Paper ML10

Smart, Digital Analysis & Reporting (A&R) Environments: They Are Upon US

Christopher J. Colangelo
Pfizer Pharmaceuticals
Groton, CT, USA

Abstract

Visions abound that leverage Artificial Intelligence (AI) applications to automate “end-to-end” within & across clinical trials. We at Pfizer share such a vision. We are dividing & conquering this vision segment-by-segment. My segment applies & imbeds computerization—AI/ML, automation, digitalization—within the context of scaled, standardized Statistical Analysis & Reporting (A&R) platforms.

This presentation will highlight the vision for such computerization, the benefits from it, the progress we have made, & what remains. We seek to inform, challenge, & motivate others to join this journey to transform A&R generation from primarily a stat programming discipline to one that leverages informatics and AI/ML to instead enable focus to be on the analytics needed to derive insights.

Introduction

I am in my twentieth (20th) year contributing to the Biopharma industry. Over seventeen (17) of them have been in a capacity where I was/am expected to strategize and drive change. Change—improvement—efforts are seemingly endemic to the industry.

Change outcomes, however—the foundation for assessing improvement, from my view, tend not to pass the “eye test.” I still see a virtual cornucopia of opportunity—opportunity to: ease unnecessary dependence on manual effort; systematize today’s tribal knowledge; exploit applications of state-of-the-art computer and information sciences; re-craft subject-matter expert roles/responsibilities to maximize their unique value; eliminate whitespace within—and across—process flows; and more.

Having experienced the industry buzz from the advent of industry standards consortiums and modernized technologies, I can say that I believe today’s optimistic buzz is louder. And possibly more-desperate. This buzz is correlated with optimism for “going digital”—particularly with novel applications of artificial intelligence, and, more specifically, machine learning (AI/ML).

At Pfizer, we are certainly part of that buzzing chorus. But, more importantly, we are investing in & focused on creation of novel AI/ML applications—and are being rewarded by promising results. We are building internal expertise across the business and IT sides of the development lifecycle; leveraging competitive “hack-a-thons” to empirically assess capability and fit, with potential vendor partners; and applying a vision—functionally prototype—assess model to de-risk decisions. Across Global Product Development, we are starting to create the change we seek.

Limiting the scope of discussion, this paper will focus on our pursuits of AI/ML within the Statistical Programming and Analysis component of Global Biometrics & Data Management. We aspire to imbed various AI/ML applications within a business architecture designed to yield obvious change to how data and analytic results are generated, as well as what expertise next-generation analytic experts will need.

SCEs in Context

Historically branded Statistical Controlled Environments (SCEs), my industry intelligence points to a pivot towards broader-scoped, more valuable, and change-yielding “Smart, Digital Analysis & Reporting (A&R) Ecosystems.” Let me take a step back to clarify. SCEs were, in essence, the core computational capabilities that enabled controlled generation and storage of both the results from statistical analysis as well as the data that were analyzed by them. Naturally, compliance to 21 CRF Part 11 became a key requirement for validated SCEs.

For context, SCEs are predominantly positioned one function, albeit a core one: Statistics or Biometrics. Predominantly in that other functions can use SCEs to review outputs, access stored content, or to import/export information. Other functions may house their own flavor of an SCE. But the main user-base of an SCE is comprised of statistical computing experts that generate the data and results used in submissions, commercialization efforts, and so on.

Back in the day, an IBM mainframe could qualify. Likewise, could PC-SAS and a controlled LAN, with customized version control, of course. Several technology providers chased the SCE market—creating products intended to span multiple companies. Likewise, consultancies filling the business integrator role have tried. Still, today, there is no industry-satisfying, gold-standard SCE.

Information sharing from a myriad of semi-formal and informal conversations that permeate the Biopharma industry highlights the need to address this gap. Thought has evolved on this front—leading to updated strategies for cracking this nut. The market will bear them out. But, in my conversations, there is a popular capability commonly mentioned nowadays: leveraging AI/ML to change the process and cost of generating results.

Current State Overview and Opportunities

A relevant, high-level conceptual overview of the current state of operations for a subject-matter expert (SME) generating statistical results and the data that are input to them can be summarized, as follows:

- Manual "judgment" is necessary to, for example:
    - Determine appropriate data standards for the given project
    - Determine appropriate analysis standards for the given project
    - Determine the detail to add in order to turn standards into study specifications
    - Determine what functionality is needed if existing re-use code is deemed invalid for use
    - Determine if best to custom code from scratch or augment existing code as a "head start"

- Manual "memorization" is necessary to, for example:
    - Recall study, integration, or asset specific details
    - Recall milestones, timelines, tasks to complete, etc.
    - Retain details about a given study protocol, integration, or submission

- Manual "legwork" is necessary to, for example:
    - Select appropriate data standards for the given project
    - Select appropriate analysis standards for the given project
    - Verify detail was correctly added to create specifications that uniquely support the project
    - Select and verifying all specification-identified re-use code is valid
    - If no usable re-use code, custom code or augment existing code used as a "head start"
    - Quality check generated data (e.g., SDTM, ADaM)
    - Record flagged data points/results, then initiating clarification or fixing
    - Quality check generated stat analysis results
    - Record flagged data points/results, then initiating clarification or fixing

Of course, further detail, additional examples, and much commentary can be added in this section, but for brevity and focus, I have limited our conversation to the above. The above scope provides rich territory for our discussion of how AI can be imbedded into technology and processes that serve the generation of statistical data and analyses.

AI Solution Thumbnails

Create databases of elements stored as controlled content—like today's documents

> If documents are the source protocols or SAPs used as reference, "scrape" them to "digitize," then database the results to store their content as electronic elements.
>
> If such content is already stored as elements in the database, one is ready to edit, manipulate, or create new content by directly editing the database—bypassing "document scraping."

Select appropriate standards

Launch our standards selection AI application. It will ingest necessary digital content from the study's protocol, the study's SAP, the library of standards, and other necessary information. Once ingested, this information will feed an algorithm designed to yield output stating which data and analysis standards it determines to be appropriate for the given study.

There will then be a SME review phase, where one's judgment will review, adjust, then approve which standards will be applied to the study. Adjustments will also provide corrective feedback to the algorithm, reducing the likelihood of subsequent incorrect determinations. This feedback loop will remain over time—continually sharpening the validity of the determinations.

Configure standards into specs

Once the appropriate standards have been selected and approved, the digital content used by the algorithm to determine the applicable standards will be re-used to determine what configurations are needed by each standard to evolve into a study specification. In essence, a standard is a template of sorts, with configuration akin to completing the editable sections of that template. Study specifications, in essence, serve as the "requirements" defining the data and statistical analysis results to be delivered by SIGMA.

As with the standards, there will then be a SME review phase, where one's judgment will review, adjust, then approve which standards will be applied to the study. Adjustments will also provide corrective feedback to the algorithm, reducing the likelihood of subsequent incorrect determinations. This feedback look will remain over time—continually sharpening the validity of the determinations.

Identify re-use code

Information in the data and analysis standards, in addition to the configured specifications, provide a roadmap to identify the applicable standard, re-usable code for the output to be generated. However, re-use code can become out of synch with its corresponding standard—either by the standard being updated without also its code, or vice versa.

This AI capability would suggest which code is appropriate for the specification-defined need, verify if it is in synch with its standard, return a usability overview of the code—key functionality, scope of validation, etc. If no re-use code is directly applicable, the capability would suggest what code is nearest in applicability, as a potential starting point for tweaking its code to create a new re-use module. If no re-use code is a candidate to serve a head start for coding, then it would suggest from-scratch coding.

Configure re-use code

Re-use code is almost always designed with parameters. Parameters allow information (often options or user-selected definitions) to be externally provided to the code and be seamlessly incorporated and processed as part of the code during runtime. Parameters provide a way to "read" pre-defined values into the code, without actually opening the code file—thus without exposing already validated and approved code to the possibility of error. Feeding parameter values to the code during runtime can be viewed as configuring the code's parameters.

This AI capability would combine specifications, parameters within identified code files, and other needed contextual information to determine the values to use to configure corresponding parameters, feed them to the code, and auto-execute that code file to generate its data or stat analysis output(s).

QC of stat output

> In progress at the time of authoring this document is an effort to deliver a custom capability to AI-identify and compare like information across statistical analysis output tables. There is demonstrable, but moderate progress to date. There is also a roadmap of future capabilities. Thus far, the two limiting factors for progress have been the myriad of input considerations one mentally references to just concordance/discordance when comparing data points across tables as well as the volume of valid training data available.
>
> Looking ahead, we aspire to continue enhancing the table-to-table compares of data points. Moreover, we look to refine metadata "fingerprinting" and contextualization to the point that we can compare data points across databases or data tables.

The fundamental backbone: Databasing key information to feed multiple AI/ML apps, aid automations, and link results upstream

- Example documents
  - Study Protocols
  - Study SAP
  - Integrated AP (IAP)

- Targeted information
  - Schedule of activities
  - Endpoints
  - Objectives
  - Inclusion/Exclusion criteria
  - Study design description
  - Same sizes
  - Key claims needing empirical evidence
  - Analytic methods
  - Analysis populations (analysis sets)

Risks, Limitations, Inertia for AI/ML

Though fraught with promise, cajoling organizations to trust AI/ML-generated outputs will require patience and depend upon success when organizational change management (OCM) challenges arise. In fact, in my experience, changes to technology infrastructure itself creates concern among end-users in the Statistics and Statistical programming community. I have worked to influence acceptance among user-communities when automating process. But the greatest inertia of concern likely lies in nurturing trust for AI/ML capabilities.

Inherently, Statisticians and Statistical Programmers are applied mathematicians who use a keyboard instead of a pen. Generalizing, we are more interested in delivering data and analytic results than we are in developing software. Simply put, we want systems and capabilities we can trust, so we can focus on data contextualization, coding appropriate methods, ensuring validity of our outputs, and contributing data-driven insights to help inform safety and efficacy. From this perspective, it seems more straightforward how changing our underlying computational systems can introduce angst. But there are still manual processes and decision-making that can help end-users gain confidence in their underlying computing engines. Likewise, there are quality processes to document testing of the technology.

Considering that perspective, we now consider the end-user pause when facing adoption of automated processes—static processes as well as ones that follow rules-based decision branches of logic; process that manipulate data as well as ones that do not. Again, a manual review of outputs by SMEs can help foster trust, and provide a feedback loop for enhancements, if needed.

When considering trust of AI-generated outputs—especially ML-generated ones, there is still bound to be a manual review step, but truly quality checking validity of the output takes on more of a black-box form. Given that (often) complex algorithms process information fed to them, gaining confidence through manual review can remain difficult.

Fortunately, in SIGMA, we will predominantly use AI/ML to output results to be used as a *workflow expediency* more than as a generator of novel data-driven *insights*. So, while AI outputs that crunch big data to yield novel insights may challenge SME ability to determine their correctness, SIGMA's use-cases for, and positioning of, AI outputs enable SME reviews that allows dispositions with certainty. This previous point is an advantage we intend to exploit during OCM and user adoptions.

Conclusion

SCEs are evolving, expanding in scope, and imbedding complex, modern capabilities like automation, AI, and ML. By initially focusing on the core computing platforms used to generate statistical data and results, it is clear to see several applications of AI/ML that will affect the user experience, the burden, and the overall operations of stat programmers.

Traditional SCEs mainly addressed the activities and artifacts related to programming, generating outputs, persisting them, and retrospectively documenting what was generated. Henceforth, SCE's will more resemble informatics & analytics ecosystems—comprised of many components, services, applications, etc., out growing the creation of a large product.

Now expanding scope, for instance, to include processes and delivery of necessary documentation used as input into coding, we surface novel application of digital and AI/ML capabilities that, in sum, would transform how the role of "statistical programmer" is executed. The several examples described above are certainly viable, but are not yet alive inside Pfizer.

What is also viable, and not strictly within the scope of SIGMA, is the novel applications of AI/ML to advance scientific and analytical insights. These present initial focus on safety and modest judgments. Our view of organizational change management is to reinforce trust over time. We have already rolled-out multiple RPA/automation and AI solutions. Taking care to start modest, mushroom over time, and let trust evolve alongside. But, I recall, something I once shared: the OCM comes in the form of a carrot—a demonstrably useful solution. And AI will help us provide that in SIGMA.

Select Road-mapped Ideas/Aspirations

AI-generate code

> Leverage study specifications defining data or statistical analyses; a full open-source language like R (likely externally accessible); the R code files used to generate data or analysis outputs; and the actual outputs generated by the executed code—leverage to begin to train an AI-capability to generate more accurate code over time when fed the specifications and the full R language.

AI-identify safety risks … then generate insights to assess them

> Leverage protocol information—including disease state, mechanism of action of the treatment arms in a trial, schedule/timing of activities, actual procedures taking place, and additional contextual information—leverage to predict risks and their likelihoods.
>
> Those safety risks—say AEs or SAEs, ranked by predicted likelihood of occurrence, can all or only some (based on some business rule) be assessed during the conduct of the study. Clever application of RPA can navigate the period where researchers must remain "blinded."

Based on results of those assessments, correlation profiles can be derived, sharing further insights into the AEs/SAEs—possibly even yield causality.

AI-identify data distributions and vet assumptions for applying statistical methods

Helpful in pursuit of statistical analytics is empirically deriving the data distribution(s) observed in one's study. While this does not require AI, it can benefit from safe access to unblinded data while researchers must remain blinded. That allows for refreshing the distribution over time, if desired. Else assessing the distribution at discrete time-points can happen.

Leveraging the SAP and specifications for statistical analyses to be conducted, as well as curating the assumptions corresponding to the methods to be used, will provide results of assessments of valid use of those methods, on a per analysis bases.

AI-identified sub-groups or sub-group influencing characteristics

Sub-group analysis is routine when assessing results of clinical trials. In essence, they use any of a myriad of characteristics to define groups of patients, then analyze results within those groups, with the hope of identifying some result that enlightens. Enlightenment may be mere explanation for a differential result, or it may be something leading to a refined study design—restricting the study population, and leading to a statistically significant result for that narrower population.

Contact Information

Christopher J. Colangelo

Pfizer Pharmaceuticals

17111 Patriot Way

West Greenwich, RI 02817

317.373.3663

Christopher.Colangelo@pfizer.com