The International Journal of Biological Markers, Vol. 19 no. 2, pp. 141-146

Statistical diagnostics emerging from external quality control of real-time PCR

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

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

On behalf of the Italian Network for Quality Assessment of Tumor Biomarkers (INQAT) and the Italian Society of Clinical Chemistry and Clinical Molecular Biology (SIBioC).

ABSTRACT: Besides the application of conventional qualitative PCR as a valuable tool to enrich or identify specific sequences of nucleic acids, a new revolutionary technique for quantitative PCR determination has been introduced recently. It is based on real-time detection of PCR products revealed as a homogeneous accumulating signal generated by specific dyes. However, as far as we know, the influence of the variability of this technique on the reliability of the quantitative assay has not been thoroughly investigated. A national program of external quality assurance (EQA) for real-time PCR determination involving 42 Italian laboratories has been developed to assess the analytical performance of real-time PCR procedures. Participants were asked to perform a conventional experiment based on the use of an external reference curve (standard curve) for real-time detection of three cDNA samples with different concentrations of a specific target. In this paper the main analytical features of the standard curve have been investigated in an attempt to produce statistical diagnostics emerging from external quality control. Specific control charts were drawn to help biochemists take technical decisions aimed at improving the performance of their laboratories. Overall, our results indicated a subset of seven laboratories whose performance appeared to be markedly outside the limits for at least one of the standard curve features investigated. Our findings suggest the usefulness of the approach presented here for monitoring the heterogeneity of results produced by different laboratories and for selecting those laboratories that need technical advice on their performance. (Int J Biol Markers 2004; 19: 141-6)

Key words: Statistical quality control, Real-time PCR, Standard curve

INTRODUCTION

Besides the traditional application of conventional qualitative PCR as a valuable tool to enrich or identify specific sequences of nucleic acids, a new revolutionary technique for quantitative PCR has been introduced recently, starting from the first experiments with real-time PCR in 1993 (1). This technique alters the original concept of molecular techniques serving mainly to provide qualitative information (i.e., the presence or absence of a genetic feature), but at the same time opens a new debate on different aspects connected to quantitative PCR. Various approaches can be followed in the choice of the fragment to amplify, the labeling of probes, the quantitative strategy, the type of external standard to be used, etc. In addition to these general problems, the analytical aspects of the assay method should be accurately checked and monitored.

External standard quantification is the method of choice for nucleic acid quantification, particularly mRNA quantification. It allows absolute and accurate quantification down to a few molecules (<10 molecules) (2). The dynamic range of a validated and optimized, externally standardized real-time PCR assay can accurately detect target mRNA up to nine orders of magnitude or a billion-fold range (3), generally with low variation (4). The reliability of an absolute real-time PCR assay depends on the presence of "identical" amplification efficiencies for both the native target and the calibration curve in the reaction and subsequent kinetic PCR (2, 3).

Figure 1 illustrates the three main analytical steps involved in this technique: amplification of standard dilu-

0393-6155/141-6$15.00/0

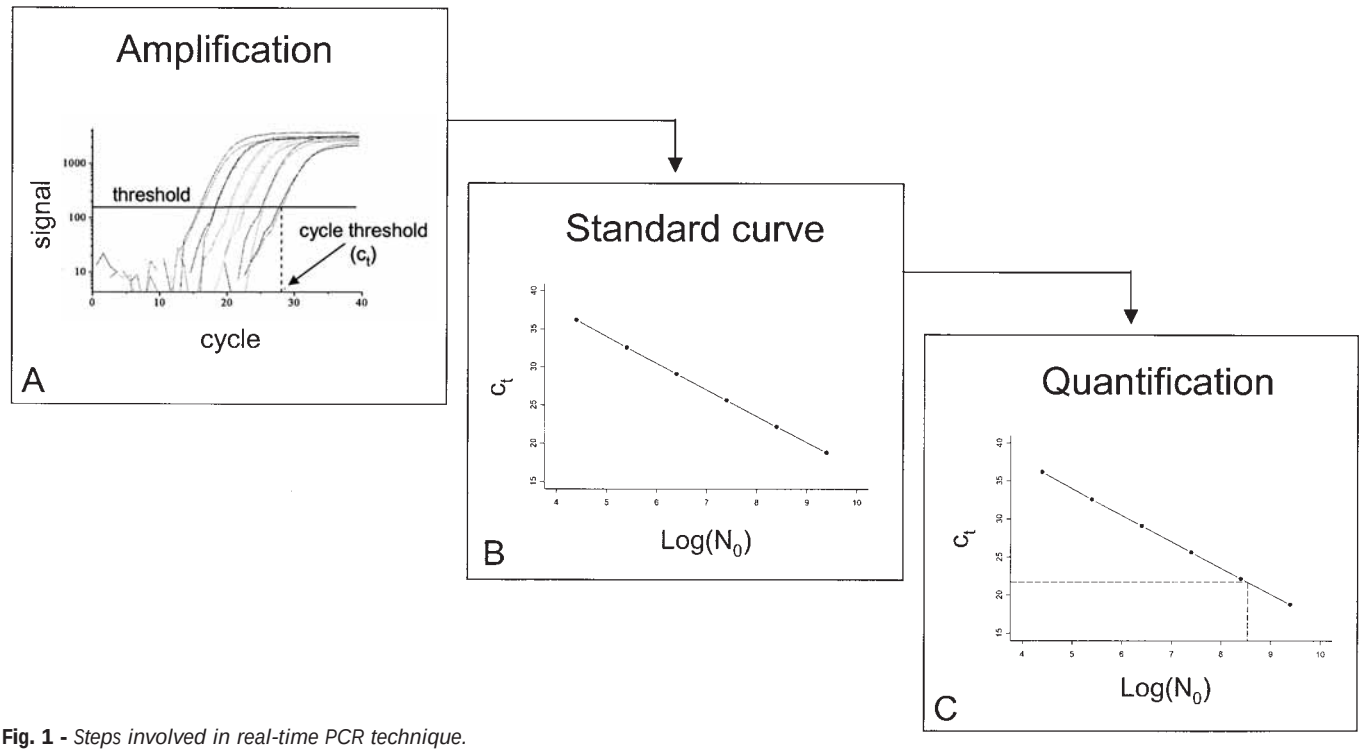

Fig. 1 - *Steps involved in real-time PCR technique.*

tions (panel A), generation of the standard curve (panel B), and target quantification in unknown samples (panel C). Typically, real-time PCR employs fluorescent probes which generate a signal that accumulates during PCR cycling at a rate proportional to the concentration of amplification products. Specifically, absolute quantification of nucleic acid targets can be achieved using a standard curve constructed by amplifying known amounts of 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 the so-called threshold cycle c_t (defined as the fractional cycle where a threshold amount of amplified cDNA is produced) are performed and a straight line is fitted to the data by plotting the c_t averages as a function of the logarithm of the known starting concentration of the standard [$Log(N_0)$]. 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 of different samples to be tested. Figure 1 shows that in the whole analytical process a fundamental role is played by the generation of the standard curve. In fact, the target concentration of the unknown sample, the only measure of interest from a clinical point of view, relies upon this straight line.

As with any biological marker based on complex technology, also in this case it is crucial to use external quality control programs to establish the influence of the variability of the technique on the reliability of the quantitative assay. A good quality control program should cover two principal aspects: monitoring and process modifying. With regard to real-time PCR, the first aspect, monitoring, concerns the quantification of unknown samples, i.e., the final measurements provided by each laboratory. In this case external quality control allows monitoring of the actual practice of measurements on a nationwide scale. The second aspect, process modifying, should be based on the standard curve features. Here the aim is to obtain indications that will help to understand the reasons for the possible poor performance of some laboratories and plan the necessary corrective action. In an attempt to cover both aspects we developed an EQA program to check the analytical performance related to real-time PCR procedures based on TaqMan™ probes. The results concerning the first aspect have been submitted elsewhere (5). In this paper we report the results related to the second aspect. The main analytical features of the standard curve were investigated in order to produce statistical diagnostics that could serve as a basis for modification of the assay procedure.

MATERIALS AND METHODS

Participants and exclusion criteria

Forty-eight laboratories joined the first round of external quality control and received the reference material. The experiment was completed within the assigned time by 42 of them including 16 public hospital labora-

tories, 13 academic laboratories, 11 private laboratories and two industrial laboratories. The experiment was performed in six AB5700 ABI PRISM® Sequence Detection Systems (SDS) from Applied Biosystems (Foster City, CA, USA), two AB7000s, 14 AB7700s, six AB7900s, 13 iCycler iQ Real-Time PCR Detection Systems (Hercules, CA, USA), and one DNA Engine Opticon® from MJ Research, Inc. (Boston, MA, USA). Laboratories were identified with a progressive random numeration to guarantee anonymity. It should be noted that the quality control schedule adopted here implies the following two preliminary exclusion criteria:

1. laboratories providing at least one no-template control (NTC) different from zero;
2. laboratories not providing values for one of the six standards.

Two (L24 and L30) and six laboratories (L25, L32, L34, L37, L40 and L41) were excluded according to the first and second criteria, respectively. Thus, of the data provided by the original 42 participating laboratories, only those of 34 laboratories were taken into account in the analysis.

Reagents

Each participating laboratory received by express courier a kit containing five vials. Vial #1 contained 70 µL of a standard cDNA solution ($5x10^6$ cDNA copies/µL) obtained by in vitro transcription of a fragment of the hTERT (telomerase catalytic subunit) gene cloned in a pCR™ II vector (Invitrogen Corporation, San Diego CA). Vial #2 included a mix of primers (300 nM each) and probe (300 nM). The mix (280 µL) was ready for use and contained the following primers and probe: forward primer 5'-ACGGCGACATGGAGAACAA-3' (2592-2610), reverse primer 5'-CACTGTCTTCCGCAAGTTCAC-3' (2767-2787) and fluorescent probe (FAM) 5'-CTCCTGCGTTTGGTGGATGATTTCTTGTTG-3' (2642-2671) according to the hTERT gene sequence AF015950 (GenBank).

Vials #3, #4 and #5 contained unknown cDNA samples from three pools of total RNA extracted from different human cancers to obtain samples with high, intermediate and low levels of hTERT mRNA expression. These samples (45 µL each) were ready for use.

Instructions and actions

Participants received detailed instructions for preparation of the experiment. The standard cDNA solution was provided at 2.5×10^7 cDNA copies/5 µL (STD1), which represented the highest point of the standard curve. Five additional points were prepared by participants by serial 1:10 dilutions in double distilled water to obtain the following concentrations: STD2 = 2.5×10^6, STD3 = 2.5×10^5, STD4 = 2.5×10^4, STD5 = 2.5×10^3, and STD6 = 2.5×10^2 cDNA copies/5 µL. Five µL of each dilution was pipetted in triplicate onto the microplate. Five µL of double distilled water in triplicate was used as NTC. For each unknown sample, 5 µL of solution contained in tubes #3, #4 and #5 was transferred in triplicate onto the plate. A master mix was prepared from the material of each participant using reagents for real-time experiments commonly used in the respective laboratories, containing 3.5 µL/well of primer/probe solution and H_20 to a final volume of 25 µL/well. The instrument was set to read FAM as the reporter dye and TAMRA as the quencher dye. The amplification protocol was: 10 min at 95°C for one cycle; 15 secs at 95°C and 1 min at 60°C for 40 cycles.

Reports and evaluation

At the end of the amplification each participant analyzed the results with the procedure used in their laboratories. Participants were free to define the cycle interval for calculation of the fluorescence baseline and to set the threshold value for c_t calculations. The complete data sheet and the experiment report were faxed or emailed to the organizers for data collection. In addition, participants provided the c_t value of each well for NTC, standard and unknown samples, respectively. Finally the concentration for each unknown sample was calculated.

Statistical methods

The approach adopted to assess the different features of the standard curves provided by the laboratories was based on graphic devices. After defining the relevant statistics, four control charts were drawn. In each of these charts specific limits were given to aid biochemists in taking possible technical decisions to improve the performance of their laboratory. Four features were investigated, including:

i) Variability measurement

This statistic, here defined as "standardized standard deviation", is the square root of the ratio of the "pure error" variance of each laboratory, s_i^2 (i = 1,2,3,…,34), with 12 degrees of freedom (d.f.) to the "pure error" variance (s_p^2) pooled on the whole set of laboratories (408 d.f.). The two horizontal lines in the chart (Fig. 2), corresponding to the 97.5*th* centile (upper line) and the 2.5*th* centile (lower line), represent the action limits.

ii) Standard curve accuracy

This statistic is the standardized difference $(b_{1_i} - \beta_1)/ SE(b_{1_i})$, where b_{1_i} is the slope of the standard curve provided by the i-*th* laboratory together with its standard error $SE(b_{1_i})$ and β_1 is the expected value corresponding to a 100% yield of amplification efficiency ($\beta_1 = -3.32193$). The two horizontal lines in the chart (Fig. 3), corresponding

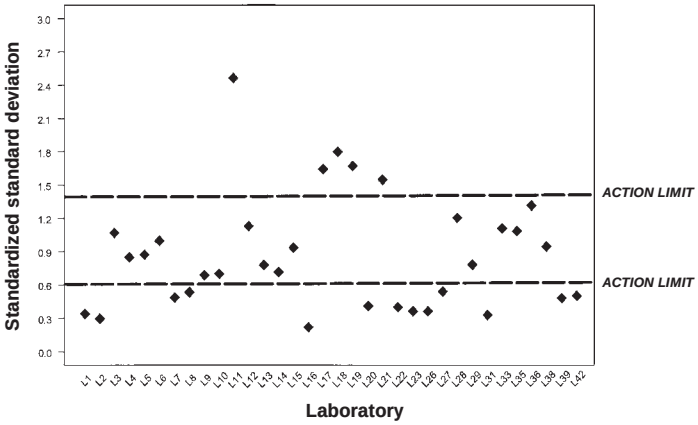

Fig. 2 - *Variability measurement.*

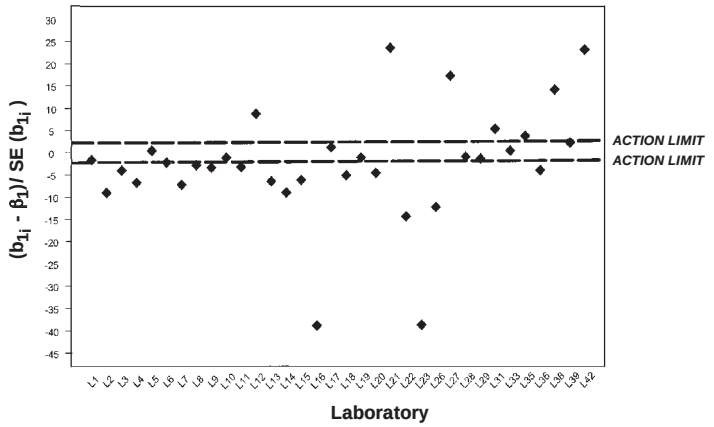

Fig. 3 - *Standard curve accuracy.*

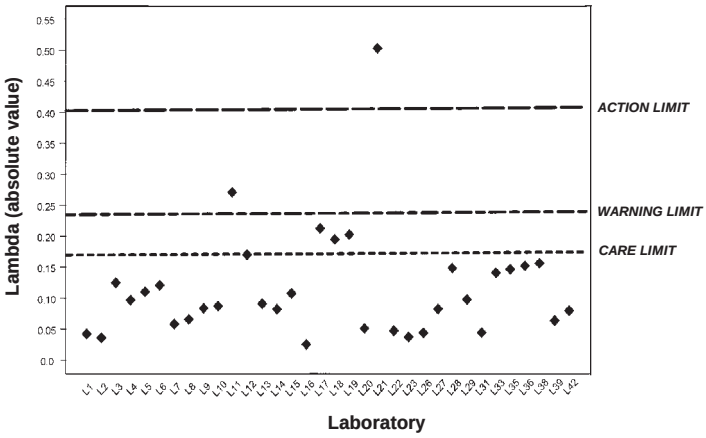

Fig. 4 - *Assay imprecision.*

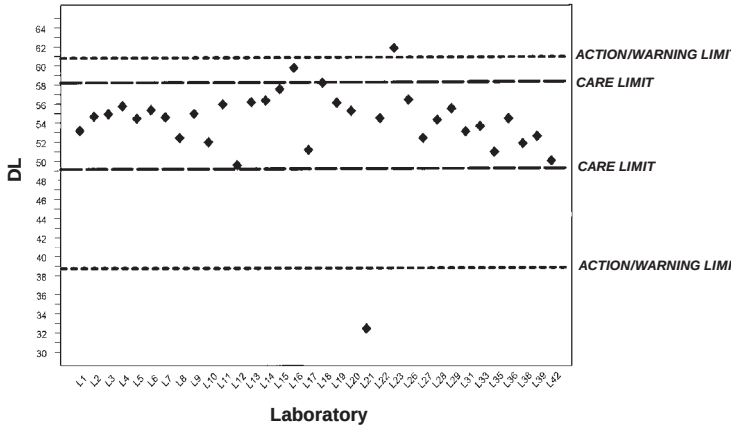

Fig. 5 - *Signal detection limit.*

to the 97.5th centile (upper line) and the 2.5th centile (lower line), represent the action limits.

iii) Assay imprecision

According to Bliss (7) the statistic (lambda) is $\hat{\lambda}_i = s_i/b_{1i}$. The corresponding absolute value was adopted to build the chart (Fig. 4). The three horizontal lines correspond to the 95th centile (upper line, action limit), 90th centile (middle line, warning limit) and 80th centile (lower line, care limit).

iv) Signal detection limit

According to Miller and Miller (8) the statistic is $DL = b_{0i} + 3.29s_i$, where b_{0i} is the intercept estimated by each standard curve. Like in i) and iii), both the upper and lower regions of the chart are relevant. Since the action and warning limits partially overlap, the chart shows four lines corresponding to the 97.5th centile (upper line, action and warning limits), the 90th centile (upper middle line, care limit), the 10th centile (lower middle line, care limit) and the 2.5th centile (lower line, action and warning limit).

It is worth noting that the statistics of charts i) and ii) take into account the within-laboratory variability only, whereas those of charts iii) and iv) take into account both the within- and between-laboratory variability. This explains the different width of the control bands presented in the two chart subsets. The message emerging from the charts must be interpreted accordingly.

The rationale and the statistical method used to construct the charts are described in the appendix.

RESULTS

The statistics considered in iii) and iv) are composite indexes which assembly, in different ways, information regarding the measurement variability and the standard curve accuracy. Therefore, when trying to interpret data produced by different laboratories, it seems reasonable to start from the two charts reported in Figures 5 and 4, respectively. One can move later to the charts of Figures 3 and 2 to find the reason for the possible poor performance of a laboratory. In particular, from Figure 5 it ap-

pears that two laboratories (L21 and L23) are outside the action limit and one (L16) is outside the upper care limit. Laboratory L21 is outside the action limit even with reference to Figure 4. Furthermore, laboratory L11 is outside the warning limit and three laboratories (L17, L18 and L19) are outside the care limit according to this Figure.

When we move to the charts of Figures 3 and 2, we see that laboratories L16, L21 and L23 are outside the limits with regard to the standard curve accuracy and variability measurement. While laboratory L21 is outside the upper action limit for accuracy (with a notably lesser slope than expected), the other two laboratories (L16 and L23) are outside the lower action limit for accuracy (Fig. 3) as a joint effect of a high slope and an unpredictably low variability measurement (Fig. 2). For laboratories L11 and L18 both the variability measurement and the standard curve inaccuracy appear to be responsible for their position in Figure 4, whereas only the variability measurement appears to be responsible for the position outside the care limit of laboratories L17 and L19.

CONCLUDING REMARKS

External standard quantification is the method of choice for nucleic acid quantification, particularly mRNA quantification. Absolute quantification relates the PCR signal to the input copy number using a calibration curve. However, the external calibration curve model has to be thoroughly validated as the accuracy of absolute quantification in real-time PCR depends entirely on the accuracy of the standards. In real-time PCR, the main analytical features of the external standard curve are the "pure error" of the straight line (an index of measurement variability), the slope (connected to the global efficiency of standard amplification) and its intercept on the y-axis (an index of the sensitivity of the curve) (6). This EQA pilot study on 42 laboratories was designed to assess the variability of these parameters and the possible influence on the reliability of the quantitative assay.

Overall, our analysis identified a subset of seven laboratories (L11, L16, L17, L18, L19, L21 and L23) that required careful attention because their performance appeared to be markedly outside the limits with respect to at least one of the standard curve features investigated.

It is important to note that the performance of each real-time experiment may be affected by several analytical variables. The type of real-time platform and reagents may influence the efficiency of amplification. Another consequence of variation in the amplification efficiency is the modification of the slope of the standard curve which, in turn, can modify the theoretical sensitivity of the assay. At the same time is important to remember that the standard curve was prepared by each participant as a 1:10 dilution of a stock solution. Inappropriate dilution of the standard may introduce an artifactual variation of the slope. Finally, as expected in any analytical procedure, the precision and accuracy are strongly affected by the skill of the operator. All these predictable and unpredictable variables may generate variable degrees of bias in sample quantification.

When the results related to the quantification of the three unknown samples were considered (data not shown), the same laboratories except one produced measurements that were out of control in terms of either precision (L11) or accuracy (L23, L16) or both (L21, L17, L18). Worthy of note is laboratory L19 which, although its measurements were not out of control for precision and accuracy for the unknown sample, proved to be outside the warning limits for assay imprecision. This suggests the usefulness of the approach presented here to identify laboratories with a problematic performance. In any case, the findings emerging from the analysis of the two investigated aspects, monitoring and process modifying, should be appropriately integrated to select laboratories that need technical advice on their performance.

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

APPENDIX

i) Variability measurement

It is well known that:

$\frac{\nu s_i^2}{\sigma^2}$ is distributed like the χ^2 variable with ν degrees of freedom (d.f.). As each s_i^2 has 12 [= 6 standard dilutions x (3 replicates – 1)] d.f., the pooled variance s_p^2 has 12 x 34 = 408 d.f. The latter was used as an estimate of σ^2. Therefore, the statistic $\sqrt{s_i^2/s_p^2}$ can be assumed to be approxi-mately distributed like: $\sqrt{\chi^2/\nu}$. The 2.5th and 97.5th centiles of the χ^2 distribution with 12 d.f. were used to draw the action limits in the chart presented in Figure 2.

ii) Standard curve accuracy

The statistic:

$\frac{b_{1i}-\beta_1}{SE(b_{1i})}$ is distributed like t_ν where:

t_ν is the Student's *t* test statistic with ν = 12 d.f. The two horizontal lines in Figure 3 therefore correspond to the 2.5th and 97.5th centiles of Student's *t* distribution.

iii) Assay imprecision

Bliss (7) suggested to draw chart limits by assuming a Student's *t* distribution for $\hat{\lambda}_i$. Unfortunately, the assumption of symmetry was not at all tenable in the present context. We therefore resorted to a bootstrap approach (9) to compute the non-parametric centiles of the $\hat{\lambda}_i$s distribution. 500,000 bootstrap samples were obtained from the original set of 34 $\hat{\lambda}_i$s. The values obtained were (see Fig. 4):
95th centile = 0.4028;
90th centile = 0.2346;
80th centile = 0.1693.

iv) Signal detection limit

Although the variance of the DL statistic can be easily computed and its sampling distribution could be approximated by a normal distribution, also in this case did we resort to the bootstrap approach to make allowance for the between-laboratory variability. 500,000 bootstrap samples were generated to obtain the non-parametric centiles of the DL distribution. The values obtained were (see Fig. 5):
97.5th centile = 60.7954;
90th centile = 58.2077;
10th centile = 49.1640;
2.5th centile = 38.7292.

REFERENCES

1. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology 1993; 11: 1026-30.
2. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 2000; 25: 169-93.
3. Orlando C, Pinzani P, Pazzagli M. Developments in quantitative PCR. Clin Chem Lab Med 1998; 36: 255-69.
4. Klein D. Quantification using real-time PCR technology: applications and limitations. Trends Mol Med 2002; 8: 257-60.
5. Raggi CC, Verderio P, Pazzagli M, Marubini E, Simi L, Pinzani P, Paradiso A, Orlando C. A national program of external quality control for quantitative assays based on real-time PCR with Taq-Man™ probes. Clin Chem (submitted).
6. Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 2002; 29: 23-39.
7. Bliss CI. The statistics of bioassay. New York: Academic Press Inc., 1952.
8. Miller JC, Miller JN. Statistics for analytical chemistry. New York: John Wiley & Sons, 1984.
9. Efron B, Tibshirana RJ. An introduction to the bootstrap. London: Chapman and Hall, 1993.

Received: June 7, 2004
Accepted: June 7, 2004