RESEARCH LETTER

Follow-Up on the Use of Machine Learning in Clinical Quality Assurance: Can We Detect Adverse Event Under-Reporting in Oncology Trials?

Timothé Ménard[1] · Björn Koneswarakantha[1] · Donato Rolo[2] · Yves Barmaz[1] · Leszek Popko[1] · Rich Bowling[3]

Dear Editor,

In a previous project [1], we developed a predictive model that enabled Roche/Genentech quality leads oversight of adverse event (AE) reporting. External clinical trial datasets such as Project Data Sphere (PDS) [2] allowed us to further test our machine learning-based approach to alleviate concerns of overfitting and to demonstrate the reproducibility of our research.

Our primary objective was to further validate our model for detection of AE under-reporting using PDS data. Our secondary objective was to build an oncology-specific model using a combined dataset of Roche and PDS data. The scope remained as predicting AEs—not adverse drug reactions—that occur in clinical trials. Good clinical practice requires all AEs (regardless of the causal relationship between the drug intake and the events) to be reported in a timely manner [3].

The curation process of downloadable PDS studies (as of November 2019) left five studies that fulfilled our data requirements, as sponsors are not required to share the full datasets. They were large phase III trials and included 742 investigator sites, 2363 subjects, and 51,847 visits. Hence, we could use PDS data to achieve our objectives.

The oncology-specific model was built using the methodology described in our previous manuscript [1]. We used a combined dataset of 53 completed oncology studies (Roche + PDS). Our final model used 38 features built from patient and study attributes.

✉ Timothé Ménard
   timothe.menard@roche.com

[1] F. Hoffmann-La Roche, Basel, Switzerland

[2] Roche Products Limited, Welwyn Garden City, UK

[3] Genentech - A Member of the Roche group, South San Francisco, USA

To test whether our model can be applied to non-Roche studies, we compared the quality of the predictions using a scatter plot (Fig. 1a) and found that, within a range of 0–150 on both axes (> 94% of all datapoints, our region of interest [ROI]), the predictions matched the observed values for both datasets equally well. To quantify the goodness of fit, we used scale-independent performance metrics (which are adequate for comparing the goodness of fit of different datasets used by the same model [4]): symmetric mean absolute percentage error (SMAPE) [5] and symmetric mean absolute poisson significance level (SMASL). The latter is calculated by subtracting 0.5 from each poisson significance level measurement, converting it to its absolute value, and taking the mean. SMASL puts equal weight on over- and under-predicting and has a range from 0 to 0.5 (i.e., The smaller the value the better the fit). Considering SMAPE, average predictions for the PDS study sites were slightly better than for the Roche study sites, whereas the reverse was true for SMASL (Fig. 1b). We concluded that the goodness of fit for both datasets using our model was very similar within the ROI.

For the secondary objective, we tested how well the oncology model (using Roche and PDS data and the same algorithm [1]) would detect simulated test cases on data not used for model training. For relevant simulation scenarios of 25%, 50%, and 75% under-reporting on the site level, our model scored an area under the curve (AUC) of the receiver operating characteristic curve of 0.60, 0.77, and 0.90, respectively. These AUC values were on par with the performance of our previous model (0.62, 0.79, 0.92).

Our main challenge was the lack of availability of external data that had all required attributes. Initiatives such as PDS or TransCelerate should be further promoted, as sharing data from historical clinical trials could serve a variety of uses [6].

Our analysis supports our approach for detection of AE under-reporting and provides rationale to integrate our model into routine clinical quality practices. Furthermore,
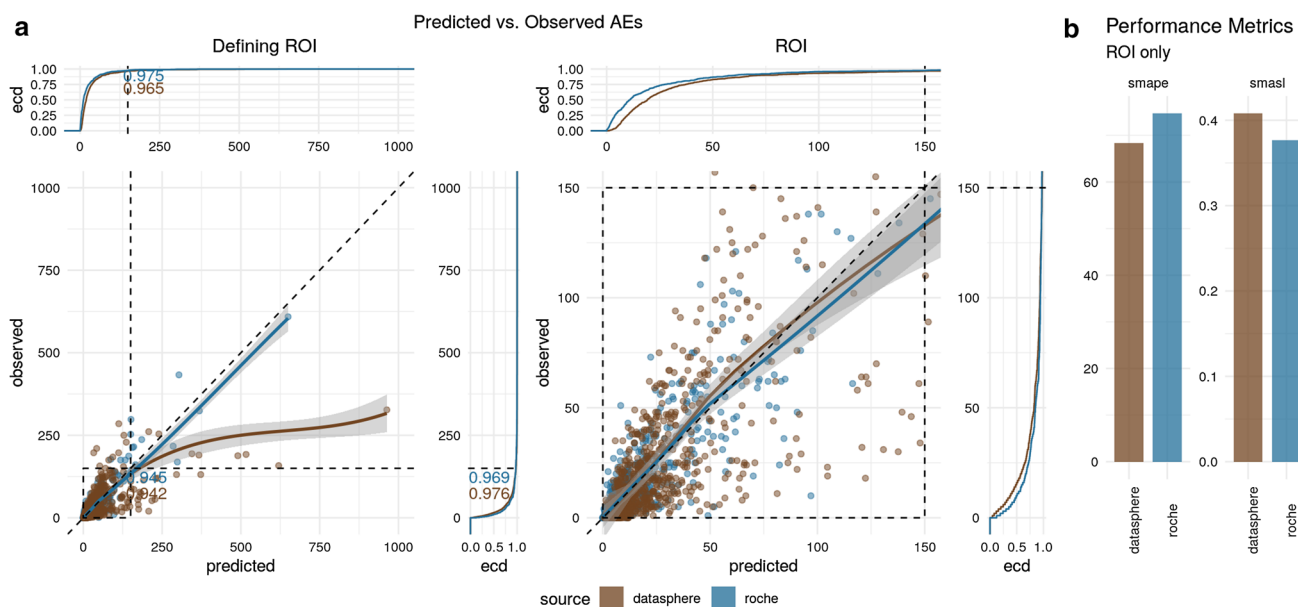
Fig. 1 Modelling performance on Roche vs. PDS studies (predictions for number of AEs were generated for Roche and PDS study sites; neither category of study site was used for model training or validation). **a** Predicted vs. observed number of AEs with locally estimated scatterplot smoothing (loess) plus standard error with marginal ecd of Roche and PDS study sites. Numbers on ecd plots denote cumulate density values at x,y = 150, and numbers on scatterplots denote the ratio of observations found within the ROI, defined by x/y range 0–150. **b** SMAPE and SMASL were calculated for ROI. *AE* adverse event, *ecd* empirical cumulative density, *PDS* Project Data Sphere, *ROI* region of interest, *SMAPE* symmetric mean absolute percentage error, *SMASL* symmetric mean absolute significance level

following the same methodology, we could produce a model for detection of AE under-reporting in oncology trials based on internal and external study data.

Compliance with Ethical Standards

References

1. Menard T, Barmaz Y, Koneswarakantha B, Bowling R, Popko L. Enabling data-driven clinical quality assurance: predicting Adverse Event reporting in clinical trials using machine learning. Drug Saf. 2019;42(9):1045–53.
2. Project Data Sphere. https://www.projectdatasphere.org/projectdatasphere/html/home. Accessed 15 Nov 2019.
3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. E26(R2) Guideline for Good Clinical Practices. 2016. https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf. Accessed 15 Nov 2019.
4. Fildes R, Makridakis S. Forecasting and loss functions. Int J Forecast. 1988;4(4):545–50.
5. Makridakis S. Accuracy measures: theoretical and practical concerns. Int J Forecast. 1993;9(4):527–9.
6. TransCelerate Placebo Standard of Care Data Sharing workstream (pSOC). https://transceleratebiopharmainc.com/initiatives/placebo-standard-of-care/. Accessed 15 Nov 2019.