Finding Influential Individuals in Drug Trial Data

UPPSALA UNIVERSITET

Authors:

Magnus Larsson

Magnus.Larsson.5876@student.uu.se

Gunnar Eggertsson

Gunnaratli.Eggertsson.8860@student.uu.se

Peili Guo

Peili.Guo.7645@student.uu.se

Supervisor:

Rikard Nordgren

rikard.nordgren@farmbio.uu.se

Project in Computational Science, 2018

Department of Information Technology

PROJECT AIM

The aim of the project is to investigate alternative methods to identify **influential individuals** in pharmacometric modelling.

Case Deletion Diagnostics (**CDD**) is the golden standard method to detect influential individuals for non-linear mixed effect models.

Bootstrap analysis is the standard way to evaluate the stability of model estimates.

We have investigated if the bootstrap data could also be used to identify influential individuals without the need to run CDD.

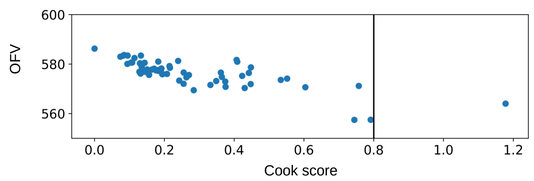
DATA

Project data: real patient data from clinical trials.

Substances (total no. individuals): phenobarbital (59), digoxin (227),nevirapine (58), paclitaxel (66).

CDD data was analysed to estimate the influence of different individuals.

Threshold value for Cook's score distinguishes the individuals deemed influential.



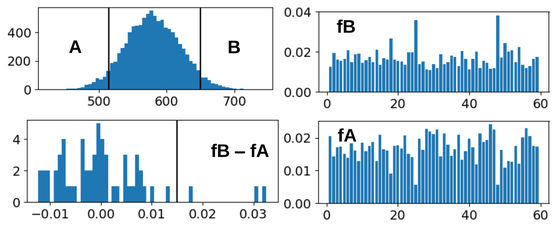
Total-OFV values from Bootstrap data exhibit normal distribution pattern. Analysis focused on the two extremes (A and B).

fA: Relative frequency of inclusion for individuals in A.

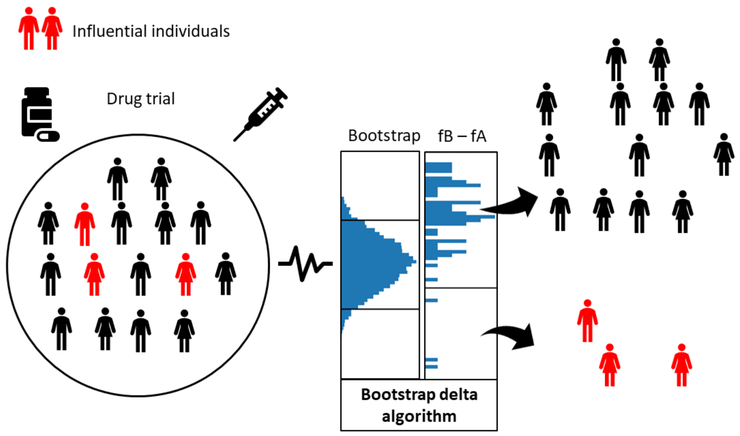
fB: Relative frequency of inclusion for individuals in B.

fB – fA: Our estimate of influence.

Highly influential individuals should contribute more towards results in B than in A.



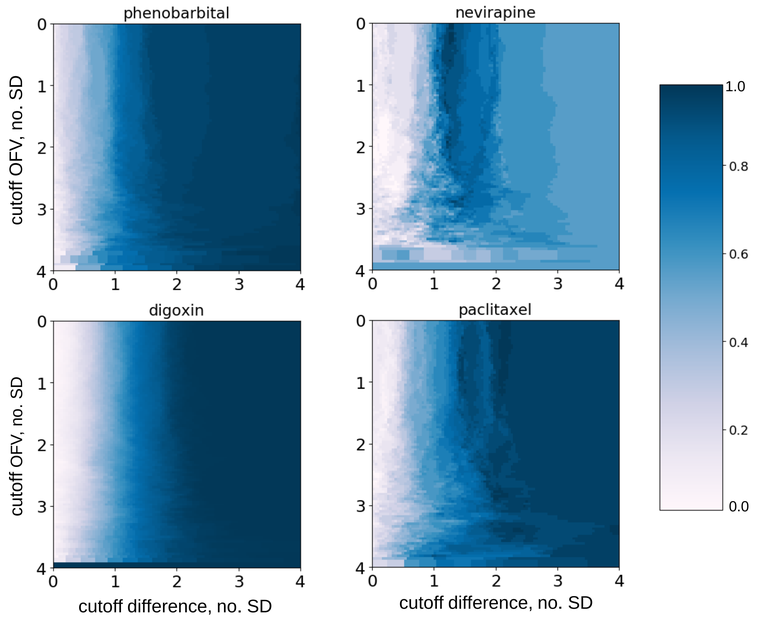
METHODOLOGY



Different cutoffs were tested for the bootstrap OFV distribution and the normalized set difference (fB - fA) to optimize the method for identification of influential individuals.

Accuracy was measured as defined in Receiver Operating Characteristics, ROC as

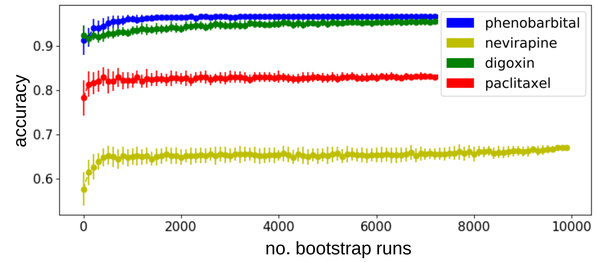
$$Accuracy = \frac{True\ positives + True\ negatives}{Total\ no.\ individuals}$$



Optimal cutoff for the set difference varied between data sets, but overall the optimal value was 1.75 SD. For OFV, the accuracy was only marginally affected by the choice of cutoff, 0.05 SD was selected.

RESULTS

The new method was run by simple random sampling from the total bootstrap set. The result showed that influential individuals could be identified with an accuracy of between 0.65 and 0.95.



Between 700 and 2000 bootstrap runs achieve close to maximal accuracy. The number of bootstrap runs lies within that range in a typical pharmacometric analysis.

The time for different substances are listed in the table below. The non-linear mixed effect models runs both CDD and bootstrap to evaluate the model.

	CDD time (s)	bootstrap 1000 time (s)	bootstrap 2000 time (s)	Bootstrap 10000 time (s)
phenobarbital	58	523	1045	5232
nevirapine	2022	31785	63038	317012
digoxin	589	3671	7302	36604
paclitaxel	274	8221	16496	82318

It is important to find the influential individual and test the robustness of the model parameters.

The new method makes this possible without having to run CDD, thus saving the runtime for that analysis.

CONCLUSIONS

- It is possible to identify the influential individuals based on available bootstrap data. This provides additional valuable information from the already generated data.
- The accuracy of our method differs among substances from 65% to 95% (average of 85 %).
- The accuracy rate converges very quick to the maximal value at between 700-2000.
- Our results imply that maximum accuracy is achieved within the limits of typical pharmacometric analyses.