Biotechnology Innovation Organization

PharmaIntelligence
Informa

QLS Quantitative Life Sciences

Clinical Development Success Rates and Contributing Factors 2011–2020

About BIO

BIO is the world's largest trade association representing biotechnology companies, academic institutions, state biotechnology centers, and related organizations across the United States and in more than 30 other nations. BIO members, from the entrepreneurial to the Fortune 500 multinationals, are involved in the research and development of innovative healthcare, agricultural, industrial, and environmental biotechnology products. BIO also produces industry-leading investor and partnering events to strengthen the innovator ecosystem.

Bio Biotechnology Innovation Organization

About Informa Pharma Intelligence

Biomedtracker, a subscription-based product of Informa Pharma Intelligence, tracks the clinical development and regulatory history of investigational drugs to assess their Likelihood of Approval (LOA) by the FDA. Biomedtracker is populated in near real-time with updated information from press releases, corporate earnings calls, investor and medical meetings, and numerous other sources. These data can be visualized, analyzed, and interpreted in the specifically designed R&D benchmarking tool Pharmapremia. This provides an exclusive window into drug development pipelines and paints a fuller, clearer picture of how the industry is performing. For more information visit www.pharmaintelligence.informa.com.

Pharma Intelligence
Informa

About QLS Advisors

QLS Advisors LLC is a technology and advisory company based in Cambridge, MA, dedicated to fostering innovation in the life sciences. QLS employs a unique blend of fundamental and quantitative tools to help clients manage risk, assess reward, and develop investment and financing strategies for portfolios of healthcare-related assets. For more information visit www.qlsadvisors.com.

QLS Quantitative Life Sciences

Executive Summary

This is both an updated study of clinical drug development success rates from our 2016 report, and an expansion into the drivers of success with the addition of machine learning modeling to analyze the predictive factors contributing to drug development. A total of 12,728 clinical and regulatory phase transitions were recorded and analyzed from 9,704 development programs over the last decade (2011–2020), across 1,779 companies in the Biomedtracker database. Phase transitions occur when a drug candidate advances into the next phase of development or is suspended by the sponsor. By calculating the number of programs progressing to the next phase vs the total number progressing and suspended, we assessed the success rate at each of the four phases of development: Phase I, II, III, and regulatory filing. Having phase-by-phase data in hand, we then compared groups of diseases, drug modalities, and other attributes to generate the most comprehensive analysis yet of biopharmaceutical R&D success.

This work was made possible due to the years of clinical program monitoring and data entry by Informa Pharma Intelligence's Biomedtracker, which subsequently populates the purpose-built Probability of Technical Success (PTS) tool, Pharmapremia. BIO has long worked with Biomedtracker to calculate success rates based on this data. More recently, BIO and Biomedtracker teamed up with QLS Advisors, a team of expert data scientists who apply machine learning (ML) and artificial intelligence (AI) to clinical trials data and drug properties to assess the features that contribute the most amount, positively or negatively, to the probability of approval. The computational analyses of over 200 different drug, trial, indication, and sponsor metrics that influence clinical success rates were applied in a predictive capacity to produce risk assessments and evaluations of R&D assets with increased accuracy.

Key Takeaways

- The overall likelihood of approval (LOA) from Phase I for all developmental candidates over 2011–2020 was 7.9%.
- Phase II development remains the largest hurdle in drug development, with just 28.9% of candidates achieving this critical phase transition.
- Of the 14 major disease areas, Hematology therapies had the highest LOA from Phase I (23.9%), representing a seven-fold increase over the least successful group, Urology (3.6%).
- Immuno-oncology therapies provide a rare pocket of success in oncology R&D with an overall LOA of 12.4% vs 5.3% for all oncology approaches.
- Rare disease therapies were notably successful with an overall LOA of 17.0%.
- Chronic, high prevalence disease therapies were less successful with an overall LOA of 5.9%.
- Biological complexity in drug modalities generally leads to higher LOA, with CAR-T and RNA interference achieving the highest LOAs of 17.3% and 13.5%, respectively.
- Development programs with trials employing patient preselection biomarkers have two-fold higher LOAs (15.9%), driven by a Phase II success rate of nearly one-in-two.
- The top contributing factors toward phase success are disease indication, target, modality, and drug novelty.
- On average, it takes 10.5 years for a Phase I asset to progress to regulatory approval. Disease areas with above-average LOAs tend to have the shortest development timelines.

Contents

Introduction

This study aimed to measure clinical development success rates, contributing factors to those outcomes, and timelines of clinical trials. With the goal of providing current benchmarking metrics for drug development, this study covers the most recent decade of individual drug program phase transitions from January 1, 2011, to November 30, 2020. A phase transition is defined as the movement out of a clinical phase – for example, advancing from Phase I to Phase II development, or being suspended after completion of Phase I development.

Three separate analyses were conducted. Part 1 of this report includes clinical development success rates based on 12,728 transitions in the Biomedtracker database. These transitions occurred in 9,704 clinical drug development programs over the last decade. Within this broad set of data, we segmented and analyzed success in drug development across levels of novelty, molecular modalities, and disease indications. As this study cuts across 1,779 companies, roughly the entire spectrum of biopharma companies was included. This is illustrated with 15 companies each contributing more than 100 transitions, while 748 smaller biotechs contributed just one transition each. Only company-sponsored, FDA registration-enabling development programs were included in this analysis. Investigator-sponsored studies and combinations with other investigational drugs were excluded. A more detailed description of the data collection, composition, and analysis methodologies is provided in the "Methods" section.

The second analysis applies big data analytics to uncover the underlying drivers of drug development success. Despite the successful application of machine learning techniques to libraries of billions of chemical and biological compounds in order to identify potential drug candidates, most drug developers and financial analysts still rely on historical observation to derive estimates of the clinical trial success rate when analyzing pipelines of investigational drugs. QLS Advisors, a technology and advisory company dedicated to fostering innovation in the life sciences, applies machine-learning techniques to predict the outcomes of randomized clinical trials. In Part 2 of this report, QLS Advisors explores how investors in the biopharma industry can use machine learning to characterize the probability of success for regulatory approval of a novel therapy.

An analysis of clinical timelines is included in Part 3. Here we measure the time it takes for drug development programs at each clinical phase to transition upon success to the next phase.

Part 1. Phase Transition Success and Likelihood of Approvals

Success rates for individual phases of the drug development process were determined by dividing the number that successfully advanced to the next phase by the total number advanced and suspended. This "advanced and suspended" number is often referred to as "n" in this report and should be taken into account when drawing conclusions from the success rate results.

Consistent with previous studies of drug development phase transition success rates, we found Phase II success rates to be far lower than any other phase.[1,2] Phase I and III rates were substantially higher than Phase II, with Phase I slightly lower than Phase III. The highest success rate of the four development phases was the New Drug Application (NDA)/ Biologic License Application (BLA) filing phase (**Figure 1**).

The Phase I transition success rate was 52.0% (n=4,414). As this phase is typically conducted for safety testing and is not dependent on efficacy for candidates to advance, it is common for this phase to have a higher success rate among the clinical phases across most categories analyzed in this report. However, Phase I success rates also may benefit from delayed reporting or omission bias, as some larger companies may not deem failed Phase I programs as material and thereby not report them in the public domain.

The Phase II transition success rate (28.9%, n=4,933) was substantially lower than Phase I, and the lowest of the four phases studied. As this is generally the first stage where proof-of-concept is deliberately tested in human subjects, Phase II consistently has the lowest success rate of all phases. This is also the point in development where industry must decide whether to pursue large, expensive Phase III studies, or to terminate development – this may be done for multiple reasons, including commercial viability. The second-highest phase transition success rate was found in Phase III (57.8%, n=1,928). This is significant as most company-sponsored Phase III trials are the longest and most expensive trials to conduct.

The probability of FDA approval after submitting an NDA or BLA, taking into account re-submissions, was 90.6% (n=1,453).

Overall phase transition success rates

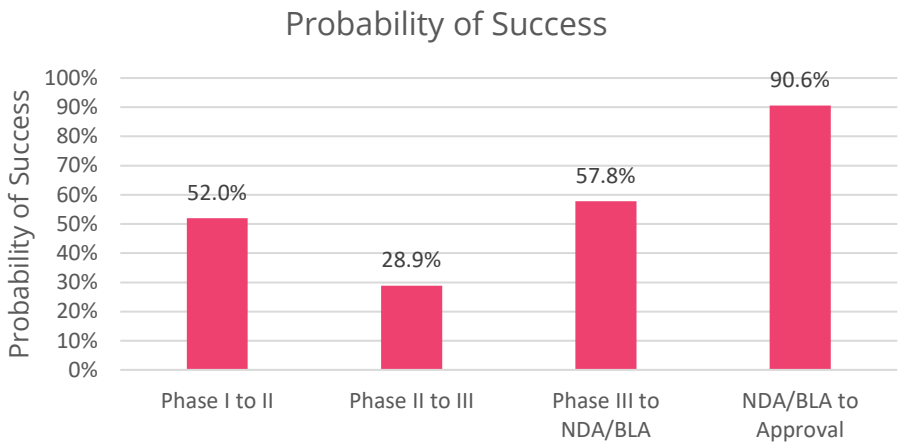


Figure 1: Phase transition success rates from Phase I for all diseases, all modalities. Source: Biomedtracker® and Pharmapremia®, 2020.

[1] Hay M, Thomas D, Craighead JL, Economides C, Rosenthal J (2014). Clinical development success rates for investigational drugs. Nature Biotechnology, 32(1), 40–51. doi: 10.1038/nbt.2786

[2] Thomas D, Burns J, Audette J, Carroll A, Dow-Hygelund C, Hay M (2016) Clinical Development Success Rates 2006-2015. Available here [Accessed 15 January 2021].

Phase Transition Success – By Disease Area

Major disease areas were segmented according to the convention used by Biomedtracker and categorized into 21 major diseases and 623 indications for the 2011–2020 timeframe. For reporting at the disease area level, in **Figure 2** we analyzed 14 major groupings: Allergy, Autoimmune, Cardiovascular, Endocrine, Gastroenterology (non-IBD), Hematology, Infectious disease, Metabolic, Neurology, Oncology, Ophthalmology, Psychiatry, Respiratory, and Urology. The remaining disease areas were placed into the "Other" category. This includes Dermatology, Renal, Obstetrics, Rheumatology (for non-autoimmune indications), ENT/Dental, and Orthopedics. Beneath these major disease areas are 573 indications, which will be analyzed and discussed in subsequent reports.

Phase transition success rates by disease area

Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	n	Phase POS	n	Phase POS	n	Phase POS	n	Phase POS
Hematology	92	69.6%	106	48.1%	82	76.8%	72	93.1%
Metabolic	136	61.8%	149	45.0%	66	63.6%	48	87.5%
Infectious disease	403	57.8%	414	38.4%	197	64.0%	156	92.9%
Others	154	63.6%	228	38.6%	90	60.0%	69	88.4%
Ophthalmology	88	71.6%	200	35.5%	82	51.2%	45	91.1%
Autoimmune	413	55.2%	471	31.4%	219	65.3%	202	94.1%
Allergy	55	56.4%	92	28.3%	34	64.7%	20	100.0%
Gastroenterology	45	46.7%	73	34.2%	35	57.1%	33	90.9%
All indications	4414	52.0%	4933	28.9%	1928	57.8%	1453	90.6%
Respiratory	179	55.9%	215	21.9%	62	64.5%	45	95.6%
Psychiatry	150	52.7%	164	26.8%	71	56.3%	57	91.2%
Endocrine	319	43.3%	293	26.6%	151	66.2%	124	86.3%
Neurology	516	47.7%	504	26.8%	226	53.1%	165	86.7%
Oncology	1628	48.8%	1732	24.6%	495	47.7%	324	92.0%
Cardiovascular	214	50.0%	252	21.0%	105	55.2%	80	82.5%
Urology	22	40.9%	40	15.0%	13	69.2%	13	84.6%

Figure 2: Phase transition success rates by disease area. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase. The ordering of disease areas is consistent with the overall likelihood of approval from Phase I, which is analyzed later in Figure 5. Source: Biomedtracker® and Pharmapremia®, 2020

Phase I Transition Success Rates

Success rates by disease area for Phase I ranged from 40.9% to 71.6%, with the average for all disease indications coming in at 52.0%. Ophthalmology and Hematology were both well above the average rate, achieving respective successful Phase I transitions of 71.6% (n=88) and 69.6% (n=92). With the exception of these two major disease areas, the remainder were all within a reasonable distance from the mean.

Phase II Transition Success Rates

In every disease area, Phase II had the lowest transition success rate of the four phases. As shown in **Figure**

3, Phase II success rates ranged from a high of 48.1% (Hematology, n=106) to a low of 15.0% (Urology, n=40). This 33% range of disparity between major disease areas at the Phase II transition is the major contributor to the observed divergences in overall likelihood of approval (LOA), as discussed in the next section. With only Hematology and Metabolic (45.0%) coming close to achieving a one-in-two success rate, these two disease areas are also the leaders when calculating the overall LOA from Phase I, as shown later in **Figure 5**.

Similar to Phase I transition success, the lowest-performing disease groups were Urology, Cardiovascular (21.0%), and Oncology (24.6%). Relative to its placement for Phase I success rate, the Gastroenterology category performed better in Phase II (34.2%), above the average for all indications.

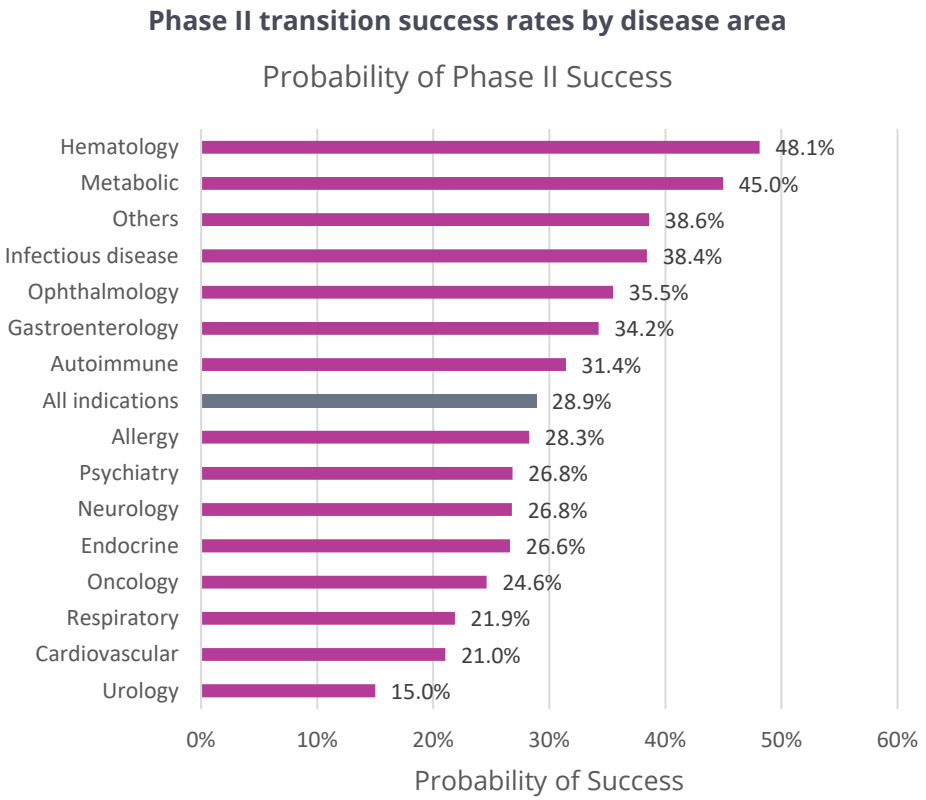


Figure 3: Phase II transition success rates by disease area. Source: Biomedtracker® and Pharmapremia®, 2020

Phase III Transition Success Rates

For Phase III transition success rates, Oncology had the lowest transition success rate (47.5%, n=495). As seen in Figure 4, the Phase III success rates for the remaining 13 specific disease areas were each above 50%.

Five other disease groups besides Oncology ranked below the average Phase III transition success rate of 57.8%, including Ophthalmology (51.2%, n=82) which in the earlier phases performed better than the average across all diseases. Conversely, Urology, which had the lowest transition rates in Phase I and II, had the second-highest Phase III success rate behind Hematology at 69.2%, albeit with a very low n of 13 transitions. All of the disease groups with below-average Phase III success rates had disease indications with large patient populations. Later in this report, we will analyze these high prevalence diseases and compare their success rates to those of therapies for rare diseases.

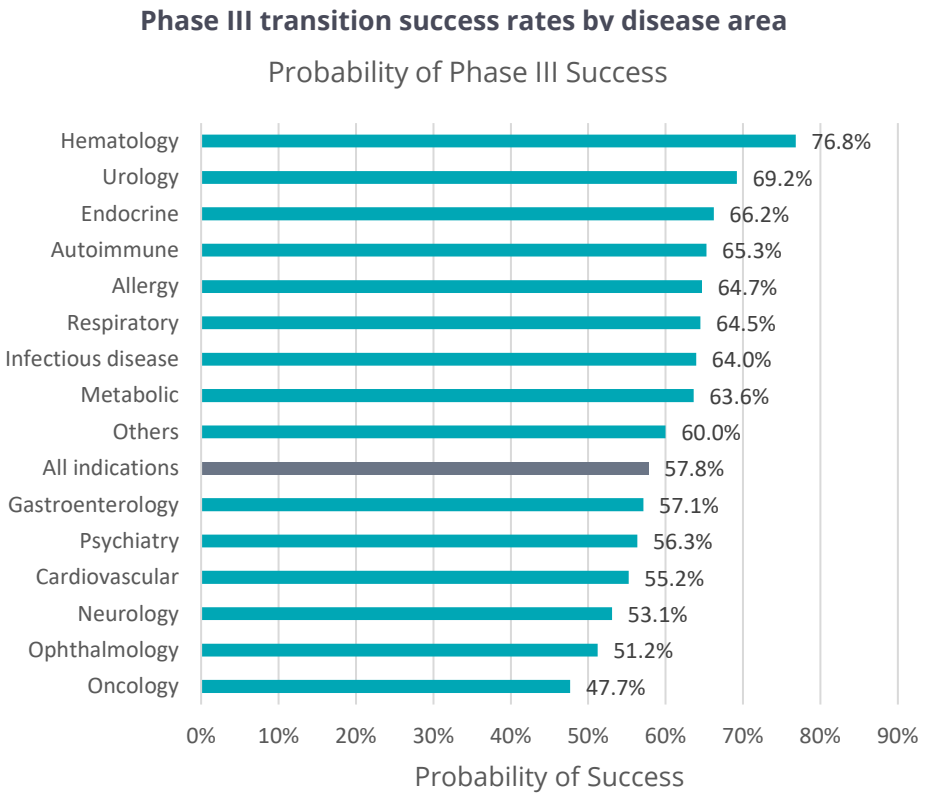


Figure 4: Phase III transition success rates by disease area. Source: Biomedtracker® and Pharmapremia®, 2020

NDA/BLA Submission Success Rates

NDA/BLA transition success rates (approval rates) for the major disease areas ranged from the low end of 82.5% (Cardiovascular) to a high of 100.0% (Allergy). This distribution had the tightest range among the four phases analyzed in this report. These rates are the result of eventual success, not success on the first review. In some cases, programs have more than two Complete Response Letters (CRLs) and attempts at approval. This unlimited allowance of submission attempts pushes the overall success above 90.6% across all diseases, with only 137 drugs suspended by their sponsors at the regulatory transition over 2011–2020 (thus, 1,316 approved).

Among the clinical development programs analyzed, there was an approximately even split of innovative drugs and non-originator products that were successful in obtaining regulatory approval. These included 740 innovative drug programs (435 new molecular entities (NMEs), 278 new biologics, and 27 vaccines) and 576 non-originator products (453 non-NMEs and 123 biosimilars). There is a more detailed discussion about success rates by for drug modality and classification later in the report.

Likelihood of Approval (LOA) – By Disease Area

One of the key measures of success used in this report is the LOA from Phase I. The LOA success rate is simply a multiplication of success rates from all four phases, a compounded probability calculation. For example, if each phase had a 50% chance of success, then the LOA from Phase I would be 0.5 x 0.5 x 0.5 x 0.5 = 6.25%.

Multiplying the individual phase probabilities across all disease areas (found in Figure 1), the compounded probability of progressing from Phase I to U.S. FDA approval reveals that only 7.9% of drug development programs (n=12,728) successfully make it to market.

As can be seen in Figure 5, there is a wide range of LOAs from Phase I. At the high end, Hematology towers over the other disease groups at 23.9% (n=352). Hematology therapies had an LOA from Phase I seven times higher than Urology therapies, which had the lowest Phase I LOA of all the major disease areas (3.6%, n=88).

After Hematology, the next highest LOA from Phase I was Metabolic with 15.5% (n=399). Five other disease areas were above the overall average of 7.9% (Infectious Disease > Ophthalmology > Autoimmune > Allergy > Gastroenterology) which ranged from 13.2% down to 8.3%. Falling under the overall LOA of 7.9%, but very close, were Respiratory (7.5%) and Psychiatry (7.3%). Therapies for five disease categories were well below the average (Endocrine > Neurology > Oncology > Cardiovascular > Urology). The fact that Oncology and Neurology have the two largest n values, while also having low LOA values, indicates that these two disease categories are significant contributors in bringing down the overall industry LOA.

Overall likelihood of approval by disease area

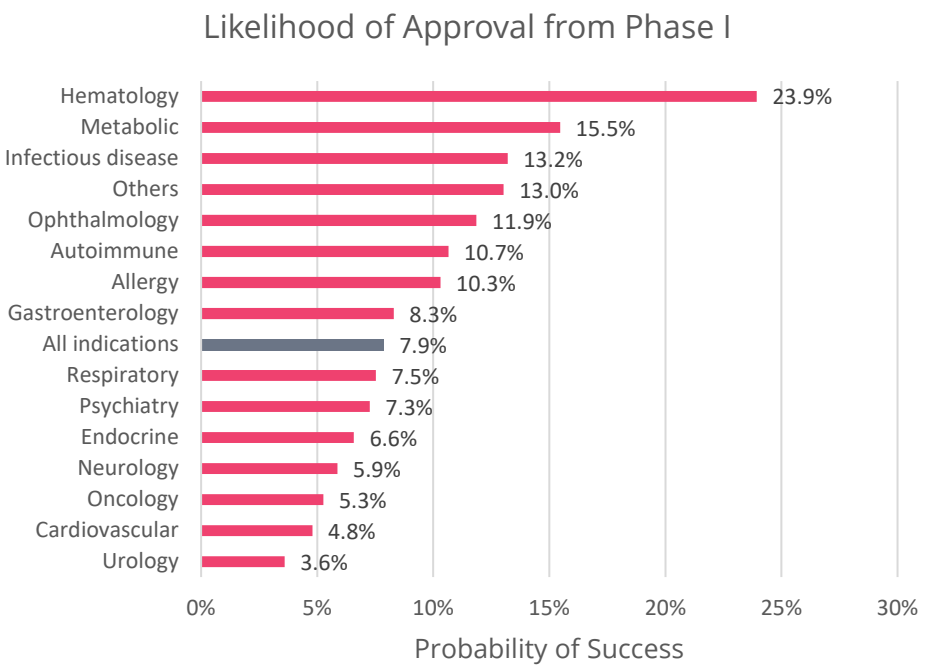


Figure 5a: Chart of LOA from Phase I, displayed highest to lowest by disease area. Source: Biomedtracker® and Pharmapremia®, 2020

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
Hematology	352	23.9%	260	34.4%	154	71.5%	72	93.1%
Metabolic	399	15.5%	263	25.0%	114	55.7%	48	87.5%
Infectious disease	1170	13.2%	767	22.8%	353	59.4%	156	92.9%
Others	541	13.0%	387	20.5%	159	53.0%	69	88.4%
Ophthalmology	415	11.9%	327	16.6%	127	46.7%	45	91.1%
Autoimmune	1305	10.7%	892	19.3%	421	61.4%	202	94.1%
Allergy	201	10.3%	146	18.3%	54	64.7%	20	100.0%
Gastroenterology*	186	8.3%	141	17.8%	68	51.9%	33	90.9%
All indications	12728	7.9%	8314	15.1%	3381	52.4%	1453	90.6%
Respiratory	501	7.5%	322	13.5%	107	61.6%	45	95.6%
Psychiatry	442	7.3%	292	13.8%	128	51.4%	57	91.2%
Endocrine	887	6.6%	568	15.2%	275	57.1%	124	86.3%
Neurology	1411	5.9%	895	12.3%	391	46.0%	165	86.7%
Oncology	4179	5.3%	2551	10.8%	819	43.9%	324	92.0%
Cardiovascular	651	4.8%	437	9.6%	185	45.6%	80	82.5%
Urology	88	3.6%	66	8.8%	26	58.6%	13	84.6%

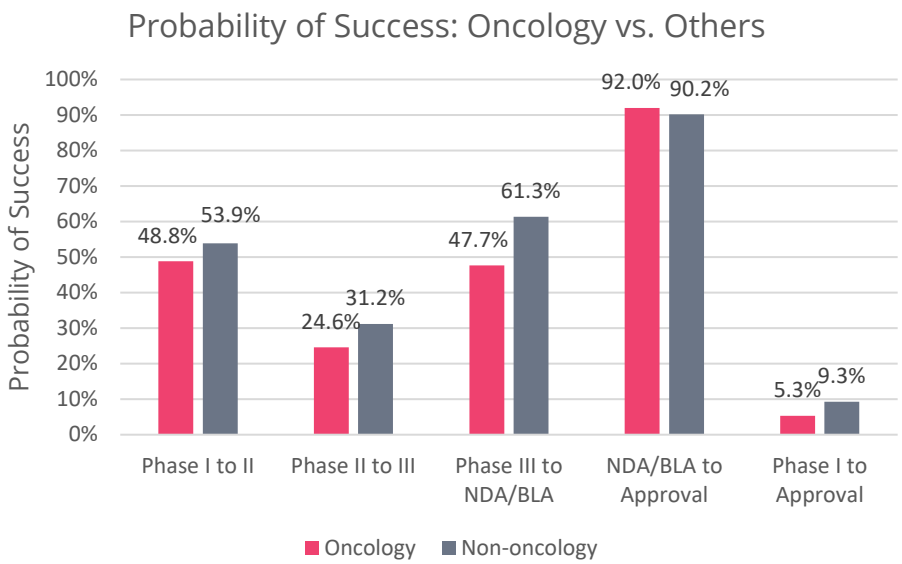
Figure 5b: Table likelihood of approval by disease area with corresponding n values. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. Source: Biomedtracker® and Pharmapremia®, 2020

Oncology and Non-Oncology Diseases

Oncology drug development program transitions in the 2011–2020 period accounted for 33% of the 12,728 total transitions. With the third-lowest LOA from Phase I (5.3%, n=4,179), Oncology had an outsized effect on the overall industry success rate. To further understand this contribution, we compared phase transition success rates and LOA for non-oncology development programs against oncology development programs **(Figure 6)**.

The LOA from Phase I across non-oncology indications, 9.3% (n=8,549), was nearly twice that for Oncology alone, at 5.3% (n=4,179). Comparing individual phase transition success rates in **Figure 2**, Oncology consistently had one of the lowest success rates compared to the other 14 disease categories for every developmental clinical transition. In particular, there were relative differentials of approximately 20% between Oncology and non-oncology groupings at the Phase II and III stages. One notable exception was the NDA/BLA to approval success rate. Oncology performed slightly better than non-oncology at the regulatory transition between NDA/BLA and approval, with a 92.0% (n=324) success rate, as opposed to 90.2% (n=1,129).

Oncology vs. non-oncology phase transition success rates and LOA



Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS
Oncology	1628	48.8%	1732	24.6%	495	47.7%	324	92.0%
Non-oncology	2786	53.9%	3201	31.2%	1433	61.3%	1129	90.2%

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
Oncology	4179	5.3%	2551	10.8%	819	43.9%	324	92.0%
Non-oncology	8549	9.3%	5763	17.2%	2562	55.3%	1129	90.2%

Figure 6: Oncology vs non-oncology phase transition success rates and LOA. Top: Chart of phase transition success rates and LOA from Phase I for Oncology vs non-Oncology. Bottom: Table of phase transition success and likelihood of approval by Oncology vs non-Oncology with corresponding n values. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase, whereas 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. Source: Biomedtracker® and Pharmapremia®, 2020

Oncology drugs were further categorized into two main types of cancer: solid tumors and hematologic cancers. Drugs for solid tumors had more than twice as many transitions in the data set (2,982 vs 1,094), but a much smaller LOA from Phase I vs hematological cancers (4.6% vs 7.5%).

The 2011–2020 time period witnessed the rise of drugs that harness the immune system to treat the underlying cancer, spanning both solid and hematologic types. The first successful immuno-oncology (IO) drugs spawned intense growth in this field over a short period of time, with 679 total phase transitions included in the Biomedtracker database. The success rates of IO drugs far exceed the traditional oncology averages, with a Phase I LOA of 12.4%. The notably high Phase II transition rate of 42.0% for IO drugs – contrasting 24.6% as a whole for oncology – is the main factor underpinning the overall LOA success for the IO class. These oncology groupings are shown in **Figure 7**.

Oncology sub-category phase transition success rates and LOA

Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS
Hematologic	425	50.1%	449	27.8%	120	60.0%	100	90.0%
Solid	1145	48.9%	1261	23.4%	364	42.9%	212	92.9%
IO	275	64.0%	244	40.2%	98	49.0%	62	98.4%

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
Hematologic	1094	7.5%	669	15.0%	220	54.0%	100	90.0%
Solid	2982	4.6%	1837	9.3%	576	39.8%	212	92.9%
IO	679	12.4%	404	19.4%	160	48.2%	62	98.4%

Figure 7: Oncology sub-category phase transition success rates and LOA. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase, whereas 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. Note: Hematologic and solid tumor types are mutually exclusive categories, while the IO drug class includes transitions from across both categories. IO drugs for solid tumors had 3x the transitions as IO hematological cancers but have similar success rates. Source: Biomedtracker® and Pharmapremia®, 2020

Rare and Chronic High Prevalence Diseases

Within the Biomedtracker database, rare diseases can be identified based on meeting either or both of the following standard criteria: affecting fewer than 200,000 people in the US, or prevalence of 1 in 2,000 people in the EU. As 43% of the Oncology transitions are for rare indications, all Oncology indications were removed to make this rare disease analysis more concentrated on inborn genetic disorders. For chronic diseases, we first obtained a list of conditions from the CMS Chronic Conditions Data Warehouse (CCW). We removed any cancer indications, then identified those diseases with greater than 1 million patients affected in the United States.

Rare disease vs. highly prevalent chronic disease success rates

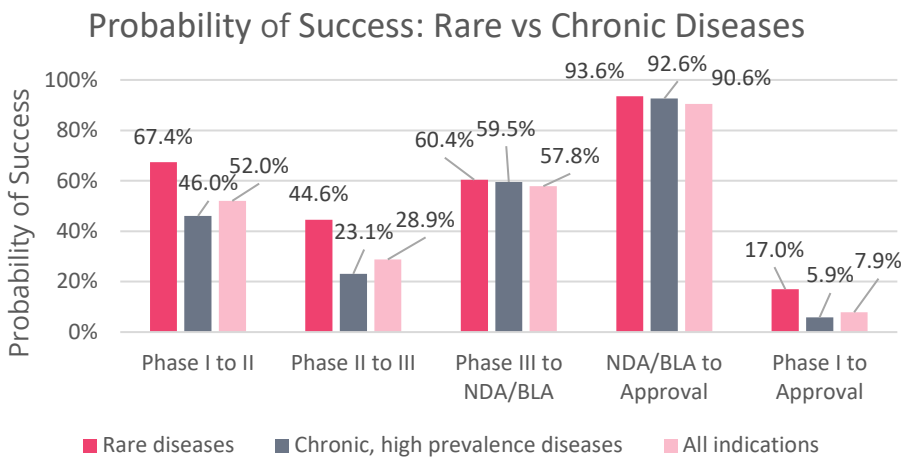


Figure 8a: Non-oncology rare disease and highly prevalent chronic disease phase transition success rates and LOA. Chart of phase transition success rates and LOA from Phase I. Source: Biomedtracker® and Pharmapremia®, 2020

Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS
Rare diseases	380	67.4%	464	44.6%	240	60.4%	172	93.6%
Chronic, high prevalence diseases	745	46.0%	737	23.1%	279	59.5%	217	92.6%
All diseases	4414	52.0%	4933	28.9%	1928	57.8%	1453	90.6%

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
Rare diseases	1256	17.0%	876	25.2%	412	56.6%	172	93.6%
Chronic, high prevalence diseases	1978	5.9%	1233	12.7%	496	55.1%	217	92.6%
All diseases	12728	7.9%	8314	15.1%	3381	52.4%	1453	90.6%

Figure 8b: Non-oncology rare disease and highly prevalent chronic disease phase transition success rates and LOA. Table of phase transition success and likelihood of approval by disease prevalence with corresponding n values. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase, whereas 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. Source: Biomedtracker® and Pharmapremia®, 2020

After identifying programs for non-oncology rare diseases and highly prevalent chronic diseases, we compared their phase transition success rates and LOA as shown in **Figure 8**. At 17.0%, the overall LOA from Phase I for rare diseases (17.0%) was almost three times higher than for chronic, high prevalence diseases (5.9%). As both of these groups exclude Oncology indications, we can compare these results with those of the overall non-oncology figures. The 9.3% LOA (n=8,549) value for all non-oncology indications is meaningfully higher than for chronic, high prevalence diseases, suggesting that these indications are having a negative impact on overall success rates outside Oncology. Chronic, high prevalence diseases accounted for 16% of total transitions, with Oncology providing a further 33%. These two groups, having LOAs well below average, together comprise half of the total dataset, thus negatively affecting the overall industry LOA of 7.9%.

Although excluded from the above analysis, rare disease Oncology indications had a higher LOA than non-rare Oncology indications (6.8% vs 4.4%), largely driven by the higher overall success for hematologic malignancies. Blood cancers typically have lower prevalence and are classified as rare diseases.

Success rates for all four transitions were higher for the rare disease group than any of the overall transition success rates. The largest difference was found in Phase II transition success rates (44.6% for rare disease vs 28.9% overall). There was also a notable gap between Phase I transition rates (67.4% vs 52.0%), although success in Phase III and in the NDA/BLA to Approval transitions were broadly comparable.

Conversely, the success rates for transitions in the highly prevalent chronic disease group were lower than the overall transition success rates in Phase I (46.0% vs 52.0%) and Phase II (23.1% vs 28.9%). The opposite was seen in Phase III and NDA/BLA transitions, where slightly higher success rates were observed: 59.5% vs 57.8% for Phase III and 92.6% vs 90.6% for NDA/BLA.

Drug Classes and Modalities

Drugs in the dataset were annotated as novel or off-patent, with the novel category including new molecular entities (NMEs; mainly small molecules), biologics, and vaccines; and the off-patent group comprised of non-NMEs and biosimilars. As noted in **Figure 9**, transition success rates for novel drugs confirm that they face a more difficult pathway to approval than for off-patent products. The overall Phase I LOA for off-patent therapies (14.7%, n=2,161) was twice as high as that for novel therapies (6.8%, n=10,527).This two-fold increase in success for non-originator products was driven by higher success rates in all three clinical phase transitions, with the greatest divergence observed during Phase III (52.9% vs 70.3%).

Within the novel drugs segment, NMEs had a lower probability of approval from Phase I than either biologics or vaccines. While there was a 5.7% (n=6,803) chance of a Phase I NME gaining approval, the probabilities for biologics and vaccines were 9.1% (n=3,412) and 9.7% (n=312), respectively. Vaccines in particular have been 100% successful in the NDA/BLA to Approval stage (n=27). In the off-patent category, biosimilars performed well, with an overall Phase I LOA of 32.3% (n=277).

Phase transition success rates and LOA from Phase I for drugs based on class and novelty

Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS
"Novel"	3933	51.3%	4394	27.9%	1374	52.9%	826	89.6%
NME	2505	50.6%	2924	25.6%	869	50.6%	505	86.1%
Biologic	1301	52.5%	1355	32.4%	462	56.7%	294	94.6%
Vaccine	127	52.0%	115	32.2%	43	58.1%	27	100.0%
"Off-Patent"	455	60.4%	527	37.6%	552	70.3%	627	91.9%
Non-NME	395	57.5%	523	37.5%	471	67.5%	495	91.5%
Biosimilar	60	80.0%	4	50.0%	81	86.4%	132	93.2%

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
"Novel"	10527	6.8%	6594	13.2%	2200	47.4%	826	89.6%
NME	6803	5.7%	4298	11.2%	1374	43.6%	505	86.1%
Biologic	3412	9.1%	2111	17.4%	756	53.6%	294	94.6%
Vaccine	312	9.7%	185	18.7%	70	58.1%	27	100.0%
"Off-Patent"	2161	14.7%	1706	24.3%	1179	64.6%	627	91.9%
Non-NME	1884	13.3%	1489	23.2%	966	61.8%	495	91.5%
Biosimilar	277	32.2%	217	40.3%	213	80.5%	132	93.2%

Figure 9: Phase transition success rates and LOA from Phase I for drugs based on class and novelty. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase, whereas 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. Source: Biomedtracker® and Pharmapremia®, 2020

Looking a little deeper into drug modality, chimeric antigen receptor T cell (CAR-T) development has seen the most success among selected novel modalities (**Figure 10**). At 17.3% (n=67), the Phase I LOA for CAR-T therapies is more than twice the 7.9% average across all diseases, and more than three times higher than the 5.3% Phase I LOA for all of oncology. Since 2017, three separate CAR-Ts have gained approval, with a total of four successful BLA transitions. The strong R&D progress in RNA interference (RNAi) therapies is reflected in the Phase I LOA for this category, second behind CAR-T at 13.5% (n=70). This class includes three approvals as of December 31, 2020, the first of which came in 2018. Monoclonal antibodies, antibody drug conjugates (ADCs), and gene therapies also each have over a 10% likelihood of approval at the Phase I stage.

Success rates by modality

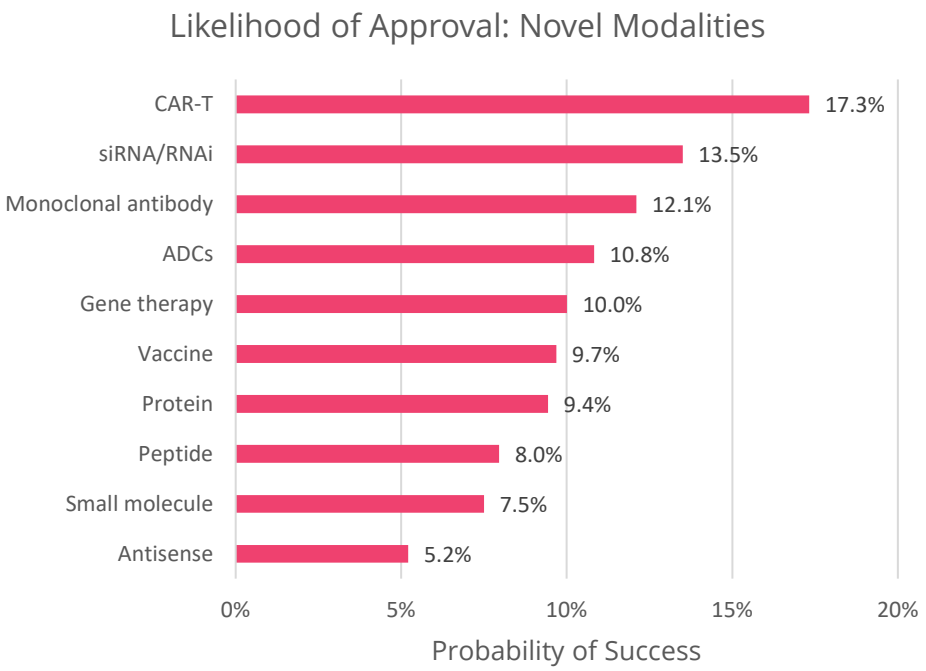


Figure 10a: LOA from Phase I for drugs based on modality. Chart of LOA from Phase I, displayed highest to lowest by drug modality.

Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS
CAR-T	43	44.2%	17	58.8%	3	66.7%	4	100.0%
siRNA/RNAi	40	70.0%	38	28.9%	6	66.7%	3	100.0%
Monoclonal antibody	804	54.7%	740	34.1%	310	68.1%	282	95.4%
ADCs	103	41.7%	53	41.5%	16	62.5%	12	100.0%
Gene therapy	27	51.9%	57	38.6%	10	50.0%	2	100.0%
Vaccine	129	52.7%	117	31.6%	43	58.1%	27	100.0%
Protein	246	51.6%	288	33.0%	149	61.7%	117	89.7%
Peptide	234	53.0%	218	28.4%	100	60.0%	67	88.1%
Small molecule	2308	52.6%	2896	28.0%	1118	56.9%	849	89.5%
Antisense	69	60.9%	70	20.0%	14	64.3%	9	66.7%

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
CAR-T	67	17.3%	24	39.2%	7	66.7%	4	100.0%
siRNA/RNAi	87	13.5%	47	19.3%	9	66.7%	3	100.0%
Monoclonal antibody	2136	12.1%	1332	22.1%	592	64.9%	282	95.4%
ADCs	184	10.8%	81	25.9%	28	62.5%	12	100.0%
Gene therapy	96	10.0%	69	19.3%	12	50.0%	2	100.0%
Vaccine	316	9.7%	187	18.4%	70	58.1%	27	100.0%
Protein	800	9.4%	554	18.3%	266	55.4%	117	89.7%
Peptide	619	8.0%	385	15.0%	167	52.8%	67	88.1%
Small molecule	7171	7.5%	4863	14.3%	1967	50.9%	849	89.5%
Antisense	162	5.2%	93	8.6%	23	42.9%	9	66.7%

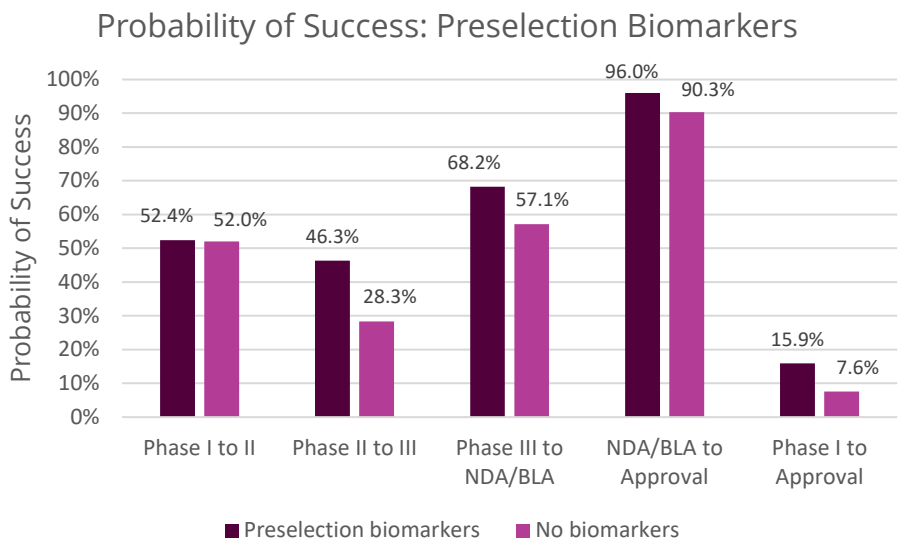
Figure 10b: LOA from Phase I for drugs based on modality. Table of phase transition success and likelihood of approval by modality with corresponding n values. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase, whereas 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. *n values for vaccine differs slightly from Figure 9 owing to variable classification of cellular vaccines. Source: Biomedtracker® and Pharmapremia®, 2020

Patient Preselection Biomarkers

A greater understanding of human disease – whether at the molecular or genomic level – ultimately leads to the investigation of personalized medicine. Indications are increasingly segmented by biomarkers in order to match patients with the treatments most likely to show the greatest benefit, according to the underlying drug mechanism and disease pathophysiology. We identified 767 phase transitions out of 12,728 (6%) that incorporated patient preselection biomarkers in their corresponding clinical trial design. This was accomplished by mapping Informa Pharma Intelligence's Biomedtracker and Trialtrove databases, to provide the supplemental level of clinical trial detail. For a detailed description please refer to the Methods section.

The LOA from Phase I segmented by biomarkers can be found in **Figure 11**. Drug development programs with trials employing patient preselection biomarkers have a two-fold higher LOA (15.9%) than those that do not (7.6%).

Success rates by use of patient preselection biomarkers



Phase Success	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS	POS n	Phase POS
Preselection biomarkers	414	52.4%	149	46.3%	129	68.2%	75	96.0%
No biomarkers	4000	52.0%	4784	28.3%	1799	57.1%	1378	90.3%

Likelihood of Approval	Phase I to Approval		Phase II to Approval		Phase III to Approval		NDA/BLA to Approval	
	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA	LOA n	Phase LOA
Preselection biomarkers	767	15.9%	353	30.3%	204	65.5%	75	96.0%
No biomarkers	11961	7.6%	7961	14.6%	3177	51.5%	1378	90.3%

Figure 11: Patient preselection biomarker phase transition success rates and LOA. Top: Chart of phase transition success rates and LOA from Phase I. Bottom: Table of phase transition success and likelihood of approval by biomarker status with corresponding n values. The n value is the total 'Advanced or Suspended' transitions of all phases used to calculate LOA. 'POS' is the probability of successfully advancing to the next phase, whereas 'Phase LOA' is the probability of FDA approval for drugs from this phase of development. Source: Biomedtracker®, Pharmapremia®, and Trialtrove®, 2020

The advantage of patient preselection biomarkers varies according to the trial stage. At Phase I, there is no discernible difference between the two discrete sets, with both biomarker and non-biomarker programs achieving close to the average success rate of 52.0%. The benefit is clearly greatest at Phase II, where the use of preselection biomarkers enables 46.3% of programs to advance (n=149), compared to just 28.3% of those without. A similar advantage, albeit smaller in magnitude, is also observed with Phase III transitions – 68.2% for biomarker-supported programs vs 57.1% for those without. Success at the NDA/BLA transition is largely contingent on Phase III trial design and so this benefit carries over from the largest clinical phase.

While this analysis only represents a small subset of the overall data, the degree of difference between individual phase success rates and overall LOA from Phase I builds confidence in the pursuit of drug development programs targeted at biomarker-enriched patient populations. Such assets are likely to advance through clinical development with lower levels of attrition, and should in theory improve patient outcomes via the advent of increasingly personalized medicine.

Part 2. Predictive Analysis of Clinical Success

In Part 1, we estimated success rates using the drug-indication pathways gathered from the Biomedtracker database. We can now incorporate additional information, called "features," for each drug development program using machine learning methods to produce forward-looking measures—forecasts—of the outcome of ongoing programs. This type of analysis provides more timely and relevant estimates than standalone analyses of historical success rates.

Some examples of features included in this predictive analysis are the characteristics of the drug, its indication, its sponsor, and its clinical trial design. There are more than 200 features used in the QLS forecasts, a sample of which can be found in **Figure 12**. In much the same way that linear regression models employ regressors, or "right-hand-side" variables, to predict the "left-hand-side" or the dependent variable, these features contain useful signals about drug development that allow the outcomes of drug development programs to be more accurately predicted than before.

Sample of the 200-plus features driving success rate probabilities

Drug Features	Indication Features	Sponsor Features	Trial Features
Drug classification	Therapeutic area	Entity type	Phase
Compound type	Disease subgroup	Headquarters location	Status
Biological target	Prevalence	Number of approvals	Site locations
Mechanism of action	Incidence	Therapeutic area track record	Actual-to-target accrual ratio
Prior approvals	Prior approvals	Phase success track record	Intervention model

Figure 12: Sample of the 200-plus features extracted from Pharmaprojects, Trialtrove, and Biomedtracker.

The QLS approach uses state of the art machine-learning techniques to analyze clinical trial data. While the details of the QLS model are proprietary, an example of a commonly used machine-learning algorithm is a random forest classifier. The random forest classifier is an algorithm that takes the average prediction from a multitude of yes-no decision trees (i.e., a "forest") that are applied to the features (see **Figure 13**).

When training a decision tree, one task is to learn the most informative questions to ask at each branch point of the tree. For example, the algorithm might ask, "Is the drug sponsor a big pharma company?" If the answer is "yes," then the algorithm would move to the "yes" branch of the decision tree and then ask, "Is the trial double-blind?" On the other hand, if the answer is "no," then the algorithm would move to the "no" branch of the decision tree, and ask, "Is the drug a small molecule?" and so on, until enough information has been collected about the drug-program to make a prediction about its success or failure. By identifying subtle patterns in a large enough database of historical drug transitions, machine-learning algorithms can often make remarkably accurate predictions. Additional details on these techniques can be found in the Methodology section.

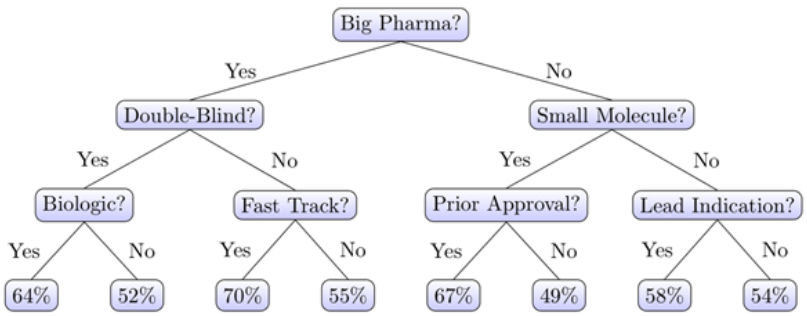


Figure 13: A hypothetical representation of a single, simplified decision tree. The percentages at the "leaves" of the tree denote the fraction of training data samples that are categorized by a given pathway.

Feature Importance

Given the stakes involved in the drug development process, machine learning forecasts must not only be accurate, but also interpretable to stakeholders charged with the responsibility of making go/no-go decisions. These decision-makers need to understand why a given forecast differs from the historical average, and which features were most important in driving the difference or delta with respect to the disease-group baseline LOA. To increase the transparency of our forecasts, the QLS machine learning algorithm reports the most important features and their individual contributions to the forecast's delta.

As an illustration, we decompose our probability of approval (POA) prediction of an oncology drug that is currently in phase 3 clinical trials into its key components. (Note that we use the term POA for the QLS machine-learning forecasts to differentiate from the empirical "LOA" presented in Part I.) **Figure 14** reports the top five features that increase this program's estimated probability of approval from its therapeutic area historical baseline of 35.0% from phase 3 to approval, to 71.8%. The top positive feature driving this higher than average probability is its breakthrough therapy designation (+20.6%). This designation is granted when the FDA has determined that preliminary clinical evidence indicates that the drug may demonstrate substantial clinical improvement over available therapies. Further positive clinical evidence in phase 2 increases the estimate by an additional +6.4%. Next, because it is a PARP inhibitor (+4.6%) that has been previously approved for another indication (+3.6%), the POA is augmented by an additional 8.2%. Finally, its treatment of a solid tumor type (-2.0%) and the net aggregate contribution of other features (-2.8%) penalize the overall POA score slightly.

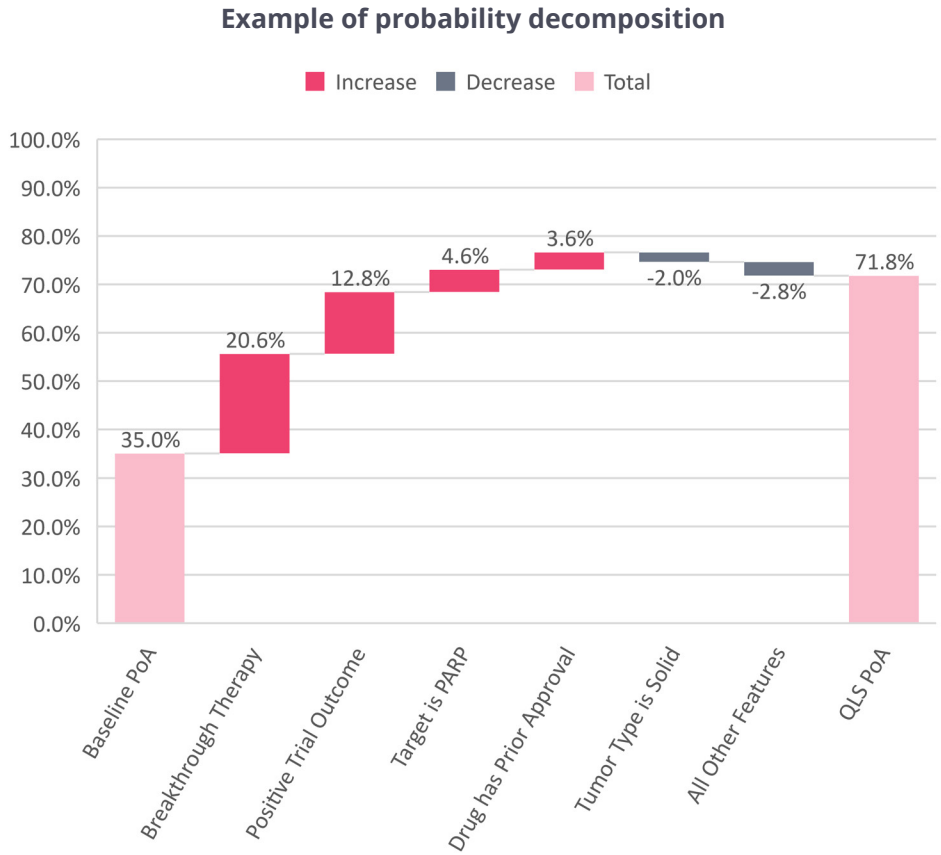


Figure 14: Decomposition of the probability of approval estimate for an anti-cancer drug that is currently in phase 3 clinical trials for prostate cancer. The top five features that increase this program's estimated probability of approval from the historical phase 3 to approval baseline in oncology of 35.0% to 71.8% are reported. Abbreviations: POA=probability of approval; PARP= poly ADP-ribose polymerase.

Although each *individual* prediction may have unique drivers, we can also extract the most informative variables across *all* predictions to gain insight into some of the most common predictors of success. **Figure 15** summarizes our results.

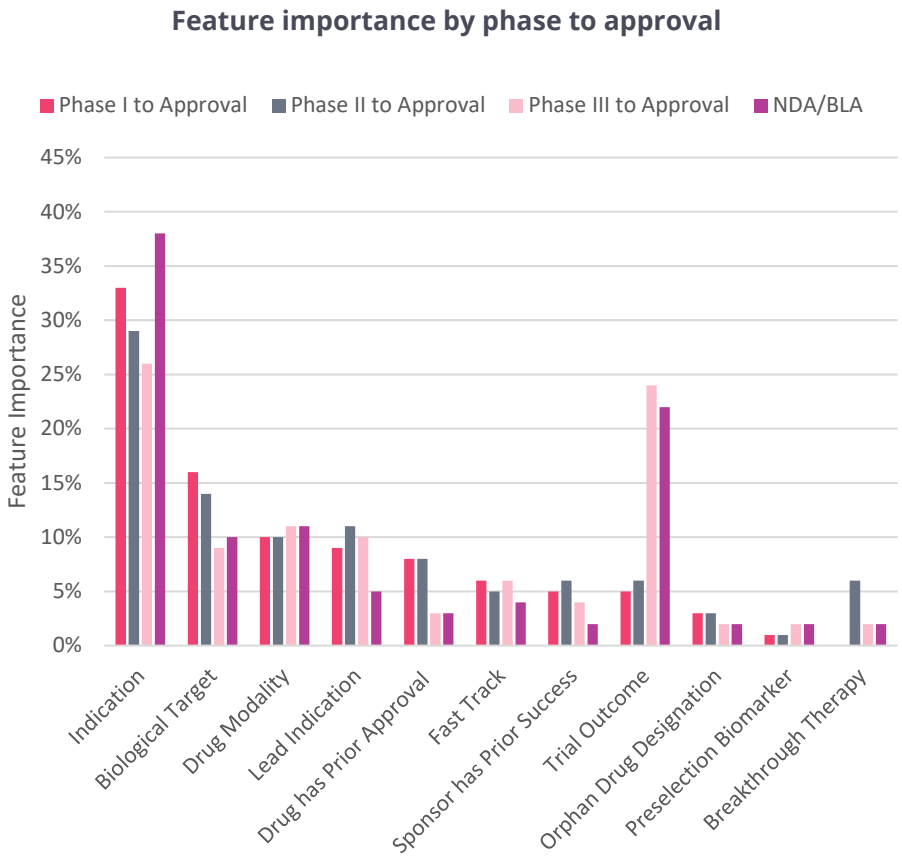


Figure 15: The most important features of our random forest classifier across all predictions. Feature importance is normalized to sum to 100%. Note that individual predictions may have unique drivers specific to a given therapeutic area, trial design, or drug profile that are not listed here. Source: QLS Advisors.

The indication was consistently ranked the top variable across all clinical development phases. Indeed, success rates vary substantially across indications even within a therapeutic area. For example, in oncology, the overall phase 1 to approval LOA ranges from a minimum of 1.1% (n=275) for pancreatic cancer to a maximum of 15.2% (n=33) for alimentary cancers. We also observe that lead indication status, prior approval of a drug for another indication, and biological target validation all have a significant impact on the probability of approval. In some cases, developing an already approved drug for a new indication—one that has a controlled manufacturing process, and has already been shown to be safe in humans—has a greater likelihood of success than a novel indication. In other cases, the success rates for lead indications may be higher if a sponsor initiates clinical trials for multiple follow-on indications for which the drug was not originally intended, and many of the initiated clinical trials for the same drug fail.

Analysis showed that the trial outcome (whether the trial was completed, with its primary endpoints met) has significant associations with late-stage clinical success. It is easy to imagine that a drug-indication pair whose trial failed to meet its endpoints has a low probability of success in advancing to approval. For example, as of December 31, 2020, our algorithm predicts that one specific BLA for Alzheimer's disease has only a 65% chance of progressing from regulatory review to approval, even though historically 83% of neurology drugs have made the transition to approval once they have reached this stage. The observation that the biologic failed to meet its primary endpoints in phase 3 is the top contributor (-15%) to this adjustment from the

baseline. In contrast, candidates that achieve positive outcomes have a higher probability of success.

The use of patient pre-selection biomarkers in clinical trials has become more common, and it has been reported that trials using biomarkers are more likely to succeed.[3] Consistent with these findings, **Figure 15** shows that the use of patient pre-selection biomarkers has an important impact on success rates. For example, in the past decade, the number of non-targeted therapies in oncology has declined, while the use of targeted agents and pre-selection biomarkers has risen dramatically. Between 2016 and 2020, 85% (n=136) of oncology approvals were targeted agents. Moreover, 37% (n=51) involved either pre-selection against a driver mutation (26%; n=36) or pre-selection using a tumor-specific antigen (11%; n=15).

In **Figure 15**, drug modality can be seen as another important predictor of success. As an illustrative example, monoclonal antibodies have had higher phase-1-to-approval success rates (5.9%; n=935) than other modalities within oncology. Moreover, immunotherapies focused on PD-1 or PD-L1 have been tremendously successful from phase I (23.9%; n=71). In contrast, other immunotherapies, while numerous, have been much more likely to result in failure from phase I (1.9%; n=945).

Finally, we find that sponsor track record, quantified by the number of previous successful phase transitions in a given therapeutic area, is also a useful indicator for clinical program predictions. The positive correlation between track record and future success is likely associated with operational experience in conducting clinical trials and navigating the regulatory review process.

Performance

How well does machine learning predict success? A number of criteria have been proposed to measure machine learning performance, some of them highly complex, but perhaps the most straightforward metric is to see how many correct and incorrect predictions there are on an out-of-sample data set. We analyzed the accuracy of our machine learning predictions by first splitting our dataset into a training set (80%) and an out-of-sample test set that consisted of the most recent 20% of data. **Figure 16** summarizes the results for predictions for drug development programs transitioning through approval. We found that drugs with higher probability estimates were more likely to be approved. Moreover, the predictions were relatively well calibrated; when the model predicted a drug-indication pair had a certain probability of approval, the drug was subsequently approved at approximately the same rate.

[3] Wong CH, Siah KW, Lo A (2019). Estimation of clinical trial success rates and related parameters. Biostatistics, 20(2), 273–286. doi: 10.1093/biostatistics/kxx069

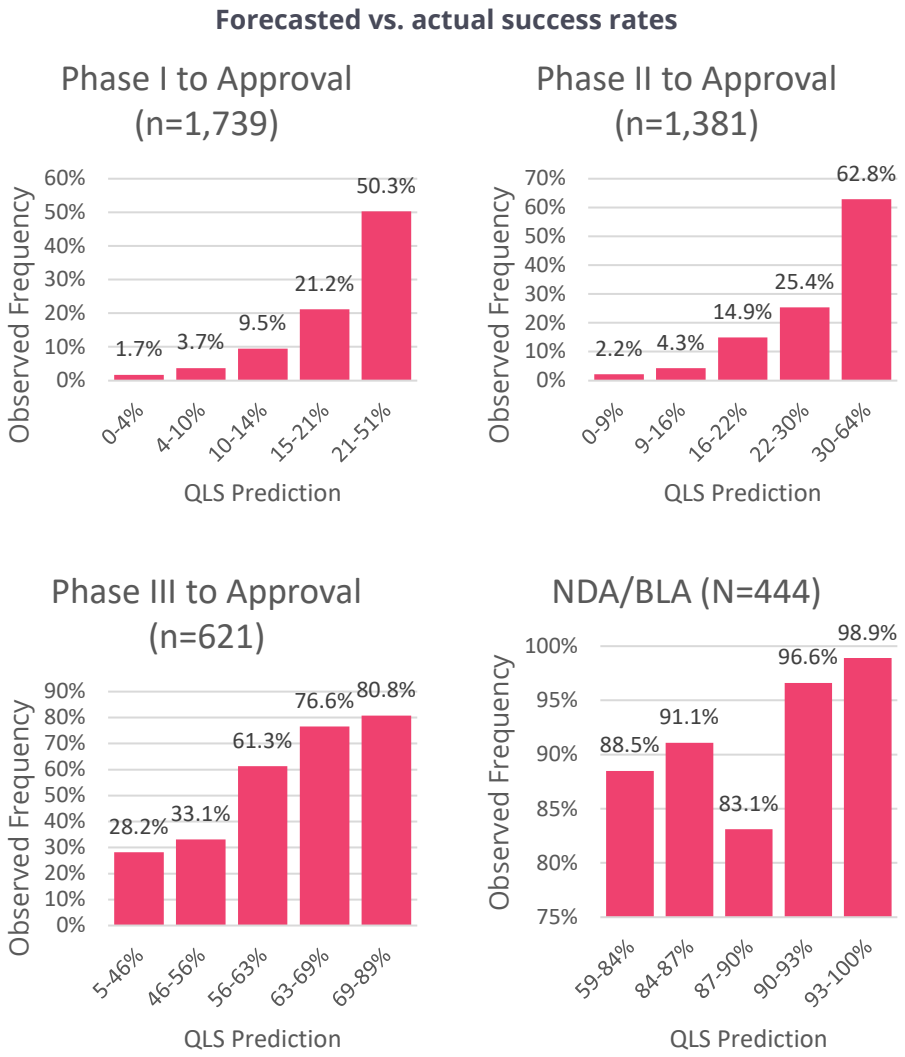


Figure 16: Histogram of machine-learning forecasts of drug development programs transitioning to approval using the most recent 20% of our dataset. Higher forecasts are associated with higher rates of approval, which indicates positive forecast power. Estimates for each phase are grouped into quintiles.

However, the results also demonstrate that these predictions are by no means perfect. There is considerable noise in these estimates. In certain cases, our model tends to slightly underestimate the approval rates in the most recent out-of-sample period. For example, success rates in oncology have been increasing, concurrently with the emergence of genomic technologies, an improved scientific understanding of cancer biology and ways to disrupt cancer cells, as well as an evolving regulatory environment that includes a higher number of accelerated approvals. As we incorporate more data, more features, and the latest scientific and medical knowledge into the algorithm, these forecasts will improve, which will occur naturally over time as these methods become more popular. The QLS model is trained on the full dataset, which includes the most recent development programs and approvals.

Part 3. Drug Development Timelines

In addition to determining whether drugs are advanced or suspended at the end of a phase transition, the Biomedtracker and Pharmapremia data can also be analyzed to yield time spent at each clinical stage, and overall drug development timelines. These are valuable metrics as there are considerable opportunity costs associated with investing in an R&D process that may take up to a decade, and a simple analysis of clinical trial durations excludes the contribution of internal decision making and strategic execution from the overall timeline.

Based on 6,151 successful phase transitions over the 2011–2020 period, it took an average of 10.5 years for a drug to successfully progress from Phase I development to regulatory approval. This includes 2.3 years at Phase I, 3.6 years at Phase II, 3.3 years at Phase III, and 1.3 years at the regulatory stage.

Phase duration can vary greatly according to numerous factors, such as disease area and indication, best practices of clinical trial design, and patient availability. Accordingly, **Figure 17** evaluates drug development timelines for the major 14 categorized disease areas.

Disease areas with above-average LOAs tend to be associated with shorter development timelines. Five of the seven best-performing disease areas by LOA fall beneath the 10.5-year average development duration, including all four groups with a duration of less than 10 years (Allergy, Metabolic, Infectious disease, and Ophthalmology). Conversely, the remaining disease areas, with below average LOAs, either have durations that lie very close to the 10.5-year average duration, or in the case of Urology, Cardiovascular and Neurology, notably exceed the average.

Phase transition durations by disease area

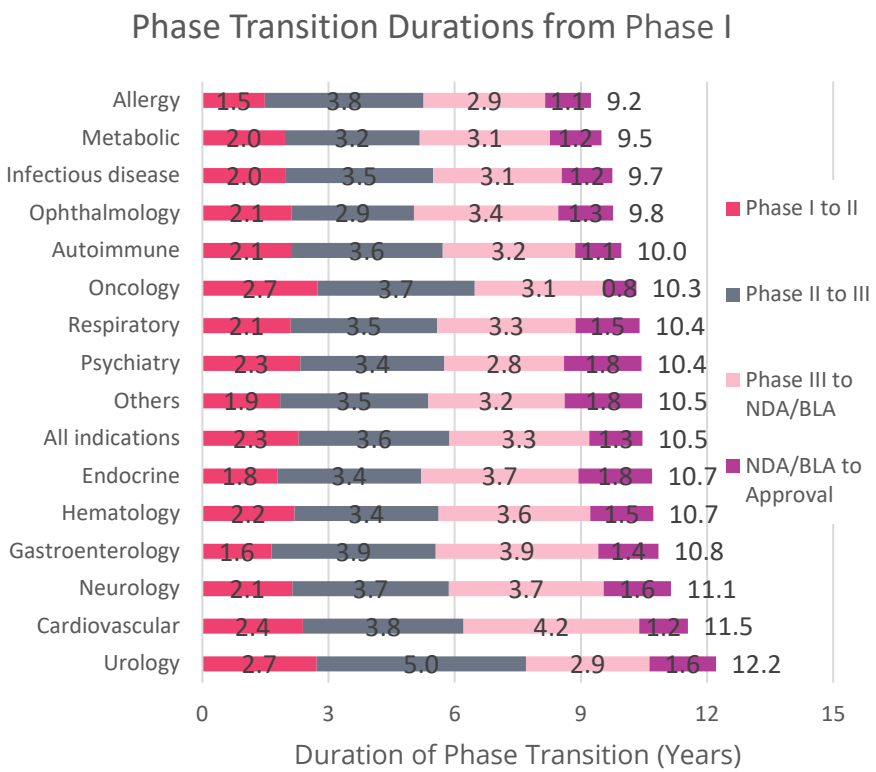


Figure 17a: Phase transition durations from Phase I by disease area. Chart of phase transition duration from Phase I for all diseases. Source: Biomedtracker® and Pharmapremia®, 2020

Phase Duration	Phase I to II		Phase II to III		Phase III to NDA/BLA		NDA/BLA to Approval	
	Advanced	Duration	Advanced	Duration	Advanced	Duration	Advanced	Duration
Allergy	31	1.5	26	3.8	22	2.9	20	1.1
Metabolic	84	2.0	67	3.2	42	3.1	42	1.2
Infectious disease	233	2.0	159	3.5	126	3.1	145	1.2
Ophthalmology	63	2.1	71	2.9	42	3.4	41	1.3
Autoimmune	228	2.1	148	3.6	143	3.2	190	1.1
Oncology	795	2.7	426	3.7	236	3.1	298	0.8
Respiratory	100	2.1	47	3.5	40	3.3	43	1.5
Psychiatry	79	2.3	44	3.4	40	2.8	52	1.8
Others	98	1.9	88	3.5	54	3.2	61	1.8
All indications	2296	2.3	1424	3.6	1115	3.3	1316	1.3
Endocrine	138	1.8	78	3.4	100	3.7	107	1.8
Hematology	64	2.2	51	3.4	63	3.6	67	1.5
Gastroenterology	21	1.6	25	3.9	20	3.9	30	1.4
Neurology	246	2.1	135	3.7	120	3.7	143	1.6
Cardiovascular	107	2.4	53	3.8	58	4.2	66	1.2
Urology	9	2.7	6	5.0	9	2.9	11	1.6

Figure 17b: Table of phase transition duration by disease area with corresponding n values. The n value is the total 'Advanced' transitions within each phase used to calculate duration. The ordering of disease areas is equivalent to the total years overall in clinical development. Source: Biomedtracker® and Pharmapremia®, 2020

Within each individual phase, there are some noteworthy performers. Oncology shares the longest Phase I transition at 2.7 years with Urology, but it is remarkably close to the Phase II and III averages with durations of 3.7 and 3.1 years, respectively. It is also the only disease area to have an average regulatory review of less than 1 year; the 0.8-year duration is almost half as short as the cumulative total for all non-Oncology indications (1.4 years).

Urology drug candidates experience the longest Phase II transition (5.0 years), although this average is calculated from a sample size of just six. Excluding Urology, the remaining disease areas all lie close to the average transition of 3.6 years. Ophthalmology emerges as the fastest disease area for Phase II research (2.9 years).

When looking at Phase III timelines, Cardiovascular drug programs have the longest duration, extending to 4.2 years. The large cardiovascular patient population sizes in trials and the long-term evaluation of cardiovascular outcomes contribute to longer timelines than seen in other highly prevalent diseases, such as Psychiatry (2.8 years), which typically assesses short-term symptomatic improvement using rating-scale questionnaires.

Discussion

Comparison with previous reports on attrition rates

In 2016, BIO published a major report on industry attrition rates using Biomedtracker's data from 2006 to 2015 for 9,985 phase transitions, revealing an overall LOA of 9.6% for an asset entering Phase I development. This report analyzed 12,728 transitions from 2011 to 2020 and revealed a lower overall LOA of 7.9%. While the underlying data and analyses are consistent, we caution against direct comparison between the two top-line LOA numbers.

Among the additional ~2,750 phase transitions in this analysis, a sizable portion can be attributed to industry expansion, with more companies (1,779 vs 1,103) and a larger pipeline (9,704 programs vs 7,455) in the dataset. Additionally, increased demand from investors for disclosures on drug pipeline progress has resulted in timelier reporting of suspended programs. Furthermore, the time period examined in this report includes more rigorous scrutiny of the Biomedtracker database. The result is a stronger account of failed transitions in the 2011–2020 analysis vs the 2006–2015 analysis. In particular, this has resulted in a notable correction in the calculation of success rates for the Phase I transition, where drugs can lie dormant or unreported for many years, lowering the success rate from 63.2% during 2006–2015 to 52.0% in 2011–2020. Transition rates at Phase II and III are closer, albeit slightly lower in this update (28.9% vs 30.7%; 57.8% vs 58.1%). Regulatory approval moved in the other direction, increasing from 85.3% to 90.6%

Comparing current success rates with prior publication

Likelihood of Approval	2016		2021	
	Phase I to Approval	N	Phase I to Approval	N
Hematology	26.1%	283	23.9%	352
Metabolic	15.3%	241	15.5%	399
Infectious disease	19.1%	916	13.2%	1170
Others	16.3%	301	13.0%	541
Ophthalmology	17.1%	267	11.9%	415
Autoimmune	11.1%	837	10.7%	1305
Allergy	14.7%	107	10.3%	201
Gastroenterology	15.1%	156	8.3%	186
All indications	9.6%	9985	7.9%	12728
Respiratory	12.8%	428	7.5%	501
Psychiatry	6.3%	451	7.3%	442
Endocrine	13.2%	791	6.6%	887
Neurology	8.4%	1304	5.9%	1411
Oncology	5.1%	3163	5.3%	4179
Cardiovascular	6.6%	632	4.8%	651
Urology	11.4%	108	3.6%	88

Figure 18: Comparing current success rates (2011–2020) with prior publication (2006–2015). Disease areas ranked by the latest LOA results from high to low. See Appendix for a breakdown of individual phase success rates

The observations and takeaways of the 2016 report remains as relevant today as they did in 2016, even accounting for the changes from these improved calculations. The leading factors that correlate with higher success rates remain consistent: clinical validation of a target (e.g., whether a drug with the same mechanism of action has been previously approved), biologic modality (e.g., mAbs), patient population and selection strategy (e.g., rare diseases with detectable gene mutations).

Oncology: largest segment of industry, modest success rates

Oncology continues to possess the largest portion of the 12,728 phase transitions over 2011–2020, even though it has the lowest Phase I LOA (5.3%). Despite our caution against direct comparisons with the 2016 paper, Oncology was one of the few major disease areas to increase its LOA, growing marginally from 5.1% in the previous analysis. In addition to the emergence of IO therapeutics, other drivers of improved success may be due to an increase in biomarker-defined patient populations, and the success of novel drug modalities (CAR-T, ADCs, etc.). That being said, the overall performance of Oncology is still considerably below the industry average (7.9%) and non-oncology indications (9.3%). There is also no substantial benefit to many Oncology indications being classified as rare diseases, as the rare Oncology subset only carries a Phase I LOA of 6.8%. These findings indicate that scientific complexity and the competitive dynamic that characterize this therapy area continue to have an adverse impact on the success rates for oncology therapies. This has not dissuaded investment within Oncology, with nearly half of all venture capital going into cancer companies by the end of the last decade.[4]

Challenges for Phase II proof-of-concept drive overall performance

The Phase II transition is clearly the main translational and limiting step from bench to bedside, as proof-of-concept is only established in 28.9% of drug programs at this stage. Phase II also sees the greatest distribution between major disease areas, with a three-fold difference spanning Urology (15.0%) and Hematology (48.1%) at opposite ends of the spectrum. Within these disease areas, it is the performance of "novel" drug classes – NMEs (New Molecular Entities), biologics, and vaccines – that have the most significant negative contribution to low overall LOA success rates, as the "off-patent" grouping relies on duplication of established science.

Hematology retains a dominant position largely on the strength of hemophilia research, which accounts for the largest number of transitions within the disease area. In the case of hemophilia A, where the underlying biology is extremely well characterized, 32 of 34 of phase transitions have been successful, including all six at Phase II development. The other main hematologic disease, anemia (which includes several different sub-indications), also has notably above-average success rates.

Conversely, Cardiovascular saw a large numbers of failed Phase II transitions for novel drug candidates, with an overall rate of just 21.0%. Compared to the other major disease areas such as Oncology and Autoimmune, there is relatively little penetration of biologics within Cardiovascular, where novel drug discovery is commonly centered on oral small molecules. Of the 507 Cardiovascular development programs in the Biomedtracker dataset, there are just 98 biologics vs 321 NMEs. This approximate 1:3 ratio is far lower than Autoimmune therapies which have a ratio of 1:1 and a LOA of 10.7%. As noted in **Figure 9**, NMEs underperform biologics considerably, including at the critical Phase II transition. For biologics, there is typically a stronger relationship between disease state and drug mechanism, enabling a stronger proof-of-concept.

Neurology is an interesting example where the Phase II transition rate (26.8%) is relatively close to the industry average, with most of the developmental risk realized later in Phase III studies. Clinical development programs for neurodegenerative diseases are often able (or designed) to find adequate justification from biomarker data to progress into late-stage trials, without having a clinical proof-of-concept on clinically validated endpoints. The resulting effect is underperformance in the Phase III and NDA/BLA transitions, dragging down the overall Phase I to LOA success rate for this disease category to 5.9%

Characteristics of rare disease R&D not found in highly prevalent chronic disease

Throughout the last decade, industry investment and drug development have pivoted towards rare, congenital diseases. Specific examples of clinical and commercial successes have encouraged this transition. Drivers of these successes include targeting molecularly defined causes of disease, regulatory incentives, and favorable reimbursement environments. However, as these drivers of success are absent from chronic, highly prevalent diseases, investment for these indications has waned over the same period, with notable

[4] BIO (2020) Emerging Therapeutic Company Investment and Deal Trends. Available here [Accessed 15 January 2021].

large company exits in psychiatric and cardiovascular medicine. BIO has covered the state of innovation in these areas in a series of industry analysis reports.[6,7,8]

One large difference between this 2011–2020 dataset and the previous 2006–2015 iteration is the intensifying focus on rare diseases. Our latest analysis includes 1,256 phase transitions within rare diseases, a considerable increase over the 521 noted in the previous study. This spans 685 different lead developers (not including those listed solely as partners). This indicates that companies view pivoting to rare disease clinical development as a sound strategy.

While notably more successful than industry averages – and in particular chronic, highly prevalent diseases – the increasing breadth of rare disease R&D is leading to a reduction in overall success. Our calculated Phase I LOA of 17.0%, slightly over double the industry average (7.9%), represents a notable fall from the 25.3% reported in 2016. As the number of programs has grown, so has the number of rare diseases that are being addressed for the first time by therapeutic modalities where less is known about their biology and pathology. In other cases, there are new difficulties in targeting the root cause. Nevertheless, rare disease R&D remains an attractive area for investment, combining a large unmet need with commercial potential, supported by regulatory incentives and above-average clinical success rates.

Complexity of modality correlates with higher LOA

As an approximate rule, the increasing chemical complexity of the therapeutic being tested is associated with higher rates of phase success, as shown in **Figure 10**. In the past ten years the proportion of complex biologics in the clinical pipeline has steadily expanded from 25% to 40%.[9] However, this trend has not been enough to boost overall industry rates beyond a 10% LOA, as simpler small molecules remain the in the majority. These account for 57% of the 12,728 phase transitions included in this analysis, with a modest 7.5% Phase I LOA.

Monoclonal antibodies, which entered routine clinical practice three decades ago, have considerably improved LOAs (12.1%) when compared to the drug classes that came before – small molecules (7.5%), peptides (8.0%), and proteins (9.4%). Since then, numerous new classes have emerged that have not yet had a comparable transformational impact, including gene therapies and antibody-drug conjugates. However, recent advances with CAR-T cell therapies and RNA interference are yielding industry-leading LOAs of 17.3% and 13.5%, respectively.

Advanced biologic therapies possess the intrinsic benefit of being precise tools for manipulating human biology and are reserved for indications where there is a clear pathophysiology. Additionally, they are also more likely to benefit indirectly by association with other factors that are highly predictive of success, including regulatory designation, biomarker-enriched patient populations, and reduced competition. On the last point, commercial maturity is one important factor that leads to pipeline attrition. Investment decisions and portfolio prioritization reflect the competitive scenario, so we should realistically expect strategic and clinical suspensions to increase over time. It will be interesting to see whether CAR-T and RNA interference approaches can sustain such remarkable LOAs as fast-followers enter the clinic and attempt to extend the technology into new, unproven treatment settings. Success rates can eventually wane over time as sample sizes increases.

Contributors of clinical success and predictive analytics

While recent breakthrough therapeutics offer new hope for patients, they have also made biomedical innovation more complex, riskier, and more expensive. In the face of multiple growing uncertainties, the

[5] Thomas D, Wessel C (2019). The State of Innovation in Highly Prevalent Chronic Diseases Volume V: Systemic Hypertension and Heart Failure. Available here [Accessed 15 January 2021].
[6] Thomas D, Wessel C (2018). The State of Innovation in Highly Prevalent Chronic Diseases Volume II: Pain and Addiction Therapeutics. Available here [Accessed 15 January 2021].
[7] Thomas D, Wessel C (2017). The State of Innovation in Highly Prevalent Chronic Diseases Volume I: Depression Therapeutics. Available here [Accessed 15 January 2021].
[8] Lloyd I (2020) Pharma RD Annual Review 2020 Whitepaper. Available here [Accessed 15 January 2021].

need for greater accuracy in predicting clinical trial outcomes has also grown. More accurate forecasts mean fewer drug failures, faster approval times, a lower cost of capital, and more funding available to bring new and better therapies to patients sooner.

When managing their portfolios of investigational drugs, biopharma companies typically use simple historical averages of regulatory approval rates, based on backward-looking relative frequencies of approval. While these historical averages contain useful insights, they do not take into account the many important predictive features related to therapeutic modality, trial design, sponsor track record, or other characteristics specific to the indication. Drug and device developers need to be able to accurately evaluate the impact of key drivers on the probability of success to efficiently allocate capital to opportunities with the highest potential. In this report, we demonstrate how machine learning techniques can be used to fully exploit all available data.

The relative importance of individual features in our predictive models find that trial outcome, regulatory designation, prior approval for another indication, the use of patient pre-selection biomarkers, drug modality, and sponsor track record are all critical features for predicting success in our set of more than 200 features. Because these classifiers are nonlinear, there is no simple interpretation of the contribution of each predictor to the forecast. However, the intuition behind some of these factors is clear. For example, drug-indication pairs with trials that achieve positive outcomes should have a greater chance of approval; candidates sponsored by companies with strong track records and diverse expertise in drug development should have a higher likelihood of success; and already approved drugs should have higher chances of approval for a second related indication.

When applied to out-of-sample data, the QLS forecasts show that candidates with higher probabilities were indeed more likely to be approved, indicating that our models are able to discriminate between high- and low-probability candidates. These promising results suggest the possibility of even more powerful predictive models with the inclusion of more refined drug development data, including patient-level clinical trial data. Ultimately, such predictive analytics have the potential to facilitate more informed data-driven decisions about the risk assessment and portfolio management of investigational drugs at all clinical stages.

Increasing duration in clinical development typically carries an elevated risk

The average timeline from Phase I to approval over 2011–2020 was 10.5 years. This lengthy timeframe is a significant risk to investors and entrepreneurs, and speaks to the current complexities of today's clinical drug development. This risk is compounded with the correlation of longer timelines in certain disease areas with greater pipeline attrition rates. For example, Urology and Cardiovascular carry both the longest clinical program durations (12.2 years and 11.5 years) and the lowest LOAs from Phase I (3.6% and 4.8%). This may partly be attributed to highly prevalent diseases requiring longer enrollment times and more complex endpoints. The average time to run an Alzheimer's Phase II trial is not going to be a short as a Phase II trial for a Covid-19 antibody. However, for some of the more complex diseases, the failure to terminate R&D programs when the data do not support further progression can add significantly to the duration recorded for a particular phase. There are countless examples of studies yielding inconclusive results, necessitating further clinical investigation and prolonging the duration within a given phase, only for the drug to subsequently fail again.

Oncology timelines were found to have unique characteristics. Although the overall oncology development time is close to the average (10.3 years vs 10.5 years), the length of each constituent phase varies considerably from the average. Its Phase I duration (2.7 years) is the longest of all major disease areas, while its NDA/BLA filing transition is the shortest (0.8 years). Oncology drug developers have broadly adopted the basket trial design at Phase I, whereby multiple different indications are studied in a single trial to evaluate potential efficacy signals, before expansion cohorts are initiated, either within the same trial or in a new Phase II study. At the US FDA, the Office of Oncologic Diseases in particular has led the use of expedited pathways and predictive endpoints, as well as Real Time Oncology Review (RTOR), facilitating the transition of new cancer drugs that address unmet clinical needs.

Methods

Drug Development Programs analyzed in this report track a specific indication for each drug. For example, Rituxan in non-Hodgkin's lymphoma (NHL) qualifies as a different development path than Rituxan in multiple sclerosis (MS). Biomedtracker assigns a unique internal identifier that can be used to isolate all development paths.

In addition to tracking the phase of development, Biomedtracker also assigns "lead" status to certain development paths. This is used to denote the most advanced indication in clinical development for a specific drug. Drugs can only have one lead development program (with some rare exceptions). For example, cancer drugs developed in multiple indications will have the most advanced program assigned as the lead, and the rest as "non-lead". However, in this report we do not differentiate the most advanced programs and analyze the data on a program/development path level, which more accurately reflects company resource utilization.

Individual Phase Transition Success Rates were calculated as the number of drugs that moved from one phase to the next phase divided by the sum of the number of drugs that progressed to the next phase and the number of drugs that were suspended. The n value associated with the phase transition success rates represents the number of drugs that have advanced plus the number of drugs that have been suspended, which we label as phase transitions. Phase transition success rates reported in this study were based on transition rates, not necessarily resulting from safety or efficacy data. Transition rates are negatively impacted by early development termination due to commercial and regulatory uncertainty as well as economic and portfolio management decisions.

Biomedtracker further classifies events by phase of development, summarized in the table below:

Biomedtracker phase	Description for purposes of the study
I	Drug is currently in Phase I
I/II, II, IIb	Drug is currently in Phase II
II/III, III	Drug is currently in Phase III
NDA/BLA	Application for approval has been submitted to the FDA and is currently under review
Approved, Withdrawn from Market, Approved (Generic Competition)	Drug has been approved for marketing in the United States
Suspended	Drug is no longer in development
Approved in Europe, Approved in other than U.S./E.U., Development, Development Outside U.S.	The company developing this drug does not plan to market it in the United States

Generic products were not included, but generic manufacturers developing novel investigational drugs were represented. Biosimilars, which require thorough clinical development, were also included.

Likelihood of Approval (LOA) denotes the probability of reaching FDA approval from the current phase and is also expressed as a percentage. LOA is calculated as the product of each phase success probability leading to FDA approval. The n value associated with LOA is the sum of the n values for each phase transition included in the LOA calculation.

For example, if a drug is currently in Phase II, and the phase success for Phase II is 50% (n=10), Phase III is 50% (n=10), and FDA approval is 50% (n=10), then the LOA for the Phase II drug would be 12.5% (50%×50%×50%=12.5%, n=30).

Data Source for Drug Program Transitions. Data used for this study were extracted from Biomedtracker

using Pharmapremia, a purpose-built Probability of Technical Success (PTS) tool, which identified all 'Advanced' and 'Suspended' drugs by development phase from January 1, 2011, to November 30, 2020. Biomedtracker and Pharmapremia, subscription-based products of Informa, track the clinical development and regulatory history of investigational drugs to assess their Likelihood of Approval (LOA) by the FDA. Biomedtracker is populated in near real-time with updated information from press releases, corporate earnings calls, investor and medical meetings, and numerous other sources. These data are recorded in Biomedtracker and tagged with a date. Biomedtracker also uses other sources, including regular communication with companies conducting clinical trials, to enhance the accuracy and timeliness of the data.

Drug Classification Methods. Biomedtracker records FDA classification (i.e., new molecular entity (NME), non-NME, biologic, or vaccine) as well as the biochemical profile (e.g., small molecule, monoclonal antibody, peptides, natural proteins, antisense, vaccine, etc.). Biologics, as defined by the FDA, can be sugars, proteins, or nucleic acids, or complex combinations of these substances, or may be living entities such as cells and tissues.

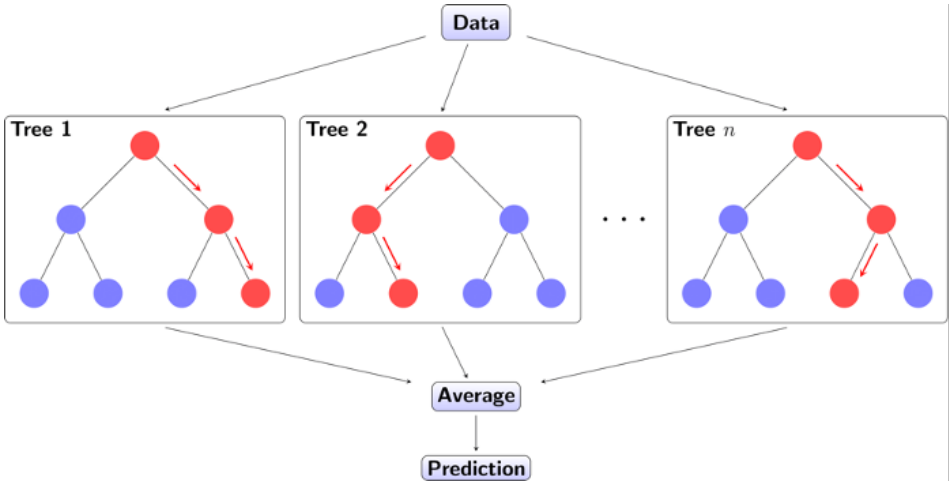
Patient Preselection Biomarkers. A subset of the Biomedtracker phase transition data was matched with corresponding entries in Trialtrove, an analyst-curated clinical trial database produced by Informa Pharma Intelligence. Where available, NCT numbers (ClinicalTrials.gov) from Biomedtracker for each drug-indication pair was matched according to the corresponding entry in Trialtrove, which contains additional information on whether patient preselection biomarkers featured in the trial design. This includes trials incorporating pharmacogenomic and/or pharmacogenetic analysis, including the use of genomic biomarkers (RNA, DNA) for patient selection or stratification. Qualifying trials were also matched by their corresponding phase of clinical development and timing, with completion dates occurring in advance of the subsequent phase transition. Phase transitions for which an associated biomarker-augmented trial could be identified were assigned accordingly, while those that either featured trials without confirmed patient preselection, or for which no information could be identified, were marked as not having patient preselection biomarkers. NDA/BLA transitions were assumed to feature patient preselection biomarker data if the Phase III transitions qualified as such.

Drug Development Timelines. In addition to denoting whether a drug was advanced or suspended at the end of a phase, Pharmapremia also calculates the time spent at each clinical or regulatory phase for all successful transitions. Suspended program timelines were not included to assess timelines. This is the duration between public disclosures of the initiation of appropriate stages, and as such does not map exactly to clinical trial timings. Each clinical phase transition may encompass more than one clinical trial, which may or may not run sequentially. Furthermore, there may be an additional delay between planning development strategy, reporting of trial results, the decision to progress or suspend, and the public disclosure of this information.

QLS Methodology. As in the case of estimating historical success rates, missing data is a significant challenge when applying machine-learning algorithms to predicting clinical trial outcomes. Gathering and filling in this missing data manually is expensive, time-consuming, susceptible to error, and, in many cases, simply not possible because the missing data no longer exist. A typical solution is to discard any clinical trials or drug-indication pathways with missing features. However, this approach eliminates a great deal of useful information, and can lead to biased inferences in the analysis. A more elegant solution involves replacing it with substituted values obtained through various imputation methods. One approach is a statistical technique called the k-nearest neighbors algorithm, in which the missing data are replaced with the average of the k "most similar" (i.e., closest) examples in the data set. For example, if one were replacing a square of missing bathroom tile, the k-nearest neighbors algorithm using a k of 8 would suggest the replacement tile should have the color average of the eight tiles surrounding the missing piece.

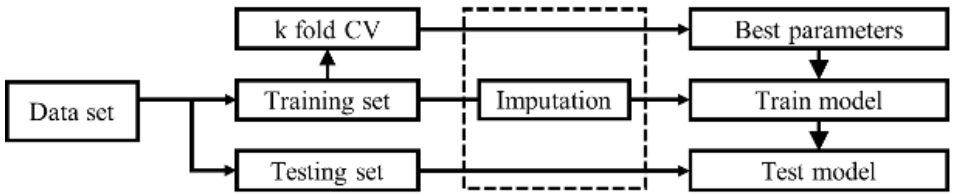
Once the missing information has been imputed, the QLS approach uses a proprietary machine-learning algorithm to analyze the clinical trial data, including its additional features. While the details of the QLS model

are proprietary, an example of a commonly used machine-learning algorithm is a random forest classifier. The motivation for a random forest is to aggregate the predictions of many different decision trees, each of which is trained over a randomized subset of variables (see figure below). While each individual tree is a weak predictor, their collective average forms a strong predictor, similar to the 'wisdom of crowds' that determines the market price of a financial asset. For example, consider a specific dataset of clinical trials in which the successful trials are more frequently associated with big pharma companies having three or more oncology drugs approved over the past five years, all of which were biologics rather than small molecules. The collection of decision trees that form a random forest might recognize this pattern and predict that a small biotech company with no previous drug approvals will have a lower probability of success than a company that more closely fit the profile described in the previous sentence.



A "forest" of decision trees. Each individual tree is a weak predictor, but similar to the wisdom of crowds, their ensemble average forms a strong predictor.

By identifying subtle patterns in a large enough database of historical drug transitions, machine-learning algorithms can often make remarkably accurate predictions. An illustration of the components of the QLS approach is provided below.[9]



Modeling methodology adopted by QLS. Abbreviations: CV=cross-validation.

[9] We use Citeline data provided by Informa Pharma Intelligence, a superset of the most commonly used data sources in the industry. It combines individual clinical trial information from Trialtrove and drug approval data from Pharmaprojects and Biomedtracker, in addition to incorporating multiple data streams, including nightly feeds from official sources such as ClinicalTrials.gov. We restrict our analysis to drugs being developed for the U.S. market and seeking U.S. FDA approval using Pharmaprojects' manually curated Drug Program Landscape data field.

Appendix

Comparison of phase success for the 2016 vs. 2021 publications, ranked by 2021 phase success

Phase Success	2016		2021	
	Phase I to Phase II	N	Phase I to Phase II	N
Ophthalmology	84.8%	66	71.6%	88
Hematology	73.3%	86	69.6%	92
Others	66.7%	96	63.6%	154
Metabolic	61.1%	95	61.8%	136
Infectious disease	69.5%	347	57.8%	403
Allergy	67.6%	37	56.4%	55
Respiratory	65.3%	150	55.9%	179
Autoimmune	65.7%	297	55.2%	413
Psychiatry	53.9%	154	52.7%	150
All indications	**63.2%**	**3582**	**52.0%**	**4414**
Cardiovascular	58.9%	209	50.0%	214
Oncology	62.8%	1222	48.8%	1628
Neurology	59.1%	462	47.7%	516
Gastroenterology	75.6%	41	46.7%	45
Endocrine	58.9%	299	43.3%	319
Urology	57.1%	21	40.9%	22

Phase Success	2016		2021	
	Phase III to NDA/BLA	N	Phase III to NDA/BLA	N
Hematology	75.0%	64	76.8%	82
Urology	71.4%	21	69.2%	13
Endocrine	65.0%	143	66.2%	151
Autoimmune	62.2%	135	65.3%	219
Allergy	71.4%	14	64.7%	34
Respiratory	71.1%	45	64.5%	62
Infectious disease	72.7%	150	64.0%	197
Metabolic	71.4%	35	63.6%	66
Others	69.6%	46	60.0%	90
All indications	**58.1%**	**1491**	**57.8%**	**1928**
Gastroenterology	60.6%	33	57.1%	35
Psychiatry	55.7%	70	56.3%	71
Cardiovascular	55.5%	110	55.2%	105
Neurology	57.4%	216	53.1%	226
Ophthalmology	58.3%	60	51.2%	82
Oncology	40.1%	349	47.7%	495

Phase Success	2016		2021	
	Phase II to Phase III	N	Phase II to Phase III	N
Hematology	56.6%	83	48.1%	106
Metabolic	45.2%	84	45.0%	149
Others	39.7%	116	38.6%	228
Infectious disease	42.7%	286	38.4%	414
Ophthalmology	44.6%	101	35.5%	200
Gastroenterology	35.7%	56	34.2%	73
Autoimmune	31.7%	319	31.4%	471
All indications	**30.7%**	**3862**	**28.9%**	**4933**
Allergy	32.5%	40	28.3%	92
Psychiatry	23.7%	169	26.8%	164
Neurology	29.7%	465	26.8%	504
Endocrine	40.1%	242	26.6%	293
Oncology	24.6%	1416	24.6%	1732
Respiratory	29.1%	196	21.9%	215
Cardiovascular	24.1%	237	21.0%	252
Urology	32.7%	52	15.0%	40

Phase Success	2016		2021	
	NDA/BLA to Approval	N	NDA/BLA to Approval	N
Allergy	93.8%	16	100.0%	20
Respiratory	94.6%	37	95.6%	45
Autoimmune	86.0%	86	94.1%	202
Hematology	84.0%	50	93.1%	72
Infectious disease	88.7%	133	92.9%	156
Oncology	82.4%	176	92.0%	324
Psychiatry	87.9%	58	91.2%	57
Ophthalmology	77.5%	40	91.1%	45
Gastroenterology	92.3%	26	90.9%	33
All indications	**85.3%**	**1050**	**90.6%**	**1453**
Others	88.4%	43	88.4%	69
Metabolic	77.8%	27	87.5%	48
Neurology	83.2%	161	86.7%	165
Endocrine	86.0%	107	86.3%	124
Urology	85.7%	14	84.6%	13
Cardiovascular	84.2%	76	82.5%	80

Contributing Authors

Biotechnology Innovation Organization (BIO): David Thomas

Informa Pharma Intelligence: Daniel Chancellor, Amanda Micklus, Sara LaFever, Michael Hay

QLS: Shomesh Chaudhuri, Robert Bowden, Andrew W. Lo