

RESEARCH

Open Access



# Introducing the $f_0\%$ method: a reliable and accurate approach for qPCR analysis

Mahmoud Gamal<sup>1\*</sup> and Marwa A. Ibrahim<sup>1</sup>

\*Correspondence:  
mahmoud.gamal@cu.edu.eg

<sup>1</sup> Department of Biochemistry and Molecular Biology, Faculty of Veterinary Medicine, Cairo University, Giza 12211, Egypt

## Abstract

**Background:** qPCR is a widely used technique in scientific research as a basic tool in gene expression analysis. Classically, the quantitative endpoint of qPCR is the threshold cycle ( $C_T$ ) that ignores differences in amplification efficiency among many other drawbacks. While other methods have been developed to analyze qPCR results, none has statistically proven to perform better than the  $C_T$  method. Therefore, we aimed to develop a new qPCR analysis method that overcomes the limitations of the  $C_T$  method. Our  $f_0\%$  [eff naught percent] method depends on a modified flexible sigmoid function to fit the amplification curve with a linear part to subtract the background noise. Then, the initial fluorescence is estimated and reported as a percentage of the predicted maximum fluorescence ( $f_0\%$ ).

**Results:** The performance of the new  $f_0\%$  method was compared against the  $C_T$  method along with another two outstanding methods—LinRegPCR and  $C_{y_0}$ . The comparison regarded absolute and relative quantifications and used 20 dilution curves obtained from 7 different datasets that utilize different DNA-binding dyes. In the case of absolute quantification,  $f_0\%$  reduced CV%, variance, and absolute relative error by 1.66, 2.78, and 1.8 folds relative to  $C_T$ ; and by 1.65, 2.61, and 1.71 folds relative to LinRegPCR, respectively. While, regarding relative quantification,  $f_0\%$  reduced CV% by 1.76, 1.55, and 1.25 folds and variance by 3.13, 2.31, and 1.57 folds regarding  $C_T$ , LinRegPCR, and  $C_{y_0}$ , respectively. Finally,  $f_0\%$  reduced the absolute relative error caused by LinRegPCR by 1.83 folds.

**Conclusions:** We recommend using the  $f_0\%$  method to analyze and report qPCR results based on its reported advantages. Finally, to simplify the usage of the  $f_0\%$  method, it was implemented in a macro-enabled Excel file with a user manual located on <https://github.com/Mahmoud0Gamal/F0-perc/releases>.

**Keywords:** Real-time PCR, Delta  $C_T$ , PCR efficiency, Calibration curve, Performance indicators, Variation between replicates, Curve analysis, Inflection cycle

## Background

Quantitative polymerase chain reaction (qPCR) allows the quantification of minute amounts of a specific target DNA by monitoring the increase in fluorescence associated with its amplification. Fluorescence signals collected in qPCR are commonly produced from DNA-binding dyes, such as SYBR Green I or fluorophore-labeled oligonucleotides



[1]. Combining qPCR with reverse transcription (RT-qPCR) extended its ability to quantify RNA, especially microRNA and mRNA. This advent unleashed the power of gene expression analysis which became one of the widely used methods in scientific research [2].

qPCR experiments are used to quantify the target nucleic acid either absolutely or relatively [3]. Absolute quantification requires building a standard curve to give a copy number for each reaction. A standard curve is a linear relationship between the threshold cycles ( $C_T$ ) and their  $\log_{10}$  transformed concentrations [4]. On the other side, relative quantification requires one or more reference genes to quantify the target gene relative to them. While absolute quantification is widely used in quantifying microbial nucleic acid, relative quantification is a basic tool in gene expression analysis [3].

The quantitative endpoint of both types of quantification is a threshold cycle ( $C_T$ ) that is defined as the PCR cycle at which the fluorescence signal crosses an arbitrary threshold [5]. Till now,  $C_T$  is the most used method to analyze and report qPCR results [6]. However, there are several limitations to this method including:

1. The efficiency of the PCR reaction ( $E$ ): The PCR reaction starts—in best cases—with ( $E=2$ , complete doubling). Then, efficiency gradually declines due to the reduced availability of the reaction substrates till ( $E=1$ , no amplification) at the plateau phase [7]. The starting efficiency of the PCR reaction can vary depending on the template, primers, and reaction conditions. However, the  $C_T$  method assumes that the PCR efficiency is the same for both the target and the reference genes. If there is a significant difference in efficiency between them, then the normalization of the target gene expression using the  $C_T$  method is invalid [8, 9].
2. Inhibition of the PCR reaction: The presence of inhibitors in the sample can affect the efficiency of the PCR reaction and lead to inaccurate  $C_T$  values. Inhibitors can be present in the sample due to various reasons such as impurities in the RNA preparation or the presence of PCR inhibitors in the sample matrix [10].
3. Accuracy of the instrument: The accuracy of the qPCR instrument can affect the precision and accuracy of the  $C_T$  values. Differences in the sensitivity and specificity of the instrument can lead to variations in the  $C_T$  values obtained [10].
4. Data analysis: The  $C_T$  method assumes that the PCR amplification is in the exponential phase, and the  $C_T$  value is determined at a fixed threshold level. However, the actual amplification kinetics can vary between samples and genes, and the choice of threshold level can affect the accuracy of the  $C_T$  values obtained [11].

Many methods have been developed to overcome the limitations of the  $C_T$  method such as the sigmoidal models,  $C_{Y_0}$ , LinRegPCR,  $C_{Y_0}^*$ , CqMAN, etc. The sigmoidal curve methods involve fitting the raw data to a four or five-parameter sigmoid equation and the initial fluorescence is calculated [12, 13]. In the  $C_{Y_0}$  method, the raw data is fitted to Richard's equation and a tangent is drawn at the inflection point where its intersection with the abscissa is considered the  $C_{Y_0}$  value that is used as a  $C_T$  [14]. LinRegPCR calculates the efficiency for each reaction through a straight line fitted through a pre-determined window of linearity. Then, the average of these efficiencies is calculated and used for each amplicon [15].  $C_{Y_0}^*$  determines the earliest amplification cycle ( $C^*$ ) as an outlier over the background fluorescence and calculates efficiency through three

amplification cycles starting with the  $C^*$  followed by calculating the initial template amount [16]. In CqMAN, the  $C_q$  is the cycle corresponding to the midpoint between the baseline and the second derivative maximum fluorescence based on a modified Gompertz model. While efficiency (averaged per amplicon) is calculated from a three-parameter exponential model fitted to the cycles from the  $C_q$  to the second derivative maximum [17].

Despite the aforementioned limitations of the  $C_T$  method and the development of many methods to overcome them, the  $C_T$  method is still the most adopted in analyzing qPCR results [6]. One of the reasons is the simplicity of the  $C_T$  method. Another important reason is the lack of statistical evidence of the advantage of using the other methods. The  $C_T$  method was compared once with 10 methods and another time with 13 methods and the Friedman test included the  $C_T$  method along with LinRegPCR and  $C_{y_0}$  in the subset with the highest rank in both studies [7, 16].

The current study aims to develop a qPCR analysis method ( $f_0\%$ ) that addresses the drawbacks of the  $C_T$  method by minimizing the quantification errors and variation between replicates. Hence, enhances the validity and robustness of the gene expression analysis. Moreover, the performance of the  $f_0\%$  method was compared with the best methods in analyzing qPCR—the  $C_T$ , LinRegPCR, and  $C_{y_0}$  methods—as reported earlier [7, 16] using datasets that depend on DNA-binding dyes. Moreover, the analysis process considered the presence or absence of a dilution curve. Finally, to facilitate the use of the  $f_0\%$  method, a model was developed and implemented in a user-friendly program.

## Methods

### Datasets

20 dilution curves obtained from 7 different datasets were used to evaluate the  $f_0\%$  method against the  $C_T$ , LinRegPCR, and  $C_{y_0}$  methods. The datasets represent various PCR instruments, DNA binding-dyes, and reaction mixtures. All datasets were imported directly from the qpcR R package [18]. In each reaction, the baseline cycles ( $C_{3;8}$ ) were averaged and their slope was calculated, after which the background fluorescence was subtracted except for the LinRegPCR method. Furthermore, the normalization was performed by dividing the fluorescence of each reaction by the maximum fluorescence of the corresponding dilution curve.

The used datasets are named after the name of their first author as follows:

1. *Boggy* contains a dilution curve with six tenfold dilutions with 2 replicates. The qPCR was performed on Chromo4 thermal cycler (Bio-Rad) using the SYTO-13 fluorescent dye. The target was a randomly generated synthetic DNA sequence that was optimized to reduce the secondary structures [19].
2. *Ruijter* contains a dilution curve with four tenfold dilutions with 94 replicates. The qPCR was performed on CFX 384 instrument (Bio-Rad) using SYBR Green I dye. The DNA target was a synthetic oligonucleotide for the human MYCN gene [7].
3. *Guescini* contains a dilution curve with seven tenfold dilutions with 12 replicates. The qPCR amplification was conducted using LightCycler<sup>®</sup> 480 (Roche) with SYBR

Green I dye. A plasmid containing a 104 bp fragment of the mitochondrial gene NADH dehydrogenase 1 served as the target DNA [14].

4. *Lievens* contains a dilution curve with five fivefold dilutions with 18 replicates. Soybean genomic DNA was used with primers targeting the lectin endogene Le1. Quantification was based on SYBR Green I [20].
5. *Spiess* contains four dilution curves denoted as Spiess\_1, Spiess\_2a, Spiess\_2b, and Spiess\_3. Spiess\_1 is a dilution curve of seven tenfold dilutions with 4 replicates. Spiess\_2a and Spiess\_2b are two dilution curves of five fourfold dilutions with 3 replicates for two different cDNA samples. Spiess\_3 is a dilution curve of seven fourfold dilutions with 3 replicates. The S27a housekeeping gene served as the target. The qPCR instruments used were LightCycler 1.0 (Roche) for Spiess\_1 and MXPro3000P (Stratagene) for Spiess\_2a, Spiess\_2b, and Spiess\_3. While SYBR Green I dye was used for the quantification of all dilution curves, only Spiess\_3 was ROX-normalized [18].
6. *Rutledge* contains Six tenfold dilutions with 4 replicates in 5 individual batches. Each batch is considered a dilution curve and denoted as Rutledge\_1:5. The primers were designed to amplify a 102 bp amplicon with the help of SYBR Green I dye using Opticon2 (MJ Research Inc) [12].
7. *Vermeulen* is a huge dataset containing the expression data of 59 genes in addition to 5 housekeeping genes. It was performed to build a multigene-expression signature to help in the prognosis of patients with neuroblastoma. Each of the 64 genes had a dilution curve of five tenfold dilutions with 3 replicates. We included only the first 7 genes (alphabetical order) to avoid bias to a single qPCR instrument. The used genes were AHCY, AKR1C1, ARHGEF7, BIRC5, CAMTA1, CAMTA2, and CD44 while ALUsq was excluded as it shows early amplification ( $C_T=21:26$ ) in the no template control. Quantification was conducted on LightCycler<sup>®</sup> 480 (Roche) using SYBR Green I dye [21].

### The $C_T$ method

The threshold cycle ( $C_T$ ) is a method to report qPCR results quantitatively as defined earlier. This method is based on placing an arbitrary threshold at the exponential phase of the amplification curve [5]. Since choosing the threshold value has a great impact on the  $C_T$  analysis results [11], we employed a strategy that links the threshold to the baseline noise of the corresponding dilution curve. We found that the maximum standard deviation of the baseline fluorescence of each dilution curve serves as a reliable indicator of its noise. Then, the threshold value for each dilution curve was set to be 100-fold the baseline noise. This threshold yielded the best performance in all datasets except for *Boggy*, *Ruijter*, and *Rutledge* datasets where a threshold value of 10, 50, and 50 folds of their baseline noise yielded better results, respectively. An amplification plot for each dilution curve with its threshold is provided in Additional file 1: Fig. S1 and Additional file 2: Fig. S2. Finally, the  $C_T$  was calculated according to Eq. (1, 2) [22].

$$C_T = x + \frac{\ln \text{Threshold} - \ln f_x}{\ln E} \quad (1)$$

where  $x$  is the cycle immediately before the threshold,  $f_x$  is the fluorescence at cycle  $x$ , and  $E$  is calculated as follows:

$$E = \frac{f_{(x+1)}}{f_x} \quad (2)$$

### The LinRegPCR method

The LinRegPCR method analyzes qPCR using efficiencies calculated from the slope of a regression line of the datapoints in the exponential phase of baseline corrected fluorescence data. The software of this method takes non-baseline corrected raw data and corrects the baseline. Then an iterative algorithm is used to allocate the exponential phase known as the window of linearity. Then, the average of these efficiencies is calculated and used for each amplicon. Finally, the software produces an  $N_0$ —initial nucleic acid amount—calculated using the mean efficiency [15].

### The $Cy_0$ method

In the  $Cy_0$  method, the 5-parameter Richard's equation was used to fit a non-linear curve using the raw data. At the inflection point of this curve a tangent is drawn where its intersection with the abscissa is considered the  $Cy_0$  value [14]. The  $Cy_0$  was calculated by the `cy0()` function of the `qpcR` R package [18] using RStudio-2023.06.1-524 [23] and R programming language v4.3.1 [24].

### The $f_0\%$ method

The  $f_0\%$  method is based on a six-parameter model Eq. (3) composed of two parts: a four-parameter sigmoid part and a two-parameter linear part. The four-parameter sigmoid part is used to fit the amplification curve with parameters to predict the values of the maximum fluorescence ( $f_m$ ), the rate of efficiency decay ( $D$ ), the starting efficiency ( $E$ ), and the inflection cycle ( $C_i$ ). While the role of the two-parameter linear part is to subtract the background noise with parameters to predict the values of the baseline slope ( $a$ ) and the baseline intercept ( $b$ ). The mathematical role of each of the equation parameters is graphically illustrated in Fig. 1.

$$f_x = f_m - \frac{f_m}{(1 + DE^{x-C_i})^{1/D}} + ax + b \quad (3)$$

where  $f_x$  represents the fluorescence at cycle  $x$ .

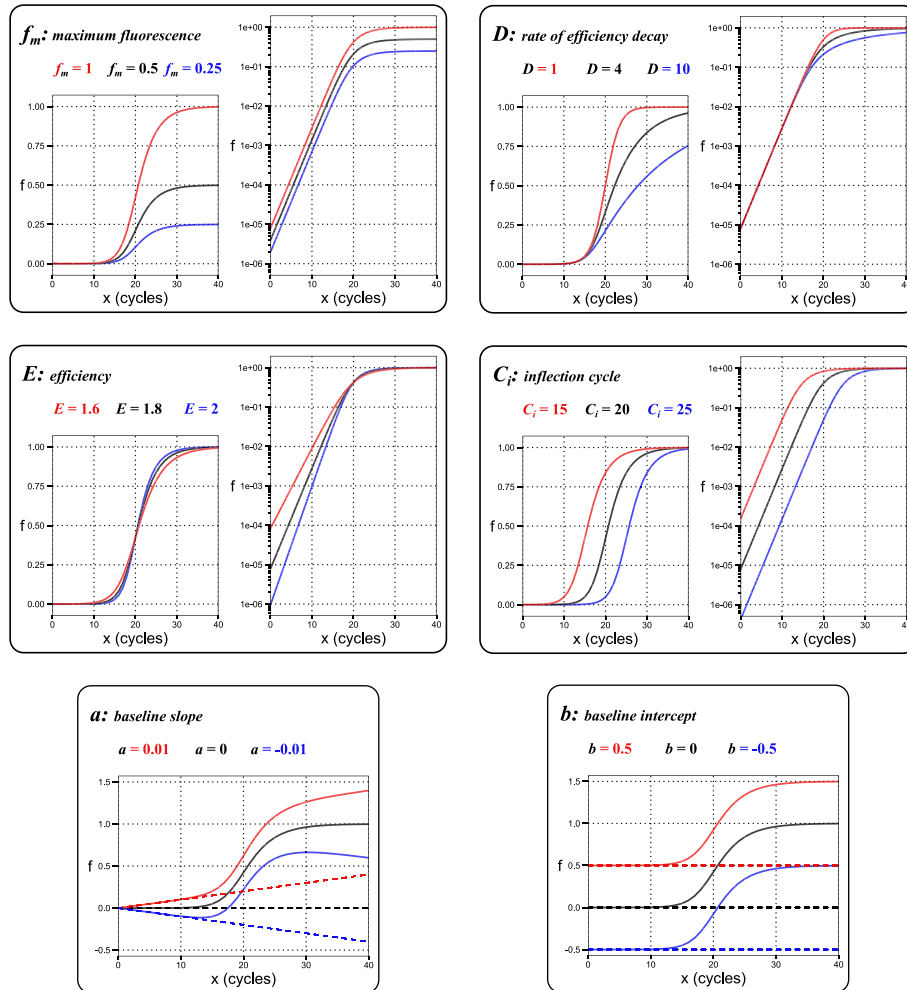
The previous equation (Eq. 3) operates in two modes:

1. Free  $E$  mode: where  $E$  is a variable that is left to be predicted. Then,  $E$ s from different reactions are averaged per amplicon to give a prediction of the starting efficiency of that amplicon. This mode of Eq. 3 is fitted to cycles ranging from the first cycle to the cycle just after the inflection cycle ( $C_i$ ).
2. Fixed  $E$  mode: where  $E$  is a constant that may be calculated in the free  $E$  mode or predetermined using the slope of a standard curve as described later. In this mode,

### Equation parameters

$$f_x = f_m + \frac{f_m}{(1 + DE^{x-C_i})^{1/D}} + ax + b$$

← Sigmoid part
→ Linear part  
 To fit the amplification curve  To subtract the background noise



**Fig. 1** Graphical illustration of the equation parameters. The parameters of the equation could be divided into sigmoidal-related parameters ( $f_m$ ,  $D$ ,  $E$ , and  $C_i$ ) to fit the amplification curve and linear parameters ( $a$  and  $b$ ) to fit the baseline of the curve. Where  $f_x$  is the fluorescence of cycle  $x$  and  $x$  is the cycle number. To illustrate the effect of each of these parameters, the value of only one parameter was varied while maintaining all other parameters to a fixed value ( $f_m = 1$ ,  $D = 2$ ,  $E = 1.8$ ,  $C_i = 20$ ,  $a = 0$ ,  $b = 0$ ). All parameters have two plots: linear (left) and logarithmic (right) except for  $a$  and  $b$  have only linear plots.  $f_m$ : the maximum predicted fluorescence.  $D$ : the rate of efficiency decay, a high value indicates a rapid reduction in efficiency from cycle to cycle.  $E$ : the starting efficiency or efficiency in the baseline (note that fixing the value of other parameters—specifically  $C_i$ —while plotting curves with variable  $E$  necessitates starting with different fluorescence. Thus, low  $E$  values were compensated with high starting fluorescence leading to an early rise of their amplification curve).  $C_i$ : the inflection cycle is the point that separates the upper and lower parts of the curve and corresponds to the point with the maximum slope.  $a$ : baseline slope is the slope of the background noise in case of drifting background.  $b$ : the baseline intercept is the starting value of the background noise

all cycles are fitted by solving for  $f_m$ ,  $D$ ,  $C_i$ ,  $a$ , and  $b$  using the constant  $E$ . Then, the initial fluorescence ( $f_0$ ) is calculated using Eq. (4).

$$f_0 = f_m - \frac{f_m}{(1 + DE^{-C_i})^{1/D}} \quad (4)$$

In all modes, the initial cycle or cycles should be discarded if they deviate obviously from the baseline. Finally,  $f_0\%$  is calculated as a percentage of the predicted maximum fluorescence as shown in Eq. (5). A flowchart of the analysis process is shown in Fig. 2.

$$f_0\% = \frac{f_0}{f_m} \times 100 \quad (5)$$

It is important to note that there are two sources of  $E$  that should be considered when using the fixed  $E$  mode:

1. In experiments lacking the data for a standard curve,  $E$  is calculated in the free  $E$  mode.
2. In experiments with the data for a standard curve, this data is converted to an approximate  $f_0\%$  using the fixed  $E$  mode (assuming  $E=2$ ). Then, the standard curve is built by regressing  $\log_{10}(f_0\%)$  on  $\log_{10}(conc.)$ . Finally,  $E$  is calculated from the slope of the regression line using Eq. (6).

$$E = 2^{1/Slope} \quad (6)$$

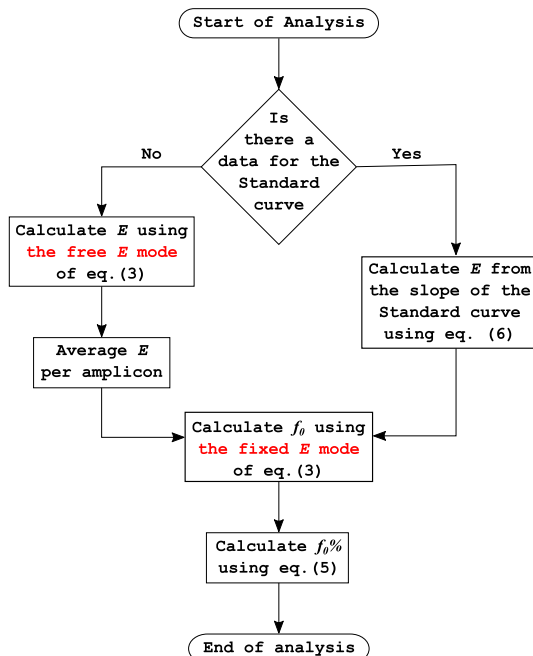
### Quantification

After calculating the  $C_T$ ,  $N_0$ ,  $C_{Y_0}$ , and  $f_0\%$  for all reactions in each of the 20 dilution curves, the  $C_T$ ,  $N_0$ ,  $C_{Y_0}$ , and  $f_0\%$  values were converted to predicted concentrations. For all methods, the predicted concentrations were calculated twice. Once considering the presence of a standard curve and once assuming the absence of a standard curve. This strategy was adopted to assess the performance of the used methods for experiments that either include or lack a standard curve, respectively.

#### **Condition 1: considering the presence of a standard curve**

The standard curve is built by regressing  $C_T$ ,  $\log_{10}(N_0)$ ,  $C_{Y_0}$ , or  $\log_{10}(f_0\%)$  on  $\log_{10}(conc.)$ . In the case of using  $C_T$  or  $C_{Y_0}$ , the slope of the regression line is negative because both  $C_T$  and  $C_{Y_0}$  proportionate inversely with the concentration. While using  $\log_{10}(N_0)$  or  $\log_{10}(f_0\%)$  produces a positive slope because  $N_0$  and  $f_0\%$  are directly proportional to the concentration. Finally, to retrieve concentrations, the regression equation is solved for  $\log_{10}(conc.)$  using the  $C_T$ ,  $\log_{10}(N_0)$ ,  $C_{Y_0}$ , or  $\log_{10}(f_0\%)$  values. And the obtained  $\log_{10}(conc.)$  is raised to the power of 10.

### Flowchart of the analysis process



**Fig. 2** Flow chart of the analysis process. *The first step* is to calculate the starting efficiency ( $E$ ). If there is data for a standard curve, approximate  $f_0\%$  is calculated using the fixed  $E$  mode of Eq. (3) assuming ( $E = 2$ ). Then,  $\log_{10}(f_0\%)$  is regressed on  $\log_{10}(conc.)$  and  $E$  is calculated from the slope of this regression according to Eq. (6). If there is no data for a standard curve,  $E$  is estimated using the free  $E$  mode of Eq. (3) and averaged per amplicon. *The next step* is calculating  $f_0$  using the fixed  $E$  mode of Eq. (3) utilizing  $E$  calculated in the previous step. Followed by calculating  $f_0\%$  using Eq. (5)

#### Condition 2: assuming the absence of a standard curve

Normally in this condition, a relative quantification is performed rather than an absolute one. However, for the aim of our study, a predicted concentration should be calculated for each reaction to be used in the performance indicators (described later). To solve this problem, the concentrations were predicted relative to the highest concentration (1st level) in each dilution curve. So, the predicted concentrations have the same scale as the true concentrations.

For the  $C_T$  and  $C_{y_0}$  methods:

$$\Delta(C_T \text{ or } C_{y_0})_{(i)} = (C_T \text{ or } C_{y_0})_{(i)} - \text{Mean}\left((C_T \text{ or } C_{y_0})_{(1^{st} \text{ level})}\right) \tag{7}$$

$$\text{Predicted conc.}_{(i)} = 2^{-\Delta(C_T \text{ or } C_{y_0})_{(i)}} \times \text{true conc.}_{(1^{st} \text{ level})} \tag{8}$$

For the  $f_0\%$  and LinRegPCR methods:

$$\text{Predicted conc.}_{(i)} = \frac{(f_0\% \text{ or } N_0)_{(i)}}{\text{Mean}\left((f_0\% \text{ or } N_0)_{(1^{st} \text{ level})}\right)} \times \text{true conc.}_{(1^{st} \text{ level})} \tag{9}$$



### Performance indicators

The predicted concentrations calculated by the  $C_T$ , LinRegPCR,  $C_{y_0}$ , and  $f_0\%$  methods were compared with the true concentrations—concentrations obtained from the datasets—to evaluate the performance of these methods. Different performance indicators were needed to measure the performance of the tested methods from different aspects. Precision, which refers to the variation between replicates, was evaluated using the coefficient of variation and variance. Accuracy, which refers to the deviation of the predicted concentrations from the true concentrations, was assessed using the relative error and bias. The performance indicators were calculated as follows:

1. **Coefficient of variation (CV%).** CV% was calculated for each level of the dilution curves as follows [17]:

$$CV\% = \frac{SD(\text{predicted conc.})}{Mean(\text{predicted conc.})} \times 100 \quad (10)$$

2. **Variance.** Variance represents the within-level variance of the  $\log_{10}(\text{predicted concentration})$  [7].
3. **Relative error (RE).** RE is the deviation of the predicted concentrations from the true concentrations [17].

$$RE = \frac{\text{predicted conc.} - \text{true conc.}}{\text{true conc.}} \quad (11)$$

The perfect value of RE is zero which indicates no error. While values greater or lesser than zero indicate error proportional to the absolute value. In this manner, if we take the average of RE for different reactions, negative values will negate the effect of positive values leading to a misleading average. Therefore, we calculated an absolute relative error that could be averaged.

$$\text{Absolute RE} = |\text{Relative error}| \quad (12)$$

4. **Bias.** Bias is the ratio of the averaged predicted concentrations of the highest to the lowest levels [7]. For example, if the true concentration of the highest level in a given dilution curve is 10,000 and the true concentration of the lowest level is 10, then the perfect ratio for bias is 1000. Obviously, the perfect ratio varies between dilution curves according to the dilution rate and the number of levels. Therefore, a normalized bias was calculated by dividing the bias ratio by the perfect expected ratio of the respective dilution curve.

$$\text{Normalized Bias} = \frac{Mean(\text{the highest predicted conc.})}{Mean(\text{the lowest predicted conc.})} \div \frac{\text{the highest true conc.}}{\text{the lowest true conc.}} \quad (13)$$

Because the normalized bias may be greater or lesser than one—the optimum value—an averaged normalized bias would be misleading as described earlier for relative error. So, we calculated an absolute bias that can be averaged with one indicating no bias and values greater than one indicating bias.

$$\text{Absolute Bias} = e^{|\ln(\text{Normalized Bias})|} \quad (14)$$

For each of the previous performance indicators, a fold reduction was calculated to quantify the effect of using  $f_0\%$  instead of  $C_T$ , LinRegPCR, or  $C_{y_0}$ . Then, the geometric mean of the fold reduction was reported.

$$\text{Fold Reduction}_{(PIx)} = \frac{C_T, \text{ LinRegPCR, or } C_{y_0(PIx)}}{f_0\%(PIx)} \quad (15)$$

where  $PIx$  is one of the performance indicators: CV%, variance, absolute RE, and absolute bias.

### Statistical analysis

Friedman test was performed to check if the distribution of each performance indicator—grouped by method, paired by dilution curve—shows a statistically significant difference. A significant Friedman test was followed by pairwise Wilcoxon signed-rank tests to identify the differences between methods. Then,  $p$  values were adjusted for alpha inflation using the Bonferroni correction. All statistical analysis was conducted using RStudio-2023.06.1-524 [23] and R programming language v4.3.1 [24]. A statistically significant difference was considered when ( $p$  value < 0.05). All R scripts containing the statistical analysis were provided in the Additional file 4.

### Results

The variation in the used qPCR datasets was intended to represent different templates, primers, master mixes, DNA-binding dyes, and qPCR instruments. Furthermore, the analysis process considered the varying objectives of the qPCR experiments from absolute to relative quantification, which influences the need for a standard curve. Finally, different performance indicators were used to measure the accuracy of the compared methods from different perspectives.

#### Performance evaluation considering the presence of a standard curve

Upon comparing the methods considering experiments using a standard curve, it was clear that the  $f_0\%$  method offers more advantages than the  $C_T$  and LinRegPCR methods. Calculating the  $f_0\%$  reduced the variation between technical replicates indicating increased precision. This was evident by reducing the CV% of the  $C_T$  and LinRegPCR methods by 1.66 folds ( $p$  value < 0.001) and 1.65 folds ( $p$  value < 0.01), respectively. Moreover, the  $f_0\%$  reduced the variance of the  $C_T$  and LinRegPCR methods by 2.78 folds ( $p$  value < 0.001) and 2.61 folds ( $p$  value < 0.01), respectively. On the other hand, there was no statistically significant difference between the  $f_0\%$  and the  $C_{y_0}$  methods in both parameters. Furthermore, the  $C_T$ , LinRegPCR, and  $C_{y_0}$  methods also didn't show a statistically significant difference between each pair of them regarding the CV% and variance.

Regarding accuracy, both  $f_0\%$  and  $C_{y_0}$  methods reduced the absolute relative error in comparison to the  $C_T$  method by 1.8 folds ( $p$  value < 0.0001) and 1.19 folds ( $p$  value = 0.022), respectively. Furthermore, only the  $f_0\%$  method decreased the absolute

relative error regarding the LinRegPCR method by 1.71 folds ( $p$  value  $< 0.01$ ). Regarding the other parameter of accuracy—absolute bias, the Friedman test was insignificant. Figure 3 outlines these results while Additional file 3: Table S1 shows the detailed performance of all methods on each dilution curve.

#### Performance evaluation assuming the absence of a standard curve

In experiments lacking a standard curve, the  $f_0\%$  method showed potential advantage when compared to other methods specifically in the precision parameters. The  $f_0\%$  method increased the uniformity of replicates by lowering the CV% by 1.25 folds ( $p$  value  $< 0.01$ ) against the  $C_{y_0}$  method, by 1.55 folds ( $p$  value  $< 0.01$ ) against the LinRegPCR method, and by 1.76 folds ( $p$  value  $< 0.0001$ ) against the  $C_T$  method. Moreover, the  $f_0\%$  method diminished the variance by 1.57 folds ( $p$  value  $< 0.001$ ) regarding the  $C_{y_0}$  method, by 2.31 folds ( $p$  value  $= 0.033$ ) regarding the LinRegPCR method, and by 3.13 folds ( $p$  value  $< 0.0001$ ) regarding the  $C_T$  method.

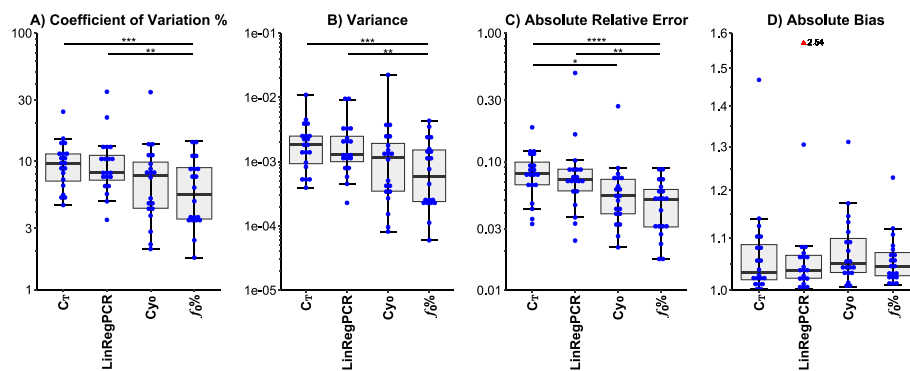
Regarding the accuracy related parameters, statistically significant differences were scarce. However, the  $f_0\%$  method minimized the absolute relative error in comparison to the LinRegPCR method by 1.83 folds ( $p$  value  $< 0.01$ ). These results are briefed in Fig. 4 while the performance on each dilution curve is provided in Additional file 3: Table S2.

#### Effect of the inflection cycle ( $C_i$ ) position on the performance of the $f_0\%$ method.

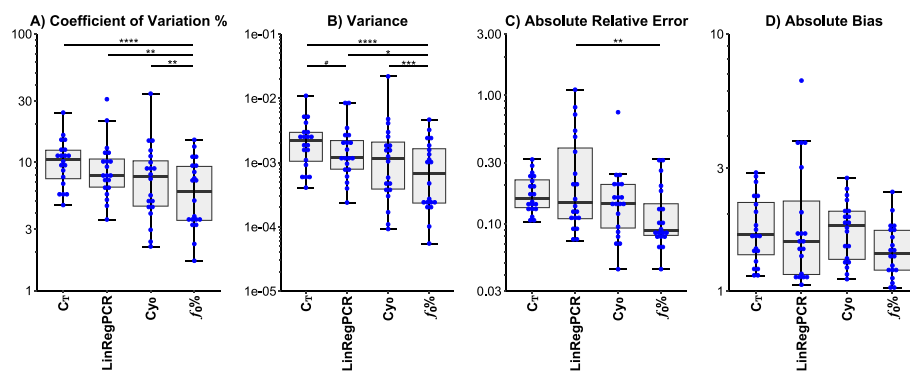
As described earlier, all available cycle readings are used to obtain the  $f_0\%$  in the Fixed  $E$  Mode, however, sometimes all available cycles are not enough to calculate a precise  $f_0\%$ . The performance of the  $f_0\%$  depends on the inflection cycle ( $C_i$ ) position. It was assumed that the earlier the  $C_i$ , the more precise the  $f_0\%$ . To check this assumption, the first levels of the 20 dilution curves were used to evaluate the performance of the  $f_0\%$  (represented as CV%) relative to the position of the  $C_i$ . For each reaction, 11 predictions were performed that differed in the final cycle ( $C_{final}$ ) to be used in the model ranging from ( $C_{final} = C_i - 5$ ) to ( $C_{final} = C_i + 5$ ). We found that our assumption is true: an earlier  $C_i$  is associated with a more precise  $f_0\%$ . Generally,  $f_0\%$  is considered precise only when the  $C_{final}$  is two cycles or more after the  $C_i$  ( $C_{final} \geq C_i + 2$ ). Figure 5 shows the relationship between the performance of the  $f_0\%$  and the position of the  $C_i$  relative to the  $C_{final}$ .

## Discussion

qPCR is a widely used technique to quantify minute amounts of specific nucleic acid. It allowed unprecedented advances in gene expression analysis and quantifying infectious pathogens [2]. The traditional method used to report the quantitative endpoint for that technique is the  $C_T$  method [5] which ignores most of the reaction data. To overcome the drawbacks of the  $C_T$  method, we introduce a new qPCR analysis method ( $f_0\%$ ). The new method proved to be statistically more robust in most situations than the other compared methods. The  $f_0\%$  method outperformed the widely used  $C_T$  method in experiments that either contain or lack a standard curve. In experiments with a standard curve, the  $f_0\%$  reduced CV%, variance, and absolute relative error by 1.66, 2.78, and 1.8 folds, respectively. Moreover, when analyzing the experiments without using



**Fig. 3** Performance evaluation considering the presence of a standard curve. **A** Coefficient of variation %: level-specific standard deviation as a percentage of the level-specific average. **B** Variance: within-level variance of the  $\log_{10}(\text{predicted concentration})$ . **C** Absolute relative error: the absolute value of the deviation of the predicted concentrations from the true concentrations. **D** Absolute bias: the absolute value of the ratio of the average of the highest to the lowest predicted concentrations divided by the ratio of the highest to the lowest true concentrations. In all indicators, lower is better. \* $p$  value < 0.05, \*\* $p$  value < 0.01, \*\*\* $p$  value < 0.001, and \*\*\*\* $p$  value < 0.0001

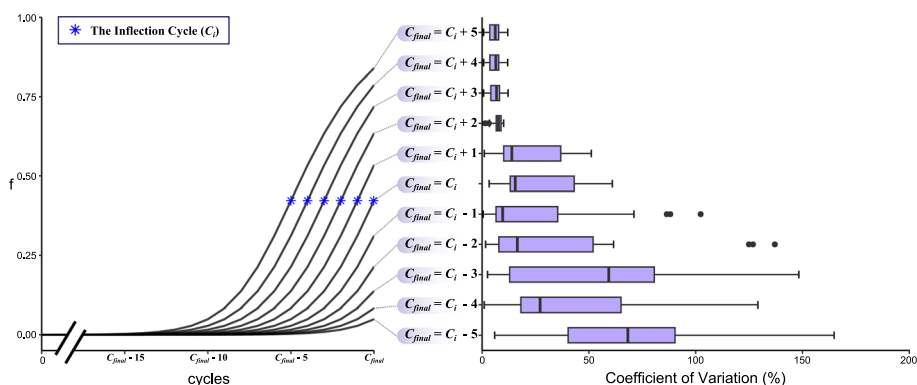


**Fig. 4** Performance evaluation assuming the absence of a standard curve. **A** Coefficient of variation %: level-specific standard deviation as a percentage of the level-specific average. **B** Variance: within-level variance of the  $\log_{10}(\text{predicted concentration})$ . **C** Absolute relative error: the absolute value of the deviation of the predicted concentrations from the true concentrations. **D** Absolute bias: the absolute value of the ratio of the average of the highest to the lowest predicted concentrations divided by the ratio of the highest to the lowest true concentrations. In all indicators, lower is better. # $p$  value = 0.05, \* $p$  value < 0.05, \*\* $p$  value < 0.01, \*\*\* $p$  value < 0.001, and \*\*\*\* $p$  value < 0.0001

the standard curve data, the  $f_0\%$  reduced CV% and variance by 1.76, 3.13 folds, respectively. Finally, the  $f_0\%$  method was implemented in a macro-enabled Excel file available at <https://github.com/Mahmoud0Gamal/F0-perc/releases> with a user manual to describe how to professionally use the  $f_0\%$  method.

### Method development

The amplification curve of the qPCR has a characteristic sigmoid nature. Thereafter, many qPCR analysis methods adopted the sigmoid functions [12–14]. In 2004, Rutledge introduced the sigmoid function as a tool to analyze the qPCR curve. He used a four-parameter sigmoid equation with a symmetric nature that struggled to fit the asymmetric sigmoid curve of the PCR. To overcome this limitation, he excluded the plateau



**Fig. 5** Effect of the position of the inflection cycle ( $C_i$ ) on the coefficient of variation (%). Reactions of the first levels of the 20 dilution curves were used to calculate  $f_0\%$  11 times. Each time the reaction was trimmed at a different cycle relative to the  $C_i$  so that the final cycle ( $C_{final}$ ) ranged from ( $C_{final} = C_i - 5$ ) to ( $C_{final} = C_i + 5$ ). On the left side, amplification curves with different  $C_i$  ranging from ( $C_{final} = C_i - 5$ ) to ( $C_{final} = C_i + 5$ ). On the right side, a horizontal box plot representing the coefficient of variations (%) of the 20 dilution curves grouped by  $C_i$

phase from the fit by choosing a cut-off cycle [12]. Later on, Spiess et al. introduced a flexible five-parameter sigmoid equation containing an asymmetry parameter. Spiess’ equation increased the accuracy of the fit, but it retained a parameter responsible for the slope [13]. This slope parameter enables the calculation of efficiency from single reaction data and inherently quantification using this reaction-specific efficiency [25]. Although it corrects for efficiency variation and intuitively seems to be advantageous, quantification based on reaction-specific efficiencies tends to dramatically increase the variation between replicates and diminishes the reliability of the quantification. On the other hand, averaging the efficiency per each target achieves greater robustness and reliability [8].

The  $f_0\%$  method also utilizes a sigmoid equation (Eq. 3) with minor but critical modifications. In our equation, we substituted the constant in the denominator (the natural base— $e$ ) of the previous sigmoid equations [12, 13] with a variable that directly represents the starting efficiency ( $E$ ). Moreover, redundancy was reduced by removing the parameter responsible for the slope, since its effect was transferred to the  $E$  parameter. Also, we retained the flexibility of the model by including a parameter for asymmetry ( $D$ ). Here,  $D$  represents the rate of efficiency decay, a high value of  $D$  indicates a rapid reduction in cycle-to-cycle efficiency, associated with increased deceleration of the amplification rate. In contrast to  $D$ ,  $E$  is the starting efficiency and it’s not related to changes in efficiency from cycle to cycle.

Using reaction-specific efficiency in quantification was reported to increase the variation between replicates [8]. After validating this report, we avoided the direct quantification using reaction-specific efficiencies by calculating and averaging efficiencies from all amplicon-specific reactions using the free  $E$  mode. Then, taking this averaged efficiency to the fixed  $E$  mode to calculate the  $f_0\%$ . This scenario provides an estimate for efficiency when no standard curve is present while preventing the drastic errors caused by using reaction-specific efficiencies.

For accurate analysis of the qPCR curve, the background noise should be properly subtracted by a process called baselining. Therefore, a linear part ( $ax + b$ ) was added to (Eq. 3) to allow the  $f_0\%$  method to subtract the background noise during the analysis. This approach was adopted instead of baselining separately before analyzing the qPCR curve to avoid reduced precision as noted earlier [25]. Indeed, our method of baselining is fairly accurate even if the background noise is shifting up or down assuming a linear shift. However, some qPCR reactions may show a complex non-linear baseline. In these reactions removing the first few cycles leaves a nearly linear baseline which is suitable for the  $f_0\%$  analysis.

The use of the  $f_0\%$  method was extended to correct for variations in the maximum fluorescence or the reaction volume. It is assumed that the maximum fluorescence is similar in all reactions containing similar amounts of the same PCR mixture. However, due to volume variations and fluctuations in the signal output from the PCR instrument, maximum fluorescence is not identical. This artifact occurs between different wells in the PCR instrument, and it is exaggerated when ROX normalization is ignored [26]. To remove the effect of this artifact on quantification, the  $f_0\%$  method reports the predicted initial fluorescence ( $f_0$ ) as a percentage of the predicted maximum fluorescence ( $f_m$ ) using (Eq. 5). Therefore, apparent differences in the maximum fluorescence would not impair the accuracy of the quantification.

### Performance

The performance of the  $f_0\%$  method was compared with the  $C_T$ , LinRegPCR, and  $Cy_0$  methods which constitute the best subset of Friedman test in analyzing qPCR data as reported earlier by two independent benchmarking studies [7, 16]. The methods were tested for precision and accuracy using different indicators regarding various qPCR platforms and experimental designs. The overall precision of the  $f_0\%$  method was statistically superior to the  $C_T$  and LinRegPCR methods evidenced by reduced CV% and variance regarding both types of quantification (absolute and relative). In addition, considering relative quantification the  $f_0\%$  method was more robust than the  $Cy_0$  method. The superiority of the  $f_0\%$  method in precision is attributed to many factors. First, the  $f_0\%$  method—unlike the  $C_T$  and LinRegPCR—takes benefit of all cycle readings recorded by the qPCR instrument like other curve-fitting methods e.g.,  $Cy_0$ . Consequently, the  $f_0\%$  method doesn't depend on an arbitrary threshold that its position may change the results [13–15, 17]. Moreover, unlike other sigmoidal models [12, 13], the  $f_0\%$  and LinRegPCR methods depends on averaged efficiency per amplicon rather than quantification directly using reaction-specific efficiencies [15]. Finally, the  $f_0\%$  method is the first—based on our information—to normalize the result to the predicted maximum fluorescence.

In terms of accuracy, absolute relative error (RE) and absolute bias were calculated. Regarding the absolute RE, the  $f_0\%$  method offered substantial improvement over the  $C_T$  and LinRegPCR methods in experiments containing a standard curve. Moreover, when considering experiments lacking a standard curve, the  $f_0\%$  method continued to outperform the LinRegPCR method. However, in both types of experiments, there was no statistical difference between the  $f_0\%$  and  $Cy_0$  methods. This might be attributed to using the whole data points in the analysis of the  $f_0\%$  and  $Cy_0$  [14] but not for the  $C_T$  and LinRegPCR methods [5, 15].

Regarding the other accuracy parameter—absolute bias, depending on a standard curve made all methods produce relatively unbiased results that were very close to one. This was predicted as quantification using a standard curve nearly eliminates bias [7]. However, in experiments lacking a standard curve, the values of absolute bias were greatly deviated. Although bias is caused—in theory—by misestimating efficiency [7], methods that correct efficiency didn't show statistically significant enhancements in absolute bias. To interpret these unexpected findings, we shall consider the effect of dilution errors on bias. As these errors make the actual bias ratio vary greatly from the perfect expected ratio [8]. Therefore, it will be obvious that the most accurate methods will produce a deviated bias in case of dilution errors. In our case, we used publicly available datasets, so we had no control over their quality, and we relied on the quality control parameters stated by their authors.

### Limitations

The  $f_0\%$  method predicts the shape of the amplification curve which has a sigmoid nature. Sigmoid curves are characterized by the presence of one inflection point that separates the upper and the lower parts of the curve. To precisely predict the shape of this curve, we should provide enough data around the inflection point or inflection cycle ( $C_i$ ) [27]. However, in very rare occasions amplification curves may show late amplification—especially those with reduced efficiency—and fail to reach the  $C_i$  before the final cycle ( $C_{final}$ ) leading to insufficient data around this critical point. Therefore, the precision of the calculated  $f_0\%$  will be reduced.

In practical situations, most amplification curves will pass the  $C_i$ . However, in very rare occasions—genes with very low expression along with reduced amplification efficiency—the amplification curve may fail to reach the  $C_i$ . If a researcher encounters this situation, the reaction efficiency shall be enhanced. If it is not possible, the number of cycles could be increased. However, increasing the number of cycles will increase the chance of amplifying non-specific products. Therefore, increasing the number of cycles should be done with caution in these rare occasions only if efficiency improvements fail. Moreover, a high-resolution melt curve analysis must be examined carefully to identify any non-specific products.

Another limitation related to data availability is using the datasets that depend on DNA-intercalating dyes only for assessing the performance of the  $f_0\%$  method. Although the widespread usage of TaqMan probes in signal detection in qPCR experiments, we didn't find enough publicly available datasets to validate the performance of the  $f_0\%$  method on them.

### Applications

The goal of qPCR experiments is either absolute or relative quantification [3]. Absolute quantification relies on the presence of a standard curve with known concentrations as described earlier. On the other side, relative quantification could be performed without using a standard curve, but it requires the presence of one or more reference genes. Relative quantification is well described using the  $C_T$  method [5] and here we will describe how to perform two modes of relative quantification using the  $f_0\%$  method.



**Fold change**

Fold change is the ratio between the concentration of the target gene and the geometric mean of the concentrations of the reference genes of the same sample. Fold change is useful in experiments lacking a control group and could be calculated using Eq. (16).

$$\text{Fold change} = \frac{f_0\%(\text{Target gene})}{\text{Geometric Mean}(f_0\%(\text{Reference genes}))} \quad (16)$$

**Normalized fold change**

Normalized fold change is the ratio between the fold change of a given sample and the mean of the fold changes of the control group samples. In this case, the mean of the normalized fold changes of the control group samples should equal one. Normalized fold change is calculated using Eq. (17).

$$\text{Normalized fold change}_{(\text{Sample } x)} = \frac{\text{Fold change}_{(\text{Sample } x)}}{\text{Geometric Mean}(\text{Fold change}_{(\text{Control group Samples})})} \quad (17)$$

**Conclusions**

Although the widespread usage of the  $C_T$  method in analyzing qPCR data, it suffers from many drawbacks. To address these limitations, we introduced the  $f_0\%$  method which utilizes all the available cycle readings to give more reliable results. The new method is based on a flexible sigmoid model that was modified to avoid the instability of the previous sigmoidal models. Indeed, our method demonstrated more robust and accurate results when compared with the  $C_T$ , LinRegPCR, and  $C_{y_0}$  methods using a compiled multi-platform dataset. However, the enhanced performance of the  $f_0\%$  method comes at the cost of requiring a well-developed amplification curve that has at least two cycles post-inflection. Moreover, we facilitated the usage of the  $f_0\%$  by implementing the method in a user-friendly macro-enabled Excel file which can be easily downloaded and used by researchers. Overall, the  $f_0\%$  method offers a more reliable and accurate approach to qPCR analysis, with the potential to improve the accuracy of quantitative measurements in a variety of applications using qRT-PCR.

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-024-05630-y>.

**Additional file 1: Fig. S1.** Graphical representation of Dilution curves (1–10).

**Additional file 2: Fig. S2.** Graphical representation of Dilution curves (11–20).

**Additional file 3: Table S1.** Results of the performance indicators of the  $C_t$ , LinRegPCR,  $C_{y_0}$ , and  $f_0\%$  methods considering the presence of the standard curve. **Table S2.** Results of the performance indicators of the  $C_t$ , LinRegPCR,  $C_{y_0}$ , and  $f_0\%$  methods assuming the absence of the standard curve.

**Additional file 4:** R scripts used in the analysis process.

**Acknowledgements**

Not applicable.



**Author contributions**

MG and MI conceptualized the research idea, interpreted the results, and wrote the article. MG wrote the R code and programmed the f0% analyzer excel sheet. All authors read and approved the final manuscript.

**Funding**

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). This research received no external funding.

**Availability of data and materials**

All data analyzed during this study are imported from the qpcR R package as shown in the R scripts in Additional file 4.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 1 November 2023 Accepted: 2 January 2024

Published online: 11 January 2024

**References**

1. Navarro E, Serrano-Heras G, Castaño MJ, Solera J. Real-time PCR detection chemistry. *Clin Chim Acta*. 2015;439:231–50.
2. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*. 2008;44:619–26.
3. Taylor SC, Nadeau K, Abbasi M, Lachance C, Nguyen M, Fenrich J. The ultimate qPCR experiment: producing publication quality, reproducible data the first time. *Trends Biotechnol*. 2019;37:761–74.
4. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: recommendations for precise and robust qPCR efficiency assessments. *Biomol Detect Quantif*. 2015;3:9–16.
5. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc*. 2008;3:1101–8.
6. Hawke DC, Watson AJ, Betts DH. Selecting normalizers for microRNA RT-qPCR expression analysis in murine preimplantation embryos and the associated conditioned culture media. *J Dev Biol*. 2023;11:17.
7. Ruijter JM, Pfaffl MW, Zhao S, Spiess AN, Boggy G, Blom J, et al. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. *Methods*. 2013;59:32–46.
8. Ruijter JM, Barnewall RJ, Marsh IB, Szentirmay AN, Quinn JC, van Houdt R, et al. Efficiency correction is required for accurate quantitative PCR analysis and reporting. *Clin Chem*. 2021;67:829–42.
9. Bustin SA, Benes V, Garson JA, Hellems J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55:611–22.
10. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nat Protoc*. 2006;1:1559–82.
11. Ruiz-Villalba A, Ruijter JM, van den Hoff MJB. Use and misuse of Cq in qPCR data analysis and reporting. *Life*. 2021;11:496.
12. Rutledge RG. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic Acids Res*. 2004;32:e178–e178.
13. Spiess A-N, Feig C, Ritz C. Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. *BMC Bioinform*. 2008;9:221.
14. Guescini M, Sisti D, Rocchi MBL, Stocchi L, Stocchi V. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC Bioinform*. 2008;9:326.
15. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res*. 2009;37:e45–e45.
16. Zhang L, Dong R, Wei S, Zhou H-C, Zhang M-X, Alagarsamy K. A novel data processing method CyC\* for quantitative real time polymerase chain reaction minimizes cumulative error. *PLoS ONE*. 2019;14:e0218159.
17. Zhang Y, Li H, Shang S, Meng S, Lin T, Zhang Y, et al. Evaluation validation of a qPCR curve analysis method and conventional approaches. *BMC Genom*. 2021;22:680.
18. Spiess A-N. Modelling and analysis of real-time PCR data—CRAN. 2018, p. 109.
19. Boggy GJ, Woolf PJ. A mechanistic model of PCR for accurate quantification of quantitative PCR data. *PLoS ONE*. 2010;5:e12355.
20. Lievens A, Van Aelst S, Van den Bulcke M, Goetghebeur E. Enhanced analysis of real-time PCR data by using a variable efficiency model: FPK-PCR. *Nucleic Acids Res*. 2012;40:e10–e10.
21. Vermeulen J, De Preter K, Naranjo A, Vercruyse L, Van Roy N, Hellems J, et al. Predicting outcomes for children with neuroblastoma using a multigene-expression signature: a retrospective SIOPEN/COG/GPOH study. *Lancet Oncol*. 2009;10:663–71.
22. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2<sup>-ΔΔCT</sup> method for quantitative real-time polymerase chain reaction data analysis. *Biostat Bioinform Biomath*. 2013;3:71–85.
23. Posit team. RStudio: integrated development environment for R. 2023.
24. R Core Team. R: a language and environment for statistical computing. 2023.
25. Tellinghuisen J. Estimating real-time qPCR amplification efficiency from single-reaction data. *Life*. 2021;11:693.

26. Wang G, Becker E, Mesa C. Optimization of 6-carboxy-X-rhodamine concentration for real-time polymerase chain reaction using molecular beacon chemistry. *Can J Microbiol.* 2007;53:391–7.
27. Pödör Z, Manninger M, Jereb L. *Advanced computational methods for knowledge engineering.* Cham: Springer; 2014.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

