BENTLEY UNIVERSITY
Center for Integration of Science and Industry

Tech Note[1] 2023-1

Title: NIH Contribution to phased clinical development of drugs approved from 2010-2019

Investigators: Edward W. Zhou, Pharm.D., Matthew J. Jackson, Ph.D., Fred D. Ledley, M.D.

Date: 12/14/2021

Keywords: basic research, applied research, drug development, phased clinical trials, NIH funding, industry funding

Submitted by: Edward Zhou, Matthew Jackson, Fred Ledley	Version history: Final
Reviewed by: Fred Ledley	Date: 1/26/2023

Short Background:

This report describes the methods and preliminary results for the manuscript titled, "NIH contribution to phased clinical development of drugs approved from 2010-2019".

Materials and Methods:

PMIDs

Brand names, generic chemical names, and active pharmaceutical ingredient names of the 387 drugs approved by the FDA 2010-2019 were obtained from FDA annual reviews. The 235 targets associated with these drugs were identified from the mechanism of action section of the FDA approval labels (available at Drugs@FDA) and were confirmed via public literature searches.

Publications (PMIDs) for the 387 drugs and their 235 targets were obtained utilizing a Python code method to search the National Center for Biotechnology Information (NCBI) PubMed access tool, Entrez Global Query Cross-Database Search System. The Drug PMIDs were identified by searching for the brand names, generic chemical names, and active pharmaceutical ingredient names. The Target PMIDs were identified by searching for the

mechanism of action or associated target receptors. Target searches were then manually optimized through the NCBI PubMed web interface, adding search term limiters (i.e., AND, OR, NOT; Medical Subject Headings (MeSH) terms) to exclude PMIDs not directly related to the mechanisms of the approved drugs.

All identified PMIDs relating to the drugs (applied research) were subdivided into Clinical Trial and Non-Clinical PMIDs using an optimized search term, a PubMed 'Publication Type' tag, MeSH terms, and title/abstract text analysis (Figure 1). Those PMIDs identified as Clinical Trials were further subdivided into Phase and Non-Phase I-IV Clinical Trials using a lower hierarchy 'Publication Type' search, combined with a text analysis of PMID abstracts and Accession Numbers for references of a National Clinical Trial (NCT) (Table 1).

Sensitivity and Specificity

Search terms for Clinical Trial identification were developed on an earlier version of the PMID collection with a total of 305,892 PMIDs. Search term sensitivity and specificity were assessed by manually reading PMID titles and abstracts or through automated text and Publication Type screening of random sample sets consisting of 200-400 PMIDs (Figure 8). Percentages of true and false captures and exclusions within the test samples were then applied to group totals to estimate overall sensitivity and specificity of searches.

An initial Clinical Trial filter with the search term 'A2' (Table 5) captured 27,866 PMIDs from the validation dataset, though our extrapolated manual validation estimated 17,863 were false positives. An expansion to search term 'A3' captured 25,614 PMIDs with 296 false positives, generating 89.51% specificity and 97.06% sensitivity. To categorize Phase Clinical Trials, search term 'A4' was applied generating a poor sensitivity of 63.74% due to an estimated 5,290 false exclusions, and specificity of 97.02%. Improved Phase extraction was seen with search term 'A5', reducing false exclusions to an estimated 907, giving sensitivity and specificity values of 92.31% and 92.37% respectively. Overall, the whole extraction search term was estimated to have a sensitivity of 85.67% and specificity of 99.64%. An assumption of correct exclusion (true negatives) without manual validation was made for PMIDs excluded with a non-relevant Publication Type, or where the title or abstract made no reference of 'clinical', 'NCT', or 'trial'.

Funding

In order to analyze funding of all Clinical PMIDs, each PMID was associated with relevant NIH-funded projects within the RePORTER database. Expanding on methods initially developed by Cleary et al.[2,3], Applied Project Years (APYs) were established to accurately assess the impact of

[2] Cleary, E. G., Beierlein, J. M., Khanuja, N. S., McNamee, L. M., & Ledley, F. D. (2018). Contribution of NIH funding to new drug approvals 2010–2016. Proceedings of the National Academy of Sciences, 115(10), 2329-2334. https://www.pnas.org/content/115/10/2329

[3] Cleary, E., Jackson, M. J., & Ledley, F. (2020). Government as the First Investor in Biopharmaceutical Innovation: Evidence From New Drug Approvals 2010–2019. Institute for New Economic Thinking Working Paper Series, (133).

each grant (project) relative to their associated PMIDs. An APY is the year in which a PMID is published relative to the fiscal year of the proposed funding period for the associated project. APYs included up to 4 years after the last fiscal year of funding to account for publication lag. In addition, APYs will not consider any PMID published before the first fiscal year of the project. Total NIH funding (which includes both direct and indirect costs associated with each project) is calculated for each subgroup by the unique APY assigned to each category (Table 2). All funding is in constant US dollars (USD) and is inflation-adjusted to 2018 using the US Bureau of Labor Statistics consumer price index (CPI).

In order to determine NIH funding involvement in each clinical phase, sub-analysis of Phase I-IV PMIDs was conducted using an additional in-house built database from clinicaltrials.gov to divide them by phase type and to recover NCT numbers that were not categorized by the initial Python PubMed search method into its phased category (Table 3). PMIDs describing phase 1/2, phase 2/3, or multiple clinical trials were assigned to the highest clinical phase.

Additional sub-analysis was conducted to identify the distribution of developmental funding in other categories of interest. Using WHO therapeutic classification of drug first indications, each PMID was assigned a therapeutic area of activity depending on its original search term (Figure 6). PMIDs, which include multiple search terms, can be represented in more than one therapeutic area in this analysis. The special approval type designations (first in class, orphan, accelerated, breakthrough, fast track, and priority) were analyzed using the search terms categorization method as the therapeutic breakdown (Table 4).

Results and Conclusions:

Main Analysis

This work has identified NIH funding contributions to basic, applied, and drug development research for publications related to FDA approved drugs for the years 2010-2019. Search terms for the 387 drugs and 258 targets identified 2,550,350 unique PMIDs via the PubMed/Entrez database. A total of 354,156 PMIDs were linked to direct NIH funding equating to 312,149 unique APYs and $247,324 million of NIH investment for all of these drugs approved in the 2010-2019 decade (Figure 1). Further analysis identified 2,278,648 PMIDs associated with Target Only (basic research), 329,368 of which were funded by the NIH ($209,979 million, 84.90% of total NIH funding). Alternatively, 71,702 PMIDs were associated with Drugs (applied research), 24,788 of which were funded by the NIH ($37,344 million, 15.10% of total NIH funding) (Table 2). From the total number of Drug PMIDs, 24,826 were associated with human only, non-review Clinical Trials, of which 2,687 were linked to NIH funding ($10,023 million, 4.05% of total NIH funding). Direct drug development research was subsequently identified as any PMID categorized as a Phase I-IV Clinical Trial or found with an NCT registration number.

This resulted in 12,340 PMIDs of which 1,967 were matched to NIH funding ($8,103 million, 3.28% of total NIH funding).

In conclusion, of the $247,324 million of NIH investment identified in our work for FDA approved drugs from 2010-2019, 84.90% ($209,979 million) was associated with basic research, 11.05% ($27,320 million) was Non-Clinical applied research, and 3.28% ($8,103 million) was dedicated to applied phased clinical trial research.

Limitations:

Our analysis is limited by a number of factors associated with the nature of the databases and the method of our interpretation of funding.

Our APY funding association method has a limit of one year of funding per APY, which presents a conflict in fully representing long-term clinical trials. While the RePORTER database provides total project funding for all associated drug search terms, it does not describe the proportion of project funding to each individual relevant PMID. APY was created to eliminate duplications of multiple PMIDs linked to one project's funds, however, an APY only represents the year of funding the PMID was published. This reduces the funding representation of larger multiyear-study PMIDs vs smaller single-year trials.

In addition, a common publication lag of research papers relative to their supporting project official end date creates a significant drop in funding associated with more recent clinical trials or studies. We estimate that on average only 63.4% of PMIDs are published within their project's planned funding period, with an additional 10% being published one year after the project ends. While we adjusted the publication lag to account for up to four years after the project ends, this does not capture the unpublished works from 2021 because the RePORTER database has not been updated since December 2020 as of the writing of this report.

Another database limitation we encountered is that the PubMed metatags for both MeSH terms and Publication Types seem to have an inherent error rate with their association. This was partly corrected by our word analysis of abstract and Trial Accession number for NCT registration. However, the real sensitivity and specificity of the PubMed metatags need further elaboration, as a significant portion of our false-negative results from our sample hand check stemming from the complete or significant lack of metatags is linked with the PMIDs.

Work reported: January 31, 2023

Funding source: National Pharmaceutical Council; National Biomedical Research Foundation.

Tables and Figures:

Table 1. Total drug and target search terms.

Search Terms	Unique Search: Drug	Funded Search: Drug	Unique Search: Target	Funded Search: Target
Target Only			235	234 (99.2%)
Drug	387	355 (91.7%)		
Clinical Trial	381 (98.5%)	355 (91.7%)	215 (91.5%)	234 (99.2%)
Phase I-V or NCT	377 (97.4%)	240 (62.0%)	208 (88.5%)	155 (65.9%)
Non-Phased I-V	322 (83.2%)	140 (36.2%)	195 (82.9%)	121 (51.5%)

Table 2. Total PMIDs, APYs and NIH funding.

PMID Subtype	Unique PMIDs	NIH-funded PMIDs	Unique APYs	NIH funding (USD, 2018)	% of total NIH funding
TOTAL	2,550,350	354,156	321,023	$247,324,576,193	100%
Target Only	2,278,648	329,368	288,366	$209,979,809,340	84.90%
Drug	271,702	24,788	32,657	$37,344,766,853	15.10%
Clinical Trial	24,826	2,687	3,975	$10,023,865,788	4.05%
Phase I-V or NCT	12,340	1,967	2,834	$8,103,178,596	3.28%

Table 3. Developmental PMIDs by clinical phase.

Phase Type	PMIDs	NIH-funded PMIDs	APY	Funding (USD, 2018 (millions))
Phase 1	1,877	451	550	$1,460 (18.0%)
Phase 2	3,344	884	1,376	$3,502 (43.2%)
Phase 3	4,278	500	664	$2,557 (31.6%)
Phase 4	450	66	124	$427 (5.27%)
Undetermined (NCT)	2,391	66	120	$159 (1.93%)
TOTAL (unique)	12,340	1,967	2,834	$8,103

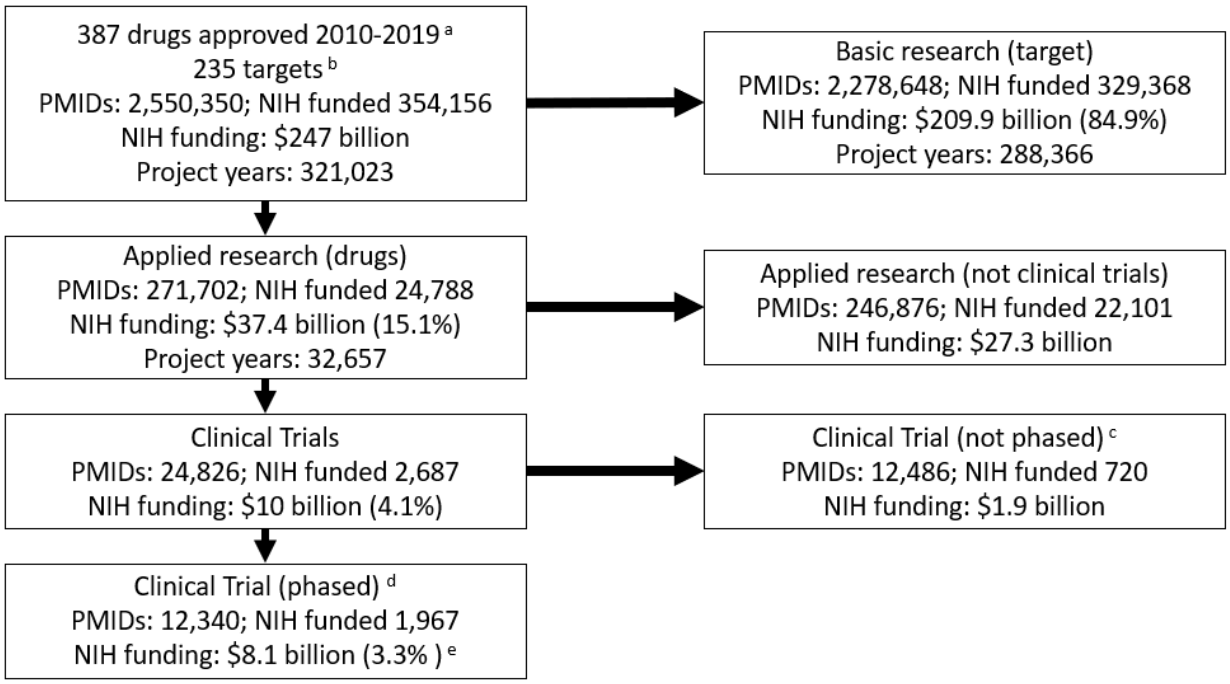
Table 4. Drug Search terms and PMIDs by approval type.

	Drugs	% Drugs with PMIDs	% PMIDs with Phase I-V, NCT
All approvals	387	92%	62%
First In Class	140	98%	73%
Orphan	167	95%	75%
Accelerated	49	100%	90%
Breakthrough	78	96%	78%
Fast Track	142	96%	73%
Priority	207	98%	73%

Table 5. Sensitivity and specificity for clinical trial and subsequent phase trials.

			Predicted Values							
Search Term Filters	**PMIDs Captured**	**PMIDs Excluded**	**Captured, True**	**Captured, False**	**Excluded, False**	**Excluded, True**	**Sensitivity**	**Specificity**	**Precision**	**Accuracy**
Step A2,…AND Clinical Trial [Publication Type]	45,729	263,603	27,866	17,863	-	-	-	-	60.938%	-
Step A3,...AND Human [MeSH] NOT Review [Publication Type]	25,910	272,195	25,614	296	3,001	9,771	89.511%	97.064%	98.859%	91.477%
Step A4,…AND Phase I-IV [Publication Type]	9,634	16,276	9,297	337	5,290	10,986	63.736%	97.022%	96.500%	78.283%
Step A5,…AND NCT [Text Search]	11,962	13,948	10,885	1,077	907	13,041	92.312%	92.374%	91.000%	92.346%
OVERALL CLINICAL TRIAL, PHASE EXTRACTION	11,962	297,370	10,885	1,077	1,820	295,550	85.673%	99.637%	91.000%	99.063%

Figure 1. Schematic of analysis and summary of PMIDs and NIH costs related to products approved 2010-2019 and phased clinical development, applied research, or basic research.



a - 355/387 (91.7%) drugs were categorized as applied research and 386/387 drugs received NIH funding through basic or applied research. b - 234/235 targets (99.6%) produced PMIDs that were categorized as basic research. c - Not phased clinical trials are PMIDs identified as human clinical trials but were not identified as part of a phase trial or have an NCT number. d - Phased clinical trials are PMIDs identified as human clinical trials reporting on phase 1-4 drug development trials. e - This $8.1 billion represents NIH investment for phase 1-4 PMIDs.

Figure 2. PMIDs and APYs by years from first drug indication FDA approval date.

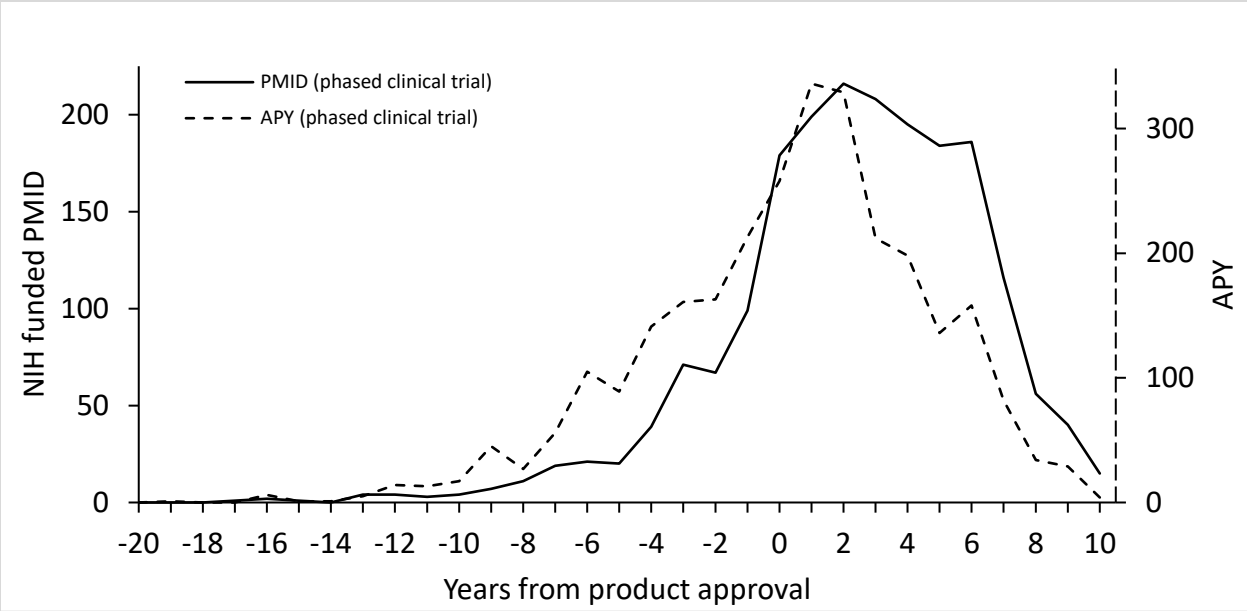


Figure 3. APYs for all PMIDs and phased PMIDs by years from first drug indication FDA approval date.

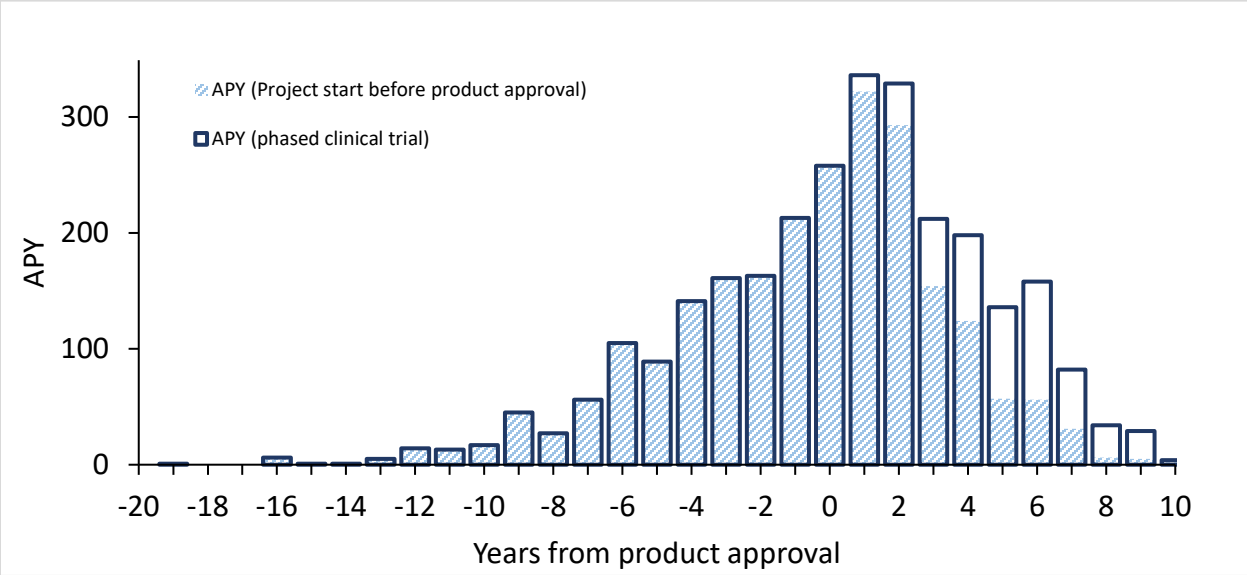


Figure 4. Cost of all PMIDs and phased PMIDs by years from first drug indication FDA approval date.

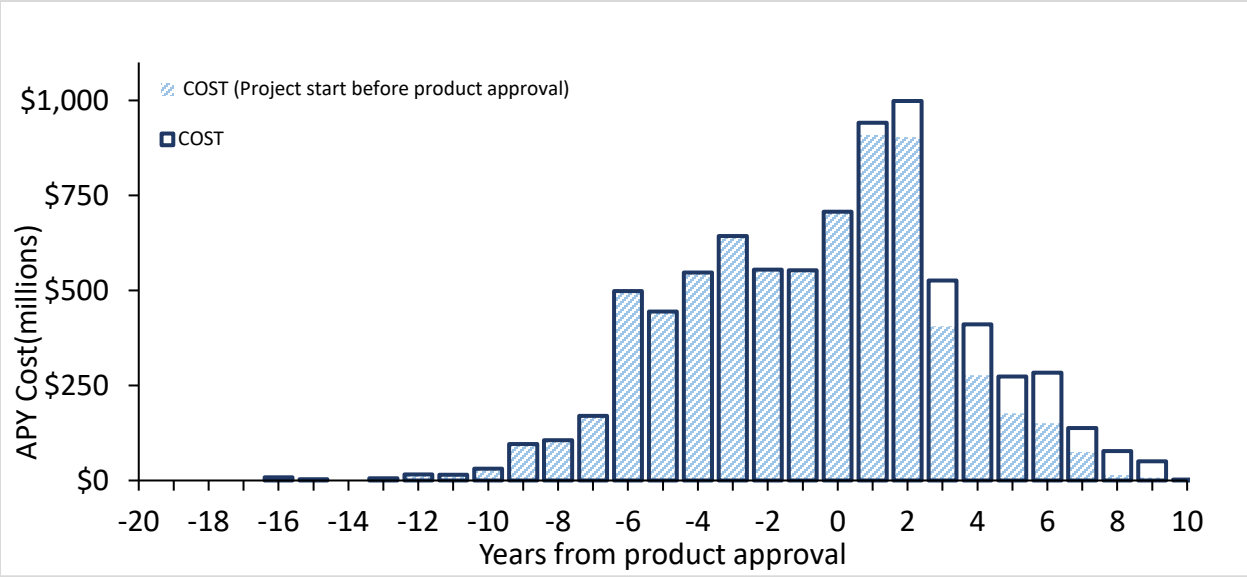


Figure 5. Products, PMIDs, APYs, and Funding by approval type.

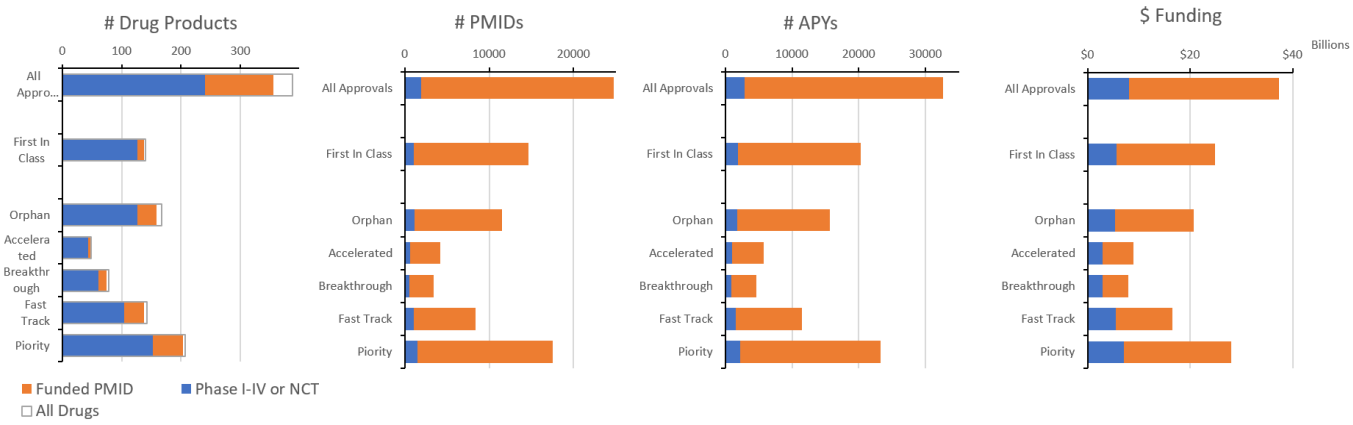


Figure 6. Drug approval distribution by therapeutic areas.

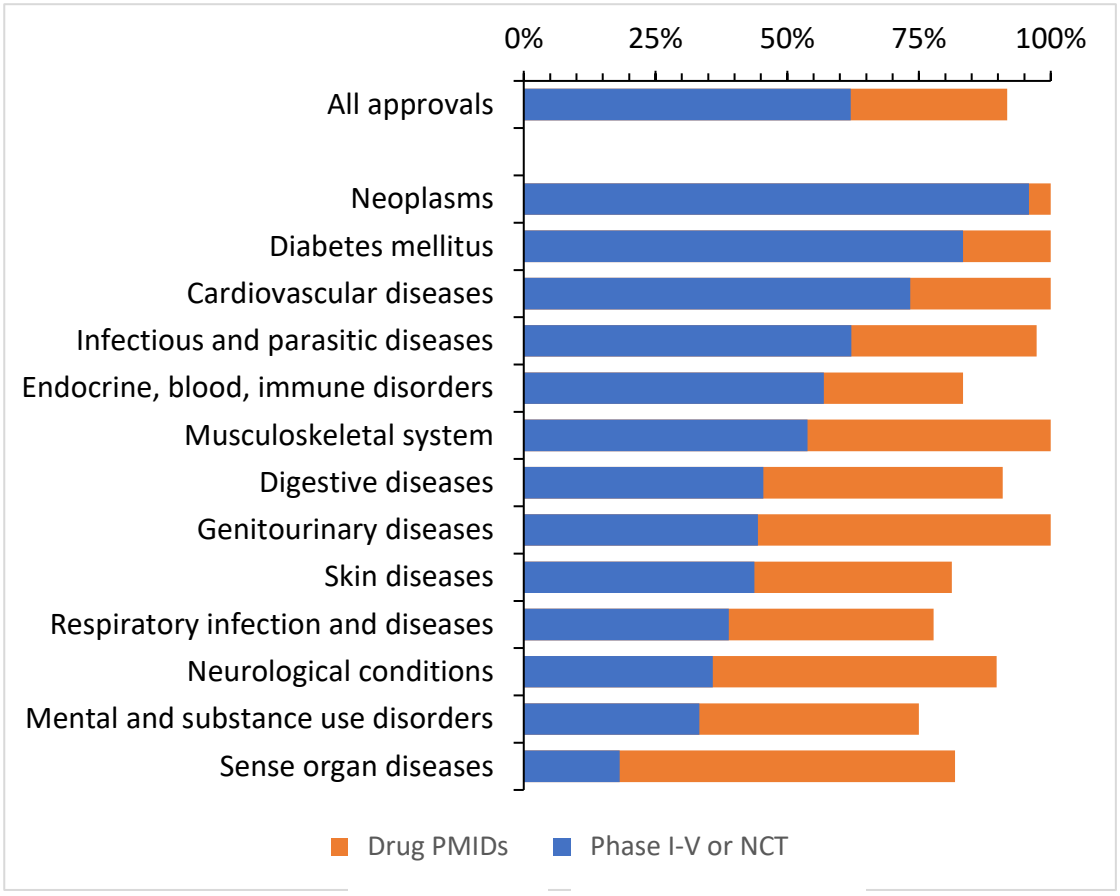


Figure 7. PMIDs, APYs, Costs by therapeutic areas.

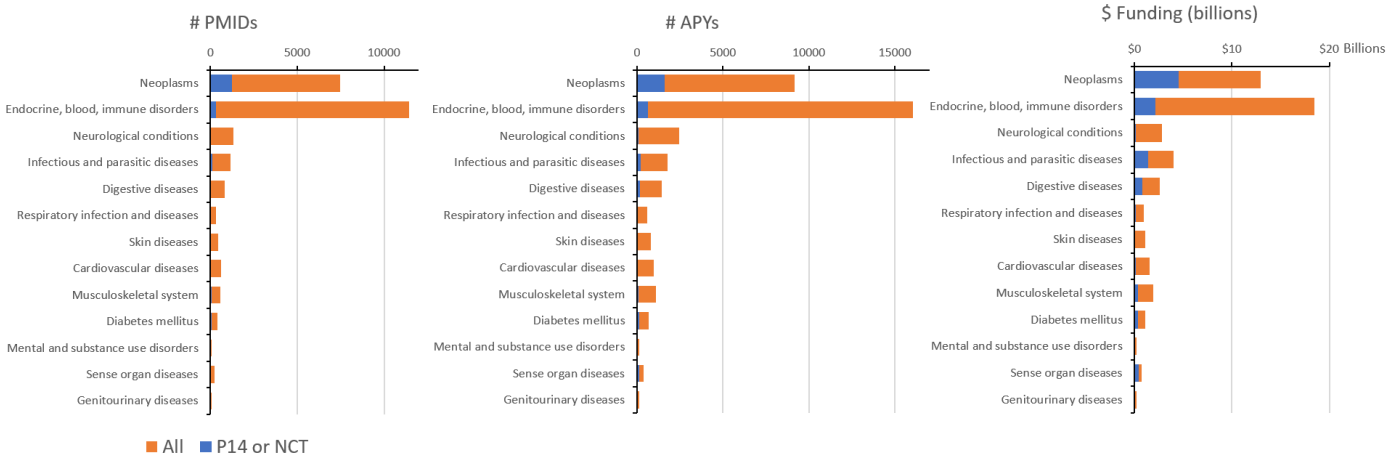


Figure 8. Clinical Trial Search term development using a validation dataset.

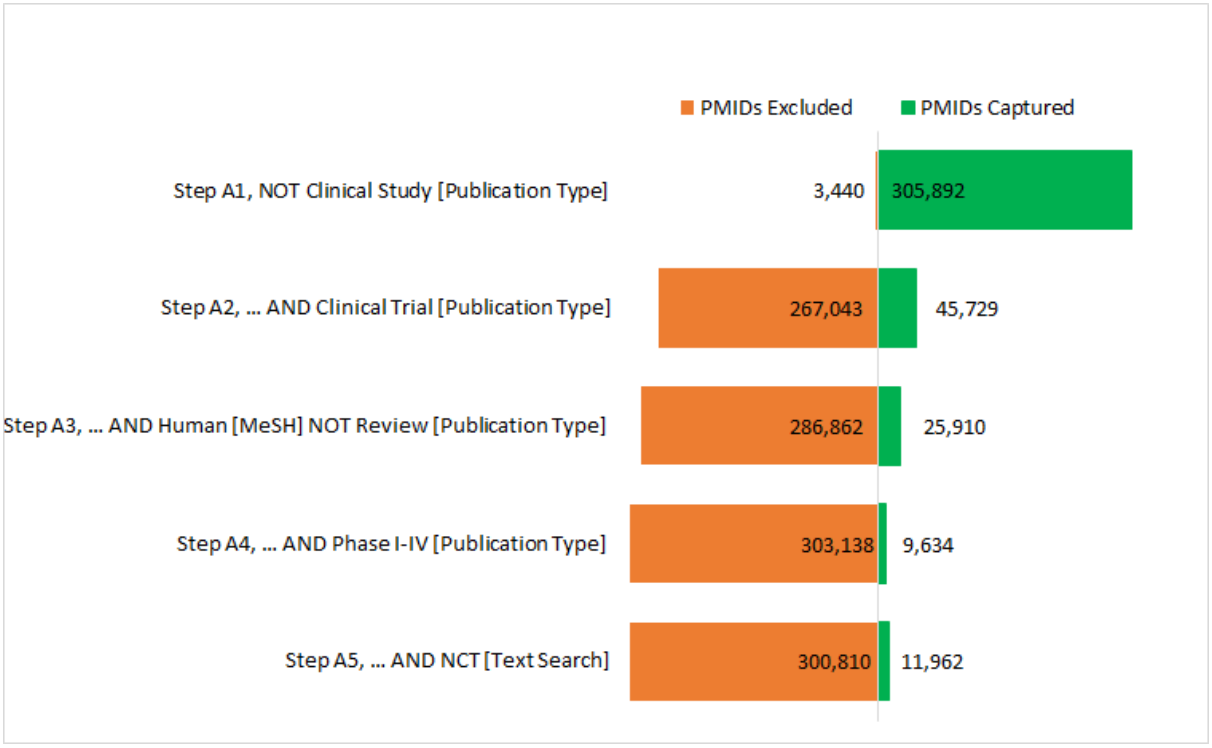


Figure 9. Phase Clinical Trial search term development on a validation dataset.

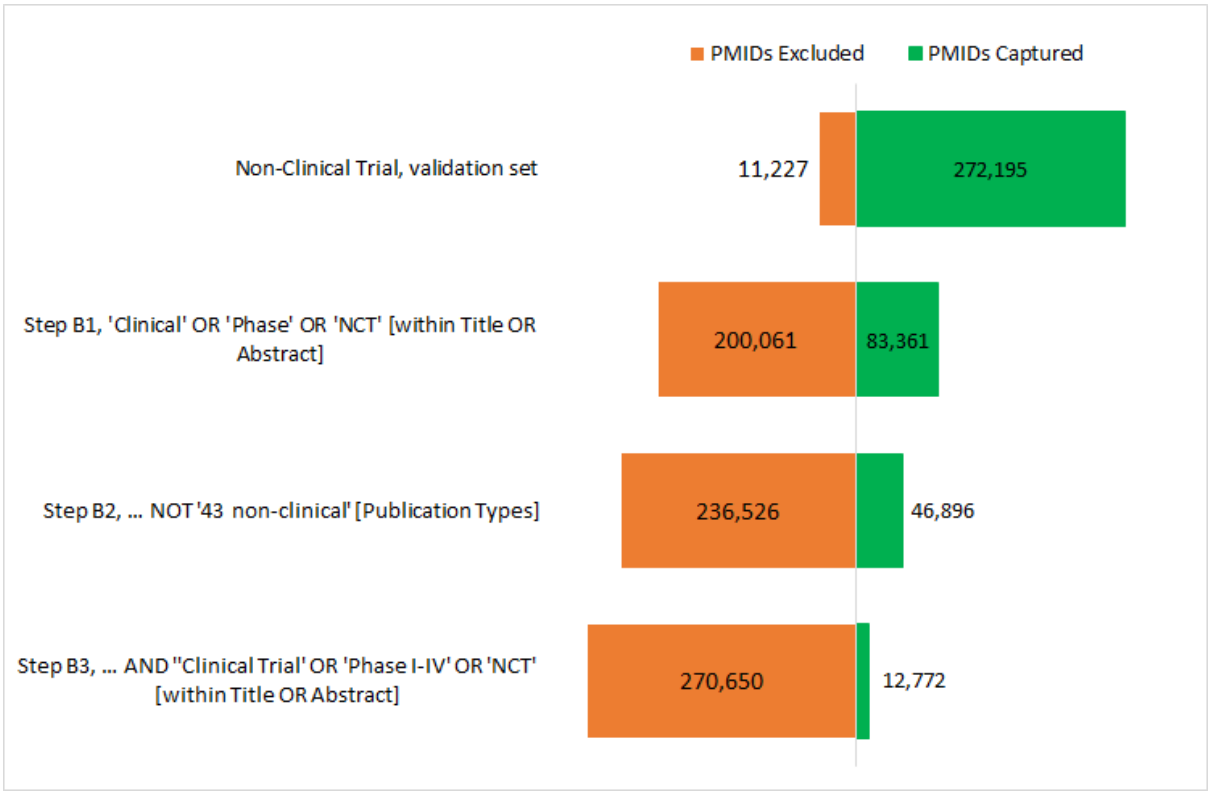
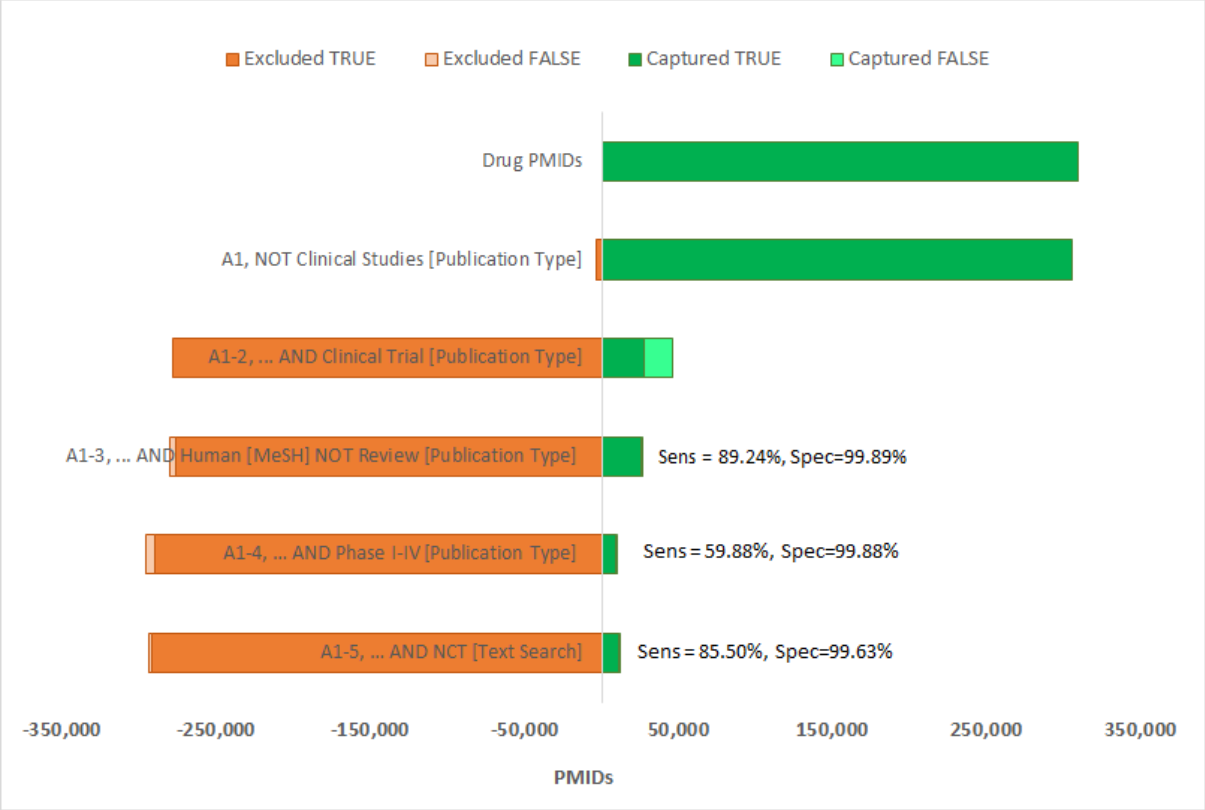


Figure 10. Search term development estimated sensitivity and specificity.



Conflict of Interest Disclosures: All authors report no conflict of interest.

Funding/Support: This work was supported by grants from the National Pharmaceutical Council and the National Biomedical Research Foundation.