**MSc and Internship proposals**

GSK, Rixensart  
TRD (Technical R&D)

Contact: bernard.x.francq@gsk.com

---

**A Shiny Application to assess the power in Significance and Equivalence approach in Linear Regression Analysis.**

*Application in Vaccines Development, i.e. Stability Study.*

Stability studies are used in setting pharmaceutical product expiration dates. It is essential to ensure safety and efficacy through all the shelf life of the vaccine. Linear degradation models are usually proposed in the statistical literature, and an expiration date is estimated following FDA guidelines (Chow 2007, Appendix B). Decay of potency over time is modelled using a linear model with random (or fixed) slopes and intercepts for the batches.

To design a new study, sample sizes and power are usually assessed according to different scenarios. In the classical significance approach, the power is the ability to reject the null hypothesis under an alternative one. The power of a statistical parameters can be assessed using the design matrix and the expected residual variance via a F distribution. In the equivalence testing, (see e.g. Schuirmann’s Two One-Sided Test (TOST) procedure), one wants to show that the slope is not higher than a given threshold (Δ). The power is then the ability to conclude equivalence, in simulations the frequency of confidence intervals that lie entirely into the margins [−Δ, Δ]. Additionally, for a given design (as proposed by ICH guidelines) and a given power (i.e. 80%), one can calculate the targeted analytical variability (the accuracy of the analytical method) to reach to detect a degradation (calculate the residual variance for a given design and a given power).

The goal of this work is to develop a Shiny Application in R to calculate the power on parameters in linear models for a given design specified by the user. Performance of the methodology will be evaluated by means of simulations and applications to case studies within CMC statistics and vaccines development.


Bayesian Optimal Design of Experiments with applications in vaccines

Current state-of-the-art vaccines development is based on the “Quality-by-Design” paradigm, where risk-based and data driven decisions are key. A prominent example is the classification of process parameters into “critical” and “non-critical” based on a series of Designs of Experiments (DoE) performed during vaccine development. This helps to understand the relationship between Critical Process Parameters (CPPs) and “Critical Quality Attributes“ (CQAs) and then to establish the “Design space”.

Optimal design of experiments are powerful tools to assess different experimental conditions with a low number of samples by taking into account several constraints. Scientists usually have some knowledge about process, i.e. the effect of a parameter or the operator-to-operator variance, or the repeatability of the process. The aim of the project is first to perform a literature search on the Bayesian optimal designs. Second, compare the Bayesian optimal DoE to the recent definitive screening designs and to the classical D- or I-optimal designs (with real data and simulations).


A-optimal Design of Experiments applied in High-Throughput experimentation. Implementation with Shiny.

A-optimal designs are good alternatives to the ‘classical’ D-optimal designs as the focus is made on the interpretation of the algorithm (minimizing the mean of the standard errors of the parameters). The aim of the project is to compare the A-, D- and I-optimal designs with simulations and real data, and to implement the A-optimal design in shiny. If results are promising, a collaboration with a local software engineering team is possible to make results available to a broader (GSK) audience in the form of a web application with graphical user interface. The main application is the High-Throughput experimentation where the different experimental conditions must be tested with 96-wells plates. The DOE and the randomization of the experiments will take into account these constraints to be applicable by the scientists and the equipment in the High-Throughput team.

https://en.wikipedia.org/wiki/High-throughput_screening

Bayesian and bootstrap probabilities of being out of specification limits in release limits

It is essential to assess product quality and vaccine potency. Vaccine shelf life is managed through determination of a minimum potency release requirement, which ensures appropriate potency throughout expiry. Release limits of drug dosage forms are the bounds on the potency at which a lot can be released with a 'guarantee' that it remains within limits throughout its shelf life.

Decay of potency over time will be modelled using a mixed model with random slopes and intercepts. The Probability to be Out Of Specification (POOS) can be evaluated with a Bayesian repeated measures model for longitudinal data. The aim of this project is to compare the Bayesian framework with a parametric bootstrap approach to calculate the POOS, but also to construct a confidence interval for this POOS. The following two definitions to clarify the concept of release limits will be compared:

The release limit should (1) ensure with a given probability that a batch with a true value at time of release above the release limit (assuming a decaying product over time) will be measured in specification at the end of shelf life. (2) to ensure that a batch with a measured value at release above the release limit will have a true value in specification at end of shelf life.

Simulations will be run to compare both approaches (Bayesian and bootstrap) for both definitions. Real data will be used to apply these techniques in a case study.


**Evaluation of Coefficient of Variation in a mixed model**

Frequentist, Bayesian and Bootstrap approach to calculate a CV and its confidence (credible) interval under mixed models. The uncertainty of an estimated coefficient of variation can be expressed by a (1-sided) confidence intervals. The calculation of such confidence interval is challenging as the CV is typically a ratio of 2 estimated statistical parameters where, in mixed models, the mean is replaced by a given fixed effect estimate and the standard deviation by the square root of a (linear combination of) variance component. In mixed models, the sum of all the variance components in a variance components mixed model is the so-called total variance that can be expressed as a CV. In pharma industry, this is related to the ‘intermediate precision’. The assessment of its uncertainty is essential to quantify for example the risk of OOS (out of specification). An adaptation of McKay formula can be used to calculate its confidence interval. The goal will be a literature search and compare different formulae by simulations and applied to some case studies.


Growth curves non-linear modeling in vaccine development

Growth curves modeling are common in pharma industry, for example to investigate infection rates over time. Such data usually present low variability at time zero and greater variability across time. The goal of this project is to compare different mathematical models (Gompertz, logistic curves…) by simulations and with real data. Statistical intervals in such models are usually calculated with asymptotic properties where the coverage probabilities are lower than the nominal level. The main goal is the growth comparison between different curves (i.e. a categorical effect…), predict and contrast between different time points. Is the infection rate difference between 2 time points equivalent or not ? Frequentist or Bayesian analysis can be investigated. R package can be built (with unit tests). If results are promising a collaboration with a local software engineering team is possible to make results available to a broader (GSK) audience in the form of a web application with graphical user interface.

http://shiny.webpopix.org/sia/nlmemGrowth/
https://en.wikipedia.org/wiki/Growth_curve_(biology)