Paper DV09

Staying Alert to Data Quality: A Shiny App to Visualise the Electronic Patient Reported Outcome (ePRO) Data

Wenyi Liu, Roche, Mississauga, Canada

ABSTRACT

The use of patient reported outcome (PRO) data in clinical trials has soared in recent years, no longer limited to being just a complement to clinical evidence. More clinical trials are directly using PRO endpoints. Specifically, electronic patient reported outcome (ePRO) data has great value in efficiency and cost-saving. However, due to the 'patient-reported' feature of ePRO data, data quality – especially missing data – has been the most challenging obstacle for clinical trials to adopt this strategy. To track the ePRO data collection status and find signals of factors that cause data quality issues, a Shiny app was created to show real-time alerts. In addition, several other visualizations were also created to show the composition of the endpoint as well as disease progression so that prompt suggestions can be sent to the study team to evaluate the impact on the study.

INTRODUCTION

The U.S. Food and Drug Administration (FDA) defines patient-reported outcome (PRO) as any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else [1]. In clinical trials, PRO can refer to subjective feelings that cannot be observed or quantified (e.g., *Feeling better* or *the pain level comparing to yesterday*), or objective signs and symptoms (e.g., *cough* or *be able to drive*). PRO data can be collected in multiple ways. The most used methods are paper-based, computer-based, telephone-based and electronic-based assessments [1]. Among them, the PRO data collected via electronic devices or mobile apps is called electronic patient-reported outcomes (ePRO) data. Since the ePRO data are usually collected on daily basis or more frequent, it is sometimes used interchangeably with the term *eDiary*.

INCREASED RECOGNITION OF PROS IN CLINICAL TRIALS

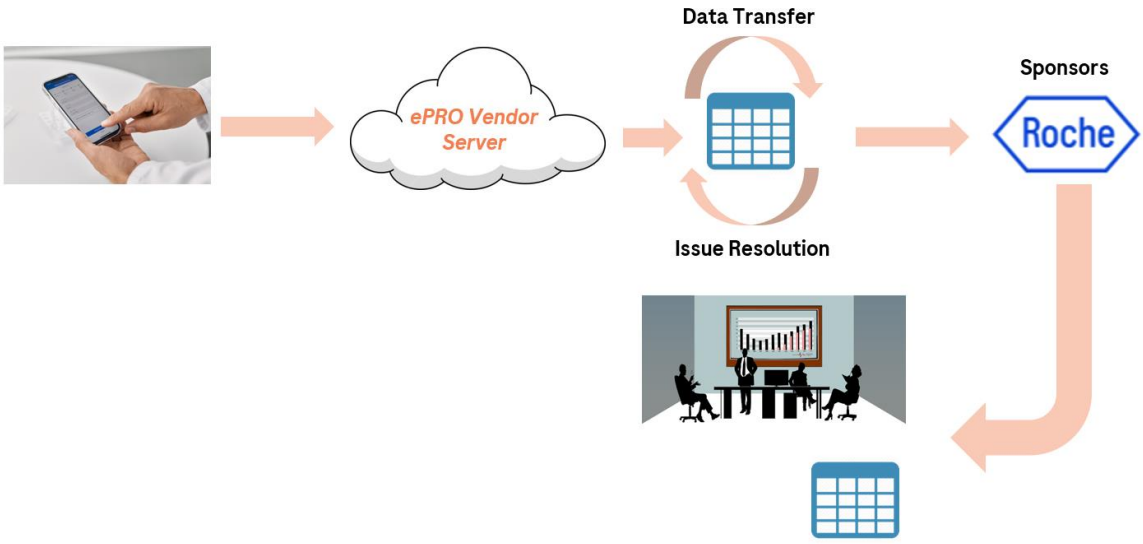
In recent years, patient-reported outcome has played a critical role in clinical trials to measure the effects of treatment from the perspective of a patient [2]. Comparing to clinical outcomes as measured by physicians, patient-reported outcomes advocate more on patients' recovery from the disease state and quality of life. A typical example during the pandemic is when measuring the recovery from COVID, one can use "COVID tested negative" or "COVID symptom-free". The former one is an objective clinical outcome which can be quantified by scientific methods, such as a PCR test, but the latter one can only be reported by patients when they no longer feel short of breath or no more coughs. In real practice, the "symptom-free" is the true recovery state. In addition, patient-reported outcomes also provide supplemental evidence for health authorities, public policy makers and payers to evaluate the cost and benefit of the treatment.

With the increased recognition of PROs in clinical trials, more studies have adopted PRO endpoints. A Roche Influenza study evaluates time to alleviation of influenza signs and symptoms maintained for 24 hours as efficacy endpoint, of which the influenza signs and symptoms were collected by Canadian Acute Respiratory Illness and Flu Scale (CARIF) Questionnaire. Another Roche study evaluates the efficacy of treatment for Ulcerative Colitis bowel movement signs and symptoms at Week 10 as assessed by the Ulcerative Colitis Patient-Reported Outcome Signs and Symptoms (UC-PRO/SS) questionnaire.

ADVANTAGES AND CHALLENGES

If electronic method was used to collect PRO data, patients are usually given an electronic device or asked to download a mobile app. The pre-defined questionnaire will be activated upon enrollment. Patients will receive reminder to answer the questionnaire in the device or app. The answer should reflect patients' true state at the time of measurement. Once submitted, the PRO data should be uploaded and stored at the ePRO vendor's server. The ePRO vendor shall process the data based on the agreed specification and transfer the data to sponsor for further analysis. Once sponsor receives the ePRO data, they need to process data and run analysis according to Statistical Analysis Plan (SAP).

Figure 1:



ADVANTAGES

The electronic method allow fast and real-time measurement and data collection as patients can answer questions any time. Especially in large phase 3 trials which usually recruit hundreds or thousands of patients, the electronic method can save a huge number of resources. Visiting doctor's office is always considered a burden to patients who participated in clinical trials, but the electronic method allows them to complete the assessment at home or during work break. With all stating above, the electronic method ultimately enables more data collection. The more data we collect, the more powered results the trial can achieve.

CHALLENGES

Like every coin has two sides, the electronic method can also cause issues. In addition, the "patient-reported" feature of PROs also faces its challenges.

Most patients do not have medical knowledge. They may not fully understand what the question is and cannot provide the accurate results. Some ePROs require patients to conduct some self-measurement, for example, to measure their body temperature. If they cannot properly use the thermometer or if they confuse the Celsius and Fahrenheit, the collected data can also be wrong.

A more common and severe issue is missing data. The causes of missing data are various. Even though patients will be reminded to answer the questions on time, there are still way more chances that a questionnaire can be missed. Just like we may encounter connection issue with our mobile phone, the electronic device can also face technical problems such as freezing window, loss of connection and out of battery. The missing data have a huge impact on the accuracy of the results.

Even though it is considered an advantage of using electronic method to collect PRO data real-time, it also means the data was collected as is. When data transfer to sponsors and data issues are found, it is already too late to query since patients usually would not keep a copy of their answers.

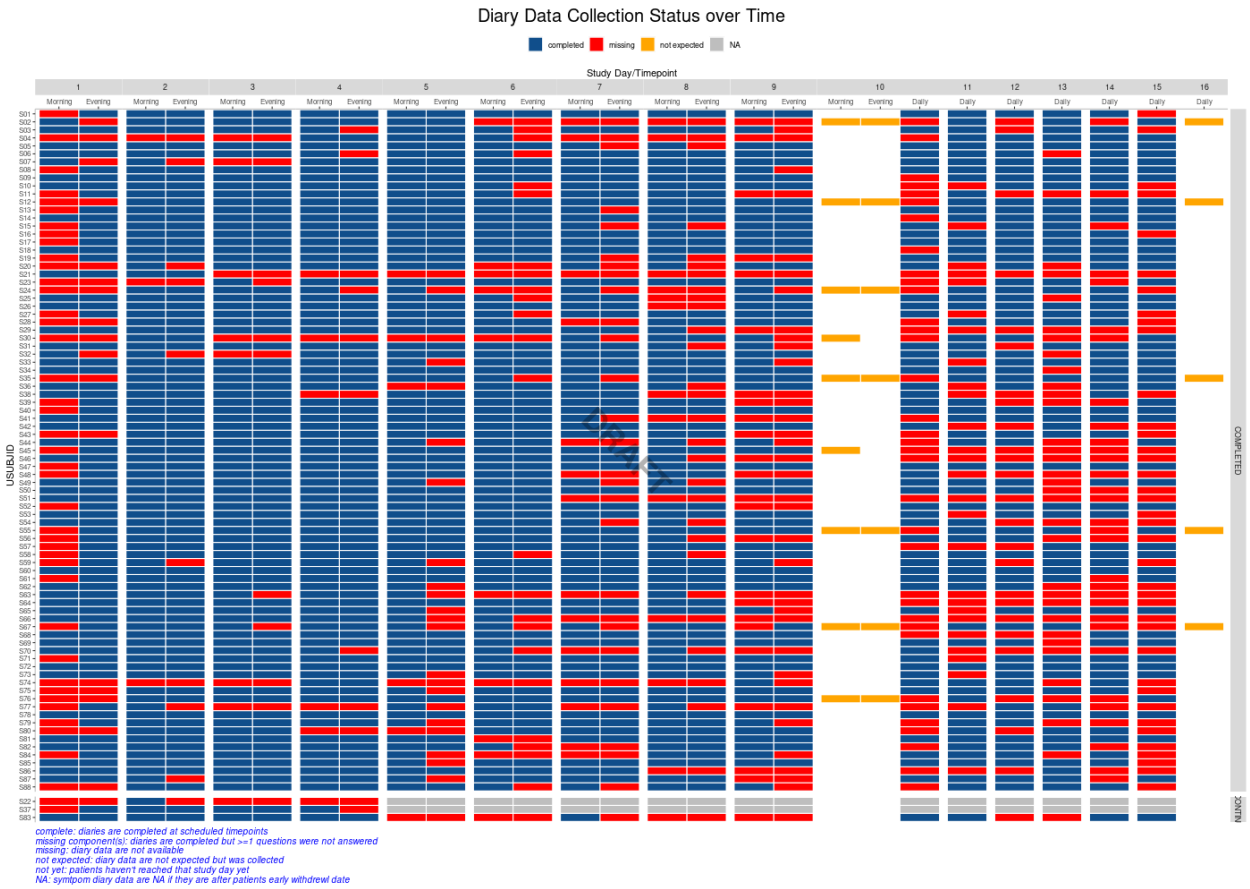
THE SHINY APP TO TRACK AND DETECT DATA ISSUES

The Data Scientist's work is to utilize the advantages of ePRO while overcoming the challenges in a clinical study. Therefore, a Shiny app was created to track and detect data issues when study was still ongoing with the objective that issues can be found and actions can be taken at an early stage. Even though data issues are unavoidable for clinical trials, Data Scientists can still do their best to deliver as reliable results as possible. The Shiny app was built based on the Roche-developed TEAL package, which is a framework that leverages the R Shiny package to scale development of Shiny apps [3]. The app comprises of several tabs which serve different purposes. The data used in the demo were randomly selected from an Influenza study and data were de-identified.

DATA TRACKING

The data tracking tab displays a heatmap-like plot which gives an overview of ePRO data collection status, especially the scale of missing data. In this plot, each row represents a patient and each column is a scheduled ePRO timepoint. It uses different colors to show different data collection status, e.g., blue for completed, red for missing etc. Users can easily notice certain types of missing schemes, for example, consecutive missing timepoints (consecutive red tiles) and systematic missing for certain sites (by filtering down site numbers). It also provides the feature to check unexpected entries. During study development, users can also check whether data were transferred on time. If not, a grey tile would be displayed. If there were a large scale of grey tiles, it would be a warning signal whether there is an issue with the data transfer.

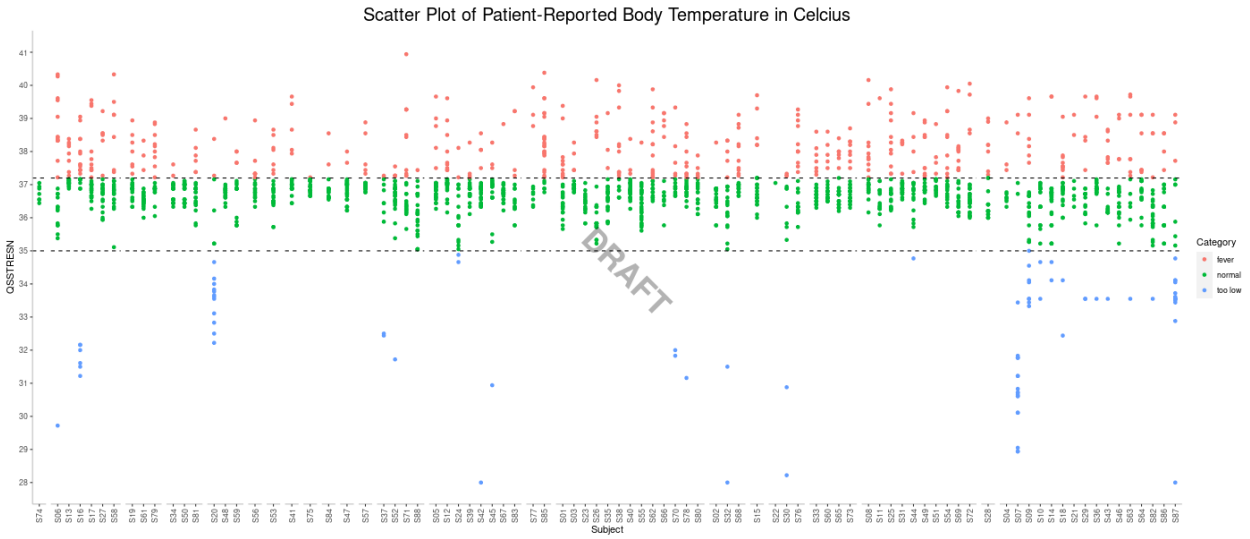
Figure 2:



BODY TEMPERATURE TRACKING

This visualization refers to the challenge that patients lack medical knowledge when completing the ePRO assessment. In the Influenza study, one of the required assessments is to record the body temperature at scheduled timepoints. In this tab, a scatter plot of body temperature by patients was displayed. By reviewing this plot, users can identify a few typical issues: 1) Confusion of temperature unit (Celsius vs. Fahrenheit); 2) Invalid numbers, for example, 28 Celsius degree of body temperature. In addition, by grouping the patients by sites, the plot would tell users which sites have the most issues, thus signaling the team that this site needs a closer monitoring. If these issues continue to come out, team need to consider taking actions on patient training and ePRO device/app UI design improvement.

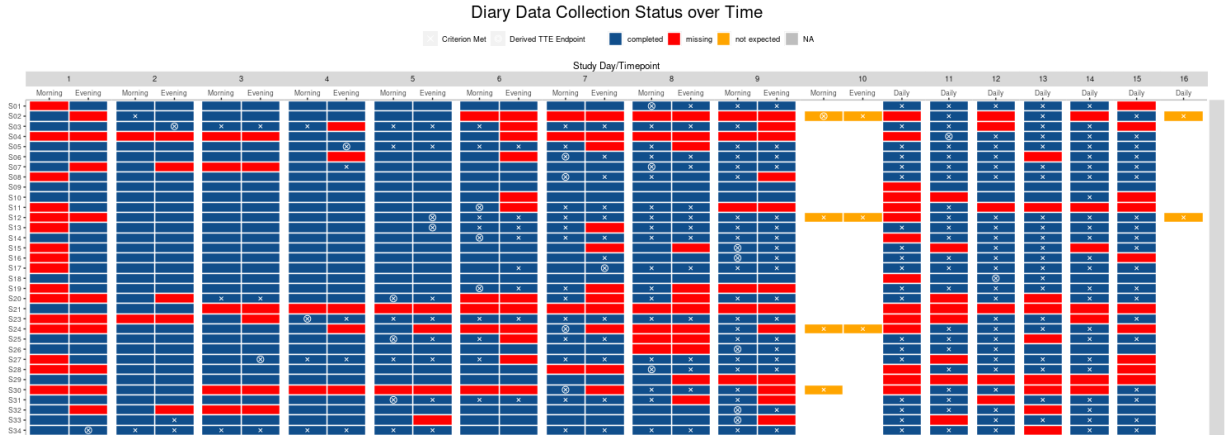
Figure 3:



ENDPOINT CRITERIA MET PROGRESS TRACKING

The efficacy endpoint for the demo study is *Time to alleviation of Influenza signs and symptoms maintained for 24 hours*, which comprises of multiple items in the ePRO assessment. To meet the endpoint, patients need to achieve the pre-defined alleviation criteria and maintain the status for at least three consecutive assessments. The tracking plot in this tab can track the two main objectives. This plot is an enhancement of the data tracking plot with additional markers on the tiles. The “x” mark marks alleviation criteria met at certain timepoint while the “o” mark marks the derived endpoint based on the current algorithm. It provides a dynamic view of patients’ achieving the endpoint and whether the missing assessment (red tile) is a roadblock for patients to achieve it. With more understanding of the data, the team can decide whether it is necessary to implement a more strict or more relaxing missing data handling rule. It also provides a visual check of the derivation code.

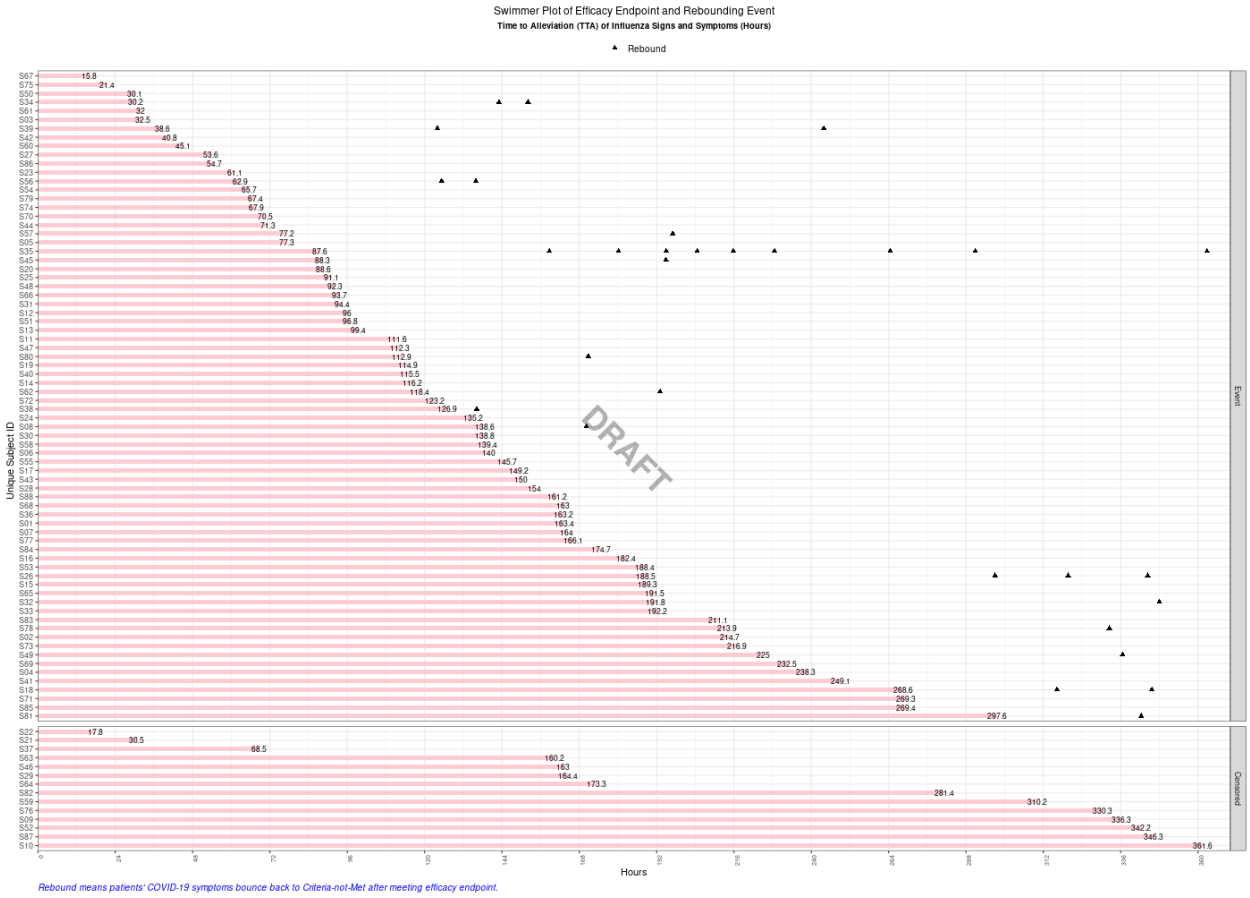
Figure 4:



SYMPTOMS REBOUND TRACKING

For Influenza study or other infectious disease studies, one of the critical scientific concerns is the rebound. Too many rebounds or too soon rebounds may incur questions on the treatment effect. This tab displays a swimlane plot that shows patients’ time to alleviation of Influenza signs and symptoms and any rebounds post endpoint until the end of the scheduled assessment period. If the plot shows too many or too soon rebounds, Scientist and Biostatisticians need to consider whether the endpoint criteria need to be strengthened. It also gives an insight to future studies whether this is a reasonable endpoint.

Figure 5:

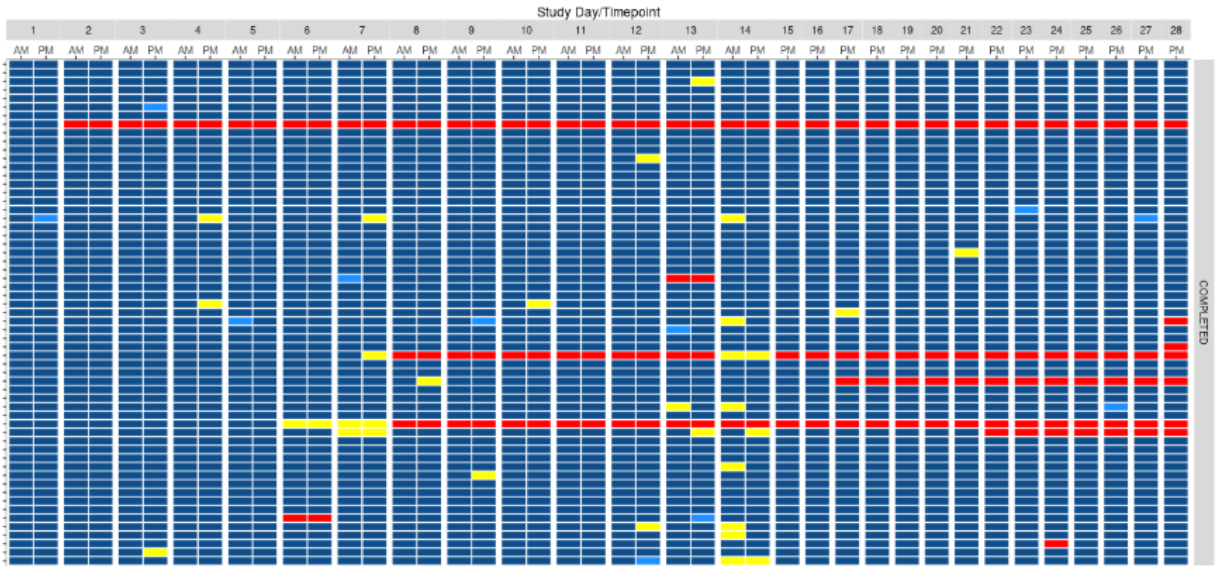


Rebound means patients' COVID-19 symptoms bounce back to Criteria-not-Met after meeting efficacy endpoint.

DISCUSSIONS

The visualizations in this demo were built based on the study needs. Some tabs may not be applicable to other studies while other studies can create visualizations that fit their purposes. However, there is one purpose which is shared, that is to minimize the missing data. It is always expected that the study can have an all-blue data tracking plot, but it can never happen. In some occasional cases, the status shown in Figure 6 may be achieved and this study uses paper-based PRO and patients need to visit office or answer the phone calls from site staff on the scheduled time. For studies that adopt ePRO, the status in Figure 2 seems to be normal. But apparently, the number of patients recruited or the human resources devoted to the two PRO collection methods were quite different. It leaves a question to the study team to evaluate the risk vs. benefit of using ePRO and other methods of PRO.

Figure 6:



For ePRO itself, there are still many aspects that are yet to improve. The goal for ePRO designer is to collect more data and more accurate data. However, it seems a dilemma in current stage. There was a scenario when ePRO data need to be collected twice a day: MORNING (8:00am – 10:59am) and EVENING (5:00pm – 7:59pm). There are two options. The first option is to allow the ePRO system to be open all day so that patients can submit their entries any time during the day, but patients will get notified when it's time to submit data. Using this method, the system can collect more data, but it cannot automatically track the time of assessment and entries. Additionally, it is more likely to get incorrect data since it relies on patients' input. The second option is to open the system only at the pre-defined time window and patients get notified when it's time to assess and submit data. Using this method, the system can automatically track the timestamp and make sure the submitted answers reflect patients' real-time assessment and experience, but apparently, it is easy for patients to miss out the time window.

CONCLUSION

PRO and ePRO are already widely used in clinical trials regardless of the challenges on itself. As Data Scientists, we need to use our expertise to overcome the challenges to the maximum extent and benefit from its designed purposes. The demo of the visualization in this paper is a Shiny app, but the tool is just a medium for realizing the purpose. The truly fundamental part is the idea and how to implement the idea. There is no need to stick to one tool or software. The use of PRO or ePRO is not just the matter of Biostatisticians or Data Managers, it involves partnership with the whole study team and even the external team such as the PRO instruments designer and the ePRO vendors.

REFERENCES

[1] Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. U.S. Food and Drug Administration. 2009. [Content current as of 10/17/2019]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims

[2] Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas*. 2018 Nov 1;9:353-367. doi: 10.2147/PROM.S156279. PMID: 30464666; PMCID: PMC6219423.

[3] Teal: Interactive Exploratory Data Analysis with Shiny Web-Applications. Available from: https://insightsengineering.github.io/teal/main/

CONTACT INFORMATION

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

Wenyi Liu
Hoffmann-La Roche Limited
7070 Mississauga Road
Mississauga, Ontario, L5N 5M8, Canada
Work Phone: +1 437 996 9347
Email: wenyi.liu@roche.com