Faster and cheaper clinical trials

The benefit of synthetic data



accenture | Phesi
The Clinical Data Science Company

Accenture Life Sciences
Patient Inspired. Outcomes Driven.

With pharmaceutical innovation in treatments reaching new heights, the traditional methods for conducting clinical trials are increasingly a barrier to cost-efficient and timely drug development, with their drawn-out timelines and billions in investment. The exponential growth in patient data, combined with advances in AI, analytics and computing power, is enabling broader adoption of in-silico clinical trial methods. This reduces the need for patient recruitment and mitigates patient burden and ethical concerns over placebo treatments. Although pharmaceutical companies are taking steps in this direction, proactive efforts to stay ahead of the game are essential. It requires a global vision, more strategic use of internal and external data, integration and automation of smart algorithms, and an operating model that effectively integrates synthetic data into clinical trial design.

Patient data and analytics capabilities growing yet clinical trials still slow and costly

In the past few years, the volume of patient data created through clinical trials and electronic medical records (EMR) has soared, creating huge data lakes.[1] The ability to store, process and analyze this data has also expanded, with technological advances in AI and analytics, cloud utilization and computing power enabling more ambitious use of patient data. These advances are occurring at a time when pharmaceutical companies are still using highly time-intensive and costly means to obtain data in the clinical trial phase of drug development. A key driver of the cost and duration of clinical development is the number of participants needed for each trial. This is particularly an issue for indications with high unmet medical need, such as those within oncology, or in rare diseases, where the limited number of patients make it even more difficult to recruit patients without delay.

A role for synthetic data

What if the control arm for clinical trials currently conducted using standard of care or placebo treatments were replaced by a virtual cohort based on historical data?

Instead of patients dealing with the burden of a clinical trial but receiving just standard of care, today a synthetic control arm can be generated from historical data. Patients who have received standard of care in the past have EMR that track their outcomes. Data on trial protocols and results from completed clinical trials, observational studies and data from registries can all be incorporated into a robust synthetic control arm or used to reduce the size of the comparator arm.

Using synthetic data for the control arm substantially reduces the demand for patient recruitment, saving time and resources. The number of patients needed for a clinical trial is a key factor driving the cost and duration of clinical development. The use of a synthetic control arm instead of a patient cohort receiving standard of care reduces the patient burden of participation in clinical trials.

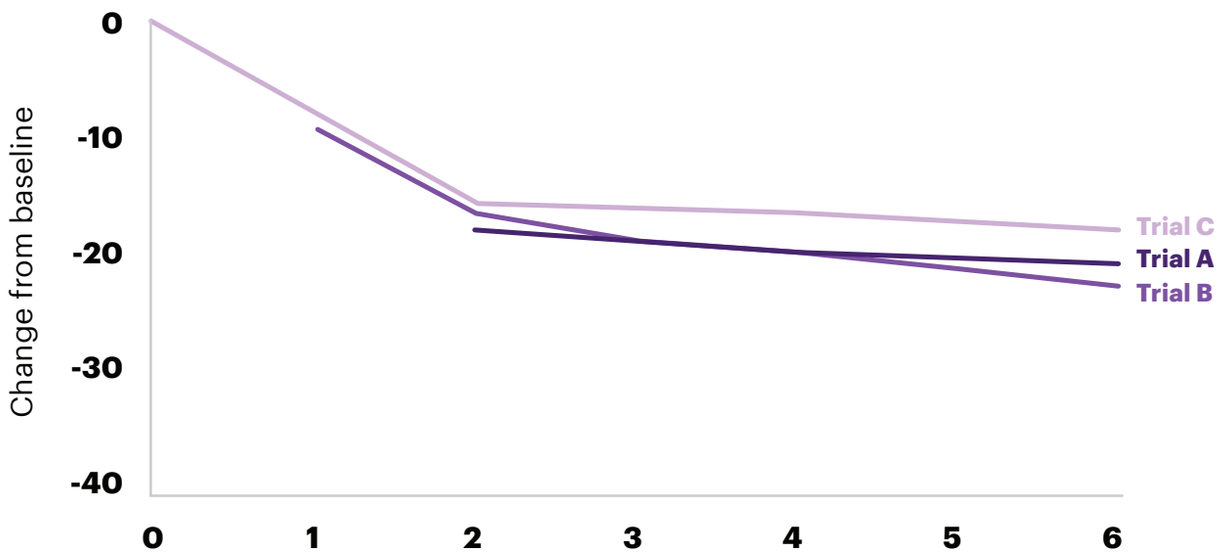
Even more, it mitigates the long-standing ethical concern that in most clinical trials, only a portion of trial participants actually get to receive the treatment being tested.

Synthetic control arms are particularly relevant for clinical trials where control group performance has been historically well-characterized, and where results have been consistent from trial to trial. The example of osteoarthritis pain in Figure 1 highlights how existing data can consistently predict the outcome for a placebo-treated patient population. There are other examples demonstrating concordance of clinical trial data with existing real-world data, for example in breast cancer.[2]

Why should a sponsor running a new study run yet another control arm trial when that data already exists? Wherever we can define patient populations across clinical trials with sufficient similarity, there is the opportunity to establish a synthetic control arm for that patient population. Leveraging the data from similar trials avoids exposing patients to unnecessary burden and risk.

In addition to minimizing the need for placebo patient enrollment, synthetic patient data can be used to define the boundaries of a trial, and to model and predict what types of patients should be included and excluded.

Figure 1: Three individual studies showing similar results in placebo control arm



Source: Phesi analysis

Study in Osteoarthritis pain of knee. Trial A, B and C included 78, 52 and 76 placebo patients respectively. Change from baseline was measured by Least squares (LS) mean change in the Western Ontario and McMaster Universities Index (WOMAC) pain score in 100 mm visual analogue scale (VAS).

Case study

Ulcerative colitis: With sufficient data and an integrated analytical platform, “what if” scenarios can be leveraged to define target patient populations

A synthetic control arm project always starts with defining the inclusion and exclusion criteria of the patient population to be included in the clinical trial. Tools and techniques that allow analysis of large data sets for different “what-if” scenarios are important, to answer questions about how the baseline looks if we include or exclude certain criteria.

The example of ulcerative colitis (UC) illustrates the differences in baseline characteristics between a specific sub-group with distal colitis versus the overall UC patient population. Table 1 shows the patient profile in Phase 2 clinical trials for UC, based on 169 trial arms from 63 trials and a total of 8,974 UC patients. The amount of patient data used to complete the analysis for each parameter is also shown.

Key differences between the targeted subgroup population and UC overall include:

- Younger age: 42.2 vs. 48.9 years
- Shorter disease duration: 5.2 vs. 6.9 years
- Less medicated: 59% vs. 75% on 5-aminosalicylate
- 12% of targeted subgroup population are drug naive
- 16% of targeted subgroup population with first UC

In-depth, integrated analysis of industry related protocols and amendments and a pharma company’s own clinical trial data, supported the creation of the targeted ‘synthetic patient’ profile. This enabled the protocol optimization and study feasibility assessment for the company’s priority UC asset, positively impacting operational success.

Table 1: Analysis of UC patients versus targeted sub-population with distal colitis.

		UC Overall					Targeted sub-population with distal colitis				
		"Synthesized" baseline patient characteristics			Data from real trials for "synthesis"		"Synthesized" baseline patient			Data from real trials for "synthesis"	
Category	**Variable**	**Mean**	**SD**	**Percentage of patients (%)**	**Derived from patients data**	**Derived from # of trial arms**	**Mean**	**SD**	**Percentage of patients (%)**	**Derived from patients data**	**Derived from # of trial arms**
	Age (years), mean (SD)	48.9	13.5		7,093	126	42.2	12.1		2,391	28
	Male			56%	7,418	139			50%	2,553	33
	Female			46%	4,432	91			51%	1,994	19
	Weight, kg	73.0	16.1		3,149	59	68.0	13.5		1,800	22
	BMI	25.2	4.8		1,863	40	26.1	5.3		546	4
	Duration of Ulcerative colitis (years)	6.9	6.7		4,770	79	5.2	5.7		1,060	17
	Smoking, current			9%	3,321	66			8%	1,441	12
	Smoking, ever			22%	2,073	30			21%	1,111	6
	Smoking, never			70%	2,094	32			70%	1,111	6
Disease extent	Proctitis			39%	664	14			39%	739	16
Disease extent				26%	995	22			61%	1,499	16
Clinical course	First attack								16%	772	13
Clinical course	Relapse								77%	1,602	20
Concomitant medications				75%	3,610	62			59%	772	13
Concomitant medications	Mesalamine			69%	1,522	36			57%	935	8
Concomitant medications	naive								12%	442	11

Source: Phesi analysis

Overcoming obstacles to synthetic data use: variety, volume and velocity

While there are compelling reasons to do more with synthetic data from both a practical and ethical perspective, the reality is that pharma companies face several implementation hurdles. Companies need to exploit the dynamics of data along three dimensions: Variety, Volume and Velocity.

A variety of data types is essential to enable robust data usage and analysis. Pharma companies typically do not have sufficient external data at hand and have been reluctant to look beyond their own clinical data to other sources, such as industry data from other clinical trials and real-world data from EMRs. Each of these data types has its own merit, and they should all be integrated to empower in-silico clinical trials and improve data confidence.

Along the same lines, the volume of data is important for robust analysis and statistical confidence. Pharma companies have their own separate data, but this is just a fraction of the potential data available for use at an indication level, and are usually not sufficient for rigorous statistical analysis. While there has been little appetite for sharing data with others, the opportunity to access a larger volume of data is strong justification for organizations to collaborate and share data.

Project Data Sphere, a consortium of pharma companies, shares patient-level clinical data from member companies.[3] Innovative companies such as Phesi have access to more than 30 million patient records from 350,000 trial protocols covering a full range of diseases and geographies. Through their recently announced collaboration with Sensyne Health, Phesi has broadened their access to a data pool of more than seven million NHS patient data records.[4]

Finally, the velocity of data matters, in terms of the speed with which a company can integrate new incoming data. Companies are often utilizing historical and distinct data sets without incorporating new data as it is generated, and have limited capacity to interpret incoming data and relate it to existing analysis efficiently. Data is often not being parsed and structured in a way that makes it usable, to allow for detecting signals through all the noise.

Each of these three components relies on access to data, which is not a given. There is no global system facilitating the exchange of patient data, and no global network of anonymized health records.

Enablers of data use: technology, skills, and leadership

Once the availability of data is addressed, using this data effectively will require an investment into technology, skills and leadership. Pharma companies need to do more to identify and adopt advanced technologies to work with the data needed for in-silico clinical trials. The algorithms exist, but they are not yet fully deployed at pharma companies in a way that enables an integrated analysis and application of data. Without this capability, companies are not able to extract meaningful insights from the data and build relationships between data sets, such as the links between patient data, investigator site performance and protocol design. Connecting different data sources is often still a manual process, limiting the potential scale of application.

Technical skill sets are another challenge for pharma, as companies have not prioritized recruitment of the best talent and tech-savvy data scientists compared to other industries such as financial services and retail. Without the technical skills, companies are unable to effectively apply digital solutions like predictive modeling and have a limited understanding of big data processing techniques such as artificial intelligence and machine learning. The global pandemic has highlighted this weakness as pharma companies struggled to predict and mitigate impacts on their clinical trial portfolios.

The other key enabler is ambition and visible leadership interested in proactively ramping up innovation in data collection, processing and practical application in clinical trial design and operations. Synthetic arms are not a new concept and have already been used sporadically over the years. But without the commitment to push forward and adopt new technologies, skills and services, and revamp existing processes where needed, the potential benefits from data use for in-silico clinical trials will remain limited. Strong leadership both within the life sciences industry and in key regulatory agencies is needed to drive transformation and build the infrastructure for utilizing historical data (both internal and external) pragmatically and systematically. R&D leaders will have to be bold in defining their pathways and their next steps.

Where to start? Four things to do now

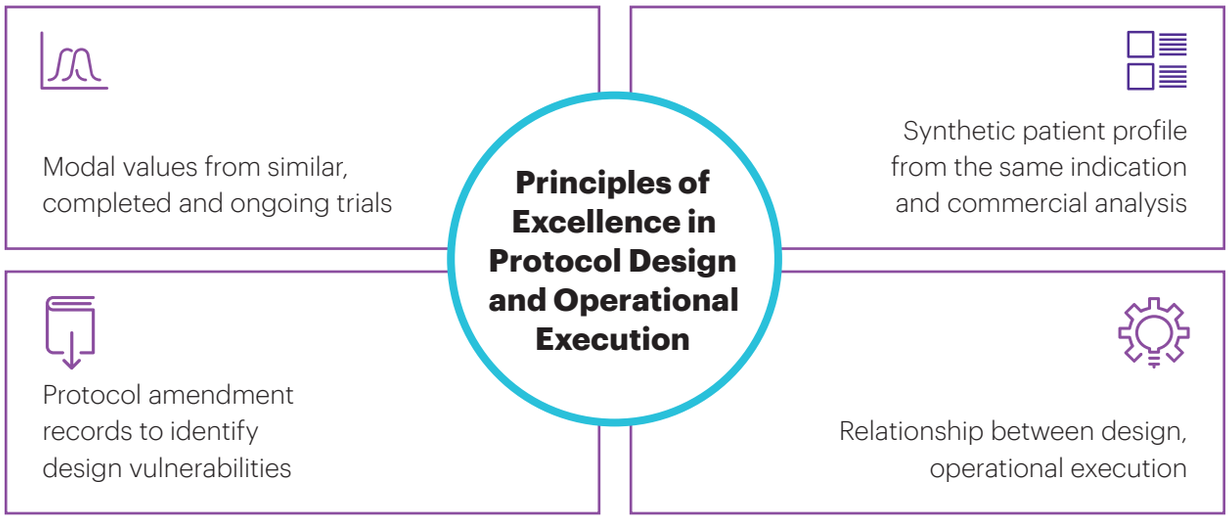
These four building blocks are crucial for pharma companies to make strides in the usage of synthetic data tools.

01 Draft your big plan while getting started with pilots

A bold global vision is needed to create a framework and shared sense of what is to be done and why, enabling the entire organization to move towards the goal of implementing in-silico clinical trials. That vision is bolstered by incorporating concrete objectives—faster development of drugs at lower cost, and an improved ethical position for patient recruitment and site burden. Prioritize assets with which to pilot the new data-led approach. Allocate resources and funding to support the transformation.

In practical terms, there are four key areas where predictive data analytics can deliver immediate value to the R&D organization, illustrated in Figure 2.

Figure 2: Four areas where predictive data analytics can deliver immediate value to the R&D organization



Source: Phesi analysis

02 Take stock of your data and look outside for more

Your own internal data is important, but it is not enough. At best, a pharma company will have only a fraction of the relevant data that is out there, so do not default to data that is controlled and at hand. External data sources offer greater value, and internal data may end up serving as more of a reference.

Leverage integrated external data sources to find the variety and volume of data that is needed for robust statistical confidence, and critically appraise their value. This should include industry views and data from other clinical trials (clinical trial design, protocols, amendments and results) in the same indication. Cross-industry performance data can validate internal data and thinking, increasing the value of n to boost statistical rigor. Healthcare data such as patient records will also further validate expectations in a real-world setting. An overlay of all three of these data sources is needed to build a robust predictive analysis approach.



03 Deploy smart algorithms for "what-if" analysis

Integrated and automated data systems, combined with analytical algorithms and statistical tools, should be a key part of any plan to accelerate the use of synthetic data. Such technologies support synthetic control arm design, protocol design, and accurate country and investigator site selection. Look for algorithms that can support "what-if" analyses and decision making, and accurately reflect what we can expect to see without testing. "What-if" scenarios might be "What if I include XYZ inclusion criteria or ABC outcome measures? Does it extend or limit my confidence on what we expect to see?" Collaboration and partnerships are ways to build and complement your in-house analytics capabilities.

Algorithms should enable very specific search criteria, so that you can easily filter and question the data. There is arguably no need to see the inner workings of the algorithms and how they arrive at the results, but it should show the level of confidence of that result. Clear access to the base data is critical.

Integration across functional boundaries is the goal, with algorithms crossing the lines between clinical, clinical operations, investigators, contract research organizations, patients, and healthcare professionals. The tools should be predictive rather than retrospective, and should look at the system as a whole rather than particular silos, as part of an integrated data analytics platform with AI and machine learning built in.

04 Evolve your operating model

Once you have a vision, and have looked at your data and tools, you can identify how your operating model may need to evolve. New governance, skills and processes are required to execute this fundamentally different approach to clinical trials and commercialization, where you first check the data and then decide on the development plan. Predictive data analytics should inform the design of the clinical development plan and target patient profile, based on an understanding of whether or not you can build a synthetic arm and what the hurdles for success are for this particular trial.

Take the innovation leap now

While pharma companies should each start making these changes individually, there is a broader cross-industry current of data ecosystem change that is also taking place. Regulators and payers will also need to adapt and prepare for this new way of leveraging data. While access to data remains challenging at the moment, the opportunity to accelerate and scale is huge.

More success stories are emerging of a synthetic data-driven approach to clinical trial design, and it's time to get ready. Getting the approach to in-silico clinical trials right will enable deployment at a meaningful scale, leading to a more reasoned process for clinical development, with lower patient burden and treatments that are much faster to market, at a lower cost. The potential of this opportunity is transformative, and companies that take the innovation leap now will find themselves in a much better position to benefit from the wealth of data that is out there.



Case study

The value of synthetic control arms is recognized by regulatory authorities

Synthetic data in oncology treatment development

Examples already exist of companies utilizing synthetic data to accelerate treatment approvals in oncology. Roche was seeking EU approval for Alecensa (alectinib) as a lung cancer treatment in 2017, and the EU's conditional approval required more evidence of the treatment's effectiveness relative to the standard of care (certinib). A synthetic control arm of 67 patients was accepted as evidence, speeding up availability of Alecensa in the EU by 18 months.[5]

In cases with short patient survival times, using a synthetic control arm is highly valuable for successful completion of clinical trials. The clinical trial for Bavencio (avelumab) from Pfizer and Merck KGaA to treat Merkel cell carcinoma, which was approved in 2018, used data from Electronic Medical Records in a synthetic control arm.[6]

Accelerated approval was also granted for Amgen's Blincyto for leukemia treatment by the FDA in 2014 and EMA in 2015, using a comparator arm of historical data from 694 patients based on 2,000 patient records for the Phase 2 study.[7]

About Accenture

Accenture is a global professional services company with leading capabilities in digital, cloud and security. Combining unmatched experience and specialized skills across more than 40 industries, we offer Strategy and Consulting, Interactive, Technology and Operations services—all powered by the world's largest network of Advanced Technology and Intelligent Operations centers. Our 514,000 people deliver on the promise of technology and human ingenuity every day, serving clients in more than 120 countries. We embrace the power of change to create value and shared success for our clients, people, shareholders, partners and communities. Visit us at **www.accenture.com.**

About Phesi

Phesi provides comprehensive clinical development analytical products and services for biopharmaceutical companies around the world. The company's integrated offerings cover the entire clinical development process—from development planning and indication assessment to protocol evaluation and design (including synthetic control arm), site selection, and trial implementation management. Phesi has the industry's most comprehensive and dynamic clinical trials database and predictive analytics tools, consisting of over 330,000 completed clinical trials, 604,000 completed research projects, data from >30 million patients, >4.2 million physicians and > 600,000 investigator sites worldwide. For more information, please visit Phesi.com.

About Accenture Life Sciences

Accenture's Life Sciences group is committed to helping our clients make a meaningful impact on patients' lives by combining new science with leading edge technology to revolutionize how medical treatments are discovered, developed and delivered to people around the world. We provide end-to-end business services as well as individual strategy, consulting, digital, technology and operations projects around the globe in all strategic and functional areas—with a strong focus on R&D, Sales & Marketing, Patient Services and the Supply Chain.

We have decades of experiences working with the world's most successful companies to innovate and improve their performance across the entire Life Sciences value chain. Accenture's Life Sciences group connects more than 15,000 skilled professionals in over 50 countries who are personally committed to helping our clients achieve their business objectives and deliver better health and economic outcomes.

References

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6508843/

[2] Bartlett et al. PLOS 2020. Concordance of real-world versus conventional progression-free survival from a phase 3 trial of endocrine therapy as first-line treatment for metastatic breast cancer

[3] https://www.projectdatasphere.org/

[4] https://www.phesi.com/phesi-and-sensyne-health-agree-collaboration-to-advance-development-of-synthetic-arms-in-clinical-trials/

[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7218288/#CIT0009

[6] https://www.clinicaloncology.com/Hematologic-Malignancies/Article/09-18/MCC-Rx-Marks-First-Use-of-Real-World-Data-for-FDA-Approval/52751

[7] https://onlinelibrary.wiley.com/doi/full/10.1002/pds.4932

Copyright © 2021 Accenture. All rights reserved.
Accenture and its logo are registered trademarks of Accenture.

Authors

Phesi

Jonathan Peachey
Chief Operations Officer, Phesi
jonathan.peachey@phesi.com

Dr. Gen Li
President and Founder, Phesi
gen.li@phesi.com

Dr. Paul Chew
Chief Medical Officer, Phesi
paul.chew@phesi.com

Dan Manak,
Business Development, Phesi
dan.manak@phesi.com

The authors wish to thank **Dr. Michelangelo Barone**, Clinical Science Director Corporate R&D, Alfasigma S.p.A for his contribution to this paper.

Accenture

Dr. Boris Bogdan
Managing Director, Global Lead Precision Oncology and Personalized Healthcare, Accenture
boris.bogdan@accenture.com

Dr. Sanjay Jaiswal
Managing Director,
R&D Analytics Lead – NA, Accenture
sanjay.s.jaiswal@accenture.com

Dr. Sandra Dietschy-Kuenzle
Senior Principal, Life Sciences, Accenture
s.dietschy-kuenzle@accenture.com

Follow us

@AccentureLifSci

www.linkedin.com/company/accenture