

Paper DV02

Optimize PRO Data Analysis with Interactive Approach

Tadeusz Lewandowski, F. Hoffmann-La Roche, Ltd., Basel, Switzerland

Xiangyun Wang, Genentech, South San Francisco, USA

Elisabeth Piau-Louis, Genentech, South San Francisco, USA

ABSTRACT

With patient-focused drug development, generating patient-relevant evidence of the treatment clinical benefit is increasingly gaining importance for regulatory, pricing and patient access decision-making. Patient-reported outcomes such as patients' reported treatment tolerability are used in most of registration trials. Yet interpretation of the results is often limited by the numerous outputs that need to be generated to inform a comprehensive understanding of patients' experience with a disease and treatment.

We piloted an innovative approach to optimize delivery and interpretation of PRO data through use of interactive tools. These tools enable easy visualization of the cross-sectional and longitudinal data across multiple dimensions and analytical approaches. Further analyses can easily be self-generated via divers clinical, biomarkers or demographic filters to inform the generalizability of the findings across different subsets of populations. The flexibility and dynamics of interactive tools provide an efficient alternative to standardized static outputs for improved evidence generation for drug risk/benefit assessment.

INTRODUCTION

In oncology trials, the focus is on documenting the quantity of life gained (overall survival), (lack of) tumor growth (Progression Free Survival) and treatment tolerability. Collection of patient-reported outcomes (PROs) such as symptoms severity, functional impact or quality of life, ensures a more comprehensive evaluation of the clinical benefit of a given treatment approach. Because patient-completed questionnaires capture several outcomes, their evaluation is associated with a high volume of outputs. In addition, PROs are often positioned as last secondary or exploratory endpoints within the testing hierarchy.

Companies are striving for faster timelines between database lock to filing of treatment benefit evidence to health authorities, and therefore it is compelling to streamline the analyses and number of outputs to be generated. In this context, analyses for PRO endpoints are either de-prioritized for endpoints that are higher in the testing hierarchy or fully outsourced to external contract research organizations precluding teams to fully use the totality of evidence available to inform treatment benefit.

PRO scientists and Programming analysts saw opportunities in using interactive tools such as Spotfire and R-Shiny to supplement traditional analytic packages for PRO analyses. This paper describes an interactive analysis framework for improved delivery of PRO outputs and demonstrates that such framework offer the flexibility to provide an efficient solution to PRO data exploration and analysis.

PATIENT-REPORTED OUTCOMES ENDPOINTS

Some concepts such as pain severity, one's ability to go through day-to-day activities and quality of life can only be reported by the patients themselves.¹ To minimize measurement heterogeneity inherent to any subjective/ human-based assessments, patient-reported outcomes often consist of close-ended questions associated with rating scales that capture severity, frequency, occurrence or impact for instance of a given concept such as a symptom (*Over the past week, how much pain did you experience?*, 0=no pain, 10=pain as worst as I could imagine). PRO instruments vary from a simple one-item scale to lengthier questionnaires that yield several multiple-item scores and often one single summary score. Instruments and their resulting scores undergo rigorous validation to confirm content validity and psychometric performance. Regulatory and Health Technology Assessment Authority reviewers have offered guidance on how to best implement these instruments in clinical trials so that PRO could be used as part of the totality of evidence available to inform their review and decision-making.¹⁻³ In oncology, analyses of PROs generate similar outputs across scales, including Kaplan-Meier curves for time to event analyses, spaghetti plots for changes in

mean population score from baseline to each visit timepoint, bar graphs to document completion compliance at each visit timepoint, etc. In addition to be run across scales, the analyses are also conducted according to different variables such as demographic characteristics (e.g., compliance according to recruiting site), clinical characteristic (change in pain according to bone metastases status), population (e.g., ITT, PRO-evaluable, Safety) and to different interpretation thresholds (e.g., 10 point minimal important difference, 20 point minimal important difference). All these combined are needed for a comprehensive interpretation of the data but the large amount of outputs associated with PRO endpoints put significant burden on the analytical teams.

PILOTING AN INTERACTIVE APPROACH FOR PRO DATA AND ANALYSIS

Considering the large volume, yet repetitive nature of PRO outputs, interactive applications that would enable users to run similar analyses with similarly structured inputs such as those generated by scale scores, and quickly visualize their outcomes is a fit for purpose solution to resource scarcity and programs acceleration.

SELECTING THE INTERACTIVE APPLICATION

The first step in defining the solution was the selection of the interactive application(s) to run and interpret PRO analyses. Internally Spotfire was already in use for safety data exploration and analysis while R-Shiny for efficacy and bio-maker analyses. We took into accounts our skill levels and existing infrastructure and decided to go with **a mixed toolkit using both Spotfire and R-Shiny for PRO pilot**:

- For plots and summaries with simple descriptive statistics such as completion rate, mean line plot with $\pm$ SE, change from baseline over time (worsened, stable, improved), Spotfire is recommended with its dynamics and drill-down capability.
- Summaries for inferential statistics such as time to event table and Kaplan-Meier plot are created using R-Shiny based on existing R programs for efficacy endpoints.

Below highlight the key features of the tools:

- **Consistency between visualization results from interactive tools and CSR outputs for PRO endpoints**

The interactive tools are built upon PRO datasets that are used to generate outputs for Clinical Study Report (CSR); in addition, analyses and visualizations are created by following the same output layout and business rules as for CSR. The interactive tools support both tabular and graphical formats with drill-down capability into patient-level listing. The summary tables correspond to CSR outputs while graphs help identify data trends easily. PRO scientists can select which scale scores to display or scores at selected visit timepoints without a need for extensive programming.

- **Flexibility of choosing analysis population and filters for further data exploration and subgroup analysis**

The interactive tools are designed to allow users to select any pre-specified population of interest such as intent-to-treat, safety evaluable population. With one click, all visualization results are generated instantly. Users can also conduct subgroup analysis via filters such as race or bio-maker subgroups.

- **Dynamic calculation of PRO responder for sensitivity analysis**

In PRO instrument, a threshold is defined that informs the clinical meaningfulness of the findings. Over the past few years, HTA (GBA in particular) and FDA have started to demand sensitivity analyses using different threshold to ascertain the clinical meaningfulness of the data. The interactive tool allows users to enter different thresholds (eg. 5 points for less conservative, or 20 points for more conservative) and dynamically re-run the main analyses as part of the sensitivity analyses.

DESIGNING THE INTERFACE

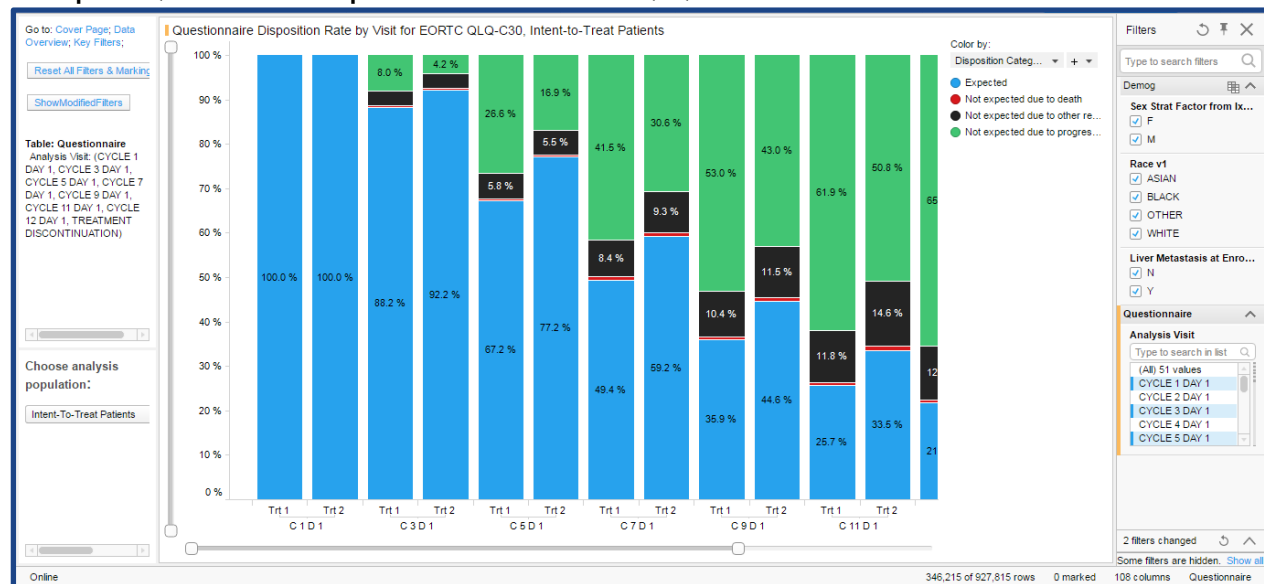
The second step was to appraise the statistical analyses plan and define which analyses would be the best candidate(s) to implement in the application. The decision was driven by the opportunity seen in efficiency gained in using the same analysis specifications and programs for each scale scores generated by the PRO questionnaire. Inferential analyses were limited to R-Shiny as the same R-Shiny application used for traditional efficacy endpoints (cox, mixed model) can be easily expanded to include PRO endpoints; Longitudinal and cross-sectional descriptive analyses were selected for Spotfire for the visualization, the friendly viewer interface, use of drill-down capability to patient level data, and linking across data domains.

More features were added to the dashboard to include demographic, clinical characteristics and other variables (stratification, countries) that would be interesting to explore to inform the interpretation of the data. Additional features were also added to dynamically test analytical constraints (e.g., event definition, patient responders, populations) to inform sensitivity analyses.

Examples of Spotfire and R-Shiny interfaces are provided in the next paragraphs to demonstrate the key functionalities as highlighted above. The PRO questionnaire used here is the EORTC QLQ-C30, a 30-item questionnaire that generates 15 single item or multi-item scores for patients with cancer tumors to inform the status of their quality of life and functioning and symptom severity.

EXAMPLE INTERACTIVE DATA DISPLAYS

Example 1: Questionnaire Disposition Rate for EORTC QLQ-C30

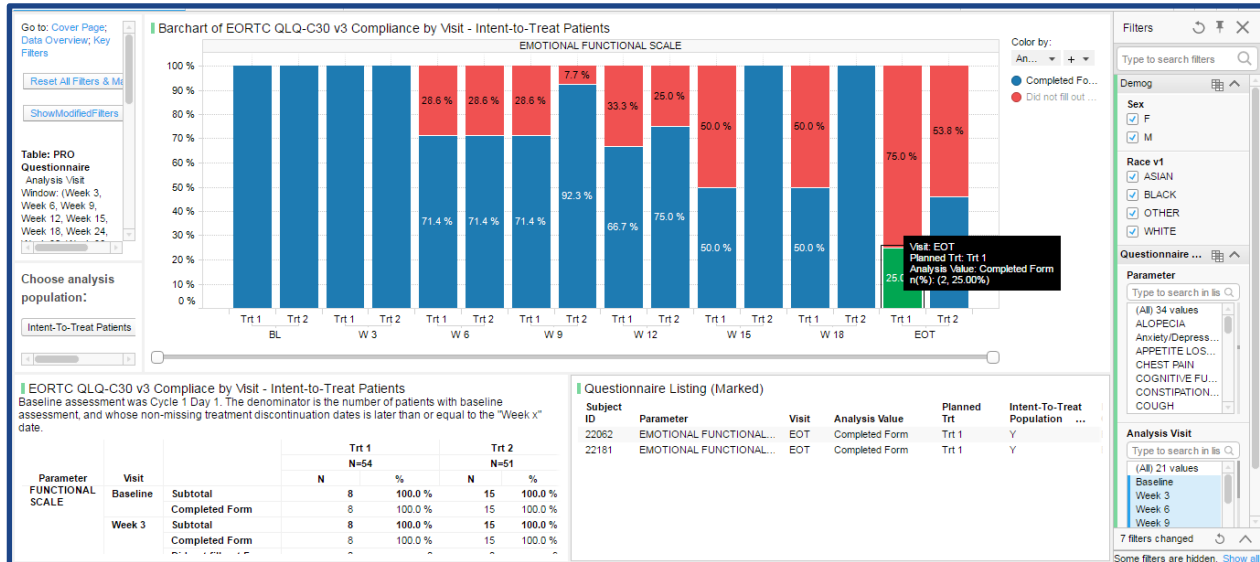


We wanted to understand which patients contributed to the data available. The figure above documents the number of patients who completed the questionnaire alongside the number of patients who were not expected to complete the EORTC QLQ C30 at each scheduled assessments. Patients who were not expected to provide data were further split into three categories:

- 'Progression of disease' for those patients that already progressed
- 'Death' for those patients who had already died by the time of visit, and
- 'Other' for the remaining patients (e.g. questionnaire not administered at the site, patient unwilling to complete the questionnaire)

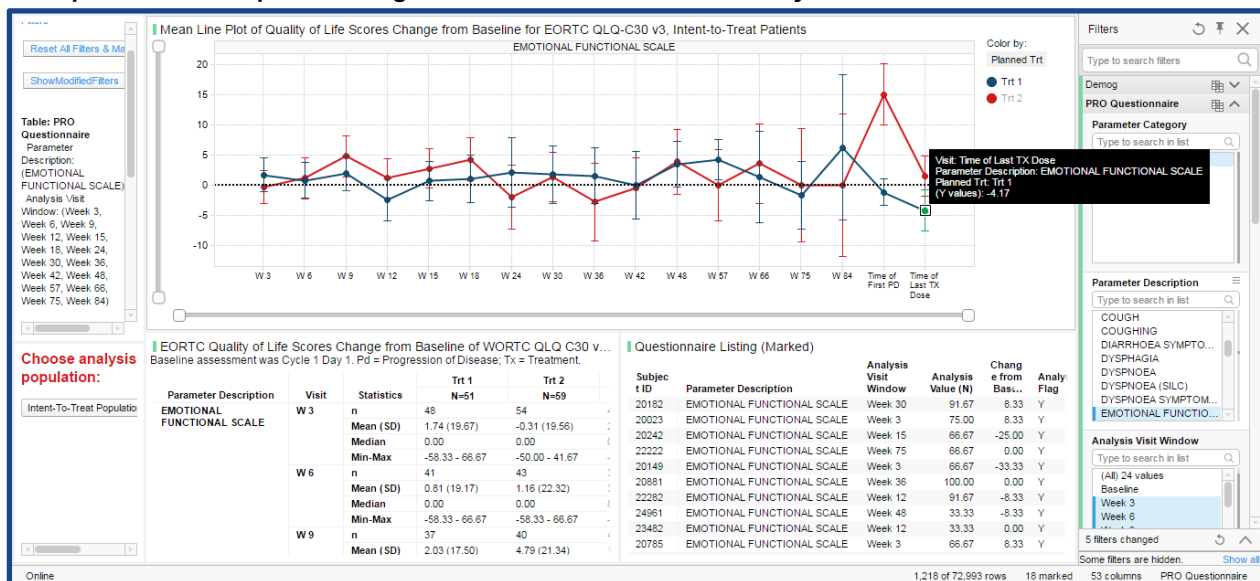
The graphical display makes it so much easier to identify the trend of patient disposition rate over time as compared to a tabular format. As expected, the percentage of patients who did not complete the questionnaire due to progression of disease increased over time much more than the percentage of patient who died or who elected not to complete the questionnaire. Furthermore, having completion rate of the questionnaire for each treatment displayed side by side, provide information on the strength of the evidence available over time and the fairness in comparing the data. By selecting different filters, similar graphs could be generated to inform trends (e.g., patients with clinical characteristic X contributing less PRO data overtime due to earlier death or progression).

Example 2: Questionnaire Completion Rate to Inform Compliance



The above screenshot displays the completion compliance rate for the Emotional function scale per treatment arm and timepoints. The patient frequency counts and percentages summary in the bottom reflect the tabular outputs systematically generated for the CSR (that runs on several pages). The bar chart is the visual display. The interactive visualization offers the flexibility to zoom to select specific visits or de-zoom for trend overtime. The Parameter (=scale scores) can be slid down so that to quickly appraise any potential difference in completion rate across scales. By using different filters (on the left part), the same can be generated for the Safety population (when using PRO symptom severity to understand treatment burden), or Efficacy PRO-evaluable population (adding a constraint of only considering patients who completed the questionnaire at baseline).

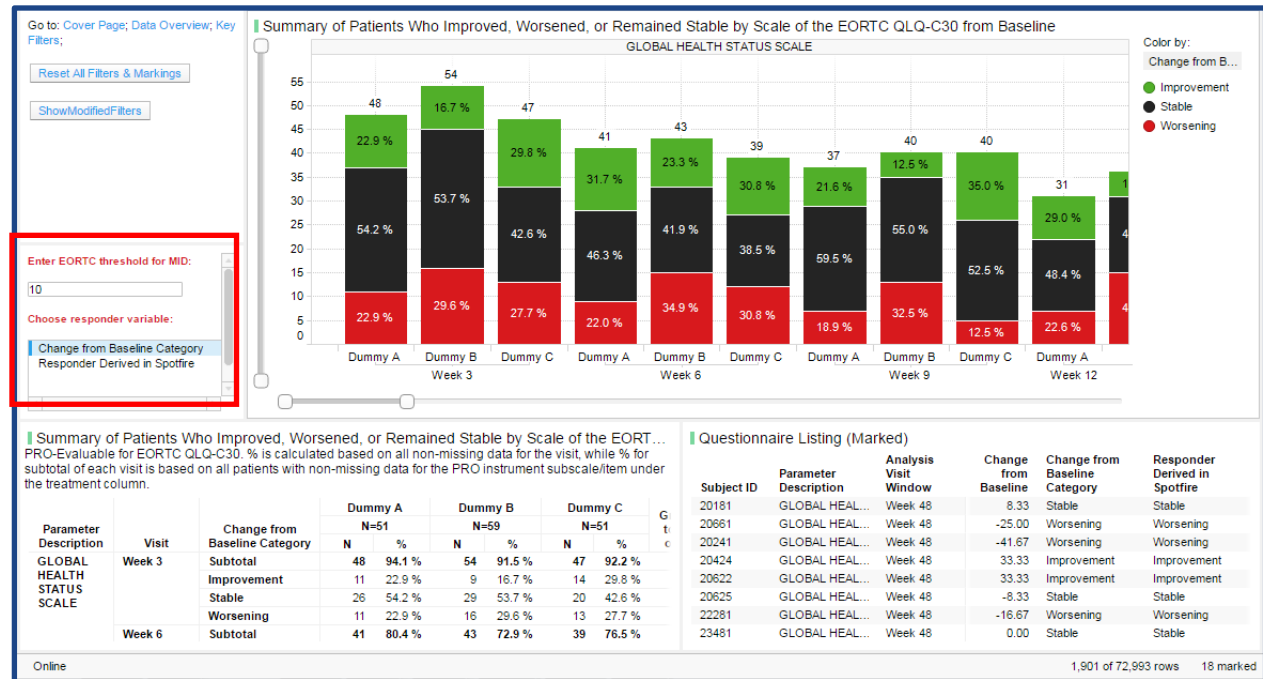
Example 3: Mean line plot of change from baseline $\pm$ standard error by visit



This screenshot displays the mean change from baseline values to each visit including specific events (i.e., time to first PD, time of Last Tx dose) for the EORTC Emotional Functional Scale. As elaborated earlier, the summary table

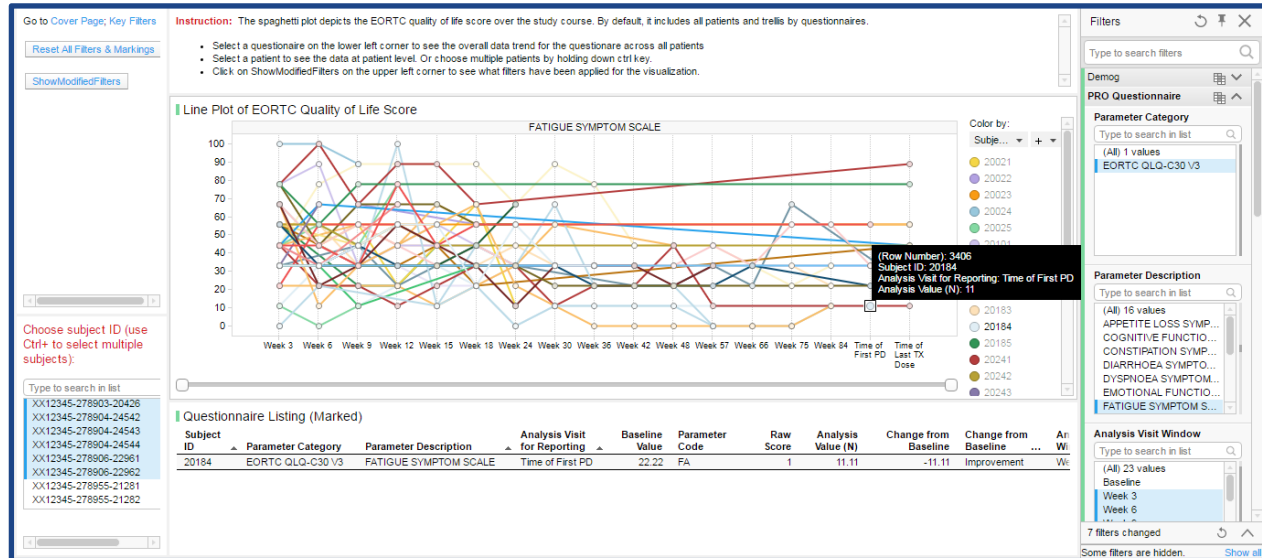
in the bottom contains the exact same results as in CSR output while the line plot on the top depicts the mean value with standard error by weeks. The same graph can be presented using the 95% Confidence interval .

Example 4: Proportion of patients with clinical meaningful change from baseline



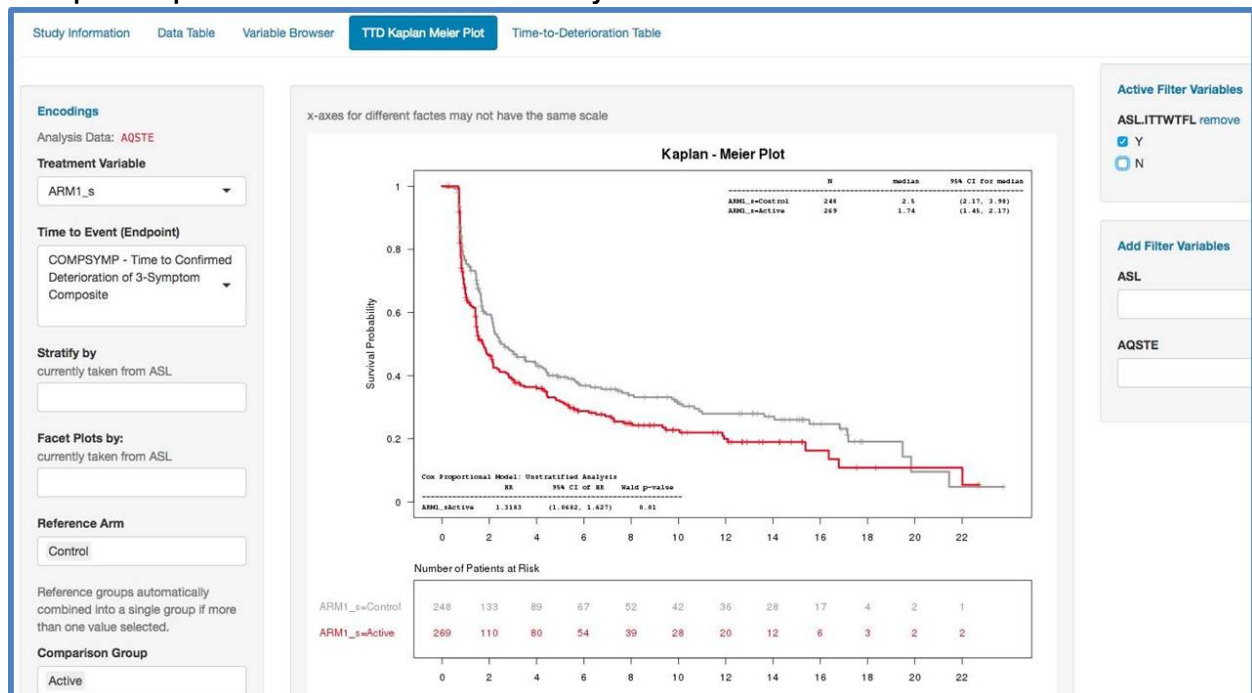
The visualization screenshot above depicts proportion of patients for which scores on the Global Health Status scale indicate improvement, worsening, or no change in health-related quality of life. The stacked bar chart shows N by visit. The pre-specified definitions for “improvement” and “deterioration” use a threshold of 10-point in score increment or decrement, respectively. Through an additional filter indicated in the rectangular box on the left panel, the user can elect to use different thresholds such as 5 or 20 enabling quick determination of the strength of the evidence. The flexibility of the dynamic calculation of pre-determined derived variable such as greatly saves time for PRO Scientists to gain comprehensive understanding of the PRO data.

Example 5: Spaghetti plot of questionnaire score



Outputs for submission reflect population-level analyses. Yet patient-relevant evidence is often meaningful not when expressed as a population mean output but rather as individual experience. Sportfire enables access to patient-level data which then enables exploration of PRO data in light of other information (e.g., decrease in pain symptoms following onset or increase in analgesic concomitant medication).

Example 6: Kaplan Meier Plot for time to event analysis



The visualization screenshot above depicts time to deterioration in symptoms and includes the estimate of the hazard ratio per cox regression model. This is the same type of output generated for the purpose of the CSR. However, as teams might prioritize certain scales, access to the interactive application, in this illustration, R-Shiny, enables PRO scientists to generate the curves and associated estimates for every single score that comprise the EORTC plus any combination as long as the event definition has been pre-programmed (e.g., time to fist deterioration of 10 points,

time to sustained deterioration of 20 points, time to deterioration in pain or cough or fatigue, time to deterioration in pain and cough and fatigue, etc).

R-Shiny provides not only the visual description of the difference between the treatment arms based on the Kaplan-Meier method to estimate the median timing of an event for each treatment arm but also provides the estimate of the treatment effect expressed as an Hazard Ratio, as well as a 95% CI as generated by the Cox regression model.

INTERACTIVE STANDARDIZED FRAMEWORK

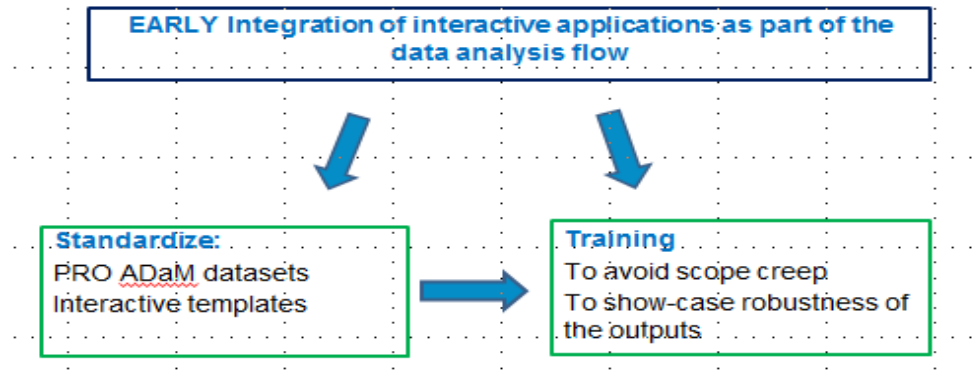
It was an iterative process of application development in which requirements and visual displays with filter options have been refined based on ongoing discussions between PRO scientists and Programming analysts as well as learning from information requests from Health Authorities and Health Technology Assessment Authorities. The semi-standardized templates with fit-for-purpose design were then used in the following months for additional study readings.

From the pilot experience and ongoing implementation, the interactive tools are proved to offer an agile solution to support PRO data exploration and decision-making. The interactive tools could be used to generate all CSR outputs, and therefore help ensure quality of CSR results by cross-checking visualization results from the interactive tools and static CSR outputs. In addition, with simple click of filter(s) of interest, the same type of analysis could run for PRO secondary or exploratory endpoints, or for subgroup results via filters such as race, clinical biomarkers. One of many successful use cases is that the subgroup analysis results for Asian population were easily obtained via the interactive tools to support initial filing consultation with health authorities.

On the other hand, the pilot went through quite a few challenges. The main challenge lies in the readiness of PRO analysis datasets, which were used as source data for the interactive tools. The PRO analysis datasets have three main areas for improvement in order to support the smooth and timely deployment of the interactive tools at project/study level. Specifically, a) PRO analysis data are currently not fully standardized across studies, which caused inefficiency due to some inconsistent variable naming conventions or data structure; b) PRO analysis data sets were typically programmed based on pre-specified analysis plan for CSR purpose, and thus may not include all PRO exploratory endpoints in order to inform main and sensitivity analyses and exploration of the data; c) PRO analysis datasets may not be available right after database lock as resources were pulled to support main analysis related to efficacy and safety.

The interactive standardized framework was therefore proposed to look into a) improvement of data standardization for commonly used PRO instruments; b) process changes to integrate PRO interactive approach as part of data analysis flow; c) development of standardized interactive tools; and d) training to promote the proper use of the interactive tools including scope of use and preventing one-on-one request reflecting formatting preference. This is shown as the diagram below. Through close collaboration to plan early and start early, we aim at front loading PRO datasets so that interactive tools can be deployed timely to empower PRO scientists with comprehensive understanding of totality of evidence results.

Interactive Standardized Framework



CONCLUSION

Interactive applications are fit for the purpose of analyzing and visualizing multidimensional yet structured patient-reported outcomes data. The pilot demonstrated that using interactive applications to complement traditional analytical suits provide a more efficient experience to analytical teams and research scientists when interpreting the PRO data and communicating the findings. Drill downs capabilities foster data exploration to better understand each patient experience (patient level data), to test the conclusions across subpopulations, and to inform trends (use of dashboard filters). Data interpretation and communication is made more efficient and less resource intensive with the user friendly data visualization.

REFERENCES

1. US Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. 2009 < <http://purl.access.gpo.gov/GPO/LPS113413>>

Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man: The use of patient-reported outcome (PRO) measures in oncology studies

2. CDER Patient-Focused Drug Development, <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm579400.htm>

3. Current FDA Approach to Describing Patient Experience Data, Mallorie H. Fiero, PhD Mathematical Statistician FDA, CDER, Office of Biostatistics, DBV <https://www.fda.gov/downloads/NewsEvents/MeetingsConferencesWorkshops/UCM620947.pdf>

ACKNOWLEDGMENTS

We thank Caroleen Quash, Agnes Hong, Katherine Kim, Thomas Karangiannis, Ying Yuan, and Sara Bodach for their time in participating in the pilots.

CONTACT INFORMATION

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

Xiangyun Wang

Genentech, a member of Roche Group

wang.xiangyun@gene.com

Elisabeth Piau-Louis

Genentech, a member of Roche Group

piaultlouis.elisabeth@gene.com

Tadeusz Lewandowski

F. Hoffmann-La Roche

tadeusz.lewandowski@roche.com

Brand and product names are trademarks of their respective companies.