Internet Appendix for "Competition and Innovation Revisited: A Project-Level View"

The Internet Appendix is structured as follows:

- Appendix I.A: ICD-10 Therapeutic Markets: Background and Description
- Appendix I.B: A Description of the Cortellis Database and the Processing Procedure
- Appendix I.C: Identifying BTD Designation Events
- Appendix I.D: Descriptions of the Samples and Measures Referenced in the Paper
- Appendix I.E: Supplementary Results for the Phase-II Development Tests
- References in the Internet Appendix

Appendix I.A: Description of ICD-10 Markets

In Section I.A.1 of this appendix, we discuss the reasoning behind our choice to define therapeutic markets using 2nd subchapter ICD-10 codes. Next, we provide a general description of these markets in Table I.A.1 in Section I.A.2. Finally, Table I.A.2 in Section I.A.3 provides examples of competition between rivals in a therapeutic market.

Section I.A.1: A Discussion on the Advantages of using 2nd Subchapter ICD-10 Codes to Define Therapeutic Markets

In our main tests, we define therapeutic markets using ICD-10 codes defined at the second subchapter (herein 2SC-ICD-10)[1]. We believe these (2SC-ICD-10) codes offer several advantages relative to other therapeutic market definitions (e.g., 1st subchapter ICD-10 (1SC-ICD-10) codes and ICD-9 codes). First, our objective is to identify rivals who operate in the same market where a BTD was awarded. It is therefore critical to ensure that these rival drug projects are substitutes of the BTD drug. 2SC-ICD-10 codes allow for the identification of rival projects that address the same exact indication. Whereas 1SC-ICD-10 codes and ICD-9 codes lump together a group of similar markets, with potentially different characteristics. For example, while myeloid leukemia (2SC-ICD-10 is "C92-0") is a very common type of cancer among adults and is therefore a very competitive indication, few drugs under development address the rare subtype promyelocytic leukemia (2SC-ICD-10 of "C92-4"). Both myeloid and promyelocytic leukemias have the same 1SC-ICD-10 ("C92") and ICD-9 code ("205"). Moreover, the FDA approves drugs for specific indications. This means that a drug cannot be prescribed for both types of leukemia unless it successfully completes separate clinical trials, and obtains FDA approval for each indication[2]. In addition, we find, in untabulated results, that the announcement returns of rivals around BTD announcements were significantly more negative when therapeutic markets were defined on 2SC-ICD-10 codes, relative to markets defined

[1] If an indication only has an ICD-10 code at the first subchapter level, (e.g., essential hypertension has an ICD-10 code of I10), we use the first subchapter designation (rather than deleting observations).

[2] Drugs can be prescribed "off-label" without FDA approval, however, with the exception of rare cases, the best available treatments for an indication are always drugs approved for that indication. Furthermore, firms with effective off-label treatments for an indication risk forgoing significant revenues because off-label uses are less known and many physicians avoid prescribing off-label therapies due to potential litigation risk.

on either of 1SC-ICD-10 or ICD-9 codes, suggesting that 2SC-ICD-10 codes may be identifying substitutes more accurately. Second, while 1SC-ICD-10 codes can vary on some dimensions (as illustrated in the leukemia example above), they also share many characteristics, e.g., subtypes of the same general disease, target the same organ or cells, use similar target-based actions (i.e., technologies) to cure the condition, cause similar symptoms…etc. This motivates our decision to stratify the sample in the main (phase-II development) Cox proportional hazard tests by 1SC-ICD-10 codes. This is also consistent with the model specifications of the phase-II hazard tests used in Krieger (2021). Doing so allows us to make "apples to apples" comparisons[3]. As a reminder, the sample in our main tests defines therapeutic markets, rivals and BTD events using 2SC-ICD-10 codes. That is, the treatment effect is based on 2SC-ICD-10 codes, while the counterfactuals are based on 1SC-ICD-10 codes.

[3] For example, Lymphoma (2SC-ICD-10 is "C85-0") has 3 other variations: high grade b-cell lymphoma (2SC-ICD10 is "C85-10"); primary mediastinal large b-cell lymphoma (2SC-ICD-10 is "C85-20"); non-hodgkins lymphoma (2SC-ICD-10 is "C85-9"). Of these, only the last two experience BTD entry. This allows the treatment effect to be estimated relative to comparable counterfactuals (i.e., phase-II projects in the first two (never-shocked) markets) of the same age, stage of development, and with similar market characteristics.

Section I.A.2: Description of the ICD-10 Therapeutic Markets

Table I.A.1: Description of ICD-10 markets.

This table provides information on the medical conditions in each broad ICD-10 category. 1SC-ICD-10 codes that address a common medical condition are grouped together in each row. Source: "International Classification of Diseases 10th Revision". World Health Organization.

ICD-10 Market Code	ICD-10 Market Code Description
A00–B99	Certain infectious and parasitic diseases
C00–D48	Neoplasms
D50–D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00–E90	Endocrine, nutritional and metabolic diseases
F00–F99	Mental and behavioral disorders
G00–G99	Diseases of the nervous system
H00–H59	Diseases of the eye and adnexa
H60–H95	Diseases of the ear and mastoid process
I00–I99	Diseases of the circulatory system
J00–J99	Diseases of the respiratory system
K00–K93	Diseases of the digestive system
L00–L99	Diseases of the skin and subcutaneous tissue
M00–M99	Diseases of the musculoskeletal system and connective tissue
N00–N99	Diseases of the genitourinary system
O00–O99	Pregnancy, childbirth and the puerperium
P00–P96	Certain conditions originating in the perinatal period
Q00–Q99	Congenital malformations, deformations and chromosomal abnormalities
R00–R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
S00–T98	Injury, poisoning and certain other consequences of external causes
V01–Y98	External causes of morbidity and mortality
Z00–Z99	Factors influencing health status and contact with health services

Section I.A.3: Examples of Competition Between Rivals in a Therapeutic Markets

Table I.A.3: Examples on Between-Patent Competition of BTD-Awarded Drugs and Rivals

This table provides examples from our sample that illustrate rival competition and entry into markets where a drug was awarded a BTD. In all three panels below, the first row (displayed in bold font) provides information on the BTD drug and the subsequent rows provide information on rival drugs targeting the same market. The BTD-awarded drugs below were also covered by patents and received FDA approval shortly after the award was granted. The information provided for each drug includes the following: same of the firm developing the drug (Firm), name of the drug (Drug), the date on which the drug received FDA approval (FDA Aprvl Date), the date on which the first patent was issued for a drug (Patent First Issued), the date on which the last patent covering the drug will expire (Patent Last Expiring), and the target action of a drug (the technology used by the drug to produce a therapeutic effect. The Hepatitis-C therapeutic market is the focus in Panel A, the Metastatic Breast Cancer therapeutic market in Panel B and the Hepatocellular Carcinoma therapeutic market in Panel C.

Panel A: Competition in the Hepatitis C Therapeutic Market

Firm	Drug	FDA Aprvl Date	Patent First Issued	Patent Last Expiring	Target action
Gilead sciences	**Harvoni (Awarded a BTD on 7/25/2013)**	**2013Q4**	**2008Q3**	**2031Q3**	**HEPATITIS C VIRUS NS5B POLYMERASE INHIBITOR; HEPATITIS C VIRUS PROTEIN NS5A INHIBITOR**
AbbVie	PARITAPREVIR + RITONAVIR + OMBITASVIR	2014Q4	2012Q2	2030Q2	HEPATITIS C VIRUS NS3 PROTEASE INHIBITOR; HEPATITIS C VIRUS NS5B POLYMERASE INHIBITOR; HEPATITIS C VIRUS PROTEIN NS5A INHIBITOR
AbbVie	DASABUVIR	2014Q4	2012Q2	2030Q2	HEPATITIS C VIRUS NS5B POLYMERASE INHIBITOR
AbbVie	PARITAPREVIR + RITONAVIR + OMBITASVIR + DASABUVIR	2016Q4	2012Q2	2030Q2	HEPATITIS C VIRUS NS3 PROTEASE INHIBITOR; HEPATITIS C VIRUS NS5B POLYMERASE INHIBITOR; HEPATITIS C VIRUS PROTEIN NS5A INHIBITOR
AbbVie	GLECAPREVIR + PIBRENTASVIR	2017Q3	2012Q2	2030Q2	HEPATITIS C VIRUS NS3 PROTEASE INHIBITOR; HEPATITIS C VIRUS PROTEIN NS5A INHIBITOR
Merck	GRAZOPREVIR + ELBASVIR	2016Q1	2011Q3	2029Q3	HEPATITIS C VIRUS NS3 PROTEASE INHIBITOR; HEPATITIS C VIRUS PROTEIN NS5A INHIBITOR

Panel B: Competition in the Metastatic Breast Cancer Therapeutic Market

Firm	Drug	FDA Aprvl Date	Patent First Issued	Patent Last Expiring	Target action
PFIZER	**Ibrance (awarded a BTD on 4/10/2013)**	**2015Q1**	**2008Q1**	**2023Q1**	**CYCLIN-DEPENDENT KINASE-4 INHIBITOR; CYCLIN-DEPENDENT KINASE-6 INHIBITOR; RETINOBLASTOMA ASSOCIATED PROTEIN MODULATOR**
NOVARTIS	Tykerb	2015Q1	2002Q2	2022Q1	EPIDERMAL GROWTH FACTOR RECEPTOR ANTAGONIST; ERBB2 TYROSINE KINASE RECEPTOR INHIBITOR
R-PHARM	EPOTHILONE	2016Q2	2003Q3	2018Q4	TUBULIN MODULATOR
NOVARTIS	Kisqali	2017Q1	2013Q2	2031Q2	CYCLIN-DEPENDENT KINASE-4 INHIBITOR; CYCLIN-DEPENDENT KINASE-6 INHIBITOR
ANI PHARMA.	ARIMIDEX	2017Q4	1990Q2	2010Q2	AROMATASE INHIBITOR
ELI LILLY & CO	VERZENIO	2017Q4	2010Q4	2029Q4	CYCLIN-DEPENDENT KINASE-4 INHIBITOR; CYCLIN-DEPENDENT KINASE-6 INHIBITOR; PIM-1 PROTEIN KINASE INHIBITOR
ASTRAZENECA	OLAPARIB	2018Q1	2006Q3	2028Q1	PARP INHIBITOR; POLY ADP RIBOSE POLYMERASE 1 INHIBITOR; POLY ADP RIBOSE POLYMERASE 2 INHIBITOR; POLY ADP RIBOSE POLYMERASE 3 INHIBITOR
ROCHE	TECENTRIQ	2019Q1	2012Q3	2030Q1	PROGRAMMED CELL DEATH LIGAND 1 INHIBITOR
ROCHE	HERCEPTIN HYLECTA	2019Q2	2010Q3	2034Q2	ERBB2 TYROSINE KINASE RECEPTOR INHIBITOR; HYALURONIDASE STIMULATOR

Panel C: Competition in the Hepatocellular Carcinoma Therapeutic Market

Firm	Drug	FDA Aprvl date	Patent First Issued	Patent Last Expiring	Target action
Merck	Keytruda (Awarded a BTD in 7/15/2018)	2019Q3	2009Q3	2031Q2	PROGRAMMED CELL DEATH PROTEIN 1 INHIBITOR
Bristol-Myers Squibb	Yervoy	2020Q1	2014Q3	2020Q3	CYTOTOXIC T-LYMPHOCYTE PROTEIN-4 INHIBITOR
Roche	TECENTRIQ	2020Q2	2012Q3	2030Q1	PROGRAMMED CELL DEATH LIGAND 1 INHIBITOR
Roche	ALTUZAN	2020Q2	2003Q2	2019Q1	VEGF LIGAND INHIBITOR

Appendix I.B: A Description of the Cortellis Drug Development Database and the Processing Procedure

Table I.B.1 of Section I.B.1 provides an example of the drug development information included in the Cortellis database. Section I.B.2 describes the procedure used to identify the drug developing firms.

Section I.B.1: An Example of Drug Development History in Cortellis

Table I.B.1: An Example of Drug Development History in Cortellis

This table provides an example of the development milestone information listed in Cortellis. Target Actions, i.e., a drug's technology, is the molecule in the body upon which a drug performs its function. Extract provides a summary on a drug's ownership status, target markets and technology. DevelopmentStatusCurrent presents the most recent development on the drug-indication pairing. DevelopmentStatusHistory lists all previous developments for a drug-indication.

Drug Name	Originator	Active Companies	Target Actions	Inactive Companies	Extract	DevelopmentStatusCurrent	DevelopmentStatusHistory
IDE-196	Novartis	IDEAYA Biosciences	Protein kinase inhibitor	Novartis	IDEAYA Biosciences under license from Novartis Pharmaceuticals is developing IDE-196 (previously LXS-196; NVP-LXS-196), an oral immediate release tablet formulation, a protein kinase C inhibitor, for the potential treatment of metastatic uveal melanoma (MUM), solid tumors including, cutaneous melanoma, and colorectal cancer.	IDEAYA Biosciences Inc: US: Phase 2 Clinical: Colorectal tumor: 25-Jun-2019 IDEAYA Biosciences Inc: US: Phase 2 Clinical: Solid tumor: 25-Jun-2019 Novartis Pharmaceuticals Corp: US: Outlicensed: Uveal melanoma: 23-Oct-2018 IDEAYA Biosciences Inc: US: Phase 2 Clinical: Uveal melanoma: 25-Jun-2019 IDEAYA Biosciences Inc: US: Phase 2 Clinical: Melanoma: 25-Jun-2019	Novartis Pharmaceuticals Corp: US: Discovery: Uveal melanoma: 06-Nov-2015 Novartis Pharmaceuticals Corp: US: Phase 1 Clinical: Uveal melanoma: 01-Feb-2016 IDEAYA Biosciences Inc: US: Preclinical: Colorectal tumor: 23-Oct-2018 IDEAYA Biosciences Inc: US: Preclinical: Melanoma: 23-Oct-2018 IDEAYA Biosciences Inc: US: Preclinical: Solid tumor: 23-Oct-2018 IDEAYA Biosciences Inc: US: Phase 1 Clinical: Uveal melanoma: 23-Oct-2018

Section I.B.2: Identifying Drug Developing Firms in the Cortellis Database

Cortellis provides information the following for a drug: drug names, target therapeutic market, originating firm, current and previous owners, sales in 2018 through 2023 and expected sales through 2029, target action (i.e., technology), regulatory designations (e.g. breakthrough designation and priority review designation), patent status, and (most importantly) detailed history of key development milestones and dates.

Coretllis Data lists the originator firm for each drug. It also lists firms that are actively developing the drug, and (inactive) firms who had previously developed the drug. In addition, the "Extract" field contains elaborate information on the ownership of the drug, and whether the originating firm was acquired, is a subsidiary of another firm, was spun-off by another firm, or whether the firm changed its name. However, it does not always list the dates on which a drug's ownership changed.

We use several resources to identify the correct owner(s) of a drug in different on a given date. We obtain merger deal records from SDC platinum. We retrieve all merger deals occurring between January 2000 to December 2021 with targets in the biopharmaceutical industry. We also search Google to identify firms and their owners.

Our matching strategy is similar in spirit to that of Cunningham et al (2021):

1. We use the comp_stnd package in Stata (Wasi and Flaeen (2013)) to standardize company names in both Cortellis and SDC. We further clean company names from European entity types (e.g., SAS, AB, NV…etc.), and from the words "Pharmaceuticals, Pharmaceutical, and Pharma."
2. We match firms listed Cortellis's Active Company- and Inactive Company fields to SDC merger deals using acquirer and target firm names, and retain only exact matches. We manually check all matches to ensure validity.
3. Using the list of Cortellis firms matched to SDC, we examine Cortellis's Extract field to identify patterns in the word descriptions of firms that change ownership. We identify several patterns, e.g. (XX, a subsidiary of YY), (XX, now YY), (XX, after its acquisition of YY), (XX, after its merger with YY) …etc. We use these patterns to flag other firms that had changed ownership.
4. We match these flagged Cortellis firms to SDC merger data using a fuzzy (non-exact) match. We manually check each match to ensure accuracy.
5. For flagged firms that did not match (or incorrectly matched) to SDC using the fuzzy matching method, we manually search Google to identify if any changes in ownership had occurred.

.

Appendix I.C: Identifying BTD Designations

Section I.C.1 of this appendix provides details on the procedure used to identify BTD designations. Figure I.C.1 of Section 1.C.2 presents an illustrative summary of the distribution of BTD events by year and therapeutic market. Section I.C.3 provides a discussion on the perception of BTDs as viewed by physicians, patients and the biopharmaceutical industry in general. Finally, Section I.C.4 discusses the results from tests that demonstrate the higher likelihood of approval for BTD drugs.

Section I.C.1: Identifying BTD Designation Events

We identify BTD events and match each to the corresponding firm, drug and indication as follows:

1. We utilize three resources to identify BTD awards. First, the "Regulatory Designation" field in Cortellis indicates whether a drug was awarded with a BTD, however, it does not identify the grant date, nor the designated indication[4]. Second, we use the Friends of Cancer Research (FOCR) website,[5]which identifies the BTD drug's name, the grant announcement date, the sponsoring firm and the designated indication(s). Finally, when possible, we also use the financial statements[6] of a firm to verify the accuracy of our matches.
2. Cortellis provides two variables that identify the drug name – 1) Drug Name, and 2) Other Drug Names. The first variable usually identifies the active ingredient of a drug (generic name), and the second identifies all names given to a drug by any of its current, or previous, developers. We match the Cortellis data to the FOCR data on exact drug names. Unmatched drugs are processed in step 3 below. Matched drugs are manually verified.
3. For the sample of unmatched drugs (from step 2), we conduct a fuzzy (non-exact) match between Cortellis and FOCR using drug names, and manually verify the matched records.
4. For unmatched drugs identified by Cortellis as BTD-recipients but not included in FOCR, we search firm financial statements, FDA documents, and business media articles to find the grant date, developing firm and designated indication.
5. We investigate the cases where a BTD drug in FOCR did not match to Cortellis. We find 42 such drugs that either don't exist in Cortellis or do exist but the designated indication does not.
6. Finally, we also verify our matches with the 143 BTDs reported in the online supplementary appendix for Hoffman et al. (2019).

[4] As a reminder, a single drug may be developed for several indications. BTD designations are awarded to a drug-indication pairing. This implies that a BTD drug may be designated for some indications, but not for others. Cortellis identifies the BTD awards at the drug-level, and does not indicate the designated indication.

[5] https://www.focr.org/breakthrough-therapies

[6] For publicly listed BTD firms.

Section I.C.2: Distribution of BTD Awards by Year and Therapeutic Market

Figure I.C.1: Distribution of BTD grants by year, ICD-10 market and firm

This figure presents illustrative descriptions of the BTD awards in our sample. Figure I.C.1.A displays the distribution of BTD awards by year. Figure I.C.1.B displays this distribution by the therapeutic market in which a BTD was granted. These markets are defined using the first letter of their ICD-10 code. Refer to Table I.A.1 in Internet Appendix I.A for the definitions of the ICD-10 markets.

Figure I.C.1.A: Distribution by Year

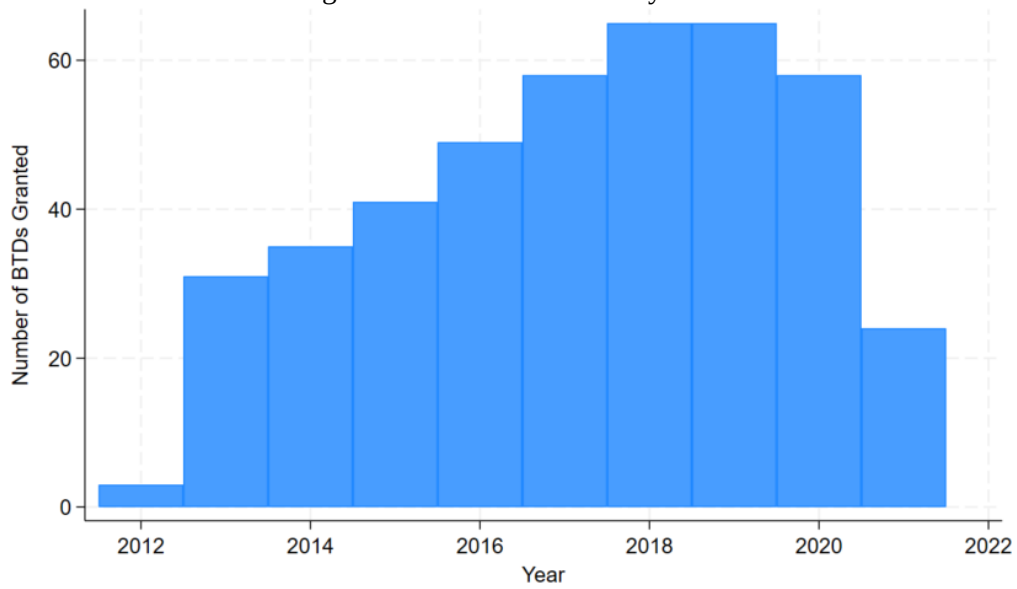
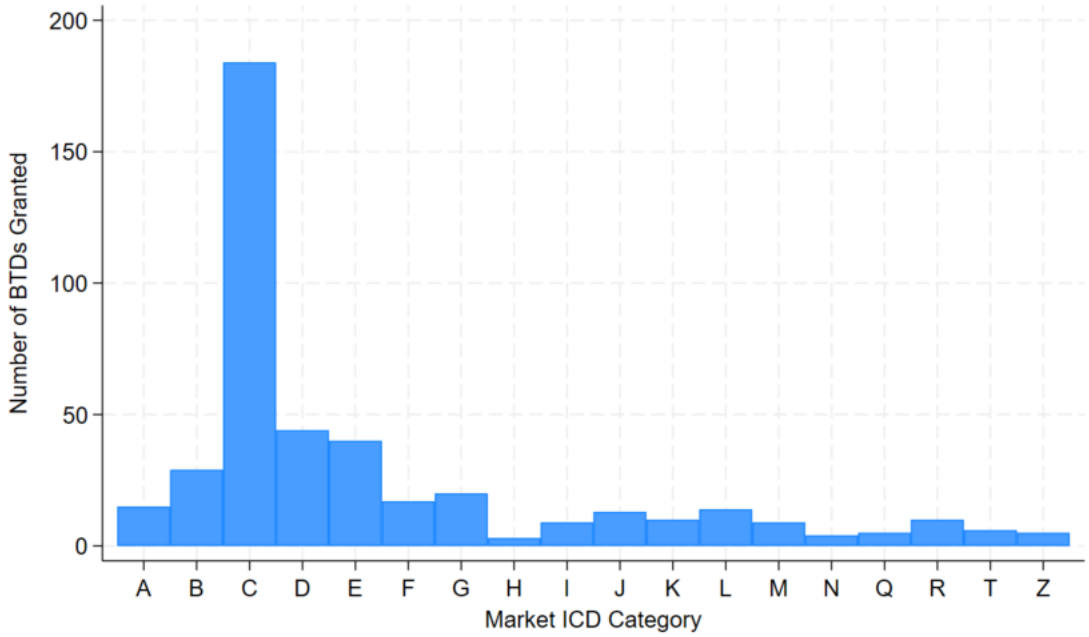


Figure I.C.1.B: Distribution by Therapeutic Market



Section I.C.3: Patient, Physician, and Industry Views of BTDs

Demand for pharmaceutical products is primarily driven by prescriptions from physician office visits.[7] In addition, patients may request physicians for specific drug prescriptions, especially for brand name drugs (Campbell et al. (2013)). This suggests that the demand for pharmaceutical drugs depends to a large extent on the perception of the best available treatment by both physicians and patients.

Abola and Prasad (2016) find that describing drugs using words such as "breakthrough" creates public perception that suggests scientific victory and miracle cures. Krishnamurti et al. (2015) survey a random sample of 597 Americans and find that the term "breakthrough" increased people's belief in a drug's effectiveness and participants were more likely to choose such a drug to treat a deadly condition over a drug without such description. This perception is not limited to the public as some studies have also found that health professionals and physicians can also perceive breakthrough drugs to be substantially better than existing therapies. For example, Kesselheim et al. (2016) analyze survey data from 692 physicians and find that physicians were more likely to prescribe the breakthrough drug for their patients than the alternative treatment and conclude that the choice of the term "breakthrough" may lead physicians to overprescribe the drug.

The perceived superiority of BTD drugs is reinforced by their extraordinary effectiveness in some cases. For example, Zoulim et al. (2015) find that Gilead Sciences' Hepatitis C BTD drug, Harvoni, cures over 95% of most patient populations while simultaneously reducing the treatment to 12 weeks compared with the 45% cure rate and the 6-12 months treatment duration for previous therapies. BTD drugs are also likely to boost the revenues of their owners. For example, Merck's Keytruda, approved in 2015, accounted for 23% of Merck's total revenues in 2019. Additionally, a report by Evaluate Vantage Pharma which ranked the drugs approved in 2017 by expected 2022 sales, found that 7 of the top 10 drugs were BTDs.[8]

[7] For example, the CDC estimates that in 2016, the number of drugs prescribed through physician office visits was about 3 billion units compared to 359 million units prescribed at hospital emergency department visits. Source: https://www.cdc.gov/nchs/fastats/drug-use-therapeutic.htm

[8] These are Ocrevus, Dupixent, Durvalumab, Niraparib, LEE011, KTE-C19, and Ingrezza. Source: Helfand, Carly. "The Top 10 Drug Launches of 2017." *FiercePharma*, 30 Jan. 2017, www.fiercepharma.com/special-report/top-10-drug-launches-2017.

Section I.C.4: FDA Approval Likelihood of BTD Projects

Given BTDs are granted earlier on in the development process, there remains uncertainty whether they will be approved-for-sale by the FDA. Nevertheless, we argue that they are still credible indicators of a threat to rivals. First, in untabulated tests we find that FDA approval-to-sell likelihood is higher for BTDs than for matched projects. The matched sample are drugs in the same ICD-10 (i.e., therapeutic) market, and having the same patent status, same initial development status, and similar drug age. Hazards for BTD drugs indicate they are 3.5 times more likely to be approved relative to non-BTD drugs.[9] Second, the literature provides evidence that BTD drugs are more likely than the average drug to receive FDA approval (see for example Hermosilla (2024)), and that they receive FDA approval in a shorter time (see Hwang et al (2018)).

[9] We also run a separate logit model using this randomly matched sample while accounting for the potential right censoring problem of our data by dropping drug projects that started development after 2019Q4, and find similar results.

Appendix I.D: Descriptions of the Various Samples and Measures Referenced in the Paper

In this appendix, we provide a detailed description of the various samples and measures that were referenced in the main paper. The sections in this appendix are ordered as follows:

- Section I.D.1: Description of the Sample used in the Chinese Import Competition Tests.
- Section I.D.2: Description of the Alternative Sales-Based HHI Competition Measure
- Section I.D.3: Descriptions of the Sample of BTD Drugs that Either Published Trial Findings or Issued Patents before the BTD Award.

Section I.D.1: Description of the Sample used in the Chinese Import Competition Tests.

This section provides details on the sample construction for the tests of firm innovation following Chinese import penetration that are reported in Table 2. In the first part of this section, we provide details on the data that was used to create the sample and the definition of the variables. We next provide details on the first stage of the 2SLS regressions.

First, a description of all the data sources used in this test:

- Data on Chinese import penetration: This data is obtained from David Dorn's website and is used in Autor et al (2013) and Acemoglu et al (2016). The data includes information on Chinese import penetration to US 4-digit SIC industries and to a set of the same industries in comparison countries' and covers the period extending from 1991 and until 2011. This data is originally obtained from the UN Comtrade database and converted from HS 6-digit codes to sic87dd codes using a crosswalk provided at www.ddorn.net/data.htm. Specifically, we obtain two variables from this data: (i) The variable "l_import_usch_YYYY" indicate the value of US imports from China in year YYYY for a given industry, and (ii) "l_import_otch_YYYY" provides the corresponding import values for a group of eight other wealthy countries in a given industry and year. Both variables are normalized by the US market volume for that same industry in 1991. The sample of comparison countries include Australia, Denmark, Finland, Germany, Japan, New Zealand, Spain, and Switzerland, which represent all high-income countries for which they can obtain disaggregated bilateral trade data at the HS level back to 1991.

- Firm scope is a measure for the breadth of product markets that a firm actively operates in. Scope data is obtained from Hoberg and Phillips (2023). The observational level of this data is firm year and a gvkey is provided for each firm.

- Data on patents and patent citations comes from the data used in Kogan et al (2017). The observational level of this data is at the patent level. For each patent, the following information is included: the Permno id for the issuing firm, the date the patent was filed, the date the patent was issued, the number of its citations and the real value of the patent.

- Data on quarterly firm financial information comes from the Compustat Annual fundamentals data. Specifically, we obtain R&D expenses and total assets.

- Data on Industry concentration: used as a control variable and is the four-firm concentration ratio obtained from the Census Bureau. The four-firm concentration ratio is the percentage of Value of

Shipments accounted for by the four largest firms in an SIC industry and includes all establishments in the industry, both privately held and public.

We construct the sample as follows. We begin with the sample of Chinese Import competition and retain a sample of each 4-digit SIC industry in each year with available information. Next, we match this sample of industry-years to firms in Compustat within the same industry and with available information in the corresponding year. We next match the resulting sample to scope data on gvkey and year and drop unmatched firms from the sample. Finally, we use the CRSP-Compustat Merge files to identify the Permno of each firm in Compustat and match these firms to the patent data. We create a citation-weighted count for the patents filed by a firm each year. The citation-weighted patent count variable is created for each firm-year and is calculated by (i) for each patent, we find the relative citation weight by dividing its citations by total citations of all patents in that industry, (ii) we then multiply this relative citation-weight by the total number of patents within that industry, (iii) finally, we sum up all citation-weighted patents of a firm that were filed in a given year. The resulting sample includes 4,230 firms in 120 4-digit SIC industries over the period covering 1991 through 2011.

The goal of our analysis is to highlight the presence of the aggregation problem that arises when using firm-level data, and especially for firms that operate in numerous product markets (i.e., firms with high scope). We wish to show that observing the inverted-U relationship between competition and innovation is more challenging with firm-level data when firms operate in numerous markets. We use Chinese imports to the US as a proxy for competition and run regressions of firm innovation, proxied by R&D spending and patents, on competition and its square.

Since Chinese imports to the US can be endogenous (see Lie and Yang (2022)), we use the identification strategy of Autor et al (2013) by instrumenting for the Chinese trade penetration in the United States using the Chinese trade penetration in a sample of comparison countries. This approach is commonly used in the literature (see Acemoglu et al (2016), Lie and (Yang 2022)).

More specifically, our strategy involves the use of a 2-stage least square regression where the first stage instruments both Chinese imports to the US and its square with Chinese Imports to comparison countries and its square. We run the following first stage regression for each 4-digit SIC industry *i* in year *t:*

$$\widehat{USImports\ from\ China}_{t,i} = \alpha_0 + \hat{\vartheta}_1 \widehat{OtherImportsfromChina}_{t-1,i} + CR4_{t,i} + \gamma_f + \theta_i + \delta_t + \epsilon_{i,t}$$

$$\widehat{USImportsfromChina}^2_{t,i} = \alpha_0 + \hat{\vartheta}_2 \widehat{OtherImportsfromChina}^2_{t,i} + CR4_{t,i} + \gamma_f + \theta_i + \delta_t + \epsilon_{i,t}$$

For firm *f*, we run the following second stage regressions:

$$Firm\ Innovation_{f,t,i} = \alpha_0 + \beta_1 \widehat{USImportsfromChina}_{t-1,i} + \beta_2 \widehat{USImportsfromChina}^2_{t-1,i} + CR4_{t,i} + \gamma_f + \theta_i + \delta_t + \epsilon_{i,t}$$

The estimation of the 2SLS regressions are done using stata's ivreghdfe function, with the following FE: 4-digit SIC, firm number of segments, and year. SE are clustered by firm. Moreover, we control for an industry's concentration by including the four-firm ratio (CR4). The results are reported in Table 2. We also test the overidentification restriction and find that the instruments are valid since we are unable to reject the null hypothesis (i.e., the p-value from Hansen's J-test is not significant of the). This is also consistent with the findings in Lie and Yang (2022).

Section I.D.2: Description of the Alternative Sales-Based HHI Competition Measure

In this section we first provide details on the construction of an alternative (inverse) competition measure; a sales-based HHI concentration measure. Then we show that this sales-based concentration measure is significantly and negatively correlated with our original competition measure.

Our original competition measure, defined as the natural log of the number of drug projects within a therapeutic market, does not include information on the level of sales within an industry. We construct the alternative sales-based concentration [i.e. inverse-competition] measure, and replicate our baseline hazard tests of table 6 in the main paper.

The sales-based HHI index is created using Drug sales data from the Cortellis database. Cortellis provides information on drug sales in 2018 through 2023 and expected sales through 2029. There are several challenges associated with the measurement of a drug's sales in a market, the following discusses some of these challenges and our approach to overcoming them. First, since drugs are often developed for multiple indications, finding the sales of a drug in a given indication is unobservable because drug sales are reported at the drug (not drug-indication) level. We overcome this challenge by assuming that each of a drug's indications contributes equally to its revenues and divide drug sales by the number of indications. Second, since drugs may be approved for some indications and still being developed for others, we assume that only those indications where a drug is approved for sale will generate revenues (this assumption also applies to the first point above). Third, because our sample starts in 2010 and drug sales are only reported in 2018 and after, we backfill the sales of a drug using the average sales of the drug-indication in 2018 through 2023, conditional on the drug-indication being approved for sale by the FDA.

We calculate the HHI index in each therapeutic market as follows. First, we compute the quarterly sales of each firm in each product market. We next compute the share of sales that each firm accounts for

in a market by dividing the firm's sales in that market by the total sales in that market. We next compute the HHI index by adding the square of each firm's market share in that market in each quarter.

The distribution of the HHI index is reported in Panel A of Table I.D.1 below. The index has high and negative correlation with our baseline competition measure at -53%. Moreover, the tests in Panel B, which regress the HHI index on the baseline competition measure, show that it is significantly correlated with the baseline measure. This provides support for the validity of our baseline competition measure.

Table I.D.1: Sales-Based HHI Measure Comparison to the Baseline Competition Measure

This table presents summary statistics on the Sales-Based HHI competition measure (Panel A) and cross validates this measure with the original competition measure (Panel B). The sample has an observation level of market-quarter and includes only markets with non-missing HHI values. Panel B presents the results from an OLS regression of the HHI competition measure on the original competition measure. The regression includes market and quarter fixed effects and clusters standard errors by market. T-stats are presented in parentheses and asterisks indicate statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1

Panel A: Summary Statistics of HHI

	(1)
Mean	0.718
25th Percentile	0.473
Median	0.801
75th Percentile	0.997
Standard Deviation	0.289
Correlation with Original Competition Measure	-0.531

Panel B: Regressing HHI on Baseline Competition Measure

	(1)
Baseline Competition Variable	-0.075***
	(-4.668)
Constant	0.935***
	-20.064
Observations	21,057
R-Squared	0.8294

Appendix I.E: Supplementary Results for the Phase-II Development Tests

In this appendix we provide supplementary results for the phase II development tests. The sections in this appendix are ordered as follows:

- Section I.E.1 Time Trends of Phase-II Development and Drug Initiations
- Section I.E.2: Estimating the Relationship Between Competition and Phase-II Development Using Non-Parametric Regressions
- Section I.E.3: Additional Tests on the Robustness of Phase-II Development Results

Section I.E.1: Time Trends in Phase-II Development Using Alternative Regression Models

Figure I.E.1: OLS replication of Main Figure 2

The figures below display results from tests that replicate those of Main Figure 2 only using OLS regressions. The dependent variable, Graduation, is an indicator equal to one in the quarter that a project graduates to phase-III, and equal to zero in all quarters where it remains in phase-II. The analysis sample is summarized in Table 5 below. The plotted coefficients are interaction variables between an indicator, ever-BTD shocked, which equals one if a phase-II project ever experiences BTD entry, with annual event time indicators for each of the 5 years before and after BTD entry. The bar caps cover the 95% confidence intervals for each coefficient. The regressions below include the following fixed effects: calendar quarter, market and project age in phase-II. Standard errors are clustered by drug project. The analysis sample in Figure 1A is the full sample of phase-II projects. The sample in panel B (panel C) includes phase-II projects in markets with low (high) competition, i.e., markets with competition levels below (above) the median competition level in the full phase-II sample. To control for the right-censoring problem, the sample excludes projects that report the start of phase-II development on or after 2020q1.

Figure I.E.1.A: Full Sample

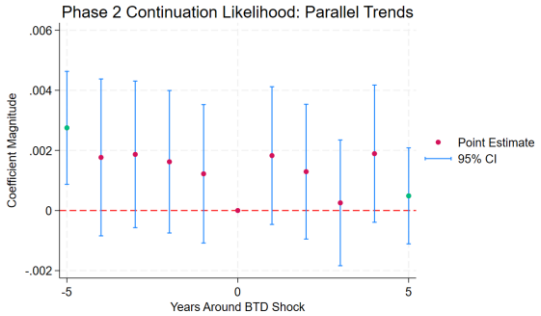


Figure I.E.1.B: Low Competition Subsample

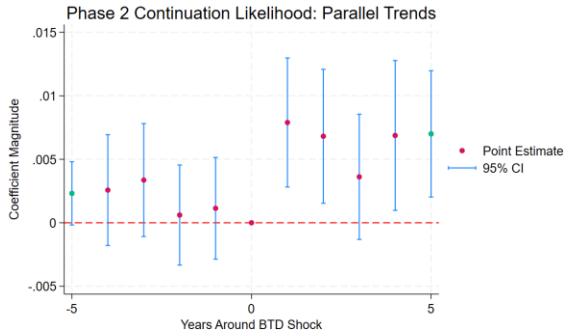


Figure I.E.1.C: High Competition Subsample

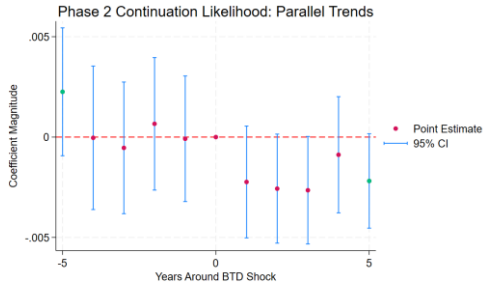


Figure I.E.2: OLS replication of Main Figure 3

The figures below display results from tests that replicate those of Main Figure 3 only using OLS regressions. The dependent variable, Graduation, is an indicator equal to one in the quarter that a project graduates to phase-III, and equal to zero in all quarters where it remains in phase-II. The analysis sample is summarized in Table 5 below. The plotted coefficients are interaction variables between an indicator, ever-SMST with annual event time indicators for each of the 5 years before and after BTD entry. Ever-SMST is an indicator that equals one if a rival's phase-II project ever experiences the entry of a BTD drug in its market that uses the same technology (i.e., the rival's phase-II project shares at least one target action with the BTD drug). The bar caps cover the 95% confidence intervals for each coefficient. The regressions below include the following fixed effects: calendar quarter, market and project age in phase-II. Standard errors are clustered by drug project. The analysis sample in Figure 2A is the full sample of phase-II projects. The sample in panel B (panel C) includes phase-II projects in markets with low (high) competition, i.e., markets with competition levels below (above) the median competition level in the full phase-II sample. To control for the right-censoring problem, the sample excludes projects that report the start of phase-II development on or after 2020q1.

Figure I.E.2.B: Full Sample

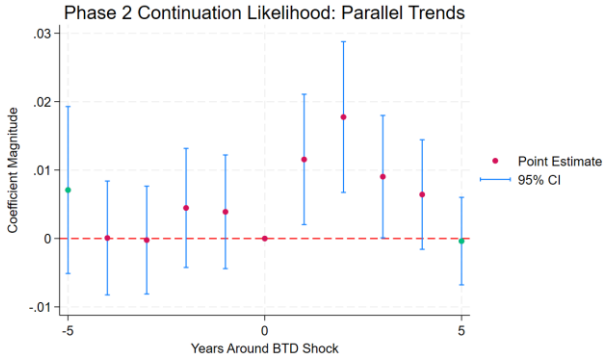


Figure I.E.2.B: Low Competition Subsample

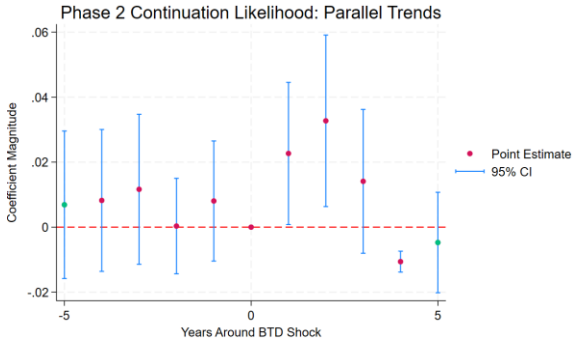


Figure I.E.2.C: High Competition Subsample

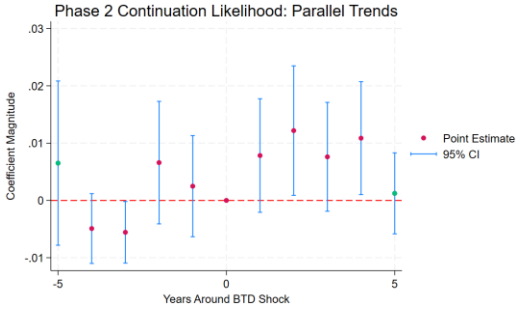


Figure I.E.3: OLS replication of Main Figure 4

The figures below display results from tests that replicate those of Main Figure 4 only using OLS regressions. The dependent variable, Graduation, is an indicator equal to one in the quarter that a project graduates to phase-III, and equal to zero in all quarters where it remains in phase-II. The analysis sample is summarized in Table 5 below. The plotted coefficients are interaction variables between an indicator, ever-SM with annual event time indicators for each of the 5 years before and after BTD entry. Ever-SM is an indicator that equals one if a rival's phase-II project ever experiences the entry of a BTD drug in its market that uses a different technology (i.e., the rival's phase-II project uses different target actions than those used by the BTD drug). The bar caps cover the 95% confidence intervals for each coefficient. The regressions below include the following fixed effects: calendar quarter, market and project age in phase-II. Standard errors are clustered by drug project. The analysis sample in Figure 3A is the full sample of phase-II projects. The sample in panel B (panel C) includes phase-II projects in markets with low (high) competition, i.e., markets with competition levels below (above) the median competition level in the full phase-II sample. To control for the right-censoring problem, the sample excludes projects that report the start of phase-II development on or after 2020q1.

Figure I.E.3.A: Full Sample

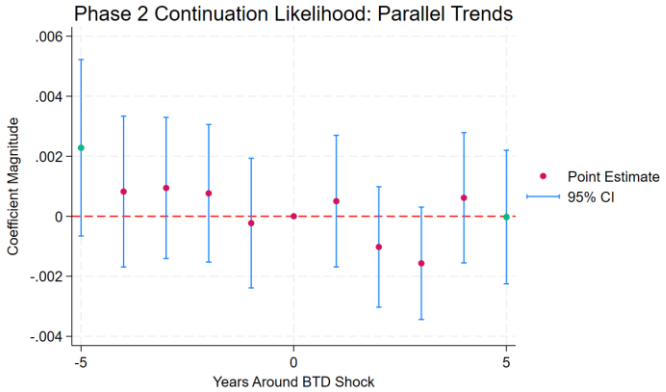


Figure I.E.3.B: Low Competition Subsample

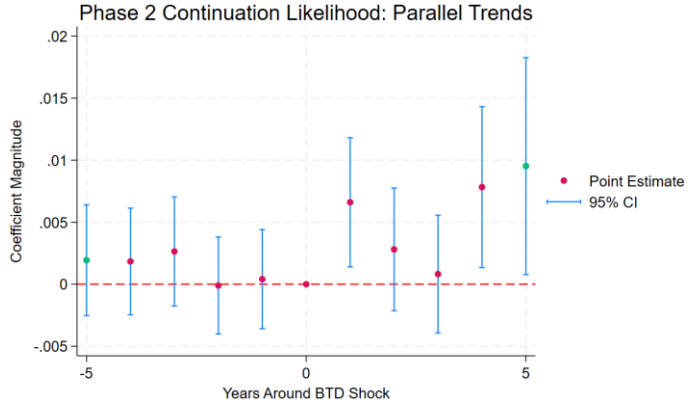


Figure I.E.3.C: High Competition Subsample

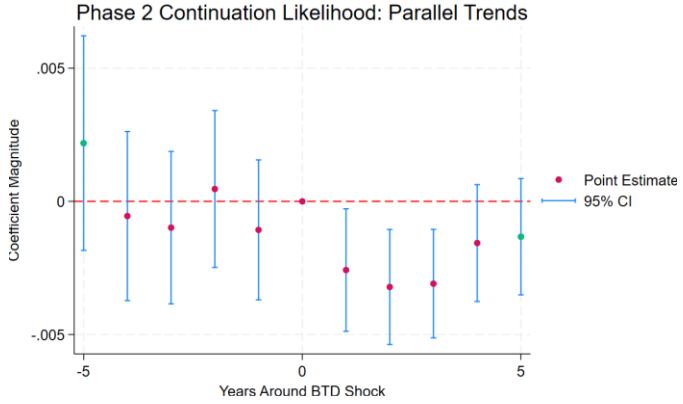


Figure I.E.4: Project Initiations following BTD entry "Does BTD Entry Expand the Market?"

This figure displays coefficients from OLS regressions of the number of new project initiations in markets that have experienced BTD entry relative to control markets that have never experienced BTD entry, in the 5 years before and 5 after, the BTD entry event. The observation level of the sample analysis is market-quarter. The dependent variable counts the number of new project initiations in a market in a given quarter. New project initiations are identified in the first quarter that a drug project appears in the sample. These new project initiations are summed up across all projects in a market on a given quarter. The plotted coefficients are interaction variables between an indicator, ever-BTD shocked, which equals one if a market ever experiences BTD entry, with annual event time indicators for each of the 5 years before and after BTD entry. The bar caps cover the 95% confidence intervals for each coefficient. The OLS model includes calendar quarter and market fixed effects, and clusters standard errors by market. The analysis sample in Figure 5A is the full sample of markets. The sample in panel B (panel C) includes markets with low (high) competition, i.e., markets with competition levels below (above) the median competition level in the full sample.

Figure I.E.4.A: Full Sample

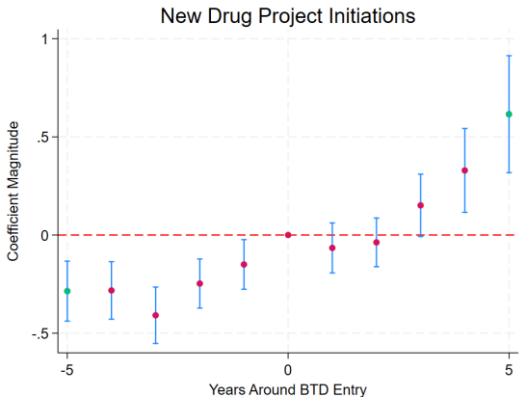


Figure I.E.4.B: Low Competition Subsample

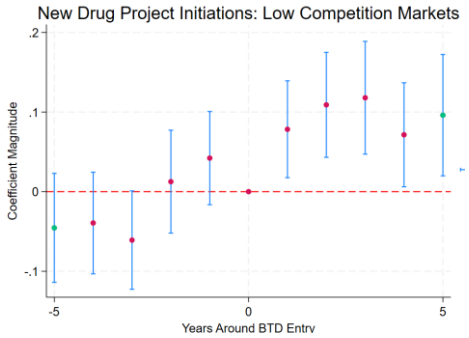
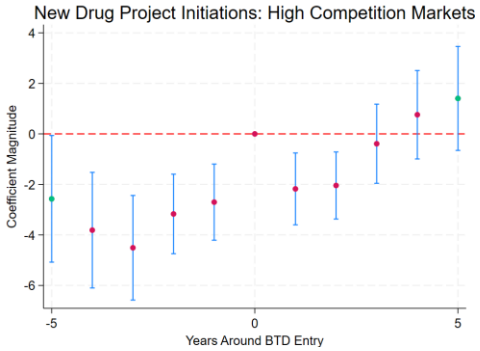


Figure I.E.4.C: High Competition Subsample



Section I.E.2: Estimating the Relationship Between Competition and Phase-II Development Using Non-Parametric Regressions

Figure I.E.5: Non-parametric Regression of Innovation on Competition

The figures below illustrate the relationship between competition and innovation using a non-parametric local polynomial smoothing function with a quadratic form. The sample used in these tests is constructed from the sample of Phase-II (Described in Table 5) as follows. First, for each drug project three variables are created: (i) average-drug-competition is the average competition level across all of a drug's quarters, and (ii) ever-graduated is an indicator equal to one if a had eventually graduated to phase-III and equal to zero if it remained in phase-II until the end of the sample, and (iii) BTD shock status, is a discrete variable equal to one if a project is never shocked by BTD entry, equal to two if a drug is eventually shocked by BTD entry and is an SM rival, and is equal to three if a drug is eventually shocked by BTD entry and is an SMST rival. The observation-level of the resulting sample is drug-project. Next, using the competition levels of the cross-sectional drug sample, competition percentiles are computed, and each drug is assigned to its corresponding competition percentile. In each percentile, the innovation variable, percentage of graduating projects, is computed by dividing the number of projects that eventually graduate to phase-III by the total number of projects in the corresponding competition percentile. The innovation measure is calculated for all drugs in a competition percentile as well as for each group (e.g., the percentage of graduating SM projects in a competition percentile is equal to the number of graduating SM projects in that competition bin divided by the total number of SM projects in the same bin). Finally, the sample is collapsed to the competition percentile-BTD shock-status level. A non-parametric local polynomial smoothing function with quadratic form, a bandwidth of 20 and an Epanechnikov kernel function estimates the relationship between the percentage of graduating projects variable and the competition percentiles. The figures below display the resulting soothed relationship in red as well as the distribution of the innovation measure across competition percentiles (displayed by the scattered blue dots). Finally, the turning point for each relationship is computed by setting the derivative of the estimated quadratic relationship to zero.

Figure I.E.5.A: All Phase-II Projects in the Sample (Rival and Control Projects)

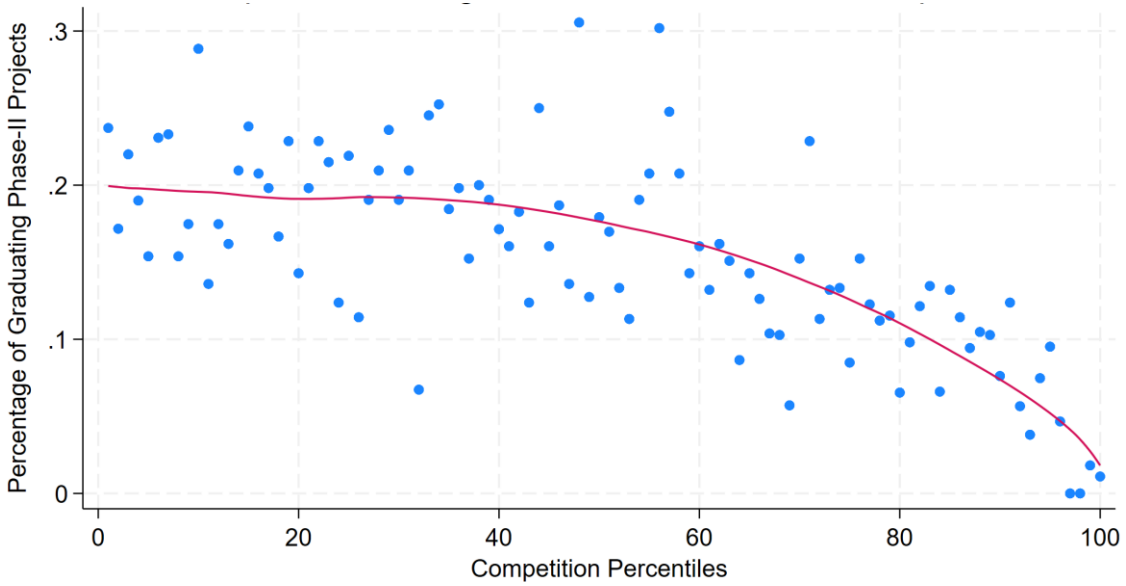


Figure I.E.5.B: Control Phase-II Projects (turning point = 48)

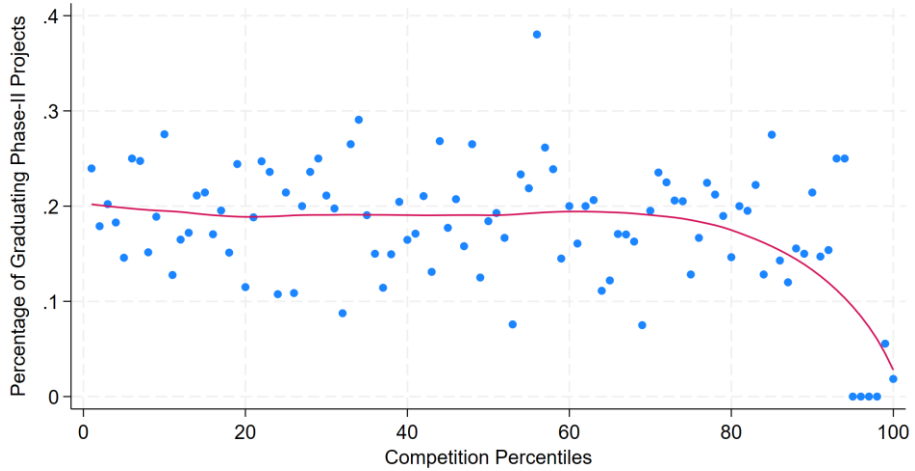


Figure I.E.5.C: All BTD Rivals' Phase-II Projects (Turning point=9)

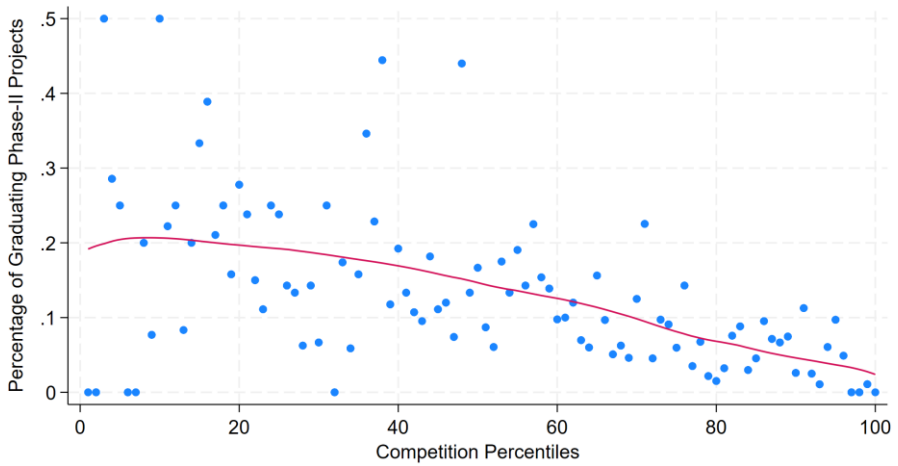


Figure I.E.5.C: SMST Rivals' Phase-II Projects (Turning Point 37)

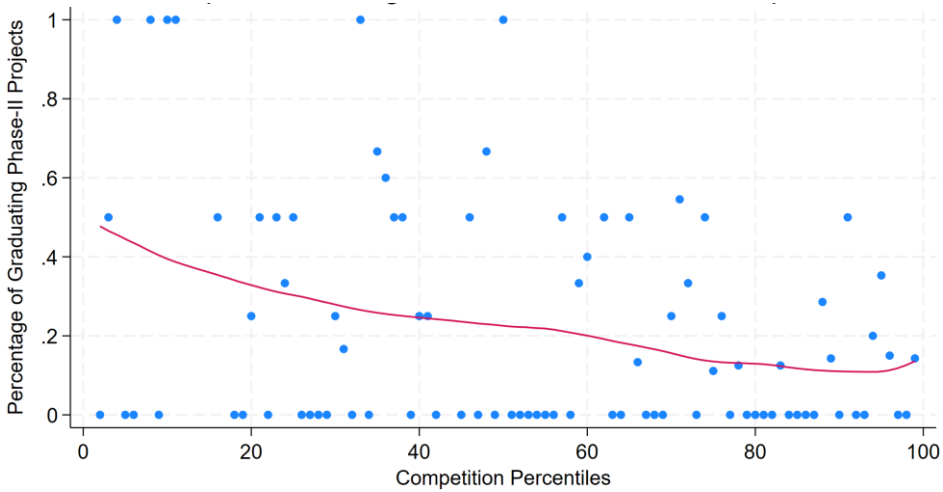
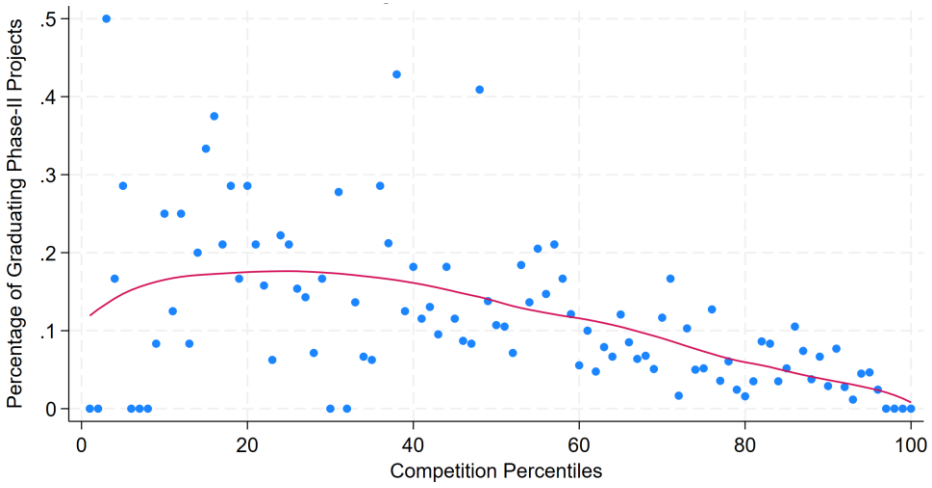


Figure I.E.5.A: SM Rivals' Phase-II Projects (Turning Point=23)



Section I.E.3: Additional Tests on the Robustness of Phase-II Development Results

Table I.E.1: Phase-II Projects Developed Before First BTD Entry

The tests in this table replicate those of Main Table 6 using only phase-II projects that had started development before the first BTD had entered the market, and control projects. The tests below use the same Cox Proportional Hazards model specifications as those used in Main Table 6, where estimates are stratified by market and the analysis time is equal to the number of quarters since the start of phase-II trials. The success event, *Graduation*, is an indicator equal to one in the quarter that a phase-II project graduates to phase-III, and equal to zero in all other quarters. z-statistics are reported in parenthesis. asterisks indicate statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1

Panel A: BTD Shocks and Phase-II Development

	Full Sample		Competition Level	
			Low	High
	(1)	(2)	(3)	(4)
BTD Shock	-0.004	1.670***	0.479***	-0.243*
	(-0.038)	(5.363)	(3.575)	(-1.829)
Competition		0.008		
		(0.428)		
BTD Shock*Competition		-0.352***		
		(-5.407)		
Observations	159,573	159,573	91,995	67,578

Panel B: BTD Shocks and Phase-II Development Conditional on the Type of Rival

	Full Sample		Competition Level	
			Low	High
	(1)	(2)	(3)	(4)
SMST	0.683***	1.542*	0.722**	0.496
	(2.974)	(1.913)	(2.118)	(1.635)
SM	-0.046	1.696***	0.375***	-0.469***
	(-0.465)	(5.111)	(2.937)	(-3.549)
Competition		0.003		
		(0.152)		
SMST*Competition		-0.186		
		(-1.075)		
SM*Competition		-0.366***		
		(-5.297)		
Observations	159,573	159,573	92,016	67,557

Table I.E.2: Phase-II Development of Acquired Projects

The tests displayed below are reported from the main Cox Hazard model which stratifies estimates by market and uses an analysis time that counts the number of quarters since the start of Phase-II development. Acquired is a dummy variable equal to one starting from the quarter in which the drug (or firm) was acquired, until the end of the sample. In Panel A, the results examine the phase-II development likelihood of acquired phase-II projects, with no information included for BTD shock. Columns 5 and 6 of Panel A partition the sample on whether the focal drug overlaps with the developing firm's portfolio. Note that for acquired projects, this variable is defined as equal to one if the acquirer had developed drugs in that market before the focal project was acquired, whereas for projects not acquired, this variable is defined as equal to one if the developing firm was developing other projects in that market before the focal drug was initiated. In both cases, if the firm's presence in the focal drug's market had begun on the day the project was initiated (or acquired), then the drug counts towards the "No-Overlap" sample.
In Panels B and C, the same tests as those used in Main Table 6 are rerun, but excluding acquired projects in columns 1-4, and focusing only on acquired projects in columns 5-8. z-statistics are reported in parenthesis. asterisks indicate statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1

Panel A: Likelihood of Developing Acquired Phase-II Projects

	Full Phase-II Sample		Mkt Comp		Acquired Project Overlaps with acquirer's portfolio	
			Low	High	Overlap	No-Overlap
	(1)	(2)	(3)	(4)	(5)	(6)
Acquired Project	-0.586***	-0.721**	-0.547***	-0.669***	-0.753***	-0.312**
	(-4.822)	(-2.309)	(-3.742)	(-3.039)	(-3.871)	(-1.987)
Mkt Comp		-0.020				
		(-1.068)				
Acquired Project*Mkt Comp		0.034				
		(0.457)				
Observations	195,736	195,736	97,900	97,836	87,007	108,729

Panel B: BTD Shocks and Phase-II Development Conditional on Including Acquired Projects

	Excludes all Acquired Projects				Includes only Acquired Projects			
			Mkt Comp				Mkt Comp	
			Low	High			Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BTD Shock	0.017	1.199***	0.458***	-0.265***	0.263	1.192	0.260	0.521
	(0.236)	(5.429)	(4.218)	(-2.643)	(0.788)	(1.263)	(0.630)	(0.771)
Mkt Comp		-0.001				0.046		
		(-0.032)				(0.488)		
BTD Shock* Mkt Comp		-0.249***				-0.208		
		(-5.449)				(-1.015)		
Observations	182,290	182,290	90,355	91,935	13,446	13,446	7,545	5,901

Panel C: SM and SMST Rivals and Phase-II Development Conditional on Including Acquired Projects

	Excludes all Acquired Projects				Includes only Acquired Projects			
			Mkt Comp				Mkt Comp	
			Low	High			Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SMST	0.618***	1.233***	0.892***	0.463***	1.059*	1.131	0.701	1.733
	(4.952)	(2.999)	(4.540)	(2.840)	(1.956)	(0.374)	(0.937)	(1.588)
SM	-0.094	1.271***	0.395***	-0.424***	0.103	1.371	0.198	0.234
	(-1.217)	(5.406)	(3.469)	(-3.962)	(0.283)	(1.423)	(0.452)	(0.309)
Mkt Comp		-0.001				0.045		
		(-0.049)				(0.477)		
SMST*Mkt Comp		-0.126				-0.021		
		(-1.522)				(-0.035)		
SM*Mkt Comp		-0.289***				-0.289		
		(-5.929)				(-1.345)		
Observations	182,290	182,290	90,355	91,935	13,446	13,446	7,545	5,901

References in the Internet Appendix

Abola, M., & Prasad, V. (2016). The Use of Superlatives in Cancer Research. JAMA Oncology, 2(1), 139-41.

Acemoglu, D., Autor, D., Dorn, D., Hanson, G.H. and Price, B., 2016. Import competition and the great US employment sag of the 2000s. Journal of Labor Economics, 34(S1), pp.S141-S198.

Autor, D.H., Dorn, D. and Hanson, G.H., 2013. The China syndrome: Local labor market effects of import competition in the United States. American economic review, 103(6), pp.2121-2168.

Campbell, E., Pham-Kanter, G., Vogeli, C., & Iezzoni, L. (2013). Physician Acquiescence to Patient Demands for Brand-Name Drugs: Results of a National Survey of Physicians. JAMA Internal Medicine, 173(3), 237-239.

Colin, Cameron., Pravin, K., Trivedi. (2005). Microeconometrics: Methods and Applications. doi: 10.1017/CBO9780511811241

Cunningham, C., Ederer, F. and Ma, S., 2021. Killer acquisitions. Journal of Political Economy, 129(3), pp.649-702.

Hermosilla, M., 2024. Regulating ethical experimentation: Impacts of the breakthrough therapy designation on drug R&D. Journal of Health Economics, 94, p.102855.

Hoberg, G. and Phillips, G.M., 2024. Scope, scale and concentration: The 21st century firm. Journal of Finance.

Hoffmann, D., Van Dalsem, S., & David, F. (2019). Stock price effects of breakthrough therapy designation. Nature Reviews. Drug Discovery, 18(3), 165.

Hwang, T., Franklin, J., Chen, C., Lauffenburger, J., Gyawali, B., Kesselheim, A., & Darrow, J. (2018). Efficacy, Safety, and Regulatory Approval of Food and Drug Administration-Designated Breakthrough and Nonbreakthrough Cancer Medicines. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 36(18), 1805-1812.

Kesselheim, A., Woloshin, S., Eddings, W., Franklin, J., Ross, K., & Schwartz, L. (2016). Physicians' Knowledge About FDA Approval Standards and Perceptions of the "Breakthrough Therapy" Designation. JAMA: The Journal of the American Medical Association, 315(14), 1516-1518.

Kogan, L., Papanikolaou, D., Seru, A. and Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Quarterly Journal of Economics, 132(2), pp. 665-712.

Krieger, Joshua L. "Trials and terminations: Learning from competitors' R&D failures." Management Science (2021).

Krishnamurti, T., Woloshin, S., Schwartz, L., & Fischhoff, B. (2015). A Randomized Trial Testing US Food and Drug Administration "Breakthrough" Language. JAMA Internal Medicine, 175(11), 1856-1858.

Lie, E. and Yang, K., 2023. Import penetration and executive compensation. The Review of financial studies, 36(1), pp.281-316.

Wasi, N., & Flaaen, A. (2015). Record linkage using Stata: Preprocessing, linking, and reviewing utilities. The Stata Journal, 15(3), 672-697.

Zoulim F, Liang TJ, Gerbes AL, et al. Hepatitis C virus treatment in the real world: optimising treatment and access to therapies. Gut. 2015;64(11):1824-1833.