The International Journal of Biological Markers, Vol. 19 no. 1, pp. 76-79
© 2004 Wichtig Editore

TUTORIAL BIOMETRY CORNER

Confidence interval estimation for DNA and mRNA concentration by real-time PCR: A new environment for an old theorem

P. Verderio[1], C. Orlando[2], C. Casini Raggi[2], E. Marubini[3]

[1]Operative Unit of Medical Statistics and Biometry, Istituto Nazionale per lo Studio e la Cura dei Tumori, Milan
[2]Clinical Biochemistry Unit, Department of Clinical Physiopathology, Università di Firenze, Florence
[3]Institute of Medical Statistics and Biometry, Università degli Studi di Milano, Milan – Italy

INTRODUCTION

Real-Time PCR (RT-PCR) is becoming the method of choice for quantification of minute amounts of nucleic acids. Different applications of real-time approaches have been widely reviewed (1-5). 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. By this principle the measurement of fluorescence in each sample provides an homogeneous signal which is specifically associated with the amplified target and quantitatively related to the amount of PCR products (1). Fluorescent monitoring of DNA/cDNA amplification is the basis of real-time PCR: the target DNA/mRNA concentration can be determined from the fractional cycle where a threshold amount of amplified DNA/cDNA is produced. The latter is defined as threshold cycle (c_t) and corresponds to the number of amplification cycles required to generate enough fluorescent signal to reach the threshold (1). These c_t values are directly proportional to the amount of starting template and are the basis for quantification of target DNA/mRNA concentration. Specifically, absolute quantification of DNA/mRNA target can be achieved using the so-called standard curve, which is constructed by amplifying known amounts of DNA/cDNA. To generate the standard curve, a set of 10-fold dilutions of a positive control template is used as standard. For each dilution, replicated determinations of c_t are performed and a straight line is fitted to the data by plotting the c_t averages as function of the logarithm of their known starting concentration. Finally, by applying a technique known as "inverse regression", the straight line is used as a "calibrator" to estimate the unknown starting DNA/cDNA concentration. As in any titration, the biologist needs to know the confidence interval of the "*true*" value of the unknown starting concentration. Surprisingly, as far as the authors of this note know, the confidence intervel is not usually provided in the papers dealing with real-time PCR determinations of DNA/mRNA. This could perhaps be attributed to the fact that no straightforward method for the computation of such interval is available. In this note a suitable statistical tool, based on the Fieller's theorem (6), is suggested to address this issue by resorting to the calculation of the roots of a second degree equation to attain the limits of the above mentioned confidence interval. The methodological background of the indirect titration approach is presented together with the introduction and justification of Fieller's theorem and a detailed example of computation on real data is provided.

METHODOLOGICAL BACKGROUND

Standard Curve

The statistical model corresponding to the equation (a4) (appendix) is the simple linear regression model:

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

where y_{ij} specifies the value of c_{ts} determined for the j – th replication ($j = 1,2,\ldots,J_i$) at $x_i[=(\text{Log}N_{0s})_i]$ ($i = 1,2,\ldots,I$) standard dilution and ε_{ij} is the corresponding random component assumed to be normally distributed and homoscedastic with common error variance σ^2. Any statistical package employed in biological laboratories allows the user to obtain the estimates b_0 and b_1 of β_0 and β_1, respectively. By these estimates the generic equation of the reverse regression is:

$$\hat{x} = \frac{(y - b_0)}{b_1} \qquad (2)$$

A sample of unknown log starting DNA/mRNA concentration (x_u) provides K replicated values of $c_{tk} = y_k$ ($k = 1,2,\ldots,K$). Therefore, by inserting in equation (2) the mean ($\bar{y}$) of these determinations, the estimate ($\hat{x}_u$) of the unknown log starting concentration is obtained.

0393-6155/076-4$15.00/0

The "pure error" variance can be estimated from both the standard and the unknown preparations as:

$$s_p^2 = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i}(y_{ij} - \bar{y}_i)^2 + \sum_{k=1}^{K}(y_k - \bar{y})^2}{\sum_{i=1}^{I}(J_i - 1) + (K - 1)} \quad (3)$$

where $\bar{y}_i$ is the mean of the c_{ts} values at the i – th standard dilution.

After defining $s_{xx} = \sum_{i=1}^{I} J_i (x_i - \bar{\bar{x}})^2$

where:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{I} J_i x_i}{\sum_{i=1}^{I} J_i},$$

the variances and covariance of b_0 and b_1 are, respectively:

$$\text{var}(b_0) = s_p^2 \left(\frac{\sum_{i=1}^{I} J_i x_i^2}{\sum_{i=1}^{I} J_i s_{xx}} \right) \quad (4)$$

$$\text{var}(b_1) = \frac{s_p^2}{s_{xx}} \quad (5)$$

$$\text{cov}(b_0, b_1) = \frac{-s_p^2 \bar{\bar{x}}}{s_{xx}} \quad (6)$$

Fieller's theorem

Equation (2) shows why the confidence interval of the *"true"* log of the starting DNA/mRNA concentration (x_u) cannot be computed by a straightforward method. Infact, as $\hat{x}_u$ is the ratio of random variables, its standard error cannot be computed directly but only by approximated approaches based on the delta method.

A better alternative to compute the aforementioned confidence interval is by using the Fieller's theorem (6). Let:

$$u = (\bar{y} - b_0) - b_1 \hat{x}_u \quad (7)$$

and

$$\text{var}(u) = \text{var}(\bar{y}) + \text{var}(b_0) + \hat{x}_u^2 \text{var}(b_1) - 2\hat{x}_u \text{cov}(b_0, b_1) \quad (8)$$

where $\text{var}(\bar{y}) = \frac{s_p^2}{K}$

By leaving $\hat{x}_u$ unspecified in equations (7) and (8), one can write:

$$u^2 - t_{f;1-\alpha/2}^2 \text{var}(u) = 0 \quad (9)$$

where $t_{f;1\text{-}\alpha/2}$ is the critical value corresponding to a prefixed 100(α/2)% level of Student's t-distribution with f degree of freedom (d.f.).

Expression (9) can be made explicit as a second degree equation in x. The coefficients of this equation in its canonical form:

$$Ax^2 + Bx + C \quad (10)$$

are:

$$A = b_1^2 - t_{f;1-\alpha/2}^2 \text{var}(b_1)$$

$$B = -2\bar{y}b_1 + 2b_0 b_1 + 2t_{f;1-\alpha/2}^2 \text{cov}(b_0, b_1)$$

$$C = \bar{y}^2 + b_0^2 - 2\bar{y}b_0 - t_{f;1-\alpha/2}^2 \text{var}(\bar{y}) - t_{f;1-\alpha/2}^2 \text{var}(b_0)$$

The two roots of equation (10) are the required 100(1–α)% confidence limits of x_u.

TABLE I - c_t DETERMINATIONS FOR THE SIX STANDARD DILUTIONS

Standard	Determination 1	Determination 2	Determination 3
S1	18.72	18.77	18.84
S2	22.10	22.17	22.22
S3	25.48	25.50	25.91
S4	28.96	29.13	29.20
S5	32.51	32.58	32.59
S6	36.12	36.13	36.27

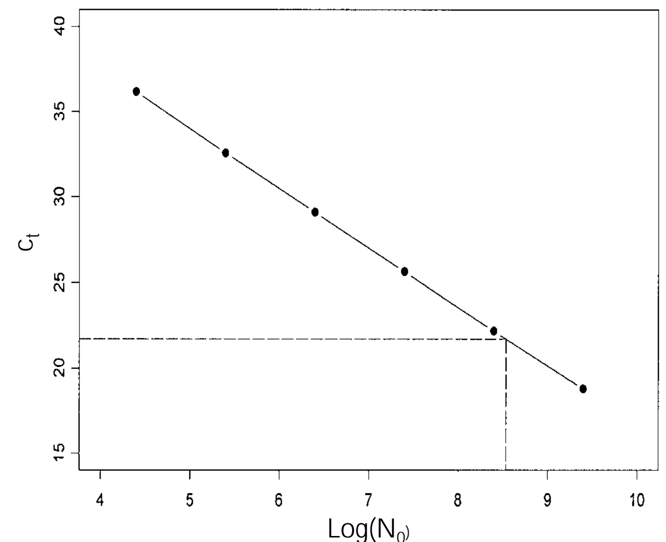

Fig. 1 - *Standard curve corresponding to the data in Table I.*

AN EXAMPLE

The example is based on real data collected from one of the 42 laboratories taking part in a national program of external quality control for quantitative assays based on real-time PCR. In this study, sponsored by the Italian Ministry of Health, standard DNA solution was provided to each laboratory at 2.5×10^9 copies DNA/5 µL (S1), which represented the highest point of the standard curve. Five additional points were prepared by serial 1:10 dilutions to obtain the following concentrations: $S2=2.5\times10^8$, $S3=2.5\times10^7$, $S4=2.5\times10^6$, $S5=2.5\times10^5$, $S6=2.5\times10^4$. For each dilution, three determinations of c_t were performed; the pertinent data are given in Table I. Using a standard statistical package (SAS/STAT software) (7), the regression coefficient estimates were:

$$b_0 = 51.37330$$
$$b_1 = -3.47543$$

The corresponding standard curve is plotted in Figure 1 together with the mean values (•) for each dilution. c_t values provided for the sample of unknown DNA concentration were 21.81, 21.60 and 21.73 with $\bar{y}=21.71$. From equation (2) the result is:

$$\hat{x}_u = \frac{21.71 - 51.37330}{-3.47543} = 8.53515$$

According to equation (3):

$$s_p^2 = \frac{0.18067 + 0.02250}{12 + 2} = 0.01451,$$

and coherently with equations:

$$(4),\ \text{var}(b_0) = 0.01451 \times \left(\frac{908.96837}{945}\right) = 0.01396\ ;$$

$$(5),\ \text{var}(b_1) = \frac{0.01451}{52.5} = 0.00028\ ;$$

$$(6),\ \text{cov}(b_0, b_1) = \frac{-0.01451 \times 6.89794}{52.5} = -0.00191\ .$$

In this example, f = 12+2=14 d.f.. The resulting Student's t is:

$$t_{14;1-0.05/2} = 2.14479\ .$$

All the ingredients to obtain A, B and C of equation (10) are now available, namely:

$$A = -3.47543^2 - 2.14479^2 \times 0.00028 = 12.07734\ ;$$

$$B = [-2 \times 21.71 \times (-3.47543)] + [2 \times 51.37330 \times (-3.47543)] + \left[2 \times 2.14479^2 \times (-0.00191)\right] = -206.20299\ ;$$

$$C = 21.71^2 + 51.37330^2 - (2 \times 21.71 \times 51.37330) - \left(2.14479^2 \times \frac{0.01451}{3}\right) - (2.14479^2 \times 0.01396) = 879.82490\ .$$

The 95% confidence limits of x_u, i.e. the roots of equation (10), are: 8.37176 |---| 8.70178.
The above results have been obtained from SAS (7) by rounding at the 5th decimal figure.

CONCLUDING REMARKS

The 95% confidence limits of the *"true"* c_t value of the unknown sample in our example, are:

$$\bar{y} \pm t_{14;1-0.05/2}\sqrt{\frac{s_p^2}{K}} = 21.71 \pm 2.14479\sqrt{\frac{0.01451}{3}} =$$

$$= 21.56083 \text{ |---| } 21.85917\ .$$

Naively, one could argue that the 95% confidence limits of x_u could be computed straightforwardly by inserting each of these values in equation (2), thus obtaining: 8.49222 |---| 8.57807. Unfortunately this approach is fallacius because of it does not make allowance for the uncertainities inherent in the estimation of the standard straight line, which the Fieller's theorem, on the other hand, does.

APPENDIX

The basic equation describing RT-PCR kinetics is:

$$N_c = N_0 E^c \qquad \text{(a1)}$$

where N_c is the template concentration at cycle c, N_0 is the starting template concentration and E is the amplification efficiency. The latter can be thought of as the yield of the amplification reaction and 2 is its ideal value, corresponding to a yield of 100%: in this case every molecule is duplicated at each cycle so that the template concentration at cycle c would be twice the template concentration at cycle c – 1. To establish a direct connection of the above relationship with the standard dilutions it is convenient to rewrite equation (a1) as:

$$N_{c_{ts}} = N_{0_s} E^{c_{ts}} \qquad \text{(a2)}$$

where the subscripts: t means *threshold* and s means *standard preparation*.

By taking the common logarithms, equation (a2) becomes:

$$\mathrm{Log}(N_{c_{ts}}) = \mathrm{Log}(N_{0_s}) + c_{ts}\mathrm{Log}(E) \quad \text{(a3)}$$

thus:

$$c_{ts} = \frac{\mathrm{Log}(N_{c_{ts}})}{\mathrm{Log}(E)} - \frac{1}{\mathrm{Log}(E)}\mathrm{Log}(N_{0_s}). \quad \text{(a4)}$$

The latter is the equation of a straight line describing the relationship between the dependent variable $c_{ts} = y$ and the independent variable $\mathrm{Log}(N_{0_s}) = x$ with the intercept $\beta_0 = \frac{\mathrm{Log}(N_{c_{ts}})}{\mathrm{Log}(E)}$ and slope $\beta_1 = \frac{1}{\mathrm{Log}(E)}$.

Address for correspondence:
Dr. Paolo Verderio
Unità Operativa di Statistica Medica e Biometria
Istituto Nazionale per lo Studio e la Cura dei Tumori
Via Venezian 1
20133 Milano, Italy
e-mail: paolo.verderio@istitutotumori.mi.it

REFERENCES

1. Orlando C, Pinzani P, Pazzagli M. Developments in quantitative PCR. Clin Chem Lab Med 1998; 36: 255-69.
2. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 2000; 25: 169-93.
3. Ginzinger DG. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. Exp Hematol 2002; 30:503-12.
4. Jung R, Soondrum K, Neumaier M. Quantitative PCR. Clin Chem Lab Med 2000; 38: 833-6.
5. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 2002; 29: 23-39.
6. Fieller EC. The biological standardization of insulin. J R Stat Soc 1940; Suppl 7: 1-64.
7. SAS Institute Inc., SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc., 1999.

Received: January 12, 2004
Accepted: February 27, 2004