R Notebook for Cq Shift

Table of Contents

# Preliminaries

## Motivation

This notebook provides the code required for defining of “Cq shift” - the difference between the quantification cycle (Cq) of a pool test and the Cq of component samples measured in individual sample tests. Cq shift arises from the effective ‘dilution’ of viral loads in positive samples when mixed (pooled) with virus-negative samples. It therefore takes more cycles of PCR to amplify the viral genetic material and make it detectable.

Alongside other variables, including the Cq distribution within tested samples (a higher % of ‘weak positive’ samples with low viral load increases risks of non-detection after Cq shift), degree of clustering of positive samples within pools, and pool size essential variable for determining the sensitivity of pooled testing, “cq shift” is a key determinant of sensitivity for pooled testing.

There were two reasons to examin “Cq shift” closely within this DTA systematic review:

### To provide population-adjusted measures of sensitivity (and specificity) for meta-analysis

The primary meta-analysis combined studies that took a true representative sample of samples (single-gate, quality-group 1) and those that took a purposive sample of positives and combined those with negative samples within pools (two-gate, case-control type, quality group 2). The latter typically over-sampled positives with low viral loads (“weak positives”) to verify limits of analytical sensitivity. However, this will lead to an underestimate of ‘real-world’ clinical sensitivity which depends on the mix (distribution) of viral loads in the test poopulation.

Several two-gate (quality group 2) studies provided “population adjusted” estimates of real-world clinical sensitivity alongside the observed results from a purposive sample. However, we wished to used a consistent method for obtaining “population-adjusted” summmary estimates of accuracy.

### To explore pool-testing protocols that may improve analytical sensitivity

## Methodology

Our overall approach mirrors that proposed by the US FDA: 1. *Observe* Base predictions on laboratory observations of change in Cq (Cq shift) and probability of detection of a purposive sample of clinical specimens when tested in pools of desired size. 2. *Model* Define the relationship between the Cq measure from individual sample testing and the expected Cq of pools containing positive sample(s), and the probability of detection at pre-defined Cq cut-offs. 3. *Predict* Apply the model in step 2 to a true representative sample of specimens from a relevant test population with a representative distribution of viral loads (Cq values). Define the likelihood of detection of the positive sample at the

## Load packages required

library(gamm4)  
library(tidyverse)  
library("earth")  
library(forecast)  
library("stats")  
library("lmtest")  
library("mgcv")

# 1. Load obervations

The code below loads the “Cq shift” data obtained from studies included in the Cochrane review, and edits to create two datasets in formats required for different methods.

shift <- read\_csv("egshift.csv", col\_types =cols(  
 gene = col\_factor(),  
 indiv= col\_factor(),  
 quality=col\_factor(),  
 eval=col\_factor(),  
 study =col\_factor(),  
 pool\_id = col\_factor(),  
 n1=col\_factor()))  
shiftex <- shift %>% group\_by(pool\_id) %>% summarise (  
 pool\_exp1 = -log2(sum(2^(-cq\_ist))/10))  
  
shiftnew <- left\_join(shift,shiftex, by="pool\_id")  
  
shiftnew <- shiftnew %>% mutate (expgap = cq\_pool-pool\_exp1)  
shiftnew <- shiftnew %>% mutate (exp\_pool=cq\_ist+expgap)  
shiftnew <- shiftnew %>% mutate (diffobs = cq\_pool-cq\_ist)

# 2. Model Cq shift

Standard methods exist in laboratory and medical science for comparing measures, including Bland-Altman plots /analysis, Probability of agreement analysis, Passing-Bablok etc. However, these have limitations for the our purpose of modelling expected Cq shift:

1. Cq measures are made (often) on multiple SARS-CoV-2 gene targets in each sample, and the impact of pooling on Cq can vary between genes.
2. Multiple positive samples can arise within the same pool.
3. The probability of multiple positive samples being in the same pool varies with prevalence, and clustering.

We therefore explored two approaches for defining Cq shift:

* *GAMM: Generalised Additive Mixed Model*
* *BLMM: Bayesian Linear Mixed Model*

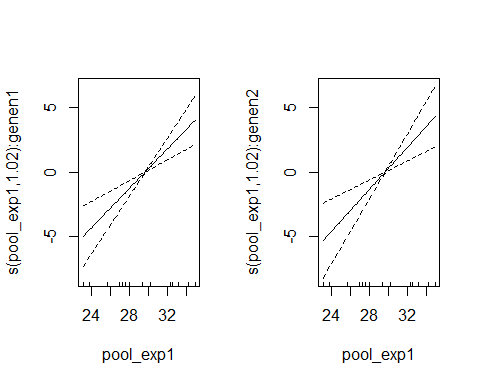
## 2.1 GAMM

We start from theoretical assumption that if a single positive sample is mixed within a pool, with other samples being negative, the pool Cq is expected to be:

where; Cqist = individual sample test Cq n = pool size

This can be generalised to a scenario of “i” positive samples within a pool as

gamm1<- gamm4(cq\_pool~s(pool\_exp1, by=gene),random=~(1+gene|pool\_id),  
 family=gaussian(),data=shiftnew)  
plot(gamm1$gam, pages=1)

 ## 2.2 BLMM