Paper DH04

A Programmer's Point of View of Current and Future PD Detection Tools

Meng Li, AstraZeneca, US

ABSTRACT

Identifying and programming for Protocol Deviations (PDs) and Important Protocol Deviations (IPDs) are critical for running clinical trials. This process is a collaboration between multiple functions. Detecting IPDs can require complex programming. There are even times when this output needs expert review from a medical team. This paper will explore a current and a possible future approach to identify IPDs, as well as their advantages and disadvantages.

The current approach uses traditional tools/macros created by programming team to identify IPDs. This approach significantly decreases human effort and increases accuracy in some PD categories. However, there are still some difficulties in implementing this approach.

A future approach would take advantage of Artificial Intelligence (AI)/-Machine Learning (ML). AI/ML driven tools can further minimize manual involvement and increase accuracy in more PD categories. A downside to this is the time and resources needed to implement the technology.

INTRODUCTION

A clinical study protocol is a document describing the rules of how a clinical trial will be conducted in every aspect. “A protocol deviation is any change, divergence, or departure from the study design or procedures defined in the protocol.” [1] Protocol deviations (PDs), especially important deviations (IPDs), may negatively impact the participants and/or quality of the trial results. Despite the efforts to minimize them, the reality is PDs do happen. Therefore, it is important to define, detect, analyze, report, and document them. This process requires tremendous collaboration between multiple functions in clinical trial team.

Manual detection and programmatic detection are the two widely used methods to identify PDs. Statistical programming team plays a key role in programmatic detection. Even though these two methods are widely used today, there are many potential advantages of using artificial intelligence (AI) and/or machine learning (ML) to detect PDs and IPDs. The unique features of PD can make it a good candidate to apply AI/-ML.

BACKGROUND INFORMATION

As described above, several study team roles have responsibilities in the process from defining to documenting PDs. Figure 1 provides an overview of this process. Statistical programmer’s responsibilities are mainly in the ‘identify and collect’ [2] phase. To be more specific, statistical programmers are the main players in programmatic detection.

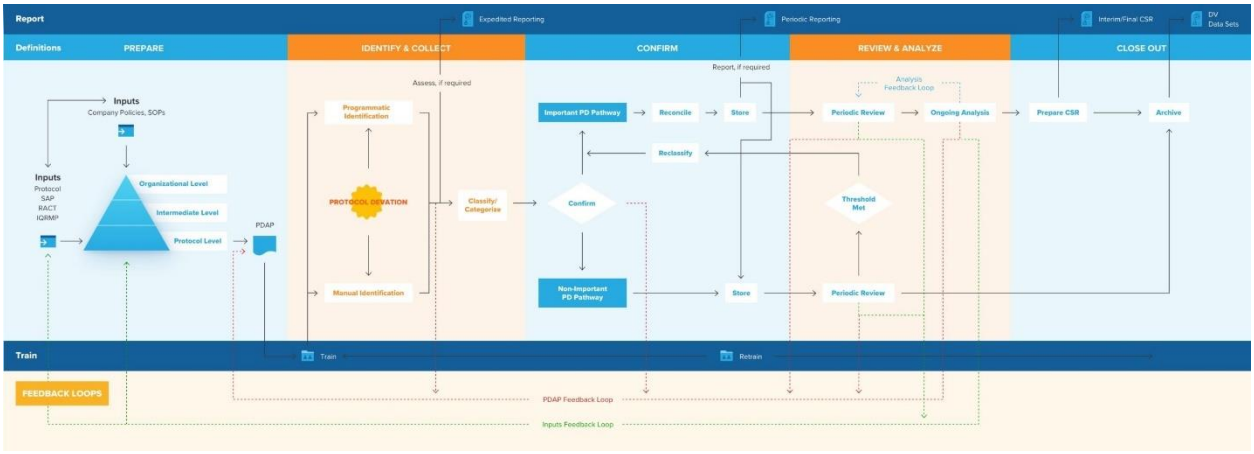

Figure 1 Protocol deviation map [2]

PD DETECTION METHODS AND PROCESS

PD detection methods are usually categorized into manual or programmatic. Manual detection relies on human interpretation. Programmatic detection is based on data captured in a database using an electronic/computerized process or program. For example, to identify investigational product (IP) storage temperature is not set at the manufacturer's requirements, the detection should be a primary focus for CRAs during on-site visits. Obviously, the identification of those elements are not programmable in most trial designs. A common programmatic detection example is identifying informed concerned form not signed before project specific assessments or procedures commence, because case report form (CRF) collects all required elements to identify this PD.

STATISTICAL PROGRAMMER'S ROLE IN THE PROCESS OF PD IDENTIFICATION AND HANDLING

Statistical programming team usually works with other core study team members to complete PD management plan, and specify an identification method for each PD. They may be responsible to develop and generate the listings if the detection method is programmatic, and/or prepare the data for data visualization tool, depending on the study's PD management plan.

ADVANTAGES AND ISSUES WITH CURRENT PROGRAMMATIC APPROACH

TRADITIONAL WAYS

It is no doubt that there are many tools to support SDTM and ADaM activities. However, because PD related task only covers a small portion of statistical programmer's responsibility list, not many tools are developed to automate this process. In reality, one common method is borrow and modify existing programs that developed for previous studies. Some automation

tools to support PD detection are statistical programming language based macros, which requires user to input parameter values. Some PD listings generated by the statistical programming team are not final PD result due to the complexity of the PD, and it requires a medical team to further manual review.

ADVANTAGES AND ISSUES

These commonly used programmatic methods decrease human effort by reducing the amount of manual review and programming from scratch. It also increases accuracy by minimizing human errors.

There are also issues with these approaches.

First of all, if PD plans require a programming team to generate PD listing, the first thing to do is to generate a specification. This process usually involves other functions, such as statistician and medical team member. Statistician and/or medical team may support programmers to provide the logic behind the PD programming. Medical team usually provides medical terms. For example, when PD is to detect a medication used for one or more medical conditions within a certain period, the drug code and the dictionary terms are from the medical team. When more parties are involved, it is inevitably that back and forth discussion happens more. Not to mention that these are additional work for other functional teams.

Secondly, these methods, no matter it is 'borrowing' programs from other studies or automated tools, still require customization in order to be used for a given study. When 'borrowing' from other studies, programmers need to modify the programs here and there. For example, replace the input dataset name in the program, or add a block of code for a PD that does not exist in previous study's program, which is borrowed from. Some automated tool can be seen as macros. When using these tools, the user needs to enter a set of parameters. This process can be more complicated than it seems, as it usually requires the user to use an Excel format spreadsheet to store these parameters. Then the tool imports the spreadsheet and generate a report based on the information in the spreadsheet. These Excel sheets are sometimes macro-enabled using VBA code. Things could go wrong because the tool is not a real automated tool, and a mistake in the spreadsheet may cause run-time errors in execution.

Last but not the least, some PDs are complex, and it is challenging to use programs alone to check. It requires combined programmatic and manual detection method. For example, PD to detect the XXX dose not maintained at the same level from Visit 2 to Visit 3 for reasons other than safety. To detect this PD, the program needs several input data, such as concomitant medication, visit, and adverse event. What makes identifying this PD a challenge is the condition 'other than safety'. The therapy reason is collected on CRF, however, when the form allows site to enter free text in 'Other, specify' field, it becomes a challenge for using programmatic method alone. Theoretically, programs cannot cover all cases in free text field. A programmatic combined with manual detection is necessary in this case.

FUTURE APPROACH

The emerging trend of AI/-ML offers great opportunities to accelerate clinical trials and lower cost. For example, AI technology is used to identify subgroups by analyzing electronic health records and genotype data. It provides insights on how specific patients would respond to a medication by analyzing clinical characteristics and disease comorbidities [5]. As can be seen from this example, AI/-ML model provides a pattern of clinical characteristics and disease comorbidities. Finding patterns and dependencies from data and apply them to new data is a key function of AI. Therefore, it also can be used in identifying PDs.

BRIEF INTRODUCTION OF AI/-ML

AI has varied definition. A simple version is that AI 'is a field, which combines computer science and robust datasets, to enable problem-solving.'[6] ML is a sub and main field of AI. ML is the ability for a machine to think and learn.

The fundamental concept of ML is building a model by analyzing the data to extract patterns and dependencies and then use these rules to gain understanding of a phenomenon or make predictions. ML models give the best result based on probabilities.

WHY CHOOSE AI/-ML FOR PD DETECTION

PD detection is a good candidate to apply AI/-ML technologies due to its unique feature. In clinical trials, each PD has standalone logic. There is no need to refer another PD when identifying a given PD. Comparing with SDTM, or ADaM logic, PD rule is simpler and more straightforward. The model should be relatively simple, and the result is easier to verify.

The data used to detect PD are usually CRF data. Cleaned CRF data are high quality data, and they are already labeled, categorized and grouped. It's a good fit for supervised learning, which compares to unsupervised learning, will have more accuracy in this case.

ADVANTAGES

AL/-ML tools can further minimize human effort, and accelerate the process.

The AL/-ML program is able to use the pattern and dependencies to identify PD in source data. It is a true automation. Statistical programmers are not necessarily involved in the process once the algorithm is finalized. Medical team or other functions are able to generate a report by a single click without requesting the programming team to run it for them. It significantly minimizes human efforts. It also further increases accuracy as there is nearly no human effort in the detection process once the model is finalized. The power of computerized calculation is able to detect PD in real-time. It can help to minimize the negative impact of PD to participant and clinical trial result. AL/-ML algorithm can also be trained to identify early warning signs of potential PD, such as frequent rescheduling of visits.

Current macro based tools limit PD programmatic check to only use tabulated data. The source data for AI/-ML could be documents, pictures or files. Some PDs, for example, identifying

dispensing accountability documentation is incorrect or out-of-date, can be detected by AI/ML even the source data is not a dataset. More PDs, which are usually manually detected, has the potential to be programmatically identified.

Lastly, the rule and logic of programmatic PD detection are similar to the ones of SDTM or ADaM but often quite less complicated. Using AI/-ML in PD detection can be seen as a testing for it to be eventually used in broader fields in statistical programming.

CHALLENGES

Similar to other areas that AI/-ML technologies are introduced, there are challenges. On the scientific side, it requires a large amount of data to train and test the model. If the model was trained in one situation, it might perform badly in another. It is important that care needs to be taken before it is used widely. On the financial side, a sponsor needs to leverage resources to enable AI/-ML tools development. It takes time, and requires investment. On the operational side, using a new way to detect PD, requires some level of changes to existing processes. For example, changes to operational procedures and staff training. Moreover, how AI/-ML regulatory impacts its usage in pharmaceutical industry is uncertain. Whether some type of AI decision making will be banned is unknown.

CONCLUSION

The paper is not designed to focus on the actual algorithm, but it emphasizes on the reason to use AI/-ML in PD detection. The purpose of this paper is to highlight the current PD detection issues, and suggest a possible PD detection way for both pharmaceutical companies and technology providers to address in the future. Applying AI/-ML in PD detection process is an area AI/-ML can shine in statistical programming field.

REFERENCES

1. ICH GCP https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-5.pdf
2. TransCelerate Protocol Deviation Map https://www.transceleratebiopharmainc.com/wp-content/uploads/2022/08/PD-Process-Map_081222-scaled.jpg
3. Laura Galuchie1 · Catherine Stewart, PhD2 · Frank Meloni, PhD3, Protocol Deviations: A Holistic Approach from Defining to Reporting *Therapeutic Innovation & Regulatory Science (2021) 55:733–742,* https://doi.org/10.1007/s43441-021-00269-w
4. Elements of AI https://course.elementsofai.com/
5. Urtė Fultinavičiūtė, AI benefits in patient identification and clinical trial recruitment has challenges in sight, https://www.clinicaltrialsarena.com/features/ai-clinical-trial-recruitment/
6. What is artificial intelligence, https://www.ibm.com/topics/artificial-intelligence

ACKNOWLEDGEMENT

The author is very grateful and thankful to Greg Silver, Statistical Programming Director at Late R &I AstraZeneca, for his support, advices, guidance and comments throughout all aspects of this paper.

The author is also very thankful to Daniel Schiff, Statistical Programmer II at Late R &I AstraZeneca, for his expertise, assistance and comments on this paper.