Diabetic retinopathy

Optimization of Brolucizumab phase II clinical program

Gilles Guillot

2020-03-04 13:03:46

Background

- Clinical trials are costly ($50,000 per patient in oncology) and time consuming
- Often carried out with delays and cost overruns
  - screening failure
  - under recruitment
  - drop out
- Clinical trial data of varying quality
  - lack of screening oversight (low screening failure rate)
  - trial patients not representative of target population (high screening failure rate)
  - departure from trial protocol

Clinical program optimization problem

Identify physicians and trial sites likely to provide

- high quality trial data
- complying with budget
- on time

Clinical program optimization problem (cont')

Best predictor of performance in the future is performance in the past.

Data analytics objectives:

- Identify *performance indicators* in database
- Extract signal from data
- Identify physicians and trial sites best in terms of those indicators

Data at hand

Excel file Health_Insights_Case_Study___TSO containing three datasets (sheets)

- Set A: 139 US physicians
  - some replicates in `gpharma_HCP_ID`
  - `Patient_Count`
  - 11 specialties incl. `OPHTHALMOLOGY`
  - admin. / geog. data

Data at hand (cont')

- Set B: 365 contributions to studies
  - Contribution = `PI` working in a `CENTER`
  - Performance indicators:
    - `pat_TARGET_TREATMENT` and `pat_ENTERED_TREATMENT`
    - `# of competitor trials by PI (active)`
    - `# of Supporting Staff`
    - `PI Tier` 1,2,3
    - `PI Risk Score` H/M/L
  - Admin. / geog. data
  - Uninformative columns:
    - `Trial_STATUS_DESC` all completed
    - `Performance` ratio of `ENTERED`/`TARGET`
    - `PI  Availability` fully missing

Data at hand (cont')

- Set C: correspondance `gpharma_HCP_ID` (set A) to `Doctor ID` (set B)
- PI Affiliation: count of occurence of physicians

Big data usually defined as: large, complex structure, real-time, messy

Method

- Set A
  - `Patient_Count` not documented
  - does not seem to match any of
    - `pat_TARGET_TREATMENT` or
    - `pat_ENTERED_TREATMENT`
    - even after various summations
- Set C document link between A and B
- Sheet “PI Affiliation”
  - `Count of HCP_Last_Name` not documented
  - does not seem to match physician occurences in A or B
- Only Set B considered in the sequel

Result

Conclusion

Supplementary material

All documents available on the github repository GiantPharma

- R code for data analysis: load.R
- Slides in R markdown under Rstudio: GPharma.Rmd

THANK YOU!