

# BCI: an R-based algorithm for computing bootstrap-t confidence interval for nucleic acid concentration by absolute real-time PCR



Sara Pizzamiglio and Paolo Verderio

Unit of Medical Statistics and Biometry, Fondazione IRCCS Istituto Nazionale Tumori, Milan - Italy

## Introduction

There are two major approaches of real-time PCR quantification: the absolute and the relative method. The latter evaluates the change in expression of the target gene relative to a reference gene, whereas the absolute method uses a standard curve to quantify unknown amount of nucleic acids in a target sample (1).

From a statistical view point, the standard curve corresponds to the simple linear regression model and thus the absolute quantification method allows to exploit the methodological background of the linear regression theory. As reported in Figure 1, the fitted standard curve is used as calibrator to estimate the unknown nucleic acid concentration in the target sample by means of a technique known as inverse regression.

Several approaches (2,3) have been proposed for constructing confidence interval (CI) in inverse regression (Delta method, Fieller's theorem, Bayesian method, etc.) and in biochemical applications the Delta method is commonly used to compute an approximate estimate of the CI. As the accuracy of this CI depends on the asymptotic normality of the concentration estimate, bootstrap methods appears to provide a reliable alternative to produce CI that does not depend on this assumption (4). In a recent work, starting from real PCR data and using 500 Monte Carlo simulations, we demonstrated that, among the various bootstrap method used to compute CIs, the most appropriate in terms of coverage probability is the bootstrap-t (5).

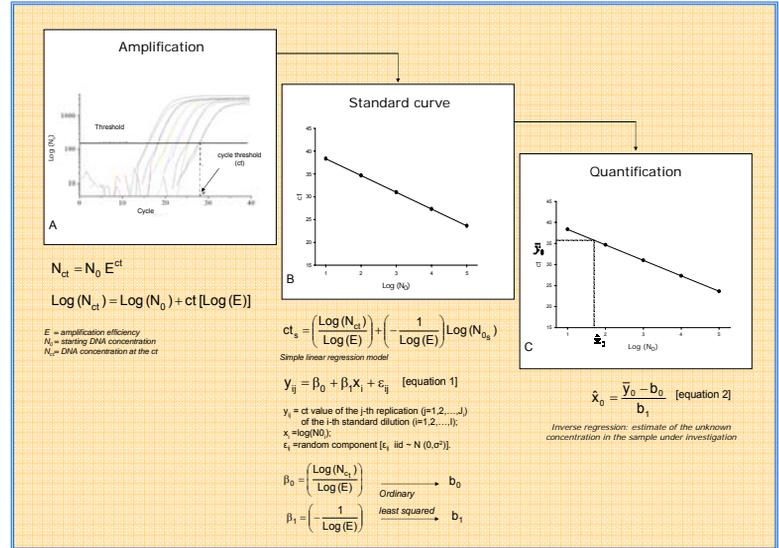


Figure 1: Steps for the absolute quantification of nucleic acids by real-time PCR

## Software implementation

We propose a user-friendly algorithm, named BCI (Bootstrap Confidence Interval), specifically designed to compute bootstrap-t confidence interval for nucleic acid concentration by absolute real-time PCR. The algorithm has been written in R language (<http://www.R-project.org>) an open-source statistical software.

The BCI algorithm works as follows:

1. starting from the threshold cycle (ct) values of the standard dilutions the linear regression model corresponding to the standard curve was fitted [equation 1] and, using the regression coefficients estimates ( $b_0$  and  $b_1$ ) together with the ct average ( $\bar{y}_0$ ) of the target sample, the unknown concentration was estimated [equation 2].

2. the residuals pool for the bootstrap resampling was generate as:

$$\tilde{r}_{ij} = \sqrt{n/(n-h)} * (y_{ij} - \hat{y}_i) \quad \text{where } n = \sum_{i=1}^I J_i \text{ and } h = \text{number of parameters}$$

$$\tilde{r}_k = \sqrt{K/(K-1)} * (y_{0k} - \bar{y}_0) \quad \text{where } y_{0k} = \text{ct value of the } k\text{-th replication (k=1,2,...,K) of the unknown sample}$$

$$\text{and } \bar{y}_0 = \frac{\sum_{k=1}^K y_{0k}}{K}$$

3. sampling with replacement B times from the pool of the above residuals, the n+K bootstrap residuals ( $\tilde{r}_{ij}^*$  and  $\tilde{r}_k^*$ ) were obtained for the computation of the bootstrap data as:

$$y_{ij}^* = b_0 + b_1 x_i + \tilde{r}_{ij}^* \quad \text{for the standard dilution}$$

$$y_{0k}^* = \bar{y}_0 + \tilde{r}_k^* \quad \text{for the unknown sample}$$

For each bootstrap sample, point 1 was repeated to obtain the bootstrap standard curve parameter estimates  $b_0^*$  and  $b_1^*$  and

the bootstrap estimate of  $x_0$ :  $\hat{x}_0^* = \frac{\bar{y}_0^* - b_0^*}{b_1^*}$  with  $\bar{y}_0^* = \frac{\sum_{k=1}^K y_{0k}^*}{K}$

4. the following pivotal statistic was computed for each bootstrap dataset to estimate the bootstrap-t CI:

$$t^* = \frac{\hat{x}_0 - \hat{x}_0^*}{\text{se}(\hat{x}_0^*)} \quad \text{where, according to Delta method, } \text{se}(\hat{x}_0) = \sqrt{\left(\frac{s^2}{b_1^2}\right) * \left(\frac{(\hat{x}_0 - \bar{x})^2}{s_{xx}} + \frac{1}{K} + \frac{1}{n}\right)}$$

$s^2 = \frac{\sum_{i=1}^I \sum_{j=1}^{J_i} (y_{ij} - \hat{y}_i)^2}{f}$  is the estimated error variance of the standard curve, with  $f = (n-h)$  degrees of freedom

$S_{xx}$  = sum of squares of the  $x_i$  values

$\bar{x}$  = mean of the  $x_i$  values

Computed in each bootstrap dataset

5. from the distribution of the  $t^*$  values the 100( $\alpha/2$ )-th and 100(1- $\alpha/2$ )-th percentile were used to estimate the lower and upper limits of the bootstrap-t CI as:

$${}_{BT} \hat{x}_L^* = \hat{x}_0 - t_{1-\alpha/2}^* \text{se}(\hat{x}_0)$$

$${}_{BT} \hat{x}_U^* = \hat{x}_0 + t_{\alpha/2}^* \text{se}(\hat{x}_0)$$

## Import

Data must be imported into BCI as Comma Separated Values (.csv) file using Microsoft Excel '97 or a more recent version (Figure 2).

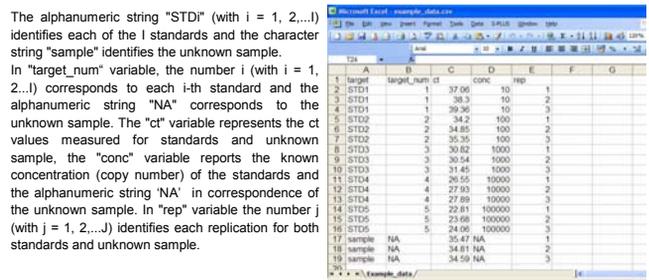


Figure 2: Screenshot of the .csv data file

## Output

The BCI algorithm provides the bootstrap estimate of the unknown concentration both in (common) logarithmic scale and in its original scale as copy number together with the lower and upper limits of the 100(1- $\alpha$ )% bootstrap-t CI of the unknown concentration in both logarithmic and original scale. Users can modify the number of bootstrap resampling and the confidence level (1- $\alpha$ ) of the bootstrap-t CI.

Figure 3 illustrate an example of output provided by BCI. The output reports a two-tailed 95% bootstrap-t CI obtained from B=999 bootstrap resampling.

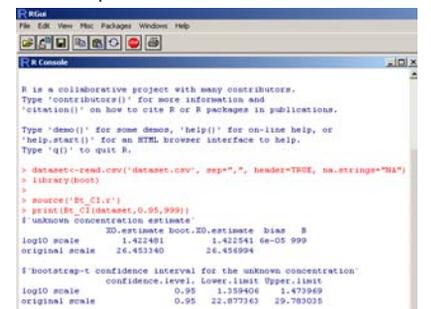


Figure 3: Example of BCI output

## References

1. Kubista M et al. The real-time polymerase chain reaction. Mol Aspects Med 2006; 27: 95-125.
2. Verderio P et al. FCI: an R-based algorithm for evaluating uncertainty of absolute real-time PCR quantification. BMC Bioinformatics. 2008;9:13.
3. Pizzamiglio S et al. Confidence Interval for DNA/mRNA concentration by real-time PCR. Int Biol Markers 2007; 22: 232-36.
4. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. London: Chapman and Hall, 1993;436pp.
5. Pizzamiglio S, Verderio P, Marubini E. Bootstrap confidence intervals for nucleic acid concentration by absolute real-time PCR. Biomedical Statistics and Clinical Epidemiology 2008; 2 (2): 109-115

Correspondence: [sara.pizzamiglio@istitutumorimi.it](mailto:sara.pizzamiglio@istitutumorimi.it)