

Software

Open Access

## FCI: an R-based algorithm for evaluating uncertainty of absolute real-time PCR quantification

Paolo Verderio<sup>1</sup>, Sara Pizzamiglio<sup>1</sup>, Fabio Gallo<sup>1</sup> and Simon C Ramsden\*<sup>2</sup>

Address: <sup>1</sup>Unit of Medical Statistics and Biometry, Istituto Nazionale dei Tumori, Milan, Italy and <sup>2</sup>National Genetics Reference Laboratory (Manchester), St. Mary's Hospital, Manchester, UK

Email: Paolo Verderio - [paolo.verderio@istitutotumori.mi.it](mailto:paolo.verderio@istitutotumori.mi.it); Sara Pizzamiglio - [sara.pizzamiglio@istitutotumori.mi.it](mailto:sara.pizzamiglio@istitutotumori.mi.it); Fabio Gallo - [fabio.gallo@istitutotumori.mi.it](mailto:fabio.gallo@istitutotumori.mi.it); Simon C Ramsden\* - [Simon.Ramsden@CMMC.nhs.uk](mailto:Simon.Ramsden@CMMC.nhs.uk)

\* Corresponding author

Published: 10 January 2008

Received: 4 October 2007

*BMC Bioinformatics* 2008, **9**:13 doi:10.1186/1471-2105-9-13

Accepted: 10 January 2008

This article is available from: <http://www.biomedcentral.com/1471-2105/9/13>

© 2008 Verderio et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** FCI is an R code for analyzing data from real-time PCR experiments. This algorithm estimates standard curve features as well as nucleic acid concentrations and confidence intervals according to Fieller's theorem.

**Results:** In order to describe the features of FCI four situations were selected from real data collected during an international external quality assessment program for quantitative assays based on real-time PCR. The code generates a diagnostic figure suitable for assessing the quality of the quantification process.

**Conclusion:** We have provided a freeware programme using this algorithm specifically designed to increase the information content of the real-time PCR assay.

### Background

Real-time PCR is widely used for the quantification of nucleic acids in a wide range of clinical and research applications including the measurement of gene dosage, detection of residual disease in haematological malignancies and detection of bacterial and viral infection.

Real-time PCR typically employs fluorescent probes which generate a signal that accumulates during PCR cycling in a manner proportional to the concentration of amplification products. Absolute quantification of a nucleic acid target can be achieved using a standard curve, which is generated by amplifying known amounts of the target DNA. The standard curve is typically generated using a series of 10-fold dilutions of a control template. For each dilution, replicated determinations of the cycle threshold (ct) are performed and a straight line is fitted to

the data by plotting the ct averages as a function of the logarithm of the starting concentration of the standards. By applying a technique known as "inverse regression," the straight line is used as a "calibrator" to estimate the unknown starting DNA concentration in the sample under examination.

As in any titration, evaluating the uncertainty in the estimated concentration of the unknown sample is critical for interpreting the data and optimizing experimental procedures. In addition to the point estimate it is important to calculate the confidence interval of the "true" value of the unknown concentration.

Commercial software specifically designed for generating standard curves and estimating nucleic acid concentrations is now available, however, there are currently no

freeware tools available for calculating the uncertainty associated with the concentration estimates.

Several approaches have been proposed for constructing confidence intervals in inverse regression [1-5], but the most frequently used method is due to Fieller [6]. Fieller's theorem provides a general procedure for the construction of confidence limits for certain ratios of parameters, most often applied to ratios of linear combinations of parameters (e.g. inverse estimate from a linear regression model). In this note we propose a statistical tool, FCI (Fieller's Confidence Interval), to estimate the confidence interval (CI) of the "true" value of each unknown concentration according to Fieller's theorem.

**Implementation**

FCI code has been tested on the WINDOWS platform with the R Software version 2.5.1 [7], an open-source statistical programming language. FCI implementation involves data import, FCI running and FCI output.

Data must be imported into FCI as Comma Separated Values (.csv) file [see Additional File 1] using Microsoft Excel '97 or a more recent version. Figure 1 shows a screenshot of real-time PCR data suitable for analysis by FCI.

Before running the FCI code it is necessary to save the file in the new directory C:\DATA and install the 'car' package from the Packages menu of R.

The FCI algorithm works as follows. First, the data are fitted to a linear regression model as reported in equation 1 and the corresponding analysis of variance table (Anova table, Table 1) is displayed in the output. Next, using the regression coefficient (Intercept and Slope) estimates, the FCI code provides the unknown concentration estimate (equation 2) both in (common) logarithmic scale and also in its original scale as copy number. Finally lower and upper limits of the 100(1- $\alpha$ )% Fieller's CI of the unknown concentration in both logarithmic and original scale are estimated as shown in [8]. By default FCI provides a two-tailed 95% confidence interval (significance level,  $\alpha =$

	A	B	C	D	E	F	G	H	I	J
1	target	target_num	ct	conc	rep					
2	STD1	1	37.06	10	1					
3	STD1	1	38.3	10	2					
4	STD1	1	39.36	10	3					
5	STD2	2	34.2	100	1					
6	STD2	2	34.85	100	2					
7	STD2	2	35.35	100	3					
8	STD3	3	30.82	1000	1					
9	STD3	3	30.54	1000	2					
10	STD3	3	31.45	1000	3					
11	STD4	4	26.55	10000	1					
12	STD4	4	27.93	10000	2					
13	STD4	4	27.89	10000	3					
14	STD5	5	22.81	100000	1					
15	STD5	5	23.68	100000	2					
16	STD5	5	24.06	100000	3					
17	sample	NA	35.47	NA	1					
18	sample	NA	34.81	NA	2					
19	sample	NA	34.59	NA	3					

**Figure 1**  
**Screenshot of PCR real time information file in .csv format.** The alphanumeric string "STDi" (with i = 1, 2,...I) identifies each of the I standards and the character string "sample" identifies the unknown sample under investigation. In "target\_num" variable, the number i (with i = 1, 2,...I) corresponds to each i-th standard and the alphanumeric string "NA" corresponds to the unknown sample. The "ct" variable represents the ct values measured for standards and unknown sample, while the "conc" variable reports the known concentration (copy number) of the standards. Note that, the alphanumeric string NA is inserted in correspondence of the unknown sample. In "rep" variable the number j (with j = 1, 2,...J) identifies each replication for both standards and sample.

**Table 1: Anova Table**

Source	df	SS	MS	F.value
Regression	$df_R = 1$	$SS_R = \sum_{i=1}^I J_i (\hat{y}_i - \bar{\bar{y}})^2$	$MS_R = \frac{SS_R}{df_R}$	$\frac{MS_R}{s_p^2}$
Error	$df_E = (I - 2) + \sum_{i=1}^I (J_i - 1)$	$SS_E = \sum_{i=1}^I \sum_{j=1}^{J_i} (y_{ij} - \hat{y}_i)^2$	$MS_E = \frac{SS_E}{df_E}$	
Lack of fit	$df_L = (I - 2)$	$SS_L = \sum_{i=1}^I J_i (\hat{y}_i - \bar{y}_i)^2$	$MS_L = \frac{SS_L}{df_L}$	$\frac{MS_L}{s_p^2}$
Pure error	$df_P = \sum_{i=1}^I (J_i - 1)$	$SS_P = \sum_{i=1}^I \sum_{j=1}^{J_i} (y_{ij} - \bar{y}_i)^2$	$s_p^2 = \frac{SS_P}{df_P}$	
Total corrected	$df_T = \left( \sum_{i=1}^I J_i \right) - 1$	$SS_T = \sum_{i=1}^I \sum_{j=1}^{J_i} (y_{ij} - \bar{\bar{y}})^2$		

Analysis of variance for the linear regression model underlying the standard curve. In the absence of repeated measurements of ct for all standard dilutions the Error Sum of Squares ( $SS_E$ ) can not be broken up into Lack of fit ( $SS_L$ ) and Pure error ( $SS_P$ ) Sum of Squares. df, degrees of freedom; SS, Sum of Square; MS, Mean Square;  $\hat{y}_i$ , predicted value corresponding to  $x_i$ ;  $\bar{y}_i$ , mean of the replicated values of ct measured for the i-th standard;

$$\bar{\bar{y}} = \frac{\sum_{i=1}^I \sum_{j=1}^{J_i} (y_{ij})}{\sum_{i=1}^I J_i}$$

0.05), however users can modify the confidence level (1- $\alpha$ ) by replacing 0.95 in the command line for calling the code with the chosen level.

The code performs a "Lack of fit" test on the fitted calibration model, the computation of 95% confidence interval for regression coefficients and, where appropriate, provides messages to assist in the interpretation of the results. In addition it generates a diagnostic plot enabling both an assessment of PCR assay quality and also a visual representation of the Fieller's CI graphical derivation [9].

**Results and Discussion**

In order to illustrate the functionality of FCI, we considered the data provided by four participants (laboratories A-D) in an international external quality assessment program (EQUAL-quant) for quantitative assays based on real-time PCR [10]. In this program laboratories received primers, labelled probe master mixture, plasmid standards (containing 10, 10<sup>2</sup>, 10<sup>3</sup>, 10<sup>4</sup>, 10<sup>5</sup> copies/5  $\mu$ L), and three samples of unknown concentration (test samples). Participants were required to measure the target copy number in all test samples and provide all ct values in triplicate. For the purposes of this illustration we consider only performance in the analysis of the test sample corre-

sponding to the lowest target copy number (56.21 copies/ 5  $\mu$ l calculated as median value of copy number from a total of 92 laboratories).

For laboratory A the linear model underlying the standard curve was inadequate. In this case the warning message "Lack of fit" appears in the output (Figure 2, panel A) as the p-value associated to the Lack of fit test is less than 0.05 (significance level  $\alpha = 0.05$ ). Although FCI code will provide the Fieller's confidence interval, caution should be taken when interpreting results and attempts should be made to discover where and how the inadequacy occurred. In fact, as we see from Figure 3, panel A, the replicates related to standards 4 and 5 appear inadequate to the linear fit.

For laboratory B the ct mean of the unknown sample ( $y_0$ ) lies outside the range of the standard dilutions used to fit the standard curve. In this situation (i.e. extrapolation) we cannot assume that the linear model on which the calibration is based will hold true outside the range of the standard dilutions. In this case the FCI code will not provide the Fieller's CI in the output (Figure 2, panel B), and the warning message "Out of standards range" appears in the legend of the corresponding figure (Figure 3, panel B).

```

RGui - [R Console]
File Edit Misc Packages Windows Help

$`anova table`
      df      SS      MS      F.value      P.value
Regression  1 336.74200 336.74200 2681.35155 <0.00001
Error       13  11.40297  0.87715
Lack of fit  3  10.14710  3.38237  26.93254  4e-05
  Pure Error 10  1.25587  0.12559
Total corrected 14 348.14497 24.86750

$message:`
[1] " Warning: Lack of fit"

$`Regression coefficient estimates table`
      Estimate Lower limit 95% CI Upper limit 95% CI
Intercept 39.45567          38.97753          39.93380
Slope     -3.35033          -3.49450          -3.20617

$`ct mean of the unknown sample`

y0 = 33.42

$`Unknown concentration estimate`

X0 = 1.80151
Conc = 63.31583

$`Fieller's confidence interval`
Confidence.level
0.95

$` `
X.lower X.upper
1.6414 1.95718

$` `
Conc.lower Conc.upper
43.79254 90.61042
    
```

**A**

```

RGui - [R Console]
File Edit Misc Packages Windows Help

$`anova table`
      df      SS      MS      F.value      P.value
Regression  1 378.34209 378.34209 2494.68504 <0.00001
Error       12  2.46235  0.20520
Lack of fit  3  1.09742  0.36581  2.41202  0.13406
  Pure Error  9  1.36493  0.15166
Total corrected 13 380.80444 29.29265

$`Regression coefficient estimates table`
      Estimate Lower limit 95% CI Upper limit 95% CI
Intercept 40.75309          40.21874          41.28743
Slope     -3.61642          -3.78021          -3.45263

$`ct mean of the unknown sample`

y0 = 40.10333

$`Unknown concentration estimate`

X0 = 1.80151
Conc = 63.31583

$`Fieller's confidence interval`
Confidence.level
0.95

$` `
X.lower X.upper
1.6414 1.95718

$` `
Conc.lower Conc.upper
43.79254 90.61042
    
```

**B**

```

RGui - [R Console]
File Edit Misc Packages Windows Help

$`anova table`
      df      SS      MS      F.value      P.value
Regression  1 244.36637 244.36637 1107.89059 <0.00001
Error       10  1.77720  0.17772
Lack of fit  3  0.23321  0.07774  0.35244  0.78908
  Pure Error  7  1.54398  0.22057
Total corrected 11 246.14357 22.37669

$`Regression coefficient estimates table`
      Estimate Lower limit 95% CI Upper limit 95% CI
Intercept 39.44469          38.69412          40.19526
Slope     -3.47383          -3.72061          -3.22704

$message:`
[1] " Optimal efficiency"

$`ct mean of the unknown sample`

y0 = 33.65333

$`Unknown concentration estimate`

X0 = 1.66714
Conc = 46.46656

$`Fieller's confidence interval`
Confidence.level
0.95

$` `
X.lower X.upper
1.44079 1.8825

$` `
Conc.lower Conc.upper
27.5925 76.29644
    
```

**C**

```

RGui - [R Console]
File Edit Misc Packages Windows Help

$`anova Table`
      df      SS      MS      F.value      P.value
Regression  1 406.05123 406.05123 699.48532 <0.00001
Error       13  6.00357  0.46181
Lack of fit  3  0.19857  0.06619  0.11402  0.94988
  Pure Error 10  5.80500  0.58050
Total corrected 14 412.05480 29.43249

$`Regression Coefficient estimates table`
      Estimate Lower limit 95% CI Upper limit 95% CI
Intercept 42.027          40.99903          43.05497
Slope     -3.679          -3.98894          -3.36906

$`ct mean of the unknown sample`

y0 = 34.95667

$`Unknown concentration estimate`

X0 = 1.92181
Conc = 83.52346

$`Fieller's confidence interval`
Confidence.level
0.95

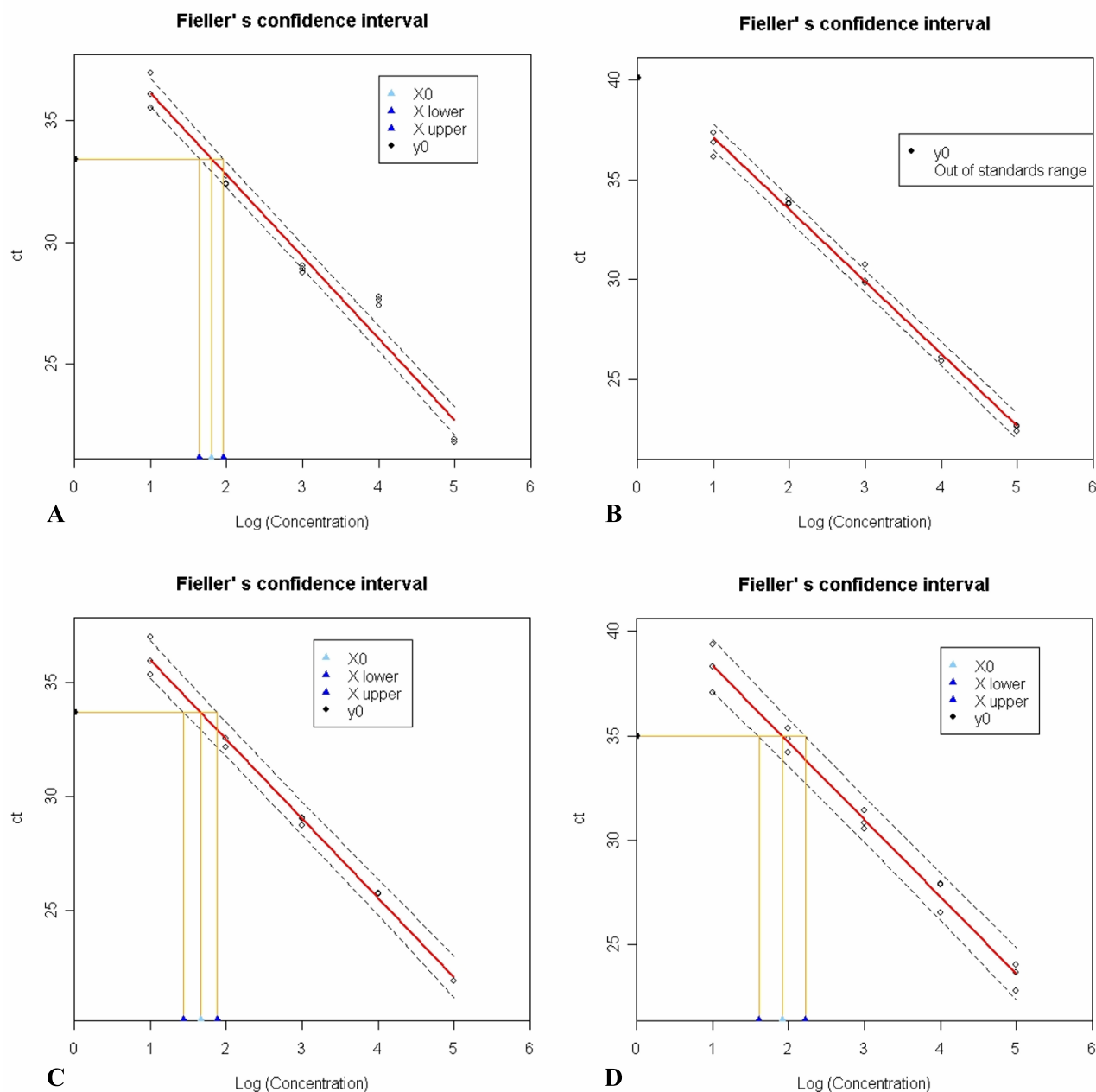
$` `
X.lower X.upper
1.60727 2.22094

$` `
Conc.lower Conc.upper
40.48234 166.3171
    
```

**D**

**Figure 2**

Panel A – D show the FCI output for example A-D, respectively. The FCI output provides the following information: Anova table, summarizes the results of the analysis of variance for the linear regression model underlying the standard curve; Regression Coefficient estimates tables, reports estimates of the standard curve parameters (Intercept and Slope) together with their 95% confidence interval; y0, ct mean of the unknown sample; X0, Unknown concentration estimate in common logarithmic scale; Conc, Unknown concentration estimate in its original scale as copy number; Confidence.level, the chosen confidence level (1- $\alpha$ ) of the Fieller's confidence interval; X.lower and X.upper, Lower and upper limits of the 100(1- $\alpha$ )% Fieller's confidence interval of the unknown concentration in logarithmic scale; Conc.lower and Conc.upper Lower and upper limits of the 100(1- $\alpha$ )% Fieller's confidence interval of the unknown concentration in original scale as copy number.



Key to the figure: ▲ is the unknown concentration estimate, X0.; ▲ are the estimates of the lower and upper 100(1-α)% confidence limits of the unknown concentration, X.lower and X.upper; ● is the ct mean of the unknown sample, y0.

**Figure 3**  
**Fieller's confidence interval.** Panel A-D show the figure provided by FCI for examples A-D, respectively. In each figure the red continuous line is the estimate standard curve and the two black dashed curves are the estimated limits of its 100(1-α)% prediction interval. A horizontal line parallel to the X-axis is drawn at the height of the mean of the ct values (y0) measured in the unknown sample. Where this line intercepts the two dashed curves, as well as the standard curve, three perpendicular lines are traced onto the X-axis giving the lower and the upper 100(1-α)% confidence limits (X.lower and X.upper) of the unknown concentration together with its estimate (X0).

For laboratory C the 95% CI of the true value of the slope includes the theoretical expected slope value (-3.32193) corresponding to 100% amplification efficiency. In such a situation the message "Optimal efficiency" appears in the output (Figure 2, panel C). This message appears only when no warning messages have been displayed. Figure 3, panel C reports the pertinent graph.

Laboratory D illustrates the situation in which no messages, either in the output or in the figure, are displayed (Figure 2, panel D; Figure 3, panel D).

When the 95% confidence interval of the true value of the slope includes zero (an unlikely eventuality in a real-time PCR experiment) the warning message "Not real confidence interval" appears in the FCI output. In this situation, not observed in our real data, the resulting Fieller's confidence interval is infinite in extent and the data have to be considered as valueless for estimating the unknown concentration.

## Conclusion

We have described an algorithm for the computation of CIs based on Fieller's theorem in the context of real-time PCR quantification. While the algorithm is designed in this case for real-time PCR experiments, it is easily adapted to other assays based on inverse estimation from a straight line model (calibration). FCI estimates the standard curve, the unknown sample concentration and its uncertainty. Furthermore it provides an insightful diagnostic figure.

## Methods

The statistical model corresponding to the standard curve is :

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij} \quad (1)$$

where  $y_{ij}$  specifies the value of ct measured for the  $j$ -th replication ( $j = 1, 2, \dots, J_i$ ) at the  $i$ -th standard ( $i = 1, 2, \dots, I$ ),  $x_i$  defines the logarithm of the starting DNA/cDNA concentration of the  $i$ -th standard and  $\varepsilon_{ij}$  is the random component assumed to be normally distributed with mean zero and constant error variance  $\sigma^2$ . The estimates  $b_0$  and  $b_1$  of  $\beta_0$  and  $\beta_1$  respectively are obtained by the Ordinary Least Squared method.

The uncertainty associated to linear regression model underlying the standard curve can be broken up according to the analysis of variance (Anova) as reported in Table 1[11].

The value of interest in a real-time PCR experiment is the logarithm of the unknown DNA/cDNA concentration in

the sample under examination. The latter ( $x_0$ ) is usually estimated by resorting to the inverse regression, as:

$$\hat{x}_0 = \frac{\bar{y}_0 - b_0}{b_1} \quad (2)$$

where  $\bar{y}_0$  is the mean of the  $K$  replicated values of ct [ $ct_k = y_{k'}$  ( $k = 1, 2, \dots, K$ )] measured for the sample under examination.

As shown in Verderio *et al.* [8] the limits of  $100(1-\alpha)\%$  confidence interval of  $x_0$  according to Fieller's theorem are obtained as roots of a second degree equation.

## Availability and requirements

For the current version of FCI code please see Additional file 2. The commands for importing data, running the code and displaying the output are provided in Additional file 3. Further development as well as extended version of FCI code will be available through our website: [http://www.istitutotumori.mi.it/INT/dipartimento/UO/biome tria](http://www.istitutotumori.mi.it/INT/dipartimento/UO/biome%20tria). FCI code use R Software <http://www.r-project.org>, an open-source statistical programming language.

## Authors' contributions

All authors contributed to the work described in the manuscript. The R-implementation of the FCI code as well as its testing was carried out by FG with advice and supervisor from PV and SP. PV provided oversight of the work, finalized the draft and optimized the final version. SR supplied the data for testing the code and improved the writing. All authors read and approved the manuscript.

## Additional material

### Additional file 1

A .csv file containing the data of example D on which are based the output in figure 2, panel D and the graph in figure 3, panel D.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-13-S1.csv>]

### Additional file 2

A text file containing the FCI code written in R language.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-13-S2.R>]

### Additional file 3

A text file containing the commands for importing data, calling the code and displaying the output.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-9-13-S3.R>]

## Acknowledgements

EQUAL-quant has been supported by the EU Sixth Framework Program (Contract 504842). This note was supported by the "Sovvenzione Globale INGENIO" awarded by the "Fondo Sociale Europeo, Ministero del Lavoro e della Previdenza Sociale and Regione Lombardia".

## References

1. Hoadley B: **A Bayesian look at inverse linear regression.** *J Am Stat Assoc* 1970, **65**:356-369.
2. Cox C: **Fieller's theorem, the likelihood and the delta method.** *Biometrics* 1990, **46**:709-718.
3. Krutchkoff RG: **Classical and inverse regression methods of calibration.** *Technometrics* 1967, **9**:425-439.
4. Grybill FA: *Theory and applications of the linear model* Massachussets USA: Duxbury press; 1976.
5. Wu CFJ: **Jackknife, bootstrap and other resampling methods in regression analysis.** *Ann Statist* 1986, **14**:1261-1295.
6. Fieller EC: **The biological standardization of insulin.** *J R Stat Soc* 1940:1-64.
7. Ihaka R, Gentleman R: **R: A language for data analysis and graphics.** *J Comp Graph Stat* 1996, **5**:299-314.
8. Verderio P, Orlando C, Casini Raggi C, Marubini E: **Confidence interval estimation for DNA and mRNA. concentration by real-time PCR: A new environment for an old theorem.** *Int J Biol Markers* 2004, **19**:76-9.
9. Pizzamiglio S, Verderio P, Orlando C, Marubini E: **Confidence Interval for DNA/mRNA concentration by real-time PCR.** *Int J Biol Markers* 2007, **22**:232-6.
10. Ramsden SC, Daly S, Geilenkeuser WJ, Duncan G, Hermitte F, Marubini E, Neumaier M, Orlando C, Palicka V, Paradiso A, Pazzagli M, Pizzamiglio S, Verderio P: **EQUAL-quant: An International External Quality Assessment Scheme for Real-Time PCR.** *Clin Chem* 2006, **52**:1584-91.
11. Draper NR, Smith H: *Applied Regression Analysis* Third edition. New York: John Wiley & Sons; 1998.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

