

# Adopting the Power of AI to Drug Development Projects



The top challenge today for life sciences and healthcare organisations is effectively extracting and operationalising information from complex sources for decision-making. Certara's deep-learning platform and model-based meta-analysis services offer integrated solutions for researchers seeking to enhance their R&D programs by leveraging all available clinical trial data sources.

**Matt Zierhut**, Vice President, Integrated Drug Development, Certara

**Nick Brown**, Associate Director of Marketing, Certara

## 1. How can artificial intelligence (AI) and deep learning help in integrating and analysing data in clinical trials? Can you share any examples?



**NICK:** A tremendous amount of documentation is created throughout the clinical trial lifecycle. These documents include recruitment forms, trial summaries, scientific publications, clinical study reports, drug labels and post-marketing materials; all of which hold valuable data points that can influence future trials.

The challenge with these documents is that it's incredibly time consuming to review and extract relevant data points from this content. That's where AI comes in.

AI in the form of generative pre-trained transformers (GPTs) and large language models (LLMs) are uniquely adept at searching and "understanding" complex content.

By applying these AI models to clinical trial documents, researchers can accelerate screening and data extraction workflows enabling them to collect highly-relevant information in a format that complements their analytics needs. For example, at Certara.AI we have a dataset creation tool that researchers use to take large corpuses of clinical trial documents and leverage AI to extract the relevant data points they need directly into a structured format. These newly structured datasets can then be used for expanding clinical outcomes databases or fed into visualisation tools to run a variety of different analyses.

## 2. How does deep learning benefit small compound analysis and generating new drugs? Can you highlight any advancements or breakthroughs in this area?



**NICK:** Just as LLMs and GPTs can be trained to "understand" language and then leveraged to analyse documents, they can also be trained on SMILES and SELFIES strings. SMILES and SELFIES are sequences of numbers and letters that represent molecules and in simple terms can be considered the "language of compounds."

Models trained on SMILES or SELFIES can perform exciting tasks such as property prediction and de novo compound generation which complement the existing analytics workflows of chemists and discovery scientists. For example, the Certara.AI team is collaborating closely with our

colleagues in medicinal chemistry, to develop these models. To date we've successfully deployed models to help users predict the toxicity, blood brain barrier permeability and lipophilicity of structures which can enhance which molecular properties to prioritise. In addition, we also have made tremendous progress in models for de novo compound generation which will deliver highly-relevant structures to complement existing discovery workflows.

## 3. What factors should organisations consider when selecting a deep-learning platform for R&D in life sciences and healthcare?



**NICK:** Flexibility, data access and model training are all factors organisations should consider when selecting a deep learning platform for R&D in life sciences. The AI landscape is rapidly evolving, so having a platform with the flexibility to handle multiple types of models will enable your team to shift as new innovative models hit the market. Secondly, data access will always be a challenge in AI. A platform that allows you to securely connect your internal data and apply AI models onto those assets enables you to create an expansive environment for leveraging AI across multiple data types and teams within your organisation. Last, but certainly not least, is the base training of the AI models. In many cases, AI models are trained on broad datasets that enable an expansive, but top-level understanding of concepts. At Certara.AI, we focus specifically on developing models that are trained specifically on life sciences content. This enables our customers to leverage these models with confidence knowing that they understand the unique complexities of the life sciences industry.

## 4. What is Model-Based Meta-Analysis (MBMA), and how does it impact drug development? Can you provide examples of how it optimises clinical trial outcomes and informs regulatory decisions?



**MATT:** Model-based meta-analysis utilises study results at the summary-level (aggregate data) to gain a deeper understanding of the landscape of both available treatments and treatments currently in development for any given indication or therapeutic area. MBMA is a broad term that can encompass almost all types of meta-analyses. These meta-analyses can range from simple pairwise meta-analyses of two treatments based on multiple similarly-designed studies (testing the same compounds), to network meta-analyses that compare multiple treatments using connected networks of studies (allowing for indirect comparisons of treatments without direct comparison data), and to full model-based meta-analyses that can add complex model structures to account for variability from dose-ranging data, longitudinal data, and data from studies with different populations or designs. MBMA can provide detailed context around any drug development program by enabling an appropriate comparison across all relevant trials, thus providing insight into the likelihood any drug in development would be able to successfully compete with other treatment options. This insight can be appreciated by all parties involved: sponsors, regulatory bodies, patients, and others.

## 5. How does MBMA help understand covariates in clinical trials? What methodologies are employed to unravel these factors and their implications for decision-making?



**MATT:** MBMA uses modeling techniques that are common to the fields of pharmacometrics and biostatistics to help explain variability in trial results that may be due to a variety of sources and factors. Usually, covariate effects are explored and described using additive, proportional, and power/exponential terms, but more complex functions can also be used. For example, to capture the influence of dose on outcome, a sigmoidal or Emax function can be applied. Exponential decay is also commonly used to explain changes in outcomes over time. Additionally, MBMA can capture the influence of both prognostic and predictive covariates, a difference that is critical to development decisions. Prognostic covariates are factors that influence trial outcomes at the general treatment arm level, not relative to any other treatments. Prognostic covariates can be thought of as covariates that affect populations receiving placebo, or control treatment, or they can help explain differences in disease progression. Predictive covariates, on the other hand, help to explain differences in relative treatment effects, and relative effects are ultimately what involved parties are interested in. The key question being: does this drug elicit a larger benefit than the control or competitor treatment?

## 6. What limitations and considerations should researchers and stakeholders keep in mind

## when applying MBMA and AI-powered analytics? Are there potential biases, challenges in data interpretation, or regulatory considerations?



**NICK:** Data quality is one of the top considerations to keep in mind. MBMA requires highly-specific datasets to ensure accuracy for any statistical analyses taking place. As discussed earlier, AI can accelerate the creation of these datasets. However, it's critical that validation workflows are in place to ensure AI predictions are accurate. This is a key focus for us at Certara. AI. We've developed human-in-the-loop validation capabilities that enable users to review AI predictions and adjudicate their accuracy and relevance before the dataset is finalised and used by MBMA experts.

## 7. What are the common challenges in applying MBMA to gain insights into disease, trial characteristics, and investigational drugs? How does Certara overcome these challenges and develop innovative solutions?



**MATT:** Each MBMA project presents its own unique challenges; however, the most common challenges typically involve data availability. This could be high-level data availability like the number of trials that have published results in the indication of interest. MBMA in rare diseases is not frequently done due to the typically



low number of drugs in development for a specific rare disease, although MBMA has been utilised in multiple rare disease cases – again it depends on the unique case and questions being asked. Other data challenges can also exist in more popular indications that have many published trial results to work with. Trials in these indications may not always publish the same details around population or trial characteristics. If these missing characteristics are potentially influential to trial outcomes, analysts would have to choose between excluding these trials or imputing the missing data. The most common solution is data imputation, thus keeping trials with other potentially relevant information. However, sensitivity analyses are always conducted to ascertain the potential influence of data imputation techniques. Other data challenges can

exist in how published results define important variables and even endpoints. If a common definition is not used across all studies, the definition may be a significant covariate itself. Finally, there are general challenges in transferring data from external (and internal) sources to an analysis-ready database, typically due to human error. These variables make it incredibly challenging to effectively collect data through manual efforts. Fortunately, advances in AI are a key solution. At Certara, we leverage advanced large language models trained on life science concepts to help our teams extract data from unstructured documents, identify semantically relevant content that is at risk of being overlooked and deliver results in a structured spreadsheet. Automating this first step with AI allows our team of expert curators to focus on data quality. With a low code AI validation tool, the team can more efficiently conduct QC workflows enabling them to be more productive while mitigating some of the tedious tasks that often lead to human error.

## 8. What are the benefits of conducting a meta-analysis by systematically gathering data from maintained databases? How does this approach enhance reliability and aid evidence-based decision-making in drug development?



**MATT:** As any modeller would (or should) tell you: garbage in, garbage out. The benefits of using systematically gathered data can be directly observed in the quality of the model output. Additionally, by maintaining these systematically developed databases, one can quickly jump into any

new analysis within an indication that has an already existing database. If you can trust that existing databases accurately portray all the relevant and available information, you can not only quickly initiate work to inform critical decisions, but you can also be confident that those decisions were properly informed. The highest priority of the Data Science team at Certara is to develop and maintain accurate databases using a systematic review and quality control process.

## 9. How does Certara's advanced deep-learning platform benefit R&D programmes in life sciences and healthcare? Can you share success stories or use cases?



**NICK:** Certara's AI analytics platform helps life science R&D teams solve two key challenges — provide a solution for applying LLMs and GPTs to structured and unstructured files and make those multiple data sources searchable and accessible in a single platform. This combination improves collaboration and accelerates insight discovery that can inform a number of go/no-go decisions across the drug discovery pipeline.

As previously mentioned, so much of the data needed for R&D efforts resides in literature-based documents. The Certara AI platform's strength in analysing and bringing new value to this content enables it to be leveraged in a number of use cases. For example, using LLMs to accelerate large scale systematic literature reviews that are used to inform trial design and development, GPTs assist in the creation of regulatory documents and concept tagging to improve insight discovery and data standardisation.

## 10. How have AI-powered analytics improved decision-making in drug development through model-based meta-analysis?



**MATT:** Effective MBMA starts with high quality data. The area that we see the greatest promise in AI complementing MBMA is in the identification and collection of relevant data needed to effectively conduct analyses. As mentioned earlier in this interview, much of the data we need to understand a clinical landscape comes from unstructured documents. AI's ability to "comprehend" this content enables us to effectively identify relevant insights, and curate more accurate datasets that can improve our understanding of the given area we're studying.

## 11. What motivated Certara to incorporate deep learning expertise? How has this collaboration benefited Certara's clients and drug development initiatives? Can you share specific success stories?



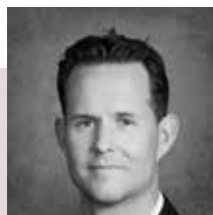
**NICK:** AI holds tremendous promise in life sciences and drug discovery. Certara AI's focus area of AI development complements Certara's product portfolio. By integrating this technology into the Certara product portfolio, we're able to add new predictive and generative AI capabilities to drug discovery workflows that arm our clients with the ability to easily access insights that can improve go/no go decisions.

## 12. What tools and services does Certara offer to help life science companies analyse and interpret data for drug development decisions? Can you provide examples of their impact on decision-making?



**MATT:** Certara offers many more tools and services than what we have discussed today. We are a full-service company that can improve drug development at all stages. To demonstrate our size and success, in each of the last nine years, 90 percent of new drug approvals by the U.S. Food and Drug Administration's (FDA) Center for Drug Evaluation and Review (CDER) were received by Certara's customers.

Focusing on Certara AI, Data Sciences, and MBMA consulting services, we offer essential tools to identify, curate, and analyse both public and proprietary data sources. Our recently developed and newly improved software is available to provide all available relevant information to the decision maker. This information could come in the form of a comprehensive database curated from multiple sources, or it could



**Dr. Matt Zierhut** advances the integration of external aggregate clinical trial data into development decisions and commercial and regulatory strategy via model-based meta-analysis (MBMA) at Certara. Matt works closely with clinical development teams to ensure MBMA is leveraged for optimal impact when making the most critical decisions. Previously, Matt was at Janssen (J & J) where over the past 5+ years he has led the development of their global MBMA capability and has worked together with many clinical development teams to ensure optimal impact of MBMA into the decision making in their programs.

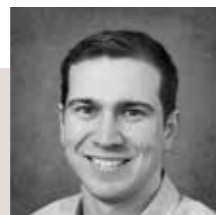
<https://www.certara.com/teams/matthew-zierhut/>

come in the form of simulations based on a model that was developed using the comprehensive database. Certara provides the tools and the recommendations to make better decisions.

## 13. How does Certara handle the limited sharing of individual-level data from in-house trials? How does it leverage internal data to enhance research and development capabilities?



**MATT:** Successful drug development depends on making wise decisions about portfolios, clinical trials, marketing, etc. We are continuously faced with the challenge of deciding whether to continue development or stop it. To support those decisions, we gather data, typically through clinical trials. We analyse the data from those clinical trials, and then we use these analyses to build models that we then use to predict what may happen in the next trial. The data collected from these in-house trials are "internal data" or "proprietary data." Companies rarely share individual level data. But, they all publish most of their aggregate level trial results.



**Nick Brown** is associate director of marketing at Certara AI. He manages customer and partner communication and engagement for the Layaar platform and AI integrations across the Certara product portfolio. Nick has over 10 years of experience in AI marketing where he focuses on educating a variety of audiences on complex technology across deep learning, AI and big data.

<https://www.certara.com/teams/nick-brown/>

AUTHOR BIO

Before publishing results from an internal study, sponsors are in a unique position where they have access to all published competitor data, and are the only ones with access to their own proprietary data. Certara utilises MBMA to put this proprietary data into the proper context of the published summary-level trial results, thus enabling the sponsor to make critical decisions before others have seen their new trial results. By fully understanding the landscape, now including their drug, they are able to make fully-informed decisions about the best next steps for their compound, whether it's advancing to the next stage or shifting resources to a different compound, with a better predicted probability of success.

## 14. How does the integration of AI in Certara's CODEX platform enhance MBMA capabilities? In what ways does AI empower the analysis of complex datasets and enable informed decision-making in drug development?



**NICK & MATT:** High-quality clinical outcomes data is at the core of the CODEX platform. By adding AI into CODEX, we're able to provide our customers with an intuitive solution for custom dataset curation that mixes our CODEX indication databases with other relevant data sources, including customer's proprietary data. As a result, our customers can dynamically expand the datasets they're analysing to quickly provide them greater insights and results that impact their most critical drug development decisions.