Enabling data-driven clinical quality assurance: predicting Adverse Event reporting in clinical trials using machine learning

Drug Safety

Timothé Ménard[1] • Yves Barmaz[1] • Björn Koneswarakantha[1] • Rich Bowling[1] • Leszek Popko[1]

1 F.Hoffmann- La Roche AG

Correspondence: timothe.menard@roche.com

Electronic Supplementary Material #3 - List of formulas

Relation between the numbers of AEs reported during visits, by patients and on investigator sites:

$$Y_{site} = \sum_{patient \in site} Y_{patient}$$

$$Y_{patient} = \sum_{visit \in patient} Y_{visit}$$

Poisson distributions of the number of AEs reported during visits, by patients and on investigator sites, and their relation:

$$Y_{visit} \sim Poi(\theta_{visit})$$

$$Y_{patient} \sim Poi(\theta_{patient}), \quad \theta_{patient} = \sum_{visit \in patient} \theta_{visit}$$

$$Y_{site} \sim Poi(\theta_{site}), \quad \theta_{site} = \sum_{patient \in site} \theta_{patient}$$

Significance level of an observed number of AEs reported from a site given the available information about this site:

$$S(x_{site}, y_{site}) = P(Y_{site} \leq y_{site} \mid x_{site}) = \sum_{k=0}^{y_{site}} \frac{\theta_{site}^{k}}{k!} e^{-\theta_{site}}$$

Loss function minimized in our machine learning model:

$$L(f) = \sum_{visit} l(y_{visit}, f(x_{visit}))$$

$$l(y_{visit}, f(x_{visit})) = 2\left(y_{visit} \log \frac{y_{visit}}{f(x_{visit})} - y_{visit} + f(x_{visit})\right)$$