Paper IC11

Beyond the Hype: A Pragmatic Approach to Leveraging AI/ML for Data Review

Robert Musterer, eClinical Solutions, Mansfield, CT, USA

ABSTRACT

There is still a lot of hype around the potential of AI/ML to transform the life sciences industry and while advancements in clinical development have accelerated in recent years, the adoption of innovative technologies has lagged. Hype aside, taking a pragmatic approach to AI/ML enables life sciences companies to adopt new capabilities and accomplish objectives in a more efficient, cost-effective manner. This session will cover how a top 20 Sponsor partnered with eClinical Solutions to leverage the AI/ML capabilities to enhance clinical data review processes for greater efficiency using fewer resources. With these capabilities, high quality data is produced in near real-time to support accelerated submission and collaboration with the FDA. We will discuss the pragmatic approach to adoption including how the AI-enabled data review use cases were selected and evaluated, the validation process for ensuring accuracy of the AI/ML models, and the resulting efficiency gains.

INTRODUCTION

As a regulated industry, Life Sciences has historically taken an extremely conservative approach to adopting new technologies. As an industry veteran with over 40 years in this field, I am skeptical of grand expectations surrounding AI/ML utilization in clinical trial data review. We have been talking about the pressures on the industry to be more cost conscious and efficient for decades, yet the progress has been painfully slow. Look at electronic data capture (EDC): I started working with the 1st commercially available EDC tool in 1985; back when it was called remote data entry (RDE), before it was called remote study monitoring (RSM), which was the moniker before it became EDC. We deployed it on numerous studies so successfully that colleagues in clinical operations stated they would have needed 3 times more resources to complete their studies without it. Yet it was well over 2 decades before EDC was a standard practice across the industry. Once EDC was well accepted, expectations for deploying ePRO (the prior term for eCOA) were that the industry had learned from the EDC experience and would embrace it rapidly. Did not happen. Fast forward to current times, Covid led to the use of decentralized trials and highlighted the value of collaboration. Expectations soared that this would be the future of clinical trials. As soon as pandemic concerns eased, the use of decentralized clinical trials (DCT) dropped so precipitously that the leading vendors were forced to have dramatic layoffs. If the refusal of sponsor companies to allow learnings from their data to be applied to ML models that others can benefit from is an example – I'd say the collaborative spirit has plummeted.

If you read enough articles, you'll find a wide range of quoted success metrics for implementation of software, unfortunately all of these metrics indicate that way too many, and often the majority, of implementations fail.
This backdrop of reality drives taking a realistic and pragmatic view of how we can best incorporate AI/ML techniques to facilitate clinical data review. A pragmatic approach balances conservativism, transparency, explainability, and innovation. As with any major initiative, the probability of success is improved by first establishing perspective and setting expectations. In this initiative, this is followed by the selection and evaluation of use cases to implement, then the refinement of the process and ultimately assessment of results.

PERSPECTIVE AND EXPECTATION SETTING

Decades of clinical trial data processing have led to a large compilation of data review objectives that were defined using the tools available and as a result are deterministic (rules based). Programs written to support such objectives result in definitive fully explainable lists of data conditions to be addressed. This is particularly true for data management review objectives, since these are focused on more precisely defined requirements. This history strongly influences expectations surrounding the output of AI/ML based approaches.

These expectations are bolstered by the excitement surrounding the promise of AI/ML, such that it is easy to lose sight of constraints when applying AI/ML to clinical trial data. AI/ML models are highly dependent on having very large data sets to be trained on. Unfortunately, in the clinical trial space, companies are understandably concerned about protection of intellectual property and patient privacy, which leads them to be extremely reluctant to allow their trial data to be used for training the models. This reluctance makes it challenging to fully comply with the Good Machine Learning Practices (GMLP) Principle #3.

> **Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population:** Data collection protocols should ensure that the relevant characteristics of the intended patient population (for example, in terms of age, gender, sex, race, and ethnicity), use, and measurement inputs are sufficiently represented in a sample of adequate size in the clinical study and training and test datasets, so that results can be reasonably generalized to the population of interest. This is important to manage any bias,

promote appropriate and generalizable performance across the intended patient population, assess usability, and identify circumstances where the model may underperform. [1]

Yet paradoxically, many expect to be able to use models trained on data from other companies. Even when a few do allow their data to be used for training, the resultant learnings achieved by the models are often not truly applicable to other clients. This is due to the limited size of the available training set, plus the high probability that the indications, drug classes, and trial designs represented in the training set may not adequately represent the running trials of other sponsor companies. When ML models are trained on smaller than ideal data volumes, a condition referred to as overfitting is likely to occur. An overfitted model performs well on the training data but fails to adequately generalize to new data. Another aspect of ML models is the condition known as model drift. For clinical trials, drift can be related to a variety of causes:

- Growth of the underlying data available to train the models,
- Introduction of a new indication to be treated,
- New investigational product is being researched,
- The pathogen being targeted is mutating and evolving.

The net effect is that ML models need to be monitored and periodically tuned and retrained. For supervised models, this includes a substantial investment of time and resources to provide reliable ground truth data.

The next request is to train the models using real world data (RWD). This raises another set of challenges:

- RWD is substantially dirtier than clinical trial data.
- RWD is collected for different purposes which biases the reporting:
  - Coding conditions to maximize billing.
  - Under reporting of many events and observations.
- Data structures vary greatly.
- Content of RWD collected per patient is typically a fraction of that collected for a clinical trial participant.
- The cost of acquiring RWD can be very expensive.

It is best to view RWD as a complement to, rather than a replacement for, clinical trial data. As a complement, it can be useful to provide a broader more diverse patient population as well as long-term outcomes and real-world effectiveness.
The level of complexity of review objectives varies considerably. Many review objectives are most efficiently managed using traditional deterministic rules-based approaches, while others are best supported using statistical methods, and others using an AI/ML based approach.

A deterministic rules-based approach is more efficient for:

- Simple Straightforward Logic or Boolean Conditions: If the decision-making process involves straight-forward logic or Boolean conditions, a rules-based system is more efficient. For example, if-else conditions based on specific criteria.
- Precise Rules: In cases where the decision-making criteria are well-understood and easily expressed such that domain experts can explicitly define decision rules based on their knowledge and experience.

Statistical approaches, such as Central Statistical Monitoring (CSM), are most efficient when the desire is to find unusual data that is defined as being significantly different from others, as opposed to being predefined as a numeric threshold. Looking at legacy review objectives, you'll often find cases where the stated objective is to find values that have changed by a defined percentage. In most cases, these stated percentages are used across all studies which ignores the context of any particular study. By using a CSM based approach, the percentage is not a set threshold. Instead, this approach compares each individual unit against all the others, and reports those that are atypical. This has the benefit of not relying on an educated guess of the threshold percentage, and only reports on those that are atypical and thus worthy of further review.

For review objectives that can be supported using deterministic or statistical approaches, the intelligent application of reusable code outperforms reliance on ML based approaches. The reference to intelligent application means having code designed to support parameter substitution and the ability to use the concepts identified by the automated data classification. This approach provides several benefits:

- Reusable code.
- Easily explainable code.
- Easily defensible validation.
- Easily interpreted output.
- Short code development time for new review objectives.

AI/ML based approaches extend the range of review objectives that can be supported, and even introduces the opportunity to perform some reviews not previously possible. Use cases include (but are not limited to) anomaly

detection, pattern recognition, natural language processing of text entries, and handling cases with a large number of permutations. For example, anomaly detection techniques can find atypical shifts in measurements taking into account the time between the shift and the direction of the shift, regardless of whether or not any of the values are outside of the normal range. Pattern recognition can find observations that occur at an atypical frequency or atypical timing. Natural language process can be used to compare the contents of open text fields for a variety of use cases such as identifying events that should be reported elsewhere in the CRF or comparing the concept against controlled terminology lists. Because these approaches do not rely on deterministic logic, the output may include false positives and/or miss some true positives. Given our industry's need to assure quality, models are engineered to minimize missing true positives, which has the side effect of increasing the number of false positives. As such, a human in the loop review process (aka manual review) of the output is required. Despite this, the benefit is the guided approach provided by identifying the entries with the greatest likelihood of exhibiting conditions that require closer scrutiny. This results in a dramatic reduction in the total number of records that require review.
For Data Managers, most of their review objectives have been honed over decades using deterministic rules-based approaches. As such, it is challenging to define review objectives that leverage AI/ML models more efficiently than a library of deterministic checks that have been designed for reusability. Data Managers have become accustomed to the checks they use providing clearly defined output. Therefore, a combination of deterministic checks complemented with AI/ML models is the preferred approach. The reality is that even with this hybrid approach, many of the AI/ML model results are finding atypical patterns which require the expertise of a clinician rather than a Data Manage to interpret.

Let's face it, the role of the Data Manager has changed considerably over time. Back in the 1980's and 1990's, Data Managers were expected to have a degree of programming proficiency and would be responsible for some clinical data plausibility review such as using references including the Physician's Desk Reference (PDR) and American Drug Index to check if concomitant medication reports were reasonable. Over time, the requirement to have programming competency was eliminated and the responsibility for clinical data plausibility review was transitioned to clinical roles. An artifact of this evolution (or perhaps devolution) was the rise of the view that data management services are a commodity that can be outsourced. Outsourcing success varied widely across companies, and there is still some pendulum swinging back and forth on this point. Regardless of the level of outsourcing success, external pressures continue to ramp up putting greater emphasis on efficient and cost-effective approaches to managing and reviewing clinical trial data. Advances in AI/ML techniques (and the associated hype around them) hold the promise of assisting in improving the process for managing and reviewing clinical trial data. However, to take advantage of using these techniques will require another evolution of the data management role. There are numerous articles and presentations about evolving from clinical data management to clinical data science. Implications of this evolution mean that data mangers either must become more technologically proficient again and able to resume some of the data plausibility review; or be at greater risk of being outsourced as a commodity plus have decreased perceived value within an organization (which has its own set of implications related to career opportunities and compensation).

These realities must be taken into account when considering how AI/ML techniques can be used to support processing data from clinical trials. It is the recognition of these realities that is guiding the approach eClinical Solutions is following for incorporating AI/ML driven approaches to further enhance the automation and data review features within the elluminate platform. This journey includes foundational capabilities as well as specific point use cases. Our philosophy is to apply the most efficient and cost-effective approach to each problem, whether that be AI/ML, statistics, or traditional deterministic logic.

PRAGMATIC APPROACH

A pragmatic approach emphasizes efficiency. It recognizes the need for explainability and validation that is most easily met using a deterministic approach. It recognizes that companies have existing libraries of such checks. Therefore, the pragmatic approach uses deterministic checks as a foundational element. It ensures this element is used in an efficient manner by taking advantage of parameter substation to allow these checks to be easily reused. It next complements this by adding ML techniques to automate the identification of fields in new source files. This then enables metadata driven automation that uses the attributes of a newly ingested file to identify components from a global library for deployment. These elements include items such as edit checks, visualizations, and transformation mapping logic.

This approach also recognizes that the classic approach to clinical data management is fraught with inefficiencies. In particular is the low percentage of queries that result in updates to the data.

> "… queries are only leading to 1.7% of eCRF changes after data has been entered by the site. This indicates that the outcome does not proportionally reward the significant effort that many CDM organizations place on querying data through traditional data review and data validation strategies especially for non-critical data."[2]
>
> SCDM The Evolution of Clinical Data Management into Clinical Data Science

Thus, the next component is a feedback loop in which edit check performance metrics are reported for adjudication. This provides for a data driven decision process to assess if any of the library elements should be retired or flagged

for use in specific scenarios (as the criticality of a field may vary based on the indication being studied or the profile of the investigational product).

Historically, the next step in the clinical data review process is to employ exception listings. These listings use deterministic logic to identify a host of potential data conditions for a person to review the identified subset of data and make a judgement call if any of it needs to be queried. Since the conservative nature of the Life Sciences industry results in resistance to adopt new approaches, this pragmatic approach progressively replaces these with AI/ML models. The progressive approach enables reviewers and quality assurance representatives to get comfortable with the use of AI/ML and appreciate the appropriateness and value of these techniques in identifying the records of greatest interest. One of the selling points is that these AI/ML models not only reduce the reliance on exception listings but can also be used to provide guidance on data to focus on within visualizations and even reduce the reliance on visualizations.

We've all heard claims of AI/ML success stories. These quotes come from users who belong to the Innovators segment in a Rogers' bell curve.

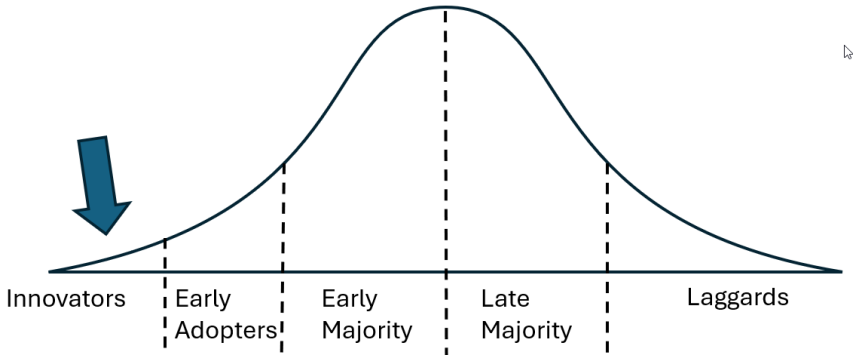

Figure 1: Bell Curve, see Reference 3 Everett Rogers

Geoffrey Moore expanded this concept to include a chasm within the early adopters for innovative solutions.

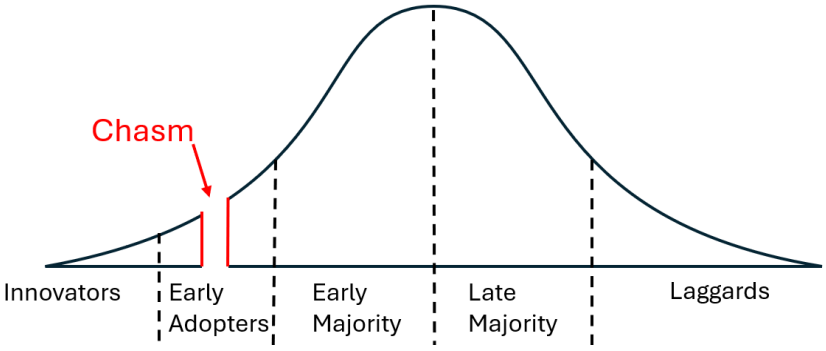

Figure 2 : Chasm, see Reference 4 Geoffrey Moore

Moore advocates for consdering two groups of customers, the visionaries and the pragmatists. These groups have very different expectations. I believe what we are seeing today are initial success stories associated with exploratory use by visionaries; and we need to face how we can cross the chasm to meet the expectations of the majority. It is my belief that taking a pragmatic approach is the way to position AI/ML in a manner that is enticing for the majority.

If we now consider the Gartner Hype Curve, the Gartner analysts placed 'AI in Clinical Development' past the peak of inflated expectations into the downward slide in expectations.

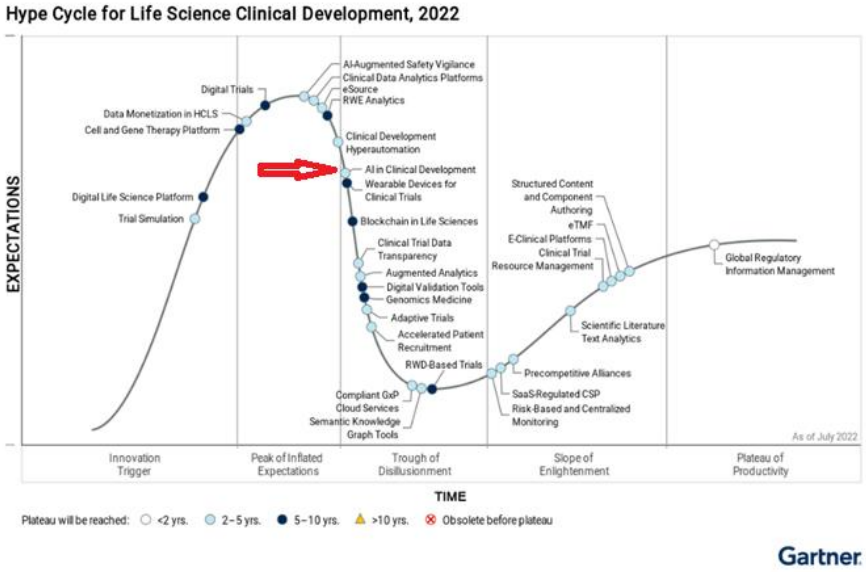

Figure 3:Hype Cycle, see Reference 5 Gartner

This is consistent with the both the conceptual framework of the innovators and the chasm. The way I think about it, the chasm coincides with the 'trough of disillusionment'. Our objective is to alter the negative slope so the decline is less precipitous, in other words, to build a bridge to the 'plateau of productivity'. This bridge is the pragmatic approach that combines legacy deterministic approaches, an intelligent utilization of reusable elements, copled with AI/ML driven techniques to improve efficiency and extend the scope of review.

LESSONS LEARNED

In working with various clients we've seen how inflated AI/ML expectations can sabatogue success. We've seen cases where there is a degree of magical thinking, assuming that AI/ML will somehow transform their work, produce 100% accurate results with complete explainability, all with minimal effort. On top of that, to do so while having been developed using someone else's data. In these cases, it is clear that a comprehensive change management process with a significant educational component is required.

This is why we promote the pragmatic approach leveraging legacy deterministic approaches that are supplmented and complemented with the introduction of AI/ML driven techniques. The following lists the primary features of relevance to this pragmatic approach.

Feature 1: Automated Data Classification

This introduces the use of AI/ML to automate the identification of the metadata for a new file. This approach uses CDISC (Clinical Data Interchange Standards Consortium, https://www.cdisc.org/) concepts from both CDASH and SDTM standards as the basis of a common data model. Why? Because some study sources have been designed using CDASH while others have been designed using an SDTM informed data model. So we're maximizing the probability of identifying the content of a new file. This is a foundational step supporting downstream automation.

Value: Improves efficiency reducing data review related cycle times.

Acceptance Driver: Inclusion of a human adjudication step provides assurance that the identification is correct.

Feature 2: Edit checks

Edit checks are a foundational tool for assuring data is clean. These deterministic checks are the 1st line of review, automating the creation of questions (issues/queries) about inconsistent and/or erroneous data conditions. AI/ML assistance is provided by taking advantage of the results of the automated data classification for metadata driven automation to select relevant checks from a global library.

Value: Improves efficiency reducing data review related cycle times.

Acceptance Drivers:

- Provides comfort of a well known and accepted process,
- Reduces time to define full set of edit checks,
- Extends scope of data that can be checked,
- Reduces reliance on manual review of exception listings,
- Automates generation of queries,
- Automates closing of queries.

Feature 3: Exception Listings

Exception listings have historically been the 2nd line of review. They list a subset of data identified using deterministic logic. The listed subset represents a constellation of related data points that require judgement to assess if there is a data condition that warrants questioning. While the nature of reviewing data within exception listings remains the same, the use of automated data classification brings efficiency by being able to select relevant listings from a global library via a metadata comparison.

Value: Improves efficiency.

Acceptance Drivers:

- Provides comfort of a well known and accepted process,
- Reduces time to define full set of exception listings.

Feature 4: AI/ML Review models

Introduces the use of ML models to provide a guided user experience in which the models have identified data with greatest likelihood of requiring human review. As such AI/ML Review emerges as the new 2nd line of review, delegating exception listings to a shrinking 3rd line of review.

Value: Improves efficiency, expands scope of data review, reduces data review related cycle times

Acceptance Drivers:

- Replaces some of the exception listings,
- Improves upon exception listings by doing a better job of narrowing the set of records to be reviewed and, with feedback over time, automate the generation of questions when appropriate,
- Expands the scope to include conditions that weren't anticipated and hence not covered by an exception listing,
- Incremental expansion of models over time provides a path for users to become comfortable with adopting this new approach.

USE CASE SELECTION

To select use cases for development, we reviewed multiple data review plans. For each review objective, we assessed if it could be reasonably met with a deterministic based approach, or whether it was a candidate for application of AI/ML. Of these, we focused on those that were rated as requiring the greatest amount of hours for manual review. We then weeded down the list of candidates based on those for which there was available data that could be used for training and testing the models. Finally, we confirmed our assessments with representative data reviewers.

SAMPLE AI/ML REVIEW MODELS

Concomitant Medication Abnormal Duration

The manual review of data, particularly through exception listings, is a time-consuming and labor-intensive process. One crucial task involves identifying concomitant medications with unusually short or long durations. While essential for patient safety, this task is prone to errors and inefficiencies. This model aims to automate parts of this process using machine learning techniques and insights from historical data.

The primary goal Is to alleviate the burden on data monitors and enhance data review efficiency In clinical trials. These capabilities can be leveraged to identify approximately 86% of anomalous data, reducing the amount of data that requires review by data managers. As a result, data monitors can avoid reviewing a significant portion of data while still capturing the anomalous data points, greatly

reducing their workload.

Anomalous Adverse Event Durations
In the realm of adverse events, occurrences that exhibit unusually short or prolonged durations often warrant a closer look. Such instances may either hold significant clinical implications or stem from data entry errors. An automated solution for detecting these anomalies can substantially enhance the efficiency of data review processes. The Automated Adverse Event Duration model harnesses an unsupervised machine learning algorithm to automatically pinpoint these exceptional adverse events and surface them to users.

The primary objective of this use case is to shine a spotlight on data points that might otherwise slip under the radar of data reviewers. This crucial task empowers data monitors to uncover data inconsistencies that might easily elude detection in a conventional study scenario. In essence this module acts as a guardian, safeguarding the integrity of your data by bringing into focus those critical data nuances that would typically go unnoticed, ensuring the highest standards of data quality.

Domain Classification
Across clinical trials, submission to the FDA requires all data to be clean and recorded in the correct forms. Accurate data recording on the correct forms/domains is essential for maintaining the integrity of clinical trials, meeting regulatory requirements, facilitating analysis, and ensuring patient safety. With the adoption of SDTM v3.2, procedures now fall under a separate form than medical history and adverse events. This entails a data reviewer to check all records across many fields of Medical History, Adverse Events, and Procedure forms to confirm correct data entry. Traditionally, manual data review – which is a highly labor-intensive and time-consuming process – is performed to guarantee this.

Domain Classification identifies records and fields that do not belong in the entered form (i.e. a procedure entered in the medical history form) to accelerate timelines and minimize errors during the data review process. This model leverages the strength of a large language model to automatically flag incorrect data entries at both a record and field level. With a near 98% reduction in records to review, it significantly reduces data monitors' workload, freeing up their time for more complex tasks.

Univariate Outlier Detection
Exception listing reports are created for every study within a clinical trial. This involves defining data review objectives, rules to accomplish the objectives, and then coding/programming these rules. Reviews of these reports are prone to manual error and require a significant number of resources. Additionally, given the variety of the diverse types of exceptions and the datapoints within the exceptions, this review process requires a high level of domain expertise.

Univariate Outlier Detection automates the identification of anomalous data points on any numerical data set to reduce the time and resources spent on data review, increase consistency, and improve data quality. This model uses a collection of unsupervised outlier detection algorithms to flag anomalous datapoints and surface them to the end user. Users can configure the numerical datasets for review, allowing customized outputs targeting specific exception listings. As new data comes in, it can quickly find and update the anomalous data points, making the review process easier.

Concomitant Medication Consistency
Concomitant Medication datasets include the medications patients have taken during a clinical trial. Each row consists of the medication name, route, frequency, dose, unit, indication, and dates along with other variables – FDA submission requires that the data is consistent and recorded correctly in order to maintain data integrity. The identification of inconsistencies is traditionally performed manually to ensure that unit, route, and frequency are consistent with the name of the medication across each row. Because there are oftentimes thousands of records within these datasets, manual review is time consuming, resource intensive, and increases the opportunity for human errors to be made.

The Concomitant Medication Consistency Model predicts the unit, route, and frequency of a concomitant medication given for each entry and compares it with the recorded value for consistency. This model uses a supervised CatBoost algorithm model to identify records that are not consistent and automatically surface these to the user at a record level. This model is designed to assist the data reviewer when checking for consistency between medication, dose, unit, route and frequency. Leveraging this model

significantly reduces the workload of the data reviewer by eliminating the need to review up to 97% of concomitant medication records.

Concomitant Medication Indications

Traditionally, medical reviewers examine and authenticate concomitant medications and their paired indications for accuracy verifying that every medication is correctly matched with its designated use as per the guidelines of the study protocols. To simplify this task and further enhance the efficiency of clinical data management, the Concomitant Medications Indication model not only evaluates the correctness of a medication-indication pair but also provides a nuanced confidence score, indicating the level of certainty in the alignment between the prescribed medication and its specified indication. This model uses an information extraction technique, fine-tuned on open-source data, to associate each medication with a curated list of correct indications from a comprehensive medication-indication dictionary. Building upon this foundation, a sophisticated clinical similarity scoring system further harnesses the capabilities of Large Language Models.

As a result, the volume of medication-indication pairs requiring manual review is reduced by 60%, all while maintaining a remarkable sensitivity rate of over 90%. In practical terms, this translates to a fast and effective streamlined workflow for medical monitors, paving the way for a more efficient and resource-conscious future in clinical trial data management.

WHAT DIDN'T IMPROVE WITH AI/ML

SHIFT IN LABS AND VITALS

We originally developed an ML model that compared numeric values for lab analytes and vital sign measurements that compared changes between two results within a comparable timeframe, to find atypical changes, regardless of whether the values were within or outside of the normal range. The initial round of ground truthing went well. However, in actual implementation users found it too challenging to understand why the shifts in values were identified as anomalies. An additional round of ground truthing with a separate set of testers led to very different results. At the end, we found the preferred and acceptable approach was to simply find atypical percentage changes between values.

VALIDATION APPROACH

Appropriate datasets are required for successful AI and ML adoption. The quality, volume, and representativeness of the data used for training affect the accuracy of the models and their ability to generalize to new unseen data. We train the bulk of our models using historical data, working throughout the development cycle with independent members of our staff who are subject matter experts in performing data review. This includes understanding the subtleties in the data and assessing our data needs. We maintain separate training, validation, and test sets, per accepted best practice, to ensure the development of machine learning models that generalize well to new data by allowing for the training of the model, tuning of hyperparameters, and unbiased evaluation of performance on unseen data. We leverage our subject matters experts again during the testing phase. Feedback from these exercises lead to model refinement iterations until the results are satisfactory. Clearly, any specific client's set of trial designs, indications, and investigational products must be taken into account. Therefore, to improve performance of supervised models, whenever the client has historical trial data, retraining the models using the client's own data should be considered.

As mentioned, we train the bulk of our models using historical data. These are studies for which some of our clients have granted us permission to use anonymized copies of their data for initial model development. As seen with the collaboration in response to the Covid-19 pandemic, everyone benefits from collaboration. Therefore, we encourage all clients to similarly allow us to include anonymized copies of their data to continue developing more robust AI/ML models to assist with data review.

The steps involved in the validation approach included:

- Sets of data identified to be used for training, validation, and testing.
- Data Science team evaluation of candidate techniques.
- Independent team develops ground truth data.
- Data Science team selects best performing solution.

- Data Science team develops, trains, and tunes preferred model using training data.
- Data Science team assesses model performance against several measures:
    - True Negatives
    - False Positives
    - False Negatives
    - True Positives
    - Recall
    - Precision
    - Prevalence
    - Predicted Negative (records not flagged for review)
    - Positive Predicted Value (records correctly flagged)
    - Negative Predicted Value (records correctly not flagged)
- SME evaluation of results by running on independent test data.
- Traditional SQA testing
- Limited rollout for feedback.

Feature 5: Data Management Workbench

The data management workbench can be viewed as a complementary tool to support review actions associated with the first 3 lines of review, and while also offering a 4th line for ad hoc review. Within the data management workbench, queries and issues can be viewed and associated data can be reviewed in workspaces. Similarly, if a user wishes to further explore data beyond that presented by default in the out of the box dashboards or exception listings, they may create ad-hoc listings and/or custom dashboards. Note: output from the first 3 review lines is presented within the data management workbench.

Value: Improves efficiency, reducing data review related cycle times.

Acceptance Drivers:

- Provides familiar option to conduct any ad-hoc enquiries assisting with comfort for adoption,
- Provides central location for managing data from all sources,
- Provides central location for tracking review progress,
- Supports sending questions to all source data providers (EDC or other)

Feature 6: Clinical Data Analytics

Analytics is the 5th line of review. This takes advantage of standardized data structures to provide visualizations of data across multiple studies for reviewing data for trends. Evaluation of trends typically requires clinical perspective and as such is commonly used by a clinician role. Use of these visualizations decreases over time as the AI/ML models are highlighting data conditions for further exploration.

Value: Improves efficiency, reducing data review related cycle times.

Acceptance Drivers:

- Provides familiar option to conduct any ad-hoc enquiries assisting with comfort for adoption.

RESULT

The primary result is the pragmatic approach was accepted and improved the efficiency of the data review process. This is exemplified by the following customer quotes:

1. **"It takes a village to develop and deploy these models."** I like this quote because:
    - It provides a rememberable spin on a cultural tag line encapsulating the GMLP principle #1 'Multi-Disciplinary Expertise Is Leveraged Throughout the Total Product Life Cycle'.
    - It highlights that it takes considerable effort across multiple roles to develop and maintain AI/ML based approaches. These are not easy quick wins.
2. **"Traditional CDM approaches won't scale."** This quote illustrates the promise that reviewers recognize that the traditional approaches won't be able to keep up with demand being placed on them. So, there is hope to be optimistic that the industry won't take 2 decades to adopt AI/ML to a significant degree if we take a pragmatic approach.

3. **"Reduced listing programs."** This quote highlights one of the efficiency gains that comes from the pragmatic approach. It is worth pointing out that this efficiency stems from 2 contributing factors. First is the metadata driven automation to pull elements from a global library for reuse. The other is that the way the ML models can be built to look for atypical patterns, one model can detect data conditions that would otherwise require multiple exception listings to surface.
4. **"Mitigated risk faster!"** This quote exemplifies why adopting AI/ML is important to pursue the evolution of the Clinical Data Management role to the Clinical Data Science role as described by SCDM in The Evolution of Clinical Data Management into Clinical Data Science [2].
5. **"Accelerated data cleaning cycle times."** I call your attention to the fact that this quote was stated in past tense. In other words, this reviewer is acknowledging that they have already seen a reduction in data cleaning cycle times.

CONCLUSION

There is strong precedent that can lead to the inference that the life sciences industry will be a laggard in implementing AI/ML to support clinical data review. However, the extreme hype and the visible utilization of it in so many aspects of our daily lives, makes it possible to be optimistic that this will not be the case. We have seen instances in which inflated unrealistic expectations have led to implementations being disappointing. In contrast, use of a pragmatic approach that builds upon legacy processes, deploying AI/ML to complement and extend capabilities can be well accepted and successful.

In light of the poor history of the life sciences industry adopting innovations, the resistance of pharmaceutical companies to allow their data to be used to train models that could be used by their competitors, and the potential value of using AI/ML to expedite clinical trial data processing; we are interested in participating in open forums to develop models for mutual benefit.

REFERENCES

1. 2021 "Good Machine Learning Practice for Medical Device Development: Guiding Principles" U.S. Food and Drug Administration (FDA), Health Canada, and the United Kingdom's Medicines and Healthcare products Regulatory Agency (MHRA)

2. SCDM, September 12, 2022 "The Evolution of Clinical Data Management into Clinical Data Science, An SCDM Position Paper on how to create a Clinical Data Science Organization, Your path toward Clinical Data Science"

3. Everett Rogers, Diffusion of Innovations, 1962, Simon and Schuster

4. Geoffrey Moore, Crossing the Chasm, 1991, Harper Business Essentials

5. Gartner Research "Hype Cycle for Healthcare Data, Analytics and AI, 2023", July 24, 2023

6. Clinical Data Interchange Standards Consortium https://www.cdisc.org/

RECOMMENDED READING

- Stokman P, et al. Risk-based Quality Management in CDM. Journal of the Society for Clinical Data Management. 2020; 1(1): 1, pp. 1–8. DOI: https://doi.org/10.47912/jscdm.20
- eClinical Solutions, "From Complexity to Clarity: Harnessing the Power of AI/ML and Risk-informed Strategies to Streamline Clinical Data Management", 2023, https://www.eclinicalsol.com/ebooks/harnessing-the-power-of-ai-ml-and-risk-informed-strategies-to-streamline-clinical-data-management/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Robert Musterer
Company: eClinical Solutions, LLC
Email: RMusterer@eclinicalsol.com