examine the diffusion of CMR-funded research into medical training and practice. To do so, we draw on a leading medical textbook (Cecil's *Textbook of Medicine*) and popular clinical reference manual (the *Merck Manual of Diagnosis and Therapy*).[2] Using successive editions of these texts, we show that CMR-funded subjects were more likely to be added to postwar editions of these textbooks, reflecting the practical value of new medical knowledge which emerged from the war—albeit with a several-year lag, reflecting the time it takes for new medical information to standardize and diffuse into practice.

[2]Our choice to probe diffusion of medical information into textbooks is motivated by research arguing that textbooks were until recently a main source of information for clinicians which "reflect the level of knowledge of the average medical practitioner and [can be used as] a consistent sampling device for measuring changing practice patterns" based on their periodic revisions (Greene 2007, p. 294; also see Catillon 2017).

Individually, these results document the impacts of public R&D on biomedicine and its application in technology—subjects of perennial scholarly interest (e.g., Cockburn and Henderson 1996, Jacob and Lefgren 2011, Toole 2012, Azoulay et al. 2019b, Myers 2020).[3] Beyond its specific impacts in these domains, the postwar takeoff of biomedical science and innovation in Figure 1 suggests a broader transformation or structural change was afoot, of a type which has been less of a focus in research on the effects of public R&D and is more difficult to explain.

To make sense of this takeoff, we turn to a long-running literature which examines bottlenecks to economic development and industrial growth. A line of argument from Rosenstein-Rodan (1943) to Murphy et al. (1989) (and beyond) has suggested that cross-sectoral interdependencies may inhibit industrialization. In essence, this literature argues that interdependence makes low development an absorbing state, and shifting the economy to a high-growth equilibrium may require coordinated investment—a so-called "big push"—to kickstart a growth process that, due to these same complementarities, can feed off itself. We adapt this logic to innovation systems (Nelson and Rosenberg 1993), where spillovers arise not from static demand (as in Murphy et al. 1989) but rather from institutional interactions and dynamic science-technology interplay.

Consistent with this logic, we find that CMR created several new R&D assets around which postwar biomedical science and technology subsequently developed. In addition to new policy institutions—up to the very idea of large-scale government funding for biomedical research itself—these included new research tools and techniques, new therapies, new drug development platforms, new firm capabilities, new collaborations, and new fundamental understanding. We also find that the CMR shock not only caused science and technology to grow: it also triggered growing linkages within and between them. These are evident in spillovers across research subjects and in the co-development of new drug classes and production methods. A deeper change was the growing connection between science and technology: whereas prewar drug developers had limited connections to science, and progressed more through trial-and-error empiricism, science became common in postwar drug development—a change which is visible in the growing rate at which drug-related patents, and particularly CMR contractor firms' patents, cited academic science.

One contribution of this paper is in understanding innovation processes and institutions. In particular, the heightened connectedness of innovation under CMR and in postwar biomedicine is different from the popular "linear" view of innovation in which science flows down to technology, production, and implementation (Godin 2006, Balconi et al. 2010). Despite frequent critiques, the linear model remains the subject of a large and interdisciplinary scholarship on innovation and a

[3]To these outcomes, we add a measure of incorporation into practice. Our research also differs from prior work in studying not "bottom-up", investigator-initiated funding for fundamental research, but the effects of a more "top-down", use-oriented biomedical R&D policy bordering on technology policy—an approach which draws recurring interest in biomedicine (Cook-Deegan 1996, Sampat 2012, D'Souza et al. 2024) and other fields (Branscomb 1992, Nelson 1997) and which has a recognized evidentiary gap (Bloom et al. 2019).

significant share of innovation policy (including much of NIH).[4] This paper offers a reminder that innovation processes involve substantial interdependence across different actors and institutions and across the R&D value chain (Rosenberg 1982). Our emphasis on this feature of innovation and its implications brings a complementary "systems" perspective to empirical research on the effects of R&D policy, which otherwise often focuses on discrete outcomes.

More generally, this paper contributes to a wider literature on the effects of biomedical research funding. Although the NIH has been the subject of most previous empirical studies assessing the returns to medical research, its emphasis on funding undirected, investigator-initiated science at universities is specific and distinct from CMR's integrated research model in World War II, which set application-oriented priorities, coordinated the activity it funded, spanned sectors and disciplines, and linked science to technology, manufacturing, and diffusion. While scholarly interest in R&D policies with these features is growing (e.g., Mowery 2009, Foray et al. 2012, Mazzucato 2018, 2021, Kattel and Mazzucato 2018, Azoulay et al. 2019a, Agarwal et al. 2021), there is limited systematic empirical evidence on what effects such approaches might have and why.[5] This paper brings new evidence to this literature, complementing our prior writing on the effects of World War II research on the non-biomedical U.S. innovation system (Gross and Sampat 2023a), where we found that the broader OSRD effort set in motion postwar growth and regional agglomeration of U.S. electronics and communications innovation. Beyond the shift in domain (to biomedicine), the present paper provides a deeper insight into the origins and structure of innovation systems and suggests a new mechanism through which R&D policy can catalyze long-run growth.

We proceed as follows. Section 2 provides institutional background, including on the importance of medicine in warfare, CMR and the World War II medical research effort, and CMR's influence on postwar research policy. Section 3 introduces our data and characterizes the CMR shock. Section 4 evaluates CMR's effects on postwar biomedical science; Section 5, on pharmaceutical innovation; and Section 6, on medical practice. Section 7 explores why CMR had such long-lived effects and what we can learn from it about the economics of science and science policy. Section 8 considers extensions and remaining questions and concludes.

[4] Though long declared "dead" (e.g., Rosenberg 1991, p. 335), the linear model is implicit in the voluminous literature examining the impact of science on technology (see Rosenberg 1974 or Brooks 1994), efforts to measure direct linkages from science to technology (e.g., Marx and Fuegi 2020, Bryan et al. 2020), and policies designed to encourage commercialization of university science (Mowery et al. 2001, 2004), among others.

[5] Though this growing body of scholarship on "mission-oriented" R&D considers the impacts of actively-managed R&D policy, we find that the "mission" label is too ambiguous for our purposes and can be difficult to generalize. For this reason we largely eschew it in this paper. For example, though NIH is a mission-oriented agency (whose mission is to seek and apply fundamental knowledge to improve human health; Sampat 2012), it functions mainly as an arms-length financier, delegating both grant review and the performance of the research it funds to a decentralized scientific community, and leaving it to the private sector to decide how to use it. Most academic writing on "mission-oriented" R&D has a more interventionist approach in mind, where the state makes or shapes the market (e.g., Mazzucato 2018, 2021)—a feature which is more characteristic of defense R&D (Mowery 2012) or crisis innovation policy (Gross and Sampat 2022) specifically than of mission-oriented policy generically.

2 World War II and Biomedicine

2.1 Military medicine: The battle against disease

Although nearly 70 countries participated in World War II, the U.S. military's greatest adversary was arguably disease: for most of history, infectious disease has killed more soldiers than battlefield wounds and incapacitated an even greater number (Hoyt 2006).[6] World War II also introduced a wider range of medical problems than the U.S. military had previously encountered—not only new diseases, but also new environmental conditions and traumatic injuries. In the early 1940s, there was thus an urgent need for knowledge and technologies that could address medical problems Allied soldiers faced, including prevention and treatment of bacterial and viral infections, malaria, wound and shock treatment, blood substitutes, mental health, and issues relating to aviation physiology, motion sickness, temperature, and nutrition, among many others.

When World War II began, there was no obvious office in the U.S. federal government to assign to this charge. Though the National Institute of Health (singular) was created in 1930, and was expanded by Congress to include a National Cancer Institute in 1937, it was small and had no serious extramural research funding capacity of the type that the war required (e.g., Swain 1962). Private foundations had previously funded medical research through grants to individual researchers, but these too were insufficient for the demands the war presented, due to their relatively small scale and their focus on fundamental research rather than applications.

In 1941, the U.S. goverment's Committee on Medical Research (CMR) was created as a subsidiary to the Office of Scientific Research and Development (OSRD)—an agency established to coordinate and fund civilian R&D for war—to fill this void. At the time, this was an unprecedented choice: before 1940, there was very little federal funding of research outside of agriculture, nor precedent or mechanism for funding extramural research. In the executive order creating OSRD, CMR was to support "the mobilization of medical and scientific personnel of the Nation" and advise and oversee contracts "with universities, hospitals, and other agencies conducting medical research activities ... related to the national defense" (Andrus 1948, p. xlii).

The R&D-funding apparatus that CMR ultimately developed to fulfill this charge was broad and multifaceted (Gross and Sampat 2023b). General research priorities were determined in partnership with the military, which brought military medical problems to its attention. CMR then shared these priorities and solicited proposals widely, including from university scientists, firms, hospitals, and independent research institutes. Rather than reviewing these proposals directly, CMR forwarded proposals to the National Research Council's (NRC) Division of Medical Sciences (DMS), where

[6] See Appendix A for further discussion. Appendix Table A.1 compares mortality from disease and injury in prior eight prior conflicts, based on reporting from the U.S. Army's Medical Department.

over thirty committees—comprised of hundreds of elite medical researchers and officers from the Army and Navy—provided peer evaluations, in an early use of peer review. Based on these reviews, DMS gave each application a letter grade and returned a recommendation for funding, which CMR typically followed. Once funded, CMR provided active project management, including organizing meetings of investigators to promote information flows, collecting and circulating progress reports, and supplementing or terminating projects as their results, and CMR's priorities, evolved. For research directed at new treatments for diseases, CMR was also active in development, evaluation, and implementation, with many of its contracts supporting experimental interventions and/or clinical trials. Like OSRD in general, it was primarily a "results-oriented" program, and prioritized relevance and speed over fundamental, scientific value.

Beyond funding: coordination and brokerage

Beyond funding, CMR played an important role in creating connective tissue in the nascent U.S. biomedical innovation system. This connective role was particularly pronounced in specific projects, such as in the effort to mass produce penicillin. At the beginning of the war, the technology did not exist to produce enough penicillin to treat a single patient, let alone for clinical testing—yet by the end of the war, there was enough for all Allied troops and civilian use (Keefer 1969). Early in the war, CMR reportedly was crucial in persuading private firms to get involved, brokering information flows between these firms and the U.S. Department of Agriculture's Northern Regional Research Laboratory (NRRL), organizing meetings, and refereeing conflicts among participants (Swann 1983, Neushul 1993). Once firms were able to produce enough penicillin for testing, CMR coordinated and funded clinical trials. After trials were complete, CMR worked with the Office of Production R&D of the War Production Board (WPB) to scale up production to the needed levels. The WPB provided needed material and equipment to firms, shared technical expertise, and provided some funding. WPB corresponded with 175 potential producers, and eventually worked with 20 in the program, chosen based on experience with penicillin, fermentation, and biological production in general, as well as the quality of staff (Neushul 1993). Although CMR's primary role in the scale-up of natural penicillin was coordination and clinical testing rather than more upstream R&D funding for biochemistry or drug discovery, CMR did invest considerable research funds in a parallel synthetic penicillin development program—which, at the beginning of the war, it viewed as a more likely path to large-scale production (Swann 1983).

2.2 Results of CMR research

Between 1941 and 1945, CMR engaged researchers in roughly 570 contracts totaling around $400 million (2024 dollars). Though it comprised only 5% of OSRD's total spending, and is less than 1% of NIH's modern budget, CMR research funding was an order of magnitude larger than previous

federal spending on medical research, much of it going into subjects that had not been a major focus of research prior to the war. As Stewart (1948, p. 102) describes:

> The shift in emphasis and even in direction was enormous. Many subjects of minor importance in peacetime become of controlling importance in war. Some subjects are born of war. Tropical medicine had been considered of rather academic interest to the health of the United States. Even the machine age had not adapted our younger generation to flying at 40,000 feet or diving at 400 miles an hour.

Though the performance of individual research programs it supported was mixed, CMR was successful enough during the war to impact its conduct. The mass production of penicillin is its most celebrated accomplishment: by 1944 the U.S. was able to produce enough natural penicillin to meet military demand, and so much by 1945 that it was made available for civilian use. In contrast, CMR's parallel effort to chemically synthesize penicillin—initially seen as more promising—was unsuccessful. Its research also extended to malaria. Though malaria had long been treatable with quinine, the Japanese invasion of Java and war in the South Pacific cut off supply routes. CMR's malaria program focused on finding quinine substitutes. It identified and synthesized over 14,000 compounds and tested promising candidates against animal models, in clinical trials, and on soldiers in the field (Keefer 1969). This effort struggled to find substitutes during the war, and when CMR established that an existing molecule, atarabine, could be used safely and effectively, it was adopted as the military's preferred preventative and treatment.[7]

There were numerous other successes and failures, most of which are detailed in Baxter (1946) and Andrus (1948). Hoyt (2006) finds that between CMR and other government agencies, wartime research helped develop new or improved vaccines for 10 of the 28 vaccine-preventable diseases identified in the 20th century. CMR had a crucial role in funding research and development on blood substitutes used to treat battlefield casualties (Creager 1999). CMR research also deepened understanding of human physiology under environmental hardships such as nutrient deficiencies, sleep or oxygen deprivation, and temperature exposure, as well as of stress disorders and mental health—all in response to the demands the war presented.

The medical impacts of this work can be seen most clearly in military statistics (Appendix A). In short, World War II R&D essentially solved the military's problem of infectious disease. The ratio of U.S. military deaths from disease vs. injury fell from 1.02 in World War I to 0.07 in World War II (0.01 in the European theater). Hospital admission and death rates for many common infectious diseases—such as pneumonia, influenza, and typhoid fever—declined nearly 100% between the two conflicts. Appendix Figure A.1 extends these comparisons by plotting the time series of U.S. Army hospital admissions and death rates per capita between 1895 and 1955, which shows significant spikes in prior wars, but no such deviations in World War II.

[7] Chloroquine was also a subject of CMR research, though it came into focus too late to be useful during the war—and instead became a revolutionary malaria treatment in the years immediately after.

2.3 CMR and postwar research policy

When CMR was disbanded in 1945, the Public Health Service took over its forty-odd open contracts, which became the kernel of NIH's extramural research program (Swain 1962, Fox 1987). Beyond providing a general model for using grants and contracts to fund research, NIH adapted specific contracting approaches developed by CMR and OSRD during the war, including indirect cost recovery policies and elements of patent policy (Rosenzweig 1998, Sampat 2020, Azoulay et al. 2025). NIH's peer review approach was also based on CMR: "study sections" of external scientists providing initial scientific/technical review were modeled on the wartime NRC/DMS review system. In 1946, the NRC Penicillin Panel transitioned into the Syphilis Study Section, marking the inception of what has grown to encompass over 250 similar study sections within the NIH. The NIH budget has increased 1000-fold in real terms since the end of the war, and the agency is sometimes known as the "crown jewel" of the federal government (Sampat 2023).

Despite these continuities, NIH also made major and explicit departures from CMR, including by emphasizing fundamental research and providing flexibility for scientists to take their research in unplanned directions when opportunities emerged (Van Slyke 1946). In addition to this freedom of exploration, topic choice in the NIH program was also primarily investigator-initiated or "bottom up" rather than the "top down" priority setting of the wartime model. In a 1962 interview, Van Slyke recounted that the targeting of biomedical research ("[T]hou shalt concern yourself with the making of anti-malarial. Thou shalt concern yourself with the survivor's suits.") was "justified by war time needs and exigencies" but was not the focus of the NIH.

Over time, and especially after Congressional investigations in the 1960s, NIH's original emphasis on scientific freedom eroded, with the growth of bureaucracy and reporting requirements (Sampat 2023). However, the NIH model continues to focus on "bottom up" investigator-initiated research with little of the explicit top-down project selection, active project management, and focus on application and diffusion that were core elements of the CMR model. NIH's approach has come under scrutiny, including during debates surrounding the War on Cancer in the 1970s, the Artificial Heart Program, concerns about NIH priority setting (Sampat 2012) and more recently in the initiative that led to ARPA-H (Sampat and Cook-Deegan 2021). Though nominally a mission agency, NIH (and its parent Department of Health and Human Services) also does not link up different aspects of its health mission (research, development, testing, diffusion into practice, procurement) as CMR's "integrated research model" (Hoyt 2006) did.

3 Data and Empirical Approach

To systematically examine the link between CMR and postwar biomedical innovation, we collected, transcribed, and harmonized a complete record of 590 CMR contracts (573 extramural, 17 intramural) from OSRD archival records.[8] For nearly all (588), we have a summary report that identifies the sponsoring CMR division and the research projects' subject, principal investigator(s) (PI) and other technical staff, institution(s), budget, and timing (see Appendix Figure B.1 for an example).[9] These summary reports also provide an extended abstract and list all resulting publications (e.g., journal articles, technical reports, progress updates). Collectively, this information provides the corpus we work with: titles and abstracts provide information on funded subject matter, publications on research output, and header data on the investigators and institutions involved. Through CMR records we identify 2,438 scientific publications produced by CMR-funded research, spanning a wide range of journals in medical science and related fields (e.g., public health, organic chemistry, entomology). We manually linked these publications to three external publication databases: Microsoft Academic Graph (MAG), Web of Science (WOS), and PubMed.

In parallel to information from CMR, we collect data on three categories of outcomes: science, innovation, and medical practice. We measure scientific activity via research publications, which are attractive for their transparency and observability over long horizons, including both before and after the war. Our publication sample begins with the universe of publications between 1930 and 1970 in MAG. We filter this sample to the 5.5 million publications in the *Natural Sciences* and *Medical and Health Sciences* published over this period, as indicated by MAG topic labels (OECD field codes). Though other publication datasets (e.g., WOS) have been used in prior research, MAG has two main advantages over other sources: (i) it has comprehensive historical coverage dating to the late 1800s, and (ii) it is open access. MAG provides additional information about individual publications that are useful for our purposes, including titles, journals, authors, and cross-paper citation linkages, and enables us to use secondary data products produced from the MAG sample, such as patent citations to science (Marx and Fuegi 2020, 2022).

Measuring pharmaceutical innovation is harder: though drug development is often measured via FDA drug approvals, there are no electronic data on drug approvals from this era, and the FDA Orange Book begins only in 1985 (Durvasula et al. 2023). We fill this gap by locating and digitizing

[8] See Appendix B for a detailed description of our data sources and data collection. We cross-validate and CMR contract and publication data against additional OSRD records, including separately-maintained CMR contract lists and publication lists (details in Appendix B). This cross-validation led to occasional minor corrections, usually attributable to minor typos in the source records. The 17 intramural contracts were entered into with other government agencies (primarily with the USDA, NIH, and FDA), which participated in a handful of CMR research programs (e.g., USDA's Bureau of Entomology in the malaria program).

[9] Extended abstracts are available for 441 (75%) of summary reports. The two contracts for which we do not have summary reports are CMR's first and last extramural contracts.

de Haen (1976)'s "Compilation of New Drugs, 1940-1975", which lists "new chemical entities or synthesized drugs not previously available in the United States" first marketed between 1940 and 1975 and serves as the basis for the FDA History Office's historical drug approval statistics.[10] The De Haen data document 1,010 drugs developed by 126 distinct firms over this period, along with their trademark name, generic name, and year introduced. Each drug is categorized into one of 42 therapeutic classes and over 150 subclasses, which we will use in our analyses below. In parts of the paper, we will also analyze drug-producing firms, for which we take two further steps. First, we link firms in de Haen (1976) to patent assignees and measure these firms' drug patents (which we define as those in NBER patent category 31, "Drugs"; Hall et al. 2001) and their characteristics.[11] Second, because some firms in our sample merged during the study period, we collect merger data from historical Federal Trade Commission tables (FTC 1980, covering 1947-1978) and dynamically assign firms to their contemporary parents after known mergers.

In examining the effects of wartime medical funding on science and pharmaceutical innovation, our analysis parallels previous work on the effects of NIH funding (e.g., Jacob and Lefgren 2011, Azoulay et al. 2019b, Myers 2020). Our third set of outcomes extends this line of work by studying diffusion. Though measuring the incorporation of publicly-funded research into medical practice is challenging, prior research points to medical reference books as a window into medical knowledge and practice. We focus on two series published both before and after the war: the *Cecil Textbook of Medicine* (henceforth CT), a staple textbook of medical training (Greene 2007), and the *Merck Manual of Diagnosis and Therapy* (MM), a popular clinical reference (Tomes 2021). For each of these book series, we digitize all editions with an index through approximately 1960 (1930 to 1959 for CT, 1940 to 1961 for MM) and compile a list of indexed subjects.

A final resource for this paper is postwar biomedical research funding. As Section 2 explains, CMR inspired and funded the creation of NIH extramural research funding, which increased sharply over the first two postwar decades. Postwar NIH funding in specific subjects may be another effect of the CMR shock. We digitize annual editions of the U.S. Public Health Service's *Research Grants and Fellowships Awarded by the National Institutes of Health* from 1948 to 1970 to collect information on all NIH grants over this period, complementing the CMR record.

3.1 Categorizing CMR contracts, publications, and other sources

We use the National Library of Medicine's (NLM) Medical Text Indexer (MTI) to map these data to a common domain: NLM's Medical Subject Headings (MeSH) vocabulary, which gives structure to biomedical research space. The MeSH vocabulary was developed as part of MEDLINE, NLM's database of biomedical journals, and consists of descriptors used to index MEDLINE articles for

[10] See https://www.fda.gov/about-fda/histories-product-regulation/summary-nda-approvals-receipts-1938-present.
[11] Patent data obtained from Google Patents and the Reliance on Science project (Marx and Fuegi 2020, 2022).

searching and retrieval. There were 29,915 unique MeSH descriptors ("MeSH terms") at our time of use (2021-2022).[12] MeSH has an accompanying hierarchical structure, and the underlying "MeSH space" is organized into 16 broad classes (Appendix Figure B.4 provides a list), each with subclasses, which have further subclasses, and so on.[13] For example, the "Diseases" branch (letter C) begins with Infections (C01), which in turn contains infections at various levels of specificity; each node in this tree we will refer to as a MeSH code (e.g., C01.221.250: Blood-Borne Infections). Individual MeSH terms can exist at multiple locations in the MeSH tree (e.g., "Blood" is listed under both Body Fluids and Hemic and Immune Systems)—a feature which shapes our preference for MeSH terms (which are unique) rather than codes as our unit of analysis.

MTI is a language processing tool that maps input text to MeSH terms and is used by NLM to provide initial indexing of MEDLINE articles based on titles and abstracts. It can also be used to categorize arbitrary biomedical text (NLM 2022). It does so in several ways, including (i) taking the words in provided text and finding similar concepts in NLM's UMLS Metathesaurus, then finding the closest MeSH headings, and (ii) by finding similar PubMed articles and extracting their MeSH terms (Mork et al. 2013). It also returns scores indicating confidence in each match. Prior research has used MTI in other applications, including to measure the breadth of scientific articles (Kolev et al. 2020) or determine the gender focus of patents (Koning et al. 2021). Here we use it to identify the subjects of contracts, grants, publications, and medical texts.

In practice, MTI requires choices over what input text to use, how to map text to terms, how to use confidence scores, and how to aggregate up to MeSH subjects or subject-years. Where possible, we use all available text: we index MAG publications on their titles, CMR contracts on their titles and abstracts, NIH grants on their titles, and medical textbooks on index entries. In using these data, we normalize the returned MeSH term confidence scores for each publication to sum to one, and drop all terms with a score below 10% to reduce noise.[14] Our analysis will at times make use of score-weighted totals at the MeSH term or term-year level, and at times binary indicators (e.g., of whether a particular MeSH term was a focus of CMR research).

For drugs we take a different approach, partly because we have less raw text to feed into MTI, and our preliminary probes indicated that MTI does not reliably return descriptors from drug names, active ingredients, or even drug categories as inputs. We instead create a manual crosswalk between de Haen (1976) drug categories and 12-digit MeSH codes on the *pharmacologic action* branch of the MeSH tree. We can then perform analysis at the level of these codes or their associated descriptors.

[12] The MeSH vocabulary is continuously revised and updated as new terms appear in the scientific literature.

[13] The MeSH tree is conceptually similar to hierarchical patent classification schemes such as the Cooperative Patent Classification (CPC), with top-level (lettered) MeSH codes conceptually analogous to top-level CPC sections, three-character MeSH codes (e.g., A01) analogous to CPC classes, and so on.

[14] We also drop check tags (MeSH descriptors that specify species, sex, or age, such as "Humans"), and supplementary concepts (terms outside of the MeSH thesaurus, many of which are chemical formulae).

We discuss this drug-MeSH linkage in more detail in Appendix B.

3.2 Characteristics of the CMR shock

Table 1 describes the shape of CMR's research investment, showing the distribution of contracts, contractors, and research funding across CMR divisions. Two patterns stand out: one is the (relatively) even distribution of contracts and funding across divisions and the myriad problems they were solving. The other is the shockingly low total cost of the program, at roughly $21 million in the 1940s—equivalent to $400 million in 2024, or less than 1% of the current NIH research budget. The table also contextualizes CMR research in MeSH space, documenting the most common MeSH terms in each division: *Syphilis* (Medicine), *Burns* (Surgery), *Oxygen* (Aviation Medicine), *Shock* (Physiology), *DDT* (Chemistry), and *Antimalarials* (Malaria). The most common MeSH term across the CMR portfolio is *Penicillins*. Figure 2 extends the last row of this table, showing the top 10 MeSH terms for each of these divisions, by term share of division contracts. The results reveal a fuller list of focal subjects of CMR research, while also providing a check on the face validity of the MTI approach to categorizing contracts and publications.

[Table 1 and Figure 2 about here]

The raw data suggest that despite the specificity of the wartime problems CMR targeted, research it funded may have been more broadly impactful. One such indication is visible in publication counts alone. Appendix Table C.1 lists the top five MeSH terms entering the publication record (based on titles in the complete MAG corpus) each year from 1939 to 1946. For each term, we also report the number of associated publications over the next 10 years and indicate whether it was a focus of CMR research. Treatments and therapies that CMR cultivated were immensely more likely to be the most heavily studied new topics of this period.

Other evidence comes from comparing CMR-funded research to contemporary work in the same subjects. In Table 2 we estimate differences in the characteristics of CMR and non-CMR publications, conditional on subject-year fixed effects.[15] We do so on three dimensions: novelty, breadth, and impact—measured by the introduction of a new MeSH term combination, the number of associated subjects, and forward citations, respectively. Column (1) reveals that CMR-funded publications are significantly more likely to introduce new combinations, with the difference a precisely estimated 25% increase on the mean rate. Columns (2) to (4) show that CMR-funded publications were also significantly broader, and a handful of its publications (e.g., surveys of antibiotics or malaria) were

[15] Concretely, we estimate the following specification: $Y_i = \beta \cdot \mathbb{1}(\text{CMR-funded})_i + \delta_{st} + \varepsilon_i$, where i indexes publications, δ_{st} are subject-year fixed effects, and standard errors are clustered by subject and year. For the purposes of these tests, publications' primary subjects are measured as their top MTI-scoring subject.

among the broadest of this era. Columns (5) to (8) show that CMR publications were also heavily cited, roughly twice as likely to be in the top 10% of cited articles in their year and nearly quadruple as likely to be in the top 1%, with relatively tight standard errors.

[Table 2 about here]

4 Effects on Biomedical Science

Our analysis of CMR's effects begins by examining its impacts on science, where a closer look at a few examples can motivate our approach. As we discussed in Section 2, wartime medical research had several thrusts, including research efforts in developing antibiotics, antimalarials, synthetic hormones, and vaccines, as well as developing new techniques and understanding for blood, blood preservation, and blood substitutes, or confronting physiological challenges the war presented (like human performance at high altitudes or in extreme temperatures). To understand the context in which CMR exists, we examine publication time series in specific MeSH subjects closely related to these efforts, which are shown in Appendix Figure C.2. Consistent with Stewart (1948)'s observation that "some subjects are born of war," we find that many of these research areas had little pre-war publication activity but took off after the war ended. *Penicillins*, for example, grew from 0 publications per year pre-war, to 450 at its wartime peak, and settled at roughly 150 publications per year afterwards, while *Anti-bacterial Agents* (often representing synthetics) was slower to grow but was the subject of roughly twice as many publications per year as natural penicillin after the war. Similar patterns are present for *Steroids* and *Blood Proteins*. Other subjects had pre-war research activity but grew significantly following the CMR shock (e.g., *Oxygen*). There are also exceptions: for example, research in *Antimalarials* temporarily spiked during World War II, but that intensity was evidently not sustained in the postwar era.

This evidence motivates the empirical comparisons we make in the rest of this section, where we systematically compare publication activity over time in subjects with and without World War II investment. Our baseline estimating equation is as follows:

$$Y_{mt} = \sum_{t=1931}^{1970} \beta_t \cdot \mathbb{1}(\text{Any CMR contracts in MeSH term } m) + \alpha_m + \delta_t + \varepsilon_{mt} \tag{1}$$

where m and t index MeSH terms and years, and the sample runs from 1930 to 1970, with standard errors clustered by MeSH term. All β_t parameters will be estimated relative to 1930, which is the omitted (reference) year. Our preferred treatment measure is an indicator of whether a MeSH term was the subject of any CMR contracts. This choice is the product of two subsidiary choices—whether to measure inputs (research contracts) or outputs (e.g., publications), and whether to do so

on the extensive versus intensive margin. We prefer inputs to outputs primarily because some CMR research may not have yielded output during the war but may have created longer-lived research assets (a theme we will return to in Section 7). We prefer the extensive to intensive margin for two reasons. The first is the potentially wide variation in the cost of research across subject areas (e.g., physiological studies vs. drug trials), which is difficult to adjust for. Second is that in a few areas (e.g., in natural penicillin), CMR primarily provided coordination rather than funding, and these contracts had only nominal legal consideration. This latter observation applies more broadly: CMR was both a research management organization and financier. For these reasons, the extensive margin is likely to be more meaningful than the intensive margin.

Our principal dependent variable will measure the inverse hyperbolic sine (IHS) of *MTI score-weighted* publications in a given MeSH term, which due to the MTI weights is a continuous rather than discrete measure.[16] We also evaluate unique authors publishing in a given MeSH term (using MAG author identifiers) and the number of new combinations created with other MeSH terms, as a measure of recombinant science. Whereas publications reflect the level of research activity in a subject, new combinations measure an expansion in its scope.

4.1 Identification

Though CMR produced the first significant U.S. government funding for medical research, a potential concern is the endogeneity of what it supported—especially the possibility that CMR funding may have flowed to growing subjects or those with emerging scientific or technological potential, which could result in upwardly-biased estimates of CMR's effects on the outcomes we study in this paper. This possibility is less likely to interfere with our analysis of the emergent interconnections between them, which (we will show) were limited before the war, but substantial after it, and we argue unlikely to arise absent an integrative, CMR-like intervention.

Our discussion of identification therefore focuses on our analysis of CMR's direct effects on science and technology. An intuitive argument for independence is that military medical needs were exogenously different than the problems scientific activity would have otherwise concentrated around absent the war. Contemporary accounts suggest this is true in part, but not in full: in the opening sentence to the official history of CMR, Andrus (1948, p. 3) observes that "War [both] augments

[16] IHS-transformed outcomes approximate logged values but are defined at zero, and in many settings support interpreting parameters as semi-elasticities. Results are quantitatively similar with unweighted publication counts. We show in Appendix C.2, that results are also quantitatively similar with log transformations (rather than IHS), and for most specifications statistically similar; where we examine publication activity in "embryonic" subjects (defined below), results under logged outcomes remain similar but grow noisier due to missing values. Appendix C.2 further shows that our results are similar for count models estimated by Poisson pseudo-maximum likelihood (PPML). We present IHS-based results in the paper in part for their flexibility and ease of interpretation (where each parameter can be interpreted without applying an exponential transformation), and in part because count model estimators (like PPML) can be sensitive to outliers. In Appendix Table C.2, we explore the contributions of the intensive and extensive margin to our results, finding comparable patterns in each.

certain problems already existent in civilian life and engenders new ones." Intuitively, problems like aviation medicine or traumatic injury may be particularly acute in a war, but in other areas, civilian and military medicine can coincide—most prominently in preventing and treating common infectious diseases. We also see this dichotomy in the data: when we separately evaluate CMR's impact on established research areas and embryonic ones, we find evidence of pre-CMR growth in the former but not the latter, and distinct postwar patterns in each.

The combined evidence thus suggests there are two classes: established subjects where CMR may represent continuity, and emerging ones where it may be an exogenous shock. Given the evidence, much of our focus will be on the latter. Although we cannot fully rule out that CMR pursued subjects which were on the precipice of new breakthroughs that might have occurred anyway, we can offer a few observations. One is that CMR's pursuit of multiple, parallel approaches in many of its research programs, and its mix of successes and failures, suggests against predictability. Even if CMR prioritized areas where breakthroughs were conceivable in 1940, this does not imply they were imminent: penicillin was long known to have therapeutic potential, yet at the time no firm was making it, and there was no known way to produce it in large enough quantities for research, let alone clinical use.[17] Moreover, the risk that CMR merely dovetailed with prewar research funders and would have a similar, confounded effect is also low: despite being the main source of university research funding at the time, foundations such as the Rockefeller Foundation were too small and concentrated to support science at CMR's scale or catalyze research ecosystems.[18] Even at NIH, pre-1940 research priorities were demonstrably different from what CMR was focused on.[19] More generally, CMR's emphasis on not only research but also drug development, clinical trials, and production techniques, and its role in building more connective tissue between them, also made it distinctive from other (more meager) research funding sources at the time.

[17] Recent history offers a parallel: mRNA vaccine research was already "nearly complete" at the dawn of the Covid-19 pandemic, which demonstrated mRNA vaccines' technical feasibility, but prior to then "the technology had been considered for many years [to be] too risky to take seriously" (Bonvillian 2024, p. 1195).

[18] The Rockefeller Foundation in particular is of singular interest, as it was the largest benefactor of natural sciences research prior to the war, and in the mid-1930s accounted for over 70% of foundation funding in the natural sciences (Geiger 1986). Though Rockefeller began supporting experimental biology in the 1930s, and much of what it funded is thought to have been impactful, Geiger reports that in 1938 (at its peak), the foundation funded only 22 project grants in the natural sciences, at a total of roughly $22 million (2024 dollars), and primarily at the most established research universities of the time (see Geiger 1986, pp. 39, 165-166). By comparison, CMR contracted with nearly 70 universities (in addition to many firms)—and funded 20x as much R&D over the war.

[19] As an illustrative case, in 1935 the NIH announced plans for "a program of attack on disease along some 70 lines" of research, using a then-windfall of $1 million (5% of CMR's future budget) authorized by Congress under the umbrella of social security legislation (Science Service 1935). Its press release at the time described many research problems which were "on the doorstep" of scientific study—indicating both their perceived scientific potential and public health value—but on hold due to a lack of funds, including problems and conditions as varied as cancer, heart disease, leprosy, and Rocky Mountain spotted fever (Minneapolis Journal 1935). These problems bore little overlap with CMR's research priorities a few years later. Systematic comparisons reinforce this view: when we map diseases discussed in the 1935 NIH priority list to MeSH codes and evaluate its overlap with the disease areas CMR supported, we find very little. To the degree there was overlap, it was mainly in malaria and venereal disease. Despite these plans, in the end, the authorized NIH funds were never appropriated.

Despite this conclusion, attributing effects we find to CMR specifically could still be challenging, due to two concurrent changes. First, the CMR shock (a supply shock) coincided with war-driven demand for specific types of research, which could produce similar effects (e.g., Clemens and Rogers 2023 show that prior U.S. wars attracted private R&D in prosthetic devices). Second, NIH's postwar expansion may have correlated (in research space) with CMR's portfolio and influenced the same outcomes. Contextual understanding suggests demand is unlikely to explain our results, especially for academic science—where there is no salable product—but also for pharmaceutical innovation, where latent civilian demand for new therapies like antibiotics was already high. We more formally consider these possibilities later in this section, by (i) evaluating research trajectories in World War I, which presented a similar demand shock but without comparable directed funding for military medical research, and (ii) controlling for postwar NIH funding.[20]

4.2 Baseline effects

Figure 3 presents our initial results, displaying the β_t estimates from Equation (1) for publication output, with 95% confidence intervals. Panel (A) estimates effects across all subjects, and finds that CMR-supported subjects were growing disproportionately before the war and continued growing during it, peaking in the 1950s before contracting. Panels (B) and (C) disaggregate the sample into established subjects (with above-median pre-war publications) and more embryonic subjects (with below-median pre-war publications), which is effectively a third layer of differencing. The evidence indicates that CMR dovetailed with trending growth in established subjects, but supported substantial growth in embryonic ones—reflecting both continuity and change. Whereas established subjects increased roughly 25% over pre-war levels by the late 1940s and subsequently contracted back to 1940 levels by 1970, emerging subjects were by 1950 generating roughly 100% more publications per year by 1950 than before the war, and these effects largely sustain through 1970. This initial set of results adds to Figure 1 in suggesting that World War II marked an inflection point in biomedical science, as not only did total research volume take off (Appendix Figure 1), but new, CMR-borne subjects grew while old subjects stalled or declined.

[Figure 3 about here]

4.3 Growth vs. substitution

In principle, these two results could be interrelated if embryonic subjects' growth pulled resources away from established subjects, rather than being strictly additive.[21] Beyond potential substitution,

[20] A third possibility is that the 1950s Korean War perpetuated the CMR shock. Though we lack comparable measures of 1950s military medical research investments, these were almost certainly been lower and far narrower in scope than in World War II, due to the Korean War being a smaller, single-front war and to the successes in World War II, which solved the military's infectious disease problem (see Appendix A).

[21] In the implicit triple difference (comparing old and new subjects), this would be a SUTVA violation.

a corollary question to the results in Figure 3 is whether this increased scientific output is a result of entry versus increased productivity. We explore these questions in Table 3, which presents a range of supplementary analyses specifically for embryonic subjects, estimating parameters in five-year intervals (rather than annually), with 1930-1934 the omitted category.

[Table 3 about here]

As a reference point, Column (1) of the table reproduces Figure 3's result with these periodic parameters. In Column (2), we evaluate differences in the number of unique authors who published in a given subject and year (conditional on $\geq$1 publication in the subject-year). The results indicate substantial postwar entry into CMR-treated subjects. Columns (3) to (5) probe this result further, evaluating to what degree publication growth is driven by old or new authors: Column (3) examines the growth of publications which had only prewar authors (defined as authors with at least one prewar publication); Column (4), publications with any prewar author; and Column (5), publications with any postwar author.[22] The evidence indicates that the growth seen in Figure 3 (and Column 1) is not driven by existing researchers substituting away from their prewar fields, but rather by new researchers pursuing opportunity in emerging fields.

Column (6) of the table provides one further perspective on this growth, re-estimating Equation (1) for new MeSH combinations. This outcome does not mechanically rise as publications do, since scientific research will often study the same subjects (or bundles of subjects) as prior literature. It instead captures the tendency for researchers to expand the scope of inquiry around a given subject. Column (6) provides evidence of a postwar expansion of the scope of research in CMR-supported subjects, beyond a mere increase in the rate of science alone.

4.4 Linkages across research space

The analysis thus far has examined localized impacts of the CMR shock within treated subjects. We use our measurement of co-occurring MeSH terms to examine how CMR shocks in each subject area affected specific combinatoric pairings across MeSH space. To do so, we assign MeSH terms to branches of the MeSH tree based on their associated MeSH codes and examine the effects of CMR on publications in pairwise branch combinations—such as Diseases (C) and Drugs (D), Organisms (B) and Phenomena and Processes (G), and so on. For parsimony, we focus on MeSH tree branches (A) to (G), which represent a large majority of research in medicine and the life sciences, excluding branches for ancillary subjects such as specific geographies or subpopulations, or research on medical professions and the healthcare system or in the social sciences.

[22] In doing so, we extend the omitted category to 1939, as there is no pre-1940 variation in the outcome.

The unit of analysis remains the same (MeSH terms, some of which are CMR-funded), but we now subsample terms on their membership in a given branch of the tree (A to G), and measure the number of associated publications with a co-occurring MeSH term in each other branch of the tree. We re-estimate Equation (1) for publications with each pairwise combination, plotting the estimates in Figure 4, where the focal branch is shown by row and the paired branch across columns. We also restore the sample to all subjects (not just embryonic ones), to explore where CMR may have extended recent prewar trends—though when the sample is restricted to embryonic subjects, all pre-trends once again disappear (as in Figure 3C).

[Figure 4 about here]

The figure offers several insights. First, it highlights that that CMR's effects were similar across all branches of the MeSH tree (column "All", which estimates effects of CMR on all publications associated with terms in the row branch). Second, we can see in which areas spillovers were stronger, such as anatomy (row A) and organisms (row B), where across columns, the estimated effects are larger in magnitude than those of other rows. We also see where they were weaker: for example, the effects on research that recombines with subjects in psychology and mental health (column F) are relatively small. Third, we get a deeper understanding of the pre-trends previously found in Figure 3. The evidence indicates these pre-trends were specifically driven by drug-related research (row D), though even here the figure shows that CMR triggered an off-trend surge in postwar science, powered by the growth of more embryonic subjects. In other areas (besides drugs), CMR appears to have supported research that was not previously growing.

Finally, the results offer suggestive evidence of an effect on both basic life sciences research and clinical medicine, as indicated by publications in more basic branches of the MeSH tree (such as B: Organisms and G: Phenomena and Processes) and more applied ones (such as D: Chemicals and Drugs), as well as indications of spillovers between them, in the form of new combinations. Given CMR's largely applied orientation driven by demands of the war, the finding that it affected postwar basic science is surprising, contradicting traditional "linear model" logic, which posits that innovation flows downstream from science to applications (Godin 2006). Appendix C.3 provides a complementary lens on CMR's relationship to basic and clinical research, classifying publications into each category using multiple methodologies (described in Appendix B) and evaluating CMR's impacts on each. Appendix Figure C.4 reinforces that despite CMR's use-oriented focus, its effects on both life sciences research and clinical medicine were large.

4.5 Additional evidence

Appendix C presents additional results. In Appendix C.4 we examine heterogeneity in these effects across CMR divisions, programs, and categories of research performers (firms, universities, hospitals, etc.). Appendix Sections C.5 to C.7 provide additional robustness checks. In Appendix C.5, we estimate a variant of Equation (1) using intensive treatment measures—grouping subjects into quantiles of CMR funding—and find monotonically greater effects for more heavily-funded subjects. For the reasons enumerated above, we continue to prefer an extensive treatment measure, but we consider these results reinforcing. In Appendix C.6 we re-estimate Equation (1) controlling for whether a given subject was funded by the postwar NIH (between 1948 and 1970), and we find these results unchanged—suggesting that the CMR effect is distinct from the postwar NIH. In Appendix C.7 we probe the possibility that the "CMR shock" may in fact be a war shock—particularly if war has a demand-pull effect that brings scientific attention and activity to new subjects (even without CMR-style funding or coordination), and that in turn triggers accumulative endogenous growth. To do so, we collected analogous data around World War I (WWI), including digitizing contemporary lists of WWI medical problems, and estimate Equation (1) around WWI in relation to these subjects. We find little evidence of a generic effect of war on science. Finally, in additional tests we have re-estimated our main results on a sample 12-digit MeSH codes (rather than MeSH terms), with quantitatively and statistically similar results. MeSH code-based results also remain similar when we cluster standard errors at higher levels of the MeSH tree to account for potential interdependency in error structures across related subjects.

5 The Postwar Pharmaceutical Industry

Given the relatively primitive condition of the pre-war pharmaceutical industry, the possibility that the war would have any impacts on pharmaceutical innovation—let alone that CMR's efforts might succeed during the war itself—is far from straightforward.[23] Despite this, the first two postwar decades produced an immediate and sustained take-off in U.S. pharmaceutical innovation, in what is now seen as a "golden age" of drug discovery. Postwar progress was powered by growing use of synthetic chemistry, rational drug design, and systematic drug screens, and by the 1960s the U.S. pharmaceutical industry already looked much more similar to its current state than its pre-war condition. Several scholars have attributed these changes to wartime research in historical analysis: Landau et al. (1999, p. 63), for example, claims that "To a great extent the U.S. government's wartime policies led to the emergence of the American pharmaceutical industry as the

[23] The U.S. pharmaceutical industry in the early 20th century was primitive and disorganized by modern standards: most drug manufacturers were chemical companies with incidental or subsidiary pharmaceutical businesses, and drug discovery was driven more by serendipity or trial-and-error empiricism than by science. Perhaps as a result, as Figure 1 shows, annual drug-related patenting grew only meagerly pre-1940.

undisputed worldwide leader," observing that "the federal war effort encouraged corporate research and development, widened and deepened the companies' cooperation with academic institutions, and catalyzed the diffusion of new technologies across the industry."[24]

Perhaps no single firm exemplifies these changes better than Pfizer. Founded in 1849 by two German emigres, Pfizer was for most of its first 100 years a fine chemical manufacturer and commodity chemical supplier whose keystone product by the 1930s was citric acid, which it produced at scale through fermentation. Despite its lack of experience in drug development, Pfizer was brought into the wartime penicillin project for its expertise in fermentation, which was needed to produce natural penicillin from the *Penicillium notatum* mold at scale. Its success led it to become the U.S. Army's biggest penicillin supplier, and after the war ended, Pfizer pivoted around this experience and new R&D capability and entered the pharmaceutical industry, focusing first on developing a wider range of antibiotics and later expanding its R&D portfolio to other drug categories. As it did so, it grew increasingly scientifically oriented, employing a large staff of biologists, mycologists, and later organic chemists and developing consulting relationships with leading academic scientists, and soon became a leading drug developer (Daemmrich 2009).

To make systematic comparisons in relation to CMR, we harness the de Haen (1976) list of new drugs introduced between 1940 and 1975, which we manually link to 12-digit MeSH codes under the *therapeutic use* or *physiological effects* sub-branches of the *pharmacologic action* branch of the MeSH tree, and then through these links we retrieve associated MeSH terms. We use this crosswalk to produce a count annual new drugs associated with individual MeSH terms. Our analysis will examine changes in the rate of aggregate drug development across MeSH terms, comparing those which were a target of CMR-funded research vs. others. We then extend this analysis to the firm level, comparing the rate of new drug introductions by pharmaceutical firms (defined as firms in the De Haen sample) which were CMR contractors against those that were not. Here we will expand our treated set to include firms which were engaged in penicillin production under contract with WPB, most (but not all) of which were CMR contractors as well.

We continue using the same specification as in our analysis of CMR's effects on science, estimating time-varying differences across subjects or firms with vs. without CMR suport (Equation 1). Table 4 presents the results, estimating five-year (rather than annual) parameters, following the format of Table 3. Our estimation sample in this case runs from 1940 (when the De Haen drug statistics begin) to 1970, with 1940-1945 the omitted category. That the sample begins during the war is a constraint of the data, but insofar as there may be a CMR effect, defining the early 1940s as the reference period will generally only attenuate postwar differences.

[Table 4 about here]

[24] Also see, e.g., Temin (1979), Cockburn et al. (1999), and Pisano (2002).

We first report estimates from our term-level analysis in Panel (A), where the unit of observation is a term-year. We estimate the impact of CMR on (i) the likelihood of any new drug introductions related to a given MeSH term in a given year, (ii) the number of new drugs, and (iii) the IHS number of new drugs. Even-numbered columns exclude anti-infectives, as a test of whether antibiotic drugs were responsible for any overall effects. We do not find an effect of CMR on the extensive margin (Columns 1 and 2), in part because the many of the areas CMR emphasized already had relatively high propensity to produce at least one related drug per year in the aggregate. However, we find large effects on the intensive margin (Columns 3 to 6), with CMR-supported areas producing on average 1-2 more new drugs per year in the 1950s than others, and with similar-magnitude effects across both anti-infective drugs and other drug categories.[25]

Table 4, Panel (B) reproduces this analysis at the firm level, with similar (if not stronger) patterns. Relative to other drug-producing firms, CMR contractors produced a differentially-large surge of new drugs over the next two decades. At its peak, these firms were an extra 25 p.p. more likely to introduce at least one new drug per year than other firms, and on average introduced one more new drug per year—in both cases doubling or tripling mean rates. As with total drug innovation (Panel A), these differences are largest during the pharmaceutical industry's golden age in the 1950s and early 1960s, and cannot be attributed solely to antibiotics.[26]

Beyond developing new drugs, the war effort also supported R&D in new ways of producing them—especially natural drugs and vaccines which are produced through industrial-scale fermentation. To evaluate the sustained effects of this effort, we shift our analysis from drugs to drug-related patents, focusing on patents filed between 1930 and 1970 by firms in the De Haen sample (see Section 3). We particularly direct our attention to fermentation process patents, identified as patents in CPC subclass C12P ("Fermentation or enzyme-using processes to synthesize a desired chemical compound"), and patents related to biologically-produced drugs, identified using CPC subgroup Y10S 435/82 and its subsidiaries ("Subcellular parts of microorganisms", which encompasses bacterial and fungal cultures from which these drugs are produced). We estimate a similar regression to that in Table 3, evaluating the growth over time in fermentation patenting and biologically-manufactured drug patenting by CMR/WPB contractors versus other firms.

[Table 5 about here]

[25] These differences are visible in the raw data (e.g., in binned scatterplots). We estimate count outcomes (Columns 3 and 4) by OLS in part to simplify interpretation of average cross-firm differences in drug counts, and in part because count models (including Poisson models) suffer from an incidental parameters problem with two-way fixed effects, due to the limited number of subjects in the sample: Table 4(A) includes 46 MeSH terms, only 25 of which have at least one associated drug in our data and would thus enter a PPML estimation sample.

[26] That the magnitudes of the effects in Panel (A) are larger than those of Panel (B) on the intensive margin (# of new drugs), and smaller on the extensive margin (1(Any new drugs)), reflects the different level of aggregation, where MeSH terms are generally more aggregated than firms are, and therefore likely to produce larger magnitudes on the intensive margin and exhibit lower variation on the extensive margin.

The table estimates three variants of each outcome at the firm-year level, measuring an indicator for (i) any patent (Columns 1 and 2), (ii) the number of patents (Columns 3 and 4), and (iii) IHS patents (Columns 5 and 6), in all cases estimated by OLS to evaluate average differences between CMR firms and others. Results for fermentation patenting are provided in the odd-numbered columns, and for biological drug patents in the even numbered columns. The evidence indicates that in the 1950s and 1960s, firms which participated in the war effort disproportionately increased biological and fermentation-related drug patenting, filing patents in these classes at a roughly 60% higher rate by 1960, with no evidence of differential pretrends.

6 Impacts on Medical Practice

The most significant potential impact of wartime research was improving human health. Appendix A.1 provides suggestive evidence of its short-run effects on military morbidity and mortality, with hospitalizations and deaths from many common infectious diseases declining to near zero. These gains held large potential for postwar civilian health as well.

A crucial intermediate step for the adoption of new knowledge, techniques, and therapies in health care is getting frontier research findings to medical professionals (Phelps 1992, 2000). Historically, medical textbooks were used in both training and clinical practice (Greene 2007, Catillon 2017, Tomes 2021), especially before the academic publishing industry began to grow in the late 1960s and practitioner journals became more widely available. We therefore use historical textbook series to study the diffusion of CMR research into practice. As we described in Section 3, we digitized two textbook series that were published both before and after World War II: the *Cecil Textbook of Medicine* (CT) and the *Merck Manual of Diagnosis and Therapy* (MM). We obtained copies of each edition with an index between 1930 and ca. 1960 (comprising nine editions of CT spanning 1930-1959, and four editions of MM spanning 1940-1961), digitized each index, and mapped index entries to MeSH using MTI. For each MeSH term, we measure associated pages in each edition to assess coverage. Merging these data with MeSH term-level measures of the CMR shock, we use the specification in Equation (1) to estimate the differential growth of CMR-funded subjects' coverage in medical textbooks and manuals before versus after the war.

Table 6, Columns (1) to (3) present results for CT, where our data begin in 1930, and Columns (4) to (6) for MM, where our data begin in 1940—which are the omitted periods in each regression (respectively). We follow the structure of Section 4 in separately estimating effects for more vs. less heavily-developed subjects, using the same definitions.

[Table 6 about here]

We find that textbook coverage of CMR-funded subjects grew significantly in the postwar period, despite no differential pre-war growth, though it also took several years to realize these impacts. Although these subjects were slow to expand in the late 1940s, by the early 1950s their coverage had grown 20-30% more than other subjects in CT and 10-20% more in MM. These differences are similar for subjects that were more- and less-developed prior to the war, and relatively stable throughout the 1950s. To our knowledge, this table provides the first empirical evidence of publicly-funded medical research entering the knowledge base of medical practice. In doing so, it offers a lens into how information about frontier research historically diffused to practitioners and the delayed horizons over which it reaches practitioners. In the context of this paper, it also illustrates how CMR research converted into productive output (health care).

7 The Emergence of an Innovation System

The evidence thus far is broadly consistent with prior research documenting public-private R&D spillovers, while introducing new evidence on public R&D's practical impact. Yet these findings also point to several puzzles. One is CMR's broad, systemic impact—a result which is uncommon in empirical innovation studies, where effects are often more localized. Another is the scale of the impact, with CMR's relatively modest-sized program by modern standards—less than 1% of NIH's current annual extramural research budget, in inflation-adjusted terms—having field- and industry-shaping effects. A third is that despite World War II being a transient shock, it had long-lasting impacts. Our goal in this section is to explore possible reasons why.

The concurrence of the CMR shock with the takeoffs seen in Figure 1 is evocative of a "big push" theory of economic development—a specific cross-sectoral dynamic which scholars have used to explain industrialization and economic growth. Rosenstein-Rodan (1943) introduced the argument that industrialization requires a coordinated, economy-wide investment by an economic planner to increase sectoral productivity and wages, which will expand internal markets and in turn stimulate private investment. Murphy et al. (1989) later formalized this argument in a multiple equilibrium model, highlighting the coordination challenge and showing that simultaneously upgrading every sector's production technology or providing connective infrastructure can shift an economy from a (bad) equilibrium with low industrialization to a (good) equilibrium with high industrialization that could not be achieved by decentralized markets on their own.

This theory intuitively seems consonant with the evidence in the preceding sections, which documented an "industrialization" of biomedicine. But the Rosenstein-Rodan and Murphy et al. logic—which operates through static demand spillovers across sectors, as rising wages or increasing market access spur private investment, in response to market growth—does not structurally match the features of the innovation economy, which is thought to benefit from dynamic supply-side exter-

nalities such as cumulative innovation. More generally, a substantial literature in innovation studies has narratively examined the systemic nature of modern innovative activity, where "[t]echnological advance proceeds through the interaction of many actors" (Nelson and Rosenberg 1993, p. 15).

While Rosenstein-Rodan's insight is not necessarily incompatible with innovation externalities, it requires refinement to argue that innovation externalities produce similar equilibrium dynamics. We explore this possibility in Appendix E, where we present a model of growth in an innovation-driven economy with complementarities between research and development. In doing so, we borrow ideas from the popular narrative "innovation systems" framework (Freeman 1987, Lundvall 1992, Nelson 1993, and others) and integrate "big push" logic, adapting the Rosenstein-Rodan and Murphy et al. argument to innovation-based industries. The essence of this model is a complementarity between research and development, and the presumption that innovation thrives when these inputs are able to build on each other. We conjecture that the 1940s is when U.S. biomedical innovation began to do so, with its constituent parts—firms, universities, and government, or science and technology—becoming more deeply connected and mutually reinforcing. Our intuition is that this was instigated by the CMR/World War II shock, which shifted biomedical innovation into a new equilibrium of long-run growth around an accelerating science-technology flywheel.

The model is one in which scientific research (R) and technology development (D) are both needed to produce a final good (which in this paper can be interpreted as medicine). We assume that these inputs improve with investments in innovation, and that they each feed off progress in themselves and each other (similar to, e.g., Romer 1990). We thus take an ecosystem view of innovation, where science and technology structurally interdepend. The model produces two main predictions. The first is that innovation effort in each sector $\{R, D\}$ is a function of both (i) the state of development in the other sector and (ii) the degree to which participants in each sector can harness the latent potential for complementarities between them. The second prediction is that there are two equilibria: one with low science and technology stocks and/or little integration between them, where R&D investment remains perpetually low, and one where these constraints are removed, and a virtuous cycle of self-perpetuating, balanced growth can take off.

Essentially, in this model, emerging science-based technologies face a cold-start problem: no input sector will invest in innovation until it can build on the others. Doing so requires both (i) the state of science and technology to surpass a minimum level of development, and (ii) enough collaboration or absorptive capacity for each sector to build on the other. Echoing "big push" intuition, a large, coordinated R&D investment can shift the economy into the high-growth equilibrium. This is the effect we conjecture CMR had on biomedicine: introducing new capabilities and resources for each sector develop around, and linkages between them—analogous to the production technologies and infrastructure Murphy et al. (1989) argue were historically needed for economies to transition from

cottage to factory production but held up by coordination failures.

7.1 Growth around new capabilities and resources

Contemporary and historical accounts (see Appendix D) identify a number of long-lived research assets borne of the CMR effort which researchers continued using or building on after the war ended. By our reading, most can be grouped into six categories: new (1) therapies and therapeutic candidates, (2) research tools and techniques, (3) technology platforms, (4) research capabilities, (5) collaboration patterns, and (6) new fundamental knowledge.

The role of these capabilities and resources in perpetuating CMR's effect are qualitatively visible for both science and technology. The penicillin project is perhaps a canonical example, as it gave birth to a new class of drugs and spawned a "golden age" of pharmaceutical innovation around it (Swann 1983, Bud 2007). A similar evolution took place in corticosteroids (Achilladelis 1999). Historical accounts document that new approaches to and capabilities in drug screening, synthesis, and production enabled continued progress in these drug classes and others. Hoyt (2012, p. 74), for example, writes that many companies involved in wartime efforts to produce new vaccines "found themselves at an advantage after the war, since they had been forced to adopt new production methods," several of which had become "state of the art" by the 1960s. On the scientific research front, new drugs opened up new possibilities for scientific research on how they worked and what diseases they could treat. Elsewhere in CMR's portfolio, different programs developed new research instruments or basic understanding of new chemicals, organisms, and physiological phenomena, which provided a launchpad for further postwar research.

In Table 7 we empirically evaluate these claims. We do so first for science, where we build on our analysis in Section 4 by examining changes in the (IHS) annual number of publications paired with MeSH terms in specific sets of subjects representative of each mechanism. Each column compares changes in CMR-supported subjects vs. others by estimating variants of Equation (1). Column (1) does so for publications paired with drugs (sampling MeSH terms in drug-related sub-branches of the *pharmacologic actions* branch of the MeSH tree (D27.505)); Column (2), for new research tools and techniques (sampling the branches for investigative techniques, equipment, and supplies (E05, E07)); and Column (3), for new fundamental knowledge (sampling the branch for fundamental phenomena (G)). The results are consistent with the narrative evidence but add nuance: CMR drove a short-run surge in research around new drugs that receded in the 1960s, whereas impacts on new fundamental science were slower to emerge but longer-lasting.

[Table 7 about here]

We then consider mechanisms connecting CMR to postwar drug innovation, focusing first on firm

capabilities and industry-wide drug discovery platforms (e.g., Cockburn et al. 1999, Pisano 2002). Column (4) estimates changes over time in (IHS) new drug introductions by *non*-CMR/WPB firms in drug-related MeSH subjects with CMR funding, relative to other subjects, on the presumption that newfound technological opportunity (e.g., new approaches to antibiotic discovery and development) would support pharmaceutical innovation by non-CMR firms in drug categories CMR supported. Column (5) estimates changes in the annual probability that a firm engaged in CMR/WPB antibiotic research specifically develops new antibiotics, relative to other firms, which we interpret as reflecting firm capabilities. We find effects in both cases, albeit somewhat short-lived, reflecting the ebb and flow of the antibiotic era over the 1950s and 1960s.

7.2 Linkages: The growth of science-based drug discovery

A second channel through which CMR may have catalyzed long-run development in U.S. biomedicine is by strengthening linkages across the system, increasing the potency of complementarities. Evidence of growing linkages was already visible for science in Figure 4, which found spillovers across CMR-supported research subjects and others. CMR also forged a deeper link between R&D and manufacturing during the war, especially in biologically-produced drugs like vaccines or penicillin. Table 5 showed that this pattern extended beyond the war, as CMR-contracted firms continued co-developing biological products and new production methods.

A different, distinguishing feature of CMR-led drug development was its use of science. Penicillin production, for example, used science in identifying productive mold strains and their optimal growing conditions, and linked science with both engineering and manufacturing in designing and running industrial-scale fermentation systems. The effort to synthesize and test thousands of potential antimalarial drugs was shaped by basic understanding (and fundamental studies) of Plasmodium parasites and quinoline biochemistry, in an early application of rational drug design, which was pursued in tandem with conventional empirical trial-and-error.

Whether this pattern continued after the war is a testable, empirical question. The De Haen list of new drugs does not provide information on the underlying discovery process and its connections to science. We instead return to drug-related patents by firms in the De Haen sample, assessing their use of science via in-text citations to scientific literature.[27] Figure 5 suggests that very little

[27] "In-text" citations are references which appear in the patent disclosure. Bryan et al. (2020) argue that in-text citations may be a better measure of an invention's knowledge inputs than "front-page" citations to non-patent literature (NPL), whose legal purpose is identifying prior art against which claims are evaluated. More pragmatically, NPL citations only began to be included in patents beginning in February 1947, which coincides with the end of the war. In-text citations, on the other hand, date to the establishment of the modern Patent Office in 1836 and can thus be observed both before and after World War II. However, classifying citations as in-text or NPL is difficult for historical patents, where citations have to be extracted from unstructured, digitized text (in contrast to modern patents with structured files), and on manual review we have noticed several miscategorized examples. We therefore refine the Marx-Fuegi PCS data for use in this paper by removing likely false-positive in-text citations, which we identify by their location in the patent. See Appendix B for details.

pre-war drug innovation was linked to science, but World War II marked a structural change in the industry's use of science, which subsequently grew rapidly. These patterns are consistent with known changes in R&D practices, such as the growing application of rational drug design anchored in physical, chemical, and biological understanding. Our question is to what degree these changes could be causally connected to the CMR effort in World War II.

[Figure 5 about here]

To evaluate this question we disaggregate the sample into CMR/WPB contractors vs. other firms and compare across them. Figure 6, Panel (A) shows the average, annual likelihood of filing any non-science versus science-citing drug patents—an indicator that a firm is engaged in science-based drug development—in five-year bins, first for non-CMR firms (left half) and then for CMR firms (right half). The raw data suggest these changes were driven by CMR/WPB firms. We formalize these comparisons in Panel (B), which estimates a triple-difference regression, comparing (i) science vs. non-science based patenting by (ii) CMR/WPB-contracted firms vs. other firms (iii) before vs. after the war. The estimating equation takes the following form:

$$Y_{ist} = \sum_{t=1931}^{1970} \beta_t \cdot \mathbb{1}(\text{CMR/WPB firm})_i \cdot \mathbb{1}(\text{Science-based})_s + \alpha_{is} + \delta_t + \varepsilon_{it} \tag{2}$$

where i, s, and t index firms, science- vs. non-science based innovation, and years, and the sample runs from 1930 to 1970, with standard errors clustered by firm. The unit of analysis is a firm-year-invention type, and the outcome (Y_{ist}) measures patenting by firm i in year t that is or is not based in science ($s \in \{0, 1\}$). Because patenting was a low frequency event (the firms in our sample have an average of 1.3 drug-related patents per year, and 0.3 science-citing drug patents), our preferred specification evaluates whether a firm had any science-based patents in a given year; in Appendix C.8 we also examine patent intensity. In short, Figure 6(B) shows CMR/WPB firms were significantly more likely to generate science-based inventions after World War II, relative to non-science based invention, with no differential pre-war trends.

[Figure 6 about here]

To our knowledge, these figures provide the first systematic documentation of the transition to science-based innovation in the U.S. pharmaceutical sector, complementing historical case studies of particular drugs or firms, as well as studies of the rise of corporate scientific research in other sectors (e.g., Arora et al. 2024). The evidence reinforces the interpretation that CMR contributed to the development of a more integrated U.S. biomedical innovation system. In further analysis, we

find that this is also visible in drug patents' citations to other patents, which also began to grow after the war, and did so disproportionately for patents of CMR/WPB firms—though not as much or as quickly as patent citations to science were growing.[28]

8 Concluding Remarks

The high returns to medical research for human health, coupled with market failures in R&D, have led to biomedical research becoming a major focus of innovation policy in the U.S. and around the world. In this paper, we study one of the largest shocks to biomedical research in history: World War II. Recognizing that disease and other ailments presented an even larger threat to America's military than enemy forces, in the 1940s the U.S. government created the Committee on Medical Research (CMR) to fund and coordinate civilian R&D into military medical problems. The CMR effort was effectively the United States' first biomedical R&D policy, marking the U.S. government's first significant extramural funding for biomedical research. We show that the wartime effort triggered large, sustained growth in postwar scientific research, a surge of postwar pharmaceutical innovation, and postwar changes in medical practitioner knowledge in subjects it supported, while also laying the groundwork for postwar research policy (vis-à-vis NIH) and creating sinew across the sectors and institutions which drive progress in biomedicine today.

CMR's research portfolio does not neatly fit into the basic/applied dichotomy that typifies academic and policy understanding of the innovation process, as it spanned sectors, disciplines, and categories of R&D, and included investments in production and implementation. To understand why this effort had long-lasting impacts, we evaluate the evidence through the prism of a "big push" theory of industrial development, arguing that CMR endowed the scientific research and pharmaceutical sectors with a range of new resources while also strengthening linkages between them. The net effect was to initiate the transformation of U.S. biomedicine from its largely disparate and underdeveloped public, private, and university sectors to an integrated system, setting in motion a virtuous cycle of investment and growth which has continued to the present.

A range of additional findings provide supportive evidence for this interpretation, including evidence of growing interconnectedness within and across science and technology, and indications of specific CMR-borne resources that postwar science and technology developed around. The indications of systemic (rather than localized) effects may in turn be a consequence of CMR's integrated research

[28] Similar to our previous analysis, this analysis uses measures of in-text patent citations (from Verluise et al. 2020), which can be observed both before and after World War II. In browsing examples, many citations are firms referring to their own and others' processes for preparing specific inputs like biological cultures, such as U.S. patent 3,024,168 by Beecham for "Preparation of polyfluoroalkoxymethyl penicillins," which explains that "Culture of the medium containing the new precursor and the isolation of the new penicillin are conducted according to the methods known in the art; e.g. U.S. Patents 2,562,410 and 2,440,359"—the latter being an Eli Lilly patent on "Process and culture media for producing penicillin" filed near the end of the war, in August 1945.

approach, which distinguishes it from most research policy today (in biomedicine or otherwise), and seems to have created opportunities for high-value but hard-to-predict innovation to emerge.[29] The findings, in turn, have bearing for research policy in other settings, including today: the evidence from CMR suggests that integrative R&D policy models may be particularly impactful through their potential to create or stimulate new innovation ecosystems—an argument which dovetails with storied histories of specific policy institutions like DARPA (Bonvillian et al. 2019), or even the wider OSRD in World War II (Gross and Sampat 2023b).

Much as CMR did, our analyses raise new questions, fundamental and applied. Although we have specifically emphasized CMR's impacts, contemporary accounts from Bush (1945), Richards (1946), and others point to the importance of interwar science to the war effort, much of which was funded by foundations. Archival records from Rockefeller and other medical research funders of this era may help in assessing their impacts, as well as the dynamic relationships between fundamental and applied research and between science and technology. Second, we are cognizant that CMR is but one case of applied biomedical research policy. We anticipate that in the coming years, an analysis similar to ours on the post-crisis impacts of Operation Warp Speed or the broader Covid-19 pandemic response will be possible as well. Our results suggest that there may be long-lasting effects not just on technology, but on science itself. Finally, a natural policy question is whether a research funding organization similar to CMR would yield high social returns today, in a very different context and innovation system. While we cannot definitively speak to this question, our view is that the results from our analyses at least support calls for more experimentation with alternative funding models for biomedical research (Azoulay 2012, Myers 2023), including for late-stage research and commercialization (Ouellette 2023).

[29] It is telling that despite the large, latent civilian demand for antibiotics, it took a war to turn Fleming's discovery of penicillin into a mass-produced drug. Prior research attributes this breakthrough to a level of coordination and knowledge transfer not ordinarily achieved in other contexts (Neushul 1993).

References

Achilladelis, Basil. 1999. "Innovation in the pharmaceutical industry." In *Pharmaceutical innovation: Revolutionizing human health*, edited by Ralph Landau, Basil Achilladelis, and Alexander Scriabine. Philadelphia: Chemical Heritage Press.

Agarwal, Rajshree, Seojin Kim, and Mahka Moeen. 2021. "Leveraging private enterprise: Incubation of new industries to address the public sector's mission-oriented grand challenges." *Strategy Science* **6**(4):385-411.

Andrus, Edwin C. 1948. *Advances in military medicine*. Boston: Little, Brown, and Company.

Arora, Ashish, Sharon Belenzon, Konstantin Kosenko, Jungkyu Suh, and Yishay Yafeh. 2024. "The rise of scientific research in corporate america." *Organization Science* .

Azoulay, Pierre. 2012. "Turn the scientific method on ourselves." *Nature* **484**(7392):31-32.

Azoulay, Pierre, Erica Fuchs, Anna P. Goldstein, and Michael Kearney. 2019a. "Funding breakthrough research: Promises and challenges of the 'ARPA Model'." *Innovation Policy and the Economy* **19**(1):69-96.

Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019b. "Public R&D investments and private-sector patenting: Evidence from NIH funding rules." *Review of Economic Studies* **86**(1):117-152.

Azoulay, Pierre, Daniel P. Gross, and Bhaven N. Sampat. 2025. *Indirect Cost Recovery in U.S. Innovation Policy: History, Evidence, and Avenues for Reform*. NBER working paper no. 33627.

Balconi, Margherita, Stefano Brusoni, and Luigi Orsenigo. 2010. "In defence of the linear model: An essay." *Research Policy* **39**(1):1-13.

Baxter, James Phinney. 1946. *Scientists against time*. Boston: Little, Brown and Company.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A toolkit of policies to promote innovation." *Journal of Economic Perspectives* **33**(3):163-84.

Bonvillian, William B. 2024. "Operation warp speed: harbinger of American industrial innovation policies." *Science and Public Policy* **51**(6):1195-1211.

Bonvillian, William B, Richard Van Atta, and Patrick Windham. 2019. *The DARPA model for transformative technologies: Perspectives on the US defense advanced research projects agency*. Open Book Publishers.

Branscomb, Lewis M. 1992. "Does America need a technology policy?" *Harvard Business Review* **70**(2):24-30.

Brooks, Harvey. 1994. "The relationship between science and technology." *Research Policy* **23**(5):477-486.

Bryan, Kevin A, Yasin Ozcan, and Bhaven Sampat. 2020. "In-text patent citations: A user's guide." *Research Policy* **49**(4):103946.

Bud, Robert. 2007. *Penicillin: Triumph and tragedy*. Oxford: Oxford University Press.

Bush, Vannevar. 1945. *Science, the Endless Frontier: A report to the President*. Washington DC: Government Printing Office.

Catillon, Maryaline. 2017. *Medical knowledge synthesis: A brief overview*. Working paper.

Clemens, Jeffrey and Parker Rogers. 2023. "Demand shocks, procurement policies, and the nature of medical innovation: Evidence from wartime prosthetic device patents." *Review of Economics and Statistics* pp. 1-45.

Cockburn, Iain and Rebecca Henderson. 1996. "Public-private interaction in pharmaceutical research." *Proceedings of the National Academy of Sciences* **93**(23):12725-12730.

Cockburn, Iain, Rebecca Henderson, Luigi Orsenigo, and Gary Pisano. 1999. "Pharmaceuticals and biotechnology." In *U.S. industry in 2000: Studies in competitive performance*, edited by David C. Mowery, pp. 363-398. Washington DC: National Academies Press.

Cook-Deegan, Robert Mullan. 1996. "Does NIH need a DARPA?" *Issues in Science and Technology* **13**(2):25-28.

Creager, Angela N. H. 1999. "'What Blood Told Dr Cohn': World War II, plasma fractionation, and the growth of human blood research." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **30**(3):377-405.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The determinants of mortality." *Journal of Economic Perspectives* **20**(3):97-120.

Cutler, David M. and Srikanth Kadiyala. 2003. "The return to biomedical research: treatment and behavioral

effects." In *Measuring the gains from medical research: An economic approach*, edited by Kevin M. Murphy and Robert H. Topel. Chicago: University of Chicago Press.

Daemmrich, Arthur. 2009. "Synthesis by microbes or chemists? Pharmaceutical research and manufacturing in the antibiotic era." *History and Technology* **25**(3):237-256.

de Haen, Paul. 1976. "Compilation of New Drugs: 1940 thru 1975." *Pharmacy Times* **March**:40-74.

D'Souza, Arielle, Kendall Hoyt, Christopher M Snyder, and Alec Stapp. 2024. *Can Operation Warp Speed Serve as a Model for Accelerating Innovations Beyond Covid Vaccines?* NBER Working Paper No. 32831.

Durvasula, Maya, C Scott Hemphill, Lisa Larrimore Ouellette, Bhaven Sampat, and Heidi L Williams. 2023. "The NBER Orange Book dataset: A user's guide." *Research Policy* **52**(7):104791.

Foray, Dominique, David C Mowery, and Richard R Nelson. 2012. "Public R&D and social challenges: What lessons from mission R&D programs?" *Research Policy* **41**(10):1697-1702.

Fox, Daniel M. 1987. "The politics of the NIH extramural program, 1937-1950." *Journal of the History of Medicine and Allied Sciences* **42**(4):447-466.

Freeman, Christopher. 1987. *Technology Policy and Economic Performance: Lessons from Japan*. London: Pinter.

Geiger, Roger L. 1986. *To advance knowledge: The growth of American research universities, 1900-1940*. Oxford: Oxford University Press.

Godin, Benoît. 2006. "The linear model of innovation: The historical construction of an analytical framework." *Science, Technology, & Human Values* **31**(6):639-667.

Greene, Jeremy A. 2007. "Statins: The Abnormal and the Pathological—Cholesterol, Statins, and the Threshold of Disease." In *Medicating Modern America: Prescription Drugs in History*, edited by Andrea Tone and Elizabeth Siegel Watkins, pp. 181-228.

Gross, Daniel P. and Bhaven N. Sampat. 2022. "Crisis innovation policy from World War II to COVID-19." In *Entrepreneurship and Innovation Policy and the Economy*, edited by Josh Lerner and Scott Stern, volume 1, pp. 135-181.

Gross, Daniel P. and Bhaven N. Sampat. 2023a. "America, jump-started: World War II R&D and the takeoff of the US innovation system." *American Economic Review* **113**(12):3323-3356.

Gross, Daniel P. and Bhaven N. Sampat. 2023b. "The World War II crisis innovation model: What was it, and where does it apply?" *Research Policy* **52**(9):104845.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg. 2001. *The NBER patent citation data file: Lessons, insights and methodological tools*. NBER Working Paper No. 8498.

Hoyt, Kendall. 2006. "Vaccine innovation: Lessons from World War II." *Journal of Public Health Policy* **27**(1):38-57.

Hoyt, Kendall. 2012. *Long shot: Vaccines for national defense*. Cambridge: Harvard University Press.

Jacob, Brian A. and Lars Lefgren. 2011. "The impact of research grant funding on scientific productivity." *Journal of Public Economics* **95**(9-10):1168-1177.

Kattel, Rainer and Mariana Mazzucato. 2018. "Mission-oriented innovation policy and dynamic capabilities in the public sector."

Keefer, Chester S. 1969. "Dr. Richards as Chairman of the Committee on Medical Research." *Annals of Internal Medicine* **71**(8):61-70.

Kolev, Julian, Yuly Fuentes-Medel, and Fiona Murray. 2020. "Gender differences in scientific communication and their impact on grant funding decisions." In *AEA Papers and Proceedings*, volume 110, pp. 245-249.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson. 2021. "Who do we invent for? Patents by women focus more on women's health, but few women get to invent." *Science* **372**(6548):1345-1348.

Laitner, Melissa H, Audry M Huang, and Shannon Takala-Harrison. 2024. "The State of the U.S. Biomedical and Health Research Enterprise: Strategies for Achieving a Healthier America." National Academies Press.

Landau, Ralph, Basil Achilladelis, and Alexander Scriabine. 1999. *Pharmaceutical innovation: Revolutionizing human health*. Philadelphia: Chemical Heritage Foundation.

Lundvall, Bengt-Ake. 1992. *National systems of innovation: Towards a theory of innovation and interactive learning*. London: Pinter.

Marx, Matt and Aaron Fuegi. 2020. "Reliance on science: Worldwide front-page patent citations to scientific articles." *Strategic Management Journal* **41**(9):1572-1594.

Marx, Matt and Aaron Fuegi. 2022. "Reliance on science by inventors: Hybrid extraction of in-text patent-

to-article citations." *Journal of Economics & Management Strategy* **31**(2):369-392.
Mazzucato, Mariana. 2018. "Mission-oriented innovation policies: Challenges and opportunities." *Industrial and Corporate Change* **27**(5):803-815.
Mazzucato, Mariana. 2021. *Mission economy: A moonshot guide to changing capitalism*. New York: Harper Business.
Minneapolis Journal. 1935. *U.S. to War on Diseases.* Published August 30.
Mork, James G, Antonio Jimeno-Yepes, and Alan R Aronson. 2013. "The NLM Medical Text Indexer System for Indexing Biomedical Literature." In *BioASQ@ CLEF*.
Mowery, David C. 2009. "What does economic theory tell us about mission-oriented R&D?" In *The new economics of technology policy*, edited by Dominique Foray. Northampton: Edward Elgar.
Mowery, David C. 2012. "Defense-related R&D as a model for 'Grand Challenges' technology policies." *Research Policy* **41**(10):1703-1715.
Mowery, David C, Richard R Nelson, Bhaven N Sampat, and Arvids A Ziedonis. 2001. "The growth of patenting and licensing by U.S. universities: An assessment of the effects of the Bayh–Dole Act of 1980." *Research Policy* **30**(1):99-119.
Mowery, David C, Richard R Nelson, Bhaven N Sampat, and Arvids A Ziedonis. 2004. *Ivory tower and industrial innovation: University-industry technology transfer before and after the Bayh-Dole Act*. Palo Alto: Stanford University Press.
Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny. 1989. "Industrialization and the big push." *Journal of Political Economy* **97**(5):1003-1026.
Murphy, Kevin M and Robert H Topel. 2003. "The economic value of medical research." In *Measuring the gains from medical research: An economic approach*, edited by Kevin M. Murphy and Robert H. Topel. Chicago: University of Chicago Press.
Myers, Kyle. 2020. "The elasticity of science." *American Economic Journal: Applied Economics* **12**(4):103-134.
Myers, Kyle R. 2023. *Experimentation at the NIH*. Building a Better NIH Project whitepaper, available at https://www.brookings.edu/collection/building-a-better-nih/.
Nelson, Richard R. 1993. *National Innovation Systems*. New York: Oxford University Press.
Nelson, Richard R. 1997. "Why the Bush Report has hindered an effective civilian technology policy." In *Science for the 21st century: The Bush Report revisited*, edited by Claude E Barfield. Washington DC: American Enterprise Institute.
Nelson, Richard R and Nathan Rosenberg. 1993. "Technical innovation and national systems." In *National Innovation Systems*, edited by Richard R Nelson, pp. 3–22. New York: Oxford University Press.
Neushul, Peter. 1993. "Science, government and the mass production of penicillin." *Journal of the History of Medicine and Allied Sciences* **48**(4):371-395.
NLM. 2022. "NLM Medical Text Indexer (MTI)." https://lhncbc.nlm.nih.gov/ii/tools/MTI.html.
Nordhaus, William. 2003. "The health of nations: The contribution of improved health to living standards." In *Measuring the gains from medical research: An economic approach*, edited by Kevin M. Murphy and Robert H. Topel. Chicago: University of Chicago Press.
Ouellette, Lisa Larrimore. 2023. *Commercialization of NIH-supported research*. Building a Better NIH Project whitepaper, available at https://www.brookings.edu/collection/building-a-better-nih/.
Phelps, Charles E. 1992. "Diffusion of information in medical care." *Journal of Economic Perspectives* **6**(3):23-42.
Phelps, Charles E. 2000. "Information diffusion and best practice adoption." In *Handbook of health economics*, edited by Anthony J Culyer and Joseph P Newhouse, volume 1A, pp. 223-264.
Pisano, Gary. 2002. "Pharmaceutical biotechnology." In *Technological Innovation and Economic Performance*, edited by Benn Steil, David G Victor, and Richard R Nelson, pp. 347-365. Princeton: Princeton University Press.
Richards, A. N. 1946. "The impact of the war on medicine." *Science* **103**(2680):575-578.
Romer, Paul M. 1990. "Endogenous technological change." *Journal of Political Economy* **98**(5, Part 2):S71-S102.
Rosenberg, Nathan. 1974. "Science, invention and economic growth." *The Economic Journal* **84**(333):90-108.
Rosenberg, Nathan. 1982. *Inside the black box: Technology and economics*. Cambridge: cambridge University

Press.

Rosenberg, Nathan. 1991. "Critical issues in science policy research." *Science and Public Policy* **18**(6):335-346.

Rosenstein-Rodan, Paul N. 1943. "Problems of Industrialisation of Eastern and South-Eastern Europe." *The Economic Journal* **53**(210-211):202-211.

Rosenzweig, Robert M. 1998. "The politics of indirect costs." *Council on Governmental Relations* **August**:1-12.

Sampat, Bhaven N. 2012. "Mission-oriented biomedical research at the NIH." *Research Policy* **41**(10):1729-1741.

Sampat, Bhaven N. 2020. "Whose drugs are these?" *Issues in Science and Technology* **36**(4):42-48.

Sampat, Bhaven N. 2023. *The history and political economy of NIH peer review*. Building a Better NIH Project whitepaper, available at https://www.brookings.edu/collection/building-a-better-nih/.

Sampat, Bhaven N and Robert Cook-Deegan. 2021. "An ARPA for Health Research." *Milbank Quarterly Opinion* **30**.

Science Service. 1935. *70 Lines of Research Planned by Federal Health Service*. Records of the National Institutes of Health (RG443), National Institute of Health 1930-1948, Records Relating to NIH Divisions ("Divisions File"), Box 189, folder "Miscellaneous Matters - Social Security funds", accessed 2025.

Stewart, Irvin. 1948. *Organizing scientific research for war: The administrative history of the Office of Scientific Research and Development*. Boston: Little, Brown, and Company.

Swain, Donald C. 1962. "The rise of a research empire: NIH, 1930 to 1950." *Science* **138**(3546):1233-1237.

Swann, John Patrick. 1983. "The search for synthetic penicillin during World War II." *The British Journal for the History of Science* **16**(2):154-190.

Temin, Peter. 1979. "Technology, regulation, and market structure in the modern pharmaceutical industry." *The Bell Journal of Economics* pp. 429-446.

Tomes, Nancy. 2021. ""Not Just for Doctors Anymore": How the Merck Manual Became a Consumer Health "Bible"." *Bulletin of the History of Medicine* **95**(1):1-23.

Toole, Andrew A. 2012. "The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry." *Research Policy* **41**(1):1-12.

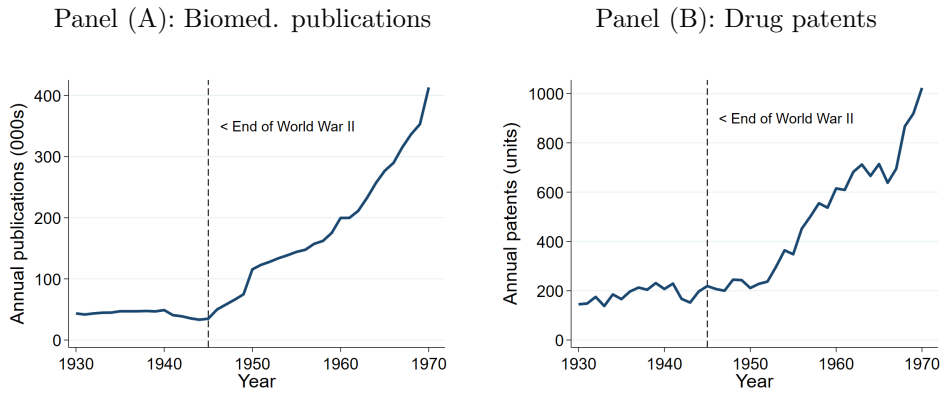
Urquiola, Miguel. 2020. *Markets, minds, and money: Why America leads the world in university research*. Cambridge: Harvard University Press.

U.S. Federal Trade Commission. 1980. *Statistical Report on Mergers and Acquisitions*. Produced by the Federal Trade Commission (FTC) Bureau of Economics.

Van Slyke, Cassius J. 1946. "New horizons in medical research." *Science* **104**(2711):559-567.

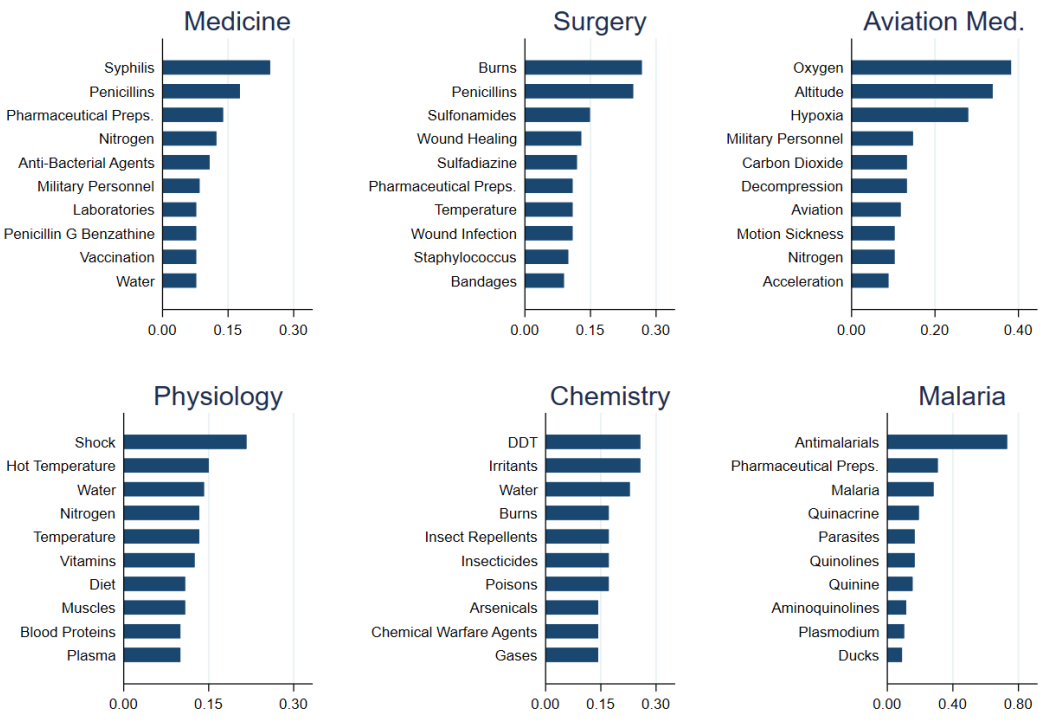
Verluise, Cyril, Gabriele Cristelli, Kyle Higham, and Gaétan de Rassenfosse. 2020. *The missing 15 percent of patent citations*. SSRN working paper 3754772.

Figure 1: Biomedical research publications and USPTO drug patents, 1930-1970



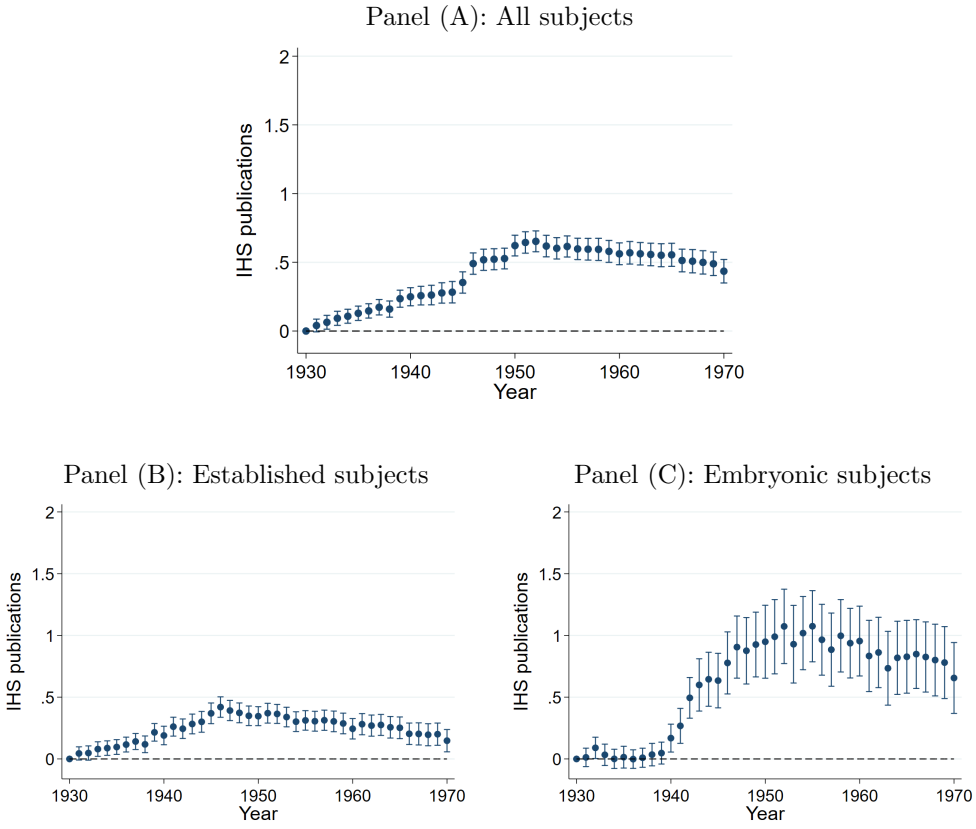
Notes: Left figure shows time series of total biomedical research publications in our core publication sample (Microsoft Academic Graph; see Appendix B). Sample consists of publications in the natural sciences and health sciences (OECD field codes 1 and 3) between 1930 and 1970. Right figure shows time series of total filings of drug patents in U.S. patent data, defined as patents in NBER category 31 ("Drugs"; Hall et al. 2001), corresponding to USPC 424 and 514. Dashed vertical line marks the end of World War II.

Figure 2: Top 10 MeSH terms by CMR division, as a fraction of division contracts



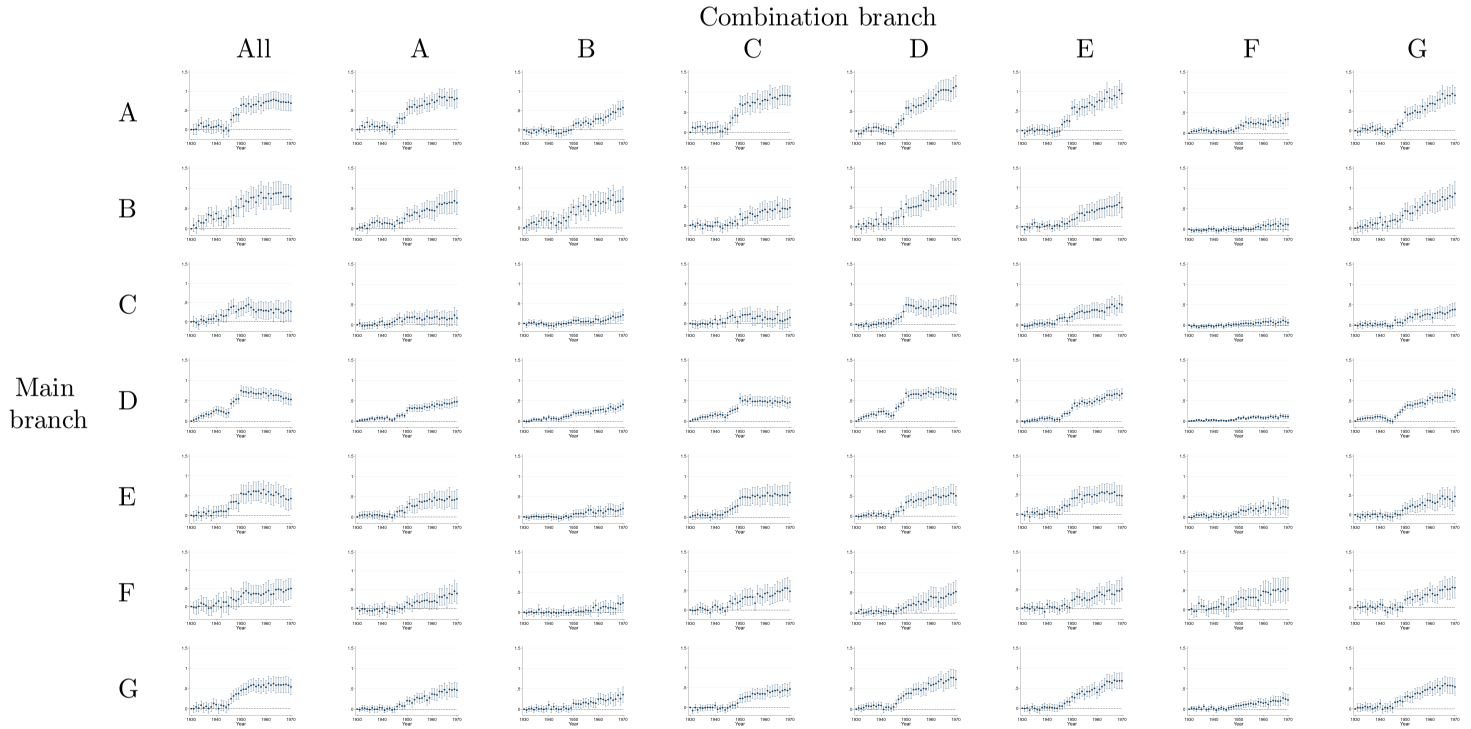
Notes: Figure lists the top 10 MeSH terms associated with CMR contracts in each of the six primary CMR divisions, showing what share of divisional contracts each term associates with. Individual contracts map to multiple MeSH terms.

Figure 3: Effects of CMR on research publications in treated subjects, 1930-1970



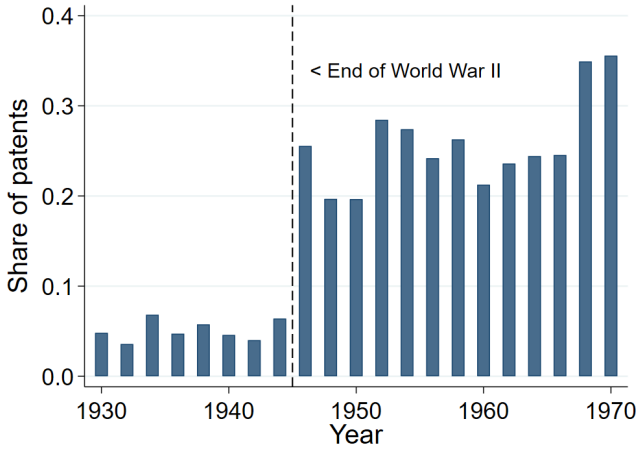
Notes: Figure shows annual estimates of the differential growth of scientific publications in MeSH terms with CMR funding, relative to others. Panel (A) does so for all MeSH terms. Panels (B) and (C) divide the sample into subjects with greater than and less than the median number of pre-1940 publications, which we label "established" and "embryonic" subjects. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

Figure 4: Effects of CMR on research publications in treated subjects, by MeSH tree branch (row) and combo branch (column)



Notes: Figure shows annual estimates of the differential growth of scientific publications in MeSH terms from each branch of the MeSH tree (rows A to G) with a CMR shock. We provide results for all publications with the term (column All) and for publications that combine the term with another from other branches of the tree (columns A to G). The MeSH tree can be browsed at https://meshb.nlm.nih.gov/treeView. Branches are defined as follows. A: Anatomy. B: Organisms. C: Diseases. D: Chemicals & Drugs. E: Techniques & Equipment. F: Psychiatry & Psychology. G: Phenomena & Processes. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

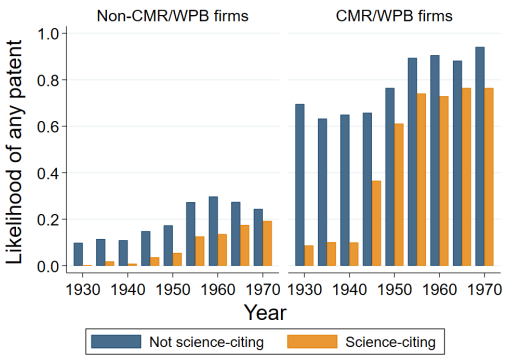
Figure 5: Science-citing share of drug patents, 1930-1970



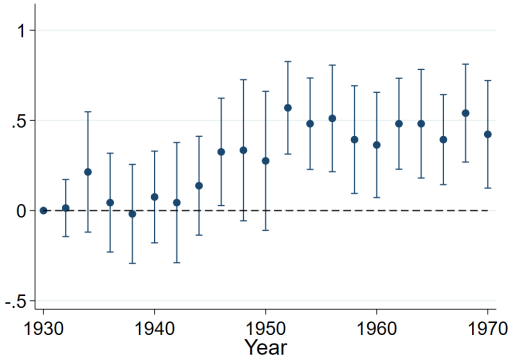
Notes: Figure shows the science-citing share of drug patents filed by firms in de Haen (1976). Drug patents defined as those associated with NBER category 31 (Hall et al. 2001), and science-citing patents as those which cite scientific literature in their text (Marx and Fuegi 2020, 2022; see text for further discussion of measures and measurement).

Figure 6: Differential growth in annual likelihood of filing science-citing vs. non-citing drug patents, by CMR/WPB-contracted firms vs. others, 1930-1970

Panel (A): Differences in the raw data



Panel (B): Regression evidence



Notes: Panel (A) shows the share of firm-years with at least one drug patent, in five-year intervals, separately reporting patent rates for CMR/WPB-contracted and other firms, and for their science-citing and non-citing patents. By the late 1950s, 75% of CMR/WPB firms were filing at least one science-citing drug patent each year (orange bars)—nearly as many as were filing non-science citing patents (blue bars)—whereas less than 20% of non-CMR/WPB firms were doing so. Drug patents defined as those associated with NBER category 31 (Hall et al. 2001), and science-citing patents as those which cite scientific literature in their text (Marx and Fuegi 2020, 2022). Panel (B) estimates the triple-difference, showing biannual estimates of the differential growth in the likelihood of filing science-citing versus non-citing patents at CMR/WPB-contracted firms versus others. Error bars represent 95% confidence intervals, with SEs clustered at the firm level.

Table 1: Extramural research contracts, contractors, and modal research subjects by CMR division

		Division						
		(1)	(2)	(3)	(4)	(5)	(6)	
	All	Medicine	Surgery	Aviat. Med.	Physiology	Chemistry	Malaria	Miscellany
CMR contracts	571	130	101	68	120	36	78	38
Percent of total	100%	23%	18%	12%	21%	6%	14%	7%
Unique contractors	128	55	39	31	56	20	49	27
Percent of total	100%	43%	31%	24%	44%	16%	38%	21%
Contract value (MMs)	$21.3	$3.4	$2.2	$2.4	$3.7	$1.2	$4.8	$3.5
Percent of total	100%	16%	11%	12%	18%	6%	22%	17%
MeSH terms per contract	11.3	11.5	13.4	10	11.8	14.9	9.1	6.5
Modal MeSH term	Penicillins	Syphilis	Burns	Oxygen	Shock	DDT	Antimalarials	Penicillins

Notes: Table provides summary statistics for CMR, overall and by division. The table lists the number of contracts, contractors, and total contract value by division, and provides the modal MeSH term from contracts in each division, weighted by value. The overall modal subject across all of CMR is *Penicillins*.

Table 2: Novelty, breadth, and impact of CMR-funded publications

	Novelty	Breadth (topics per publication)			Above given f. citation pctile.			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	New combo	# topics	≥3 topics	≥5 topics	75th pct.	90th pct.	95th pct.	99th pct.
CMR-funded	0.066***	0.114***	0.042***	0.006**	0.164***	0.103***	0.068***	0.024***
	(0.014)	(0.025)	(0.012)	(0.003)	(0.017)	(0.013)	(0.010)	(0.006)
N	493760	493760	493760	493760	230370	230370	230370	230370
R^2	0.20	0.23	0.19	0.14	0.25	0.23	0.21	0.19
Subj-Year FEs	Y	Y	Y	Y	Y	Y	Y	Y
Y mean	0.251	1.773	0.162	0.003	0.237	0.096	0.049	0.010

Notes: Table estimates differences between CMR-funded and contemporary publications in 20-year forward citations. Column (1) estimates differences in their propensity to include a new MeSH term combination (novelty); Columns (2) to (3), their number of associated MeSH topics (breadth); and Columns (5) to (8), their propensity to be top 25%, 10%, 5%, or 1% cited publications (impact). Sample is restricted to publications between 1940 and 1950. All specifications include fixed effects for publications' primary subject (highest-scoring MeSH term) and year. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by subject and year in parentheses.

Table 3: Effects of CMR on scientific publications, authors, and combinatoric innovation

			Author team composition:			
	(1)	(2)	(3)	(4)	(5)	(6)
	Pubs	Uniq. auths.	Only prewar	Any prewar	Any postwar	New combos
Any CMR * 1(1935-1939)	-0.006	-0.048				-0.067
	(0.019)	(0.111)				(0.089)
Any CMR * 1(1940-1945)	0.441***	0.517***	0.144***	0.245***	0.385***	0.305***
	(0.072)	(0.136)	(0.040)	(0.054)	(0.059)	(0.107)
Any CMR * 1(1946-1950)	0.860***	0.852***	0.121***	0.344***	0.867***	0.493***
	(0.118)	(0.163)	(0.046)	(0.077)	(0.114)	(0.129)
Any CMR * 1(1951-1955)	0.990***	0.799***	0.067*	0.269***	1.011***	0.537***
	(0.142)	(0.165)	(0.038)	(0.076)	(0.139)	(0.132)
Any CMR * 1(1956-1960)	0.920***	0.773***	0.005	0.169***	0.945***	0.526***
	(0.135)	(0.158)	(0.018)	(0.053)	(0.134)	(0.121)
Any CMR * 1(1961-1965)	0.788***	0.594***	0.005	0.106**	0.813***	0.352***
	(0.138)	(0.163)	(0.018)	(0.042)	(0.137)	(0.136)
Any CMR * 1(1966-1970)	0.755***	0.532***	-0.008	0.046	0.776***	0.355***
	(0.134)	(0.161)	(0.015)	(0.029)	(0.133)	(0.134)
N	335626	161708	335626	335626	335626	161708
R^2	0.55	0.63	0.14	0.17	0.58	0.42
Y mean	0.593	2.190	0.056	0.088	0.545	1.193

Notes: Table estimates differences over time in scientific activity in MeSH terms with CMR funding, relative to others. Evaluated outcomes are the number of scientific publications (Column 1); unique authors (Column 2); number of publications with only prewar authors (Column 3), any prewar authors (Column 4); and any postwar authors (Column 5); and new combinations (Column 6). Unique authors measured using author identifiers in the MAG data. Prewar authors defined as authors with at least one pre-1940 publication, and postwar authors those with no pre-1940 publications. New combinations are defined as new co-occurring MeSH terms in articles with the given MeSH term. All columns restrict the sample to "embryonic" subjects (with below-median pre-1940 publications), and Columns (2) to (6) condition to subject-years with $\geq$1 publication. Results are estimated relative to the 1930-1934 period. Note that Columns (3) to (5) omit the "Any CMR * 1935-1939" parameter due to a lack of pre-1940 variation in the outcome. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term in parentheses.

Table 4: Effects of CMR on new drug introductions, 1940-1970

Panel A: MeSH subject-year level

	1(Any new drugs)		# of new drugs		IHS(New drugs)	
	(1)	(2)	(3)	(4)	(5)	(6)
	All drugs	Excl. A-I	All drugs	Excl. A-I	All drugs	Excl. A-I
Any CMR * 1(1946-1950)	-0.011	-0.011	0.468	0.271	0.131	0.103
	(0.069)	(0.069)	(0.393)	(0.360)	(0.145)	(0.149)
Any CMR * 1(1951-1955)	0.141	0.141	1.420***	1.178**	0.469***	0.444**
	(0.087)	(0.087)	(0.518)	(0.487)	(0.163)	(0.170)
Any CMR * 1(1956-1960)	0.105	0.105	2.359**	2.078**	0.630***	0.612***
	(0.092)	(0.092)	(0.924)	(0.939)	(0.211)	(0.223)
Any CMR * 1(1961-1965)	0.146	0.146	0.543*	0.584*	0.267**	0.289**
	(0.100)	(0.100)	(0.301)	(0.317)	(0.127)	(0.132)
Any CMR * 1(1966-1970)	0.039	0.039	0.397	0.378	0.138	0.139
	(0.081)	(0.081)	(0.376)	(0.401)	(0.152)	(0.162)
N	1426	1426	1426	1395	1426	1395
R^2	0.57	0.57	0.62	0.57	0.67	0.63
Y mean	0.264	0.264	0.795	0.680	0.413	0.371

Panel B: Firm-year sample

	1(Any new drugs)		# of new drugs		IHS(New drugs)	
	(1)	(2)	(3)	(4)	(5)	(6)
	All drugs	Excl. A-I	All drugs	Excl. A-I	All drugs	Excl. A-I
CMR/WPB firm * 1(1946-1950)	0.205**	0.144**	0.446***	0.342***	0.295***	0.227***
	(0.080)	(0.065)	(0.140)	(0.121)	(0.097)	(0.083)
CMR/WPB firm * 1(1951-1955)	0.254***	0.303***	0.930***	0.768***	0.511***	0.475***
	(0.069)	(0.069)	(0.209)	(0.164)	(0.116)	(0.099)
CMR/WPB firm * 1(1956-1960)	0.222***	0.253***	0.977***	0.747***	0.514***	0.439***
	(0.072)	(0.076)	(0.246)	(0.210)	(0.127)	(0.119)
CMR/WPB firm * 1(1961-1965)	0.140**	0.123*	0.250**	0.261*	0.172**	0.172**
	(0.058)	(0.063)	(0.125)	(0.137)	(0.079)	(0.087)
CMR/WPB firm * 1(1966-1970)	0.111	0.089	0.079	0.090	0.074	0.077
	(0.074)	(0.059)	(0.111)	(0.072)	(0.080)	(0.055)
N	3686	3686	3686	3686	3686	3686
R^2	0.30	0.27	0.39	0.33	0.38	0.32
Y mean	0.183	0.157	0.274	0.220	0.204	0.169

Notes: Panel (A) estimates differences in the annual number of new commercially-marketed drugs in de Haen (1976) associated with MeSH terms with CMR funding, relative to others, restricting the sample to terms in drug-related sub-branches of the *pharmacologic actions* branch of the MeSH tree (see text for details). Panel (B) estimates differences in the annual number of new drugs brought to market by firms engaged in the CMR (and WPB) medical research effort vs. others. Even-numbered columns (labeled "Excl. A-I") exclude anti-infective agents to examine the degree to which the results are driven by antibiotics. In preparing the firm sample in Panel (B), we dynamically reassign a small number of firms which merged or were acquired during the sample frame to their subsequent owners using data from FTC (1980). Estimation sample begins in 1940 (the beginning of the de Haen (1976) historical drug data, which follow the FDA's initiation of drug safety approvals). Results are estimated relative to the 1940-1945 period. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term (Panel A) or firm (Panel B) in parentheses.

Table 5: Growth of biological approaches to drug development and manufacture

	Any patents?		Num. patents		IHS(Patents)	
	(1)	(2)	(3)	(4)	(5)	(6)
	Ferment.	Bio. drugs	Ferment.	Bio. drugs	Ferment.	Bio. drugs
CMR/WPB firm * 1(1935-1939)	-0.072	-0.066*	-0.233	-0.125	-0.110	-0.083
	(0.058)	(0.034)	(0.218)	(0.225)	(0.090)	(0.083)
CMR/WPB firm * 1(1940-1945)	-0.007	-0.008	-0.075	-0.022	-0.025	-0.004
	(0.065)	(0.058)	(0.256)	(0.209)	(0.110)	(0.093)
CMR/WPB firm * 1(1946-1950)	0.098	0.067	0.128	0.139	0.108	0.095
	(0.092)	(0.072)	(0.330)	(0.297)	(0.147)	(0.131)
CMR/WPB firm * 1(1951-1955)	0.203**	0.266***	0.590	0.746*	0.334*	0.418**
	(0.083)	(0.083)	(0.428)	(0.419)	(0.180)	(0.179)
CMR/WPB firm * 1(1956-1960)	0.307***	0.309***	1.415**	1.607**	0.607**	0.662***
	(0.105)	(0.097)	(0.710)	(0.719)	(0.255)	(0.248)
CMR/WPB firm * 1(1961-1965)	0.317***	0.358***	1.091**	1.545***	0.560**	0.710***
	(0.108)	(0.093)	(0.489)	(0.571)	(0.216)	(0.221)
CMR/WPB firm * 1(1966-1970)	0.265***	0.400***	0.520	1.341**	0.353**	0.681***
	(0.097)	(0.089)	(0.378)	(0.525)	(0.172)	(0.210)
N	5172	5172	5172	5172	5172	5172
R^2	0.38	0.42	0.38	0.45	0.44	0.50
Y mean	0.067	0.090	0.150	0.208	0.091	0.124

Notes: Table estimates differences in firms' annual number of drug patents on fermentation processes and on drugs produced from microorganisms, comparing engaged in the CMR (and WPB) medical research effort vs. others. Drug patents defined as those associated with NBER category 31 (Hall et al. 2001). Patents on fermentation processes are those classified into CPC C12P ("Fermentation or enzyme-using processes to synthesise a desired chemical compound or composition"), and patents associated to microorganisms are those classified into CPC Y10S/82 to Y10S/94 (all under "Microorganisms"). The firm smaple consists of firms in de Haen (1976). In preparing this sample, we dynamically reassign a small number of firms which merged or were acquired during the sample frame to their subsequent owners using data from FTC (1980). Results are estimated relative to the 1930-1934 period. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term (Panel A) or firm (Panel B) in parentheses.

Table 6: Cecil Textbook and Merck Manual coverage of subjects with CMR funding

	Cecil Textbook			Merck Manual		
	(1)	(2)	(3)	(4)	(5)	(6)
	All terms	Existing	New	All terms	Existing	New
Any CMR * 1933	-0.008	-0.019	-0.006			
	(0.018)	(0.021)	(0.033)			
Any CMR * 1937	0.058**	0.018	0.082			
	(0.026)	(0.029)	(0.064)			
Any CMR * 1940	-0.001	-0.018	0.017			
	(0.030)	(0.035)	(0.053)			
Any CMR * 1943	-0.014	-0.011	0.033			
	(0.032)	(0.037)	(0.058)			
Any CMR * 1947	0.055	0.010	0.129			
	(0.038)	(0.043)	(0.086)			
Any CMR * 1951	0.257***	0.179***	0.236***			
	(0.040)	(0.045)	(0.091)			
Any CMR * 1955	0.381***	0.223***	0.325***			
	(0.045)	(0.051)	(0.099)			
Any CMR * 1959	0.288***	0.185***	0.241***			
	(0.041)	(0.047)	(0.086)			
Any CMR * 1950				0.153***	0.093***	0.087***
				(0.019)	(0.022)	(0.032)
Any CMR * 1956				0.242***	0.155***	0.193***
				(0.023)	(0.026)	(0.045)
Any CMR * 1961				0.241***	0.152***	0.164***
				(0.024)	(0.027)	(0.045)
N	222885	73674	149211	99060	32744	66316
R^2	0.69	0.72	0.44	0.76	0.77	0.64
Y mean	0.207	0.485	0.070	0.080	0.202	0.020

Notes: Table estimates differences over time in whether the Cecil Textbook (Columns 1 to 3) and Merck Manual (Columns 4 to 6) cover individual MeSH terms with CMR funding, relative to others. The table divides the sample into MeSH terms above the below the median number of pre-1940 publications in Columns (2) to (3) and (5) to (6), respectively. The Cecil Textbook sample (Columns 1 to 3) consists of consecutive editions from 1930 (2nd ed., when an index is first observed) to 1959 (nearest edition to 1960). The Merck Manual sample consists of consecutive editions from 1940 to 1961 (on the same basis). Results are estimated relative to the 1930 and 1940 editions, respectively. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term in parentheses.

Table 7: Mechanisms: Example CMR-linked changes in postwar science and innovation

	Effects of CMR on science			Effects of CMR on new drugs	
	New combinations w/ CMR-funded:			Drug discovery	Firm-specific
	Drugs	Tools/Techs.	Knowledge	platforms	capabilities
	(1)	(2)	(3)	(4)	(5)
Any CMR * 1(1935-1939)	0.205	0.038	0.058*		
	(0.124)	(0.075)	(0.032)		
Any CMR * 1(1940-1945)	0.371**	0.155	0.066		
	(0.178)	(0.114)	(0.049)		
Any CMR * 1(1946-1950)	0.633***	0.392***	0.361***	0.086	0.076**
	(0.199)	(0.145)	(0.059)	(0.135)	(0.029)
Any CMR * 1(1951-1955)	0.754***	0.487***	0.526***	0.353***	0.116***
	(0.246)	(0.135)	(0.064)	(0.130)	(0.043)
Any CMR * 1(1956-1960)	0.503*	0.462***	0.495***	0.494**	0.138**
	(0.266)	(0.159)	(0.073)	(0.215)	(0.054)
Any CMR * 1(1961-1965)	0.423	0.353*	0.460***	0.212	0.018
	(0.300)	(0.183)	(0.085)	(0.133)	(0.051)
Any CMR * 1(1966-1970)	0.374	0.123	0.428***	0.054	0.008
	(0.301)	(0.227)	(0.090)	(0.148)	(0.040)
N	7216	34686	60065	1426	338892
R^2	0.77	0.79	0.85	0.58	0.07
Y mean	1.402	1.353	1.577	0.312	0.003
LHS variable	Combos	Combos	Combos	Drugs	Drugs
Transform.	IHS	IHS	IHS	IHS	Any

Notes: Table examines ways in which postwar biomedical science and drug development changed in relation to CMR. Columns (1) to (3) evaluate changes the annual rate of new MeSH term pairings (i.e., recombination) with CMR-funded subjects, relative to others. We do so separately for three categories of CMR-led innovation: new drugs (Column 1, sampling MeSH terms in drug-related sub-branches of the *pharmacologic actions* branch of the MeSH tree); new tools and techniques (Column 2, sampling terms on the branches for investigative techniques, equipment, and supplies); and new fundamental knowledge (Column 3, sampling on branches for phenomena and processes). Columns (4) to (6) evaluate changes in new drug introductions and drug patenting. Specifically, Column (4) estimates changes over time in new drug introductions *by non-CMR/WPB firms* in drug-related MeSH subjects with CMR funding, relative to other subjects, which we interpret as reflecting broader postwar technological opportunity. Column (5) estimates changes over time in the annual probability that a firm engaged in CMR/WPB antibiotic research specifically develops new antibiotics, relative to other firms, which we interpret as reflecting firm capabilities. In both these columns, results are estimated relative to the 1940-1945 period, when the de Haen (1976) sample begins. All specifications include fixed effects for publications' primary subject (highest-scoring MeSH term) and year (Column 1 to 4) or firm and year (Column 5). *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by subject or firm and year in parentheses.

Online Appendix

The Therapeutic Consequences of the War: World War II and the 20th Century Expansion of Biomedicine

Daniel P. Gross and Bhaven N. Sampat

Table of Contents

A Historical Context

In this section we provide historical context for this paper, including on the motivations for medical research in World War II, the origins of OSRD's Committee on Medical Research (CMR), and the data we collect on CMR's World War II-era research investments.

A.1 The need for military medical research

As we note in the paper, disease killed more soldiers than battlefield injuries in nearly every major war prior to World War II. The top panel of Table A.1 shows statistics on deaths from disease versus injury in eight military conflicts in various regions between the Mexican-American War (1846-48) and World War I (1914-1918; U.S. at war 1917-1918), where the ratio of deaths from disease:injury varied from 0.9 to 12, and averaged 1.7—including a ratio >1 for the U.S. in World War I. In World War II, this ratio was 0.07 for the U.S. Army overall and 0.01 for the U.S. Army in Europe. In essence, disease was effectively eliminated as a cause of military mortality.

Table A.1: Deaths from disease versus battlefield injury in prior wars

		Deaths from...		
	Years	Disease	Injury	Ratio
Pre-World War II				
Mexican-American War (U.S.)	1846-48	11,155	1,721	6.48
Crimean War (France)	1854-56	70,000	7,500	9.33
U.S. Civil War (Union)	1861-65	199,720	138,154	1.45
Franco-Prussian War (Germany)	1870-71	14,904	17,225	0.87
Sino-Japanese War (Japan)	1894-95	15,850	1,311	12.09
Philippine-American War (U.S.)	1899-1902	4,356	1,061	4.11
Boer War (British)	1899-1901	11,377	6,425	1.77
World War I (U.S.)	1917-18	51,447	50,510	1.02
Weighted average				1.69
World War II				
U.S. Army, total	1941-45	15,779	234,874	0.07
U.S. Army in Europe	1941-45	1,779	135,576	0.01

Notes: Table reports deaths from disease and battlefield injuries, and the ratio of the former to the latter, for the U.S. Army and foreign armies in major wars of the late 19th century, World War I, and World War II. Data from Coates and Hoff (1958), page 21, Table 7.

Table A.2 puts these patterns in further context for the U.S. Army specifically, comparing per-capita, per-year hospital admissions and deaths from infectious diseases in the Civil War (1861-65), Spanish-American War (1898), World War I (1917-18), and World War II (1942-45). Hospital admissions declined from >1 per soldier-year in the Civil War to roughly 0.1 per soldier-year in World War II, and deaths from 0.03 to 0.0015—a 99.5% decline.

Table A.2: U.S. Army admissions and deaths from infectious diseases, by war

War	Years	Per 1,000 soldiers/year Admissions	Deaths
U.S. Civil War (Union)	1861-65	1,030.34	34.77
Spanish-American War	1898	986.89	20.81
World War I	1917-18	427.03	10.43
World War II	1942-45	112.46	0.15

Notes: Table reports U.S. Army hospital admissions and deaths from infectious diseases in four major wars of the 19th and 20th centuries, per 1,000 soldiers per year. Data from Coates and Hoff (1958), page 11, Table 2.

Tables A.3 and A.4 provide a more detailed accountings of (i) per-capita hospital admissions rates and (ii) death rates per hospital admission in World Wars I and II, in both cases drawing from military medical statistics published in the same U.S. Army Medical Department series. We present admissions and death rates for specific major infectious diseases (which were subjects of CMR medical research in World War II) and compare them to cancer (which was not). The right-most columns in each table compute percentage changes from World War I to II for the U.S. Army worldwide, in the U.S., and in Europe (geographies broadly shared across the conflicts, albeit with some residual differences, such as northern vs. southern Europe—where the latter was a theater of war in the second World War but not the first).

Despite the implicit association to the research effort, direct comparisons of admissions and death rates across the two wars is complicated by changing theater of war (e.g., World War II was fought in more tropical environments with higher incidence of mosquito-borne diseases like malaria or Japanese B encephalitis) and changes in diagnostics. Similarly, attribution to World War II medical research specifically is also complicated by the possibility of interwar progress in specific disease areas (e.g., malaria). The evidence is nevertheless provocative and potentially suggestive of causal effects. Hospital admissions rates for nearly all infectious diseases declined dramatically, albeit with a few exceptions, primarily malaria and encephalitis (Table A.3). Death rates (per admission) also declined dramatically for most infectious diseases (Table A.4). Mortality from influenza, for example, declined 100%, and mortality from pneumonia, meningitis, and encephalitis—among the most lethal diseases for hospital admits in World War I—declined nearly as much. Large declines were also observed for other diseases like tuberculosis, typhoid fever, and scarlet fever. Cancer death rates, by comparison, declined only incrementally.

Table A.3: Disease hospital admissions in World War II vs. World War I

			Per 1,000 average strength, per annum								
			World War I (1917-19) Admissions			World War II (1942-45) Admissions			Pct. change Admissions		
Disease category	Etiology	Disease	Global	USA	EUR	Global	USA	EUR	Global	USA	EUR
Infectious diseases	Bacterial	Diphtheria	2.670	2.630	2.920	0.190	0.040	0.530	-93%	-98%	-82%
Infectious diseases	Bacterial	Scarlet fever	2.850	4.040	1.420	1.120	1.710	0.510	-61%	-58%	-64%
Infectious diseases	Bacterial	Tuberculosis	9.300	13.520	4.290	1.110	1.430	0.620	-88%	-89%	-86%
Infectious diseases	Bacterial	Typhoid fever	0.370	0.240	0.530	0.020	0.010	0.010	-95%	-96%	-98%
Infectious diseases	Viral	Influenza	191.560	238.700	137.150	7.430	9.290	3.460	-96%	-96%	-97%
Infectious diseases	Viral	Measles	23.650	38.200	5.500	2.380	3.650		-90%	-90%	
Infectious diseases	Viral	Mumps	56.120	63.360	49.140	4.020	5.530		-93%	-91%	
Infectious diseases	Viral	Rubella	4.160	7.230	0.350	5.300	8.400		27%	16%	
Infectious diseases	Bacterial or viral	Encephalitis	0.020	0.020	0.020	0.070	0.070	0.010	250%	250%	-50%
Infectious diseases	Bacterial or viral	Meningitis	1.300	1.400	1.260	0.560	0.740	0.430	-57%	-47%	-66%
Infectious diseases	Bacterial or viral	Pneumonia	19.030	20.540	18.030	10.680	12.990	8.090	-44%	-37%	-55%
Infectious diseases	Venereal	Chancroid	9.470	9.650	7.990	3.710	1.370	3.750	-61%	-86%	-53%
Infectious diseases	Venereal	Gonococcus	61.300	94.670	18.730	33.350	30.460		-46%	-68%	
Infectious diseases	Venereal	Syphilis	16.250	23.050	7.610	11.750	14.390	0.150	-28%	-38%	-98%
Infectious diseases	Parasitic	Malaria	3.450	4.700	0.570	15.930	3.530	6.470	362%	-25%	1035%
Neoplastic diseases		Cancer	0.150	0.180	0.100	0.300	0.330	0.270	100%	83%	170%

Notes: Table compares U.S. Army hospital admissions per 1,000 soldiers per year from select infectious diseases in World War I and World War II, with an added comparison to cancer (final row). Data reported for global personnel, U.S.-based personnel, and the European theater. World War I data reported for 1917-1919 only and World War II data for 1942-1945. Data from Love (1925), Tables 47 and 49 and Reister (1975), Tables 29, 29a, 31a.

Table A.4: Disease death rates in World War II vs. World War I

			Per 1,000 average strength, per annum								
			World War I (1917-19) Deaths:Admissions			World War II (1942-45) Deaths:Admissions			Pct. change Deaths:Admissions		
Disease category	Etiology	Disease	Global	USA	EUR	Global	USA	EUR	Global	USA	EUR
Infectious diseases	Bacterial	Diphtheria	0.015	0.015	0.017	0.026	0.025	0.032	76%	64%	87%
Infectious diseases	Bacterial	Scarlet fever	0.032	0.030	0.028	0.002	0.002	0.002	-94%	-92%	-93%
Infectious diseases	Bacterial	Tuberculosis	0.072	0.048	0.172	0.029	0.018	0.074	-60%	-62%	-57%
Infectious diseases	Bacterial	Typhoid fever	0.162	0.125	0.170	0.050	0.000	0.100	-69%	-100%	-41%
Infectious diseases	Viral	Influenza	0.031	0.031	0.032	0.000	0.000	0.000	-100%	-100%	-100%
Infectious diseases	Viral	Measles	0.025	0.023	0.038	0.000	0.001		-98%	-98%	
Infectious diseases	Viral	Mumps	0.000	0.000	0.000	0.000	0.000				
Infectious diseases	Viral	Rubella	0.005	0.004	0.000	0.000	0.000		-100%	-100%	
Infectious diseases	Bacterial or viral	Encephalitis	0.500	0.500	0.500	0.057	0.057	0.300	-89%	-89%	-40%
Infectious diseases	Bacterial or viral	Meningitis	0.392	0.343	0.444	0.045	0.043	0.047	-89%	-87%	-90%
Infectious diseases	Bacterial or viral	Pneumonia	0.244	0.216	0.282	0.004	0.004	0.005	-98%	-98%	-98%
Neoplastic diseases		Cancer	0.200	0.222	0.200	0.183	0.182	0.167	-8%	-18%	-17%

Notes: Table compares U.S. Army death rates per hospital admission from select infectious diseases in World War I and World War II, with an added comparison to cancer (final row). Data reported for global personnel, U.S.-based personnel, and the European theater. World War I data reported for 1917-1919 only and World War II data for 1942-1945. Data from Love (1925), Tables 47 and 49 and Reister (1975), Tables 29, 29a, 31a.

The impact of war on military health, and the improvements achieved by or in World War II, can be seen most clearly in the time series. Figure A.1 plots U.S. Army annual hospital admissions and deaths from infectious disease per capita between 1895 and 1955. Both series show significant spikes in earlier wars but no such spike in World War II.

Figure A.1: Admission and death rates for infectious diseases, U.S. Army, 1895-1954
(measured as number of admissions or deaths per 1,000 soldiers per year)

Notes: Figure shows time series of U.S. Army admission and death rates from infectious diseases, per 1,000 soldiers per year. Reproduced from Coates and Hoff (1958), page 20, Chart 6.

A.2 The Committee on Medical Research

As we recount in Section 2, the World War II research effort effectively began in June 1940, when U.S. President Franklin Roosevelt authorized and funded the creation of the National Defense Research Committee (NDRC) to coordinate civilian R&D in military science and technology. Based on its early successes, the NDRC was expanded into the Office of Scientific Research and Development (OSRD) in July 1941 by executive order and given formal appropriations. OSRD subsumed NDRC and created and added CMR as a second unit focused on coordinating and funding military medical research. In Gross and Sampat (2023b) we explain:

> [CMR] was charged with mobilizing medical researchers and identifying "the need for and character of contracts to be entered into with universities, hospitals, and other agencies conducting medical research activities," and was equally radical for its time.[1] Though the National Institute of Health (NIH) had existed since 1930, its budget was small and mostly spent in its own labs. Private foundations had previously funded medical research through block grants, and later (after the Depression made these financially infeasible) through grants to specific researchers. But as we discuss below, these were different in important ways from the CMR model, including their focus on fundamental research. CMR also drove a major shift in emphasis in medical research, away from peacetime problems to specific wartime medical needs ... Though there was some internal reorganization over the war, CMR's main divisions were General Medicine, Surgery, Aviation Medicine, Physiology, Chemistry, and Malaria.

In this prior work we explored more deeply how CMR operated. As a wartime research funding agency, CMR faced several basic questions which are characteristic of such efforts, including what to fund (priority-setting), whom to fund, and how to support it—including top-down vs. bottom-up priority-setting (whether to specify vs. solicit proposals), how large a role to play in managing and coordinating research efforts (vs. a more hands-off approach), and whether to invest in downstream activities like production and diffusion (vs. only R&D). Here we summarize CMR's general approach to these questions, drawing liberally from (Gross and Sampat 2023b), where we also observe that agency also accommodated exceptions when needed.

In contrast to NDRC, which identified specific R&D problems internally (based on the expertise of in-house staff and input from military representatives) and farmed out the work to suitable performers, CMR adopted a more bottom-up approach. It did so by circulating to the medical research community throughout the war bulletins identifying research priorities and soliciting "Proposals for Contract". Investigators submitting proposals were required to describe the subject of the proposed investigation, present the state of knowledge, explain its significance to national defense, and provide a research plan. These proposals were sent to a partner organization—the National Research

[1] Chester Keefer, the "penicillin czar", later described it as "a novel experiment in American medicine, for planned and coordinated medical research had never been essayed on such a scale" (Keefer 1969).

Council's Division of Medical Sciences (DMS), which has been created a year earlier in anticipation of war—where over thirty committees (with hundreds of elite medical researchers) reviewed applications for feasibility, in consultation with medical officers from the Army and Navy. Peer review was an "unprecedented approach" at the time, and CMR represented "the first sustained, large-scale exercise of the function in a biomedical context" (Mandel 1996). Based on the review feedback it received, the DMS gave each application a letter grade and submitted these reviews back to CMR, which screened them further for their possible impact on the war effort (in addition to feasibility) but typically funded what DMS recommended.

The executive order creating OSRD explicitly tasked it with coordinating wartime medical research, including across research performers and with the military. CMR undertook several activities towards this end. Several of the projects it funded or participated in were coordinated, multi-party or cross-sectoral attacks on specific problems, like blood separation and the preparation of blood substitutes. In these research programs, CMR actively managed the work it funded, including by organizing meetings of investigators to facilitate their cooperation, circulating technical reports, and continuously rebalancing its project portfolio (Stewart 1948).

CMR was also active in development, evaluation, and implementation. Even when there was initial evidence of the therapeutic benefits of new treatments from theory or animals, a key question was whether they worked in humans. Many of its contracts involved testing (e.g., of antimalarials, or an influenza vaccine). Members of the Army and Navy also helped arrange field trials on soldiers and reported back results. This user perspective helped facilitate bi-directional feedback, and ultimately utilization. In some cases, CMR supported manufacturing as well—most famously in the penicillin program, which we explore in depth in our prior work.

Further reading

Beyond our own work, several contemporary and historical accounts describe CMR's operation and the research it supported—many of which we have consulted in this and prior papers. Baxter (1946) and Stewart (1948) provide official histories of OSRD, and Andrus (1948) provides a recounting of OSRD medical research specifically. More recent writing has examined specific CMR-funded research efforts, such as in penicillin and antibiotics (e.g., Swann 1983, Neushul 1993), steroids (e.g., Achilladelis 1999), blood (e.g., Creager 1999), vaccines (e.g., Hoyt 2006), and malaria preventatives and treatments (e.g., Slater 2009), as well as links between CMR and postwar biomedical research policy (e.g., Strickland 1988, Mandel 1996, Sampat 2023).

The most comprehensive account of wartime medical research is, in our view, Andrus (1948), which is referenced heavily in Appendix D when we examine how features of specific CMR programs may explain their persistent effects. The data introduced in Appendix B are used to produce a more systematic accounting, complementing narrative histories.

B Data Appendix

B.1 OSRD/CMR data sources

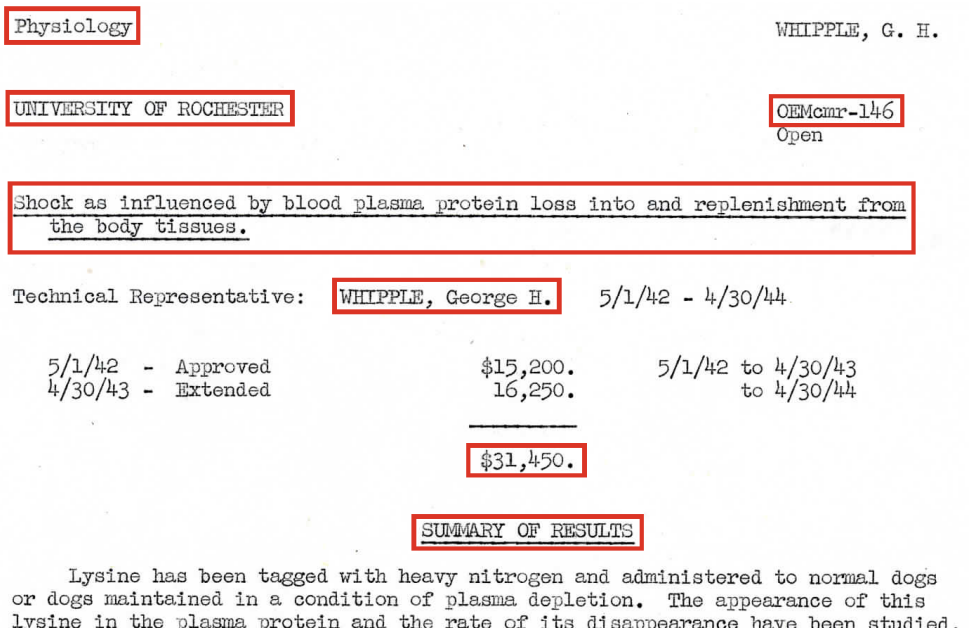
At the heart of this paper are new data on CMR-funded research obtained from OSRD archival records at the U.S. National Archives and Records Administration (NARA).[2] The principal data source for this paper is a set of CMR contract summaries, which identify the CMR division, contract number, principal investigator, institution, project title, value, award period, and a list of all technical reports, interim and final progress reports, and publications produced under the contract. For most contracts it also provided a summary of results. We cross-validate the information in these records against three other sources: (i) contract index in OSRD archival records, (ii) a contract list in the archival records, and (iii) a corresponding contract list in Andrus (1948). Figure B.1 provides a screenshot from an example CMR contract summary, highlighting key fields. Figure B.2 shows the corresponding index card from the OSRD contract index.

From the collective records we identify a total of 590 CMR research contracts: 573 extramural contracts to investigators at 128 universities, firms, hospitals, and institutes (identified in records as contracts OEMcmr-1 to OEMcmr-573), and 17 intramural contracts (primarily with the USDA, FDA, and NIH). The extramural contracts bore a total value of $21.3 million in 1940-45 dollars—around 5% of OSRD's total spending across all research contracts, most of which was comprised of NDRC developing physical war technologies like radar (see Gross and Sampat 2023a,b, Gross and Roche 2024)—equivalent to roughly $400 million in 2024.

We are also able to identify scientific publications produced under these contracts, using the union of (i) publications reported in contract summaries (like the one shown in Figure B.1), (ii) a "CMR bibliography" reported in Andrus (1948), and (iii) an independent CMR bibliography discovered in OSRD records which appears to have been the input for the Andrus list, though these sources mostly intersect. Figure B.3 provides excerpts from the archival and published bibliographies. We digitize and consolidate these bibliographies and link the results to Microsoft Academic Graph (MAG; 82% link rate), Web of Science (79% link rate), and PubMed (55% link rate, due to PubMed's limited coverage in this period). Though CMR publications provide insight into the shape of its research portfolio, throughout our analysis we prioritize contract-based measures over publication-based measures, as they indicate research inputs rather than output.

[2]See NARA Record Group 227, "Records of the Office of Scientific Research and Development". The key record set for this paper are the Committee on Medical Research *Contract Ledgers* (NC-138, Entry 164, Stack area 130, Row 22, Compartment 18, Shelf 2-3, Boxes 1-6). Also relevant were OSRD's *Index to Contracts* (NC-138, Entry 27, Stack area 130, Row 20, Compartment 11, Shelf 1, Boxes 1-5) and lists of CMR contracts and CMR-funded publications found in NARA records which we cross-validate against lists presented in Andrus (1948).

Figure B.1: Example medical research contract summary



Physiology

WHIPPLE, G. H.

UNIVERSITY OF ROCHESTER

OEMcmr-146
Open

Shock as influenced by blood plasma protein loss into and replenishment from the body tissues.

Technical Representative: WHIPPLE, George H. 5/1/42 - 4/30/44

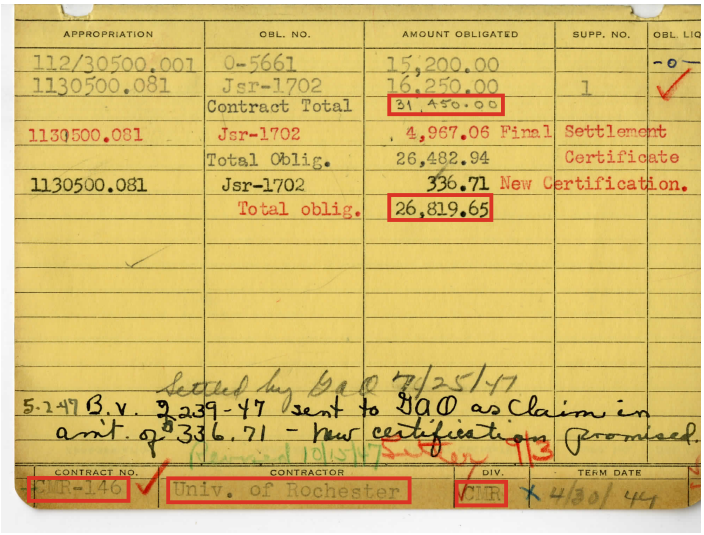
5/1/42 -	Approved	$15,200.	5/1/42 to 4/30/43
4/30/43 -	Extended	16,250.	to 4/30/44
		$31,450.	

SUMMARY OF RESULTS

Lysine has been tagged with heavy nitrogen and administered to normal dogs or dogs maintained in a condition of plasma depletion. The appearance of this lysine in the plasma protein and the rate of its disappearance have been studied.

Notes: Figure shows an excerpt from an example contract summary report from the CMR contract ledgers, for contract OEMcmr-146. The headings provide basic contract information, including the contract number, institution, subject, principal investigator, dates, and total obligated value. Also included in these reports were extended abstracts and publication lists.

Figure B.2: Corresponding contract index card



APPROPRIATION	OBL. NO.	AMOUNT OBLIGATED	SUPP. NO.	OBL. LIQ
112/30500.001	0-5661	15,200.00		-0-
1130500.081	Jsr-1702	16,250.00	1	✓
	Contract Total	31,450.00		
1130500.081	Jsr-1702	4,967.06 Final	Settlement	
	Total Oblig.	26,482.94	Certificate	
1130500.081	Jsr-1702	336.71 New	Certification.	
	Total oblig.	26,819.65		

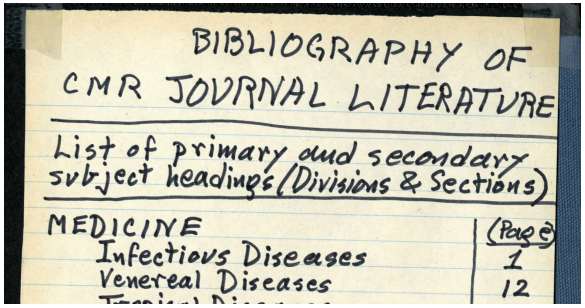
Settled by GAO 7/25/47
5-1-47 B.V. 9239-47 sent to GAO as claim in amt. of $336.71 - new certification promised.

CONTRACT NO.	CONTRACTOR	DIV.	TERM DATE
OEMcmr-146 ✓	Univ. of Rochester	CMR	X 4/30/44

Notes: Figure shows the corresponding index card from the OSRD contract index. The card identifies the contract number, contractor, OSRD division (in this case, "CMR"), and the total contracted and obligated value.

Figure B.3: CMR bibliography, archival and published copies

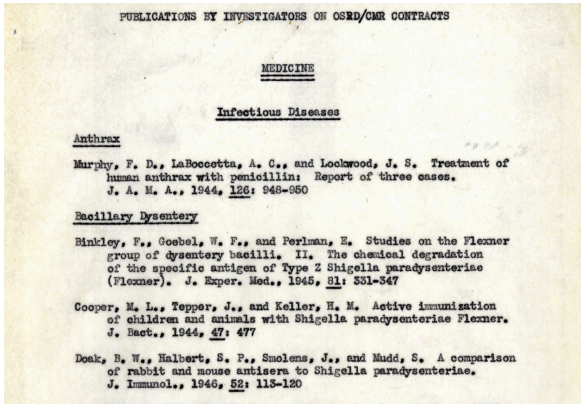
Archival records:



BIBLIOGRAPHY OF
CMR JOURNAL LITERATURE

List of primary and secondary subject headings (Divisions & Sections)

	(Page)
MEDICINE	
Infectious Diseases	1
Venereal Diseases	12



PUBLICATIONS BY INVESTIGATORS ON OSRD/CMR CONTRACTS

MEDICINE

Infectious Diseases

Anthrax

Murphy, F. D., LaBoccetta, A. C., and Lockwood, J. S. Treatment of human anthrax with penicillin: Report of three cases. J. A. M. A., 1944, 126: 948-950

Bacillary Dysentery

Binkley, F., Goebel, W. F., and Perlman, E. Studies on the Flexner group of dysentery bacilli. II. The chemical degradation of the specific antigen of Type Z Shigella paradysenteriae (Flexner). J. Exper. Med., 1945, 81: 331-347

Cooper, M. L., Tepper, J., and Keller, H. M. Active immunization of children and animals with Shigella paradysenteriae Flexner. J. Bact., 1944, 47: 477

Doak, B. W., Halbert, S. P., Smolens, J., and Mudd, S. A comparison of rabbit and mouse antisera to Shigella paradysenteriae. J. Immunol., 1946, 52: 113-120

Andrus (1948):

BIBLIOGRAPHY

Publications by Investigators on OSRD/CMR Contracts

MEDICINE

Infectious Diseases

Anthrax

Murphy, F. D., LaBoccetta, A. C., and Lockwood, J. S. Treatment of human anthrax with penicillin: Report of three cases. *J.A.M.A.*, 1944, *126:* 948–950.

Bacillary Dysentery

Binkley, F., Goebel, W. F., and Perlman, E. Studies on the Flexner group of dysentery bacilli. II. The chemical degradation of the specific antigen of Type Z Shigella paradysenteriae (Flexner). *J. Exper. Med.*, 1945, *81:* 331–347.

Cooper, M. L., Tepper, J., and Keller, H. M. Active immunization of children and animals with Shigella paradysenteriae Flexner. *J. Bact.*, 1944, *47:* 477.

Doak, B. W., Halbert, S. P., Smolens, J., and Mudd, S. A comparison of rabbit and mouse antisera to Shigella paradysenteriae. *J. Immunol.*, 1946, *52:* 113–120.

Dubos, R. J., Hoberman, H. D., and Pierce, C. Some factors affecting the toxicity of cultures of Shigella dysenteriae. *Proc. Nat. Acad. Sc.*, 1942, *28:* 453–458.

Dubos, R. J., Straus, J. H., and Pierce, C. The multiplication of bacteriophage in vivo and its protective effect against an experimental infection with Shigella dysenteriae. *J. Exper. Med.*, 1943, *78:* 161–168.

Ehrich, W. E., Halbert, S. P., Mertens, E., and Mudd, S. Mechanism of the augmenting acting of mineral oil on antibody production. Tissue reactions and antibody response to dysentery vaccine in saline, and in saline-lanolin-mineral oil emulsion. *J. Exper. Med.*, 1945, *82:* 343–360.

Goebel, W. F., Binkley, F., and Perlman, E. Studies on the Flexner group of dysentery bacilli. I. The specific antigens of Shigella paradysenteriae (Flexner). *J. Exper. Med.*, 1945, *81:* 315–330.

Goebel, W. F., Perlman, E., and Binkley, F. Antibody response in man to injection of the specific antigen of Type V Shigella paradysenteriae. *Science*, 1944, *99:* 412–413.

Notes: Figure shows excerpts from the CMR bibliography (publication list). On the left are archival records; on the right is the matching published version in Andrus (1948).

B.1.1 Linking CMR to the MeSH vocabulary

As we explain in Section 3 of the paper, we use the National Library of Medicine's (NLM) Medical Text Indexer (MTI) to connect CMR contracts to medical subjects, measured via MeSH descriptors (which we alternatively call "MeSH terms"). MeSH is the NLM's controlled and hierarchically-organized vocabulary used for indexing and cataloging biomedical and health-related research, and MTI is a natural language processing tool which identifies candidate MeSH terms for journal article indexing and serves as the first-line indexer for a large number of journals. We repurpose MTI to index CMR contracts in our setting, on the basis of their titles and contract summaries. The resulting output identifies associated medical subjects, a relevance score for each subject, and associated MeSH tree codes, identifying subjects' hierarchical relationship.[3]

MTI indexing of CMR contracts (and of MAG publications, as we discuss below) was performed between 2021 and 2022.[4] Figure B.4 provides a screenshot showing the structure of the MeSH tree; though most of our analysis occurs at the MeSH term level, in robustness checks we also make use of this hierarchical taxonomy for medical research space.

Figure B.4: Sample from Medical Subject Heading (MeSH) tree

Notes: Figure displays the MeSH tree, which assigns hierarchical structure to medical subject headings (see paper for a more complete discussion). Left image shows the main branches. Right image illustrates select subcategories of the 'A' branch.

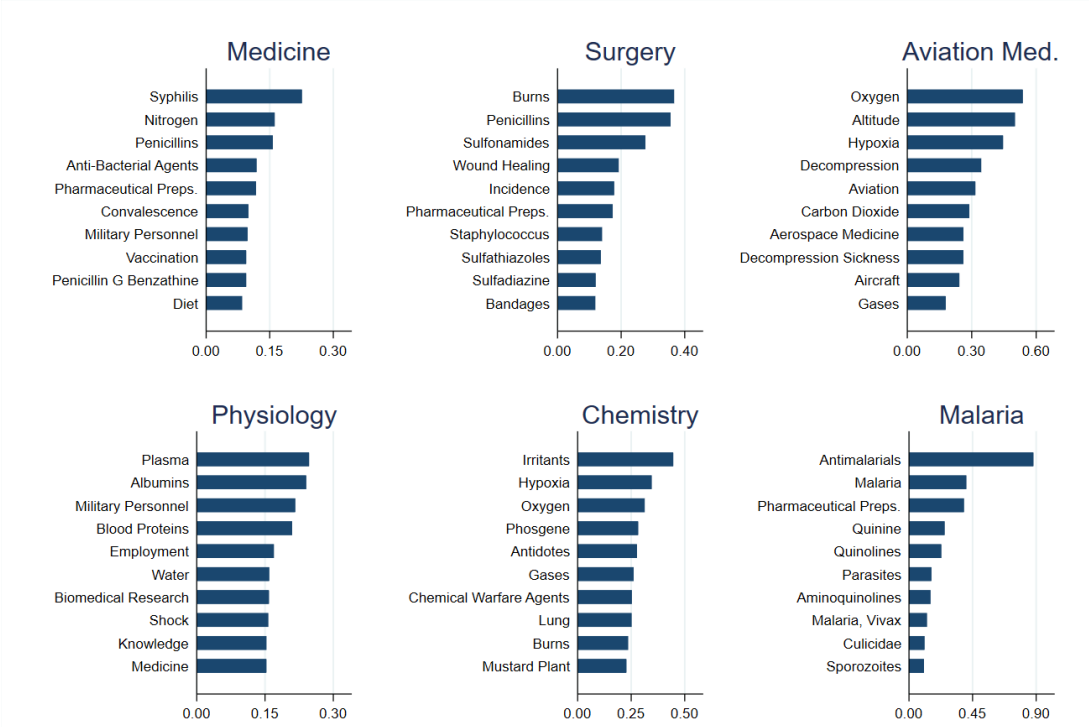
[3]The MeSH vocabulary has roughly 30,000 unique terms (descriptors) and 60,000 tree codes. Some terms link to multiple tree codes, reflecting relationships to distinct regions of biomedical idea space.

[4]Note that as of January 2025, MTI has been decommissioned, and is no longer available to the public. NLM has transitioned internal indexing to a "next-generation" replacement indexer named "MTIX" which uses neural network-based prediction methods, but MTIX is not currently available to the public. In an email exchange, NLM staff explained that there is no timeline for a public release.

B.1.2 What subjects did CMR support?

Figure 2 of the paper provides a lens into research subjects CMR funded, listing subjects in each CMR division with the most associated contracts. Here we provide two additional views of CMR's portfolio, empirical and qualitative. Figure B.5 reproduces Figure 2, showing the top subjects weighted by funding. Table B.1 presents the table of contents to Andrus (1948)—a collection of postwar summaries of CMR research programs—as a qualitative window into the research subjects CMR funded, including areas where it failed to make meaningful progress.

Figure B.5: Top 10 MeSH terms by CMR division, value-weighted



Notes: Figure lists the top 10 MeSH terms associated with CMR contracts in each of the six primary CMR divisions, showing what share of divisional contract value each term associates with. Individual contracts map to multiple MeSH terms.

Table B.1: Andrus (1948) Table of Contents

Section	Chapter
PART ONE—Medicine	
	Chapter 01 – Introduction
	Chapter 02 – Infectious Diseases
	Chapter 03 – Venereal Diseases
	Chapter 04 – Tropical Diseases
	Chapter 05 – Medical Problems of Convalescence
PART TWO—SURGERY	
	Chapter 06 – Introduction
	Chapter 07 – The Prevention of Infection in Accidental Wounds
	Chapter 08 – Experimental Wound Healing
	Chapter 09 – The Application of Penicillin to Surgical Problems
	Chapter 10 – Orthopedic Problems and Prosthetics
	Chapter 11 – The Problem of Gas Gangrene
	Chapter 12 – The Burn Problem
	Chapter 13 – The Repair of Peripheral Nerve Lesions
	Chapter 14 – Experimental Studies on Concusion
	Chapter 15 – Frostbite and Trench Foot
	Chapter 16 – New Surgical Plastics and Hemostatics
	Chapter 17 – Improvements in X–Ray Devices
	Chapter 18 – Studies on Wound Ballistics
PART THREE—AVIATION MEDICINE	
	Chapter 19 – Introduction
	Chapter 20 – The Study of Crash Injuries and Prevention of Aircraft Accidents
	Chapter 21 – The Effects of Acceleration and Their Amelioration
	Chapter 22 – Visual Problems
	Chapter 23 – Motion Sickness
	Chapter 24 – Anoxia and Oxygen Equipment
	Chapter 25 – Altitude Decompression Sickness
PART FOUR—PHYSIOLOGY	
	Chapter 26 – Introduction
	Chapter 27 – Shock
	Chapter 28 – The History of Plasma Fractionation
	Chapter 29 – Blood Substitutes
	Chapter 30 – Methods of Preservation of Whole Blood
	Chapter 31 – Problems of Nutrition
	Chapter 32 – Acclimatization to Heat and Cold
	Chapter 33 – Protective Clothing
	Chapter 34 – Water Disinfection and Allied Subjects
PART FIVE—CHEMICAL WARFARE AGENTS	
	Chapter 35 – Introduction
	Chapter 36 – Systemic Agents: Action and Treatment
	Chapter 37 – Recent Research on Respiratory Irritants
	Chapter 38 – Protection and Treatment of the Skin Exposed to Blister Gases
	Chapter 39 – The Effects of Toxic Chemical Agents on the Eye and Their Treatment
PART SIX—ANTI—PEST AGENTS	
	Chapter 40 – The Development of New Insecticides
	Chapter 41 – The Action of DDT on Invertebrates
	Chapter 42 – The Toxicology and Mechanism of Action of DDT in Mammals
	Chapter 43 – The Dispersal of Insecticides
	Chapter 44 – The Development of New Insect Repellents
	Chapter 45 – The Development of New Rodenticides
PART SEVEN—ADRENOCORTICOL STEROIDS	
	Chapter 46 – The Synthesis of Adrenocortical Steroids
PART EIGHT—MALARIA	
	Chapter 47 – Introduction
	Chapter 48 – The Synthesis of Antimalarial Drugs
	Chapter 49 – The Biology and Biochemistry of the Malarial Parasites
	Chapter 50 – The Screening Program
	Chapter 51 – The Clinical Testing of Antimalarial Drugs
PART NINE—PENICILLIN	
	Chapter 52 – Penicillin: A Wartime Achievement
	Chapter 53 – Research in the Development of Penicillin
PART TEN—SENSORY DEVICES	
	Chapter 54 – Sensory Devices

B.2 Scientific publication data

The core outcome data for this paper are biomedical publications, aggregated to the MeSH term-year level. Our base data consists of scientific publications from Microsoft Academic Graph (MAG) which we obtained from the Reliance on Science data repository (Marx and Fuegi 2020, 2022).[5] We filter this sample to publications in the natural and health sciences (OECD field codes 1 and 3) between 1930 and 1970. We then retrieved the titles of these publications and processed them through MTI to obtain MeSH subjects and associated weights. We then work with the MTI output. After dropping check tags (MeSH descriptors that specify species, sex, or age), supplementary concepts (terms outside of the MeSH thesaurus), low-scoring returned terms (specifically, those with less than 10% of a publications' remaining MeSH terms' total score) to reduce noise, we then aggregate to count score-weighted publications by subject and year. Using these data we also produce derivative measures, such as the number of unique other MeSH terms a given MeSH term co-publishes with in a given year, which we label "combinations". We measure the subset of these combinations which are new to the publication record in a given year as "new combinations", and count for each MeSH term-year its number of new combinations.

An important added step in our data collection is also to classify articles as basic or clinical (applied). As we explain in the body of the paper, there is no commonly agreed method of systematically distinguishing basic and clinical research in biomedicine. Prior research has proposed categorizations based on journal (Narin et al. 1976), title (Lewison and Paraje 2004), or content (Li et al. 2017, Ke 2019), or via machine learning (Boyack et al. 2014). We consider three ways of doing so. First, we rely on OECD subfield codes included in the MAG data which identify articles as basic medical research (3.01) and clinical medicine (3.02), via journals. Second, we identify journals indexed by two historical publications, Current Contents: Life Sciences (CC:LS) and Current Contents: Clinical Practice (CC:CP), from a leading commercial indexing service (the Institute for Scientific Information, or ISI) that were intended to cover basic and clinical research, respectively (Garfield 1972, Cardoni 1973). Concretely, we obtained a list of 702 journals from CC:CP as of January 1973 (when it was first published, and shortly after our sample period ends), and over 1,000 journals from CC:LS in 1973 (for consistency). We hand-matched these journals to those in MAG, successfully linking 80% of the former to the latter. Articles in the linked journals are then categorized as basic, clinical, mixed, or neither based on whether they appear in CC:LS or CC:CP. Our third approach applies the term lists in Lewison and Paraje (2004) to identify articles as basic or clinical based on the presence of those terms in their titles.[6]

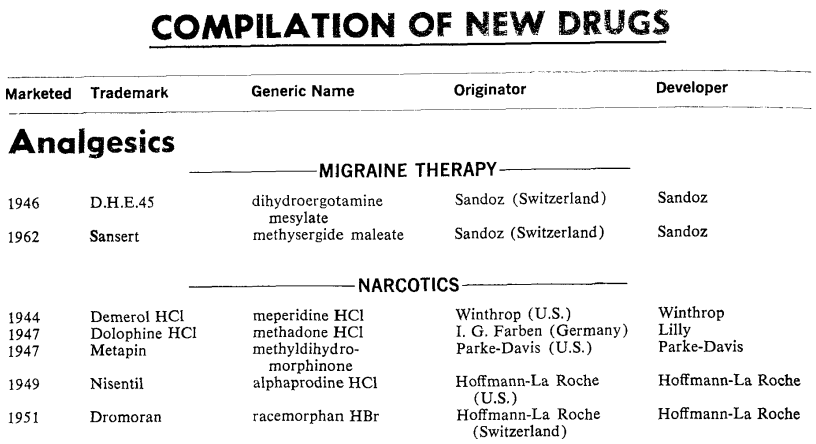
[5] Though the focus of Marx and Fuegi is measuring patent citations to science, early versions of the Marx and Fuegi repository included a complete MAG dataset, including publications which were not cited by patents. We obtaind our MAG data files from these early data deposits. MAG (and its successor OpenAlex, which developed after this project began) is to our knowledge the most complete record of historical science, with roughly 100,000 publications each year in the 1930s. By comparison, NIH's MEDLINE database (the core data behind PubMed) begins in 1966, and historical MEDLINE (a backwards extension of MEDLINE) only goes back to 1946. MAG/OpenAlex is also preferred to Clarivate's Web of Science database, which in our experience appears to have lower historical coverage than MAG/OpenAlex, in addition to being commercially gated and expensive.

[6] The MAG-based basic and clinical subject headings are easy to use, but are opaque in how they are assigned. In

B.3 Pharmaceutical innovation

Our second set of outcomes measure drug development. We digitize de Haen (1976)'s "Compilation of New Drugs, 1940-1975" to produce a list of new drugs introduced over this period (see Figure B.6 for an excerpt), reporting 1,010 drugs developed by 126 distinct firms, including information on the drug class, trademark name, generic name, and year of introduction. We use these data in two ways. First, we manually crosswalk drug classes to 12-digit MeSH codes on the "therapeutic use" and "physiological effects" subbranches of the "pharmacologic action" branch of the MeSH tree (codes D27.505.954 and D27.505.696, respectively), obtain associated terms, aggregate up to the MeSH code-year of term-year level—in analogous format to our analysis of scientific publications. Second, we aggregate up by firm-year and study effects on firms.

Figure B.6: Sample from de Haen (1976) drugs list



COMPILATION OF NEW DRUGS

Marketed	Trademark	Generic Name	Originator	Developer
Analgesics				
		MIGRAINE THERAPY		
1946	D.H.E.45	dihydroergotamine mesylate	Sandoz (Switzerland)	Sandoz
1962	Sansert	methysergide maleate	Sandoz (Switzerland)	Sandoz
		NARCOTICS		
1944	Demerol HCl	meperidine HCl	Winthrop (U.S.)	Winthrop
1947	Dolophine HCl	methadone HCl	I. G. Farben (Germany)	Lilly
1947	Metapin	methyldihydro-morphinone	Parke-Davis (U.S.)	Parke-Davis
1949	Nisentil	alphaprodine HCl	Hoffmann-La Roche (U.S.)	Hoffmann-La Roche
1951	Dromoran	racemorphan HBr	Hoffmann-La Roche (Switzerland)	Hoffmann-La Roche

Notes: Figure provides an except from de Haen (1976)'s "Compilation of New Drugs, 1940-1975," which lists chemical entities and synthesized drugs available in the U.S.

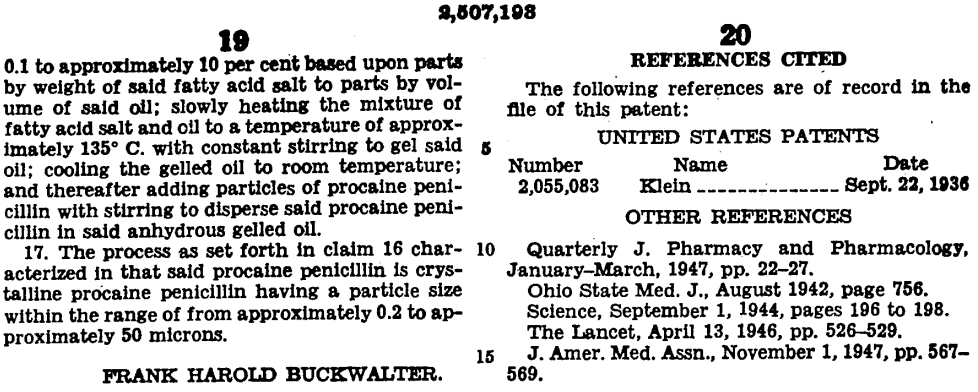
Our firm analysis involves two further steps. Because some firms in our sample merged during our study period, we collect information on M&A from U.S. Federal Trade Commission (1980), which reports acquisitions of manufacturing firms with at least $10 million in assets between 1947 and 1978. We dynamically reassign firms to their contemporary parent in the years after a known merger. Second, we manually crosswalk these firms to assignees in patent data, which we obtain from Google Patents (through Google BigQuery). We then identify drug patents belonging to these firms filed between 1930 and 1970, measuring "drug patents" as patents in NBER patent category 31 (Hall et al. 2001), corresponding to USPC 424 and 514.

comparison to MAG headings, ISI-based classification is better understood, contemporary to the period we study, and evidently was commercially valuable. As Cardoni (1973) writes in a review of CC:CP: "Current Contents/Life Sciences provides tables of contents of over 1,000 medical science journals and includes broad coverage of basic science and medical specialty publications. Current Contents/Clinical Practice is a new service covering 700 journals and emphasizing clinical practice journals while providing minimal coverage of basic science journals." Cardoni goes on to explain that "the majority of journals falling into this category are basic science in nature and are not directly related to clinical topics," which reinforces our preference for these lists.

Measuring drug patents' connections to science

We further measure whether these drug developers' patents reference scientific literature, focusing on "in-text" citations made in the patent disclosure. Bryan et al. (2020) argue that in-text citations may be a better measure of an invention's knowledge inputs than "front-page" citations to non-patent literature (NPL), whose legal purpose is identifying prior art against which claims are evaluated. Though NPL can nevertheless signal proximity to science, an important drawback to NPL citations for this paper is that they only began to be included in patents granted after February 1947 (at that time, at the end of the patent; see example in Figure B.7), which coincides with the end of the war. Evaluating changes around this period in the rate at which patents filed in the 1940s cite NPL thus risks attributing growth in NPL to the war, when it is a mechanical effect of the introduction of a new requirement to include prior art references in patent publications. In-text citations, on the other hand date back to the establishment of the modern Patent Office in 1836 and can therefore be observed both before and after World War II.

Figure B.7: Example list of patent and non-patent citations from the last page of a 1950 patent (patent 2,507,193, for "Penicillin product", filed by Bristol Laboratories)



2,507,193

19

0.1 to approximately 10 per cent based upon parts by weight of said fatty acid salt to parts by volume of said oil; slowly heating the mixture of fatty acid salt and oil to a temperature of approximately 135° C. with constant stirring to gel said oil; cooling the gelled oil to room temperature; and thereafter adding particles of procaine penicillin with stirring to disperse said procaine penicillin in said anhydrous gelled oil.

17. The process as set forth in claim 16 characterized in that said procaine penicillin is crystalline procaine penicillin having a particle size within the range of from approximately 0.2 to approximately 50 microns.

FRANK HAROLD BUCKWALTER.

20

REFERENCES CITED

The following references are of record in the file of this patent:

UNITED STATES PATENTS

Number	Name	Date
2,055,083	Klein	Sept. 22, 1936

OTHER REFERENCES

Quarterly J. Pharmacy and Pharmacology, January–March, 1947, pp. 22–27.
Ohio State Med. J., August 1942, page 756.
Science, September 1, 1944, pages 196 to 198.
The Lancet, April 13, 1946, pp. 526–529.
J. Amer. Med. Assn., November 1, 1947, pp. 567–569.

Notes: Figure provides an example list of non-patent literature cited in a post-1947 patent as prior art. Excerpted last page of US patent 2,507,193 ("Penicillin product").

To identify patents which cite science, we use the 2019 version of the Marx and Fuegi (2020, 2022) patent citations to science (PCS) file, which we found to be more complete for this period than the latest file. We restrict to citations measured with a confidence score of 5 (out of 10) or higher, based on manual validation of a subset of the data, though our results are not sensitive to this choice. We further restrict to referenced articles which are identified by the data providers as cited in-text or both in-text and front-page (in our period, final page).

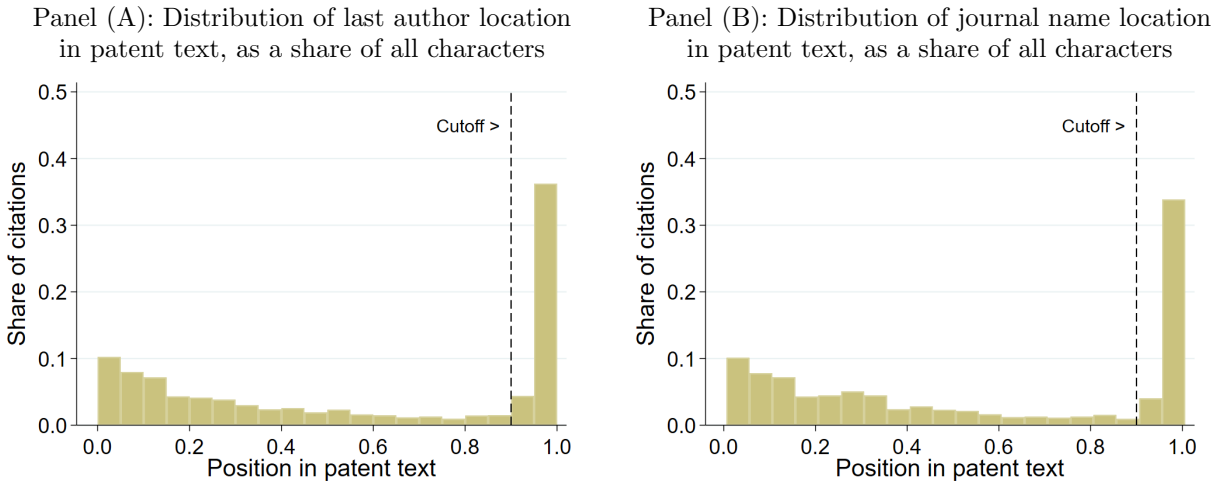
For older patents, the PCS data (in both the 2019 file and the current file) appear to mistakenly count some NPL citations as in-text. Through manual inspection of a random sample of 50 in-text PCS citations over the 1930-1970 period, we found that approximately 40% of the reported in-text citations we reviewed were actually NPL. Mismeasurement is a particular concern prior to USPTO's

transition to structured electronic data in the 1970s, since NPL citations in these years have to be extracted from unstructured, digitized patent publications, where algorithmically distinguishing between in-text and front-page (or final-page) citations is more difficult.

To investigate further, we obtained from Google Patents (via Google BigQuery) the full text of each patent in our sample. For each patent-paper citation, we then merged in publication information (first author, journal, publication year; retrieved from MAG) and patent text, extracted the first author's surname, and looked for the character position of each of these features in the text. When found, we took these positions and divided them by the patent length.

As Figure B.8 shows, the distribution of these positions has a significant mass near the end of a patent, which points to citations which are likely to be NPL rather than in-text. To limit the extent of these false positives (which, again, are only reported after 1947, and thus a particular concern for our study) we set a rule that if any of the publication features (first author surname, journal name, or publication year) is found in the patent text and its first position is in the last 10% of the patent, it should be considered an NPL citation and dropped from our in-text citation sample. Out of the initial set of 4,500 reported in-text PCS in the Marx-Fuegi data made by the patents in our sample, this procedure drops approximately 1,650 (36%) of them.

Figure B.8: Distribution of cited author/journal position in patent text

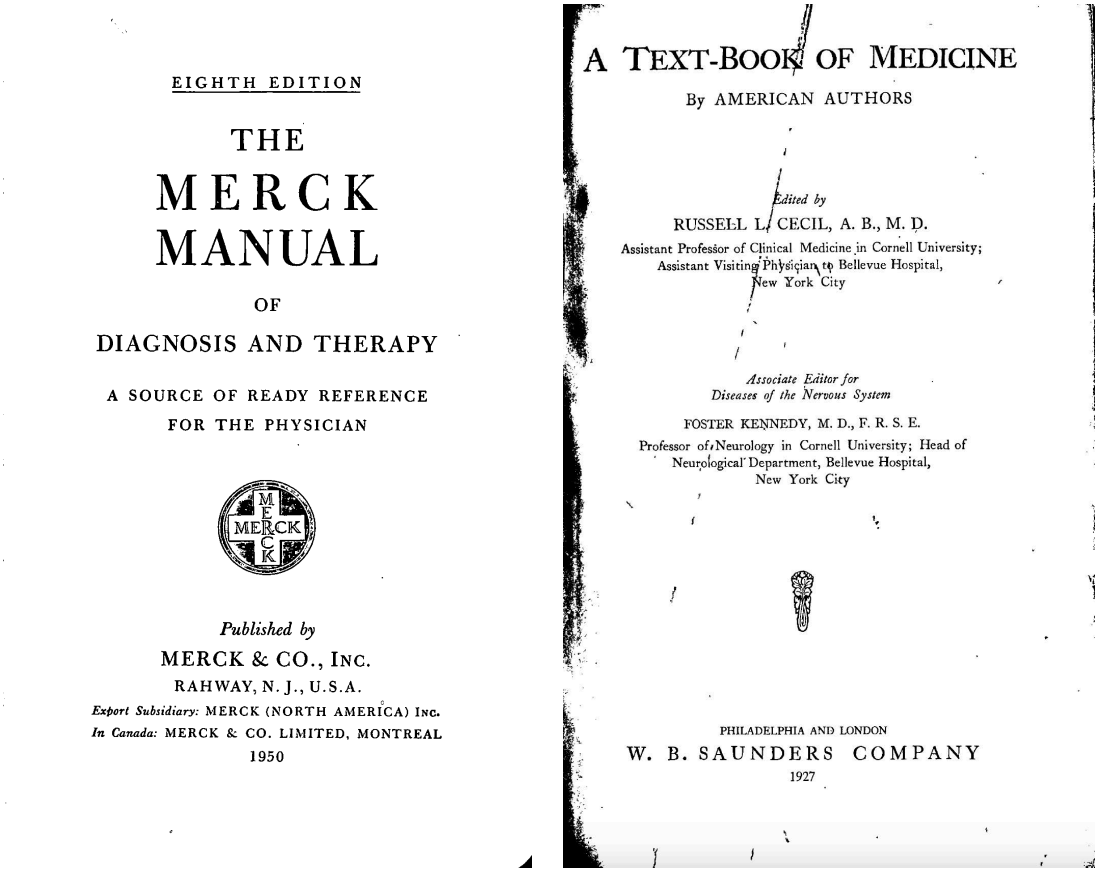


Notes: Figure provides distribution of the character position of cited author/journal names in the text of patents which are measured by Marx and Fuegi (2020, 2022) as science-citing. The mass near one suggests NPL mismeasured as in-text citations. We apply a cutoff of 90% to remove these cases from our in-text citation sample.

B.4 Mid-century medical textbooks

Our third set of outcomes studies knowledge diffusion. We use medical textbooks to measure the diffusion of biomedical research subjects into medical practice, focusing on two textbook series which were published both before and after the war: the *Merck Manual of Diagnosis and Therapy* (MM) and the *Cecil Textbook of Medicine* (CT).[7] Figure B.9 shows the covers of the Merck Manual and Cecil Textbook. We collect four editions of the Merck Manual covering 1940, 1950, 1956, and 1961 (editions 7 to 10) and nine editions of the Cecil Textbook covering 1930, 1933, 1937, 1940, 1943, 1947, 1951, 1955, and 1959 (editions 2 to 10). From each edition, we digitize the index and process it through MTI to identify covered subjects.

Figure B.9: Merck Manual and Cecil Textbook practitioner texts: sample volumes



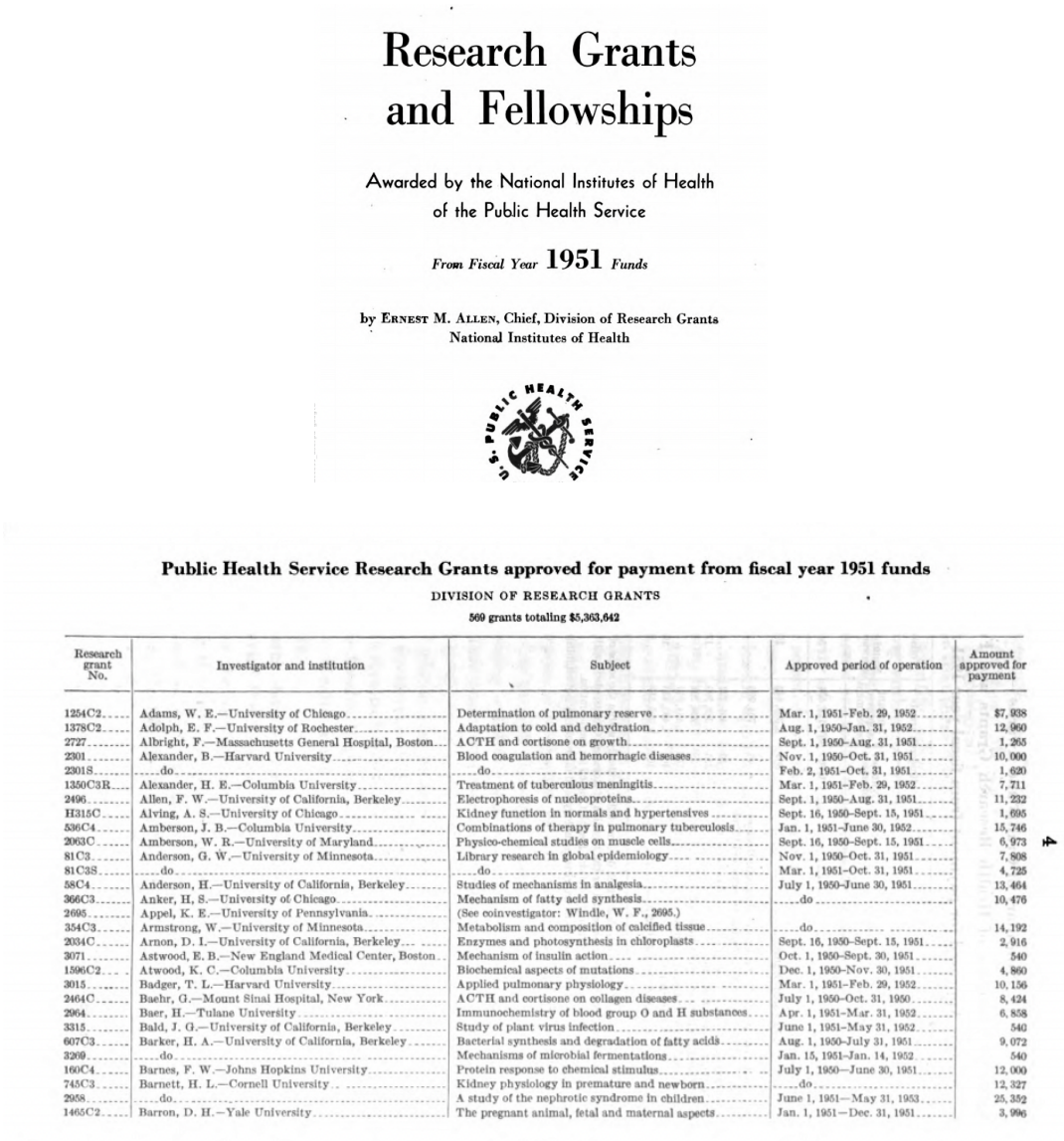
Notes: Figure shows the cover of example editions of the textbooks used in our analysis.

[7] Our interpretation of textbooks as an indicator of medical knowledge is based on textbooks historically being a principal source of guidance for practicing physicians (Catillon 2017).

B.5 NIH extramural research grants

Our final set of outcomes measures postwar extramural NIH grants. We digitize annual editions of the U.S. Public Health Service's *Research Grants and Fellowships Awarded by the National Institutes of Health* from 1948 to 1970 to collect information on all NIH grants over this period, including the grant number, PI, institution, title, amount and funding institute (see Figure B.10 for an example). Similar to other data sources, we process grant titles through MTI to associate them to medical subjects and measure subject-level NIH support.

Figure B.10: NIH research grant listing: sample volume



Research Grants and Fellowships

Awarded by the National Institutes of Health of the Public Health Service

From Fiscal Year **1951** *Funds*

by Ernest M. Allen, Chief, Division of Research Grants
National Institutes of Health

Public Health Service Research Grants approved for payment from fiscal year 1951 funds

DIVISION OF RESEARCH GRANTS

569 grants totaling $5,363,642

Research grant No.	Investigator and institution	Subject	Approved period of operation	Amount approved for payment
1254C2	Adams, W. E.—University of Chicago	Determination of pulmonary reserve	Mar. 1, 1951–Feb. 29, 1952	$7,938
1378C2	Adolph, E. F.—University of Rochester	Adaptation to cold and dehydration	Aug. 1, 1950–Jan. 31, 1952	12,960
2727	Albright, F.—Massachusetts General Hospital, Boston	ACTH and cortisone on growth	Sept. 1, 1950–Aug. 31, 1951	1,265
2301	Alexander, B.—Harvard University	Blood coagulation and hemorrhagic diseases	Nov. 1, 1950–Oct. 31, 1951	10,000
2301S	....do	....do	Feb. 2, 1951–Oct. 31, 1951	1,620
1350C3R	Alexander, H. E.—Columbia University	Treatment of tuberculous meningitis	Mar. 1, 1951–Feb. 29, 1952	7,711
2496	Allen, F. W.—University of California, Berkeley	Electrophoresis of nucleoproteins	Sept. 1, 1950–Aug. 31, 1951	11,232
H315C	Alving, A. S.—University of Chicago	Kidney function in normals and hypertensives	Sept. 16, 1950–Sept. 15, 1951	1,695
536C4	Amberson, J. B.—Columbia University	Combinations of therapy in pulmonary tuberculosis	Jan. 1, 1951–June 30, 1952	15,746
2063C	Amberson, W. R.—University of Maryland	Physico-chemical studies on muscle cells	Sept. 16, 1950–Sept. 15, 1951	6,973
81C3	Anderson, G. W.—University of Minnesota	Library research in global epidemiology	Nov. 1, 1950–Oct. 31, 1951	7,808
81C3S	....do	....do	Mar. 1, 1951–Oct. 31, 1951	4,725
58C4	Anderson, H.—University of California, Berkeley	Studies of mechanisms in analgesia	July 1, 1950–June 30, 1951	13,464
366C3	Anker, H. S.—University of Chicago	Mechanism of fatty acid synthesis	....do	10,476
2695	Appel, K. E.—University of Pennsylvania	(See coinvestigator: Windle, W. F., 2695.)		
354C3	Armstrong, W.—University of Minnesota	Metabolism and composition of calcified tissue	....do	14,192
2034C	Arnon, D. I.—University of California, Berkeley	Enzymes and photosynthesis in chloroplasts	Sept. 16, 1950–Sept. 15, 1951	2,916
3071	Astwood, E. B.—New England Medical Center, Boston	Mechanism of insulin action	Oct. 1, 1950–Sept. 30, 1951	540
1596C2	Atwood, K. C.—Columbia University	Biochemical aspects of mutations	Dec. 1, 1950–Nov. 30, 1951	4,860
3015	Badger, T. L.—Harvard University	Applied pulmonary physiology	Mar. 1, 1951–Feb. 29, 1952	10,156
2464C	Baehr, G.—Mount Sinai Hospital, New York	ACTH and cortisone on collagen diseases	July 1, 1950–Oct. 31, 1950	8,424
2964	Baer, H.—Tulane University	Immunochemistry of blood group O and H substances	Apr. 1, 1951–Mar. 31, 1952	6,858
3315	Bald, J. G.—University of California, Berkeley	Study of plant virus infection	June 1, 1951–May 31, 1952	540
607C3	Barker, H. A.—University of California, Berkeley	Bacterial synthesis and degradation of fatty acids	Aug. 1, 1950–July 31, 1951	9,072
3209	....do	Mechanisms of microbial fermentations	Jan. 15, 1951–Jan. 14, 1952	540
160C4	Barnes, F. W.—Johns Hopkins University	Protein response to chemical stimulus	July 1, 1950—June 30, 1951	12,000
745C3	Barnett, H. L.—Cornell University	Kidney physiology in premature and newborn	....do	12,327
2958	....do	A study of the nephrotic syndrome in children	June 1, 1951—May 31, 1953	25,352
1465C2	Barron, D. H.—Yale University	The pregnant animal, fetal and maternal aspects	Jan. 1, 1951—Dec. 31, 1951	3,996

Notes: Figure shows the contents of an example issue of U.S. Public Health Service's publication *Research Grants and Fellowships Awarded by the National Institutes of Health*, from which we collect information on all NIH grants from 1948 to 1970.

C Supplementary Results

C.1 Evidence from the raw data

Our analysis of CMR effects on science (Section 4) effectively compares the postwar growth of CMR-funded research subjects against others. Even absent regressions, descriptive evidence from raw publication counts in individual subjects is revealing of the CMR effect, including of the utility in drawing distinctions between existing and new subjects.

The raw data suggest that despite the specificity of the wartime problems CMR targeted, research it funded may have been more broadly impactful. One such indication is visible in publication counts alone. Table C.1 lists the top five MeSH terms entering the publication record (based on titles in the complete MAG corpus) each year between 1939 and 1946. For each term, we also report the number of associated publications over the next 10 years and indicate whether the subject was a focus of CMR research. The final row of each panel shows averages for lower-ranked new subjects for comparison. Treatments and therapies that CMR cultivated were immensely more likely to be the most heavily studied new subjects of this period. Despite that CMR cost only a few hundred million dollars (in 2024 dollars) and produced only around 2,500 scientific publications (out of roughly 280,000 published between 1939 and 1946), nearly half of the eventual top research subjects from this era were introduced or supported by CMR.

Figure C.2 shows the time series of publication activity for example MeSH terms in five CMR-funded research areas—antibiotics, steroids, blood, oxygen and lung function, and malaria. Consistent with Stewart (1948)'s observation that "some subjects are born of war," several specific terms in these areas have little pre-war publication activity but take off after the war ends. Others had existing, pre-war research activity but nevertheless grew significantly after World War II (e.g., Oxygen).[8] As we discuss in Appendix D, historical accounts of specific CMR research programs at times point out pre-war knowledge they built on, but even more often remark on the subjects of wartime research having been essentially new and unexplored (Andrus 1948). This was the case not only for a range of specific drugs discovered or developed during the war (like penicillin or new classes of antimalarials), but also for the many new physiological problems that World War II presented (like hypoxia or decompression sickness at 50,000-ft. heights).

The data are also revealing of the importance of using inputs, rather than outputs, as a measure of the CMR shock (as we discuss in Appendix B)—especially when read in conjunction with history. Several CMR programs made failed to achieve breakthroughs during the war itself but nevertheless may have created a platform for postwar innovation. This was most notably the case for synthetic antibiotics, but also applied in other areas. CMR likewise funded efforts to synthesize cortisone which were unsuccessful during the war itself, but which supported a postwar explosion of research on cortisone and corticosteroids, visible in Figure C.2.

[8] An exception is malaria, which had roughly constant publication activity over time; though research in antimalarials temporarily spiked in World War II, that intensity was not sustained in the postwar era.

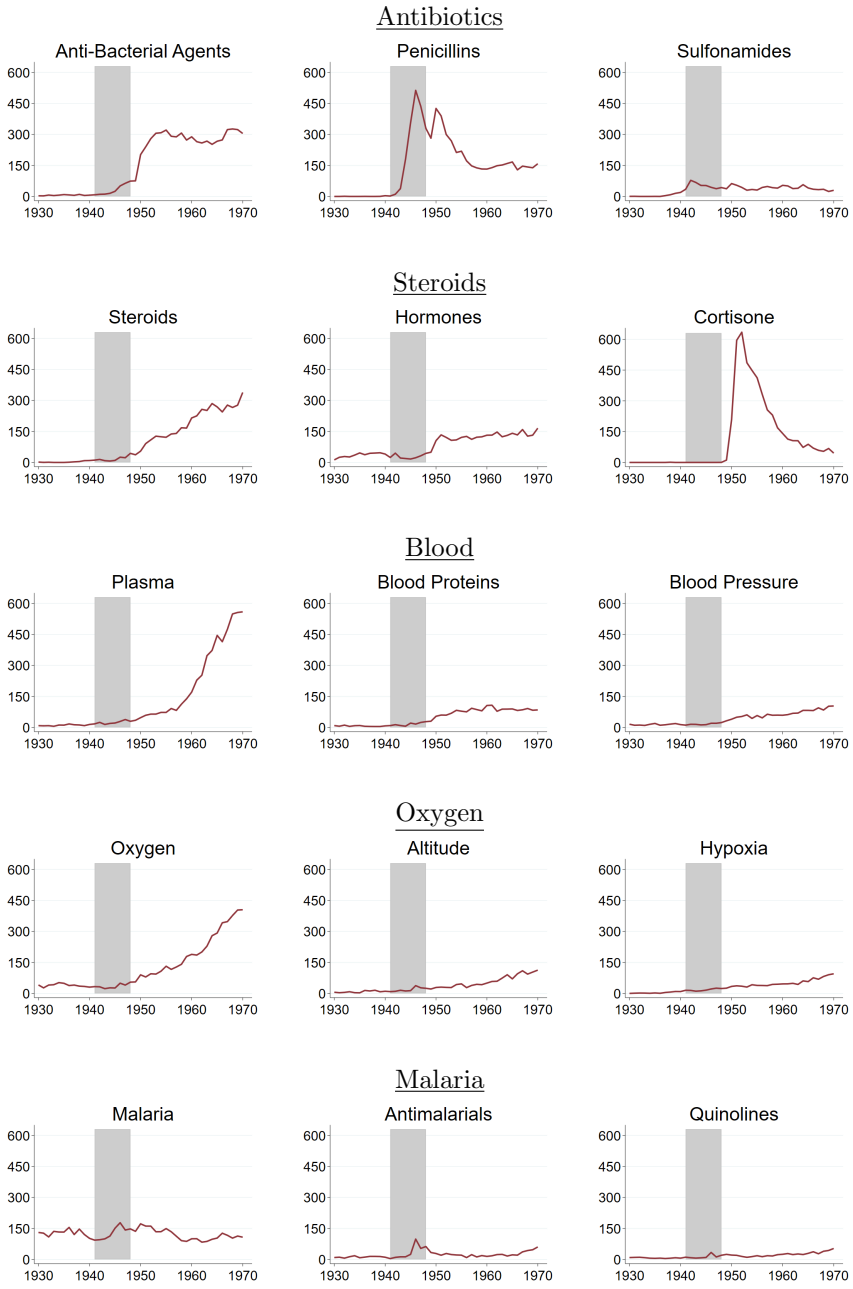
As a point of comparison, in Figure C.2 we plot similar charts of publications over time in subjects which in 1935, the NIH—then intramural, and with very limited funding—identified as research priorities for civilian health. We see little evidence of growth comparable to that in some of the faster-growing subjects in Figure C.1—though the comparison may be strained by the NIH priority list being organized around specific diseases (leprosy) or problem-oriented research fields (bacteriology, nutrition), and the CMR subject list including both diseases and therapies, the latter of which are the subjects with the largest postwar growth.

Table C.1: Top new MeSH terms entering the publication record, by year, 1939-1946

Rank	MeSH term	Pubs (10-yr)	CMR	Rank	MeSH term	Pubs (10-yr)	CMR
	1939				1943		
1	Sulfathiazole	407	1	1	Chromatography, Paper	367	0
2	Diphenhydramine	118	0	2	Penicillin G	244	1
3	Phenytoin	101	1	3	Angiocardiography	155	0
4	17-Ketosteroids	96	1	4	Coxsackievirus Infections	99	0
5	Fenestration, Labyrinth	87	0	5	Penicillin V	60	1
average, lower ranked subjects		4.4	0.03	average, lower ranked subjects		5.7	0.02
	1940				1944		
1	DDT	558	1	1	Streptomycin	4685	1
2	Thiouracil	425	0	2	Penicillin G Procaine	249	0
3	Sulfathiazoles	191	1	3	Parathion	212	0
4	Hyaluronoglucosaminidase	181	1	4	Disulfiram	179	0
5	Sulfaguanidine	94	1	5	Polyethylene	149	0
average, lower ranked subjects		4.9	0.02	average, lower ranked subjects		10.4	0.06
	1941				1945		
1	Histamine Antagonists	301	0	1	Nuclear Weapons	225	0
2	Tyrothricin	107	1	2	Dimercaprol	223	1
3	Radioactive Tracers	51	0	3	Methylthiouracil	200	0
4	Fontan Procedure	50	0	4	Nucleons	169	0
5	Folic Acid Deficiency	47	0	5	Histoplasmin	155	0
average, lower ranked subjects		5.2	0.03	average, lower ranked subjects		9.9	0.04
	1942				1946		
1	Dicumarol	371	1	1	Chloroquine	290	1
2	Newcastle Disease	116	1	2	Hypotension, Controlled	205	0
3	Sulfamerazine	84	1	3	Propylthiouracil	202	0
4	Penicillium chrysogenum	67	1	4	Tripelennamine	131	1
5	Stilbamidines	53	0	5	Promethazine	126	0
average, lower ranked subjects		5.0	0.02	average, lower ranked subjects		9.5	0.03

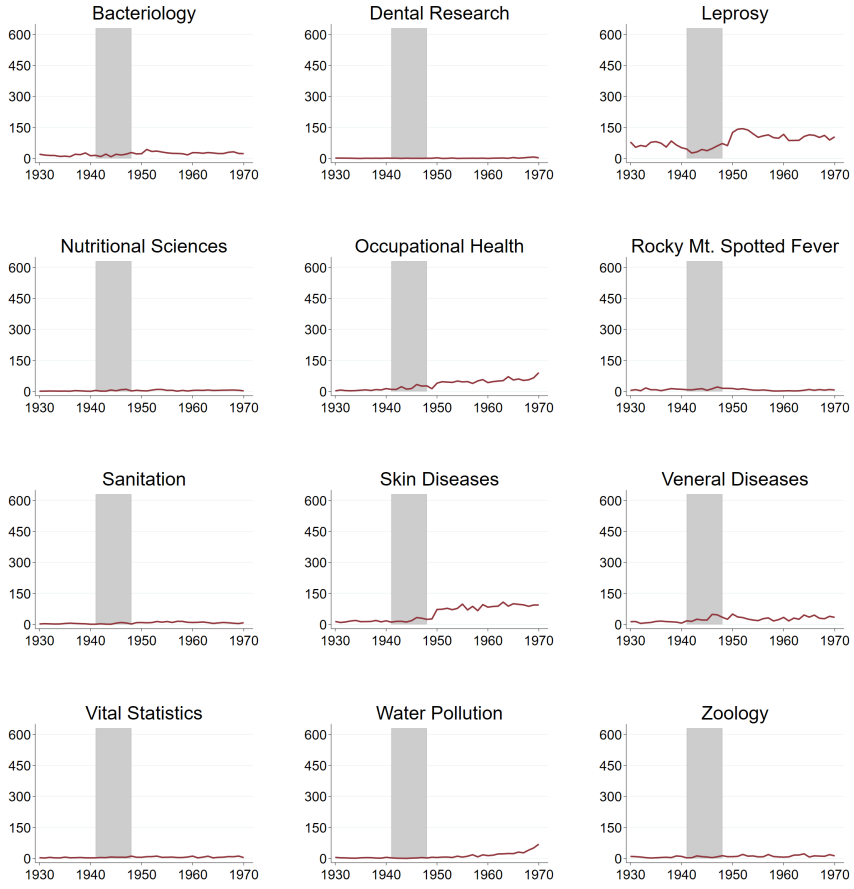
Notes: Table list the annual top 5 new MeSH terms in our publication data from 1939 to 1946, as determined by our indexing of publication titles. For each term we provide the number of associated publications over the next 10 years, and we indicate whether the subject was CMR-funded, measured as having an associated CMR contract or publication. The last row for each year provides averages for all lower-ranking terms. The evidence in the table suggests that many of the most important new biomedical research subjects and therapies emerging at this time were subjects of CMR research.

Figure C.1: Annual publications in example CMR research subjects around World War II



Notes: Figure shows the total publications for select MeSH terms which were major subjects of CMR research. Shaded region in each subfigure marks the war era (1941 to 1948, when most CMR articles were published).

Figure C.2: Annual publications in example research subjects which were NIH priorities in 1935



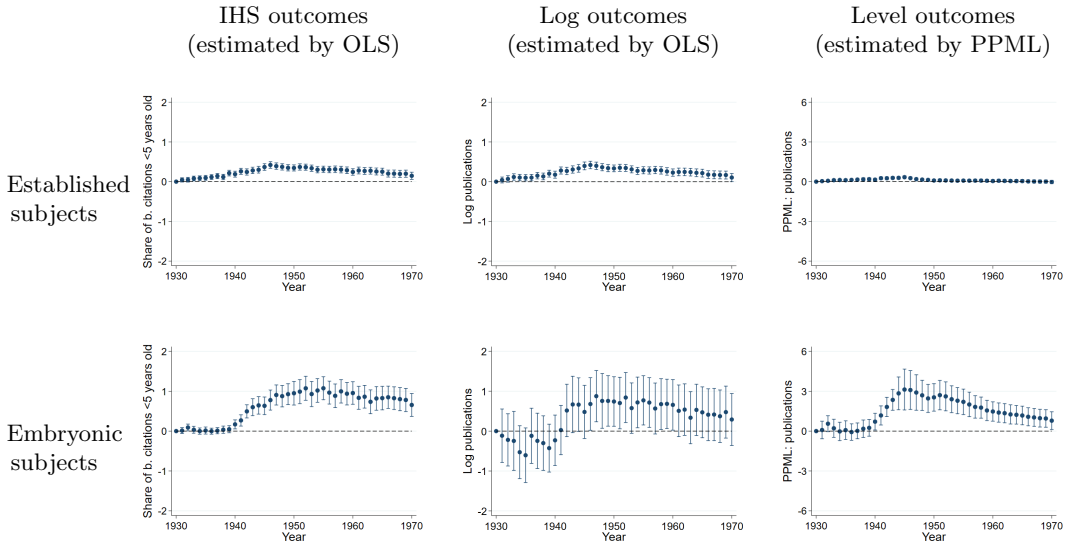
Notes: Figure shows the total publications for select MeSH terms which were NIH priorities in 1935. Shaded region in each subfigure marks the war era (1941 to 1948).

C.2 Alternative specifications: Logged and level outcomes

Figure C.3 presents variants on Figure 3 from the body of the paper using different transformations of the outcome variable (MTI score-weighted publication counts). The first column shows OLS estimates from a specification with inverse hyperbolic sine (IHS) outcomes, as in the body of the paper. The second column presents OLS estimates for log outcomes, dropping MeSH term-years with zero publications. The third column presents PPML estimates for level (i.e., score-weighted count) outcomes. The top and bottom rows of the figure separately report results for "established" and "embryonic" subjects (see paper or figure notes for definitions).

In Table C.2 we present a tabulate version of these results with periodic parameters, essentially re-estimating Column (1) of Table 3 for four different transformations of the outcome: IHS publications, log publications (dropping zeros), number of publications (the MTI score-weighted count), and 1(Any publications). Count outcomes are estimated by PPML, and others by OLS, with fixed effects for MeSH term and year. The left half of Table C.2 does so for established subjects (corresponding to the top row of Figure C.3), and the right half for embryonic subjects (corresponding to the bottom row of Figure C.3); we evaluate changes in the binary 1(Any publications) outcome only for embryonic subjects, as establish subjects have at least one publication in nearly all subject years throughout the sample, and especially after the war. The results are qualitatively consistent across specifications, with more-developed subjects which CMR supported continuing pre-war growth trends before contracting, and less-developed subjects which CMR supported beginning to grow. Columns (5) and (7) provide some insight into how much of this effect is driven by the intensive and extensive margin—where both appear to be operative.

Figure C.3: Effects of CMR on research publications in treated subjects, 1930-1970; comparison of IHS, logged, and level outcomes



Notes: Figure shows annual estimates of the differential growth of scientific publications in MeSH terms with CMR funding, relative to others. The top and bottom row divide MeSH terms into subjects with greater than and less than the median number of pre-1940 publications, which we label "established" and "embryonic" subjects. New combinations are defined as new co-occuring MeSH terms in an article with the given MeSH term. Column (1) shows OLS estimates from a specification with inverse hyperbolic sine outcomes, reproducing the contents of Figure 3 in the paper. Column (2) shows OLS estimates from a specification with logged outcomes. Column (3) shows Poisson pseudo maximum likelihood (PPML) estimates from a specification with level outcomes. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

Table C.2: Effects of CMR on scientific publications, measured four ways: IHS publications, log publications, # of publications (estimated via PPML), and 1(Any publications)

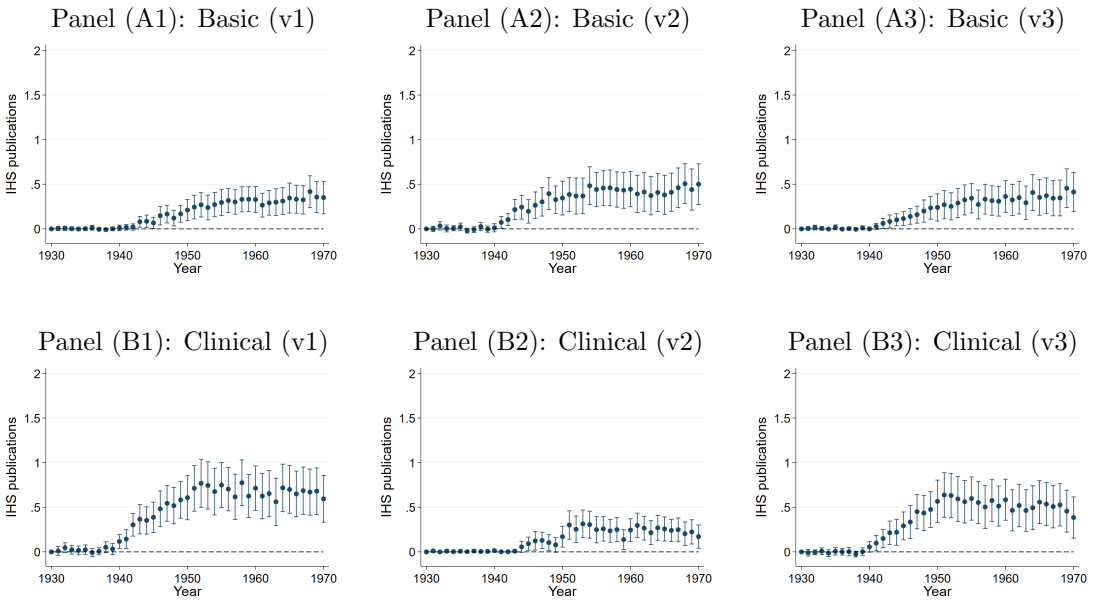
	More-developed subjects			Less-developed subjects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IHS	Log	Count	IHS	Log	Count	Any
Any CMR * 1(1935-1939)	0.086***	0.070***	0.082**	-0.006	-0.101	-0.087	0.019
	(0.018)	(0.021)	(0.032)	(0.019)	(0.125)	(0.130)	(0.026)
Any CMR * 1(1940-1945)	0.223***	0.227***	0.186***	0.441***	0.615***	2.161***	0.284***
	(0.030)	(0.030)	(0.039)	(0.072)	(0.150)	(0.502)	(0.038)
Any CMR * 1(1946-1950)	0.324***	0.304***	0.081**	0.860***	1.005***	2.512***	0.304***
	(0.032)	(0.032)	(0.037)	(0.118)	(0.183)	(0.531)	(0.037)
Any CMR * 1(1951-1955)	0.286***	0.245***	0.015	0.990***	0.967***	2.252***	0.257***
	(0.033)	(0.033)	(0.044)	(0.142)	(0.182)	(0.458)	(0.032)
Any CMR * 1(1956-1960)	0.239***	0.202***	-0.000	0.920***	0.904***	1.527***	0.192***
	(0.035)	(0.035)	(0.048)	(0.135)	(0.179)	(0.319)	(0.032)
Any CMR * 1(1961-1965)	0.215***	0.167***	-0.017	0.788***	0.721***	1.084***	0.123***
	(0.038)	(0.038)	(0.052)	(0.138)	(0.183)	(0.254)	(0.032)
Any CMR * 1(1966-1970)	0.138***	0.091**	-0.067	0.755***	0.637***	0.768***	0.083***
	(0.040)	(0.040)	(0.055)	(0.134)	(0.181)	(0.220)	(0.028)
N	335626	306414	335626	335626	161605	335626	335626
R^2	0.79	0.75		0.55	0.54		0.38
Y mean	2.100	1.507	11.919	0.593	0.222	1.500	0.482
Estimation	OLS	OLS	PPML	OLS	OLS	PPML	OLS

Notes: Table estimates differences over time in scientific activity in MeSH terms with CMR funding, relative to others. Evaluated outcome in all columns is the number of scientific publications, under four transformations: inverse hyperbolic sine (Columns 1 and 4), log (Columns 2 and 5), counts (Columns 3 and 6), and any (0/1; Column 7). Columns (1) to (3) restrict the sample to "established" subjects (with above-median pre-1940 publications), and Columns (4) to (7) to "embryonic" subjects (with below-median pre-1940 publications). We evaluate the binary 1(Any publications) outcome for embryonic subjects only, as established subjects have publications in >90% of all subject years in our sample (>95% after 1950, and >98% in the 1960s), such that there is little postwar variation. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term in parentheses.

C.3 Effects of CMR on basic versus clinical research: Results under different approaches to measuring basic and clinical research

Figure C.4 presents variants on Figure 3 for basic and applied (clinical) research, using different approaches to distinguishing types. "v1" refers to measures based on OECD subfields in the MAG data, which are used as given; "v2", measures based on historical commercial journal indexing lists; "v3", measures based on keywords in publication titles. Our preferred measure is the variant based on contemporary journal lists (v2), due to its transparency, coincidence with our study period, and apparent commercial value. See Appendix B for detailed definitions.

Figure C.4: Effects of CMR on research publications in treated subjects, 1930-1970
Clinical vs. basic research



Notes: Figure shows annual estimates of the differential growth of scientific publications in MeSH terms with CMR funding, relative to others. The sample is restricted to "embryonic" subjects (i.e., those with below-median pre-1940 publications) and the figure divides the sample into basic and clinical biomedical publications (see text for details). Each column uses a distinct approach to identifying basic and clinical research: the left column relies on MAG subject headings that identify journals as basic medical research and clinical medicine; the middle column distinguishes publications in journals indexed by two indexing lists, Current Contents: Life Sciences (CC:LS) and Current Contents: Clinical Practice (CC:CP); the right column applies the term lists in Lewison and Paraje (2004) to identify articles as basic or clinical biomedical research based on the presence of those terms in their titles. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

C.4 Heterogeneity by CMR division, program, and contractor

The following tables examine heterogeneity in CMR's effects on science. We re-estimate Equation (1) in Section 4) in several ways. Table C.3 estimates effects by CMR division (Medicine, Surgery, Aviation Medicine, Physiology, Chemistry, Malaria, and a residual "Miscellaneous" category we also observe in CMR records, which primarily funded R&D in synthetic antibiotics and hormones). Differences across divisions will reflect variation in subject matter and funding approaches, especially in the type of research supported (basic vs. applied—with, e.g., physiology on average more basic, medicine more applied, and the "miscellaneous" category largely reducing to drug development) and the relationship of CMR to the performance of the funded work.

To make these comparisons, we measure indicators for whether a MeSH subject had any associated CMR contracts from each division, and we run a single horserace regression of the effects of being funded by each division. The columns of Table C.3 represent division-specific effects from this horserace (i.e., with all parameters in the table estimated in one combined regression). Following the structure of our analysis in the paper, we also limit the sample to subjects which were less-developed at the dawn of World War II (i.e., those with below-median pre-1940 publications). We find substantial heterogeneity in the effects associated with each division. The largest effects (by far) are generated by contracts in the "Miscellaneous" category, where shocked subjects on average roughly quintupled scientific activity by the 1950s. The Physiology and Surgery divisions also had large and persistent effects. The Medicine and Chemistry divisions, by comparison, had shorter-lived effects on science, and Malaria effectively had none.

Table C.3: Heterogeneous effects of CMR on research publications, by CMR division (single horserace regression, presenting estimates for each division)

	Less-developed subjects						
	(1a)	(1b)	(1c)	(1d)	(1e)	(1f)	(1g)
Parameters for:	Medicine	Surgery	Aviat. Med.	Physiology	Chemistry	Malaria	Miscellaneous
Any CMR * Column division * 1(1935-1939)	0.014	-0.002	0.032	-0.005	-0.083	-0.143**	0.041
	(0.031)	(0.025)	(0.058)	(0.041)	(0.076)	(0.071)	(0.077)
Any CMR * Column division * 1(1940-1945)	0.607***	0.531***	0.222	0.354***	0.272	0.094	1.255
	(0.167)	(0.121)	(0.136)	(0.122)	(0.205)	(0.151)	(0.925)
Any CMR * Column division * 1(1946-1950)	0.829***	0.897***	0.307	0.770***	1.253**	-0.161	3.098***
	(0.216)	(0.174)	(0.204)	(0.221)	(0.564)	(0.125)	(1.043)
Any CMR * Column division * 1(1951-1955)	0.465*	0.898***	0.296	1.104***	1.376**	-0.379**	5.346***
	(0.258)	(0.215)	(0.247)	(0.244)	(0.693)	(0.148)	(0.768)
Any CMR * Column division * 1(1956-1960)	0.512**	0.783***	0.392*	1.107***	0.940	-0.388*	4.455***
	(0.247)	(0.218)	(0.238)	(0.250)	(0.657)	(0.218)	(0.666)
Any CMR * Column division * 1(1961-1965)	0.241	0.604**	0.293	1.221***	0.606	-0.236	4.037***
	(0.273)	(0.235)	(0.262)	(0.264)	(0.587)	(0.258)	(0.243)
Any CMR * Column division * 1(1966-1970)	0.249	0.495**	0.404	1.251***	0.440	-0.355	3.468***
	(0.264)	(0.221)	(0.304)	(0.262)	(0.473)	(0.292)	(0.154)

Notes: Table estimates differences in the (IHS) annual number of scientific publications in MeSH terms with versus without CMR funding, estimating effects by CMR division. All estimates in the table are produced from a single horserace regression with CMR division-specific treatment indicators. A given MeSH term can be treated by multiple divisions. These division-specific coefficients are reported across columns. Estimation sample is restricted to "embryonic" subjects (those with below-median pre-1940 publications). *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term in parentheses.

CMR records also associate contracts to specific topics within its divisions, which by our reading roughly correspond to research programs. In Table C.4 we estimate a similar horserace regression by CMR program, which we organize into five categories: disease studies (infectious disease, tropical disease, malaria), injury studies (burns and wounds, gas casualties), research into therapeutic techniques (neurosurgery, convalescence), research on specific therapies (antibiotics, hormones), and physiology (blood, shock, nutrition and acclimatization). The columns in this table represent program-specific effects estimated from a single regression, with the top panel reporting estimated effects of disease- and injury-driven CMR research, and the bottom panel continuing this reporting with estimated effects of research in techniques, therapies, and physiology. The largest and most persistent effects appear to have been produced by CMR investments in specific therapies and in physiology. The categorical contrast between the two—studies of blood, shock, and nutrition were relatively more fundamental, and R&D in antibiotics and hormones very applied—speaks to not only the complex bundle that CMR as a whole represented (the agency does not fit neatly into the basic-applied dichotomy), but also the possibility that investments in both research and technology can propel science, eluding a traditional linear model of innovation.

Our final tests cut the CMR shock by research performer. Specifically, we examine the effects of a MeSH term being a subject of CMR research performed by investigators in each of five sectors: (i) universities, (ii) firms, (iii) hospitals, (iv) research institutes, or (v) government. Table C.5 presents the results, estimated via a single horserace regression (as in the prior tables). We find large effects for all sectors but research institutes, with particularly large effects for government-performed research, as well as large effects for industry-performed research. Due to subject differences across contractors, we cannot attribute differential effects to the contracting sector per se (i.e., to the quality or public value of industrial or government-performed research), but the differences are nevertheless revealing of the heterogeneity across CMR's portfolio.

Table C.4: Heterogeneous effects of CMR on research publications, by CMR program (single horserace regression, presenting estimates for each program)

	Less-developed subjects					
	Disease				Injury	
	(1a)	(1b)	(1c)	(1d)	(1e)	(1f)
Parameters for:	Infect. Disease	Malaria	Trop. Disease	Ven. Disease	Burns/Wounds	Gas Casualties
Any CMR * Column program * 1(1935-1939)	0.034	-0.152**	0.087	0.027	0.027	-0.157*
	(0.040)	(0.073)	(0.081)	(0.049)	(0.044)	(0.086)
Any CMR * Column program * 1(1940-1945)	1.245***	-0.019	0.194	0.286**	0.309**	0.193
	(0.252)	(0.110)	(0.118)	(0.127)	(0.123)	(0.277)
Any CMR * Column program * 1(1946-1950)	0.958***	-0.233	0.705***	1.128**	0.579**	1.424*
	(0.202)	(0.150)	(0.265)	(0.499)	(0.279)	(0.777)
Any CMR * Column program * 1(1951-1955)	0.254	-0.396**	0.403***	1.128*	0.547	1.822**
	(0.438)	(0.164)	(0.102)	(0.637)	(0.344)	(0.894)
Any CMR * Column program * 1(1956-1960)	0.168	-0.363*	0.278**	0.978	0.520	1.455*
	(0.476)	(0.199)	(0.127)	(0.668)	(0.337)	(0.812)
Any CMR * Column program * 1(1961-1965)	-0.343	-0.228	-0.407**	0.941	0.330	1.046
	(0.501)	(0.247)	(0.200)	(0.643)	(0.376)	(0.732)
Any CMR * Column program * 1(1966-1970)	-0.128	-0.349	-0.215	0.936*	0.312	0.921*
	(0.480)	(0.293)	(0.448)	(0.526)	(0.376)	(0.524)

	Less-developed subjects						
	Techniques		Therapies		Physiology		
	(1g)	(1h)	(1i)	(1j)	(1k)	(1l)	(1m)
Parameters for:	Neurosurgery	Convalescence	Antibiotics	Hormones	Blood	Shock	Nut. & Temp.
Any CMR * Column program * 1(1935-1939)	0.014	-0.023	-0.241**	0.147***	0.041	-0.055	-0.000
	(0.039)	(0.060)	(0.107)	(0.002)	(0.056)	(0.040)	(0.086)
Any CMR * Column program * 1(1940-1945)	0.395***	0.198	1.261***	-0.043***	0.341***	0.401**	0.331
	(0.126)	(0.210)	(0.286)	(0.003)	(0.129)	(0.179)	(0.216)
Any CMR * Column program * 1(1946-1950)	0.138	0.301	2.776***	1.636***	0.585**	0.820**	0.658
	(0.119)	(0.332)	(0.559)	(0.005)	(0.249)	(0.346)	(0.407)
Any CMR * Column program * 1(1951-1955)	0.099	0.298	3.214***	6.411***	1.042***	0.860*	0.799**
	(0.235)	(0.367)	(0.661)	(0.008)	(0.388)	(0.441)	(0.373)
Any CMR * Column program * 1(1956-1960)	-0.029	0.556*	2.916***	5.375***	1.011**	0.902*	0.779**
	(0.298)	(0.327)	(0.744)	(0.009)	(0.411)	(0.489)	(0.369)
Any CMR * Column program * 1(1961-1965)	0.187	0.333	3.913***	4.297***	1.244***	1.101**	0.654*
	(0.335)	(0.376)	(0.770)	(0.011)	(0.392)	(0.482)	(0.382)
Any CMR * Column program * 1(1966-1970)	-0.109	0.168	3.419***	3.464***	1.248***	1.004**	0.807**
	(0.297)	(0.421)	(0.795)	(0.012)	(0.385)	(0.487)	(0.401)

Notes: Table estimates differences in the (IHS) annual number of scientific publications in MeSH terms with versus without CMR funding, estimating effects by CMR program. All estimates in the table are produced from a single horserace regression with CMR program-specific treatment indicators. A given MeSH term can be treated by multiple programs. These program-specific coefficients are reported across columns. Estimation sample is restricted to "embryonic" subjects (those with below-median pre-1940 publications). *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term in parentheses.

Table C.5: Heterogeneous effects of CMR on research publications, by contractor sector (horserace regression, presenting estimates for each sector)

	Less-developed subjects				
	(1a)	(1b)	(1c)	(1d)	(1e)
Parameters for:	University	Firm	Hospital	Institute	Government
Any CMR * Column sector * 1(1935-1939)	-0.005	-0.097	0.050	-0.061	-0.098**
	(0.021)	(0.111)	(0.043)	(0.076)	(0.046)
Any CMR * Column sector * 1(1940-1945)	0.355***	0.175	0.652***	0.554*	2.684***
	(0.062)	(0.165)	(0.233)	(0.334)	(0.249)
Any CMR * Column sector * 1(1946-1950)	0.763***	0.630	0.709***	0.492	4.814***
	(0.113)	(0.397)	(0.248)	(0.396)	(0.267)
Any CMR * Column sector * 1(1951-1955)	0.878***	1.026**	0.679**	0.501	4.087***
	(0.150)	(0.401)	(0.267)	(0.598)	(0.277)
Any CMR * Column sector * 1(1956-1960)	0.792***	1.048*	0.862***	0.484	3.178***
	(0.143)	(0.597)	(0.269)	(0.619)	(0.272)
Any CMR * Column sector * 1(1961-1965)	0.630***	1.422***	0.830**	0.492	3.161***
	(0.146)	(0.433)	(0.332)	(0.565)	(0.340)
Any CMR * Column sector * 1(1966-1970)	0.611***	1.433***	0.666*	0.590	2.939***
	(0.140)	(0.343)	(0.378)	(0.596)	(0.385)

Notes: Table estimates differences in the (IHS) annual number of scientific publications in MeSH terms with versus without CMR funding, estimating effects by the sector(s) of associated contractor(s). All estimates in the table are produced from a single horserace regression with sector-specific treatment indicators. A given MeSH term can be treated by multiple sectors. These sector-specific coefficients are reported across columns. Estimation sample is restricted to "embryonic" subjects (those with below-median pre-1940 publications). *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by MeSH term in parentheses.

C.5 Intensive margin: Value-weighted shock

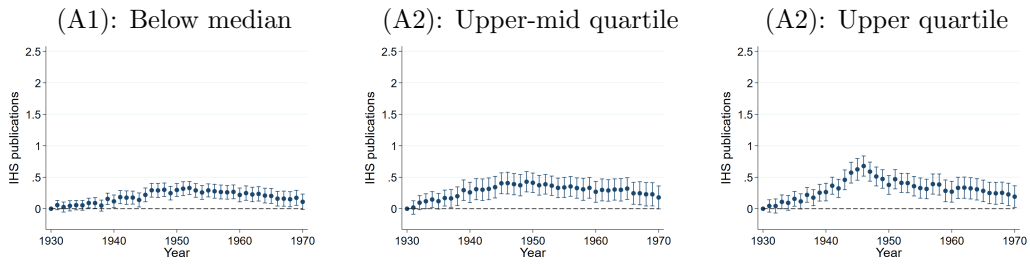
Figure C.5 below re-evaluates Figure 3 in the paper, estimating the effects of value-weighted CMR shock (the intensive margin) rather than an indicator for whether a subject was the target of any CMR contracts (the extensive margin). We group subjects into three treatment quantiles, based on (MTI score-weighted) total funding: below-median, upper-middle quartile, and upper quartile. We then estimate the following regression, where q indexes treatment quantiles, and the omitted category consists of subjects with no CMR contracts:

$$Y_{mt} = \sum_{q=1}^{3} \sum_{t=1931}^{1970} \beta_{qt} \cdot \mathbb{1}(\text{Term } m \text{ in treatment quantile } q) + \alpha_m + \delta_t + \varepsilon_{mt}$$

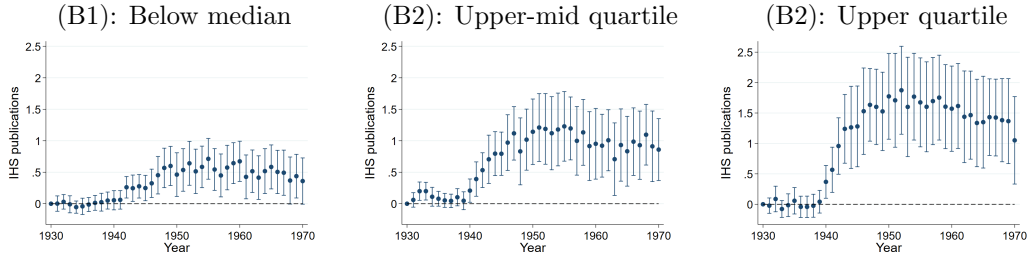
This specification essentially estimates heterogeneous effects in Equation (1) for three groups of subjects: those with the lowest, higher, and highest CMR funding level. Figure C.5, finds monotonically increasing effects across treatment quantiles. Our preferred treatment measure is nevertheless an extensive one, for reasons we describe in the paper (Section 4).

Figure C.5: Effects of CMR on research publications in treated subjects, 1930-1970, by quantile of CMR funding (0-50th, 50-75th, 75-100th percentiles)

Panel (A): Established subjects (compare to Figure 3(A1))



Panel (B): Embryonic subjects (compare to Figure 3(A2))



Notes: Figure shows annual estimates of the differential growth of scientific publications in MeSH terms with different levels of CMR funding, relative to others. Rows (A) and (B) divide MeSH terms into subjects with greater than and less than the median number of pre-1940 publications, which we label "established" and "embryonic" subjects. New combinations are defined as new co-occuring MeSH terms in an article with the given MeSH term. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

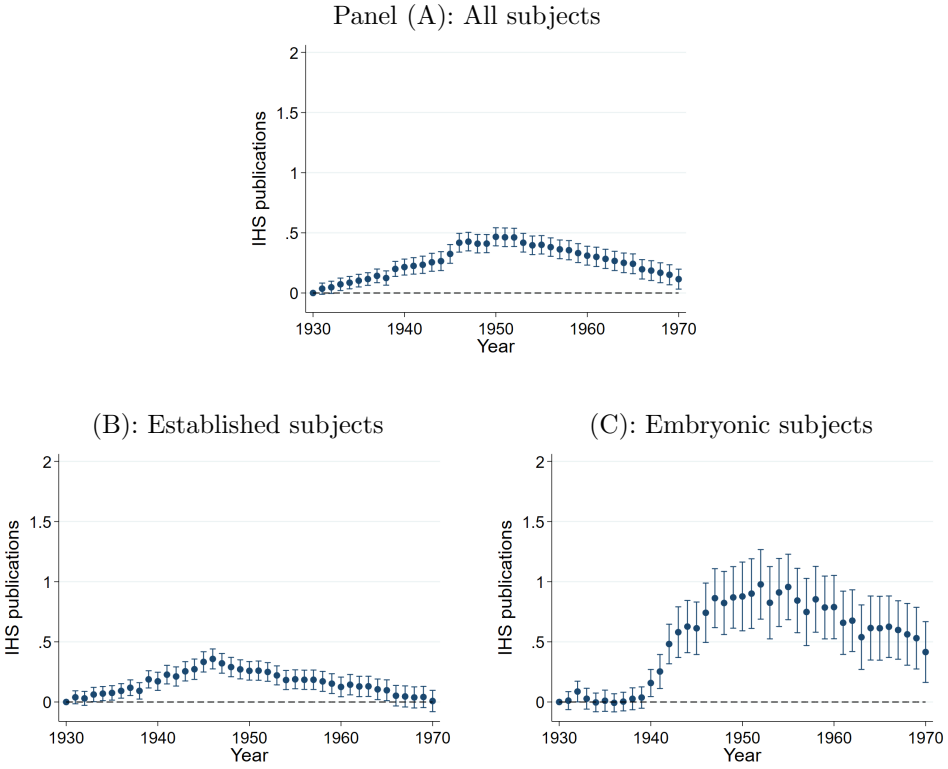
C.6 Controlling for postwar NIH funding

Figure C.6 below presents robustness checks on (i.e., re-estimations of) Figure 3 in the paper, controlling for indicators of whether a given MeSH term was funded by NIH interacted with observation year. Concretely, we estimate the following specification:

$$Y_{mt} = \sum_{t=1931}^{1970} [\beta_t \cdot \mathbb{1}(\text{Any CMR contracts in MeSH term } m) + \gamma_t \cdot \mathbb{1}(\text{Any postwar NIH grants in MeSH term } m)] + \alpha_m + \delta_t + \varepsilon_{mt}$$

The estimated CMR effect are quantitatively and statistically similar, suggesting the CMR effect is neither explained by nor confounded by postwar research funding.

Figure C.6: Effects of CMR on research publications in treated subjects, 1930-1970, controlling for postwar NIH funding (see text for discussion)



Notes: Figure shows annual estimates of the differential growth of scientific publications in MeSH terms with CMR funding, relative to others. Panel (A) does so for all MeSH terms. Panels (B) and (C) divide the sample into subjects with greater than and less than the median number of pre-1940 publications, which we label "established" and "embryonic" subjects. All results are based on specifications which control for an indicator of whether the given MeSH term received postwar NIH funding (through 1970), interacted with years. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

C.7 Placebo test: World War I

Though World War II produced the first large-scale government funding of medical research, it potentially represents a bundle of two shocks: war (a demand shock for new science and technology) and medical R&D funding (a supply shock). Given their coincidence, the results in Section 4 may be difficult to attribute to CMR specifically, particularly if war has a demand-pull effect that brings research attention (and research activity) to subjects that were previously understudied, and that in turn triggers accumulative endogenous growth in science.

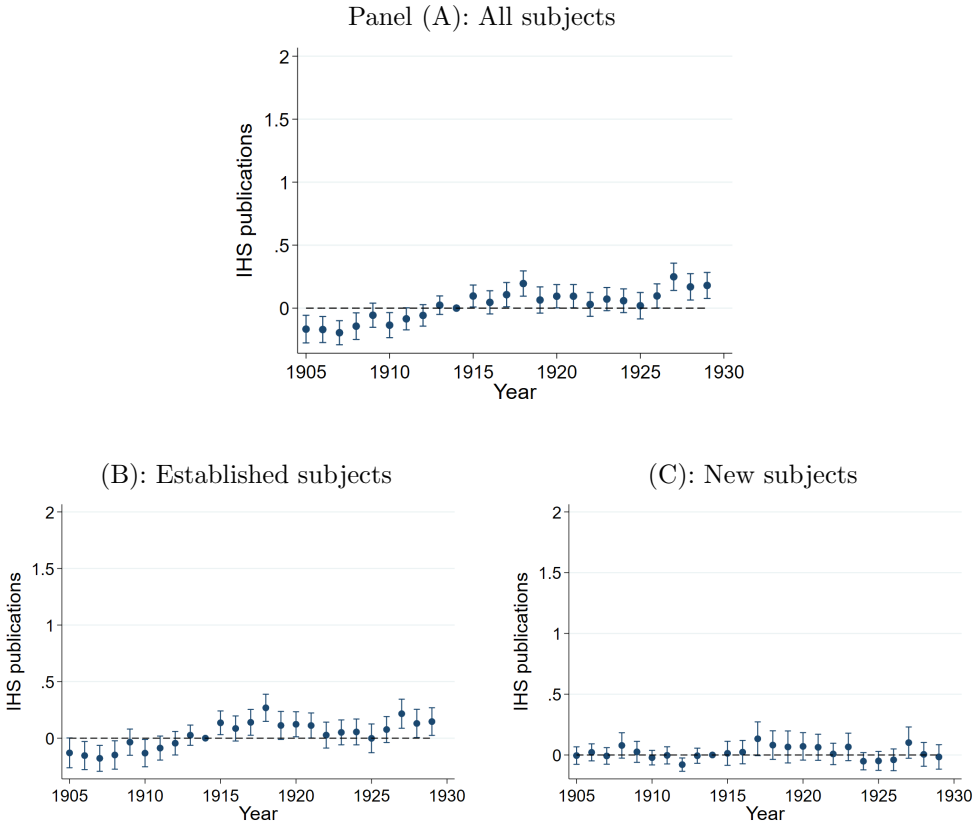
To separate CMR's effects from any general effects of war, we take our research strategy to World War I. We first identify medical subjects implicated in World War I by digitizing the index to American Red Cross (1918) titled *War Medicine*, a publication of the American Red Cross Society in France for medical officers of the U.S. Army presenting war-induced medical conditions, and processing it with MTI (see Appendix B) to link it to MeSH space. We then extend our MAG publication sample backwards to 1905 and run analogous regressions to those in Section 4 of the paper on a sample of MeSH terms between 1905 and 1929 (immediately preceding our World War II sample). Formally, we estimate the following specification:

$$Y_{mt} = \sum_{t=1906}^{1929} \beta_t \cdot \mathbb{1}(\text{MeSH term } m \text{ relevant to WWI}) + \alpha_m + \delta_t + \varepsilon_{mt} \tag{3}$$

where m and t index MeSH terms and years, and standard errors clustered by MeSH term. Following Section 4, we estimate this regression for two outcomes—publications and new combinations—and we do so separately for all terms, existing terms, and new terms, where terms are defined as existing or new based on whether they have above or below median pre-1914 publications (respectively). The resulting estimates are plotted in Figure C.7, and can be compared to the analogous World War II-era effects of CMR shown in Figure 3 in the paper.

In contrast to the effects of CMR estimated in Section 4, we find little evidence of a generic effect of war on science. Panel (A) suggests there may have been some modest pre-war growth in publications in existing war-related subjects, but this growth plateaus after World War I and appears (to our eyes) to be incidental. We find no evidence of growth in new subjects, with relatively tight standard errors small enough to rule out even moderate effect sizes.

Figure C.7: Pseudo-effects of "war shock" in World War I, 1905-1929



Notes: Figure shows annual estimates of the differential growth of scientific publications around World War I (sample period: 1905-1929, with the war taking place in 1914-1918) in MeSH subjects implicated in World War I (according to American Red Cross publications), relative to others. Panel (A) does so for all MeSH terms. Panels (B) and (C) divide the sample into subjects with greater than and less than the median number of pre-1940 publications, which we label "established" and "new" subjects. Omitted (reference) year in all figures is 1914. Error bars represent 95% confidence intervals, with SEs clustered at the MeSH term level.

C.8 Growth of the pharmaceutical industry

This appendix provides additional context, descriptive data, and results with respect to the postwar pharmaceutical industry and its relation to World War II. Section C.8.1 examines the growth of the industry, and Section C.8.2 changes in the nature of drug discovery.

C.8.1 The "golden age" of drug discovery

As we explain in the paper, several important changes to pharmaceutical science and drug innovation are understood to have taken place in the mid-20th century, such as the rise of synthetic chemistry and molecular biology, the growth of rational drug design and systematic drug screens, or advances in clinical testing. Many of these changes ostensibly have links to the war, including in efforts to synthesize new treatments for specific diseases or in large-scale testing, such as the first application of high-throughput screening in the search for new antimalarials.

The degree to which these changes were in fact triggered by the war is potentially an empirically-detectable question, and data on 20th century drug innovation (from de Haen 1976) may provide a window into the answer. Some initial indications are shown in Table C.6, which lists CMR- and WPB-contracted pharmaceutical firms in the De Haen data, along with their total number of drugs introduced in 5-year intervals from 1940 to 1970. By sheer counts, the late 1940s and 1950s was a revolutionary period for drug discovery, especially among CMR/WPB firms, whose average annual new drug introductions more than tripled to over 10 per firm every five years. Other firms increased their average 5-year drug introductions from 0.7 to 1.5 per firm.

Table C.6: Postwar new drugs developed by CMR- and WPB-contracted firms

	New drugs introduced in:					
Firm	1940-45	1946-50	1951-55	1956-60	1961-65	1966-70
Abbott	5	10	14	11	3	2
Bristol	0	0	2	6	4	1
Hoffmann-La Roche	7	7	15	10	7	4
Lilly	4	9	17	11	11	2
Merck, Sharp & Dohme	8	11	12	19	10	4
Parke-Davis	6	8	4	10	2	3
Pfizer	0	2	7	11	1	5
Squibb	0	6	9	7	2	3
Upjohn	1	6	10	12	4	4
Warner	0	2	5	5	5	1
Winthrop	3	3	7	6	2	3
Wyeth	2	5	7	18	6	6
CMR/WPB, Average	3.0	5.8	9.1	10.5	4.8	3.2
Others, Average	0.7	0.9	1.0	1.5	0.6	0.4

Notes: Table reports the number of new drugs developed and brought to market by CMR and WPB pharmaceutical contractors between 1940 and 1970, according to de Haen (1976), in five-year intervals. A small number of firms which merged during this period are combined throughout the sample (e.g., Merck and Sharpe & Dohme merged in 1953, and are reported as a single unit). The bottom row of the table compares these counts to the other-firm average.

Table 4 of the paper evaluates this growth econometrically in Panel (B), and formalizes this result: new drug development grew much more quickly in the first two postwar decades at CMR/WPB firms than others. By 1960, these firms were 20% more likely to introduce a new drug each year than non-CMR/WPB firms (Column 1), with similar effects for antibiotics and hormones (Column 2) and other drugs (Column 3). In percentage terms, the effect was to increase annual drug innovation by roughly 50% at these firms, relative to peers (Columns 4 to 6).

C.8.2 Growth of science-based drug discovery

Beyond its effects on postwar drug innovation, we can also explore whether CMR changed how firms approached drug discovery—i.e., the drug development process itself. To do so we shift our analytical unit from drugs to drug-related patents, focusing on patents filed between 1930 and 1970 by the firms in de Haen (1976). For the purposes of this analysis, we define "drug patents" as patents in the NBER patent category for drugs (Hall et al. 2001). This choice excludes other categories where pharmaceutical firms may have filed patents (such as organic or inorganic chemistry), in order to focus the sample on potential therapeutically-relevant inventions and exclude others—particularly because many drug-developing firms in this era (including the De Haen sample) were industrial chemical companies or had a parallel chemical business.

Through patent data we can get more granular insight into any potential changes in the nature pharmaceutical R&D. Our focus is on its relationship to science, for which we use existing data on patent citations to science (Marx and Fuegi 2020, 2022). For each patent in our sample, we measure whether it makes any in-text citations to scientific literature, limiting to in-text citations (rather than front-page) because (i) these are more likely to reflect intellectual inputs to the invention and (ii) front-page citations were only written into patents beginning in 1946. Patents which cite science we will call "science-citing" or "science-based".

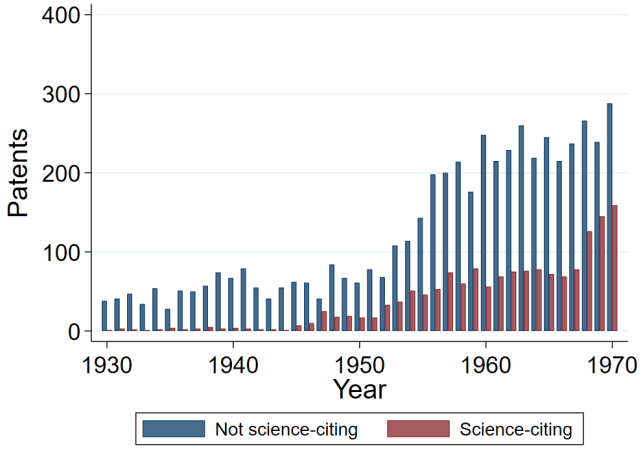
Figure C.8 plots annual counts of non-science and science-based patents in our sample. The figure provides two insights: first, prior to the late 1940s, total annual drug patenting by these firms was more or less constant; second, essentially none of these patents were based in science. Both patterns change after World War II: starting in the late 1940s, drug patenting begins to grow, with a rapid increase in science-based invention—whose share of patents grows substantially. Figure C.9 disaggregates this pattern further, showing average annual non-science and science-based patenting by non-CMR firms and CMR firms, respectively (presented in five-year intervals), where we see that these changes were mainly driven by CMR firms.

In Table C.7, we evaluate these differences econometrically. Specifically, for the De Haen (1976) firm sample, we estimate differences between CMR/WPB-contracted firms and other pharmaceutical firms in their (i) science-based patenting and (ii) non-science based patenting with parameters in 5-year intervals. We estimate a two-way fixed effects specification as follows:

$$Y_{it} = \sum_{t} \beta_t \cdot \mathbb{1}(\text{CMR/WPB firm})_i + \alpha_i + \delta_t + \varepsilon_{it}$$

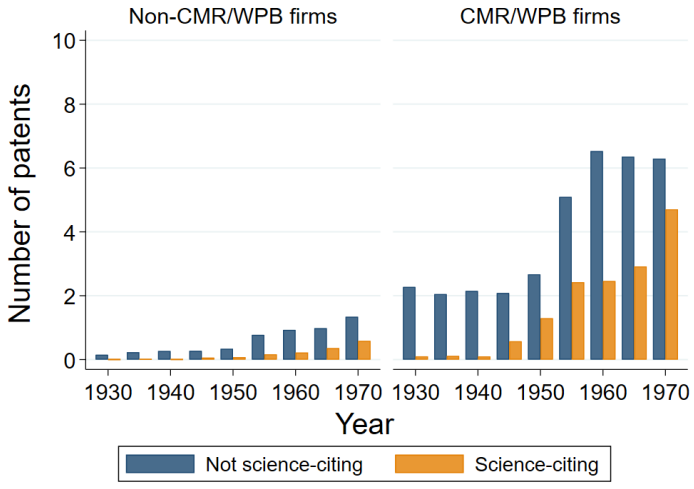
where i and t index firms and years, and the sample runs from 1930 to 1970, with standard errors clustered by firm. For precision we limit the panel to firm-years between each firm's first and last observed drug or patent, to ensure it only includes firm-years when the firm was known to be alive, though this restriction has little impact on the results.

Figure C.8: Total annual drug patents, science-citing and non-citing, 1930-1970



Notes: Figure shows the number of drug patents filed by firms in de Haen (1976), separately reporting science-citing and non-citing patents. Drug patents defined as those associated with NBER category 31 (Hall et al. 2001), and science-citing patents as those which cite non-patent literature (typically, scientific literature) in their text (Marx and Fuegi 2020, 2022).

Figure C.9: Average annual drug patents by CMR/WPB-contracted firms vs. others



Notes: Figure shows the average annual number of drug patents filed by firms in de Haen (1976), separately reporting CMR/WPB-contracted firms and other firms, and science-citing and non-citing patents. Drug patents defined as those associated with NBER category 31 (Hall et al. 2001), and science-citing patents as those which cite non-patent literature (typically, scientific literature) in their text (Marx and Fuegi 2020, 2022).

We estimate this specification for three outcomes: (i) an indicator for whether the firm filed any science and non-science based patents (Columns 1 and 2, respectively), (ii) the number of such patents (Columns 3 and 4), and (iii) the inverse hyperbolic sine (IHS) number of such patents (Columns 5 and 6). In all columns, time-varying parameters are estimated relative to the omitted 5-year interval of 1930-1934. The table shows several patterns. First, CMR/WPB firms rapidly increased their propensity to file science-based drug patents after World War II, with no discernable pre-trends pre-1940. Second, we find no such difference for non-science based patenting. On the intensive margin, we find growth in science-based patenting that roughly equals or exceeds growth in non-science based patenting. By the 1960s, CMR/WPB firms were roughly 50% more likely to file at least one drug-based patent in a given year (Columns 1 vs. 2) and increased their drug-based patenting nearly 50% more than other firms (Columns 5 vs. 6). The combination of results in Columns (1) and (2) in a triple-difference forms the basis for Figure 6(B) of the paper, which effectively shows the Column (1)-(2) difference in graphical form.

Table C.7: Science-citing patents by CMR- and WPB-contracted firms

	Any patents?		Num. patents		IHS(Patents)	
	(1)	(2)	(3)	(4)	(5)	(6)
	PCS	No PCS	PCS	No PCS	PCS	No PCS
CMR/WPB firm * 1(1935-1939)	0.010	-0.044	0.009	-0.067	0.010	-0.060
	(0.066)	(0.089)	(0.099)	(0.455)	(0.067)	(0.164)
CMR/WPB firm * 1(1940-1945)	0.052	-0.024	0.061	0.040	0.049	-0.001
	(0.053)	(0.097)	(0.099)	(0.779)	(0.056)	(0.210)
CMR/WPB firm * 1(1946-1950)	0.247***	-0.059	0.490*	-0.238	0.322**	-0.108
	(0.080)	(0.111)	(0.250)	(0.866)	(0.133)	(0.270)
CMR/WPB firm * 1(1951-1955)	0.528***	0.010	1.390***	0.772	0.826***	0.209
	(0.080)	(0.111)	(0.245)	(1.044)	(0.126)	(0.311)
CMR/WPB firm * 1(1956-1960)	0.491***	0.015	2.165***	2.650*	0.976***	0.510
	(0.093)	(0.097)	(0.615)	(1.581)	(0.217)	(0.316)
CMR/WPB firm * 1(1961-1965)	0.554***	0.043	2.277***	3.906**	1.043***	0.749**
	(0.071)	(0.100)	(0.619)	(1.512)	(0.201)	(0.310)
CMR/WPB firm * 1(1966-1970)	0.457***	0.015	2.821***	2.994**	1.099***	0.575*
	(0.095)	(0.104)	(0.710)	(1.479)	(0.236)	(0.313)
N	5172	5172	5172	5172	5172	5172
R^2	0.39	0.46	0.45	0.51	0.49	0.59
Y mean	0.133	0.275	0.307	0.987	0.180	0.445

Notes: Table estimates differences in pharmaceutical firms' annual science-based and other drug patenting, comparing CMR/WPB-contracted pharmaceutical firms to others. Sample consists of firms which introduced at least one new drug between 1940 and 1970, according to de Haen (1976). A small number of firms which merged or were acquired during the sample frame are dynamically reassigned to their subsequent owners using data from FTC (1980) (see text for details, including specific exceptions). The unit of analysis is a firm-year, and outcomes measure: (i) whether a firm filed any patents in drug classes (NBER category 31; Hall et al. 2001) in the given year (Columns 1 and 2), (ii) the number of such patents (Columns 3 and 4), and (iii) inverse hyperbolic sine (IHS) patents (Columns 5 and 6). Table separately estimates changes in patents that do vs. do not cite non-patent literature (typically, scientific literature) in their text (Marx and Fuegi 2020, 2022). Results are estimated relative to the 1930-1934 period. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by firm in parentheses.

D Contemporary Accounts

The results of Sections 4 and 5 leave us with the deeper question of why CMR—an intrinsically temporary project, that started and ended with the war—had such large and long-lasting effects. Why was this limited research investment apparently so impactful? Though the mechanics of its research effort, and the specific links between wartime and postwar research and innovation, can be difficult to measure empirically, contemporary and historical accounts provide a range of clues as to what made CMR's approach distinctive and impactful.

Why were CMR's effects long-lived?

New therapies and therapeutic candidates

Table C.1 illustrates several subjects "born of war" (Stewart 1948): mass-produced penicillin, streptomycin, DDT, chloroquine, and dimercaprol each effectively launched research in new directions to understand their chemistry, physiological effects, mechanisms of action, applications to myriad organisms and diseases, and more. Geer et al. (1948, p. 636), for example, explain that the development of DDT as an insecticide "opened a new era in insect control," prompting a surge of new investigation. Similar observations could be made for penicillin and later other antibiotics, which drove an (even larger) surge in clinical studies on their effects. Much of this research had a combinatoric flavor: for example, Lockwood (1948, p. 92) observes "the proper combination of surgery and chemotherapy would permit the surgeon far greater latitude"—reflecting new opportunities for studying the interaction of old procedures and new therapies.

Even when CMR research was not successful during the war itself, but it may have opened up new lines of inquiry which postwar investigators continued: Carden Jr (1948, p. 670) anticipated that although "many of the most promising [antimalarial] compounds were developed too late in the program" to be useful during the war, these "unfollowed leads and current loose ends will undoubtedly be explored." Slater (2009) confirms that research on these "lead compounds" continued—one of which (chloroquine) became a revolutionary malaria treatment in the years after the war, and a major new subject of study (as Table C.1 clearly illustrates).

New research tools and techniques

CMR also developed new research methods to meet the war's specific demands, many of which were useful in new problems after it ended. For example, CMR's massive drug screening efforts—including in the hunt for antimalarials, treatments for other tropical diseases, and insect repellents, which collectively screened nearly 20,000 compounds—later became a model for cancer chemotherapy research and other efforts at the NIH (Slater 2009). Techniques developed by the blood substitute research program "provided a technical framework for [a] productive research field" after the war (Creager 1999) on blood-related disorders and blood-derived therapeutics. Cohn (1948, p. 436) writes that the separation of blood "[made] possible the [further] discovery of the function and the uses in therapy" of different components in blood fractions.

Though malaria and blood were among CMR's largest programs, CMR developed new research tools and methods across its portfolio to address questions the war presented. For example, Windle (1948, p. 174) explains that "The introduction of methods of reproducing standard degrees of cerebral concussion and of measuring both the quantity of injury inflicted and the amount of effect produced have opened the possibility of measuring the relation of concussion to nervous metabolism, ... to shock, to fatigue, and to changes induced by anoxia, acidosis, electrical shock, and so forth." Millikan (1948, p. 317) writes of respiratory research that "The technics [sic] and instruments developed for aviation are already widely employed. Few researches on respiratory problems now fail to make use somewhere of a gas-analysis device, or of an instrument for measuring blood gases, or of a demand valve, developed by [OSRD] for war purposes."

New technology platforms

For several major drug categories, the CMR effort also introduced new techniques and platforms that supported continued development after the war ended. Chief among these was in the synthesis of new antibiotics: as Swann (1983, p. 189) points out, although CMR's synthetic penicillin program was a flop during the war, knowledge developed through this work "paved the way for [the] general synthesis of penicillins in the 1950s, [leading] to the development of the therapeutically invaluable semisynthetic penicillins," including dozens of antibiotics introduced in the 1950s (many of which form the basis for results in Section 5; also see de Haen 1976).

In vaccines, methods developed during the war (e.g., centrifugation techniques) subsequently became "state of the art" to the industry by the 1960s (Hoyt 2006, p. 47), contributing to the surge in new vaccines introduced in the 1950s and 1960s. Similar dynamics applied to steroids: the methods developed for producing the "miracle drug" cortisone (a general purpose steroid) during the war were used for developing other corticosteroids afterwards, and Achilladelis (1999, p. 62) writes that "because the technology had diffused among participants of the OSRD project, all of [the firms involved] introduced corticosteroid drugs in the 1950s."

New research capabilities

Beyond specific drugs, several scholars have argued that the war significantly expanded American pharmaceutical companies' general research capabilities, drawing them closer to science, and that this was an important catalyst for a more innovative and competitive U.S. pharmaceutical industry (e.g., Temin 1979, Cockburn et al. 1999, Pisano 2002). Landau et al. (1999), for example, claim that "To a great extent the U.S. government's wartime policies led to the emergence of the American pharmaceutical industry as the undisputed worldwide leader," observing that "the federal war effort encouraged corporate research and development, widened and deepened the companies' cooperation with academic institutions, and catalyzed the diffusion of new technologies across the industry." This was the case for both incumbent drug companies and new ones, as some non-pharmaceutical firms that became involved in drug development during the war continued in it afterwards—most notably Pfizer, which prior to the war was a chemical manufacturing company, became involved in

the natural penicillin program due to its experience with fermentation, and after the war pivoted to drug development, initially focusing on antibiotics.

New collaboration patterns

As we discussed in Section 2, CMR was more than a passive funding agency: it took an active role in organizing research to attack specific military needs. In many cases this required creating networks of academic researchers, hospitals, firms, and (at times) military partners around military medical problems, coordinating the acquisition of inputs, synthesis of drug candidates, fundamental research, clinical testing, and transitions to large-scale manufacturing.

This pattern can be seen across CMR's portfolio. For example, the American Red Cross' blood collection efforts supplied researchers with inputs for their investigation of blood plasma fractionation, the composition of blood fractions, and blood preservation—work which was centered at Harvard (in the laboratory of and under the direction of Edwin Cohn, a physical chemist at Harvard Medical School) but also distributed to and coordinated with researchers at the University of Wisconsin, Stanford University, Columbia University, and the Massachusetts Institute of Technology, with clinical testing units in cities across the country—while the Harvard team operated a pilot plant in conjunction with its laboratory research and worked with several pharmaceutical companies to transition plasma fractionation into production at scale. CMR's efforts to develop new insect repellents drew synthesized candidates from several universities, firms, and government laboratories, which were forwarded to the Department of Agriculture's Bureau of Entomology and Plant Quarantine's testing facilities in Orlando, whose results informed research on fundamental mechanisms at several other universities; here, Scholz (1948, p. 651) observed the "cooperation of government, private industry, and university groups," and noted that "although born of necessity during the war, the usefulness of such coordination of diverse interests for peacetime need is clearly evident." In aviation medicine, Bronk (1948, p. 209) describes how CMR organized a "cooperative scientific effort" which engaged "hundreds of medical scientists in more than a score of universities," and how it brought into "frequent conference groups of scientists working on related problems", and in nutrition Youmans and Guest (1948, p. 473) describe how CMR organized regular "conferences in which investigators and military liaison representatives met to discuss work in progress and practical applications of new information as it was gained."

The continuation of even a portion of these collaborative structures may have contributed to CMR's sustained effects. Although in many cases these fizzled, in some cases under pressures of competition—not only in the product market, but also for scientific talent, which was in high demand after the war and underdeveloped during it (Winternitz 1948)—in other cases collaborations persisted. In the study of blood, for example, relationships between Cohn's lab, blood banks, and firms with fractionation capabilities which Cohn established during the war endured after it ended, and supported Cohn's own postwar research (Creager 1999).

New fundamental knowledge

The most consistent pattern in contemporary accounts is that despite CMR's applied focus, its work produced fundamental understanding which could enable postwar research in new directions—for example, knowledge of the composition of separated blood fractions, the etiology of specific ailments like motion sickness, the epidemiology of diseases like streptoccal infections and rheumatic fever, and many insights into human, animal, and insect physiology.[9]

This anticipated postwar scientific opportunity is described in several contemporary accounts. For example, Youmans and Guest (1948, p. 487) write of nutrition: "The research program ... produced an enormous amount of new scientific information which, aside from its immediate practical value, forms a firm basis for advancing research to be continued in peacetime." Of aviation research, Millikan (1948, p. 316) writes: "As a result of the intensive study of the mechanism of respiratory processes stimulated by aviation's need, there is now a much wider understanding [of] these problems. The application of this knowledge to respiratory abnormalities of disease, in poliomyelitis, emphysema, tuberculosis, and pneumonia, is widespread." Hirshfeld (1948, p. 126) similarly observes of penicillin that "as [its] supply ... became more abundant and the knowledge concerning its action became available, it was possible to expand the categories of infections on which it could be tried" as well as the contexts where it could be applied.

These opportunities also emerged from unlikely topics like chemical warfare. Gilman and Cattell (1948, p. 546) explain that "[CMR's] approach has yielded unsuspected and fruitful byproducts ... not only have valuable research tools for the solution of many of the problems of cellular physiology and metabolism been uncovered, but also many potential therapeutic agents have appeared." With respect to the study of toxic chemicals on the eyes, Friedenwald and Hughes Jr (1948, p. 620) argue that "the net result ... has been to disclose that mustard and related compounds are useful tools in the study of a number of recondite fields of cellular and tissue physiology. In these fields more questions have been made accessible to study than have so far been answered." Regarding research on treatments for respiratory irritants, Gerard (1948, p. 566) explains: "Although the immediate practical results from the current studies have been few ... it has given us far deeper understanding ... and it has pointed the directions in which further research may be expected to yield profit." He later goes on to explain that "much knowledge on the physiology of respiration and circulation has accrued" that was valuable in attacking related medical problems.

Given CMR's heavy focus on developing new chemicals and drugs, a common theme emerges around opportunities to study mechanisms of action. This was especially the case with regards to malaria, parasites, and insect control. On insect repellents, Scholz (1948, p.651) writes: "It is only now becoming possible to correlate the chemical properties and physical characteristics of compounds with their insect-repellency effectiveness. Thousands of candidates remain to be

[9]Other examples (among many) include "fundamental knowledge related to clothing and climatic problems" Robinson and Belding (1948, p. 519), "the fundamental biochemistry of water disinfection" Fair (1948, p. 521), "fundamental studies [of] toxic chemicals in the chemistry and physiology of the cornea" Friedenwald and Hughes Jr (1948, p. 603) and "the fundamental mechanisms" of malaria Carden Jr (1948, p. 666).

analyzed in this light." Haller and Cristol (1948, p. 626) similarly explain of new insecticides: "No satisfactory correlations between chemical structure and toxicity to insects have yet been discovered ... [and] until the effect of various materials on biologic processes is fully understood, it will not be possible to systematize research in the chemistry of insecticides." Yeager (1948, p. 631) continues by explaining that "a generally acceptable theory of the lethal action of DDT must await the acquisition of more experimental data," but that "promising lines of attack ... have been opened up." Regarding the hunt for antimalarials, Carden Jr (1948, p. 670) notes that "intense efforts were made to coordinate the relation between the action of a compound and its chemical configuration, [and] although the complete answer was not found, interesting correlations of the accumulated data, which may bear fruit in the future, were brought to light. Likewise, intense efforts were made to understand the basic biochemical and biologic characteristics of the various malarial parasites, [and] much knowledge was gained that not only throws light on the basic biology of this disease but may also add to knowledge in other lines of investigation."

Section 7 of the paper empirically assesses these mechanisms within the context of our data. Table 7 there provides evidence consistent with these explanations for CMR's enduring effects in both science and technology, with evidence of biomedical science growing around CMR-developed drugs, tools and techniques, and knowledge, and of the postwar pharmaceutical industry harnessing CMR-supported research capabilities and technology platforms. In addition to these mechanisms, the postwar growth of science-driven drug discovery—which we connect to CMR in Section 5 and Appendix C.8—appears to mark a structural break for the U.S. pharmaceutical industry with respect to how new drugs were developed. Beyond new capabilities and platforms, this adoption of a scientifically-driven approach to drug discovery (versus trial-and-error empiricism) was likely also a catalyst of the postwar acceleration of drug innovation.

E Theoretical Perspective

Figure 1 indicates that the 1940s marked a turning point in U.S. biomedical innovation, which transitioned from stagnation into growth. The concurrence of the CMR shock is evocative of "big push" dynamics which scholars have evaluated in the context of industrialization and economic growth. Rosenstein-Rodan (1943) introduced the argument that industrialization may require a simultaneous, economy-wide investment to unlock external economies that arise vis-à-vis demand spillovers or shared infrastructure. Murphy et al. (1989) formalized this argument with a multiple equilibrium model, showing that a big push can move an economy from a (bad) equilibrium with low industrialization to a (good) equilibrium with high industrialization.

However, how Murphy et al. (1989)'s structure and dynamics translate to an industrial innovation system—and especially to the CMR shock—is unclear. For example, Murphy et al. (1989) examine a static investment problem, but innovation is dynamic. Their mechanism operates through demand spillovers across sectors, but CMR was largely a supply shock.

In this section, we introduce a simple model of an innovation system (in the spirit of Nelson and Rosenberg 1993) which integrates "big push" logic, adapting the Rosenstein-Rodan and Murphy et al. argument to an innovation-driven economy. We conjecture that the 1940s is when U.S. biomedical innovation began showing system dynamics, with its constituent parts—firms, universities, and government, or science and technology—becoming more deeply connected and mutually reinforcing as an engine of growth. Our intuition is that this was instigated by the CMR (and broader World War II) shock, which created this sinew and transitioned the economy into a new equilibrium of long-run growth around the science-technology flywheel. In doing so, a secondary goal of this exercise is to introduce Nelson and Rosenberg (1993)'s narrative "innovation systems" framework into economic models of innovation-driven growth.

We are specifically interested in innovation problems where scientific research, technology development, and manufacturing capabilities (inputs) are needed to produce a final good. We assume each of the inputs is supplied by a monopolist, that each input feeds off of progress in the others, that all inputs' level of development is initially zero, and no output can be produced if *either* (i) any of these inputs is missing or (ii) they are unable to integrate with each other. The assumption of input sector monopolists is to simplify the algebra but can be replaced by imperfectly competitive sectors where innovation is cumulative and enters the public after one period. We thus take an ecosystem view of innovation, where science, technology, and production interdepend, and emerging science-based technologies therefore face a cold-start problem: no input sector will begin to invest in improvement until it can build on innovation in the others.

E.1 Baseline model of an innovation system

E.1.1 Basic environment

The innovation economy we study thus consists of a final goods sector (for example, health care) which combines science and technology into output (human health). To simplify the exposition, we will assume technology development and manufacturing occur in the same firm (e.g., a pharmaceutical firm) and collapse the inputs to two sectors, which we call *research* (R) and *development* (D).[10] Each of these input sectors consists of a monopolist supplying innovation to the final goods sector, which is perfectly competitive. Each input sector has an evolving knowledge stock R_t and D_t, which evolve according to the following symmetric processes:

$$\begin{aligned} R_{t+1} &= (1-\delta_K)\, R_t + e_{R,t}\lambda(R_t D_t)^{\gamma} \\ D_{t+1} &= (1-\delta_K)\, D_t + e_{D,t}\lambda(R_t D_t)^{\gamma} \end{aligned} \tag{4}$$

where $\delta_K \in [0,1]$ is depreciation and $e_{j,t} \geq 0$ (for $j \in \{R, D\}$) is an investment in incrementing the knowledge stock (i.e., effort), which each firm must choose in every period t. We assume the impact of effort on innovation is increasing in the existing same-sector knowledge stock (in the spirit of Romer 1990, motivated by the recombination which can take place within each sector—echoing the evidence in Sections 4 and 5) and complementary to the other sector's knowledge, and with $\gamma > 0$ embodying a science-technology interplay. $\lambda \in [0,1]$ measures the degree to which each sector can harness these complementarities, and represents the degree of connective tissue in the system (through collaborative relationships, absorptive capacity, or other means). Further note that we are assuming in Equation (4) that effort has no standalone contribution to R and D outside of the complementarity term, which simplifies algebra while still allowing us to explore the consequences of interdependence—a choice we will return to later in this appendix.

These knowledge stocks are combined one period later in the final goods sector according to:

$$Y_{t+1} = A(R_t D_t)^{\beta}\,, \tag{5}$$

with $A > 0$ and $\beta < 1$. Factor prices on the inputs are in turn:

$$p_{R,t+1} = \beta \frac{Y_{t+1}}{R_t}\,, \quad p_{D,t+1} = \beta \frac{Y_{t+1}}{D_t}\,. \tag{6}$$

[10] The choice to collapse to two sectors is in part to reflect the structure of the mid-century pharmaceutical industry and in part to simplify the exposition, but it is not be material to the logic or results. The modern biomedical innovation system is arguably better characterized by a three-sector model of research (R), development (D), and manufacturing (M), reflecting the division of labor between biotech startups (which develop science into new drugs) and large pharmaceutical companies (which often externally acquire promising candidates and commercialize them). A three-sector model of this flavor generates similar dynamics to what we will find below.

E.1.2 The input sectors' problem

Each period, the firm in each input sector j makes an R&D investment to maximize discounted, next-period profits—a presumption was can also generate with a finite horizon (e.g., two period) economy. Without loss of generality we consider the R sector and drop the j subscript, as the R and D sectors are symmetric. The firm in each input sector earns rents from the available knowledge stock each period, which it sells at price $p_{j,t}$ (defined in Equation 6).

The firm thus has the following objective, in which it chooses $e_{R,t}$ to maximize discounted profits ($\delta\, p_{R,t+1}\, R_{t+1}$, for $\delta < 1$) on its innovation net of effort costs $C(e_{R,t})$:

$$\max_{e_{R,t}} \delta\, p_{R,t+1}\, R_{t+1}(e_{R,t}) - C(e_{R,t})\ .$$

Substituting in factor prices, effort costs, and the law of motion, and taking first-order conditions, we arrive at following equilibrium investment $e^*_{R,t}$ in Lemma 1:

Lemma 1. *Equilibrium innovation effort in R is:*

$$e^*_{R,t} = \lambda\, \delta\, \beta\, A\, (R_t D_t)^{\beta+\gamma}\, R_t^{-1} \tag{7}$$

Lemma 1 has two important implications. The first is that when R&D asset stocks (R_t and D_t) are low, research investment ($e_{R,t}$) will be low. The second is that if the R and D sectors are unable to integrate ($\lambda = 0$), research investment will be zero. Both are necessary conditions for the research sector to grow (and by symmetry, the technology sector as well).

E.1.3 Implications for growth

Substituting (7) into (4), recognizing that $R_t = D_t \equiv X_t$ in equilibrium (due to symmetry), and defining algebraic constants $\kappa \equiv \lambda\delta\beta A$ and $\theta \equiv 2\beta + 4\gamma - 1$, we can rewrite the equilibrium law of motion for each sector in a more simplified fashion as:

$$\begin{aligned} X_{t+1} &= (1-\delta_K)\, X_t + \left(\delta\, \beta\, A\, (X_t X_t)^{\beta+\gamma}\, X_t^{-1}\right) (X_t X_t)^{\gamma} \\ &= (1-\delta_K)\, X_t + \delta\beta A X_t^{2\beta+4\gamma-1} = (1-\delta_K)\, X_t + \kappa X_t^{\theta} \end{aligned} \tag{8}$$

In the steady state, $X = (1-\delta_K)\, X + \kappa X^{\theta}$, or equivalently $f(X) \equiv \kappa X^{\theta} - \delta_K X = 0$. It is then straightforward to identify equilibria, which we summarize in Proposition 1:

Proposition 1. *Let g_X denote the growth rate of knowledge stocks and (analogously) g_Y the growth rate of final good output. The innovation system has two steady states:*

1. *Stagnation: the input sectors are stuck in their initial state, with zero innovation ($X = X_0 = 0$, for $X \in \{R, D\}$). Accordingly, output never grows ($g_X = g_Y = 0$).*

2. *Runaway growth: if a coordinated investment increases knowledge stocks above* $\bar{X} = \left(\frac{\delta_K}{\kappa}\right)^{\frac{1}{\theta-1}}$, *the economy will enter a permanently accelerating state of growth.*

The innovation system thus has two equilibria: one where science and technology are stuck at low levels of development with no investment, and one where both receive regular investment, which drives sustained growth. Moreover, when $\theta > 1$, X_0 is locally stable, and the economy is stuck in a no-innovation equilibrium absent a coordinated investment or systemic shock that brings innovation stocks above zero (and for sustained growth, above $\bar{X}$). The machinery of the model implies that if the complementarity of science and technology in biomedical innovation (γ) is large and these inputs are reasonably productive in the production of health care (β), such that $\theta = 2\beta+4\gamma-1 > 1$, X_0 will be an absorbing state. The degree of complementarity between science and technology need not be large for this economy to get stuck: for example, if the output sector has constant returns to scale ($\beta = 0.5$), γ need only be above 0.25 for it to be lodged in the no-growth equilibrium. The intuition is that with high complementarity, each input sector will have little incentive to invest in its innovation stock without coordinated action from the other. Alternatively, if $\lambda = 0$, the economy will also be stuck in a no-growth equilibrium—even if R&D assets stocks are non-zero—because the R and D sectors are unable to harness their complementarity.

Proposition 1, however, also reveals that the key to achieving a positive-growth steady state is a coordinated effort to endow the input sectors with enough knowledge and connective tissue (the push; Rosenstein-Rodan 1943) to kick off an innovation cycle where science and technology can endogenously begin to build on each other. Once this process is kickstarted, it can become self-sustaining, requiring only that $\lambda > 0$ and the initial investment exceeds the threshold $\bar{X}$ identified in Proposition 1. The more integration there is (the closer λ is to 1), the faster growth will be. In this steady state, the innovation sectors operate as an interdependent system, each encouraging further investment in, and building on the output of, the other.

It is useful to also acknowledge that this model predicts a zero-investment equilibrium, where $X \in \{R, D\}$ is permanently stuck in an undeveloped state at $X = 0$, whereas Figure 1 of the paper makes clear that biomedicine existed pre-1940 and produced a steady flow of papers and patents, which simply weren't growing—a condition we interpret as an inertial, low-development state. In the model, the no-investment equilibrium is a result of firms not having any standalone returns to effort, which is fully loaded onto the complementarity term $\lambda(R_t D_t)^\gamma$ in Equation (4). The model can be written to produce steady state positive but stagnant R&D investment (matching the pre-1940 historical pattern) by allowing firms to earn direct monopoly returns to R&D independent of the complementarity—for example, by modeling the process for R_t (in Equation 4) to include a term $\phi e_{R,t}$ (and analogously for D_t)—but at the cost of more complicated algebra, without meaningfully different conclusions. For this reason, we keep the model as written.

E.2 Extension: Achieving balanced growth

Though the second equilibrium in Proposition 1 generates growth, it lacks a balanced growth path: complementarities dominate knowledge depreciation in the law of motion for R_{t+1} (and analogously, D_{t+1}) and lead to ever-increasing, unconstrained per-period investment in innovation, which accordingly delivers unconstrained growth. The real economy, however, has not had any such takeoff (yet). Prior research has posited multiple reasons why innovation may be so constrained, including the possibility that as knowledge develops, further progress grows more costly (e.g., Jones 2009, 2010). To accommodate a burden of knowledge we redefine incremental innovation costs to be a function of the existing knowledge stock. Concretely, let:

$$C(e_{R,t}, R_t) = \frac{1}{2}\left[1 + R_t^{\theta-1}\right] e_{R,t}^2 \tag{9}$$

where θ is defined as before—such that the burden is increasing in the interdependence (i.e., complexity; Simon 1962) of scientific and technological knowledge.

The change in costs implies a slightly different equilibrium effort, which we denote as $\hat{e}_{R,t}^*$ (to distinguish it from $e_{R,t}^*$ in Lemma 1) and characterize in Lemma 2:

Lemma 2. *With a burden of knowledge, equilibrium innovation effort in R is:*

$$\hat{e}_{R,t}^* = \frac{\lambda\,\delta\,\beta\,A\,(R_t D_t)^{\beta+\gamma}\,R_t^{-1}}{1 + R_t^{\theta-1}} \tag{10}$$

Applying the same substitutions and constants as we used to derive (8), we obtain:

$$X_{t+1} = (1-\delta_K)\,X_t + \kappa\left(\frac{X_t^\theta}{1 + X_t^{\theta-1}}\right) \tag{11}$$

As X grows large, $\lim_{X\to\infty} X_{t+1} = (1-\delta_K)\,X_t + \kappa X_t$. Defining $g_X \equiv \frac{X_{t+1}}{X_t}$, we then get a possibility of long-run balanced growth of $g_X = 1 - \delta_K + \kappa$, as in Proposition 2:

Proposition 2. *An innovation system with a burden of knowledge also has two steady states:*

1. *Stagnation: the absence of innovation and output growth* (as in Proposition 1)
2. *Balanced growth: if a coordinated investment increases knowledge stocks above $\bar{X} = \left(\frac{\delta_K}{\kappa - \delta_K}\right)^{\frac{1}{\theta-1}}$, then provided $\kappa > 2\delta_K$, the economy will enter a state of growth which will converge to a balanced growth path, with $g_X = 1 - \delta_K + \kappa$ and $g_Y = g_X^{2\beta}$.*

E.3 Proofs

Lemma 1. Equilibrium innovation effort in R is:

$$e_{R,t}^* = \lambda\,\delta\,\beta\,A\,(R_t D_t)^{\beta+\gamma}\,R_t^{-1} \tag{12}$$

Proof. Recall that the firm's problem is:

$$\max_{e_{R,t}} \delta\, p_{R,t+1}\, R_{t+1}(e_{R,t}) - C(e_{R,t})\ .$$

Taking first-order conditions and solving for $e^*_{R,t}$:

$$\begin{aligned}
0 &= \frac{\partial}{\partial e_{R,t}} \left[\delta\, p_{R,t+1}\, R_{t+1}(e_{R,t}) - C(e_{R,t})\right] \\
&= \delta\, p_{R,t+1}\, R'_{t+1}(e_{R,t}) - C'(e_{R,t}) \\
&= \delta\, p_{R,t+1}\, \lambda\, (R_t D_t)^{\gamma} - e_{R,t} \\
&= \delta \left(\beta \frac{Y_{t+1}}{R_t}\right) \cdot \lambda\, (R_t D_t)^{\gamma} - e_{R,t} \\
&= \delta\, \beta \left(A\, (R_t D_t)^{\beta}\, R_t^{-1}\right) \cdot \lambda\, (R_t D_t)^{\gamma} - e_{R,t} \\
&= \lambda\, \delta\, \beta \left(A\, (R_t D_t)^{\beta+\gamma}\, R_t^{-1}\right) - e_{R,t}
\end{aligned}$$

Hence $e^*_{R,t} = \lambda\, \delta\, \beta \left(A\, (R_t D_t)^{\beta+\gamma}\, R_t^{-1}\right)$.

Proposition 1. Let g_X denote the growth rate of knowledge stocks and (analogously) g_Y the growth rate of final good output. The innovation system has two steady states:

1. Stagnation: the input sectors are stuck in their initial state, with zero innovation ($X = X_0 = 0$, for $X \in \{R, D\}$). Accordingly, output never grows ($g_X = g_Y = 0$).

2. Runaway growth: if a coordinated investment increases knowledge stocks above $\bar{X} = \left(\frac{\delta_K}{\kappa}\right)^{\frac{1}{\theta-1}}$, the economy will enter a permanently accelerating state of growth.

Proof. That $X_0 = 0$ is a steady state is straightforward. For the positive steady state, it suffices to show that it exists and is unique, and to evaluate stability. To prove the existence and uniqueness of the positive steady state, let us define $F(X) \equiv (1 - \delta_K)X + \kappa X^{\theta}$, and $G(X) = \kappa X^{\theta} - \delta_K X$. We seek fixed points such that $F(X) = X$, or equivalently $G(X) = 0$.

Recognize that $G(X)$ is continuously differentiable with $G(0) = 0$, $\lim_{X \longrightarrow 0^+} G(X) < 0$, $\lim_{X \longrightarrow \infty} G(X) > 0$, and $G''(.) > 0$ (i.e., $G(.)$ is convex). By the intermediate value theorem, $G(.)$ has a unique root in the $(0, \infty)$ interval, which we can solve the law of motion to obtain as: $\bar{X} = \left(\frac{\delta_K}{\kappa}\right)^{\frac{1}{\theta-1}}$. To determine stability we evaluate whether $\left|F'(\bar{X})\right| \gtrless 1$ as follows:

$$\begin{aligned}
F'(X) &= (1 - \delta_K) + \kappa\theta X^{\theta-1} \\
F'(\bar{X}) &= (1 - \delta_K) + \kappa\theta \bar{X}^{\theta-1} \\
&= (1 - \delta_K) + \kappa\theta \left(\left(\frac{\delta_K}{\kappa}\right)^{\frac{1}{\theta-1}}\right)^{\theta-1}
\end{aligned}$$

$$
\begin{aligned}
&= (1-\delta_K) + \theta\delta_K \\
&= 1 + (\theta-1)\delta_K > 1\,,
\end{aligned}
$$

because $\theta - 1 > 0$ and $\delta_K > 0$. Hence, the fixed point $\bar{X}$ is unstable, and when X is endowed at $X > \bar{X}$, innovation will subsequently endogenously grow.

Finally, we examine the limit growth rate: $\lim_{X\to\infty} X_{t+1} = (1-\delta_K)\,X_t + \kappa X_t^{\theta}$, and $g_X = \frac{X_{t+1}}{X_t} = 1 - \delta_K + \kappa X^{\theta-1}$. Because $\theta > 1$, g_X will accelerate unbounded as the innovation stock X increases, and both input and output growth are therefore unbalanced.

Lemma 2. With a burden of knowledge, equilibrium innovation effort in R is:

$$
\hat{e}^*_{R,t} = \frac{\lambda\,\delta\,\beta\,A\,(R_t D_t)^{\beta+\gamma}\,R_t^{-1}}{1 + R_t^{\theta-1}} \tag{13}
$$

Proof. Recall that the firm's problem is:

$$
\max_{e_{R,t}} \delta\, p_{R,t+1}\, R_{t+1}(e_{R,t}) - C(e_{R,t})\,.
$$

Taking first-order conditions and solving for $e^*_{R,t}$:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial e_{R,t}}\left[\delta\, p_{R,t+1}\, R_{t+1}(e_{R,t}) - C(e_{R,t})\right] \\
&= \delta\, p_{R,t+1}\, R'_{t+1}(e_{R,t}) - C'(e_{R,t}) \\
&= \delta\, p_{R,t+1}\, \lambda\,(R_t D_t)^{\gamma} - \left[1 + R_t^{\theta-1}\right] e_{R,t} \\
&= \delta\,\left(\beta\frac{Y_{t+1}}{R_t}\right)\cdot \lambda\,(R_t D_t)^{\gamma} - \left[1 + R_t^{\theta-1}\right] e_{R,t} \\
&= \delta\,\beta\,\left(A\,(R_t D_t)^{\beta}\,R_t^{-1}\right)\cdot \lambda\,(R_t D_t)^{\gamma} - \left[1 + R_t^{\theta-1}\right] e_{R,t} \\
&= \lambda\,\delta\,\beta\,\left(A\,(R_t D_t)^{\beta+\gamma}\,R_t^{-1}\right) - \left[1 + R_t^{\theta-1}\right] e_{R,t}
\end{aligned}
$$

Hence $\hat{e}^*_{R,t} = \frac{\lambda\,\delta\,\beta\left(A(R_t D_t)^{\beta+\gamma} R_t^{-1}\right)}{1+R_t^{\theta-1}}$.

Proposition 1. An innovation system with a burden of knowledge also has two steady states:

1. Stagnation: the absence of innovation and output growth *(as in Proposition 1)*
2. Balanced growth: if a coordinated investment increases knowledge stocks above $\bar{X} = \left(\frac{\delta_K}{\kappa-\delta_K}\right)^{\frac{1}{\theta-1}}$, then provided $\kappa > 2\delta_K$, the economy will enter a state of growth which will converge to a balanced growth path, with $g_X = 1 - \delta_K + \kappa$ and $g_Y = g_X^{2\beta}$.

Proof. That $X_0 = 0$ is a steady state is straightforward. For the positive steady state, it suffices to show that it exists and is unique, and to evaluate stability. To prove the existence and uniqueness of the positive steady state, let us define $F(X) \equiv (1-\delta_K)\,X + \kappa\left(\frac{X^\theta}{1+X^{\theta-1}}\right)$, and $G(X) = \left(\frac{\kappa X^\theta}{1+X^{\theta-1}}\right) - \delta_K X$. We seek fixed points such that $F(X) = X$, or equivalently $G(X) = 0$.

Recognize that $G(X)$ is continuously differentiable with $G(0) = 0$, $\lim_{X\longrightarrow 0^+} G(X) < 0$, $\lim_{X\longrightarrow\infty} G(X) > 0$, and $G''(.) > 0$ (i.e., $G(.)$ is convex). By the intermediate value theorem, $G(.)$ has a unique root in the $(0,\infty)$ interval, which we can solve the law of motion to obtain as: $\bar{X} = \left(\frac{\delta_K}{\kappa-\delta_K}\right)^{\frac{1}{\theta-1}}$. To determine stability we evaluate whether $\left|F'(\bar{X})\right| \gtrless 1$ as follows:

$$
\begin{aligned}
F'(X) &= (1-\delta_K) + \kappa\frac{(\theta X^{\theta-1})(1+X^{\theta-1}) - (X^\theta)((\theta-1)X^{\theta-2})}{(1+X^{\theta-1})^2} \\
&= (1-\delta_K) + \kappa\frac{\theta X^{\theta-1} + \theta X^{2(\theta-1)} - \theta X^{2(\theta-1)} + X^{2(\theta-1)}}{(1+X^{\theta-1})^2} \\
&= (1-\delta_K) + \kappa X^{\theta-1}\frac{\theta + X^{\theta-1}}{(1+X^{\theta-1})^2} \\
&= (1-\delta_K) + \kappa\left(\frac{\delta_K}{\kappa-\delta_K}\right)\frac{\theta + \left(\frac{\delta_K}{\kappa-\delta_K}\right)}{(1+\left(\frac{\delta_K}{\kappa-\delta_K}\right))^2} \\
&= (1-\delta_K) + \kappa\left(\frac{\delta_K}{\kappa-\delta_K}\right)\frac{\theta + \left(\frac{\delta_K}{\kappa-\delta_K}\right)}{\left(\frac{\kappa}{\kappa-\delta_K}\right)^2} \\
&= (1-\delta_K) + \delta_K\left(\frac{\kappa-\delta_K}{\kappa}\right)\left(\theta + \left(\frac{\delta_K}{\kappa-\delta_K}\right)\right) \\
&= (1-\delta_K) + \delta_K\left(1-\frac{\delta_K}{\kappa}\right)\theta + \frac{\delta_K^2}{\kappa} \\
&= (1-\delta_K) + \theta\delta_K - \theta\frac{\delta_K^2}{\kappa} + \frac{\delta_K^2}{\kappa} \\
&= 1 + (\theta-1)\delta_K - (\theta-1)\frac{\delta_K^2}{\kappa} \\
&= 1 + (\theta-1)\delta_K\left(1-\frac{\delta_K}{\kappa}\right) > 1
\end{aligned}
$$

because $\theta - 1 > 0$, $\delta_K > 0$, and $\kappa > \delta_K$ (and in turn, $1 - \frac{\delta_K}{\kappa} > 0$).

Hence, the fixed point $\bar{X}$ is unstable, and when X is endowed at $X > \bar{X}$, innovation will subsequently endogenously grow. At the limit, $\lim_{X\to\infty} X_{t+1} = (1-\delta_K)\,X_t + \kappa X_t$, and $g_X = \frac{X_{t+1}}{X_t} = 1 - \delta_K + \kappa$. In turn, $g_Y = \frac{Y+1}{Y_t} = \left(\frac{R_{t+1}D_{t+1}}{R_t D_t}\right)^\beta = \left(\frac{X_{t+1}}{X_t}\right)^{2\beta} = g_X^{2\beta}$.

Appendix references

Achilladelis, Basil. 1999. "Innovation in the pharmaceutical industry." In *Pharmaceutical innovation: Revolutionizing human health*, edited by Ralph Landau, Basil Achilladelis, and Alexander Scriabine. Philadelphia: Chemical Heritage Press.

American Red Cross. 1918. *War Medicine*.

Andrus, Edwin C. 1948. *Advances in military medicine*. Boston: Little, Brown, and Company.

Baxter, James Phinney. 1946. *Scientists against time*. Boston: Little, Brown and Company.

Boyack, Kevin W, Michael Patek, Lyle H Ungar, Patrick Yoon, and Richard Klavans. 2014. "Classification of individual articles from all of science by research level." *Journal of Informetrics* **8**(1):1-12.

Bronk, Detlev. 1948. "Introduction: Aviation Medicine." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 207-221. Boston: Little, Brown, and Company.

Bryan, Kevin A, Yasin Ozcan, and Bhaven Sampat. 2020. "In-text patent citations: A user's guide." *Research Policy* **49**(4):103946.

Carden Jr, George A. 1948. "Introduction: Malaria." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 665-670. Boston: Little, Brown, and Company.

Cardoni, Alex A. 1973. "Comparison of Current Contents' Life Sciences and Clinical Practice Editions." *American Journal of Health-System Pharmacy* **30**(8):721-722.

Catillon, Maryaline. 2017. *Medical knowledge synthesis: A brief overview*. Working paper.

Coates, John Boyd Jr. and Ebbe Curtis Hoff (eds.). 1958. *Preventative Medicine in World War II, Volume IV: Communicable Diseases Transmitted Chiefly Through Respiratory and Alimentary Tracts*. Washington DC: Department of the Army.

Cockburn, Iain, Rebecca Henderson, Luigi Orsenigo, and Gary Pisano. 1999. "Pharmaceuticals and biotechnology." In *U.S. industry in 2000: Studies in competitive performance*, edited by David C. Mowery, pp. 363-398. Washington DC: National Academies Press.

Cohn, Edwin J. 1948. "The History of Blood Plasma Fractionation." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 364-443. Boston: Little, Brown, and Company.

Creager, Angela N. H. 1999. "'What Blood Told Dr Cohn': World War II, plasma fractionation, and the growth of human blood research." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **30**(3):377-405.

de Haen, Paul. 1976. "Compilation of New Drugs: 1940 thru 1975." *Pharmacy Times* **March**:40-74.

Fair, Gordon M. 1948. "Water Disinfection and Allied Subjects." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 520-531. Boston: Little, Brown, and Company.

Friedenwald, Jonas S and W P Hughes Jr. 1948. "The Effects of Toxic Chemical Agents on the Eyes and Their Treatment." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 603-620. Boston: Little, Brown, and Company.

Garfield, E. 1972. "Current Contents/Clinical Practice: New tool for optimum selection in maintaining current awareness."

Geer, Harriet A, Randall Latta, and Arthur W Lindquist. 1948. "The Dispersal of Insecticides." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 636-645. Boston: Little, Brown, and Company.

Gerard, Ralph W. 1948. "Recent Research on Respiratory Irritants." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 565-587. Boston: Little, Brown, and Company.

Gilman, Alfred Z and McKeen Cattell. 1948. "Systemic Agents: Action and Treatment." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 546-564. Boston: Little, Brown, and Company.

Gross, Daniel P. and Maria P. Roche. 2024. *Coordinated R&D Programs and the Creation of New Industries*. Working paper.

Gross, Daniel P. and Bhaven N. Sampat. 2023a. "America, jump-started: World War II R&D and the takeoff of the US innovation system." *American Economic Review* **113**(12):3323-3356.

Gross, Daniel P. and Bhaven N. Sampat. 2023b. "The World War II crisis innovation model: What was it,

and where does it apply?" *Research Policy* **52**(9):104845.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg. 2001. *The NBER patent citation data file: Lessons, insights and methodological tools*. NBER Working Paper No. 8498.

Haller, H L and Stanley J Cristol. 1948. "The Development of New Insecticides." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 621-626. Boston: Little, Brown, and Company.

Hirshfeld, John Winslow. 1948. "Introduction: Surgery." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 123-133. Boston: Little, Brown, and Company.

Hoyt, Kendall. 2006. "Vaccine innovation: Lessons from World War II." *Journal of Public Health Policy* **27**(1):38-57.

Ke, Qing. 2019. "Identifying translational science through embeddings of controlled vocabularies." *Journal of the American Medical Informatics Association* **26**(6):516-523.

Keefer, Chester S. 1969. "Dr. Richards as Chairman of the Committee on Medical Research." *Annals of Internal Medicine* **71**(8):61-70.

Landau, Ralph, Basil Achilladelis, and Alexander Scriabine. 1999. *Pharmaceutical innovation: Revolutionizing human health*. Philadelphia: Chemical Heritage Foundation.

Lewison, Grant and Guillermo Paraje. 2004. "The classification of biomedical journals by research level." *Scientometrics* **60**(2):145-157.

Li, Danielle, Pierre Azoulay, and Bhaven N. Sampat. 2017. "The applied value of public investments in biomedical research." *Science* **356**(6333):78-81.

Lockwood, John S. 1948. "Introduction: Surgery." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 91-94. Boston: Little, Brown, and Company.

Love, Albert G (ed.). 1925. *The Medical Department of the United States Army in the World War, Volume XV: Medical and Casualty Statistics*. Washington DC: Department of the Army.

Mandel, Richard. 1996. *A half century of peer review, 1946-1996*. Bethesda: National Institutes of Health.

Marx, Matt and Aaron Fuegi. 2020. "Reliance on science: Worldwide front-page patent citations to scientific articles." *Strategic Management Journal* **41**(9):1572-1594.

Marx, Matt and Aaron Fuegi. 2022. "Reliance on science by inventors: Hybrid extraction of in-text patent-to-article citations." *Journal of Economics & Management Strategy* **31**(2):369-392.

Millikan, G A. 1948. "Anoxia and Oxygen Equipment." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 296-317. Boston: Little, Brown, and Company.

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny. 1989. "Industrialization and the big push." *Journal of Political Economy* **97**(5):1003-1026.

Narin, Francis, Gabriel Pinski, and Helen Hofer Gee. 1976. "Structure of the biomedical literature." *Journal of the American Society for Information Science* **27**(1):25-45.

Nelson, Richard R and Nathan Rosenberg. 1993. "Technical innovation and national systems." In *National Innovation Systems*, edited by Richard R Nelson, pp. 3–22. New York: Oxford University Press.

Neushul, Peter. 1993. "Science, government and the mass production of penicillin." *Journal of the History of Medicine and Allied Sciences* **48**(4):371-395.

Pisano, Gary. 2002. "Pharmaceutical biotechnology." In *Technological Innovation and Economic Performance*, edited by Benn Steil, David G Victor, and Richard R Nelson, pp. 347-365. Princeton: Princeton University Press.

Reister, Frank A (ed.). 1975. *Medical Statistics in World War II*. Washington DC: Department of the Army.

Robinson, Sid and H S Belding. 1948. "Protective Clothing." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 497-519. Boston: Little, Brown, and Company.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of Political Economy* **98**(5, Part 2):S71-S102.

Rosenstein-Rodan, Paul N. 1943. "Problems of Industrialisation of Eastern and South-Eastern Europe." *The Economic Journal* **53**(210-211):202-211.

Sampat, Bhaven N. 2023. *The history and political economy of NIH peer review*. Building a Better NIH Project whitepaper, available at https://www.brookings.edu/collection/building-a-better-nih/.

Scholz, Roy O. 1948. "The Development of New Insect Repellents." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 646-651. Boston: Little, Brown, and Company.

Slater, Leo Barney. 2009. *War and disease: Biomedical research on malaria in the twentieth century*. New Brunswick: Rutgers University Press.

Stewart, Irvin. 1948. *Organizing scientific research for war: The administrative history of the Office of Scientific Research and Development*. Boston: Little, Brown, and Company.

Strickland, Stephen Parks. 1988. *The story of the NIH grants program*. Lanham: University Press of America.

Swann, John Patrick. 1983. "The search for synthetic penicillin during World War II." *The British Journal for the History of Science* **16**(2):154-190.

Temin, Peter. 1979. "Technology, regulation, and market structure in the modern pharmaceutical industry." *The Bell Journal of Economics* pp. 429-446.

U.S. Federal Trade Commission. 1980. *Statistical Report on Mergers and Acquisitions*. Produced by the Federal Trade Commission (FTC) Bureau of Economics.

Windle, William F. 1948. "Experimental Studies on Concussion." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 168-175. Boston: Little, Brown, and Company.

Winternitz, Milton C. 1948. "Introduction: Chemical Warfare Agents." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 532-545. Boston: Little, Brown, and Company.

Yeager, J Franklin. 1948. "The Action of DDT on Invertebrates." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 627-631. Boston: Little, Brown, and Company.

Youmans, John B and George M Guest. 1948. "Problems of Nutrition." In *Advances in Military Medicine*, edited by Edwin C. Andrus, pp. 473-487. Boston: Little, Brown, and Company.