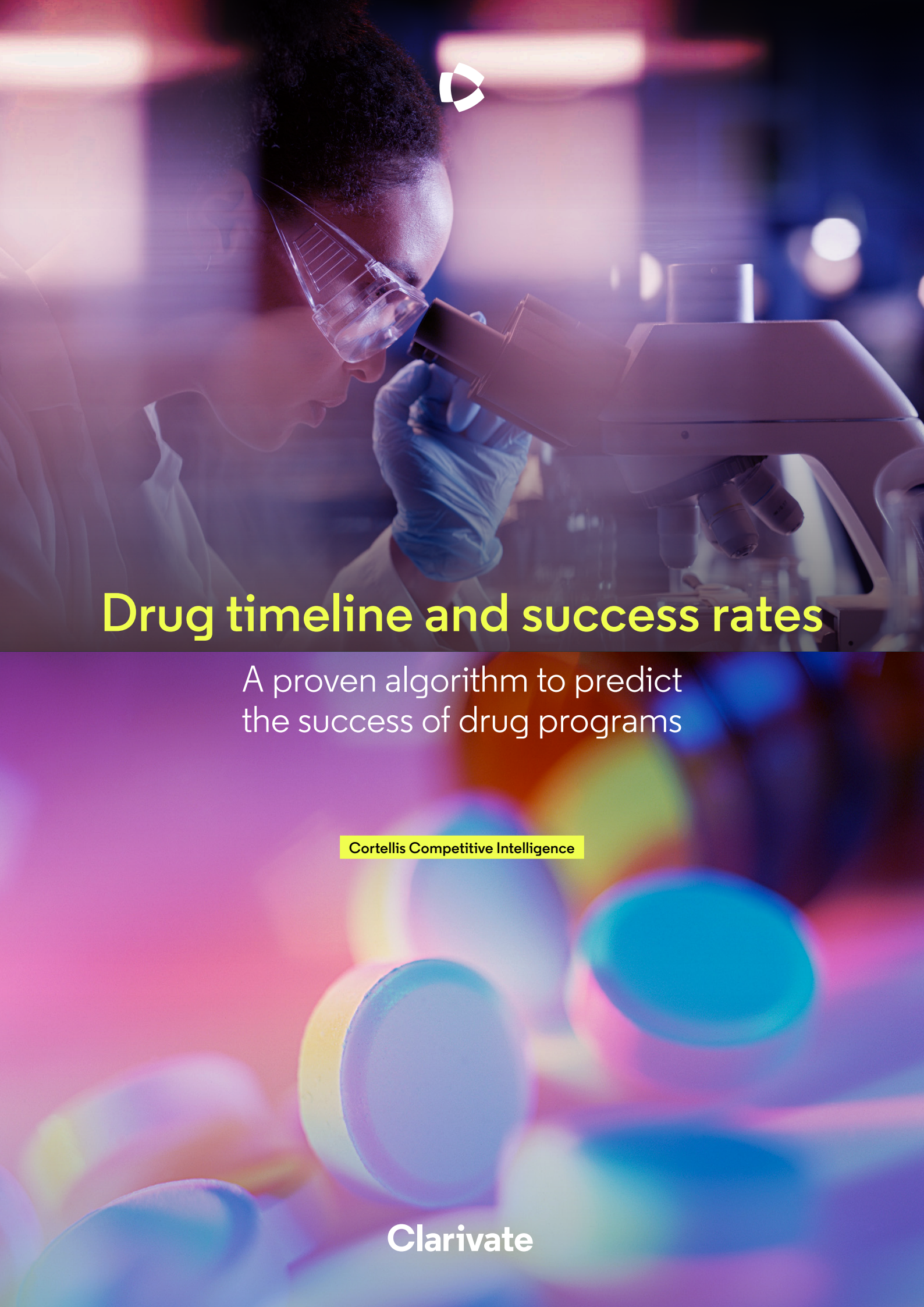


Drug timeline and success rates

A proven algorithm to predict the success of drug programs

Cortellis Competitive Intelligence

Clarivate

Contents

04 A path to better predictions

04 Our data

05 Developing a statistical model to predict clinical success

05 Turning to modeling and machine learning

06 Validation of the algorithm

06 Comparison of the algorithm's performance with benchmarking

08 Generating probabilities of success and predicted timelines

09 Key Drug Timeline & Success Rates Highlights

12 Looking to the future

12 Data available in the Cortellis suite

A path to better predictions

With 90% of clinical trial assets failing to reach the market and R&D costs continuing to soar (Figure 1), it remains critically important for pharmaceutical companies to understand the value of their portfolios. High failure rates significantly inflate the cost of successful therapeutics, as a substantial portion of R&D expenditures are consumed by programs that ultimately do not reach the market.

Figure 1: R&D expenditure and failure rates

+6.7%

Average R&D expenditure growth rate per annum (2018 to 2022)

>31%

R&D spend occurs in Phase 3 (2022)

<1/10

drugs make it to launch

Source: 2024 CMR International Pharmaceutical R&D Factbook

Many pharmaceutical companies use in-house data management techniques to guide pipeline decisions, but there are limitations to this approach. Competitive intelligence and business development teams may use databases or spreadsheet-based reporting processes built from sources such as analyst reports, news outlets, press releases and other data streams. Smaller companies without such resources may use consultants or outside analysts to provide forecasts. However, these methods are largely manual, time-intensive to curate and can be expensive to maintain; as a result, the data are often incomplete or quickly become out of date.

While the use of industry benchmarking, learning reviews, scientific literature and qualitative assessments can be helpful for a subset of assets or a specific therapeutic area, it can be difficult to replicate at scale and translate to the portfolio level.

Our data

At Clarivate, our vision is to improve the way the world creates, protects and advances innovation. Information from across our suite of products can be used separately or integrated into solutions such as Drug Timeline & Success Rates that provide a more comprehensive picture of the industry landscape:

Cortellis Competitive Intelligence provides access to data such as drug pipeline, deals, patents, global conferences and company content, along with the latest industry news and press releases.
Cortellis Deals Intelligence combines the industry's largest source of deals intelligence with enhanced visualizations of the highest-quality data, to quickly find the optimal deal without compromising due diligence.
Cortellis Clinical Trials Intelligence is a comprehensive resource of key competitive intelligence to assist with clinical trial planning, including site selection, protocol design and biomarker identification.

Developing a statistical model to predict clinical success

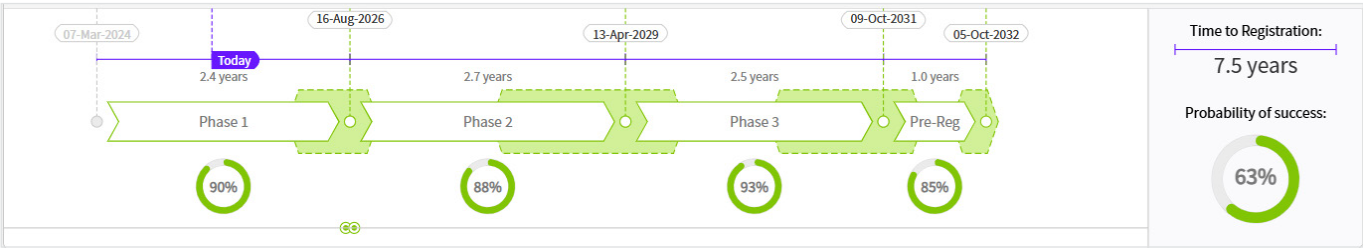
Turning to modeling and machine learning

Knowing the limitations of benchmarking, we applied data science techniques to large sets of drug development data to build an advanced analytic that would more accurately predict the clinical success and timelines of drug projects. Drug and clinical trial records spanning more than 20 years of drug development across all therapeutic areas and from all phases of clinical development through to registration are available through Cortellis Competitive Intelligence, Cortellis Deals Intelligence and Cortellis Clinical Trials Intelligence. Because the data in the Cortellis products are curated with machine and human intelligence, using a combination of automation and manual curation to populate drug and clinical trial records, our data set includes more data points than what is publicly available.

We used a predictive analytics approach to develop an algorithm that would forecast timelines and success rates based on all factors contained within the Cortellis drug and clinical trial records over a period of 24+ years (2000-2024). At that time, this included inputs from 570,000 reports on clinical trials completed or in progress; 98,000 drug records identified in pharmaceutical pipelines; 310,000 profiles of companies, institutes and research organizations developing drugs globally; 11 million patents; and 140,000 reports on life sciences licensing deals and mergers & acquisitions.

Our predictive analytics approach, patented in the US, China, and Japan, incorporates multivariant analysis, statistical modeling, and machine learning. We identified 15 areas of qualitative traits that can be used to describe a drug development program, its developer and the characteristics of the drug itself. Machine learning is then applied to the data to identify and weigh each trait for its ability to be predictive of success. This is used to build a statistical model that can identify paths and time to approval for individual drug projects (Figure 2): the algorithm in Drug Timeline & Success Rates.

Figure 2: Example of a predicted timeline and success rate



Source: Drug Timeline & Success Rates.

Validation of the algorithm

To assess the prediction accuracy, we set aside 2022-2024 phase transition data as the test set, after building our predictive analytic on the 22 years of phase transition data from 2000-2021. We determined the ability of our model to predict the success or failure of drug projects through phase transitions and found that:

91%

of the drugs for which the model predicted a 90% or greater probability of successful phase transition did indeed transition.

79%

of all Phase 2 drugs for which the model predicted a 70% or greater probability of entering Phase 3 successfully transitioned.

79%

of all Phase 1 drugs for which the model predicted an 80% or greater probability of entering Phase 2 successfully transitioned.

85%

Of the drugs with a 70% or greater predicted chance of failure did indeed fail.

Comparison of the algorithm's performance with benchmarking

As benchmarking is the current standard for assessing portfolio performance, we compared the results of our algorithm against industry benchmarks. The Centre for Medicines Research (CMR) International, the leading provider of pharmaceutical R&D performance metrics (and a wholly owned subsidiary of Clarivate), reports that the Phase 1 to Phase 2 transition successfully occurred for 57% of drug projects during the 2002-2020 period. The success rate dropped to 28% for Phase 2 to Phase 3. The Phase 3 to pre-registration transition was 69%, and 94% for pre-registration to registration (see Figure 3).

Figure 3: Benchmarking data.

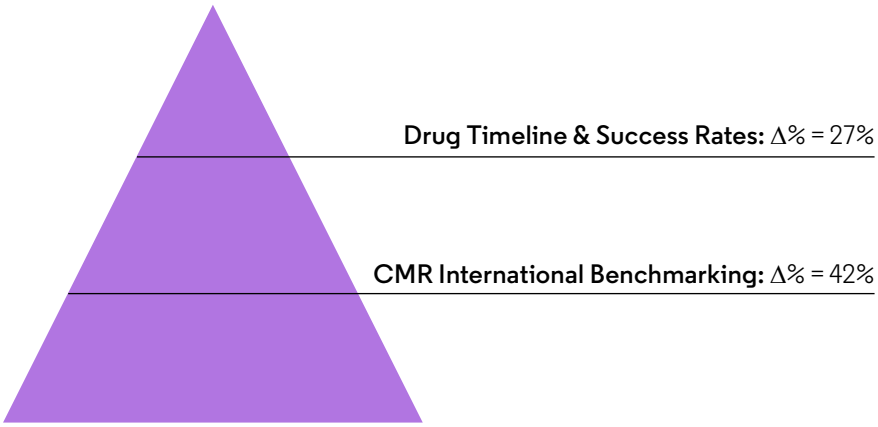
Phase transitions	Success rates (2002-2020)
Phase 1 --> Phase 2	57%
Phase 2 --> Phase 3	28%
Phase 3 --> Pre-registration	69%
Pre-registration --> Registration	94%

Source: 2023 CMR International Pharmaceutical R&D Factbook

Comparing the performance of our algorithm to CMR benchmarks, we calculated the difference between prediction success rates and actual successes, an accuracy measure known as Δ%. For this measure, the smaller the number, the more accurate the methodology. When tested across all therapy areas, the Δ% for the Drug Timeline & Success Rates algorithm was 27%, and for CMR benchmarks, it was 42% (see Figure 4). Calculating the difference between the two methodologies, we determined our analytic has improved over time from 25% to over 35% more accurate than industry benchmarking.

Our analytic has improved over time from 25% to over 35% more accurate than industry benchmarking.

Figure 4: The Drug Timeline & Success Rates algorithm is considerably more accurate than benchmarking.



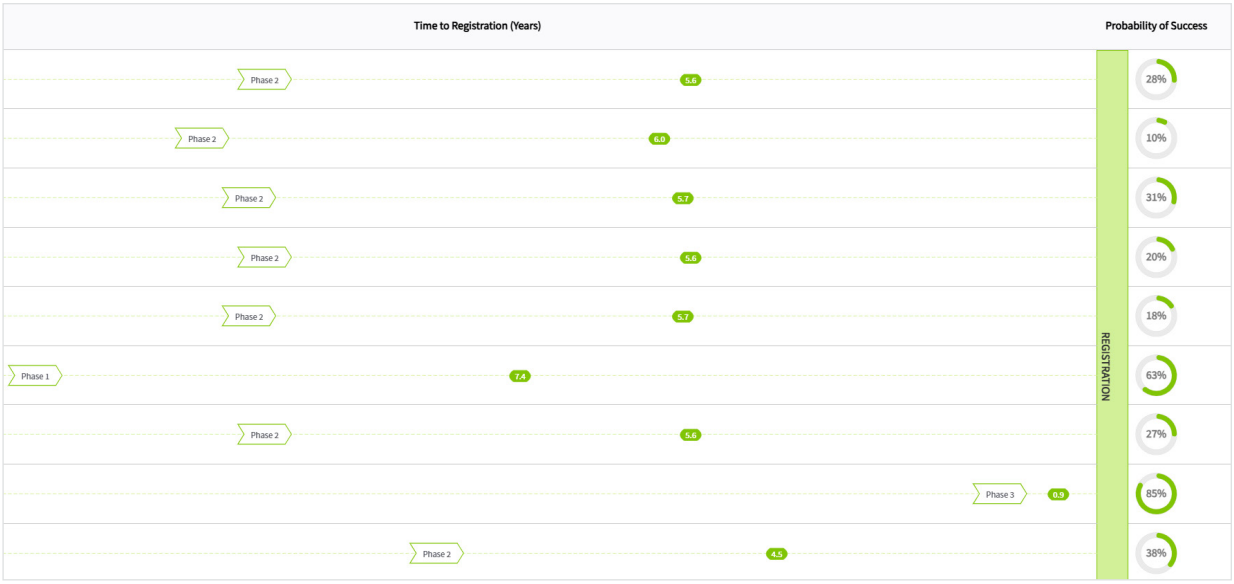
In addition, we compared Drug Timeline & Success Rates predictions with 50 novel drug approvals for 2024 by the FDA[1]. We found that, among 49 drugs with Drug Timeline & Success Rates predictions by end of year 2023, 31 of these drugs had over 90% approval probability. Furthermore, we accurately predicted the approval date within six months of the actual approval for 33 of these drugs, and within three months for 23 of these drugs.

[1] Novel Drug Approvals for 2024 [online], Available at https://www.fda.gov/drugs/novel-drug-approvals-fda/novel-drug-approvals-2024 (Accessed on March 2, 2025).

Generating probabilities of success and predicted timelines

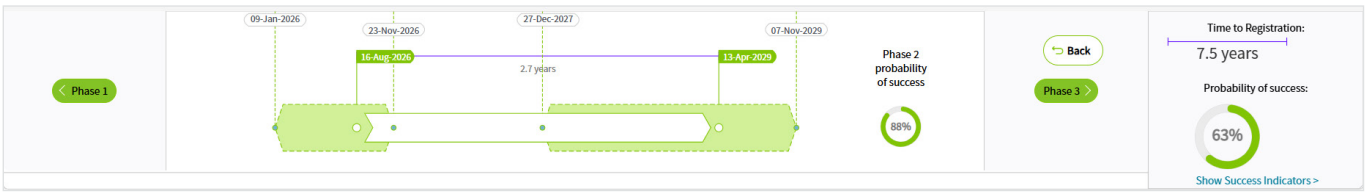
Using Drug Timeline & Success Rates, probabilities of success and predicted timelines to registration can be generated for individual drug projects undergoing clinical trials in the US, the EU, Japan, Mainland China, Singapore and South Korea (Figure 5) — across the development cycle as well as by phase (Figure 6). The timelines and success rates probabilities are based on data-driven evidence and statistical modeling, are more accurate than benchmarking and can be applied at the therapeutic area or whole portfolio level. A better understanding of timelines and forecasts enables the appropriate due diligence prior to making portfolio or investment decisions.

Figure 5: Comparison of drug timelines and success rates by drug(s) and location.



Source: Drug Timeline & Success Rates.

Figure 6: Prediction of timelines and success rates by study phase.



Source: Drug Timeline & Success Rates.

Critically, the output of the Drug Timeline & Success Rates algorithm is real-time probabilities for a given drug project, as it generates predictions based on today's information. The success indicators used in the model do not fall into a pattern because these qualitative traits and milestones are unique to each drug project. Therefore, the forecast is specific for that project and program, rather than relying on a fixed template across all drug programs.

Key drug timeline & success rates highlights

Real-time probabilities:

The model provides real-time probabilities of success based on the latest information.

Specific predictions:

Predictions are tailored to a given drug project and program as the algorithm considers unique qualitative traits and milestones.

Region-specific outputs:

Predictions can be generated for drug projects undergoing clinical trials in the U.S., the E.U., Japan, Mainland China, Singapore and South Korea.

Forecasts are dynamic and change as new milestones are achieved. Furthermore, the algorithm continues to learn from new data inputs and drug projects as the pharmaceutical landscape changes. As a result, the results are continuously refined for greater accuracy. Because of the complexity and number of possible factors influencing each drug project, the system was configured to output the 100+ predictive elements into a maximum of 15 success indicator categories, and the algorithm considers different data elements within these categories depending on the specific project and phase (Table 1).

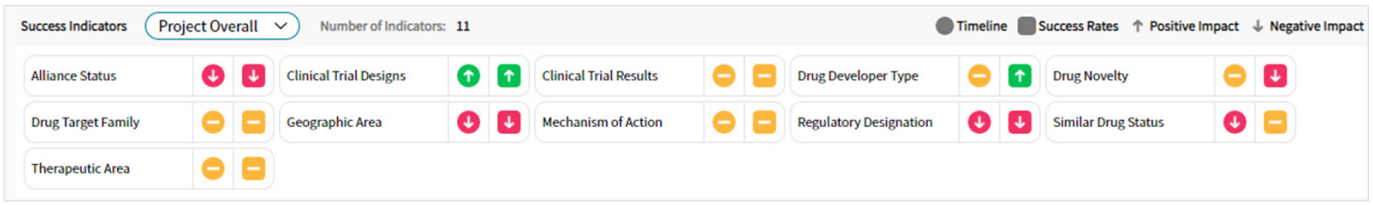
Table 1: Groupings of success indicators in drug timeline and success rates, with a small sample of the traits/data elements used in each category.

Type of indicator	Indicator group	Sample of example data elements
Cortellis data elements/traits	Geographic area	Country and region(s) in concurrent development
	Mechanism of action	Any mechanism of action associated with the drug
	Drug target family	Any protein or gene target families associated with the drug
	Therapeutic area	Any indication 'ancestors' associated with the drug project
	Drug developer type	Organization type (public/private) and size (large/medium/small) Patents count
	Regulatory designation	Any special regulatory designation
	Drug novelty	Whether drug is novel Number of active ingredients

Type of indicator	Indicator group	Sample of example data elements
	Clinical trial design	Biomarkers Patient count
	Clinical trial results	Clinical trial met/unmet primary end point Estimated trial completion date
	Alliance status	Being developed by originator
	Similar drug status	Discontinuation reason
	Type of drug compound	Biological therapeutic
Milestone events	Drug developer projections	Expected approval
	Regulatory milestone(s)	Prescription Drug User Fee Act (PDUFA)/ Biosimilar User Fee Act (BsUFA) date (new drug application [NDA]/ biologics license application [BLA])
	Clinical milestone(s)	Trial completion/final data reported

To provide transparency around the algorithm and its outputs, Cortellis Analytics – Drug Timeline & Success Rates provides several outputs. The first illustrates the impact of each of the success indicators (i.e., clinical results, mechanism of action) on the prediction (Figure 7), which is available for the project overall as well as by trial phase. The positive, negative or neutral impacts on timelines versus success rates are shown separately, with intuitive color-coding for rapid understanding.

Figure 7: Contribution of the success indicators to the prediction



Source: Drug Timeline & Success Rates.

In the Change Tracking feature, side-by-side comparisons of previous and current predictions of success rates and timelines are provided to monitor the changes in predictions over time, as industry trends evolve. Changes in drug development are easily tracked with customizable alerts on changes in development timelines, new milestone events, actual phase transition, new forecasted drug project, success probabilities, new lead development phase and actual project approval.

Benchmarking data available through Cortellis can also be used to compare historical performance of drugs within a therapeutic area, for example, with the algorithm's predictions (Figure 8) or to compare the performance of an ongoing project with that of drugs within the same therapeutic area, franchise, region, drug novelty or drug type (Table 2). In addition, you can track the progress of a drug's development to evaluate its potential as a partner or competitive threat.

Figure 8: Comparison against benchmarking data in Cortellis, with an example detail panel for one of the drug projects for the given condition.

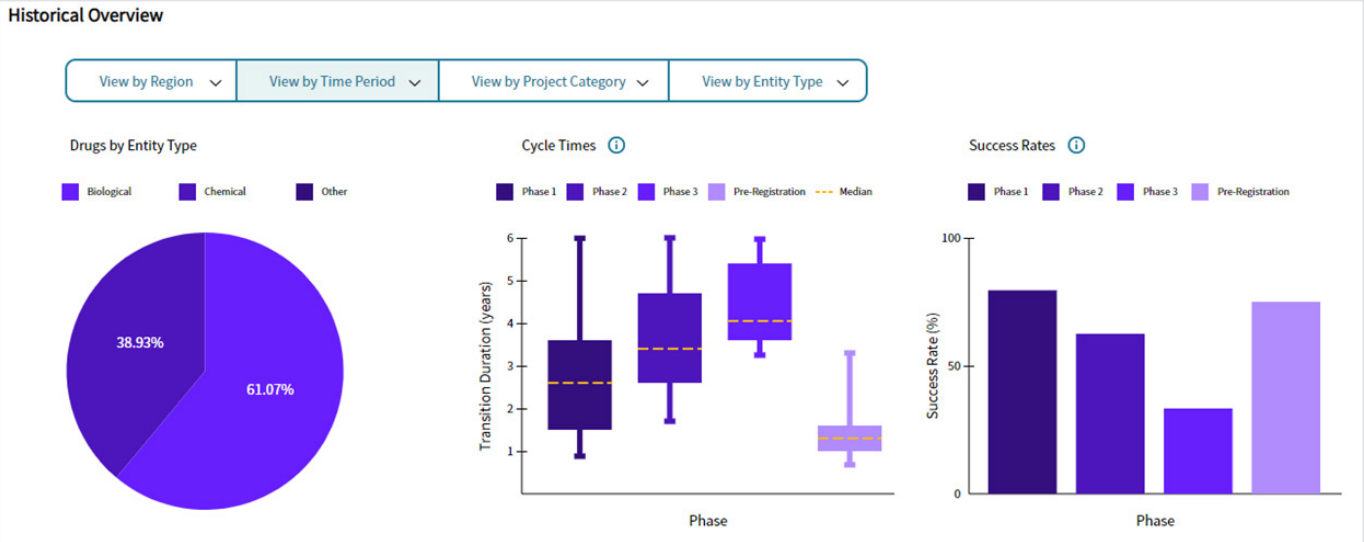


Table 2: Historical overview of success rates for drugs for cystic fibrosis, based on data in Cortellis.

Phase	Success rate	Successful drugs	Total	Unknown
Phase 1	92.68%	38	41	39
Phase 2	46.43%	13	28	74
Phase 3	76.92%	10	13	12
Pre-registration	83.33%	10	12	2

For additional analysis, Drug Timeline & Success Rates outputs can be downloaded as an Excel file for further manipulation or importing into other software systems.

Looking to the future

Continual robust testing, fresh infusions of drug development data and further refinements of the model will translate into ever more confidence in our analytic. Harnessing the power of large-scale clinical data helps deliver greater predictive accuracy to drug developers, sharpen strategic portfolio decisions and ultimately increase the efficiency of R&D programs to provide greater benefit to patients and developers.

With the seamless integration of Drug Timeline & Success Rates into the broader Cortellis suite, you can generate forecasts and gain a deeper understanding of the market based on gold-standard drug development, clinical, regulatory and deals data.

Data available in the Cortellis suite

569K+
Global clinical trials

3K+
Diseases and therapeutic areas

45K+
Trials related to rare diseases

277K+
Clinical sites across 200+ countries

2.2M+
Broker research reports

6M+
Global patents

98K+
Pipeline programs

32K+
Assets in discovery/ preclinical phase

283K+
Company profiles

1.4K+
Drug-level SWOT analyses

135K+
Life science deals

38K+
Contracts, including unredacted versions

BXD1634633160 / 05

About Clarivate

Clarivate is a leading global provider of transformative intelligence. We offer enriched data, insights & analytics, workflow solutions and expert services in the areas of Academia & Government, Intellectual Property and Life Sciences & Healthcare. For more information, please visit **clarivate.com**.

Discover solutions at:

clarivate.com