Replication files for
"Regulatory Incentives for Innovation: the FDA's Breakthrough Therapy Designation"

Date written: January 26, 2024

This Zip-file *BTD_Code_Data* contains Stata code to reproduce regression tables and figures reported in the paper and in the online appendix for "Regulatory Incentives for Innovation: the FDA's Breakthrough Therapy Designation" by Amitabh Chandra, Jennifer Kao, Kathleen L. Miller, and Ariel D. Stern.

To reproduce the main regression tables and figures, extract the ZIP file and run the file *paper_figuretable_public.do* after adjusting the global macro "path" in the file to match the folder structure in your system.

To reproduce the appendix regression tables and figures, run the file *appendix_figuretable.do*, again after adjusting the global macro "path."

The results will be placed in the folder containing the Stata code. The code was tested with Stata 16.

Our analysis makes use of restricted data on drug claims records from the Optum Labs Data Warehouse. As a result, our datasets and code exclude drug claims counts and variables (e.g., adverse event rates) that can be used to indirectly calculate them. However, any researcher wishing to replicate our results can obtain the drug claims directly from Optum Labs (https://www.optumlabs.com/). We would also be happy to provide all programs and instructions to replicate the paper once the researcher has access to the data.

In the following, we describe the datasets and variables used in the paper.

1 Datasets and Variables

1.1 analytic_full_public.dta

Position	Variable Name	Variable Description
1	applno	FDA application number
2	drugname	Established name
3	brandname	Proprietary name
4	year_app	Approval year
5	btd	Breakthrough Therapy
6	date_btd	BTD designation date
7	date_sub	FDA submission date
8	nda	Small Molecule
9	priority	Priority Review
10	fast	Fast Track
11	accel	Accelerated Approval
12	bbw	Black box warning
13	private	Private sponsor
14	orphan	Orphan drug
15	nda_app_day	Days between NDA submission and approval
16	atc_level1	ATC Code
17	atc_level1_id	ATC Code ID: group(atc_level1)
18	firm_drug	Firm experience (number of approved drugs)
19	firm_drug_rank	Ranking of firm drug experience, by year-ATC distribution
20	p2_sub_day	Days between Phase 2 trial start and NDA submission
21	p3_sub_day	Days between Phase 2 trial start and NDA submission
22	ae_2m_postap_tot	Adverse event counts: 2 months
23	ae_3m_postap_tot	Adverse event counts: 3 months
24	ae_4m_postap_tot	Adverse event counts: 4 months
25	ae_5m_postap_tot	Adverse event counts: 5 months
26	ae_6m_postap_tot	Adverse event counts: 6 months
27	p2_start	Phase 2 start date
28	p2_enrollment	Phase 2: Number of patients
29	p2_number_of_arms	Phase 2: Number of arms
30	p2_number_of_facilities	Phase 2: Number of facilities
31	p2_random	Phase 2: Randomized
32	p2_maskdoub	Phase 2: Double blinded masking
33	p3_start	Phase 3 start date
34	p3_enrollment	Phase 3: Number of patients
35	p3_number_of_arms	Phase 3: Number of arms
36	p3_number_of_facilities	Phase 3: Number of facilities
37	p3_random	Phase 3: Randomized

38	p3_maskdoub	Phase 3: Double blinded masking
39	any_new_drug_action	Any new mechanism of action
40	any_p4	Any phase 4 trial

1.2 analytic_matched_public.dta

Position	Variable Name	Variable Description
1	applno	FDA application number
2	drugname	Established name
3	brandname	Proprietary name
4	year_app	Approval year
5	date_sub	FDA submission date
6	btd	Breakthrough Therapy
7	date_btd	BTD designation date
8	nda	Small Molecule
9	priority	Priority Review
10	fast	Fast Track
11	accel	Accelerated Approval
12	bbw	Black box warning
13	orphan	Orphan drug
14	private	Private sponsor
15	nda_app_day	Days between NDA submission and approval
16	atc_level1	ATC Code
17	atc_level1_id	ATC Code ID: group(atc_level1)
18	firm_drug	Firm experience (number of approved drugs)
19	firm_drug_rank	Ranking of firm drug experience, by year-ATC distribution
20	p2_sub_day	Days between Phase 2 trial start and NDA submission
21	p3_sub_day	Days between Phase 2 trial start and NDA submission
22	ae_2m_postap_tot	Adverse event counts: 2 months
23	ae_3m_postap_tot	Adverse event counts: 3 months
24	ae_4m_postap_tot	Adverse event counts: 4 months
25	ae_5m_postap_tot	Adverse event counts: 5 months
26	ae_6m_postap_tot	Adverse event counts: 6 months
27	p2_start	Phase 2 start date
28	p2_enrollment	Phase 2: Number of patients
29	p2_number_of_arms	Phase 2: Number of arms
30	p2_number_of_facilities	Phase 2: Number of facilities
31	p2_random	Phase 2: Randomized
32	p2_maskdoub	Phase 2: Double blinded masking
33	p3_start	Phase 3 start date
34	p3_enrollment	Phase 3: Number of patients
35	p3_number_of_arms	Phase 3: Number of arms
36	p3_number_of_facilities	Phase 3: Number of facilities
37	p3_random	Phase 3: Randomized
38	p3_maskdoub	Phase 3: Double blinded masking
39	any_new_drug_action	Any new mechanism of action
40	any_p4	Any phase 4 trial

41	sampA	Sample: imputed non-BTD
42	sampB	Sample: imputed BTD
43	sampaY	Sample: true BTD
44	sampaX	Sample: true non-BTD
45	sampaW	Sample: pre-2012, other
46	did_samp	DID Sample
47	post	Post-2012
48	btd_new	True + Imputed BTD
49	post_btd_new	BTD x Post-2012

1.3 analytic_matched_fast_public.dta

Position	Variable Name	Variable Description
1	applno	FDA application number
2	drugname	Established name
3	brandname	Proprietary name
4	year_app	Approval year
5	date_sub	FDA submission date
6	btd	Breakthrough Therapy
7	date_btd	BTD designation date
8	nda	Small Molecule
9	priority	Priority Review
10	fast	Fast Track
11	accel	Accelerated Approval
12	bbw	Black box warning
13	orphan	Orphan drug
14	private	Private sponsor
15	nda_app_day	Days between NDA submission and approval
16	atc_level1	ATC Code
17	atc_level1_id	ATC Code ID: group(atc_level1)
18	firm_drug	Firm experience (number of approved drugs)
19	firm_drug_rank	Ranking of firm drug experience, by year-ATC distribution
20	p2_sub_day	Days between Phase 2 trial start and NDA submission
21	p3_sub_day	Days between Phase 2 trial start and NDA submission
22	ae_2m_postap_tot	Adverse event counts: 2 months
23	ae_3m_postap_tot	Adverse event counts: 3 months
24	ae_4m_postap_tot	Adverse event counts: 4 months
25	ae_5m_postap_tot	Adverse event counts: 5 months
26	ae_6m_postap_tot	Adverse event counts: 6 months
27	p2_start	Phase 2 start date
28	p2_enrollment	Phase 2: Number of patients
29	p2_number_of_arms	Phase 2: Number of arms
30	p2_number_of_facilities	Phase 2: Number of facilities
31	p2_random	Phase 2: Randomized
32	p2_maskdoub	Phase 2: Double blinded masking
33	p3_start	Phase 3 start date
34	p3_enrollment	Phase 3: Number of patients
35	p3_number_of_arms	Phase 3: Number of arms
36	p3_number_of_facilities	Phase 3: Number of facilities
37	p3_random	Phase 3: Randomized
38	p3_maskdoub	Phase 3: Double blinded masking
39	any_new_drug_action	Any new mechanism of action
40	any_p4	Any phase 4 trial

41	sampA	Sample: imputed non-BTD
42	sampB	Sample: imputed BTD
43	sampaY	Sample: true BTD
44	sampaX	Sample: true non-BTD
45	sampaW	Sample: pre-2012, other
46	did_samp	DID Sample
47	post	Post-2012
48	fast_new	True + Imputed Fast Track
49	post_fast_new	Fast Track x Post-2012