NBER WORKING PAPER SERIES

MARKET SIZE AND RESEARCH: EVIDENCE FROM THE PHARMACEUTICAL INDUSTRY

Dennis Byrski
Fabian Gaessler
Matthew J. Higgins

Working Paper 28858
http://www.nber.org/papers/w28858

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2021

We thank Dietmar Harhoff, Kyle Myers, Ariel D. Stern, and Lucy Wang for helpful comments and suggestions. We are grateful to Manuel Hermosilla for sharing the mapping between Cortellis indications and ICD-9 codes with us. We thank Pablo Espinosa-Uriel for his excellent research assistance. Byrski thanks the Laboratory for Innovation Science at Harvard University for their hospitality while writing parts of this paper. Gaessler gratefully acknowledges financial support from the DFG through the Collaborative Research Center TRR 190 "Rationality and Competition". Higgins acknowledges support from the Sorenson Center for Discovery and Innovation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Market Size and Research: Evidence from the Pharmaceutical Industry
Dennis Byrski, Fabian Gaessler, and Matthew J. Higgins
NBER Working Paper No. 28858
May 2021
JEL No. I13,I18,L65,O31,O32

ABSTRACT

Prior literature has established a link between changes in market size and pharmaceutical innovation; whether a link exists with scientific research remains an open question. If upstream research is not responsive to these changes, the kinds of scientific discoveries that flow into future drug development could be disconnected from downstream demand. We explore this question by exploiting the effects of quasi-experimental variation in market size introduced by Medicare Part D. We find no causal relationship between market size and biomedical research in the decade following the implementation of Medicare Part D. While many factors have been shown to motivate scientists to conduct research, this result suggests that changes in market size provide no such incentive. We do find, however, limited support for a response by corporate scientists conducting applied research. Implications for pharmaceutical innovation policy are discussed.

Dennis Byrski
Max Planck Institute for
Innovation and Competition
80539 Munich
Germany
dennis.byrski@ip.mpg.de

Fabian Gaessler
Max Planck Institute for
Innovation and Competition
80539 Munich
Germany
fabian.gaessler@ip.mpg.de

Matthew J. Higgins
David Eccles School of Business
University of Utah
1731 E Campus Center Drive
Salt Lake City, UT 84112
and Max Planck Institute for
Innovation and Competition
and also NBER
matt.higgins@eccles.utah.edu

1 Introduction

"Progress in combating disease depends upon an expanding body of new scientific knowledge."
Vannevar Bush, 1945

What factors drive innovation? This question has actively engaged scholars dating back to Schumpeter (1939). While the importance of this question is obvious for firms and policymakers, the answer is neither singular nor simple. Among others, Mowery and Rosenberg (1979) suggested that both supply-side ('technology push') and demand-side ('demand pull') factors may impact the innovation process. In general, the technology push view suggests that accumulated knowledge from research and development (R&D) activities drives the introduction of new products. In contrast, the demand pull view relies on the notion that demand characteristics (*e.g.*, population, disposable income, and preferences) shape the pattern of investments in innovation and in research (Kyle, 2020).

Focusing on demand-side factors, early work by Griliches (1957) and Schmookler (1966) recognized the importance of profit incentives and market size as drivers of innovation. This recognition has been carried forward into more recent work. Two seminal papers have established a causal link between market size and pharmaceutical innovation (Acemoglu and Linn, 2004; Finkelstein, 2004). Since then, a steady stream of empirical studies has examined how demand affects drug development (*e.g.*, Agarwal and Gaule, 2021; Blume-Kohout and Sood, 2013; Dranove et al., 2020; Dubois et al., 2015; Garthwaite et al., 2021; Kyle and McGahan, 2012; Lichtenberg and Waldfogel, 2009).[1]

These studies, however, focus on the 'D' as opposed to the 'R' in pharmaceutical R&D. They consider molecules entering (pre-)clinical trials, new drug approvals, or various other definitions that generally fall within the traditional rubric of 'development' as opposed to 'research'. Heretofore, efforts to establish a link between downstream demand and upstream research have been met with limited success. Acemoglu and Linn (2004) and Finkelstein (2004), for example, were unable to find a relationship between patenting and demographic-driven expansions or policy-induced expansions in market size, respectively.[2,3] In contrast, Bhattacharya

[1] A few studies confirm these results on innovation responses following market size shocks outside the pharmaceutical industry (*e.g.*, Aghion et al., 2018).

[2] They provide a variety of reasons for this result. First, they highlight the imperfection of their patent match. Second, they describe attenuation issues resulting from the delay in the research process. Third, they point towards companies being more responsive to profit incentives in later development stages.

[3] In her paper, Finkelstein (2004) argues that "[..] the quick initial response in development suggests the existence of a substantial reservoir of technologically feasible products 'on the shelf'. The decision to begin clinical trials is responsive, on the margin, to increases in the expected economic return to the clinical trial."

and Packalen (2011) use a similar identification strategy as Acemoglu and Linn (2004) and identify a positive relationship between disease prevalence and upstream research, but cannot disentangle the increase in societal importance from pure profit incentives. Their finding was driven by applied research and argued to be conducted at academic medical centers.[4] This is precisely where one would expect clinical studies to occur and the corresponding results to be published. Hence, the broader link, if it exists, between downstream market size and upstream (basic) scientific research remains elusive.

In this paper, we fill this gap and use the introduction of Medicare Part D to examine the effects of market size on science. For this purpose, we use novel data on scientific publications, patent-paper links, and drug development efforts mapped to disease categories. Moreover, while we use similar measures of disease exposure to Medicare Part D (*i.e.*, Medicare market share) as the extant literature, the use of scientific publications necessitates a categorization of research at the disease level, as opposed to the therapeutic class level.[5] To demonstrate that our data construction choice does not bias the results, we start by replicating the main findings of the prior literature exploring the impact of the introduction of Medicare Part D on drug development (*e.g.*, Blume-Kohout and Sood, 2013; Dranove et al., 2020).

We make several important contributions to the literature. First, over the decade following the implementation of Medicare Part D, we find no evidence for a causal relationship between market size and research. An increase of one standard deviation in the exposure to Medicare Part D leads to an insignificant increase in scientific publications by only 6.9%. This is substantially smaller than any effect on drug 'development' activities found in the prior literature and in our replication (+20.2%). These findings support Finkelstein's (2004) assertion that the link between market size and increases in drug development is driven by a reordering of products already 'on the shelf'. This is also consistent with Dranove et al. (2020), who show that the upsurge in development is driven by clinical trials of less scientifically novel drugs.

Second, there exists extensive literature on scientist motivations. For example, scientists respond to external funding or rewards (*e.g.*, Cohen et al., 2020; Foray and Lissoni, 2010; Hvide and Jones, 2018; Thursby and Thursby, 2011), altruism and prestige (Stern, 2004), recognition by the scientific community (Stephan, 2012), research opportunities and academic freedom (Aghion et al., 2008), public funding (Azoulay et al., 2019; Myers, 2020), and the desire to work on topics useful for society (Merton, 1973). It also appears that some scientists are motivated by monetary incentives (Levin and Stephan, 1991; Stephan, 1996), incentivizing

[4]Academic medical centers are hospitals that are linked to medical schools and engage in clinical trials.

[5]In the course of the analysis, we account for demographic changes, public research funding, and new research opportunities.

them to engage in patenting (Lach and Schankerman, 2008; Owen-Smith and Powell, 2001; Thursby et al., 2001). Our findings suggest that these upstream motivations are disconnected from the direct link to the opportunities created by shifts in downstream demand in the form of market size changes.

Third, the degree of product market orientation depends on the *type* of research affiliation. For example, on one end of the spectrum, the objectives of corporate scientists will be aligned with their firms, while on the other end of the spectrum, scientists at the National Institutes of Health (NIH) may be more interested in basic science research. To explore this variation, we categorize research activities by four different types of affiliation: the NIH, universities, academic medical centers, and corporations. Consistent with our core findings, we illustrate substantial differences in semi-elasticities. A statistically significant increase can only be found in corporate research (+22.7% in the tenth year after the Medicare Part-D introduction) and decreases in magnitude by distance to the market (*e.g.*, universities show an insignificant demand response of only +5.8%).

Fourth, we further refine our analysis by also focusing on the type of research (*e.g.*, applied or basic). With this refinement, we find that Medicare Part D primarily caused an increase in corporate affiliated publications linked to both clinical trials and pharmaceutical products, which are residuals from drug development activities (*i.e.*, applied research). The increase disappears for corporate affiliated publications that are more basic, which is broadly consistent with Arora et al. (2018). Similarly, we do not find any causal relationship between the type of research and market size for universities or academic medical centers. The divergence between our results and those of Bhattacharya and Packalen (2011) suggests that the differences in the *types* of downstream demand matter. In their case, research appears to respond to disease prevalence, while in our case, it does not respond to changes in disposable income within those diseases.

Fifth, not all publications are equal so we generate three different measures of impact. Our first measure are publications weighted by journal impact factors. Second, we map publications to patents to approximate whether scientific research was referenced in commercially relevant applications (Marx and Fuegi, 2020). Finally, we weight the number of publications by the patent family size associated with the publication. Overall, results remain robust with our core findings – changes in downstream demand have no impact on upstream research.

Interestingly, there is one exception. In the years directly following the enactment of Medicare Part D, a one standard deviation increase in the exposure to Medicare Part D caused an *immediate but transitory* 15.8 percent increase in corporate-affiliated patent-weighted research.

This is consistent with the idea that corporate publishing is used strategically in commercialization activities such as patenting (Della Malva and Hussinger, 2012). Finally, we conclude with a series of robustness tests that redefine the dependent variable, use alternative calculations of Medicare market size, alternative controls, alternative specifications, alternative event windows, and different aggregation levels. In all cases, our core results hold.

Our work has important implications for firms and policymakers. The pharmaceutical industry is highly dependent upon external technology markets (Higgins and Rodriguez, 2006), with much of that research emanating from universities (Cockburn and Henderson, 2000). While drug development (*i.e.*, drugs in clinical trials) appears to respond to downstream shifts in market demand, our results suggest that upstream research fails to do so. Firms face the prospect that the flow of research may not meet the kinds of development needs they may require. This problem is even more significant given the slow decline in corporate-level basic science research (Arora et al., 2018). This disconnect suggests that a more active role for policymakers may be needed. For example, in their recent work analyzing the innovation response to COVID-19, Agarwal and Gaule (2021) argue that policymakers may want to complement a market expansion with early-stage research incentives.

2 Medicare Modernization Act

In the United States, Medicare is the national health insurance program for the elderly. Prior to 2006, it only covered drugs administered during in-patient hospital stays or at doctor offices, but it did not cover out-patient prescription drugs. In December 2003, Congress passed the Medicare Prescription Drug, Improvement, and Modernization Act (MMA), which implemented the Medicare Part D prescription drug benefit as of January 2006.[6] This coverage is available for U.S. residents age 65 and older who fulfill the eligibility criteria of Medicare Part A and B. In contrast to other Medicare programs, Part D contracts with private companies that are authorized to sell insurance coverage. It is both regulated and subsidized, especially for low-income individuals.[7]

Medicare Part D covers all drugs that are also covered under Medicaid, which is a federal program that assists with medical costs for people with limited income, and that fulfill the following criteria. First, the drug has been approved by the Food and Drug Administration

[6]More details can be found here: https://www.congress.gov/bill/108th-congress/house-bill/1 [last accessed on March 8, 2021].

[7]See for more information: https://www.medicareadvocacy.org/medicare-info/medicare-part-d/ [last accessed on March 8, 2021].

(FDA). Second, it must be available only by prescription. Third, the drug is medically necessary for on-label indications (*i.e.*, this limits off-label usage). Finally, Medicare Part D also covers biological drugs, insulin, smoking cessation drugs, and vaccines. While insurance plans do not have to cover all drugs mentioned above, there are certain 'protected classes' for which most drugs are required to be included (*e.g.*, anti-cancer, anti-convulsant, anti-depressants, anti-psychotic, immuno-suppressant, HIV and AIDS drugs). The program excludes, for example, drugs that may be covered under Medicare Part A or B and over-the-counter drugs.[8]

The implementation of Medicare Part D was one of the most significant recent changes in the U.S. healthcare system. It was projected to benefit 29 million people in 2006 and 44 million people by 2015. The expected total public expenditures in the first 10 years were estimated to be $800 billion.[9] This expenditure corresponded to approximately 0.42% of GDP in 2006, increasing to 0.76% in 2015.[10] The program can be categorized as a demand subsidy.

As expected, Medicare Part D considerably increased prescription drug use by elderly patients. In Figure 1 we illustrate this development. At the extensive margin, drug use by Medicare-insured patients increased substantially after 2006, especially in the quartile of diseases most likely to afflict older patients. This implies that previously uninsured elderly are now able to purchase prescription drugs. The same applies when looking at prescription quantities for both existing and newly insured patients (Appendix Figure A-4). Thus, we can see that the MMA differentially increased the market size for drugs that are developed to treat diseases more prevalent among older individuals. This will be a fundamental aspect of our empirical strategy, which we outline in the next section.

3 Data and Methodology

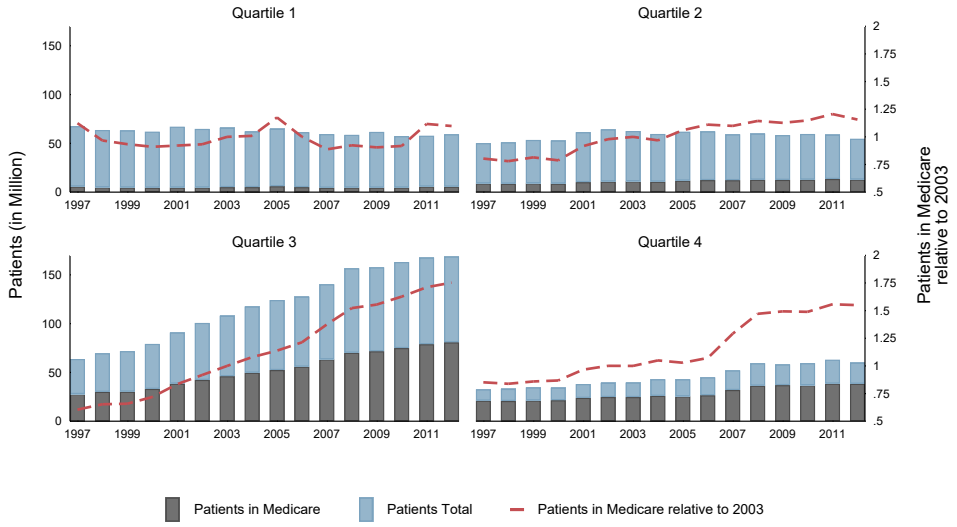
We are interested in the causal effect of Medicare Part D (*i.e.*, changes in downstream market demand) on upstream research. If upstream research responded to demand pull effects in the same way as downstream drug development, we would expect to find an increase in the number of scientific publications, all else equal. Further, we would also expect to see an increase in patent-paper links, which can be viewed as an output of basic science research.

[8] It further excludes drugs for weight loss or gain, cough and cold preparations, fertility, erectile dysfunction, cosmetic and hair growth, as well as vitamins and minerals. For more information, see: https://www.medicareadvocacy.org/medicare-info/medicare-part-d/ [last accessed on March 8, 2021].

[9] Own calculations based on the 2006 Medicare Trustees Report using an annual inflation rate of 5%.

[10] https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/ReportsTrustFunds/downloads/tr2006.pdf [last accessed on March 8, 2021].

Figure 1: Evolution of Medicare beneficiaries



Notes: The figures show the evolution of patient counts of each ICD-9 group aggregated by Medicare market share quartiles. These are patients who received a prescription drug designated to a ICD-9 group disease at least once in a given year. Our exposure measure to Medicare Part D is described in Section 3.3. The grey bars display the number of Medicare patients, the blue bars display the non-Medicare patients. Patients are counted multiple times if they appear in more than one ICD-9 group. MMS quartiles are based on the pre-2004 weighted average of patient-based MMS. The red line represents the relative increase in the number of Medicare patients with respect to the baseline year 2003. The figures depict a discrete increase in Medicare patients after 2006 in the highest age quartile.

3.1 Sample Selection

To create a link between Medicare Part D and R&D activities, we combine data on biomedical scientific publications from the NIH's MEDLINE/PubMed database and Web of Science (WoS), patent information from Patstat, and drug development activities from the Clarivate Analytics' Cortellis Investigational Drugs database by ICD-9 disease categories and their exposure to Medicare Part D based on the Medical Expenditure Panel Survey (MEPS).

However, matching publications to disease categories is not straight forward because the keywords in biomedical publication databases do not correspond one-to-one to standard international disease classifications. We take advantage of an existing crosswalk introduced by Bhattacharya and Packalen (2011), which we update and present in Table A-1. This crosswalk provides a mapping of Medical Subject Headings (MESH) terms with a range of ICD-9 three-digit codes.[11] Since some MeSH terms relate to multiple ICD-9 three-digit codes and

[11]MeSH is a hierarchical medical vocabulary administered by the NLM and consists of approximately

vice versa, MeSH and ICD-9 three-digit codes are grouped at the level of mutually exclusive ICD-9 groups that comprise similar diseases.[12] The crosswalk is not exhaustive. It excludes the majority of ICD-9 three-digit categories that comprise the words 'other' or 'unspecified' since those categories typically include various very distinct diseases.[13]

Our final sample includes 129 separate matches at the ICD-9 group level, which corresponds to 272 unique ICD-9 three-digit codes and 192 high-level MeSH terms.[14] We take advantage of the stringent MeSH hierarchy to extend the initial set of MeSH terms to all synonyms and lower-level terms. Ultimately, our sample contains 1,563 MeSH terms. Our independent variables are calculated based on the selected mutually exclusive ICD-9 groups.

3.2 Empirical Strategy

We illustrate our empirical strategy through a case study (Figure 2). We select two disease categories from our panel, one with a very high Medicare market share (MMS), *Alzheimer's disease*, and one with a very low MMS, *inflammatory skin diseases*. We show that the trends in the number of scientific publications and number of preclinical trials (drug discoveries) related to either Alzheimer's or inflammatory skin diseases are parallel before the passage of Medicare Part D in 2003. While the low-MMS inflammatory skin diseases seem to be uncorrelated with the MMA in the science and technology sphere, the high-MMS Alzheimer's disease increases disproportionately from 20 to 60 drug candidates (*i.e.*, drug development). This, however, does not occur with publications (*i.e.*, research), where the trends remain parallel after 2003.

In our multivariate analyses, we exploit the passage of the MMA by using the variation in exposure measured by the 1997-2003 MMS. To this end, we use a panel data model that

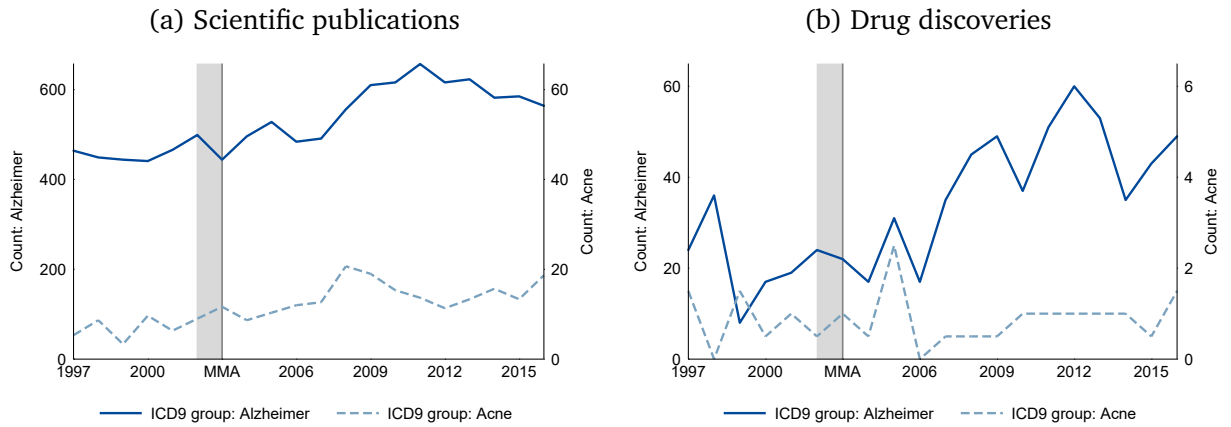
30.000 different terms in 2020 (https://meshb.nlm.nih.gov/search) [last accessed on March 8, 2021]. The International Statistical Classification of Diseases and Related Health Problems (ICD) is administered by the World Health Organization. The 9th version (ICD-9) comprises around 13,000 codes (https://www.cdc.gov/nchs/icd/icd9cm.htm) [last accessed on March 8, 2021].

[12] There exists no official publicly administrated crosswalk between ICD-9 and MeSH terms. NIH resources like the UMLS Metathesaurus (https://www.ncbi.nlm.nih.gov/books/NBK9684/ [last accessed on March 8, 2021]) only include selected 1-to-1 matches. The usage of the PubMed search algorithm, which searches for terms in titles and abstracts, is infeasible because it requires the user to search for all possible synonyms. Furthermore, identifying appropriate MeSH terms for each ICD-9 category using the MeSH on demand algorithm (https://meshb.nlm.nih.gov/MeSHonDemand [last accessed on March 8, 2021]) requires an expert assessment in case of multiple results.

[13] The match is further limited to ICD-9 codes that appear more than 100 times in the MEPS data from 2007. It excludes the following ICD-9 categories: pregnancy (class 11), congenital (class 14), perinatal (class 15), symptoms (class 16), injuries (class 17), and services (class V).

[14] In addition to the 127 disease categories in Bhattacharya and Packalen (2011), we include two major diseases that they excluded: HIV and Alzheimer's disease. Our results are not sensitive to their inclusion.

Figure 2: Trends in R&D activities in Alzheimer's and inflammatory skin diseases



Notes: The left-hand Figure presents the number of publication counts, the right-hand Figure the number of newly discovered new molecular entities for the Alzheimer's disease (ICD-9 code: 331; MMS is 97%) versus inflammatory skin diseases like Acne (ICD-9 codes: 690/706; MMS is 4%) between 1997 and 2016.

observes each disease-related MeSH term by year. The MeSH terms m are nested within ICD-9 disease groups $i(m)$ (defined as i in the following), which in turn constitute the level of treatment exposure. To capture the dynamics of the effect, we amend the standard Difference-in-Differences framework by replacing the post-period dummy with three-year binned sets of leads from 2004 onward. Hence, we compare research-related outcome variables before and after the introduction of Medicare Part D between more and less affected MeSH terms. Our empirical model can be written as follows:

$$\begin{aligned}\mathbf{E}[N_{mt}|\, X_{it}] \;=\; \exp[\alpha + \sum_{t=2004}^{2016} \beta^t \; Medicare\, Market\, Share_i \times \; \mathbf{1}_{\{\delta_t = t\}} \\ + \mu \left(\sum_{\text{lead}=1}^{\overline{f}} M_{i,t+lead} \right) + \lambda \left(\sum_{\text{lag}=1}^{\overline{l}} NIH_{i,t\text{-}lag} \right) + \gamma \left(\sum_{\text{t}=0}^{\overline{t}} K_{i,t} \right) \quad + \delta_t \; + \; \theta_m],\end{aligned} \tag{1}$$

where N_{mt} represents the dependent variable (*e.g.*, the number of publications per MeSH term in year t). The interaction terms $MMS_i \times \mathbf{1}_{\{\delta_t = t\}}$ indicate the exposure to the MMA and whether we are in the pre- or in the post-MMA periods (*i.e.*, $t \geq 2004$). Consistent with prior literature on market size and R&D, our empirical model includes controls for the future demographic driven market size, M_{it}, past public research funding, NIH_{it}, and research opportunities, K_{it}. Given the differences in the level of R&D activities between diseases, we include MeSH term fixed effects, θ_m. The inclusion of this large set of fixed effects (>1,500) reduces our sample of

analysis: 29 MeSH terms are dropped from our main specification due to lack of variation in the dependent variable.[15] We show the robustness of our results to the full sample by substituting the MeSH term fixed effects with the smaller set of ICD-9 group fixed effects, θ_i. Furthermore, we control for time trends using calendar year effects, δ_t. In alternative specifications, we interact the MMS with a full set of two-year binned leads and lags, $\sum_{t=1997}^{2016} \beta^t MMS_i \times \mathbf{1}_{\{\delta_t=t\}}$.[16] In this setting, we normalize the coefficient $\beta^{t=2002/3}$ to zero and express the dynamic treatment effects relative to this pre-treatment year. This will also allow us to examine whether the parallel trends assumption holds.

We estimate the relative change of R&D activity using Poisson pseudo-maximum likelihood regressions. Since our dependent variables are count data, Poisson is the preferred econometric model in panels (Hausman et al., 1984). Under our identification assumption, β^t gives us the average causal effect of the MMA in year t. The coefficients can be interpreted as semi-elasticities and we cluster standard errors at the MeSH term level.[17]

When replicating prior findings relating the effect of Part D on drug development, we turn to an analysis at the ICD-9 three-digit code level, N_{ct}. This is the most fine-grained level of observation for our clinical trial data. ICD-9 three-digit codes, c, are again nested in ICD-9 groups $i(c)$. In these specifications, we include ICD-9 three-digit code fixed effects, θ_c (as a robustness check: ICD-9 group fixed effects, θ_i), while all other parts of Equation 1 remain the same.[18]

Lastly, we investigate other variation by splitting the dependent variable along several categories, such as affiliation, journal type, clinical relevance, and funding source. We use this battery of dependent variables to explore treatment effects along the entire scientific and innovation value chain.

3.3 Medicare Market Share

We exploit variation across disease categories in their pre-Medicare Part D market share (MMS), expecting larger increases in scientific research in disease categories with higher MMS. Following prior studies on Medicare Part D (*e.g.*, Blume-Kohout and Sood, 2013; Dranove et al., 2020;

[15] When investigating sub-samples of the outcome variable, in some cases a larger number of MeSH terms are dropped due to lack of variation.

[16] Our post-treatment period overlaps with the passage of the Affordable Care Act (ACA) in 2010. However, this is not a concern because ACA involved only low reimbursement for pharmaceuticals and, thus, small revenue increases and incentives from market size (Garthwaite et al., 2021).

[17] A one percentage point higher MMS_i leads to a change of the dependent variable N_{mt} of $\beta^t \times 100$ percent.

[18] Again, some ICD-9 three-digit codes are dropped from our main specification due to lack of variation in the dependent variable.

Duggan and Scott Morton, 2010; Hermosilla and Wu, 2018; Krieger et al., 2018), we build a measure of a disease category's exposure to Medicare Part D based on MEPS.[19] Unlike most of these studies, we use ICD-9 disease categories as our level of observations and not therapeutic classes because scientific articles are typically indexed by keywords corresponding to diseases. In MEPS, each patient-level drug prescription is associated with a designated disease, an indicator of whether the beneficiary was insured by Medicare, and the patient's age. Using this data we calculate the MMS for each disease at the ICD-9 three-digit level and the more aggregated ICD-9 group level. The latter corresponds with the level of our match between ICD-9 and MeSH terms, and, thus, with the level of our empirical analysis on biomedical science.

It is plausible that researchers could anticipate the market size increase since the authorization of the MMA in December 2003. Thus, we calculate the fraction of patient counts, prescription counts, and quantity weighted prescription counts filed by Medicare-covered individuals compared to all individuals for each disease category as a weighted-average between 1997 and 2003.[20] The categories with the highest MMS are *Alzheimer, Retinal Diseases, Cataract, and Angina Pectoris*. Among the diseases with the lowest MMS are *Hyperkinetic Syndrome of Childhood, Scarlet Fever, Infantile Cerebral Palsy, and Inflammatory Skin Diseases*. The distribution of our MMS at the ICD-9 group level is presented in Figure 3, has a mean of 32%, and is in line with other studies (*e.g.*, Dranove et al., 2020; Hermosilla and Wu, 2018). Critically, it does not vary with the ICD-9 level, sample selection, or the use of (quantity weighted) prescription counts (Appendix Figure A-3). We are confident in our sample of 129 ICD-9 disease groups as our level of analysis since we are able to replicate the main results of the prior literature.[21]

3.4 Dependent Variables

Scientific Publications

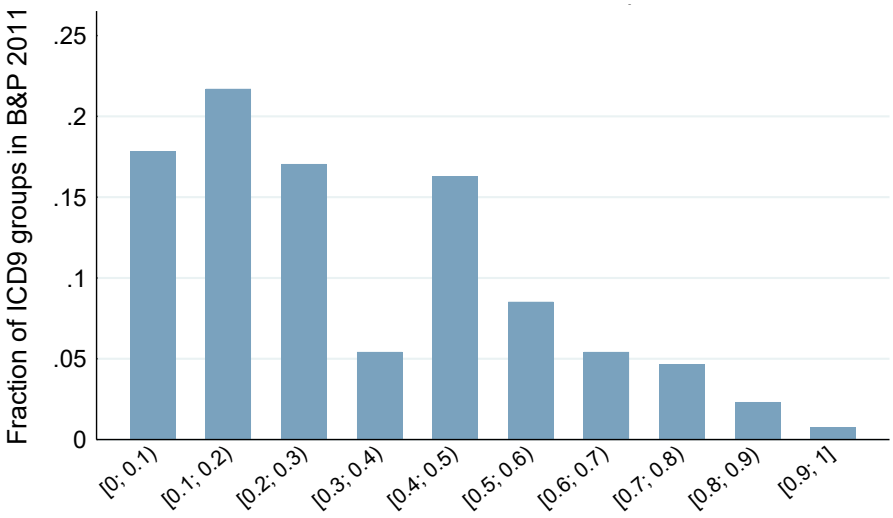
NIH's MEDLINE/Pubmed database includes the entire universe of references to journal articles in the biomedical sciences from the early 20th century to the present. We retrieve all publica-

[19] MEPS data is available here: https://www.meps.ahrq.gov/mepsweb [last accessed on March 8, 2021].

[20] We weight each survey respondent in MEPS by their representativeness, thereby creating a person-level sampling weight.

[21] Our disease level data allows us to replicate the development of quantities, drug prices, and revenues. We document the positive effect of Medicare Part D on drug consumption (Appendix Figure A-4). Moreover, in line with Duggan and Scott Morton (2010) and Duggan and Scott Morton (2011), drug prices decrease between 2006 and 2009 since patients were able to switch to cheaper insurance plans. However, prices increased after 2009 (holding the 2003-2005 drug basket for each disease constant). Despite the initial price declines, by 2006 revenues increased disproportionately for high MMS diseases. This suggests that the quantity increase outweighs the initial price decline (Appendix Figure A-5).

Figure 3: Distribution of Medicare market shares (1997-2003)



Notes: The figure presents the distribution of MMS scores among ICD-9 groups that are included in the MeSH-ICD-9 crosswalk by Bhattacharya and Packalen (2011). We use the patient-weighted average of each year between 1997-2003. The annual MMS are calculated using the total number of patients in Medicare relative to all patients for each ICD-9 group.

tions (also referred to as PMIDs) with at least one of the 1,563 MeSH terms linked to the 129 ICD-9 groups. We then restrict our sample to U.S. publications. Next, we match these publications with bibliographic data from Web of Science (WoS) to take advantage of WoS's proper author name and affiliation disambiguation. The WoS data enables us to look at various splits in the publication data, such as the affiliation type (*i.e.*, NIH, university, corporate, academic medical centers),[22] and appliedness of the journal.[23] Moreover, the WoS bibliographic information allows us to add information regarding forward citations and journal impact factors.

The coverage of PMIDs in WoS is high, which gives confidence in capturing all relevant papers related to medical science. Moreover, we extract for each paper all indexed MeSH terms from the Pubmed database to classify whether a publication is related to disease terms that are not in our sample. Since we do not know the exposure of these additional disease-related MeSH to the MMA, we treat them as potentially confounding and cautiously drop them from the sample. Beyond that, we use all indexed MeSH terms to classify whether a publication

[22] We infer the affiliation type from the disambiguated Web of Science publication data based on the string name of each affiliation. Academic medical center are identified using the string 'hospital'. Corporate affiliations are identified using legal forms like 'Corp.' or 'Inc.'.

[23] We use a classification of journals based on the proportion of published research coming from a general hospital and industry using the publicly available data set provided by Tijssen (2010). For more information, see: https://www.vosviewer.com/journal-application-domain-map [last accessed on March 8, 2021].

is related to clinical trials or pharmaceutical products.[24,25]

Next, we measure the extent of the research efforts related to a disease group by counting the number of matched scientific publications at the MeSH term level. Publications might be associated with more than one ICD-9 group, so that we account for this in two alternative ways: we treat them either as simple counts for each disease group separately or weight them by the inverse number of linked diseases that sum up to one across all disease groups. The resulting final data set spans from 1997 to 2016 and includes 449,996 unique publications.

Patents

We use references in patents to the scientific non-patent literature (SNPL) to identify knowledge diffusion between upstream research and more downstream innovation activities, like patenting (*e.g.*, Ahmadpoor and Jones, 2017; Marx and Fuegi, 2020; Poege et al., 2019; Watzinger and Schnitzer, 2019).[26] We locate publications that are directly cited in a patent by matching the Pubmed-patent link constructed by Marx and Fuegi (2020) to our sample of scientific publications. Thus, we weight U.S. scientific publications by an indicator variable that specifies whether a publication was cited as a SNPL within a 5-year window.[27] This applies to 19,891 biomedical scientific publications.[28]

Clinical Drug Development

To measure the impact of Medicare Part D on drug development activities, we use time-series data from Cortellis on all clinical drug development events by disease categories at each stage in the pharmaceutical development process. We link the Cortellis *targeted conditions* to ICD-9

[24]The category 'clinical trials' includes all MeSH terms that are related to the MeSH ID 'D016430' (Clinical Trial) and the entire set of MeSH terms at the hierarchy levels below, such as 'Adaptive', 'Phase I', 'Phase II', 'Phase III', 'Phase IV', 'Controlled Clinical Trial' or 'Randomized Controlled Trial'.

[25]The category 'pharmaceutical products' includes all MeSH terms that are related to the MeSH ID 'D004364' (Pharmaceutical Preparations) and the entire set of MeSH terms at the hierarchy levels below, such as 'Dosage Forms', 'Drug Combinations', 'Drugs, Generic', 'Drugs, Investigational', 'Pharmaceutic Aids' or 'Prescription Drugs'.

[26]Patents reference various types of documents that relate to the protected invention by either determining novelty (prior art) or explaining the content of the underlying invention. A subset of these references relates to scientific articles, called SNPL references (Poege et al., 2019).

[27]We aggregate all citing patent applications at the DOCDB family level and calculate the 5-year window from the year of the scientific publication to the year of the priority year of the patent family.

[28]The publication-patent link may suffer from attrition because late publications have not yet been cited in patents. However, we have little reason to expect that the time to patent varies systematically by MMS. This is supported by our finding that the minimum time lag between a scientific article's publication year and a patent's priority year is uncorrelated with the MMS (*i.e.*, the pairwise correlation is -0.0365).

codes using the crosswalk by Dranove et al. (2020) and identify unique new molecular entities (NME) entering Phase I, Phase II, or Phase III clinical testing, as well as being submitted to the FDA for approval.[29] The data comprises 201 ICD-9 three-digit codes and 121 ICD-9 groups. We then limit the sample to NMEs that are discovered/tested/approved in the U.S., and include information on the target-based action and at least one designated disease. If a clinical trial occurs more than once for a NME-disease link, we use the first event. To be consistent with our science level analysis, we count the number of NMEs at the more fine-grained ICD-9 three-digit code level, which is nested within our ICD-9 groups. In total, we identify 9,943 NMEs entering at least one phase of the drug development process.

3.5 Control Variables

The empirical model includes a set of additional determinants of R&D: projected market size, public research funding, and research opportunities. A detailed description of the control variables construction can be found in the Appendix A. First, we control for the projected market size due to U.S. population growth (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013). For this purpose, we use demographic (projection) data from the UN World Population Prospects. Figure 4a illustrates the average development of the population-growth driven projected market size before and after the introduction of Medicare Part D. While market size will increase in all disease categories, diseases in the highest quartile of MMS exhibit the sharpest growth due to an aging population.[30]

Second, we control for previous years' public research funding related to each disease category. Therefore, we calculate for each of our 129 ICD-9 groups the exposure to the NIH budget over time by linking each group to the relevant NIH Institute/Center (*e.g.*, ICD-9 162 *malignant neoplasm of trachea, bronchus, and lung* to the *National Cancer Institute*).[31] Figure 4b presents the normalized average NIH spending by ICD-9 groups in each MMS quartile. It becomes apparent that diseases in the lowest quartile of MMS are associated with the largest relative increase of NIH funding. However, high MMS diseases receive a substantially higher level of NIH funding.[32]

[29] Dranove et al. (2020) had two expert medical coders independently code the concordance between Cortellis indications and ICD-9 codes.

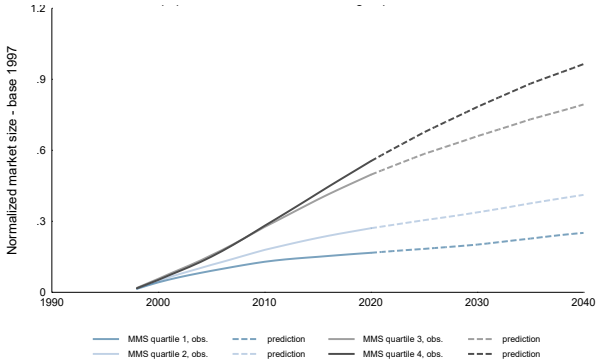
[30] We also calculate projected market size at the OECD level. Figure A-1 illustrates the evolution.

[31] Since grants are distributed within Institutes primarily by scientific merit (see discussion on NIH funding rules by Azoulay et al. (2019)) and not by allocation to narrower disease categories, we attribute the full Institute's budget to each ICD-9 group. We retrieve NIH spending data (Mechanism Detail by IC, FY 1983-2019) from https://officeofbudget.od.nih.gov/spending_hist.html [downloaded on February 17, 2020].

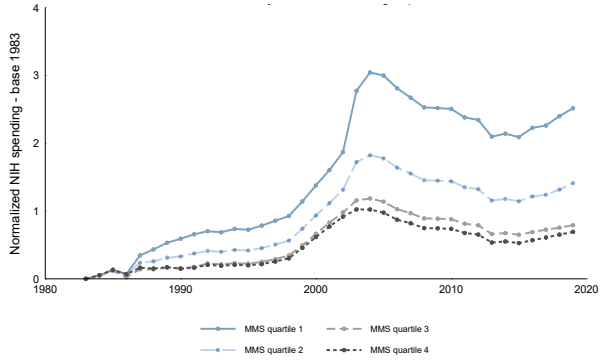
[32] In an alternative approach we attribute budgets based on the share of all publications in a disease category that

Figure 4: Control variables over time

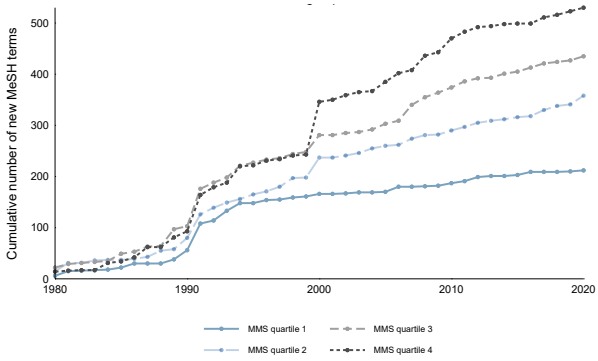
(a) Projected market size



(b) Average NIH funding (normalized)



(c) New MeSH terms



Notes: The three figures present the normalized annual control variables by MMS quartiles over time. In Figure 4a, we aggregate the U.S. population-growth driven market size (in 2003 values) of each ICD-9 group. In Figure 4b, we average the NIH spending (in 2003 values) of all Institutes/Centers, which are related to our ICD-9 groups. In Figure 4c, we accumulate the number of new MESH terms associated with our ICD-9 groups.

Third, we account for the *availability of research opportunities*.[33] We develop a direct measure of new research opportunities in a disease area based on the introduction of new terms in the respective branch of the MeSH tree. New terms are added for emerging diseases, breakdowns of existing diseases, and additional terminology to reflect topical areas that are not well represented in MeSH. Figure 4c shows substantial heterogeneity in new research opportunities across ICD-9 groups. High MMS diseases exhibit greater increases in research opportunities around 2000.

acknowledge a specific Institute. Appendix Figure A-2 illustrates the evolution.

[33] In similar fashion, Bhattacharya and Packalen (2011) construct a measure of research opportunities based on the content of research inputs and the first appearance of the idea in a scientific publication. The disadvantage of the approach is that it relies on a very narrow set of research inputs that relate primarily to drug-related medical research but not basic science.

4 Descriptive Analysis

Our final data set consists of 129 disease groups from 1997 to 2016. This allows us to investigate the possible effects of Medicare Part D on scientific research over a period of 13 post-MMA years. Each of the 129 ICD-9 groups in our sample is associated with, on average, 12.1 disease MeSH terms. For 121 of these ICD-9 groups, we find drug development activities in around 1.7 ICD-9 three-digit codes per group.

Table 1 provides summary statistics for the full set of independent and dependent variables in the year 2003. The average MeSH term has 16.86 scientific publications, which totals around 204 per ICD-9 group. The majority of publications (77.8%) have at least one author with a university affiliation, 13.6% have at least one author with an academic medical center affiliation, 5.4% have a corporate affiliation, and 4.9% are published with NIH participation. Moreover, 6.2% of these publications are cited in patent applications. This share, however, is substantially larger for corporate publications (14.3%). At the same time, 2.03 NMEs enter pre-clinical trials per ICD-9 three-digit code. This number decreases throughout the clinical trial process resulting in 0.09 new drug approvals per ICD-9 group in 2003.

A simple comparison between ICD-9 groups split at the MMS median illustrates that both groups are very similar regarding the pre-MMA levels in the majority of dependent variables (Appendix Table A-2). This also applies to the distribution of dependent variables (Appendix Figure A-6). An important exception is that diseases prevalent among the elderly are more related to clinically relevant journals and patents. This is not surprising since the overall market size is larger, in levels, before the MMA. Moreover, the MMS is positively correlated with the level of all independent variables. This supports our decision to control for these factors in our multivariate analysis.

Descriptively, the total yearly log-transformed number of scientific publications in low or high MMS ICD-9 disease groups develops in a parallel fashion until 2003. This provides visual support for our parallel trends assumption of our Difference-in-Differences framework (see Figure 5a).[34] These trends also hold across other dependent variables, like corporate publications (see Figure 5b).[35] After 2004, the log number of publications associated with high-MMS diseases increase disproportionately, but only to a small degree. This divergence is more pronounced for publications from corporate affiliates.

Parallel trends are also supported by the annual count of drug discoveries and approvals

[34] Formal tests for parallel trends, *e.g.*, by splitting the pre-period and testing whether there are differential changes in the slope, are employed and found supportive (Appendix Tables A-3 and A-7).

[35] Further univariate graphs with other dependent variables can be found in Appendix Figure A-7.

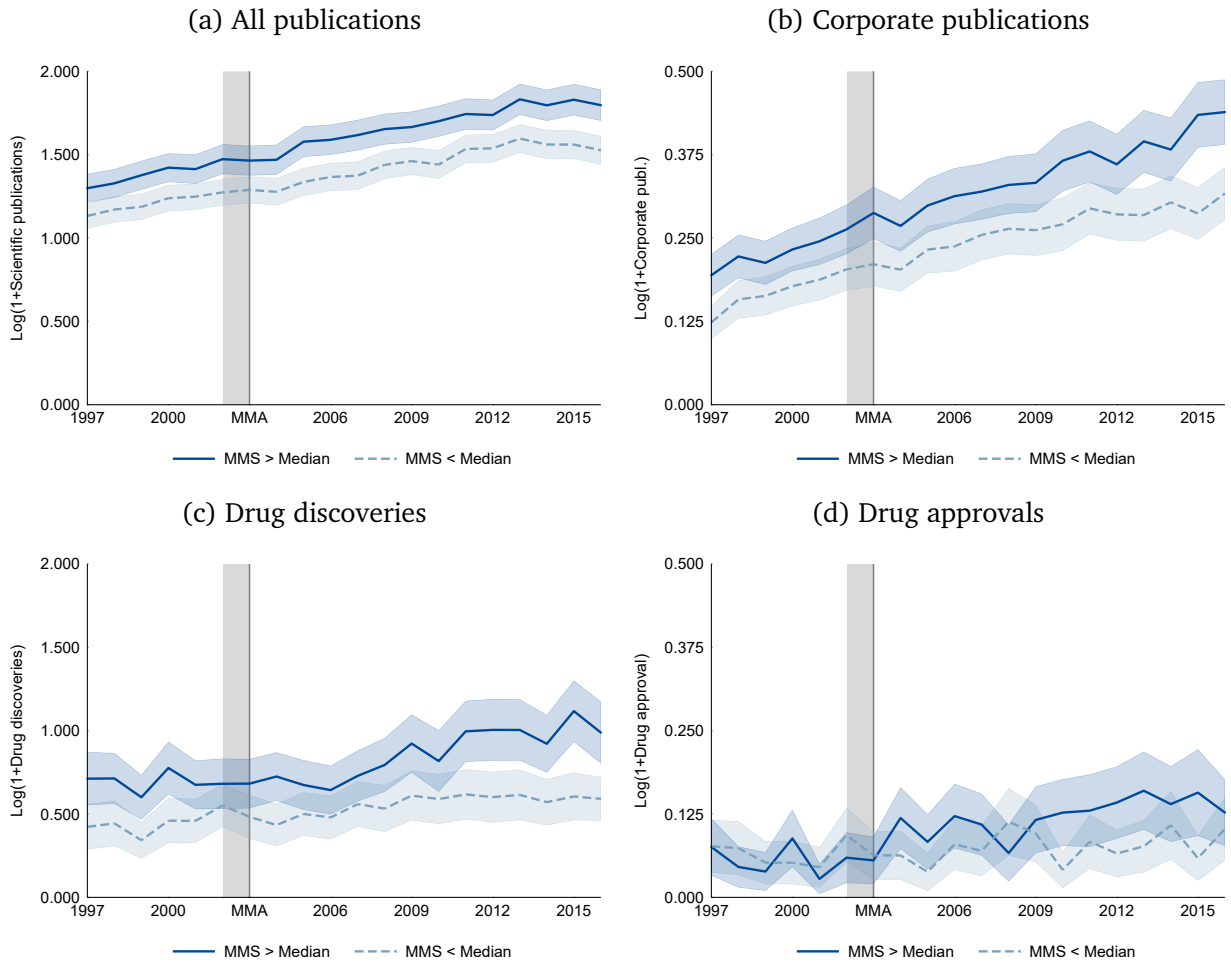
Table 1: Summary statistics

	N	Mean	Median	Std. Dev.	Min	Max
ICD9 group level						
MMS (cases)	129	31.81	27.32	23.22	0	97
MMS (prescription counts)	129	34.42	30.21	24.53	0	98
MMS (prescription quantity)	129	35.18	29.28	25.31	0	97
Cumul. US Market Size$_{t \text{ to } t+12}$	129	13575.14	2695.54	40242.85	29	343211
Cumul. NIH funding$_{t-1 \text{ to } -12}$	129	16.72	14.38	10.97	3	46
Cumul. New MeSH terms$_{t}$	129	0.23	0.00	0.70	0	4
MeSH term level						
Scientific publications	1563	16.86	2.00	69.33	0	1128
Publications - fractional	1563	13.11	1.00	58.95	0	1000
NIH publications	1563	0.82	0.00	4.14	0	70
University publications	1563	13.28	1.00	54.87	0	890
Hospital publications	1563	2.29	0.00	8.63	0	151
Corporate publications	1563	0.91	0.00	4.41	0	80
NIH funded publications	1563	6.67	0.00	32.87	0	565
Clinical trial university publications	1563	1.27	0.00	5.75	0	110
Clinical trial corporate publications	1563	0.21	0.00	1.16	0	25
Pharmaceutical university publications	1563	0.33	0.00	1.74	0	27
Pharmaceutical corporate publications	1563	0.07	0.00	0.46	0	9
Citation-weighted publications	1563	353.33	16.00	1646.59	0	29351
JIF-weighted publications	1563	60.88	4.28	266.30	0	4518
Patent-weighted publications	1563	1.05	0.00	6.26	0	125
Patent-weighted university publications	1563	0.83	0.00	4.99	0	102
Patent-weighted corporate publications	1563	0.13	0.00	0.75	0	14
Patent family size-weighted publications	1563	9.17	0.00	53.45	0	1082
ICD9 3-digit code level						
Drug discoveries	201	2.03	0.00	5.43	0	47
Phase 1 clinical trials	201	0.54	0.00	1.32	0	8
Phase 2 clinical trials	201	0.82	0.00	1.97	0	17
Phase 3 clinical trials	201	0.26	0.00	0.67	0	4
Drug approval	201	0.09	0.00	0.37	0	3

Notes: This table presents summary statistics linked to the 129 ICD-9 groups in 2003. The unit of observation is at the ICD-9 group level for both the treatment and control variables; at the MeSH term level for the dependent publication variables; and at the ICD-9 three-digit code level for dependent drug development variables. Some MeSH terms and ICD-9 three-digit codes lack variation in the dependent variable so that the corresponding observations are dropped from the estimations.

in Figures 5c and 5d, which evolve in a similar fashion for low- and high-MMS diseases until 2003. The number of drug discoveries and drug approvals increases after 2003. As such, we can replicate the prior literature (Blume-Kohout and Sood, 2013; Dranove et al., 2020) descriptively within our sample of ICD-9 groups.

Figure 5: Trends in scientific publications and drug development by MMS



Notes: Figure 5a presents the log-transformed average number of annual publication counts associated with all affiliations and Figure 5b selects only publications from corporate affiliations. Figure 5c displays the log-transformed average number of annual drug discoveries and Figure 5d the log-transformed average number of annual drug approval. For reasons of comparability, the unit of observation is the unique ICD-9 group level in all four graphs.

This descriptive analysis suggests that the introduction of Medicare Part D and, thus, the sudden insurance-induced increase of market size for diseases more prevalent among the elderly, has led to more commercial drug development activities. Upstream research activities, in general, seem to be more resilient to these changes in downstream market size. However, upstream research conducted by corporations appears more elastic in high MMS disease categories after the MMA.

The following multivariate analysis will investigate this pattern in more detail, accounting

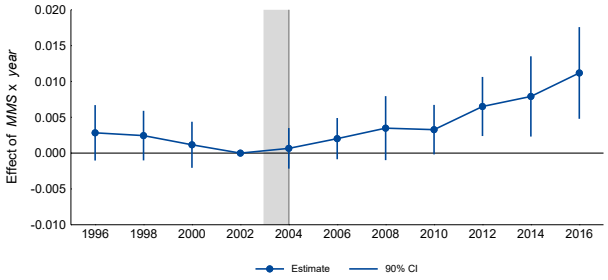
for other factors like demographic trends, public funding, and new research opportunities that may have an impact on R&D outcomes besides Medicare Part D.

5 Empirical Results – Clinical Drug Development

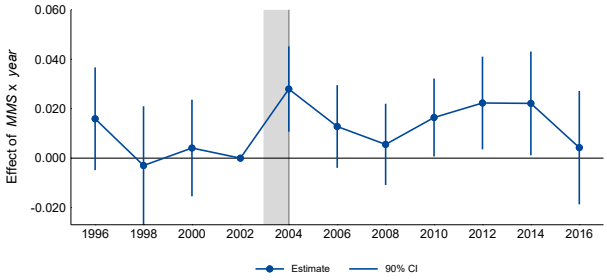
We start by replicating prior results showing the effect of the MMA on clinical drug development. This replication exercise provides validation for our sample selection and variable construction. Moreover, the results of this analysis will enable us to compare the effects among scientific research and drug development activities within our sample. Figure 6 shows the event study results similar to Equation 1. The dependent variables are the number of newly discovered NMEs and the number of drug approvals, respectively.

Figure 6: Event study – drug development

(a) Drug discoveries



(b) Drug approvals



Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the ICD-9 three-digit code level, with MMS being calculated based on patient counts at the ICD-9 group level. The sample includes all ICD-9 groups appearing in the Bhattacharya and Packalen (2011) crosswalk. Standard errors are clustered at the ICD-9 three-digit code level.

In line with our descriptive analysis, the number of new drug discoveries and the number of drug approvals display broadly similar patterns independent of the pre-MMA MMS. This suggests the absence of confounding pre-trends. After the passage of the MMA in 2003, disease categories with higher exposure to Medicare Part D exhibit a relative increase in drug discoveries. The effect becomes larger over time and is statistically significant. The same holds true for drug approvals. However, there is a significant up-tick directly after the introduction of Medicare Part D in 2004 and 2005.[36]

[36]We show the event study results of other drug related outcome variables in Appendix Figure A-8.

Next, Table 2 presents the dynamic treatment effect of our Difference-in-Differences estimation in the post-period. Columns 1 and 2 present the results on drug discoveries with and without control variables. In Columns 3 and 4 we present NMEs in all stages of the clinical development process. Finally, Columns 5 and 6 present clinical trials in Phase I-III and drug approvals separately. We find a positive and significant effect of a higher MMS on early drug development, accelerating over time. In our preferred specification, in which we control for the future market size, past NIH funding, and research opportunities, the effect becomes significant after 2011. This is consistent with the long discovery process in the pharmaceutical industry. The point estimate in 2015-2016 has a magnitude of 0.87%. This implies that one standard deviation (23.2 percentage points) increase in MMS leads to 20.2% more drug discoveries. These effects are similar in magnitude compared to Dranove et al. (2020) and slightly smaller than the results of Blume-Kohout and Sood (2013).

The effects are smaller and occur later when looking at drug development in all stages of clinical development or NMEs in Phase I-III clinical trials. However, this is not surprising given the staggered fashion of drug development (from preclinical to clinical trials to approval) and supported by other studies on drug development (Blume-Kohout and Sood, 2013). An exception are drug approvals, which show a positive significant increase immediately after the introduction of the MMA in 2004-2005 by 2.5% per additional percentage point of MMS. These results strongly support Finkelstein (2004) and suggest that pharmaceutical companies reacted by pushing forward advanced drug candidates already in their development pipeline.

6 Empirical Results – Biomedical Research

In this section, we evaluate whether Medicare Part D affected scientific research (measured by publications) for disease categories with higher MMS. Importantly, we differentiate by the type of affiliation as well as type of research and investigate the (commercial) impact.

6.1 Main Effect

First, we evaluate whether Medicare Part D differentially affected scientific research in MeSH categories associated with high MMS ICD-9 disease groups from all types of affiliations. Our event study results in Figure 7 are based on a Poisson pseudo-maximum likelihood regression with the full set of control variables, MeSH term and calendar year fixed effects (adapting Equation 1). Overall, we see no pre-MMA effect on scientific publications suggesting the ab-

Table 2: Drug development

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Drug Development					
	Early Development		All Development Stages		Phase 1-3	Approval
MMS × 2004-05	−0.0011	−0.0007	−0.0011	−0.0007	−0.0044*	0.0256***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.009)
MMS × 2006-08	0.0006	0.0013	0.0000	0.0007	−0.0017	0.0070
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.008)
MMS × 2009-11	0.0009	0.0020	0.0001	0.0012	−0.0004	0.0138
	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)	(0.009)
MMS × 2012-14	0.0035	0.0053*	0.0024	0.0040	0.0016	0.0198*
	(0.002)	(0.003)	(0.002)	(0.003)	(0.004)	(0.011)
MMS × 2015-16	0.0065**	0.0087**	0.0046	0.0066*	0.0043	0.0191
	(0.003)	(0.004)	(0.003)	(0.004)	(0.006)	(0.012)
Cumul. US Market Size$_{t \text{ to } t+12}$	No	Yes	No	Yes	Yes	Yes
Cumul. NIH funding$_{t-1 \text{ to } -12}$	No	Yes	No	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	No	Yes	No	Yes	Yes	Yes
ICD9 code FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3340	3340	3740	3740	3360	2380
ICD9-codes	167	167	187	187	168	119
ICD9-groups	110	110	114	114	110	87
Log-likelihood	−4811	−4801	−6644	−6634	−4353	−1184

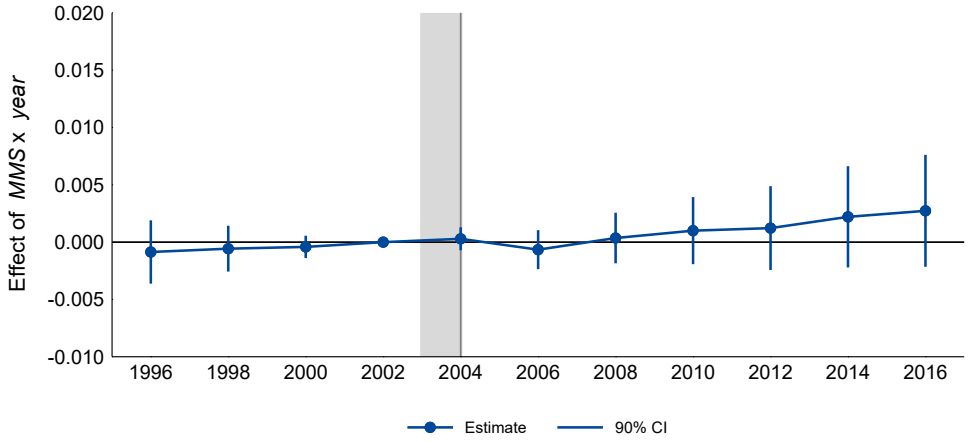
Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the ICD-9 three-digit code by year. The dependent variable is the annual number of newly discovered NMEs in Columns (1) and (2), NMEs in all clinical development stages (*i.e.*, preclinical, clinical trials, registrations, approvals) in Columns (3) and (4), NMEs in Phase I-III clinical trials in Column (5), and approved NMEs in Column (6). The control variables are log-transformed. Standard errors are clustered at the ICD-9 three-digit code level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

sence of confounding pre-trends. After the passage of Medicare Part D, we observe only a slight divergence for the average MeSH category, which is neither significantly different from zero nor large in magnitude.

We quantify the changes in scientific research in response to the Part D demand shock in Table 3. In Columns 1 and 2, we report the results from a simple Difference-in-Differences regression. The post-treatment period is defined to start in 2004 and to last until the end of our sample's observation period in 2016. Again, there is no significant effect on scientific publications, independent of the usage of control variables.

In Columns 3 to 6, we estimate the dynamic changes in science. The first time period shows the transitional effect between the passage and the implementation of Medicare Part D. The

Figure 7: Event study – scientific publications



Notes: The figure shows the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the MeSH term level, with MMS being calculated based on patient counts at the ICD-9 group level. Standard errors are clustered at the MeSH term level.

post-implementation effects are divided in three-year bins. Column 3 presents the effect of the MMA on MeSH categories related to high MMS diseases without controls. Results are close to zero and insignificant. Adding control variables in Columns 4 to 6 increases the coefficients but not the interpretation of the results. Our preferred specification (Equation 1) in Column 6 includes the full set of controls and serves as the baseline for the further analysis.

Under the assumption that there was no relationship between MMS and scientific activity prior to 2003, positive coefficients would indicate that the Part D demand shock led to an increased number of scientific publications in a given time period. This does not seem to be the case. Ten years from the passage of Medicare Part D, the point estimate can be interpreted as one additional percentage point in MMS resulting in 0.3% additional publications. This is considerably below the effect size on drug discoveries.[37] Taking the point estimate at face value, a MMS increase of one standard deviation (23.2 percentage points) leads to only 6.9% additional scientific publications compared to 20.2% additional drug discoveries. This can be considered as fairly inelastic.

[37]The 95th percentile confidence interval rules out an increase greater than 0.9%, which is approximately the effect of drug discoveries in Table 2.

Table 3: Scientific publications

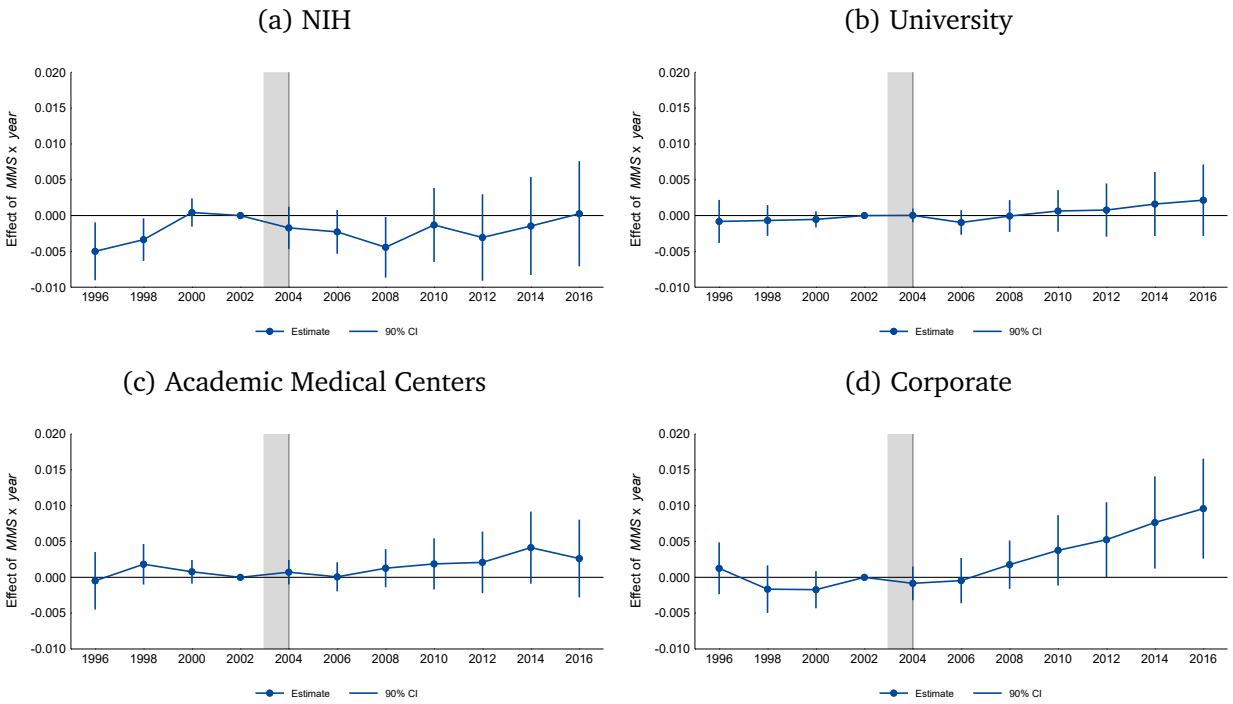
All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	DV: Scientific Publications					
MMS × post 2003	0.0003	0.0002				
	(0.001)	(0.001)				
MMS × 2004-05			0.0002	0.0006	0.0006	0.0006
			(0.001)	(0.001)	(0.001)	(0.001)
MMS × 2006-08			−0.0008	−0.0002	−0.0001	−0.0001
			(0.001)	(0.001)	(0.001)	(0.001)
MMS × 2009-11			0.0003	0.0013	0.0013	0.0013
			(0.001)	(0.002)	(0.002)	(0.002)
MMS × 2012-14			0.0004	0.0018	0.0017	0.0017
			(0.002)	(0.003)	(0.003)	(0.003)
MMS × 2015-16			0.0016	0.0032	0.0030	0.0030
			(0.002)	(0.003)	(0.003)	(0.003)
Cumul. US Market Size$_{t \text{ to } t+12}$	No	Yes	No	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	No	Yes	No	No	Yes	Yes
Cumul. New MeSH terms$_{t}$	No	Yes	No	No	No	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30680	30680	30680	30680	30680	30680
MeSH terms	1534	1534	1534	1534	1534	1534
ICD-group	129	129	129	129	129	129
Log-likelihood	−90077	−90007	−90000	−89974	−89923	−89923

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications. The control variables are log-transformed. Standard errors are clustered at the MeSH term level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

6.2 Type of Affiliation

We expect the relationship between market size and scientific research to depend on product market orientation. Biomedical scientists with corporate affiliations have direct financial ties to the market for prescription drugs. Their objectives should be aligned with those of the pharmaceutical industry (Henderson and Cockburn, 1996). Moreover, corporate scientists participate in the publication of the results of clinical trials in scientific journals. The latter also applies to scientists and practitioners at academic medical centers, who play an intermediary role between industry and academia (Lander and Atkinson-Grosjean, 2011; Lander, 2013). For scientists at universities is the relationship more subtle since market orientation differs across scientists and depends on a variety of factors (elaborated in Foray and Lissoni, 2010), but should be overall less pronounced compared to corporate scientists.

Figure 8: Event study – scientific publications by affiliation type



Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the MeSH term level. The dependent variable is the annual number of scientific publications split by at least one author's affiliation. Standard errors are clustered at the MeSH term level.

In the following analysis, we split our dependent variable by whether the publication was coauthored by at least one scientist who was affiliated with the NIH, a university, an academic medical center, or a corporation. Figure 8 shows the event study results, displaying the yearly excess publications in high MMS relative to low MMS MeSH categories. The effects go along with our predictions: the increase is least pronounced in the public research sphere at the NIH and most pronounced in the private research sphere at corporations. This is consistent with the idea that the market orientation of scientists matters.

Table 4 quantifies the Medicare Part D demand responses by affiliation subgroup. We find a statistically significant positive effect only among scientific publications with at least one corporate affiliation following the introduction of Medicare Part D, which increases gradually over time. The timing of the effect is consistent with the results of Blume-Kohout and Sood (2013), where the response magnifies through 2009 and appears to stabilize after 2012. In contrast, scientific publications with at least one university scientist have substantially smaller

coefficients.[38] The demand response from academic medical centers sits plausibly in-between universities and industry.

The point estimates of our analysis suggest that scientific research at non-corporate affiliations is less responsive to changes in market size compared to drug development within the same sample of ICD-9 groups. For instance, a MMS increase of one standard deviation (23.2 percentage points) leads to only 5.8% (7.9%) of additional scientific publications from university (academic medical center) scientists. The effect on scientific research directly conducted at the NIH is essentially zero. However, the same market size expansion leads to an increase of scientific publications coming from industry by 22.7%. This resembles the magnitude of our findings on drug development from Section 5. In the following sections, we will investigate which type of research drives this effect.

As outlined in Section 2, private insurance plans, which fall under the scope of Medicare Part D, do not have to cover all approved drugs. However, there are certain 'protected drug classes' for which most drugs are required to be included (*e.g.*, anti-cancer, anti-convulsant, anti-depressants, anti-psychotic, immuno-suppressant, HIV and AIDS drugs). Hence, we distinguish between ICD-9 groups that correspond to 'unprotected' or 'protected' drug classes. Our results in Appendix Figure A-12 correspond with our expectations and the previous literature (Blume-Kohout and Sood, 2013; Dranove et al., 2020): the only specification that contains statistically significant effects is the model focused on corporate affiliated scientists publishing in 'protected' ICD-9 groups.

6.3 Type of Research

Given the rise of corporate science in response to Medicare Part D, we investigate the vertical orientation of these research publications. To this end, we differentiate between scientific publications that are related to the development of drugs (*i.e.*, more applied in nature) and more basic science. We add the full set of MeSH terms to each publications and identify those MeSH terms that are related to clinical trials and those that are related to pharmaceutical products. We interpret the residual as fairly basic research. Finally, we split the dependent variable by the appliedness of the journal.

We show in Table 5 that the increases in corporate scientific publishing are more articulated among publications that relate to clinical trials and pharmaceutical products but less so in basic

[38] These results are quantitatively similar but less precise for publications from only corporate affiliations. This is consistent with the idea that pharmaceutical industry and university research are interlinked (Henderson and Cockburn, 1996).

Table 4: Scientific publications by affiliation type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications					
	NIH	University	No University	Hospital	Corporate	All US
MMS × 2004-05	−0.0006	0.0004	0.0007	0.0001	−0.0001	0.0004
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
MMS × 2006-08	−0.0021	−0.0005	−0.0002	−0.0002	0.0010	−0.0004
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
MMS × 2009-11	−0.0014	0.0010	0.0014	0.0011	0.0036	0.0009
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
MMS × 2012-14	−0.0018	0.0012	0.0020	0.0020	0.0059	0.0011
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
MMS × 2015-16	−0.0001	0.0025	0.0034	0.0030	0.0098**	0.0018
	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } \text{-}12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16880	30560	30220	27500	19880	30460
MeSH terms	844	1528	1511	1375	994	1523
ICD-group	111	129	129	128	125	129
Log-likelihood	−15174	−79980	−70035	−36845	−18220	−77282

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications split by at least one author's affiliation. In Column (1) at least one author is affiliated with the NIH, in Column (2) with a university, in Column (4) with an academic medical center, and in Column (5) with a firm. Column (3) includes publications that have at least one author not affiliated with a university. In Column (6), we count only publications, for which all authors have U.S. affiliations. The control variables are log-transformed. Standard errors are clustered at the MeSH term level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

science. The latter is supported by the fact that there is no increase in publications, for which the scientists received a NIH grant, a good proxy for the basicness of research. There is also no demand response on clinical trial and pharmaceutical product-related scientific research at universities or academic medical centers (Appendix Table A-17).[39] These results are supported by our event study analyses (Appendix Figure A-13) indicating that the disproportionate increases in clinical trials and pharmaceutical product publications for corporation are not driven by pre-existing trends but by the introduction of Medicare Part D. The magnitudes of the effects on corporate applied science are substantially larger compared to all (other) types of research.

[39] If at all, there is evidence for crowding out in those areas with the strongest increase among corporate publications: pharmaceutical products related publications.

Table 5: Scientific publications by type of research

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
		University			Corporate	
	Basic	CT	Pharma	Basic	CT	Pharma
MMS × 2004-05	0.0008	−0.0010	−0.0044**	−0.0020	0.0045	0.0090**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
MMS × 2006-08	−0.0002	−0.0015	−0.0056**	0.0001	0.0042	0.0060
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)
MMS × 2009-11	0.0013	−0.0006	−0.0056	0.0008	0.0104**	0.0161***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)	(0.006)
MMS × 2012-14	0.0015	0.0010	−0.0052	0.0028	0.0125**	0.0209***
	(0.003)	(0.004)	(0.005)	(0.003)	(0.005)	(0.008)
MMS × 2015-16	0.0029	0.0018	−0.0059	0.0066	0.0162***	0.0236**
	(0.003)	(0.005)	(0.006)	(0.004)	(0.006)	(0.009)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30520	19960	14600	18900	10820	6800
MeSH terms	1526	998	730	945	541	340
ICD-group	129	124	120	124	109	95
Log-likelihood	−75698	−21005	−11027	−15285	−7589	−3811

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1) to (3), the dependent variable is the number of university scientific publications and in Column (4)-(6) the number of corporate scientific publications, both split by the type of research. Columns (2) and (5) include only publications that are associated with MeSH terms related to clinical trials, and Columns (3) and (6) with MeSH terms related to pharmaceutical products. Column (1) and (4) include the residual. The control variables are log-transformed. Standard errors are clustered at the MeSH term level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

These results are consistent with our findings on the appliedness of journals. The demand response is most pronounced for research published in clinical practice, industry practice, and clinical-industrial journals driven by corporate research activity (Appendix Table A-16) and not by universities (Appendix Table A-15). We view our findings as evidence for the interpretation that a majority of corporate scientific publications, which result from the increase in market size, are related to actual drug development activities (*e.g.*, published clinical trial results) and do not constitute basic research. It supports the notion that demand pull effects are not strong enough to encourage true basic science – not even within industry.

6.4 Research Impact

In the last part of the analysis, we explore the impact of scientific research both within the scientific domain and beyond. As outlined in Section 3.4 we trace scientific publications to patents. These patent-paper linkages approximate whether scientific research got referenced in commercially relevant applications (Marx and Fuegi, 2020), in our context pharmaceutical and biomedical patents. Thus, we weight scientific publications by the journal impact factor (which is less affected by truncation compared to forward citations), by the 5-year availability of patents, and by the patent family size.

Our impact-weighted results in Table 6 show a similar but more noisy pattern with smaller magnitudes than the overall unweighted publication counts. Within the science domain, there is a disproportionate increase in corporate JIF-weighted publications (although not significant), which is substantially smaller for university publications. The magnitude of the corporate effect is around 50% of the effect on simple counts. This suggests that the effects on corporate science are more pronounced at the extensive (quantity) margin and less so when quality-weighted.

Patent-weighted publications increase primarily among corporate affiliations. The strongest effect is found initially after the passage of Medicare Part D. Corporate scientific research that ends up cited in a patent, increases by 0.68% in the years 2004-2005. This corresponds to an increase of 15.8% given a MMS increase of one standard deviation. The effect size decreases in the following years but reaches similar levels after 2011. The quick initial response suggests the existence of a reservoir of scientific research 'on the shelf' available for commercialization. This is consistent with prior literature, which suggests that publishing corporate science is used strategically in the patenting process (Della Malva and Hussinger, 2012). An alternative explanation is that industry became more likely to patent ideas from scientific publications as a reaction to the discrete increase in market size.[40]

7 Robustness Checks

We conduct a variety of robustness checks, which can be found in the Appendix. The demand response of corporate scientific research to the increase in market size is robust across a variety of changes, unless otherwise stated. The same applies to the inelastic response of all publications, across all other types of affiliation.

[40] Family-size weighted publications show a similar pattern, except for an arbitrary decrease in 2015-2016. All results are more pronounced when looking at event studies (Appendix Figure A-14).

Table 6: Impact/patent-weighted scientific publications

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	JIF-Weighted		Patent-Weighted		Family Size-Weighted	
	Uni	Corporate	Uni	Corporate	Uni	Corporate
MMS × 2004-05	−0.0001	−0.0002	0.0020	0.0068**	0.0001	0.0049
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
MMS × 2006-08	−0.0015	−0.0011	0.0009	0.0016	0.0000	0.0045
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
MMS × 2009-11	0.0004	0.0028	0.0011	0.0025	−0.0003	0.0044
	(0.002)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)
MMS × 2012-14	0.0006	0.0039	0.0008	0.0058*	−0.0013	0.0078*
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.005)
MMS × 2015-16	0.0017	0.0058	0.0003	0.0062	−0.0019	−0.0013
	(0.003)	(0.005)	(0.004)	(0.006)	(0.005)	(0.006)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30420	19720	16460	7960	16460	7960
MeSH terms	1521	986	823	398	823	398
ICD-group	129	125	114	92	114	92
Log-likelihood	−246132	−55995	−13166	−4419	−66536	−25953

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1) to (2), the dependent variable is the journal impact factor-weighted number of university/corporate scientific publications. In Columns (3) to (4), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). In Columns (5) to (6), we weight the number of scientific publications by the size of the average patent family associated with the publication. A patent/family size-weight is calculated based on the patent family's first application being filed within five years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the MeSH term level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

First, we *redefine the dependent variable* as a count of the annual number of scientific publications weighted by the inverse number of linked diseases (fractional counts) to account for multiple disease MeSH terms per publication. We also winsorize the dependent variable to deal with outliers (Figure A-10, Table A-8 and A-12).

Second, we calculate *different exposure variables*, for example, MMS based on prescription counts/quantity, binary indicators, and MMS based on 2003 values only (Figure A-11).

Third, we use *alternative control variables* such as the OECD market size or NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute. Moreover, we include control variables that do not accumulate future/past periods but

only consider year t (Table A-9 for all publications, Table A-13 for university publications, and Table A-14 for corporate publications).

Fourth, we estimate our model with *different specifications*, *e.g.*, using ICD-9 group fixed effects and cluster standard errors at the ICD-9 group level as elaborated in Section 3.2. Additionally, we employ linear regression with count dependent variables. In the latter specification, we find a significant increase in overall scientific publications, which suggests that MeSH terms with a high number of pre-MMA publications profit from a larger market in absolute counts but not relatively (Table A-9). The magnitude of this effect is substantially lower for university publications (Table A-13) than for corporate publications (Table A-14) when comparing the coefficients to the pre-MMA sample mean.

Fifth, we *restrict our dependent variable* to publications in which *all* authors are affiliated with universities or firms, respectively. While the effect sizes remain quantitatively similar, the estimation becomes less precise (Figure A-10) since we lose variation in the dependent variables. We also restrict the sample to publications in which all authors have a U.S. affiliation. This does not change the results (Table A-11).

Sixth, we re-investigate the effects including the year of the *MMA implementation (2005) into our baseline period*. This does not change the results regarding the type of affiliations or the type of research. The effects on patent-weighted publications disappear. This, however, is not surprising given that our initial findings showed primarily a disproportionate increase in 2004-2005 (Table A-18).

Seventh, we include *all* publications in our sample regardless of whether they include additional disease-related MeSH terms with unknown MMS. In our default specifications, we cautiously drop these publications since we do not know the exposure of these additional disease-related MeSH terms to the MMA (potentially confounding). Including them does not change the results (Table A-19).

Lastly, we chose *different aggregation levels* for our analysis. In Table A-20, we aggregate the dependent variable to the ICD-9 group level. Our results are robust to this aggregation.

Our results on drug development are robust to estimations using all ICD-9 three-digit codes available instead of the subset mentioned in Bhattacharya and Packalen (2011), ICD-9 group level fixed effects instead of ICD-9 three-digit code level fixed effects (both in Table A-4), alternative controls (Table A-5), alternative MMS calculations (Figure A-9), and including the MMA implementation in 2005 into our baseline period (Table A-6).

8 Thought Experiment – Incentivizing Upstream Research

We conclude with a thought experiment. For this example and the ensuing back-of-the-envelope calculation, we will take our results at face value. The goal is to explore the magnitude of the linkage between changes in downstream demand and upstream research. We start by assuming that scientific publications are mutually exclusive, meaning they are either categorized as a university, academic medical center, or corporate publication. From our preferred specification, a MMS increase of one standard deviation (23.2 percentage points) in exposure to Medicare Part D, leads to around 1,830 additional publications per year.[41] These additional publications break down broadly as follows: 1,200 authored by scientists with university affiliations, 250 with academic medical center affiliations, and 320 with corporate affiliations. Although the quantitative majority of these publications are focused in the subcategories where one would expect basic science to occur, it remains open whether this response is meaningful.

The direct costs of Medicare Part D, paid as subsidies to private insurances, during the program's first ten years was expected to be $80 billion annually (Medicare Trustees Report, 2006). When comparing our estimates to these direct costs, our results suggest that a subsidy-driven expansion of market size by $43 million would only lead to one additional scientific publication. This is substantially lower than, for example, the direct benefits of public funding, for which Myers (2020) reports that the average cost per publication is between $344,000 and $665,000 depending on the grant regime. At the midpoint of this range, this suggests that direct public funding of research would generate about 85 publications for each additional publication from our findings.

In the pre-MMA period, only about 5.4% of publications were authored by scientists affiliated with corporations. This rises to about 17.6% in the post-MMA period. If the whole scientific domain was as responsive as corporations, the 'cost' of one additional publication would fall from $43 million to $13.4 million. Direct public funding of research would still generate about 26 publications for every one from our findings. Thus, even considering the most responsive case, it does not appear that changes in downstream demand serve as sufficient incentives for upstream research.

Putting the results from this thought experiment into a broader context, Finkelstein's (2004) assertion appears to be correct – the post-MMA change in development was driven by a reordering of technology already in the development pipeline. Even in the context of our thought experiment, the impact on research seems insufficient, especially compared to public funding.

[41] We take the point estimate of our preferred specification and multiply it with the standard deviation in MMS and the 2003 number of scientific publications from Table 1.

9 Conclusion

R&D consists of two separate, but equally important components: research and development. The extant literature has conclusively found a link between changes in downstream market size and drug development – *i.e.*, 'D' – (*e.g.*, Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013; Dranove et al., 2020; Finkelstein, 2004). Heretofore, however, efforts to extend this linkage back to research – *i.e.*, 'R' – have been met with limited success. Acemoglu and Linn (2004), for example, were unable to find a relationship between demographic-driven expansions in market size and patenting. Using a similar identification strategy, Bhattacharya and Packalen (2011) find a positive relationship between disease prevalence and upstream research, but do not explicitly account for profit incentives. By exploiting the effects of quasi-experimental variation in market size introduced by Medicare Part D, we do not find a causal relationship between market size and research. We identify one limited exception to our core findings, however, and that involves corporate scientists conducting applied research.

Why does it appear that scientists are not incentivized by these changes in downstream market size? We can only conjecture as to what might explain the inelastic response. First, this type of demand pull incentive may attenuate by market distance so that it never reaches scientists (*e.g.*, Acemoglu and Linn, 2004). For example, if firms respond to demand shocks with 'off the shelf' projects (Finkelstein, 2004), new scientific discoveries may not be necessary to fuel the clinical pipeline. Similarly, the existing knowledge stock may be large enough to accommodate (for some time) the industry's higher demand for scientific discoveries. Our findings, however, suggest no response from university science even more than a decade later. Thus, a disconnect appears to exist between the kind of research industry uses as knowledge inputs and the kind of research upstream scientists conduct. Such exploration of this disconnect is left for future work.

Second, scientists may react to market size changes only if these changes affect the scientists' incentives through indirect channels, such as altruism, funding, or prestige. This is perhaps more likely the case if the market size increase is due to disease prevalence (*e.g.*, epidemics) as opposed to insurance coverage (*e.g.*, Medicare Part D).

Third, scientists may respond to demand pull differently, for example, by providing tacit knowledge through training (junior) scientists for industry (Roach and Sauermann, 2010). Unfortunately, we are unable to explore the above mechanisms with our data. Future research is necessary to understand the inelastic response of scientists. Survey evidence would be most helpful and may shed light on the mechanisms driving the disconnect between research and market size.

Finally, our study has important implications for policymakers. To the extent that there is a disconnect between the direction of academic research and the requirements of downstream markets, additional incentives may be needed to close this gap. NIH grants and public sector funding would appear to be the obvious choice as they are effective tools in fostering scientific research (Azoulay et al., 2019). It may also be the case that an expansion of R&D tax credits could be used to help incentivize companies to re-engage in and reverse the trend away from basic science research (Arora et al., 2018). These issues are left for future research.

References

Acemoglu, D. and J. Linn (2004). Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry. *The Quarterly Journal of Economics 119*(3), 1049–1090.

Adams, C. P. and V. V. Brantner (2006). Estimating the Cost of New Drug Development: Is it Really $802 million? *Health Affairs 25*(2), 420–428.

Agarwal, R. and P. Gaule (2021). What Drives Innovation? Lessons from COVID-19 R&D. IZA Discussion Paper No. 14079.

Aghion, P., A. Bergeaud, M. Lequien, and M. J. Melitz (2018). The Heterogeneous Impact of Market Size on Innovation: Evidence from French Firm-Level Exports. NBER Working Paper No. 24600.

Aghion, P., M. Dewatripont, and J. C. Stein (2008). Academic Freedom, Private-Sector Focus, and the Process of Innovation. *The RAND Journal of Economics 39*(3), 617–635.

Ahmadpoor, M. and B. F. Jones (2017). The Dual Frontier: Patented Inventions and Prior Scientific Advance. *Science 357*(6351), 583–587.

Arora, A., S. Belenzon, and A. Patacconi (2018). The Decline of Science in Corporate R&D. *Strategic Management Journal 39*(1), 3–32.

Azoulay, P., J. S. Graff Zivin, D. Li, and B. N. Sampat (2019). Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules. *The Review of Economic Studies 86*(1), 117–152.

Bhattacharya, J. and M. Packalen (2011). Opportunities and Benefits as Determinants of the Direction of Scientific Research. *Journal of Health Economics 30*(4), 603–615.

Blume-Kohout, M. E. and N. Sood (2013). Market Size and Innovation: Effects of Medicare Part D on Pharmaceutical Research and Development. *Journal of Public Economics 97*, 327–336.

Cockburn, I. M. and R. M. Henderson (2000). Publicly Funded Science and the Productivity of the Pharmaceutical Industry. *Innovation Policy and the Economy 1*, 1–34.

Cohen, W. M., H. Sauermann, and P. Stephan (2020). Not in the Job Description: The Commercial Activities of Academic Scientists and Engineers. *Management Science 66*(9), 4108–4117.

Della Malva, A. and K. Hussinger (2012). Corporate Science in the Patent System: An Analysis of the Semiconductor Technology. *Journal of Economic Behavior & Organization 84*(1), 118–135.

Dranove, D., C. Garthwaite, and M. I. Hermosilla (2020). Expected Profits and The Scientific Novelty of Innovation. NBER Working Paper No. 27093.

Dubois, P., O. De Mouzon, F. Scott-Morton, and P. Seabright (2015). Market Size and Pharmaceutical Innovation. *The RAND Journal of Economics 46*(4), 844–871.

Duggan, M. and F. Scott Morton (2010). The Effect of Medicare Part D on Pharmaceutical Prices and Utilization. *The American Economic Review 100*(1), 590–607.

Duggan, M. G. and F. Scott Morton (2011). The Medium-Term Impact of Medicare Part D on Pharmaceutical Prices. *The American Economic Review 101*(3), 387–92.

Finkelstein, A. (2004). Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry. *The Quarterly Journal of Economics 119*(2), 527–564.

Foray, D. and F. Lissoni (2010). University Research and Public - Private Interaction. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 1, pp. 275–314. Elsevier.

Garthwaite, C., R. Sachs, and A. D. Stern (2021). Which Markets (Don't) Drive Pharmaceutical Innovation? Evidence from U.S. Medicaid Expansions. NBER Working Paper No. 28755.

Griliches, Z. (1957). Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica, Journal of the Econometric Society*, 501–522.

Griliches, Z. (1981). Market Value, R&D, and Patents. *Economics Letters 7*(2), 183–187.

Hall, B. H. and J. Mairesse (1995). Exploring the Relationship between R&D and Productivity in French Manufacturing Firms. *Journal of Econometrics 65*(1), 263–293.

Hausman, J. A., B. H. Hall, and Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica 52*(4), 909–938.

Henderson, R. and I. Cockburn (1996). Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. *The RAND Journal of Economics*, 32–59.

Hermosilla, M. and Y. Wu (2018). Market Size and Innovation: The Intermediary Role of Technology Licensing. *Research Policy 47*(5), 980–991.

Higgins, M. J. and D. Rodriguez (2006). The Outsourcing of R&D through Acquisitions in the Pharmaceutical Industry. *Journal of Financial Economics 80*(2), 351–383.

Hvide, H. K. and B. F. Jones (2018). University Innovation and the Professor's Privilege. *The American Economic Review 108*(7), 1860–98.

Jacob, B. A. and L. Lefgren (2011). The Impact of Research Grant Funding on Scientific Productivity. *Journal of Public Economics 95*(9), 1168–1177.

Krieger, J. L., D. Li, and D. Papanikolaou (2018). Missing Novelty in Drug Development. NBER Working Paper No. 24595.

Kyle, M. K. (2020). The Alignment of Innovation Policy and Social Welfare: Evidence from Pharmaceuticals. *Innovation Policy and the Economy 20*(1), 95–123.

Kyle, M. K. and A. M. McGahan (2012). Investments in Pharmaceuticals Before and After TRIPS. *The Review of Economics and Statistics 94*(4), 1157–1172.

Lach, S. and M. Schankerman (2008). Incentives and Invention in Universities. *The RAND Journal of Economics 39*(2), 403–433.

Lander, B. (2013). Sectoral Collaboration in Biomedical Research and Development. *Scientometrics 94*(1), 343–357.

Lander, B. and J. Atkinson-Grosjean (2011). Translational Science and the Hidden Research System in Universities and Academic Hospitals: A Case Study. *Social Science & Medicine 72*(4), 537–544.

Levin, S. G. and P. E. Stephan (1991). Research Productivity over the Life Cycle: Evidence for Academic Scientists. *The American Economic Review*, 114–132.

Lichtenberg, F. R. and J. Waldfogel (2009). Does Misery Love Company? Evidence from Pharmaceutical Markets Before and After the Orphan Drug Act. *Michigan Telecommunications and Technology Law Review 46*(4), 844–871.

Marx, M. and A. Fuegi (2020). Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles. *Strategic Management Journal*.

Medicare Trustees Report (2006). Annual Report on the Board of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. Technical report.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Mowery, D. and N. Rosenberg (1979). The Influence of Market Demand upon Innovation: A Critical Review of some Recent Empirical Studies. *Research Policy 8*(2), 102–153.

Myers, K. (2020). The Elasticity of Science. *American Economic Journal: Applied Economics 12*(4), 103–34.

Owen-Smith, J. and W. W. Powell (2001). To Patent or Not: Faculty Decisions and Institutional Success at Technology Transfer. *The Journal of Technology Transfer 26*(1), 99–114.

Packalen, M. and J. Bhattacharya (2018). Does the NIH Fund Edge Science? NBER Working Paper No. 24860.

Pakes, A. and M. Schankerman (1984). The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources. In Z. Griliches (Ed.), *R&D, Patents, and Productivity*, pp. 73–88. University of Chicago Press.

Poege, F., D. Harhoff, F. Gaessler, and S. Baruffaldi (2019). Science Quality and the Value of Inventions. *Science Advances 5*(12). eaay7323.

Roach, M. and H. Sauermann (2010). A Taste for Science? PhD Scientists' Academic Orientation and Self-Selection into Research Careers in Industry. *Research policy 39*(3), 422–434.

Rogers, F. B. (1963). Medical Subject Headings - Communication to the Editor. *Bulletin of the Medical Library Association 51*(1), 114–116.

Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press.

Schumpeter, J. A. (1939). *Business Cycles*. McGraw-Hill.

Smith, P. W. (2006). The National Institutes of Health (NIH): Organization, Funding, and Congressional Issues. CRS Report for Congress RL33695. Library of Congress, Washington, DC.

Stephan, P. E. (1996). The Economics of Science. *Journal of Economic literature 34*(3), 1199–1235.

Stephan, P. E. (2012). *How Economics Shapes Science*. Harvard University Press.

Stern, S. (2004). Do Scientists Pay to be Scientists? *Management Science 50*(6), 835–853.

Thursby, J. G., R. Jensen, and M. C. Thursby (2001). Objectives, Characteristics and Outcomes of University Licensing: A Survey of Major US Universities. *The Journal of Technology Transfer 26*(1), 59–72.

Thursby, J. G. and M. C. Thursby (2011). Has the Bayh-Dole Act Compromised Basic Research? *Research Policy 40*(8), 1077–1083.

Tijssen, R. J. (2010). Discarding the 'Basic Science/Applied Science'Dichotomy: A Knowledge Utilization Triangle Classification System of Research Journals. *Journal of the American Society for Information Science and Technology 61*(9), 1842–1852.

Toole, A. A. (2012). The Impact of Public Basic Research on Industrial Innovation: Evidence from the Pharmaceutical Industry. *Research Policy 41*(1), 1–12.

Watzinger, M. and M. Schnitzer (2019). Standing on the Shoulders of Science. CEPR Discussion Paper No. DP13766.

Appendix

Content

A Data Construction

A.1 Control Variables

This data appendix describes the construction of our control variables: projected market size, NIH funding, and research opportunities.

Projected Market Size

We build a measure of the exogenous components of U.S. market size (Acemoglu and Linn, 2004; Blume-Kohout and Sood, 2013). Each disease group has a different age profile and, hence, is differently affected by both domestic and global demographic trends. Therefore, we use demographic (projection) data from the UN World Population Prospects[42] for the United States between 1997 and 2040 in order to calculate how the potential future market size would develop if only population growth mattered.

To this end, we keep the age profile of each disease constant and calculate the average expenditure share of drugs associated with each five-year age bin for each ICD-9 group in the pre-MMA period. Drug expenditures are measured in real-terms (base year of 2003) based on the MEPS data. We then attribute the US population growth to each age bin until 2040. Hence, our measure for projected US market size M_{it} displays the annual expected drug expenditures in each ICD-9 group i in year t. In concordance with Blume-Kohout and Sood (2013), we accumulate the projected market size M_{it} over a period of 12 lead years as of year t. This reflects the average market exclusivity term of new drugs (Adams and Brantner, 2006).

Since pharmaceutical markets are typically considered as global (Acemoglu and Linn, 2004), we build the same measure for the potential market size in all OECD countries.

NIH Funding

We control for previous years' public research funding related to each disease category. Many scholars have shown the importance of public research funding, *e.g.*, from the NIH, for progress in biomedical research and pharmaceutical innovation.[43,44] Since Congress doubled the NIH budget in the five years preceding the MMA from $13.6 billion in 1998 to $27.1 billion in

[42] The data can be found here: https://population.un.org/wpp/Download/Standard/Population/.

[43] See Jacob and Lefgren (2011) on scientific productivity, Myers (2020) on the direction of science, and Packalen and Bhattacharya (2018) on novelty.

[44] See Azoulay et al. (2019) on patenting and Blume-Kohout and Sood (2013) on NMEs entering clinical trials, and Toole (2012) on new drugs approved.

2003, it becomes an especially important determinant of R&D in any analysis of Part D (Smith, 2006). The NIH consists of twenty-seven Institutes and Centers, where each receives its own Congressional appropriation (Azoulay et al., 2019). However, the historical doubling of research funding was distributed unequally between these Institutes. In order to control for any disproportionate increase in funding correlated with the MMA, we calculate for each of our 129 ICD-9 groups the exposure to the NIH budget over time.[45]

To this end, we assign each ICD-9 group to one of these Institutes (*e.g.*, ICD-9 162 *malignant neoplasm of trachea, bronchus, and lung* to the NCI *National Cancer Institute*). Since research grants are distributed within Institutes primarily by scientific merit (see discussion on NIH funding rules by Azoulay et al. (2019)) and not by allocation to narrower disease categories, we attribute the full annual Institute's budget to each ICD-9 group.[46] In an alternative approach we attribute budgets based on the share of all publications in a disease category that acknowledge a specific Institute. According to aforementioned studies, the effect of funding on research typically materializes within the first years from the grant. Therefore, we accumulate the NIH_{it} funding over a period of 12 lagged years until year t.[47]

Research Opportunities

Scientists may switch research projects to take advantage of greater research opportunities. We therefore account for the *availability of research opportunities*.[48]

We develop a direct measure of new research opportunities taking advantage of the development of the MeSH hierarchy over time.[49] The MeSH vocabulary in its current form was introduced in 1963 (Rogers, 1963) and was intended as a dynamic list that incorporates new concepts in the medical field.[50] The NLM introduces annually hundred new MeSH terms based

[45] We retrieve NIH spending data (Mechanism Detail by IC, FY 1983-2019) from https://officeofbudget.od.nih.gov/spending_hist.html [downloaded on February 17, 2020].

[46] In rare cases we assign more then one Institute or Center to an ICD-9 group. In these cases we attribute both budgets to the disease category.

[47] We use the Biomedical Research and Development Price Index in order to calculate real values with the base year of 2003. The data can be found here: https://officeofbudget.od.nih.gov/gbipriceindexes.html.

[48] Bhattacharya and Packalen (2011) construct a measures of research opportunities based on the content of research inputs and the first appearance of the idea in a scientific publication. Using the set of approved active ingredients as an input factor for future scientific research, they estimate structural productivity parameters, which takes into account diffusion and exhaustion of knowledge, in order to infer the quality of associated opportunities. The disadvantage of the approach is that it relies on a very narrow set of research inputs that relate primarily to drug-related medical research but not basic science.

[49] MeSH terms are organised into a hierarchy called the MeSH tree. Disease groups are first defined very broadly, but become more narrow with every sub-type of a disorder. The bulk data can be found here: https://www.nlm.nih.gov/databases/download/mesh.html.

[50] See https://www.nlm.nih.gov/mesh/intro_preface.html#pref_rem.

on the need to appropriately describe concepts being discussed in the literature. New terms are added for emerging diseases, breakdowns of existing diseases, and additional terminology to reflect topical areas that are not well represented in MeSH.[51] We interpret the introduction of a new MeSH term as an emerging research opportunity since NLM employees collect new terms that begin to appear in the scientific literature, for example in emerging areas of research.[52] In 2003, the NLM added, for instance, the following MeSH terms to the vocabulary: *Retinoschisis* reflects a more detailed conceptualisation of an existing disease (Retinal Degeneration) and *Severe Acute Respiratory Syndrome* describes a newly occurring disease related to the 2000s outbreaks of the SARS-Coronavirus. All together, the introduction of each term approximates the beginning of a new research field.

We measure new research opportunities K_{it} that are associated with an ICD-9 group by calculating the number of *new* MeSH terms that occur below the hierarchy level of our ICD-9-MeSH crosswalk, introduced in a given year t.[53] Since new research opportunities likely become obsolete over time, we add a discount factor of 0.8.[54] This approach is novel to the literature, which typically uses MeSH terms statically as keywords to understand shifts in the direction of science, but not the dynamic development of opportunities.

[51] The list of new MeSH Headings for 2020 published by the NLM is available here: https://www.nlm.nih.gov/mesh/2020/download/2020NewMeSHheadingsSingleColumn.pdf.

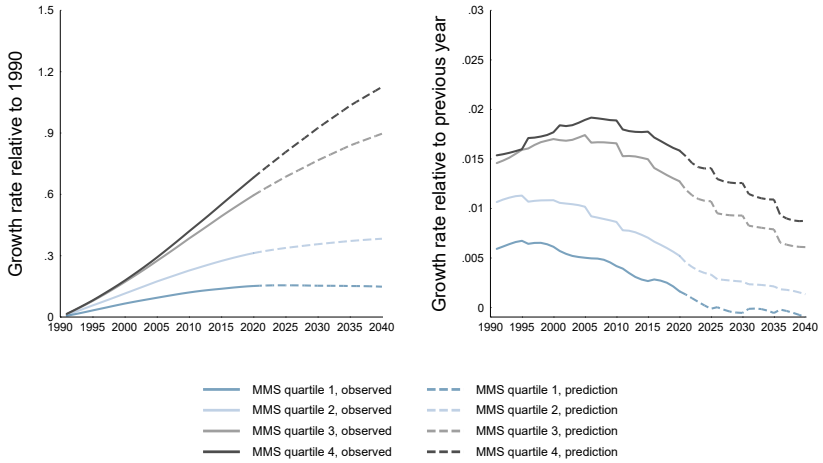
[52] See for more information: https://www.nlm.nih.gov/pubs/factsheets/mesh.html.

[53] We use the date of establishment since this is not sensitive to the transformation of the analogue MeSH vocabulary to the digital vocabulary in 1999. For more details on the variables, see: https://www.nlm.nih.gov/mesh/xml_data_elements.html.

[54] Estimated depreciation rates of knowledge capital vary in the literature. Common values lie between 15% (Griliches, 1981; Hall and Mairesse, 1995) and 25% (Pakes and Schankerman, 1984). Our results are robust to applying different depreciation rates.

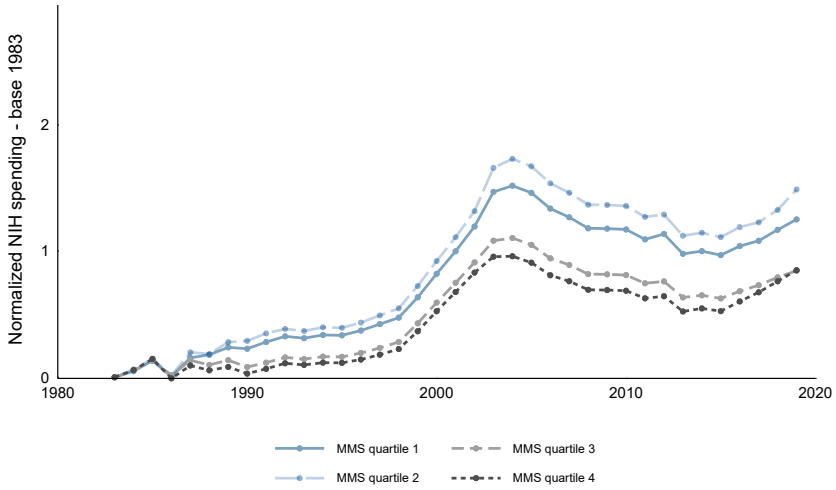
A.2 Evolution Alternative Controls

Figure A-1: OECD market size growth



Notes: The left figure presents the annual OECD population-growth driven market size (in 2003 values) of each ICD-9 group, aggregated by MMS quartiles, and normalized in 1990. The right figure shows the annual change in OECD market size relative to the prior year.

Figure A-2: NIH funding trends (proportionally by the share of acknowledgements)



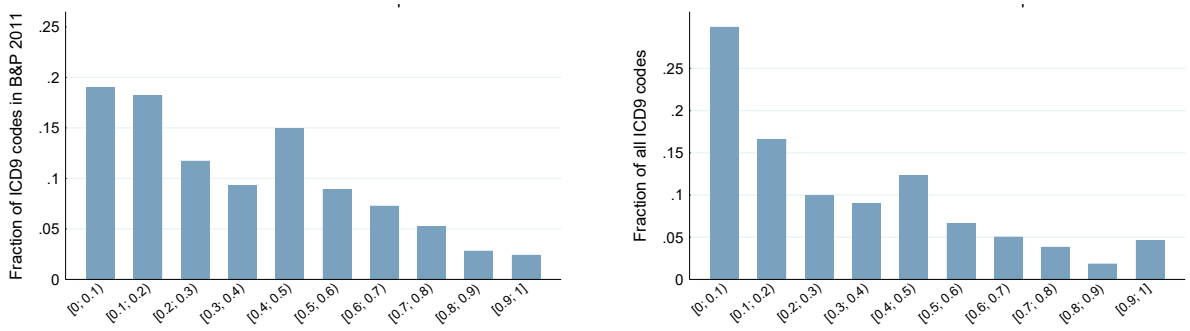
Notes: The figure presents the annual NIH spending (in 2003 values) attributed to each ICD-9 group, averaged by MMS quartiles, and normalized in 1990. We attribute NIH budgets based on the share of all publications in an ICD-9 group that acknowledge a specific Institute/Center.

B Figures

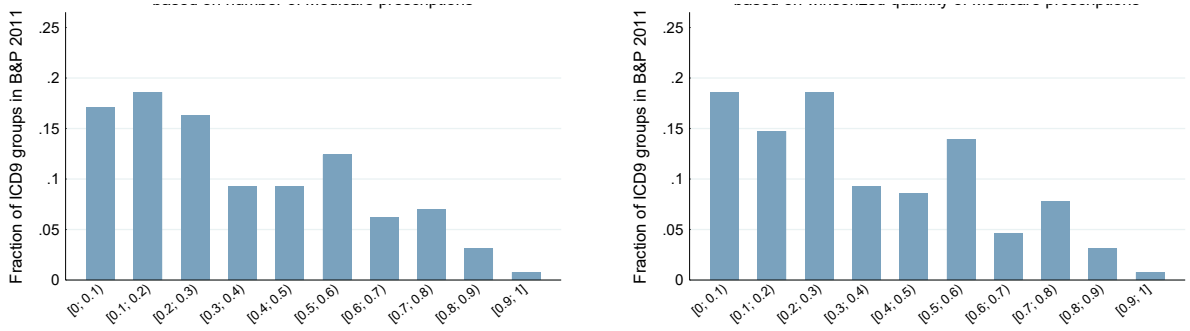
B.1 Figures – Medicare Market Shares

Figure A-3: Distribution of Medicare market shares in 1997-2003

(a) Based on number of Medicare patients

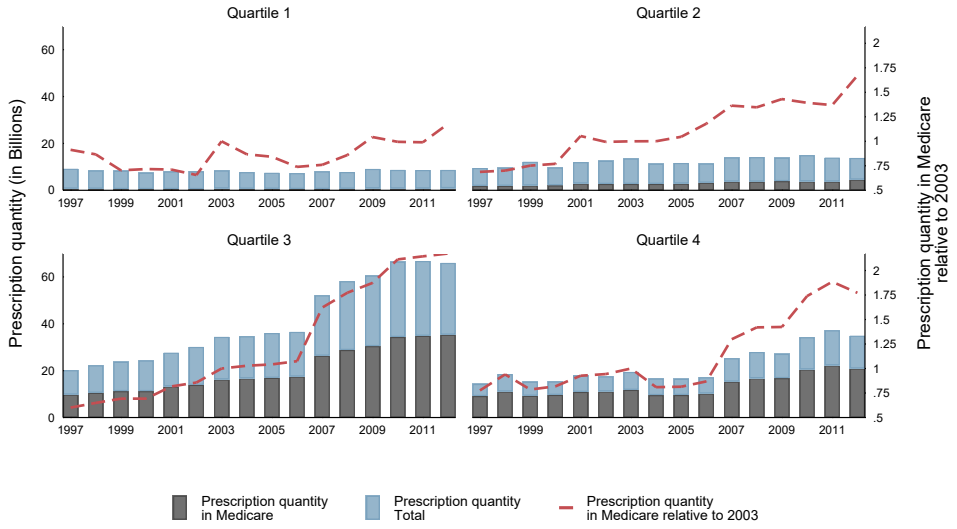


(b) Based on number of Medicare prescriptions



Notes: The figures in the top row present the distribution of MMS scores among ICD-9 three-digit codes. Figure (a) shows all 272 ICD-9 three-digit codes, which are included in the MeSH-ICD-9 crosswalk by Bhattacharya and Packalen (2011). Figure (b) shows all 752 ICD-9 three-digit codes in the 1997-2003 MEPS. We use the patient-weighted average of each year between 1997-2003. The annual MMS are calculated using the total number of patients in Medicare relative to all patients for each ICD-9 three-digit code. Both figures in the bottom row present the distribution of MMS scores among the 129 ICD-9 groups, which are included in the MeSH-ICD-9 crosswalk by Bhattacharya and Packalen (2011). In Figure (c) we use the prescription count-weighted average of each year between 1997-2003. The annual MMS are calculated using the total number of prescriptions financed by Medicare relative to all prescriptions for each ICD-9 group. In Figure (d) we use the prescription quantity-weighted average of each year between 1997-2003. The annual MMS are calculated using the quantity-weighted prescriptions financed by Medicare relative to all quantity-weighted prescriptions for each ICD-9 group.

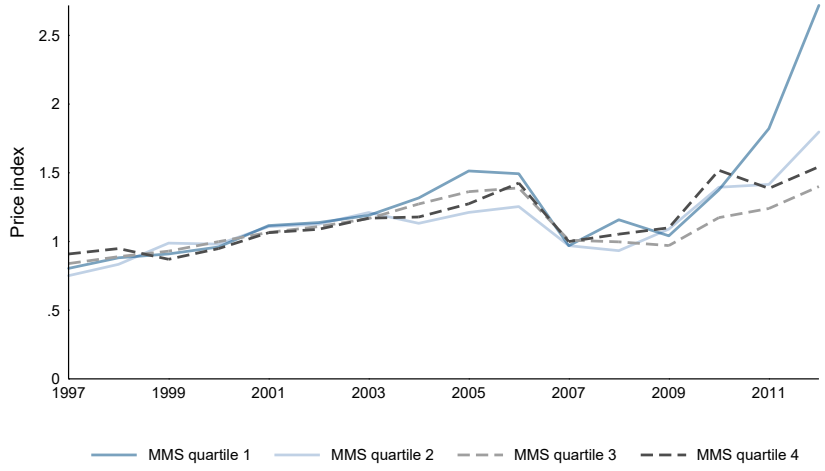
Figure A-4: Evolution of Medicare drug prescriptions by MMS quartiles



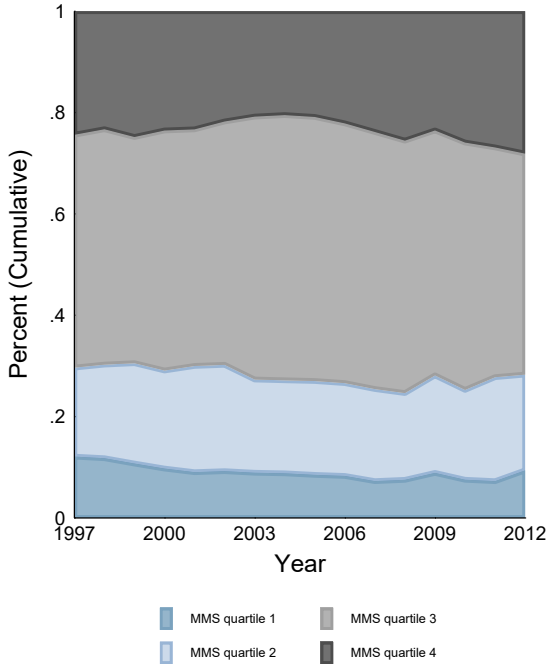
Notes: The figure shows the evolution of prescription quantities for each ICD-9 group aggregated by MMS quartiles. The grey bars display the quantity of prescriptions financed by Medicare, the blue bars display the non-Medicare prescription quantities. Prescriptions are counted multiple times if they appear in more than one ICD-9 group. The red line represents the relative increase in the quantity of Medicare prescriptions with respect to the baseline year 2003.

Figure A-5: Evolution of drug prices and total revenue

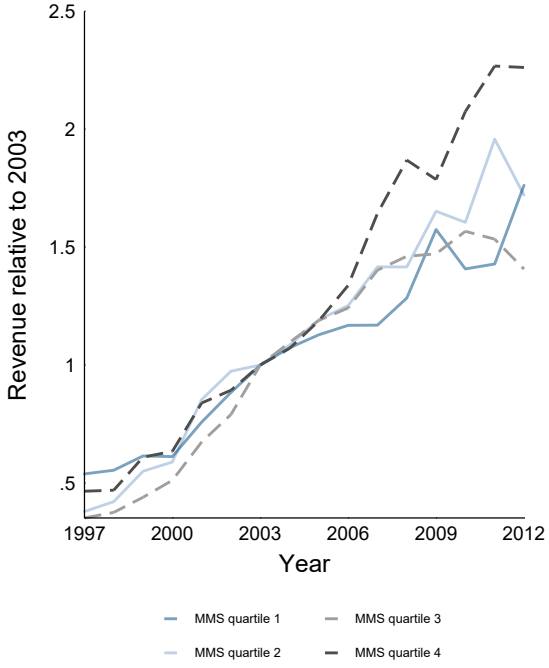
(a) Evolution of drug price indices



(b) Composition of total revenue



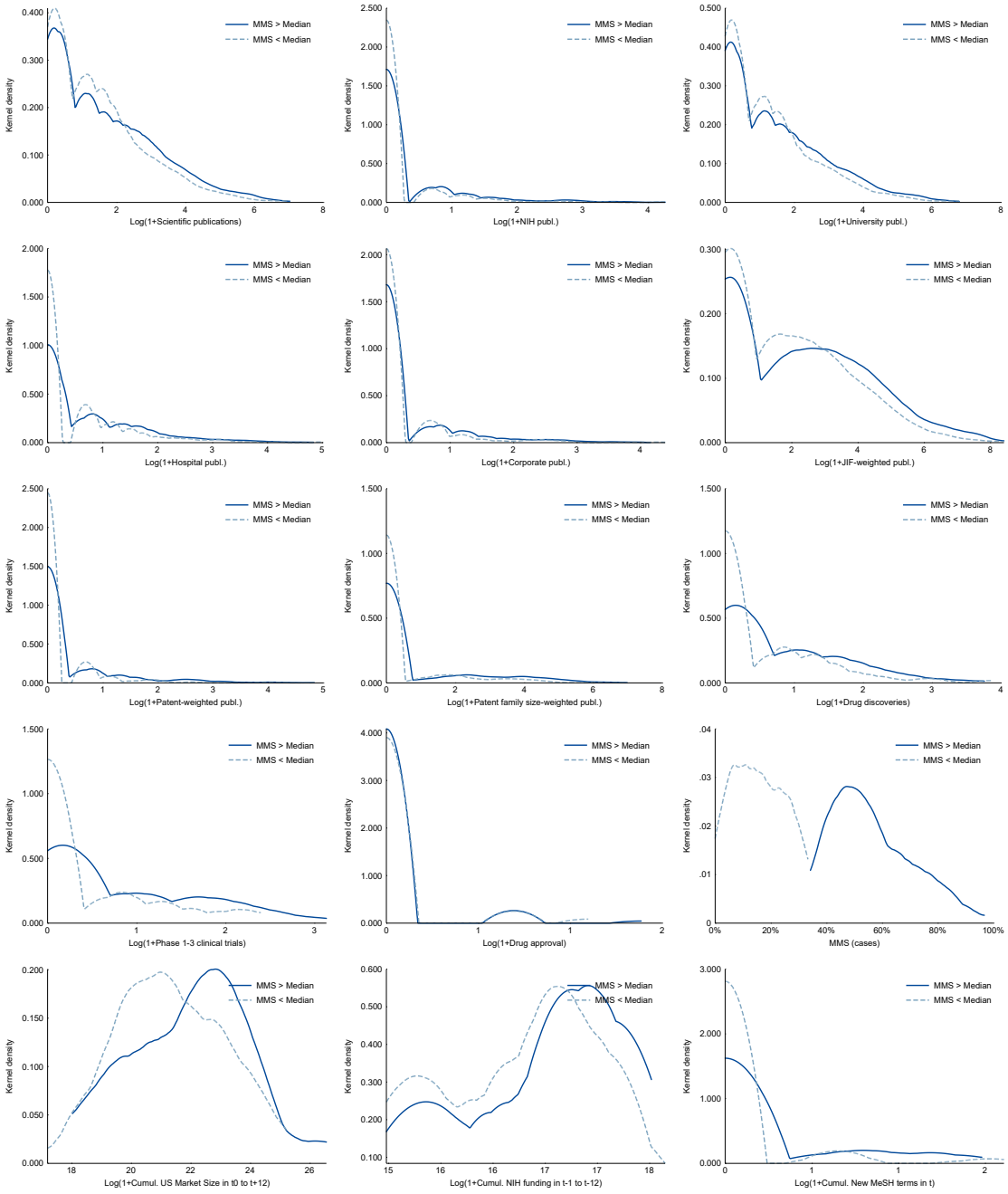
(c) Evolution of total revenue



Notes: The top figure presents the evolution of a price index by MMS quartile. It is calculated based on a prescription quantity-weighted basket of drugs (1000 most sold drugs in total) following the procedure outlined in Duggan and Scott Morton (2010). Drug-level prices are inferred from the MEPS, winsorized, and imputed/extrapolated if missing. We drop a drug-form-disease combination if it does not appear in at least 2/3 of the sample periods, in 3 consecutive years, and exhibits price growth in the top 1% of the distribution. The index is set to one by using a 1997-2003 divisor. The bottom left figure presents the composition of total drug revenues by MMS quartile over time. Revenues are calculated based on the total payment of all drugs which are prescribed for a certain ICD-9 group. The bottom right figure shows the normalized evolution of total revenues. 2003 serves as the baseline year.

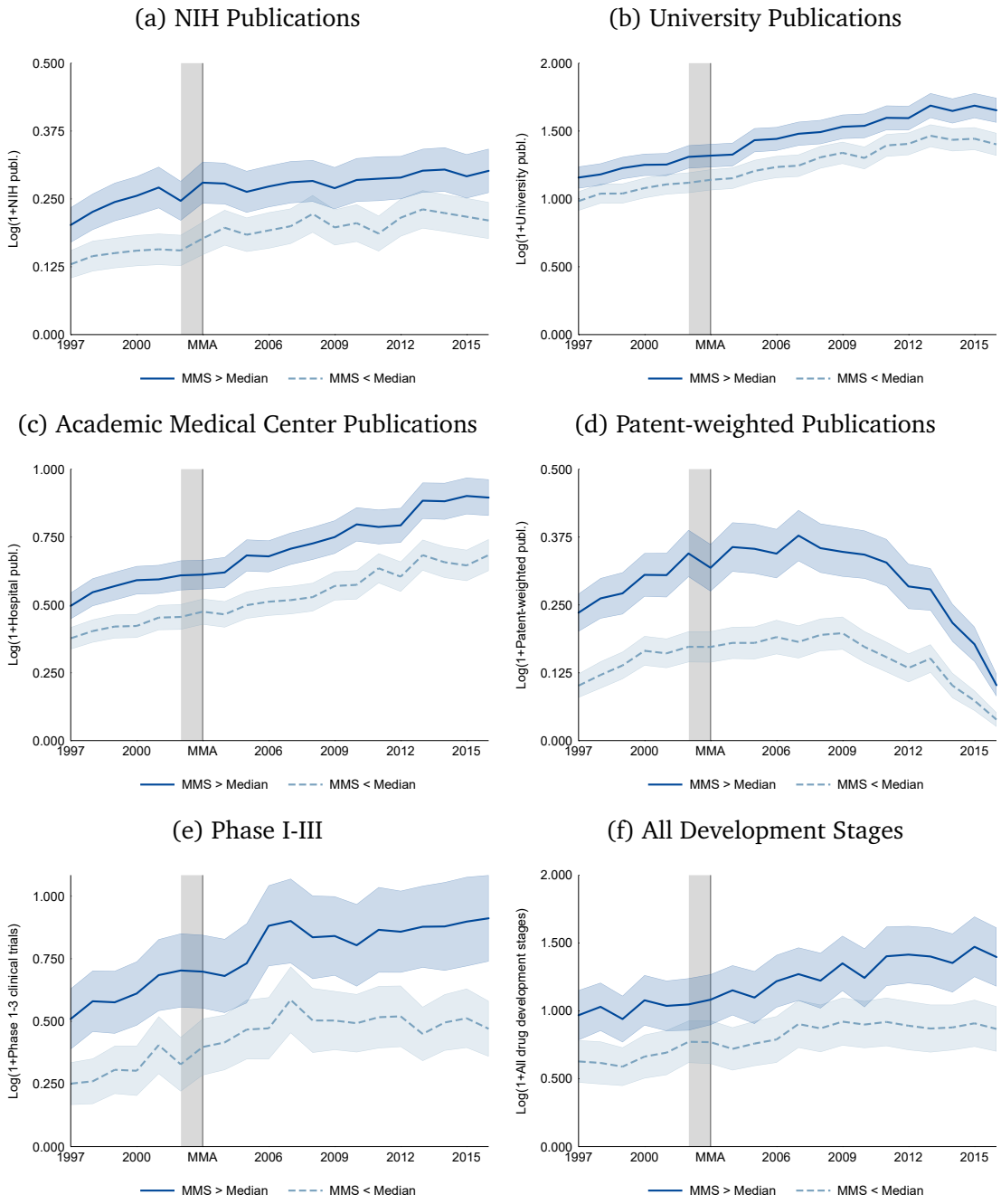
B.2 Figures – Descriptive Analysis

Figure A-6: Distribution of pre-MMA dependent/independent variables



Notes: The figures compare the Kernel density of dependent and independent variables split at the Median MMS in the year 2003. The unit of observation is the MeSH level for publication related variables, the ICD-9 three-digit code level for drug related variables, and the ICD-9 group level for MMS/control variables.

Figure A-7: Trends in scientific publications and drug development by MMS

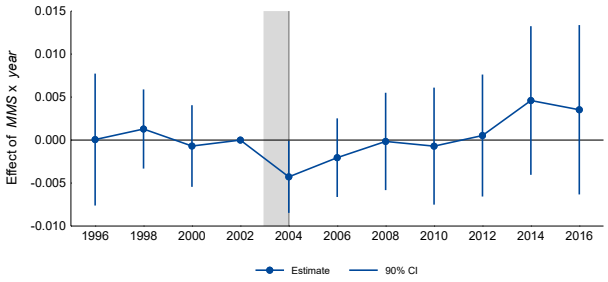


Notes: Figures (a)-(c) present the log-transformed average number of annual publication counts (a) with NIH participation, (b) with university participation, and (c) with academic medical center participation. Figure (d) presents the log-transformed average number of patent-weighted publications from all affiliations. Figures (e)-(f) present the log-transformed average number of annual NME (e) in phase I-III clinical trials and (f) in all drug development stages (preclinical, phase I-III clinial trials, registration, approval). For reasons of comparability, the unit of observation is the unique ICD-9 group level in all graphs.

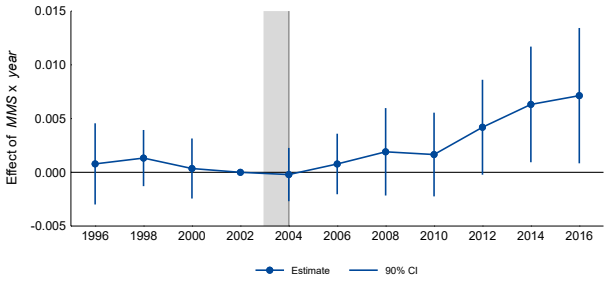
B.3 Figures – Multivariate Analysis Clinical Drug Development

Figure A-8: Event study – drug development (alternative outcomes)

(a) Phase I-III clinical trials



(b) All development stages



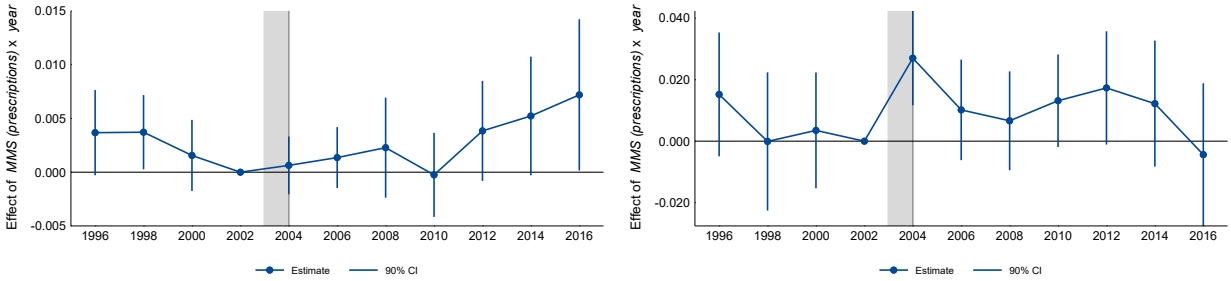
Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the ICD-9 three-digit code level, with MMS being calculated based on patient counts at the ICD-9 group level. Standard errors are clustered at the ICD-9 three-digit code level.

Figure A-9: Event study – drug development (alternative MMS)

MMS based on prescription counts

(a) Drug discoveries

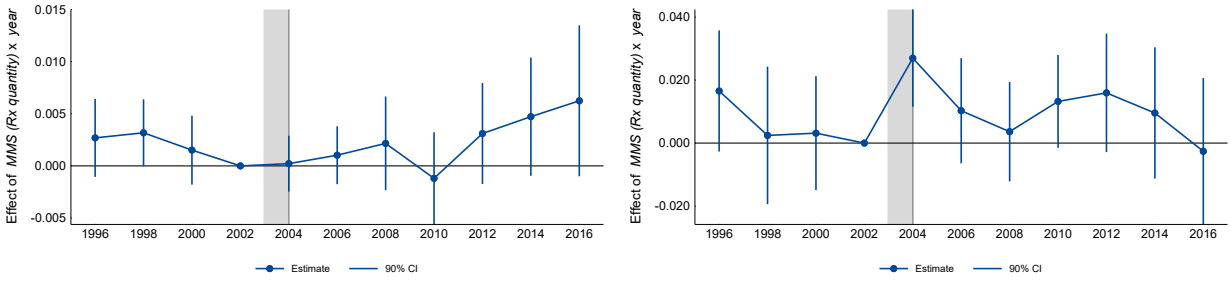
(b) Drug approvals



MMS based on prescription quantity

(c) Drug discoveries

(d) Drug approvals

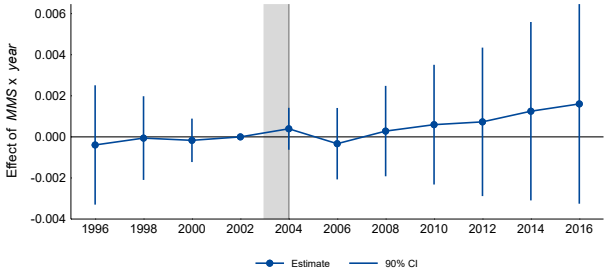


Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the ICD-9 three-digit code level. In the top figures, the MMS is calculated based on the number of prescriptions, in the bottom figures the MMS is calculated based on prescription quantity (Rx-quantity), both at the ICD-9 group level. Standard errors are clustered at the ICD-9 three-digit code level.

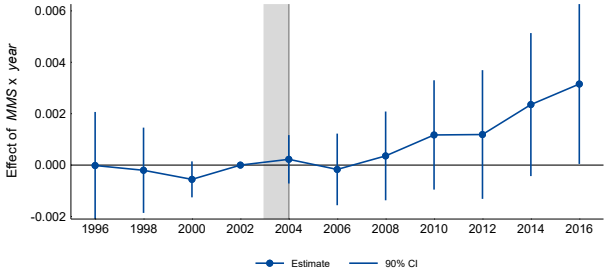
B.4 Figures – Multivariate Analysis Biomedical Science

Figure A-10: Event study – scientific publications (alternative outcomes)

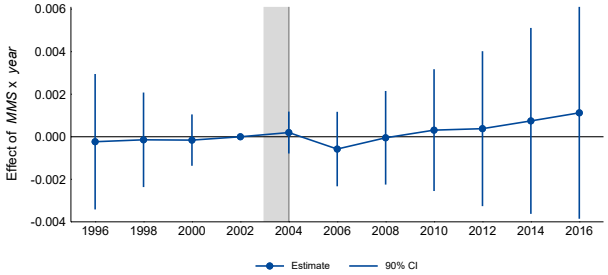
(a) Publications (fractional)



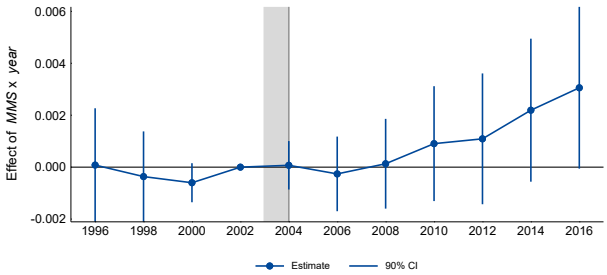
(b) Publications (winsorized)



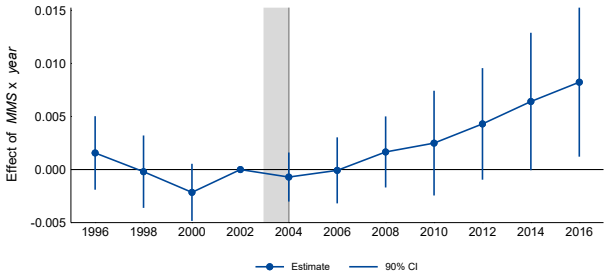
(c) University publications (fractional)



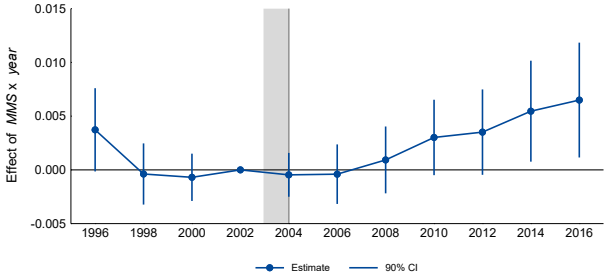
(d) University publications (winsorized)



(e) Corporate publications (fractional)



(f) Corporate publications (winsorized)

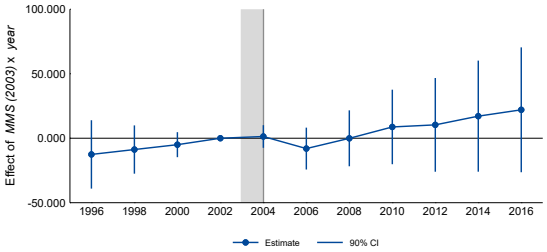


Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the MeSH term level. In the left figures, the dependent variable is the annual number of scientific publications weighted by the number of distinct ICD-9 per publications (thus, counted fractional). In the right figures, the number of scientific publications is winsozired at the annual 99th percentile. Standard errors are clustered at the MeSH term level.

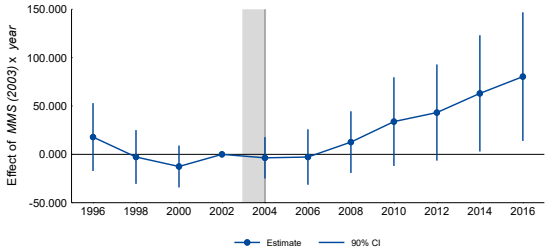
Figure A-11: Event study – scientific publications (alternative MMS)

MMS based on patient counts (2003 only)

(a) University publications

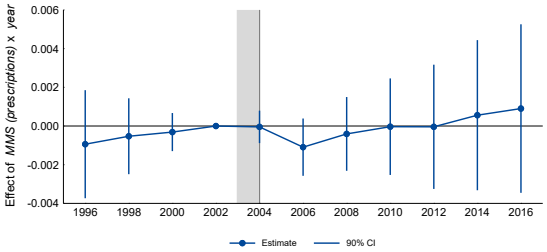


(b) Corporate publications

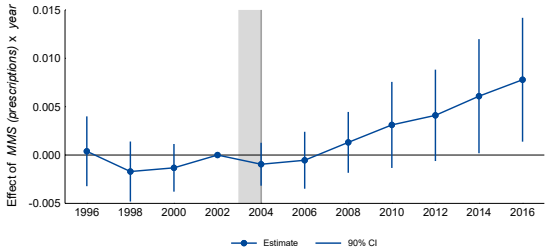


MMS based on prescription counts

(c) University publications

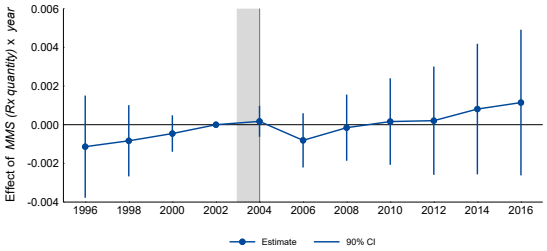


(d) Corporate publications

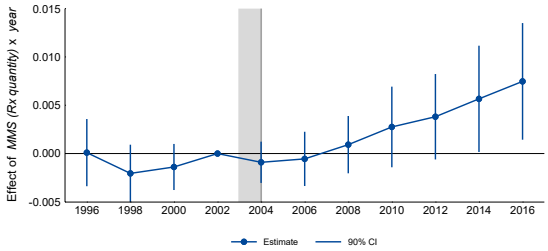


MMS based on prescription quantity

(e) University publications

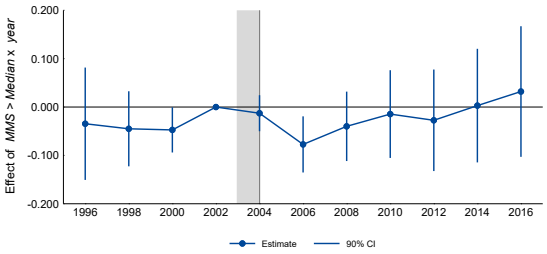


(f) Corporate publications

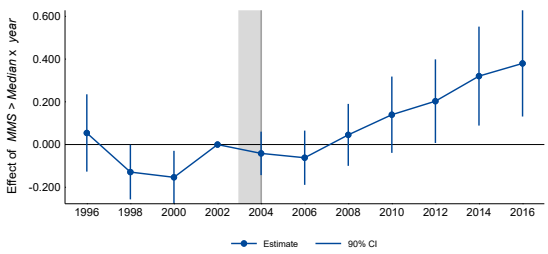


MMS based on patient counts (above median)

(g) University publications

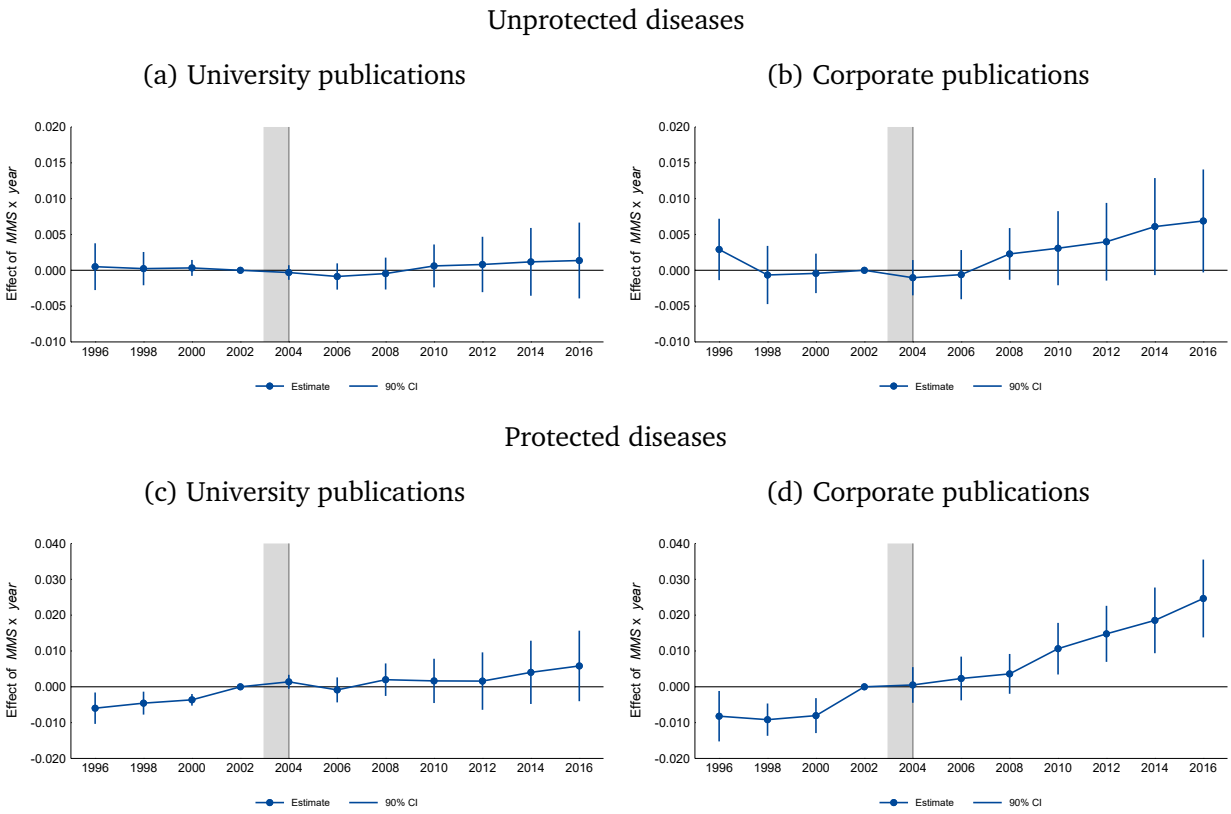


(h) Corporate publications



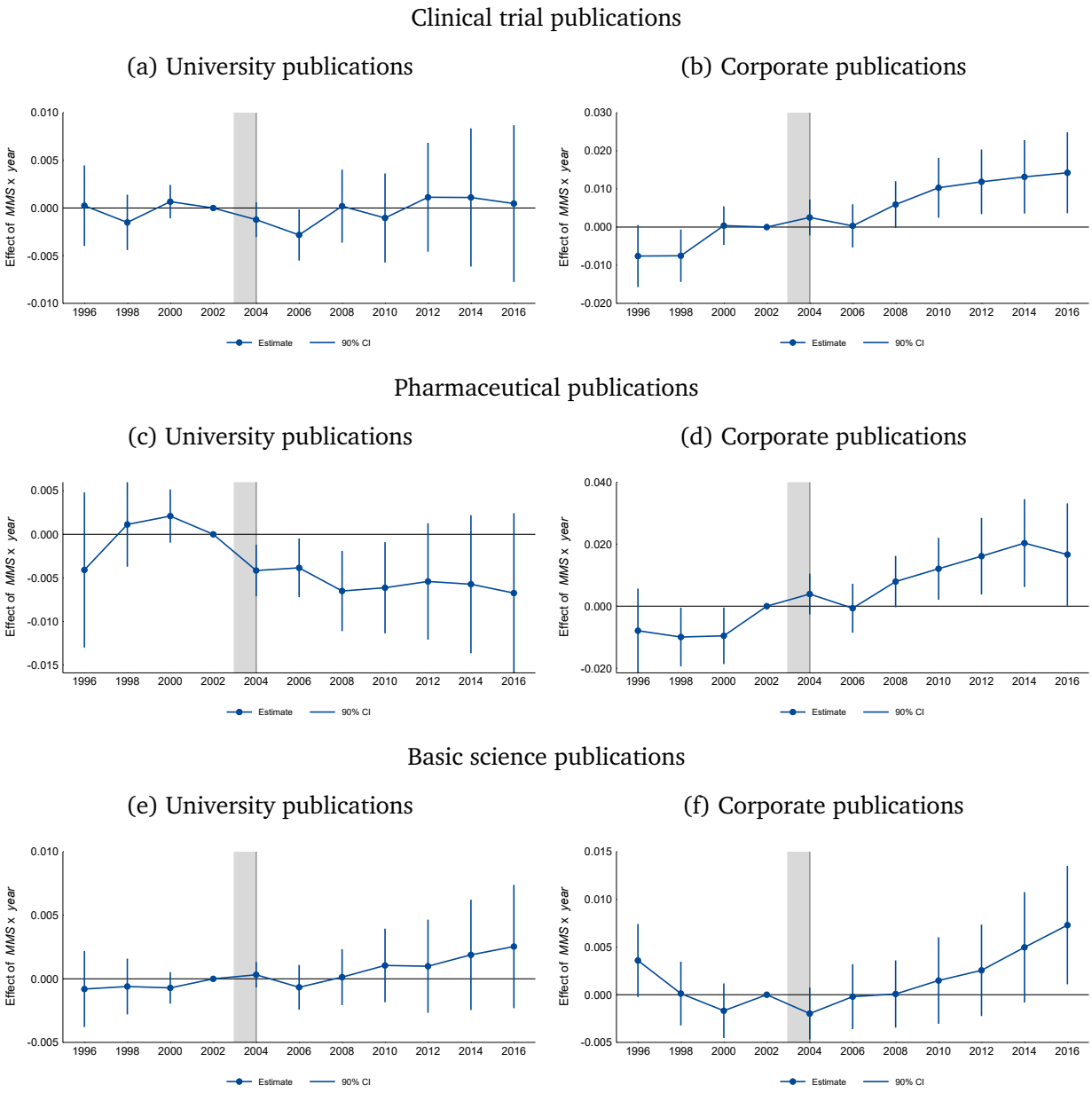
Notes: The figures show the event study estimates of Poisson pseudo-maximum likelihood regressions. The dependent variable is the annual number of scientific publications. In the top figures, the MMS is calculated based on 2003 patients counts. In the middle figures, the MMS is calculated based on prescription counts/quantity. In the bottom figures, the treatment variable is binary based on patients counts. Standard errors are clustered at the MeSH term level.

Figure A-12: Event study – scientific publications (protected classes)



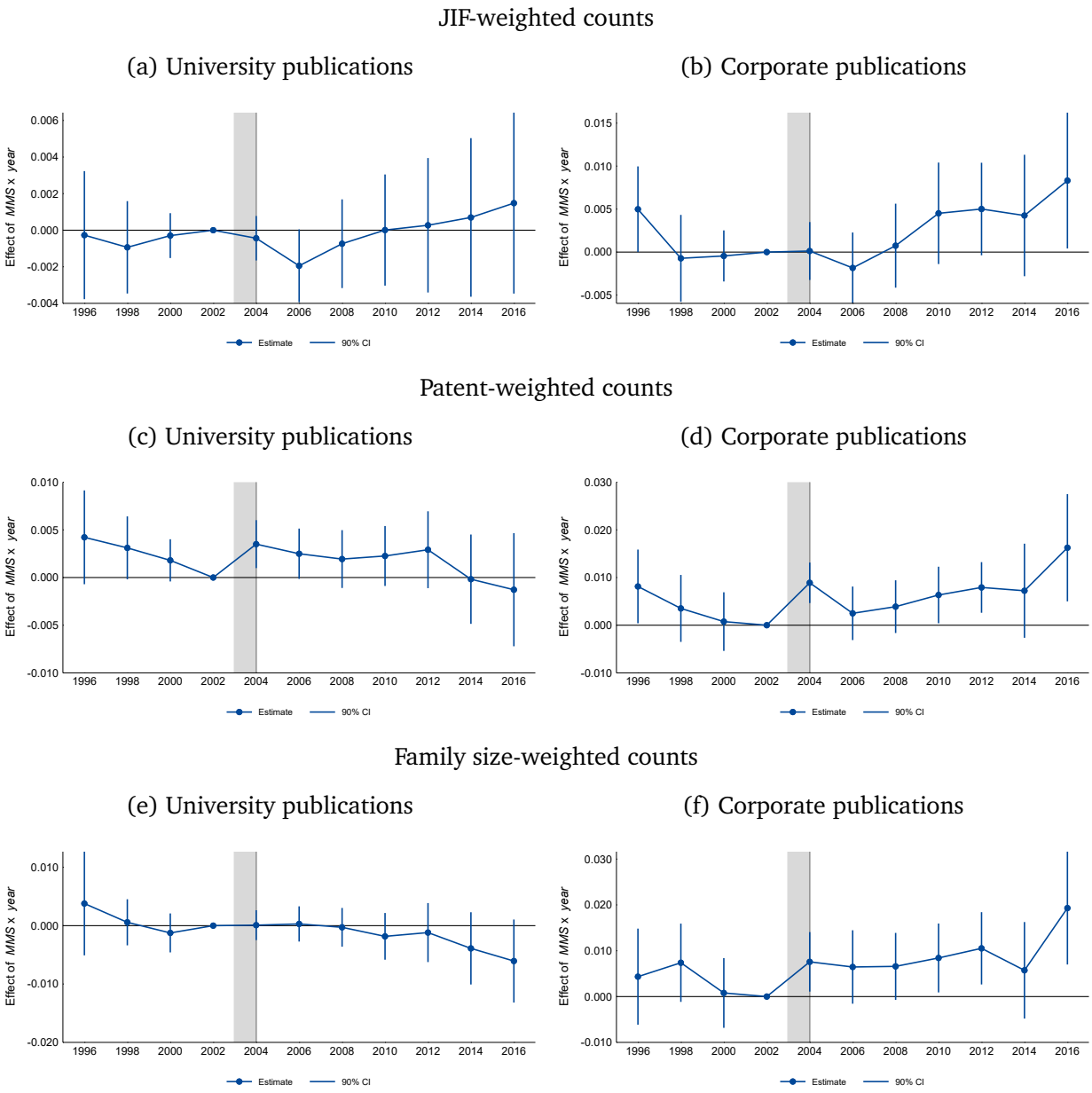
Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the MeSH term level. The dependent variable is the annual number of scientific publications. In the top figures, we *exclude* all ICD-9 groups related to protected drug classes. In the bottom figures, we *include* only ICD-9 groups related to protected drug classes. Standard errors are clustered at the MeSH term level.

Figure A-13: Event study – scientific publications by type of research



Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the MeSH term level. In the top figures, the dependent variable is the annual number of scientific publications, which are associated with a MeSH term related to 'clinical trials'. In the middle figure, the dependent variable is the annual number of scientific publications, which are associated with a MeSH term related to 'pharmaceutical products'. In the bottom figures, the dependent variables is the residual, thus scientific publications neither related to clinical trials nor pharmaceutical products. Standard errors are clustered at the MeSH term level.

Figure A-14: Event study – weighted scientific publications



Notes: The figures show the event study estimates and the 90 percent confidence bands of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects following Equation 1. The unit of observation is the MeSH term level. In the top figures, the dependent variable is the journal impact factor-weighted number of university/corporate scientific publications. In the middle figure, the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). In the bottom figures, we weight the number scientific publications by the size of the average patent family associated with the publication. A patent/family size-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. Standard errors are clustered at the MeSH term level.

C Tables

C.1 Tables – Data Overview

Table A-1: Updated ICD-9/MeSH crosswalk – based on Bhattacharya and Packalen (2011)

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
Infectious and parasitic diseases				
011	011	Pulmonary tuberculosis	Tuberculosis, Pulmonary	D014397
034	034	Streptococcal sore throat/scarlet fever	Scarlet Fever	D012541
052	052	Chickenpox	Chickenpox	D002644
053	053	Herpes zoster	Herpes Zoster	D006562
054	054	Herpes simplex	Herpes Simplex	D006561
070	070	Viral hepatitis	Hepatitis	D006505
075	075	Infectious mononucleosis	Infectious Mononucleosis	D007244
110	110	Dermatophytosis	Tinea	D014005
	111	Dermatomycosis (unspecified)	Tinea Versicolor	D014010
112	112	Candidiasis	Candidiasis	D002177
132	132	Pediculosis/phthirus infestation	Lice Infestations	D010373
133	133	Acariasis	Mite Infestations	D008924
No Match	038	Septicemia		
	074	Specific diseases due to Coxsackie virus		
Neoplasms				
150	150-159	Malignant neoplasm digestive organs	Digestive System Neoplasms	D004067
	211	Benign neoplasm digestive system	Abdominal Neoplasms	D000008
			Anal Gland Neoplasms	D000694
162	162	Malignant neoplasm bronchus/lung	Respiratory Tract Neoplasms	D012142
	163	Malignant neoplasm pleura		
171	171	Malignant melanoma skin	Soft Tissue Neoplasms	D012983
	214	Lipoma		
	215	Benign neoplasm connective tissue		
172	172	Malignant melanoma skin	Skin Neoplasms	D012878
	173	Malignant neoplasm skin		
	216	Benign neoplasm skin		
174	174	Malignant neoplasm female breast	Breast Neoplasms	D001943
	175	Malignant neoplasm male breast		
	217	Benign neoplasm breast		
179	179	Malignant neoplasm uterus	Genital Neoplasms, Female	D005833
	180	Malignant neoplasm cervix uteri	Genital Neoplasms, Male	D005834
	181	Malignant neoplasm placenta	Urologic Neoplasms	D014571

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
	182	Malignant neoplasm body of uterus		
	183	Malignant neoplasm ovary		
	184	Malignant neoplasm female genitals		
	218	Uterine leiomyoma		
	219	Benign neoplasm uterus		
	220	Benign neoplasm ovary		
	221	Benign neoplasm female genitals		
	185	Malignant neoplasm prostate		
	186	Malignant neoplasm testis		
	187	Malignant neoplasm penis/male genitals		
	222	Benign neoplasm male genital organs		
	188	Malignant neoplasm bladder		
	189	Malignant neoplasm kidney		
	223	Benign neoplasm kidney		
200	200-208	Malignant neoplasm lymphatic tissue	Leukemia	D007938
			Lymphoma	D008223
230	230-234	Carcinoma in situ	Carcinoma in Situ	D002278
Endocrine, nutritional and metabolic diseases, and immunity disorders				
240	240	Simple goiter	Goiter	D006042
	241	Nontoxic nodular goiter		
242	242	Thyrotoxicosis with/without goiter	Hyperthyroidism	D006980
243	243	Congenital hypothyroidism	Hypothyroidism	D007037
	244	Acquired hypothyroidism		
250	250	Diabetes mellitus	Diabetes Mellitus	D003920
265	265	Thiamine/niacin deficiency states	Vitamin B Deficiency	D014804
	266	Deficiency B-complex components		
272	272	Disorders lipoid metabolism	Lipid Metabolism Disorders	D052439
274	274	Gout	Gout	D006073
275	275	Disorders mineral metabolism	Hemochromatosis	D006432
			Hepatolenticular Degeneration	D006527
			Hypophosphatemia, Familial	D007015
			Hypercalcemia	D006934
			Hypocalcemia	D006996
276	276	Disorders acid-base balance	Hypokalemia	D007008
			Hypernatremia	D006955
			Acidosis	D000138
			Alkalosis	D000471
279	279	Disorders immune mechanism	Agammaglobulinemia	D000361
			DiGeorge Syndrome	D004062
			Dysgammaglobulinemia	D004406
			Wiskott-Aldrich Syndrome	D014923
No Match	256	Ovarian dysfunction		

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
Diseases of blood and blood-forming organs				
280	280	Iron deficiency anemias	Anemia	D000740
	281	Deficiency anemias		
	282	Hereditary hemolytic anemias		
	283	Acquired hemolytic anemias		
	284	Aplastic anemia		
	285	Anemias		
288	288	Diseases white blood cells	Agranulocytosis	D000380
			Granulomatous Disease, Chronic	D006105
			Eosinophilia	D004802
			Leukocytosis	D007964
Mental disorders				
295	295	Schizophrenic psychoses	Schizophrenia	D012559
296	296	Affective psychoses	Mood Disorders	D019964
	309	Adjustment reaction	Adjustment Disorders	D000275
299	299	Psychoses with origin in childhood	Child Development Disorders	D002659
300	300	Neurotic disorders	Anxiety Disorders	D001008
			Dissociative Disorders	D004213
			Feeding/Eating Disorders	D001068
			Somatoform Disorders	D013001
301	301	Personality disorders	Personality Disorders	D010554
302	302	Sexual deviations/disorders	Sexual and Gender Disorders	D019968
303	303	Alcohol dependence syndrome	Substance-Related Disorders	D019966
	304	Drug dependence		
	305	Nondependent drug abuse		
314	314	Hyperkinetic syndrome (childhood)	Attention Deficit Disorder	D001289
315	315	Specific delays in development	Developmental Disabilities	D002658
			Communication Disorders	D003147
No Match	308	Acute reaction to stress		
	306	Physiological malfunction		
Diseases of the nervous system and sense organs				
320	320	Bacterial meningitis	Meningitis	D008581
	321	Meningitis (other organisms)	Central Nervous System - Viral Diseases	D020805
	322	Meningitis	Myelitis	D009187
	323	Encephalitis/myelitis/encephalomyelitis		
332	332	Parkinson's disease	Parkinsonian Disorders	D020734
340	340	Multiple sclerosis	Multiple Sclerosis	D009103
343	343	Infantile cerebral palsy	Cerebral Palsy	D002547

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
345	345	Epilepsy	Epilepsy	D004827
346	346	Migraine	Migraine Disorders	D008881
350	350-359	Disorders peripheral nervous system	Peripheral Nervous - System Diseases	D010523
361	361 362	Retinal detachments/defects Retinal disorders	Retinal Diseases	D012164
363	360	Disorders of the globe	Uveal Diseases	D014603
363	363	Chorioretinal inflammations/scars		
363	364	Disorders iris/ciliary body		
365	365	Glaucoma	Glaucoma	D005901
366	366	Cataract	Cataract	D002386
367	367	Disorders of refraction	Refractive Errors	D012030
368	368 369	Visual disturbances Blindness/low vision	Vision Disorders	D014786
371	371	Corneal opacity/disorders of cornea	Corneal Diseases	D003316
372	372	Disorders conjunctiva	Conjunctival Diseases	D003229
373	373 374	Inflammation eyelids Disorders eyelids	Eyelid Diseases	D005141
375	375	Disorders lacrimal system	Lacrimal Apparatus Diseases	D007766
380	380	Disorders external ear	Otitis Externa	D010032
381	381 382 383	Nonsuppurative otitis media Suppurative otitis media Mastoiditis/related conditions	Otitis Media	D010033
386	386	Vertiginous synd. vestibular system	Labyrinth Diseases	D007759
389	389	Hearing loss	Hearing Loss	D034381
Diseases of the circulatory system				
401	401-405	Hypertensive disease	Hypertension	D006973
410	410 412	Acute myocardial infarction Old myocardial infarction	Myocardial Infarction	D009203
413	413	Angina pectoris	Angina Pectoris	D000787
414	414 440 441 442	Chronic ischemic heart disease Atherosclerosis Aortic aneurysm/dissection Aneurysm	Arteriosclerosis Aneurysm	D001161 D000783
426	426 427	Conduction disorders Cardiac dysrhythmias	Arrhythmias, Cardiac	D001145
428	428	Heart failure	Heart Failure	D006333

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
430	430-438	Cerebrovascular disease	Cerebrovascular Disorders	D002561
444	444	Arterial embolism/thrombosis	Embolism and Thrombosis	D016769
	451	Phlebitis/thrombophlebitis	Phlebitis	D010689
	452	Portal vein thrombosis		
	453	Venous embolism/thrombosis		
454	454	Varicose veins lower extremities	Varicose Veins	D014648
	456	Varicose veins other sites		
455	455	Hemorrhoids	Hemorrhoids	D006484
458	458	Hypotension	Hypotension	D007022
Diseases of the respiratory system				
460	460	Acute nasopharyngitis	Nasopharyngitis	D009304
	462	Acute pharyngitis	Pharyngitis	D010612
	472	Chronic pharyngitis/nasopharyngitis		
461	461	Acute sinusitis	Sinusitis	D012852
	473	Chronic sinusitis		
463	463	Acute tonsillitis	Tonsillitis	D014069
	474	Chronic disease tonsils/adenoids		
464	464	Acute laryngitis/tracheitis	Laryngitis	D007827
	476	Chronic laryngitis/laryngotracheitis	Tracheitis	D014136
			Epiglottitis	D004826
			Croup	D003440
466	466	Acute bronchitis/bronchiolitis	Bronchitis	D001991
	490	Bronchitis		
	491	Chronic bronchitis		
477	477	Allergic rhinitis	Rhinitis	D012220
480	480	Viral pneumonia	Pneumonia	D011014
	481	Pneumococcal pneumonia		
	482	Bacterial pneumonia		
	483	Pneumonia (other specified organism)		
	484	Pneumonia in infectious diseases		
	485	Bronchopneumonia		
	486	Pneumonia		
	514	Pulmonary congestion/hypostasis		
487	487	Influenza	Influenza, Human	D007251
492	492	Emphysema	Emphysema	D004646
493	493	Asthma	Asthma	D001249
511	511	Pleurisy	Pleurisy	D010998
No Match	470	Deviated nasal septum		
Diseases of the digestive system				
520	520	Disorders tooth development	Tooth Abnormalities	D014071

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
	521	Diseases hard tissues of teeth	Tooth Erosion	D014077
	524	Dentofacial anomalies	Tooth Abrasion	D014072
			Malocclusion	D008310
522	522	Diseases pulp/periapical tissues	Periapical Diseases	D010483
	523	Gingival/periodontal diseases	Dental Pulp Diseases	D003788
			Periodontitis	D010518
			Gingival Diseases	D005882
526	526	Diseases jaws	Jaw Cysts	D007570
			Granuloma, Giant Cell	D006101
527	527	Diseases salivary glands	Salivary Gland Diseases	D012466
528	528	Diseases oral soft tissues	Stomatitis	D013280
			Noma	D009625
530	530	Diseases esophagus	Esophageal Diseases	D004935
531	531	Gastric ulcer	Peptic Ulcer	D010437
	532	Duodenal ulcer	Peptic Ulcer Hemorrhage	D010438
	533	Peptic ulcer	Gastrointestinal Hemorr.	D006471
	534	Gastrojejunal ulcer		
	578	Gastrointestinal hemorrhage		
535	535	Gastritis/duodenitis	Gastritis	D005756
	555-558	Noninfective enteritis/colitis	Duodenitis	D004382
			Enteritis	D004751
			Colitis	D003092
536	536	Disorders function of stomach	Achlorhydria	D000126
			Gastric Dilatation	D013271
			Dyspepsia	D004415
540	540	Acute appendicitis	Appendicitis	D001064
	541	Appendicitis, unqualified		
	542	Appendicitis		
550	550-553	Hernia of abdominal cavity	Hernia	D006547
560	560	Intestinal obstruction	Intestinal Obstruction	D007415
562	562	Diverticula of intestine	Diverticulum, Colon	D004241
			Diverticulum, Stomach	D013273
574	574	Cholelithiasis	Cholelithiasis	D002769
577	577	Diseases pancreas	Pancreatitis	D010195
			Pancreatic Cyst	D010181
No Match	571	Chronic liver disease/cirrhosis		
Diseases of the genitourinary system				
590	590	Infections kidney	Nephritis	D009393
592	592	Calculus kidney/ureter	Nephrolithiasis	D053040
			Ureterolithiasis	D053039

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
595	595	Cystitis	Cystitis	D003556
600	600	Hyperplasia prostate	Prostatic Diseases	D011469
	601	Inflammatory diseases prostate		
	602	Disorders prostate		
607	607	Disorders penis	Penile Diseases	D010409
610	610	Benign mammary dysplasias	Breast Diseases	D001941
	611	Disorders breast		
614	614	Inflammatory disease ovary	Adnexal Diseases	D000291
	620	Noninflammatory disorders ovary		
615	615	Inflammatory diseases uterus	Uterine Diseases	D014591
	616	Inflammatory disease cervix/vagina/vulva	Vaginal Diseases	D014623
	618	Genital prolapse	Vulvar Diseases	D014845
	621	Disorders uterus		
	622	Noninflammatory disorders cervix		
	623	Noninflammatory disorders vagina		
	624	Noninflammatory disorders vulva/perineum		
	625	Pain associated with female genital organs		
617	617	Endometriosis	Endometriosis	D004715
628	628	Infertility, female	Infertility, Female	D007247
No Match	627	Menopausal/postmenopausal disorders		
Diseases of the skin and subcutaneous tissue				
680	680	Carbuncle/furuncle	Furunculosis	D005667
681	681	Cellulitis/abscess finger/toe	Cellulitis	D002481
	682	Cellulitis/abscess		
684	684	Impetigo	Impetigo	D007169
690	690	Erythematosquamous dermatosis	Dermatitis, Seborrheic	D012628
	706	Diseases sebaceous glands	Acne Vulgaris	D000152
691	691	Atopic dermatitis/related conditions	Dermatitis, Atopic	D003876
	692	Contact dermatitis/eczema	Dermatitis, Contact	D003877
696	696	Psoriasis/similar disorders	Psoriasis	D011565
			Pityriasis	D010915
			Parapsoriasis	D010267
698	698	Pruritus/related conditions	Pruritus	D011537
			Prurigo	D011536
			Neurodermatitis	D009450
700	700	Corns/callosities	Callosities	D002145
703	703	Diseases nail	Nail Diseases	D009260
704	704	Diseases hair/hair follicles	Hair Diseases	D006201
705	705	Disorders sweat glands	Sweat Gland Diseases	D013543

ICD group	ICD code	ICD entry/entries	MeSH entry/entries	MeSH ID
708	708	Urticaria	Urticaria	D014581
No Match	707	Chronic ulcer skin		
	695	Erythematous conditions		
	693	Dermatitis (substances taken internally)		
Diseases of the musculoskeletal system and connective tissue				
710	710	Diffuse diseases connective tissue	Sjogrens Syndrome	D012859
	728	Disorders muscle/ligament/fascia	Scleroderma, Systemic	D012595
			Scleroderma, Localized	D012594
			Dermatomyositis	D003882
			Myositis	D009220
715	715	Osteoarthrosis/allied disorders	Osteoarthritis	D010003
	721	Spondylosis/allied disorders		
722	722	Intervertebral disc disorders	Intervertebral Disk Displac.	D007405
726	726	Peripheral enthesopathies	Bursitis	D002062
734	734	Flat foot	Flatfoot	D005413
735	735	Acquired deformities toe	Hallux Valgus	D006215
			Hallux Varus	D050488
737	737	Curvature spine	Spinal Curvatures	D013121
No Match	717	Internal derangement knee		

Table A-2: Summary statistics by high/low MMS

High MMS vs Low MMS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	High MMS (N = 778)			Low MMS (N = 785)				
	Mean	Median	Std. Error	Mean	Median	Std. Error	Diff.	p-value
MMS (cases)	56.22	54.73	14.27	18.29	18.38	9.22	−37.94	0.0000***
MMS (prescriptions)	58.41	61.08	16.01	20.40	18.28	10.60	−38.02	0.0000***
MMS (Rx quantity)	59.23	58.15	17.49	21.60	19.48	12.26	−37.64	0.0000***
Scientific publications	19.16	2.00	69.47	14.59	2.00	69.17	−4.57	0.1922
Publications - fractional	14.85	1.33	60.16	11.38	1.00	57.72	−3.47	0.2454
University publications	15.09	2.00	55.31	11.48	1.00	54.41	−3.61	0.1940
Corporate publications	1.04	0.00	4.25	0.79	0.00	4.56	−0.25	0.2624
Hospital publications	2.63	0.00	8.63	1.97	0.00	8.63	−0.66	0.1299
NIH publications	1.01	0.00	4.32	0.63	0.00	3.94	−0.38	0.0678*
Academic journal publications	0.17	0.00	0.72	0.26	0.00	1.04	0.09	0.0564*
Clinical journal publications	7.41	1.00	26.88	6.36	1.00	26.68	−1.06	0.4359
Clinical-practice journal publications	4.08	0.00	13.16	2.34	0.00	9.90	−1.74	0.0032***
Industry-Clinical journal publications	4.96	0.00	21.49	3.35	0.00	22.34	−1.61	0.1471
Industrial journal publications	0.44	0.00	2.45	0.23	0.00	1.93	−0.20	0.0697*
Industry-practice journal publications	0.17	0.00	0.83	0.09	0.00	0.66	−0.08	0.0387**
NIH funded publications	7.31	0.00	30.65	6.03	0.00	34.94	−1.28	0.4405
Clinical trial university publications	1.25	0.00	5.20	1.28	0.00	6.25	0.02	0.9399
Clinical trial corporate publications	0.21	0.00	0.93	0.21	0.00	1.35	−0.01	0.9222
Pharmaceutical university publications	0.31	0.00	1.50	0.35	0.00	1.95	0.04	0.6346
Pharmaceutical corporate publications	0.07	0.00	0.39	0.07	0.00	0.52	0.00	0.8903
Citation-weighted publications	429.17	19.50	1777.20	278.17	15.00	1503.38	−151.01	0.0698*
JIF-weighted publications	76.29	5.33	285.15	45.61	3.34	245.43	−30.68	0.0227**
Patent-weighted publications	1.47	0.00	7.05	0.63	0.00	5.32	−0.83	0.0085***
Patent-weighted university publications	1.15	0.00	5.57	0.51	0.00	4.33	−0.64	0.0118**
Patent-weighted corporate publications	0.17	0.00	0.82	0.09	0.00	0.68	−0.09	0.0224**
Patent family size-weighted publications	12.79	0.00	59.95	5.59	0.00	45.87	−7.20	0.0077***
Cumul. US Market Size$_{t \text{ to } t+12}$	28 700.49	6105.04	64 374.47	10 778.88	3040.69	20 202.34	−17 921.61	0.0000***
Cumul. NIH funding$_{t-1 \text{ to } -12}$	22.57	21.46	12.50	12.75	11.98	8.21	−9.82	0.0000***
Cumul. New MeSH terms$_{t}$	0.83	0.00	1.01	0.30	0.00	0.66	−0.53	0.0000***

Notes: This table compares observations split at the MMS Median with t-tests. The unit of observation is the MeSH term level in the year 2003. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

C.2 Tables – Multivariate Analysis Clinical Drug Development

Table A-3: Drug development – parallel trends

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)
	Drug Development			
	Early	All Stages	Phase 1-3	Approval
MMS × 2000-02	−0.0014	−0.0008	−0.0017	−0.0035
	(0.002)	(0.002)	(0.003)	(0.009)
ICD9 code FE	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	750	876	702	348
ICD9-codes	125	146	117	58
ICD9-groups	93	100	83	52
Log-likelihood	−1099	−1458	−927	−226

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The time period of these regressions is the pre-MMA period 1997-2002. We employ placebo tests (2000-2002 x MMS with 1997-1999 as the baseline period) in order to test the parallel trends assumption. The dependent variable is the annual number of newly discovered NME in Column (1), NME in all development stages in Column (2), NME in phase I-III clinical trials in Column (3), and the annual number of approved drugs in Column (4). The unit of observation is the ICD-9 three-digit level by year. Standard errors are clustered at the ICD-9 three-digit code level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-4: Drug development – alternative ICD-9 levels

All ICD9 codes Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Drug Development					
	Early Development		All Development Stages		Approval	
MMS (ICD9 level) × 2004-05	−0.0005		0.0005		0.0195***	
	(0.002)		(0.001)		(0.006)	
MMS (ICD9 level) × 2006-08	0.0007		0.0001		0.0052	
	(0.002)		(0.002)		(0.005)	
MMS (ICD9 level) × 2009-11	0.0015		0.0006		0.0112**	
	(0.002)		(0.002)		(0.005)	
MMS (ICD9 level) × 2012-14	0.0032		0.0018		0.0066	
	(0.002)		(0.002)		(0.005)	
MMS (ICD9 level) × 2015-16	0.0051*		0.0037		0.0104**	
	(0.003)		(0.003)		(0.005)	
MMS × 2004-05		−0.0007		−0.0007		0.0256***
		(0.002)		(0.002)		(0.009)
MMS × 2006-08		0.0013		0.0007		0.0070
		(0.003)		(0.002)		(0.008)
MMS × 2009-11		0.0020		0.0012		0.0138
		(0.003)		(0.003)		(0.010)
MMS × 2012-14		0.0053*		0.0040		0.0198*
		(0.003)		(0.003)		(0.012)
MMS × 2015-16		0.0087**		0.0066*		0.0191
		(0.004)		(0.004)		(0.012)
Cumul. US Market Size$_{t \text{ to } t+12}$	No	Yes	No	Yes	No	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	No	Yes	No	Yes	No	Yes
Cumul. New MeSH terms$_{t}$	No	Yes	No	Yes	No	Yes
ICD9 code FE	Yes	No	Yes	No	Yes	No
ICD9 group FE	No	Yes	No	Yes	No	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6980	3800	7740	3880	4260	3220
ICD9-codes	349	190	387	194	213	161
ICD9-groups	349	110	387	114	213	87
Log-likelihood	−9333	−6723	−12764	−10449	−2067	−1316

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. Columns with odd numbers include a **broader set of ICD-9 codes**, namely all ICD-9 three-digit codes available in Cortellis (>300). Here the MMS is calculated at the ICD-9 three-digit code level and controls are excluded for data availability reasons. Columns with even numbers are at the ICD-9 group level again and **include ICD-9 group level fixed effects** instead of ICD-9 three-digit code level fixed effects. Here the MMS is calculated at the ICD-9 group level and log-transformed ICD-9 group level controls are included. The dependent variable is the annual number of newly discovered NME in Columns (1) and (2), NME in all development stages (preclinicals, clinical trials, registrations, approvals) in Columns (3) and (4), and approved NME in Column (5) and (6). In Columns with odd numbers standard errors are clustered at the ICD-9 three-digit code level, in Columns with even numbers at the ICD-9 group level. They are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-5: Drug development – alternative controls

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Drug Development					
	Early Development		All Develop. Stages		Approval	
MMS × 2004-05	−0.0004	−0.0006	−0.0010	−0.0007	0.0275***	0.0241***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.009)	(0.008)
MMS × 2006-08	0.0019	0.0011	0.0004	0.0004	0.0098	0.0046
	(0.003)	(0.002)	(0.002)	(0.002)	(0.009)	(0.007)
MMS × 2009-11	0.0029	0.0012	0.0008	0.0004	0.0179*	0.0096
	(0.003)	(0.002)	(0.003)	(0.002)	(0.010)	(0.008)
MMS × 2012-14	0.0063*	0.0039*	0.0035	0.0027	0.0253**	0.0140
	(0.004)	(0.002)	(0.004)	(0.002)	(0.012)	(0.009)
MMS × 2015-16	0.0099**	0.0066**	0.0060	0.0046	0.0261**	0.0117
	(0.005)	(0.003)	(0.004)	(0.003)	(0.013)	(0.010)
Cumul. OECD Market Size$_{t \text{ to } t+12}$	Yes	No	Yes	No	Yes	No
Cumul. NIH funding (Share)$_{t\text{-}1 \text{ to } -12}$	Yes	No	Yes	No	Yes	No
Cumul. New MeSH terms$_t$	Yes	No	Yes	No	Yes	No
US Market Size$_t$	No	Yes	No	Yes	No	Yes
NIH funding$_{t-1}$	No	Yes	No	Yes	No	Yes
New MeSH terms$_t$	No	Yes	No	Yes	No	Yes
ICD9 code FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3320	3340	3720	3740	2360	2380
ICD9-codes	166	167	186	187	118	119
ICD9-groups	109	110	113	114	86	87
Log-likelihood	−4759	−4795	−6563	−6632	−1176	−1183

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. Column (1), (3), and (5) include alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (2), (4), and (6) include control variables that do not accumulate future/past periods but only consider year *t*. In both cases, control variables are log-transformed and the MMS is calculated at the ICD-9 group level. The unit of observation of the dependent variable is the ICD-9 three-digit code by year. It is the annual number of newly discovered NME in Columns (1) and (2), NME in all development stages (preclinicals, clinical trials, registrations, approvals) in Columns (3) and (4), and approved NME in Column (5) and (6). Standard errors are clustered at the ICD-9 three-digit code level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-6: Drug development – pre-MMS period until 2005

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)
	Drug Development			
	Early	All Stages	Phase 1-3	Approval
MMS × 2006-08	0.0015	0.0009	−0.0003	−0.0025
	(0.002)	(0.002)	(0.002)	(0.007)
MMS × 2009-11	0.0022	0.0014	0.0012	0.0028
	(0.002)	(0.002)	(0.003)	(0.007)
MMS × 2012-14	0.0056**	0.0043*	0.0034	0.0073
	(0.002)	(0.002)	(0.004)	(0.009)
MMS × 2015-16	0.0090***	0.0070**	0.0063	0.0054
	(0.003)	(0.003)	(0.005)	(0.010)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes
ICD9 code FE	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	3340	3740	3360	2380
ICD9-codes	167	187	168	119
ICD9-groups	110	114	110	87
Log-likelihood	−4801	−6634	−4355	−1188

Notes: Columns (1) to (4) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The pre-MMA period includes the years 1997-2005 (until the implementation of the MMA). The dependent variable is the annual number of newly discovered NME in Column (1), NME in all development stages in Column (2), NME in phase I-III clinical trials in Column (3), and the annual number of approved drugs in Column (4). The unit of observation is the ICD-9 three-digit level by year. The control variables are log-transformed. Standard errors are clustered at the ICD-9 three-digit code level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

C.3 Tables – Multivariate Analysis Biomedical Science

Table A-7: Scientific publications – parallel trends

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications					
	All	All	Fractions	University	Corporate	Patent-Weighted
MMS × 2000-02	0.0002		−0.0002	0.0002	−0.0007	−0.0008
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
MMS > Median × 2000-02		0.0239				
		(0.043)				
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7440	7440	7440	7314	3342	3444
MeSH terms	1240	1240	1240	1219	557	574
ICD-group	127	127	127	127	108	107
Log-likelihood	−17026	−17024	−14660	−15270	−3586	−3694

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The time period of these regressions is the pre-MMA period 1997-2002. We employ placebo tests (2000-2002 x MMS with 1997-1999 as the baseline period) in order to test the parallel trends assumption. The dependent variable is the annual number of scientific publications from all affiliations (1)-(2), disease weighted (fractional) number of scientific publications from all affiliations (3), number of scientific publications from at least one author affiliated with an university (4), affiliated with a corporation (5), and patent-weighted number of scientific publications (6). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The unit of observation is the MeSH term by year. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-8: Scientific publications – fractional counts/winsorized

All ICD9 Groups Count/PPML	(1) Scientific Publications	(2)	(3)	(4) Scientific Publications - fractions	(5)	(6)
	winsorized	winsorized	count	count	winsorized	winsorized
MMS × post 2003	0.0001		0.0001		0.0002	
	(0.001)		(0.001)		(0.001)	
MMS × 2004-05		0.0004		0.0005		0.0003
		(0.001)		(0.001)		(0.001)
MMS × 2006-08		0.0000		−0.0001		0.0002
		(0.001)		(0.002)		(0.001)
MMS × 2009-11		0.0012		0.0007		0.0013
		(0.001)		(0.002)		(0.001)
MMS × 2012-14		0.0016		0.0009		0.0015
		(0.002)		(0.003)		(0.002)
MMS × 2015-16		0.0030		0.0018		0.0026
		(0.002)		(0.003)		(0.002)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30680	30680	30680	30680	30680	30680
MeSH terms	1534	1534	1534	1534	1534	1534
ICD-group	129	129	129	129	129	129
Log-likelihood	−87272	−87212	−76337	−76313	−74086	−74055

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications winsorized at the 99 percentile in Columns (1) and (2), the number of scientific publications weighted by the inverse number of linked diseases (fractional counts) in Columns (3) and (4), and the winsorized number of scientific publications weighted by the inverse number of linked diseases in Columns (5) and (6). The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-9: Scientific publications – alternative controls/different models

All ICD9 Groups DV: Scientific Publications	(1)	(2) Count/PPML	(3)	(4) Count/Linear
	ICD Group FE	Different Controls		OLS
MMS × 2004-05	0.0006	0.0014	0.0005	0.0908**
	(0.001)	(0.001)	(0.001)	(0.044)
MMS × 2006-08	−0.0001	0.0010	−0.0004	0.1174*
	(0.001)	(0.002)	(0.001)	(0.068)
MMS × 2009-11	0.0013	0.0030	0.0009	0.2071**
	(0.002)	(0.002)	(0.002)	(0.101)
MMS × 2012-14	0.0017	0.0039	0.0012	0.2695**
	(0.003)	(0.003)	(0.002)	(0.127)
MMS × 2015-16	0.0030	0.0058*	0.0024	0.3414**
	(0.003)	(0.003)	(0.003)	(0.150)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	No	No	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	No	No	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	No	Yes
Cumul. OECD Market Size$_{t \text{ to } t+12}$	No	Yes	No	No
Cumul. NIH funding (Share)$_{t\text{-}1 \text{ to } -12}$	No	Yes	No	No
US Market Size$_t$	No	No	Yes	No
NIH funding$_{t\text{-}1}$	No	No	Yes	No
New MeSH terms$_t$	No	No	Yes	No
ICD9 group FE	Yes	No	No	No
MeSH FE	No	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	31260	30600	30680	31260
MeSH terms	1563	1530	1534	1563
ICD-group	129	126	129	129
Log-likelihood	−894319	−89792	−89963	−149546

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications. In Column (1), we use ICD-9 group level fixed effects instead of MeSH level fixed effects and cluster standard errors at the ICD-9 group level instead of the MeSH level. Column (2) includes alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (3) includes control variables that do not accumulate future/past periods but only consider year *t*. In Column (4), we show the estimates of linear regressions with high-dimensional fixed effects. The control variables are always log-transformed. Standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-10: Scientific publications by affiliation type – all authors from same type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)
	Scientific Publications				
	NIH only	Uni only	No Uni only	Hospital only	Corp only
MMS × 2004-05	−0.0007	0.0003	0.0011	0.0027	−0.0022
	(0.003)	(0.001)	(0.001)	(0.002)	(0.003)
MMS × 2006-08	−0.0056	−0.0003	0.0003	0.0028	0.0028
	(0.004)	(0.002)	(0.001)	(0.003)	(0.003)
MMS × 2009-11	−0.0029	0.0007	0.0012	0.0036	0.0009
	(0.005)	(0.002)	(0.002)	(0.004)	(0.004)
MMS × 2012-14	−0.0035	0.0008	0.0018	0.0042	0.0058
	(0.006)	(0.003)	(0.003)	(0.004)	(0.004)
MMS × 2015-16	−0.0047	0.0022	0.0030	0.0063	0.0081
	(0.008)	(0.003)	(0.003)	(0.006)	(0.005)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes
Observations	9420	29580	28060	17580	8840
MeSH terms	471	1479	1403	879	442
ICD-group	91	129	127	120	111
Log-likelihood	−5704	−53080	−38851	−10902	−4783

Notes: Columns (1) to (5) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications split by all author's affiliations. In Column (1) all authors are affiliated with the NIH, in Column (2) with a university, in Column (4) with an academic medical center, and in Column (5) with a firm. Column (3) includes publications that have all authors not affiliated with a university. The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-11: Scientific publications by affiliation type – US only

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)
	Scientific Publications - US only				
	NIH	University	No University	Hospital	Corporate
MMS × 2004-05	−0.0011	0.0003	0.0006	0.0003	−0.0002
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
MMS × 2006-08	−0.0042	−0.0006	−0.0004	−0.0002	0.0013
	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)
MMS × 2009-11	−0.0026	0.0008	0.0009	0.0011	0.0037
	(0.004)	(0.002)	(0.002)	(0.003)	(0.003)
MMS × 2012-14	−0.0032	0.0007	0.0014	0.0026	0.0064**
	(0.005)	(0.003)	(0.003)	(0.004)	(0.003)
MMS × 2015-16	−0.0027	0.0014	0.0022	0.0030	0.0113***
	(0.006)	(0.003)	(0.003)	(0.004)	(0.004)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes
Observations	14880	30260	29640	25340	17440
MeSH terms	744	1513	1482	1267	872
ICD-group	108	129	128	125	121
Log-likelihood	−11779	−67210	−57989	−27910	−14092

Notes: Columns (1) to (5) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications split by at least one author's affiliation. All authors have US affiliations. In Column (1) at least one author is affiliated with the NIH, in Column (2) with a university, in Column (4) with an academic medical center (AMC), and in Column (5) with a firm. Column (3) includes publications that have at least one author not affiliated with a US university. The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-12: Scientific publications by affiliation type – fractional count/winsorized

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications - fractional				winsorized	
	NIH	University	Hospital	Corporate	University	Corporate
MMS × 2004-05	−0.0008	0.0003	−0.0003	−0.0003	0.0003	−0.0005
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
MMS × 2006-08	−0.0016	−0.0004	−0.0003	0.0012	−0.0001	−0.0002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
MMS × 2009-11	−0.0017	0.0005	0.0003	0.0024	0.0010	0.0026
	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)
MMS × 2012-14	−0.0023	0.0005	0.0009	0.0048	0.0015	0.0037
	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)
MMS × 2015-16	−0.0013	0.0013	0.0016	0.0081*	0.0029	0.0068**
	(0.005)	(0.003)	(0.004)	(0.004)	(0.002)	(0.003)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16880	30560	27500	19880	30560	19880
MeSH terms	844	1528	1375	994	1528	994
ICD-group	111	129	128	125	129	125
Log-likelihood	−12559	−67979	−30470	−15438	−77648	−17708

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. In Columns (1) to (4), the dependent variable is the annual number of scientific publications weighted by the inverse number of linked diseases (fractional counts) split by at least one author's affiliation. In Columns (5) to (6), the dependent variable is the annual number of scientific publications winsorized at the 99 percentile. The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-13: University publications – alternative controls/different models

All ICD9 Groups DV: University Publications	(1)	(2) Count/PPML	(3)	(4) Count/Linear
	ICD Group FE	Different Controls		OLS
MMS × 2004-05	0.0004	0.0012	0.0004	0.0808**
	(0.001)	(0.001)	(0.001)	(0.040)
MMS × 2006-08	−0.0005	0.0009	−0.0006	0.1059*
	(0.001)	(0.002)	(0.001)	(0.062)
MMS × 2009-11	0.0010	0.0029	0.0007	0.1885**
	(0.002)	(0.002)	(0.002)	(0.092)
MMS × 2012-14	0.0012	0.0038	0.0009	0.2456**
	(0.003)	(0.003)	(0.002)	(0.116)
MMS × 2015-16	0.0025	0.0057*	0.0021	0.3103**
	(0.003)	(0.003)	(0.003)	(0.136)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	No	No	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	No	No	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	No	Yes
Cumul. OECD Market Size$_{t \text{ to } t+12}$	No	Yes	No	No
Cumul. NIH funding (Share)$_{t\text{-}1 \text{ to } -12}$	No	Yes	No	No
US Market Size$_{t}$	No	No	Yes	No
NIH funding$_{t\text{-}1}$	No	No	Yes	No
New MeSH terms$_{t}$	No	No	Yes	No
ICD9 group FE	Yes	No	No	No
MeSH FE	No	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	31260	30480	30560	31260
MeSH terms	1563	1524	1528	1563
ICD-group	129	126	129	129
Log-likelihood	−742561	−79872	−80013	−146052

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications with at least one author affiliated with an university. In Column (1), we use ICD-9 group level fixed effects instead of MeSH level fixed effects and cluster standard errors at the ICD-9 group level instead of the MeSH level. Column (2) includes alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (3) includes control variables that do not accumulate future/past periods but only consider year *t*. In Column (4), we show the estimates of linear regressions with high-dimensional fixed effects. The control variables are always log-transformed. Standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-14: Corporate publications – alternative controls/different models

All ICD9 Groups DV: Corporate Publications	(1)	(2) Count/PPML	(3)	(4) Count/Linear
	ICD Group FE	Different Controls		OLS
MMS × 2004-05	−0.0001	0.0012	−0.0012	0.0047
	(0.002)	(0.002)	(0.002)	(0.003)
MMS × 2006-08	0.0010	0.0031	−0.0005	0.0088*
	(0.002)	(0.002)	(0.002)	(0.005)
MMS × 2009-11	0.0036	0.0065**	0.0014	0.0151**
	(0.003)	(0.003)	(0.002)	(0.007)
MMS × 2012-14	0.0059	0.0096***	0.0031	0.0206**
	(0.004)	(0.003)	(0.003)	(0.009)
MMS × 2015-16	0.0098*	0.0142***	0.0064*	0.0295**
	(0.005)	(0.004)	(0.004)	(0.011)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	No	No	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	No	No	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	No	Yes
Cumul. OECD Market Size$_{t \text{ to } t+12}$	No	Yes	No	No
Cumul. NIH funding (Share)$_{t\text{-}1 \text{ to } -12}$	No	Yes	No	No
US Market Size$_{t}$	No	No	Yes	No
NIH funding$_{t\text{-}1}$	No	No	Yes	No
New MeSH terms$_{t}$	No	No	Yes	No
ICD9 group FE	Yes	No	No	No
MeSH FE	No	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes
Observations	31120	19860	19880	31260
MeSH terms	1556	993	994	1563
ICD-group	125	124	125	129
Log-likelihood	−68090	−18191	−18233	−68469

Notes: Columns (1) to (3) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications with at least one author affiliated with a corporation. In Column (1), we use ICD-9 group level fixed effects instead of MeSH level fixed effects and cluster standard errors at the ICD-9 group level instead of the MeSH level. Column (2) includes alternative control variables such as OECD market size and NIH funding calculated as a share of all publications in a disease category that acknowledge a specific Institute/Center. Column (3) includes control variables that do not accumulate future/past periods but only consider year *t*. In Column (4), we show the estimates of linear regressions with high-dimensional fixed effects. The control variables are always log-transformed. Standard errors are shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-15: University publications by journal type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications					
	Academic	Clinical	Clin. Pract	Ind./Clin.	Industrial	Ind. Pract
MMS × 2004-05	0.0012	−0.0011	0.0017	0.0020	0.0006	0.0050
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)
MMS × 2006-08	−0.0052**	−0.0023	0.0011	0.0002	0.0011	0.0015
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
MMS × 2009-11	−0.0038	−0.0001	0.0017	0.0008	0.0026	−0.0012
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.005)
MMS × 2012-14	−0.0027	−0.0004	0.0027	0.0011	−0.0012	0.0020
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)
MMS × 2015-16	−0.0037	−0.0002	0.0049	0.0018	−0.0025	0.0030
	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.006)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	12020	28540	26660	23640	11840	6560
MeSH terms	601	1427	1333	1182	592	328
ICD-group	111	127	128	124	107	86
Log-likelihood	−7563	−49038	−34207	−31761	−7866	−3667

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the number of scientific publications with at least one author affiliated with an university in academic journals in Column (1), in clinically relevant journals in Column (2), in clinical practice journals in Column (3), in industry-clinical journals in Column (4), in industrial journals in Column (5), and in industry practice journals in Column (6). Journal classification is based on the proportion of published research coming from general hospitals and industry using the publicly available data set provided by Tijssen (2010). The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-16: Corporate publications by journal type

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications					
	Academic	Clinical	Clin. Pract	Ind./Clin.	Industrial	Ind. Pract
MMS × 2004-05	0.0015	−0.0037	0.0029	0.0017	−0.0003	−0.0046
	(0.015)	(0.002)	(0.004)	(0.002)	(0.005)	(0.005)
MMS × 2006-08	−0.0048	−0.0011	0.0069	0.0003	−0.0054	−0.0109*
	(0.013)	(0.003)	(0.004)	(0.003)	(0.004)	(0.006)
MMS × 2009-11	−0.0129	0.0031	0.0038	0.0021	0.0056	−0.0127**
	(0.014)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)
MMS × 2012-14	−0.0113	0.0024	0.0085	0.0062	−0.0080	−0.0046
	(0.017)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
MMS × 2015-16	−0.0205	0.0060	0.0133*	0.0101	−0.0131	0.0035
	(0.023)	(0.006)	(0.007)	(0.007)	(0.008)	(0.008)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2260	14280	9940	11860	3820	3620
MeSH terms	113	714	497	593	191	181
ICD-group	53	109	111	111	76	66
Log-likelihood	−598	−9329	−5287	−9054	−1452	−1587

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the number of scientific publications with at least one author affiliated with a corporation in academic journals in Column (1), in clinically relevant journals in Column (2), in clinical practice journals in Column (3), in industry-clinical journals in Column (4), in industrial journals in Column (5), and in industry practice journals in Column (6). Journal classification is based on the proportion of published research coming from general hospitals and industry using the publicly available data set provided by Tijssen (2010). The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-17: Academic Medical Center publications – type of research/impact

All ICD9 Groups Count/PPML	(1)	(2) MeSH-Weighted	(3)	(4) JIF-Weighted	(5) Patent-Weighted	(6)
Hospital	Basic	CT	Pharma	JIF	Patent	Family Size
MMS × 2004-05	0.0005	−0.0015	−0.0029	−0.0012	−0.0011	−0.0053
	(0.001)	(0.002)	(0.004)	(0.002)	(0.003)	(0.004)
MMS × 2006-08	−0.0001	−0.0003	−0.0037	−0.0015	0.0028	0.0042
	(0.002)	(0.003)	(0.005)	(0.002)	(0.004)	(0.004)
MMS × 2009-11	0.0014	0.0005	−0.0016	−0.0005	0.0023	0.0030
	(0.002)	(0.004)	(0.005)	(0.003)	(0.005)	(0.006)
MMS × 2012-14	0.0019	0.0029	0.0025	0.0010	0.0049	0.0061
	(0.003)	(0.005)	(0.007)	(0.004)	(0.005)	(0.006)
MMS × 2015-16	0.0029	0.0038	0.0015	0.0021	0.0019	−0.0003
	(0.003)	(0.006)	(0.008)	(0.005)	(0.006)	(0.008)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27220	14880	8100	27400	9780	9780
MeSH terms	1361	744	405	1370	489	489
ICD-group	128	117	105	128	104	104
Log-likelihood	−34582	−10813	−3947	−116402	−5186	−26975

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the number of scientific publications from at least one author affiliated with an academic medical center that are associated with MeSH terms related to clinical trials in Column (2) and pharmaceutical products in Column (3). In Column (1), we include the residual academic medical center publications (non CT/non pharmaceutical products). In Columns (4), the dependent variable is the journal impact factor-weighted number of scientific publications from at least one author affiliated with an academic medical center. In Column (5), the dependent variable is the number of academic medical center publications that are associated with at least one patent application (patent-weighted). In Column (6), we weight the number of academic medical center publications by the size of the average patent family associated with the publication. A patent/family size-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-18: Scientific publications – pre-MMS period until 2005

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications				Patent-Weighted	
	All	Fractions	University	Corporate	University	Corporate
MMS × 2006-08	−0.0004	−0.0003	−0.0006	0.0011	0.0001	−0.0009
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
MMS × 2009-11	0.0010	0.0005	0.0008	0.0037	0.0002	−0.0003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
MMS × 2012-14	0.0014	0.0006	0.0010	0.0060**	−0.0002	0.0025
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
MMS × 2015-16	0.0027	0.0015	0.0023	0.0098**	−0.0009	0.0026
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.005)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_t$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30680	30680	30560	19880	16460	7960
MeSH terms	1534	1534	1528	994	823	398
ICD-group	129	129	129	125	114	92
Log-likelihood	−89928	−76315	−79982	−18220	−13168	−4423

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The pre-period lasts until 2005, which is the year before the implementation of the MMA. The dependent variable is the annual number of scientific publications from all affiliation types in Column (1), weighted by the inverse number of linked diseases in Column (2), with at least one author being affiliated with an university in Column (3), and with a corporation in Column (4). In Columns (5) and (6), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-19: Scientific publications – all PMID (incl. confounded)

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications				Patent-Weighted	
	All	Fractions	University	Corporate	University	Corporate
MMS × 2004-05	0.0014*	0.0009	0.0013*	0.0011	0.0034***	0.0044***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
MMS × 2006-08	0.0007	0.0003	0.0005	0.0006	0.0024*	0.0028
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
MMS × 2009-11	0.0014	0.0010	0.0014	0.0018	0.0019	0.0022
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
MMS × 2012-14	0.0016	0.0011	0.0014	0.0035	0.0016	0.0063**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
MMS × 2015-16	0.0026	0.0019	0.0025	0.0065**	0.0033	0.0098**
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.004)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes	Yes	Yes
MeSH FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31200	31200	31160	24460	20940	10820
MeSH terms	1560	1560	1558	1223	1047	541
ICD-group	129	129	129	128	122	107
Log-likelihood	−142854	−95031	−125441	−28223	−20699	−6861

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the MeSH term by year. The dependent variable is the annual number of scientific publications including PMID, which are associated with a disease MeSH term with unknown MMS (confounded). It includes publications from all affiliation types in Column (1), weighted by the inverse number of linked diseases in Column (2), with at least one author being affiliated with an university in Column (3), and with a corporation in Column (4). In Columns (5) and (6), the dependent variable is the number of university/corporate scientific publications (including confounded PMID) that are associated with at least one patent application (patent-weighted). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the MeSH level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A-20: Scientific publications – ICD-9 group level

All ICD9 Groups Count/PPML	(1)	(2)	(3)	(4)	(5)	(6)
	Scientific Publications				Patent-Weighted	
	All	Fractions	University	Corporate	University	Corporate
MMS × 2004-05	0.0005	0.0004	0.0003	−0.0002	0.0013	0.0056*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
MMS × 2006-08	−0.0003	−0.0002	−0.0006	0.0010	0.0002	0.0018
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
MMS × 2009-11	0.0007	0.0008	0.0005	0.0030	0.0003	0.0013
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
MMS × 2012-14	0.0010	0.0011	0.0006	0.0052	−0.0003	0.0038
	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
MMS × 2015-16	0.0021	0.0022	0.0016	0.0082*	−0.0011	0.0031
	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.005)
Cumul. US Market Size$_{t \text{ to } t+12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. NIH funding$_{t\text{-}1 \text{ to } -12}$	Yes	Yes	Yes	Yes	Yes	Yes
Cumul. New MeSH terms$_{t}$	Yes	Yes	Yes	Yes	Yes	Yes
ICD group FE	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2580	2580	2580	2500	2280	1840
ICD-group	129	129	129	125	114	92
Log-likelihood	−13270	−12351	−12072	−5187	−3917	−1931

Notes: Columns (1) to (6) show the estimates of Poisson pseudo-maximum likelihood regressions with high-dimensional fixed effects. The unit of observation is the ICD-9 group level (MeSH terms aggregated to the ICD-9 group in the Bhattacharya and Packalen (2011) ICD-9-MeSH crosswalk presented in Table A-1). The dependent variable is the annual number of scientific publications from all affiliation types in Column (1), weighted by the inverse number of linked diseases in Column (2), with at least one author being affiliated with an university in Column (3), and with a corporation in Column (4). In Columns (5) and (6), the dependent variable is the number of university/corporate scientific publications that are associated with at least one patent application (patent-weighted). A patent-weight is calculated based on the patent family's first application being filed within 5 years from the scientific publication. The control variables are log-transformed. Standard errors are clustered at the ICD-9 group level and shown in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.