Published in final edited form as:

Stat Methods Med Res. 2018 September ; 27(9): 2581–2595. doi:10.1177/0962280216683204.

Modeling qRT-PCR dynamics with application to cancer biomarker quantification

Inna Chervoneva^{1,2,*}, Boris Freydin^{1,2}, Terry Hyslop³, and Scott A. Waldman²

¹Division of Biostatistics, Thomas Jefferson University, Philadelphia, PA 19107, USA

²Department of Pharmacology and Experimental Therapeutics, Thomas Jefferson University, Philadelphia, PA 19107, USA

³Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27710, USA

Abstract

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) is widely used for molecular diagnostics and evaluating prognosis in cancer. The utility of mRNA expression biomarkers relies heavily on the accuracy and precision of qRT-PCR quantification, which is still challenging for low abundance transcripts. The critical step for qRT-PCR quantification is accurate estimation of efficiency needed for computing relative qRT-PCR expression.

We propose a new approach to estimating qRT-PCR efficiency based on modeling dynamics of PCR amplification. In contrast, only models for fluorescence intensity as a function of PCR cycle have been used so far for relative qRT-PCR quantification. The dynamics of qRT-PCR efficiency is modeled using an ordinary differential equation (ODE) model, and the fitted ODE model is used to obtain effective PCR efficiency estimates needed for efficiency-adjusted relative qRT-PCR quantification.

The proposed new qRT-PCR efficiency estimates were used to quantify GUCY2C mRNA expression in the blood of colorectal cancer patients. Time to recurrence and GUCY2C expression ratios were analyzed in a joint model for survival and longitudinal outcomes. The joint model with GUCY2C quantified using the proposed PCR efficiency estimates provided clinically meaningful results for association between time to recurrence and longitudinal trends in GUCY2C expression.

Keywords

quantitative RT-PCR; PCR efficiency; relative qRT-PCR quantification; ordinary differential equations models; robust nonlinear regression

*Corresponding author. Inna.chervoneva@jefferson.edu.

8 Conflict of Interest Statement

Inna Chervoneva, Terry Hyslop and Scott A. Waldman are inventors of licensed technology, U.S. Patent No.20110195415 “Guanylyl Cyclase C qRT-PCR”, which is the subject of a license agreement between Thomas Jefferson University and Targeted Diagnostics & Therapeutics, Inc. to commercialize inventions related to this work (qRT-PCR quantification of GUCY2C expression).

1 Introduction

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) technology is a valuable tool to identify molecular variations in specific biomarkers between normal, precancerous, cancerous, and metastatic cells that can serve as targets for detection, diagnosis, therapy, and prevention of cancer. Quantification of metastatic tumor burden beyond the presence of tumor cells or molecular markers may be useful as a prognostic factor. Another important area of application of qRT-PCR is accurate quantification of transcripts that represent receptors for novel targeted cancer therapies in order to predict response to treatment. While detection of biomarkers (presence vs. absence) is becoming a part of routine clinical practice, the actual quantification of disease-associated mRNA expressed by circulating tumor cells remains limited. One critical barrier is unsatisfactory quantification accuracy of low abundance biomarkers in limited clinical samples.

The present study was motivated by the need to improve the accuracy of quantifying mRNA expression of the GUCY2C gene in blood of colorectal cancer patients. The GUCY2C gene encodes guanylyl cyclase C (GCC), which is an enzyme found in the intestinal epithelium. GCC is a key receptor for bacterial heat-stable enterotoxins which produce secretory diarrhea. It is selectively expressed by intestinal epithelium, which makes it a suitable biomarker for colorectal cancer micro-metastases not detectible by conventional histopathology. Previously, the mRNA expression of GUCY2C in histologically negative lymph nodes was shown to be associated with time to recurrence and disease-free survival in patients with histologically node-negative colorectal cancer [1]. The data that motivated the present study were collected during the NIH funded study conducted to examine the utility of GUCY2C expression for post-operative surveillance in the blood of patients with colorectal cancer to identify patients at increased risk for disease recurrence. This study is further referred to as the GUCY2C study. The underlying hypothesis is that GUCY2C mRNA expression may detect occult circulating tumor cells in blood, which may be an early indicator of colorectal cancer recurrence. As part of the GUCY2C study, blood samples were collected a minimum of two different times after surgery from colorectal cancer patients. qRT-PCR was used to quantify relative GUCY2C expression normalized by using two reference genes, GAPDH and TFRC. These reference genes were selected as optimal using a previously developed method of selecting reference genes for normalization in qRT-PCR [2].

Relative qRT-PCR expression is obtained by normalizing the expression of a target biomarker by the expression of endogenous reference genes. Such normalization to reference genes is required for application of qRT-PCR technology to most biomarker investigations. A relative or normalized mRNA expression is computed as a ratio between the target biomarker expression and normalizing factor based on expression of endogenous reference gene(s). The accuracy of quantification substantially depends on the accuracy of PCR efficiency estimates needed for computing relative qRT-PCR expression. Small reaction efficiency differences between target and reference gene may translate into grossly over- or under-estimated expression ratios if the difference in efficiencies is unaccounted for [3]. Therefore, state-of-the-art relative quantification methods adjust for actual reaction efficiencies by utilizing methods that explicitly incorporate estimated PCR efficiency into calculations of the relative/normalized expression [3–7].

In most qRT-PCR experiments, the empty wells (so called non-template controls) yield nonzero fluorescence signals. This background signal has to be subtracted to ensure the fundamental underlying assumption of qRT-PCR quantification that at each PCR cycle x , the fluorescence intensity is directly proportional to the number of copies of amplified templates. The background-subtracted fluorescence intensity is further denoted by $F(x)$. Figure 1 shows examples of kinetic qRT-PCR data (fluorescence intensity $F(x)$ as a function of PCR cycle x) for 4 patient samples in the GUCY2C study. Kinetic qRT-PCR data represents the amplification history through the entire range of the PCR cycles. The phases of kinetic data from a typical qRT-PCR reaction are shown in Figure 2. The fluorescence signal becomes distinguishable from the background noise at the beginning of the exponential phase, when the number of product templates becomes sufficiently large. Respectively, the length of the linear ground phase is longer when the initial number of target templates is smaller. The steepness of the slope in the exponential and log-linear phase is related to the reaction efficiency. For example, higher slopes correspond to higher amplification efficiencies. In figure 1, the observed differences in the slopes of the red kinetic curves for GUCY2C and blue and green curves for the reference genes (GAPDH and TFRC) highlight the apparent differences in PCR efficiencies between the low abundance target and much more abundant reference genes. Thus, to achieve accurate relative quantification of GUCY2C mRNA expression, it is necessary to use efficiency-adjusted relative quantification with precise PCR efficiency estimates.

The blood samples yield limited amounts of mRNA, which make it impossible to use the “gold standard” estimate of the PCR efficiency based on the slope of the calibration/standard curve constructed from a serial dilution series (see section 2.2 for detail on computing slope efficiency). For clinical samples, it is usually not feasible to have sufficient material for a serial dilution, and efficiency may be estimated only for each individual reaction. The standard methods of estimating individual PCR reaction efficiency are based on models for fluorescence intensity $F(x)$ as a function of PCR cycle x . The majority of these methods employ either an exponential growth model [8], which is valid only in the early exponential phase, or logistic/sigmoid models [5, 9, 10].

The main drawback of the exponential growth model is the necessity to identify the early exponential phase of the qRT-PCR reaction, for which the exponential model is appropriate [8]. The logistic models imply monotone decrease of relative PCR efficiency during the reaction, but all methods of relative qRT-PCR quantification are based on a constant qRT-PCR efficiency (see section 2.1). Such constant qRT-PCR efficiency is estimated either at its maximum or at some point along the fluorescence intensity curve (e.g. the point of the second derivative maximum). The use of the maximum efficiency is justified when it is reasonable to assume that the growth of the PCR product is approximately exponential up to the threshold cycle (see 2 and section 2.2 for definition) that is used for computing the relative expression [9]. But if the efficiency decrease by the threshold cycle is substantial, then different choices of a constant efficiency estimate at some point along the fluorescence intensity curve do not have any theoretical justification.

Schnell and Mendoza [11] deduced from enzymological consideration that the Michaelis-Menten model for derivative of the PCR efficiency is appropriate to describe dynamics of

PCR efficiency for the early exponential and log-linear phase of the reaction. However, the model has not been used to estimate PCR efficiency for practical purposes. One of the reasons is that Michaelis-Menten model does not have a closed form expression for fluorescence as a function of PCR cycle. Another reason is that the qRT-PCR efficiency does not follow the Michaelis-Menten model in the later PCR cycles. Hence, similar to the case of exponential model, using the Michaelis-Menten model to estimate efficiency would require for every qRT-PCR reaction, solving a challenging problem of identifying a small subset of PCR cycles where the model is appropriate. Plus, the models based on small numbers of data points are known to be less accurate and efficient. To overcome these problems, we propose to model the entire range of PCR cycles using a more general model that includes Michaelis-Menten model as a particular case but also fits well the entire range of the PCR cycles. This eliminates the need to find a subset of PCR cycles where Michaelis-Menten model is valid and substantially increases the number of data points available for estimating the model. In particular, we consider a modification of the EMAX model (a.k.a. Hill model), which includes the Michaelis-Menten model as a special case [12] but takes in account possible degradation of templates and allows efficiency to become equal to zero at a finite point (when all reagents are consumed, new templates can not be created). The fitted modified EMAX model is used to compute a PCR efficiency estimate for each individual reaction to be used for relative quantification. Thus, this report provides the first account of using appropriate models for dynamics of PCR efficiency to obtain the PCR efficiency estimates necessary for relative quantification.

Using serial dilution data for synthetic GUCY2C standards, we demonstrate that the proposed estimates of PCR efficiency have smaller bias and mean squared errors than currently used standard estimates of PCR efficiency. The bias and mean squared error are determined by comparing to the “gold standard” slope efficiency estimates. The improved estimates of efficiency are shown to improve accuracy of estimates of the ratios between known concentrations of the GUCY2C standards. The new PCR efficiency estimates were used to compute the normalized GUCY2C expressions in blood samples of colorectal cancer patients, and the normalized GUCY2C expressions and time to recurrence were modeled jointly to examine the utility of the GUCY2C expression to identify patients at increased risk of disease recurrence. Using the new method of efficiency adjusted relative quantification, it was possible to obtain clinically meaningful result about the association between the time to recurrence and trends in GUCY2C expressions.

Section 2 of this article reviews efficiency-adjusted relative quantification and standard methods for estimating PCR efficiency. In Section 3, we introduce the proposed family of modified EMAX models and describe procedures for their estimation. Numerical studies comparing the new and standard efficiency estimates and corresponding quantification results are reported in Section 4. Section 5 describes the results of the joint analysis of time to recurrence and GUCY2C expression levels in blood of colorectal cancer patients. Concluding remarks are included in Section 6.

2 Relative quantification of mRNA using qRT-PCR

2.1 Efficiency adjusted qRT-PCR quantification

For simplicity, first assume that only one replicate of target and reference reactions are available for each biological sample. Let us denote by $N_T(0)$ and $N_R(0)$ the initial number of copies of the target and reference templates, respectively. We are interested in estimating $\ln(N_T(0)/N_R(0))$ for each sample, since it is expected that log-transformed expression ratios from independent biological samples have approximately normal distribution.

The underlying assumption of the qRT-PCR is that at each PCR cycle x , $x \in [1, M]$ (M is usually between 40 and 50), the fluorescence intensity $F(x)$ is directly proportional to the number of copies of amplified templates $N(x)$, $F(x) = PN(x)$, where P is an unknown very small proportionality constant common for all reactions in the same assay. Since target and reference gene reactions usually utilize different reagents (e.g. primers and probes), we have $F_T(x) = P_T N_T(x)$ for the target gene reaction and $F_R(x) = P_R N_R(x)$ for the reference gene reaction, with generally $P_R \neq P_T$, so that $\ln(N_T(0)/N_R(0)) = \ln F_T(0) - \ln P_T - \ln F_R(0) + \ln P_R$. The constant $(\ln P_R - \ln P_T)$ may be estimated using additional control samples with known ratios between the initial concentrations [14]. In comparative studies, the same assay is used for all samples, and the constant $(\ln P_R - \ln P_T)$ does not effect any between-group comparisons. Therefore, for the purpose of analyzing data from a comparative study, log-transformed relative expression ratios may be computed as $\ln R_{R/T} = \ln F_T(0) - \ln F_R(0)$. The values $F(0)$ can not be measured directly because the noise in fluorescence data is a few magnitudes higher than these quantities of interest. Instead, the qRT-PCR quantification is effectively based on the number of cycles required for the fluorescence signal to become detectable above the noise level.

The PCR reaction is described by a branching process [19], and in the beginning of the reaction, the number of templates grows approximately exponentially with the base E , $E \geq 2$. The corresponding exponential growth model is given by

$$F(x) = F(0)E^x. \quad (1)$$

The phase of the PCR reaction where this approximation is valid is called the early exponential phase (Figure 2). In the molecular biology literature, E is usually referred to as the PCR efficiency. A more accurate, but less common, term is the amplification rate. Let C_{tR} and C_{tT} denote the early PCR cycle numbers such that $F_T(C_{tT}) = F_R(C_{tR}) = t$ for some a priori chosen t . Such C_{tR} and C_{tT} are called threshold cycles (corresponding to threshold t). Then, the exponential model (1) implies that $F_T(0)E^{C_{tT}} = F_R(0)E^{C_{tR}}$, and hence the log-transformed relative expression ratio is given by

$$\ln R_{R/T} = \ln F_T(0) - \ln F_R(0) = C_{tR} \ln E - C_{tT} \ln E.$$

Respectively, for biological samples with multiple replicates of target and reference reactions, the log-transformed relative expression ratio is computed as

$$\ln R_{R/T} = \overline{C_{tR} \ln E_R} - \overline{C_{tT} \ln E_T}, \quad (2)$$

where $\overline{C_{tT} \ln E_T}$ represents the average of quantities $C_{tTj} \ln E_{tTj}$, $j = 1, \dots, J_T$ across J_T replicates of the target gene and $\overline{C_{tR} \ln E_R}$ represents the average of quantities $C_{tRj} \ln E_{tRj}$, $j = 1, \dots, J_R$ across J_R replicates of (multiple) reference genes. Computation of cycle numbers is straightforward using linear interpolation between discretely observed values of $F(x)$, but estimation of PCR efficiencies needed in (2) may be challenging, especially for low abundance targets.

2.2 Standard methods of estimating PCR efficiency

The “gold standard” estimate of the PCR efficiency is based on the slope of the calibration/standard curve constructed from a serial dilution. In serial dilution experiments, an initial stock solution is diluted sequentially with the same dilution ratio r , which usually varies between 2 and 10. Let us denote by N_s the true concentration of templates in the initial stock solution and by $N_{dk}(r)$, the concentration in k^{th} replicate after the d^{th} dilution with dilution ratio r . Ignoring the dilution errors, $N_{dk}(r) = (N_s/r^d)$ and

$$\ln N_{dk}(r) = \ln N_s - d \ln r. \quad (3)$$

Let C_{tdk} be the threshold cycle for k^{th} replicate after the d^{th} dilution. The exponential growth model with common efficiency E implies that $t = PN_{dk}E^{C_{tdk}}$, where P is defined above, so that

$$\ln(t) = \ln P + \ln N_{dk} + C_{tdk} \ln E. \quad (4)$$

Combining (3) and (4) one can express $C_{tdk} = \alpha + \beta d$, where

$$\alpha = \frac{\ln t - \ln P - \ln N_s}{\ln E}, \beta = \frac{\ln r}{\ln E}.$$

The standard approach is to estimate the calibration curve as the regression model

$$C_{tdk} = \alpha + \beta d + \epsilon_{dk} \quad (5)$$

where ε_{dk} represent the dilution errors. Technically ε_{dk} are not independent and identically distributed errors, and the model is not a standard linear regression [15, 16]. However, it is known that the ordinary least squares point estimate $\hat{\beta}$ obtained by treating model (5) as a standard linear regression is still consistent even when the errors ε_{dk} are correlated [17]. Hence, using the regression slope estimate $\hat{\beta}$, the slope efficiency E_{sl} is computed as

$$E_{sl} = r^{1/\hat{\beta}},$$

If N_s and all N_{dk} are known, then from (4) we obtain $C_{tdk} \ln E = \ln t - \ln P - \ln N_{dk}$, where $\ln t - \ln P$ is a constant. Hence, one can express $C_{dk} = \delta + \gamma \ln N_{dk}$, where

$$\delta = \frac{\ln t - \ln P}{\ln E}, \gamma = -\frac{1}{\ln E},$$

and linear regression model

$$C_{dk} = \alpha + \gamma \ln N_{dk} + \varepsilon_{dk}$$

may be used to estimate $\gamma = -1/\ln E$ and compute the slope efficiency E_{sl} as

$$E_{sl} = \exp(-1/\hat{\gamma}). \quad (6)$$

The slope efficiency estimate is considered the most accurate, but it requires multiple reactions for the same target, which is not feasible for clinical samples with small amounts of mRNA available for analysis. The alternative is to estimate the PCR efficiency of each individual reaction. Most current methods for estimating constant exponential efficiency of an individual reaction require fitting an exponential growth model to a limited subset of the early exponential phase of the kinetic data [8]. An alternative approach is to use logistic models for the full range of PCR cycles. The 5-parameter logistic model considered for kinetic qRT-PCR data [10] is given by

$$F(x) = \frac{L + ((U - L))}{[1 + \exp\{k(x - x_0)\}]^b} \quad (7)$$

where L and U are the lower and upper asymptotes, respectively, k is the slope, x_0 is the inflection point, and b is the asymmetry parameter. With $b = 1$, the model (7) reduces to the 4-parameter logistic model proposed earlier for kinetic qRT-PCR data by [5]. Logistic models imply that efficiency is declining through the course of reaction, although in the early cycles it stays approximately constant at the maximum (Figure 2). Hence, the

estimated maximum may be used as PCR efficiency [9]. Logistic models are implemented in the R package qpcR [18], which is the most comprehensive library implementing standard methods for qRT-PCR quantification. In qpcR package, the PCR efficiency can be calculated from any point of the logistic model fit, but by default, the efficiency is estimated by $E = F(x_{d2})/F(x_{d2} - 1)$, where x_{d2} is the point of the second derivative maximum computed from the fitted logistic model.

3 Dynamics of PCR amplification and efficiency

3.1 Modeling dynamics of PCR amplification

Using enzymological considerations, Schnell and Mendoza [11] justified that the Michaelis-Menten model is appropriate for modeling PCR amplification dynamics without limitations due to consumption of reagents. Let us denote $T(x) = N(x)/K$, where $N(x)$ is the number of amplified templates at cycle x . Here $T(x)$ is the so called *reduced concentration* and K is the *Michaelis-Menten constant*. The Michaelis-Menten differential equation model for the reduced concentration $T(x)$ is given by (omitting the argument x)

$$\frac{dT}{dx} = \frac{T \ln E_m}{1 + T}, \quad (8)$$

where E_m is the maximum possible amplification rate or PCR efficiency (since $T(x) \geq 0$). In the early exponential phase, with $N(x) \ll K$, $T(x) \approx T(0)E^x$, and for $T(x) = T(0)E^x$,

$$\frac{dT}{dx} \frac{1}{T} = \frac{d}{dx} \ln(T(x)) = \ln E. \quad (9)$$

On the other hand, model (8) may be rewritten as

$$\frac{dT}{dx} \frac{1}{T} = \frac{\ln E_m}{1 + T},$$

and for x such that $N(x) \ll K$, $\frac{dT}{dx} \frac{1}{T} \approx \ln E_m$. Hence, in the early exponential phase,

Michaelis-Menten model (8) is approximately equivalent to the exponential model with base E_m . However, it does not describe the entire qRT-PCR reaction, because $T(x)$ in (8) grows without a limit at a constant asymptotic rate [19], but in reality each well has a limited amount of reagents, and $T(x)$ can grow only up to a certain maximum.

The right side of model (8) may be viewed [12] as a particular case of the EMAX model (also known as Hill model),

$$\frac{dT}{dx} = \frac{T^\lambda \ln E}{1 + T^\lambda}$$

where $\lambda > 0$. The EMAX model is common for modeling dose-response relationships in pharmacology and is also based on Michaelis-Menten kinetics. The classic EMAX model is also not flexible enough for modeling the entire range of PCR cycles, including the saturation phase, where the derivative of $T(x)$ eventually becomes zero when all reagents are consumed. Therefore, we consider a modified EMAX model for PCR dynamics:

$$\frac{dT}{dx} = \frac{T^\lambda \ln E}{1 + T^\lambda} - h(T), \quad (10)$$

where $\lambda > 0$ and term $h(T)$ may be viewed as a harvesting function in the context of population growth models (e.g. [13]). For kinetic qRT-PCR data that motivated the present work, using harvesting function $h(T) = hT^2$, where $h > 0$ is a constant, provides the best fit for most reactions in terms of the standard BIC goodness of fit criteria for nonlinear models. For other kinetic qRT-PCR data, different harvesting functions $h(T)$ may be more appropriate. Using that $T(x) = F(x)P^{-1}K^{-1}$ and $h(T) = hT^2$, model (10) is rewritten for fluorescence intensity $F(x)$ as

$$\frac{dF}{dx} = \frac{aF^\lambda}{1 + bF^\lambda} + cF^2, \quad (11)$$

where $a > 0$, $b > 0$, and $c > 0$. Model (11) is further referred to as the modified EMAX model with shorthand notation EMAX. This model also may be formulated in terms of the log-derivative of $F(x)$:

$$\varepsilon(F) = \frac{1}{F} \frac{dF}{dx} = \frac{aF^{\lambda-1}}{1 + bF^\lambda} + cF. \quad (12)$$

Function $\varepsilon(F(x))$ may be viewed as a continuous approximation to log-transformed discrete PCR efficiency. Indeed, the discrete PCR efficiency at cycle x is defined as $E(x) = T(x+1)/T(x) = F(x+1)/F(x)$. Hence,

$$\ln E(x) = \ln F(x+1) - \ln F(x) = \left. \frac{d \ln F(x)}{dx} \right|_{x=x^* \in [x, x+1]} = \left. \frac{1}{F(x)} \frac{dF(x)}{dx} \right|_{x=x^* \in [x, x+1]}$$

and $\ln E(x) = \varepsilon(F(x^*))$ for some $x^* \in [x, x+1]$. If $E(x) = E$ is constant, then clearly $\varepsilon(F(x)) = \ln E$. Thus, function $\varepsilon(F(x)) = [F(x)]^{-1} F'(x) = [T(x)]^{-1} T'(x)$ may be viewed as an *instantaneous log-efficiency*. Instantaneous log-efficiency functions vs. reduced

concentration $T(x)$ corresponding to logistic, Michaelis-Menten, and (modified) EMAX model (11) are shown in Figure 3.

3.2 Estimation of modified EMAX model

Model (10) may be viewed as particular cases of the nonlinear ordinary differential equations (ODE) models in the general form

$$\frac{dy}{dx} = f(y(x), \theta, x) \quad (13)$$

where θ is a vector of all unknown parameters in $f(y(x), \theta, x)$.

Generally, (13) does not have a closed form solution for $y(x)$. Therefore, the standard estimation of unknown parameters using nonlinear least squares methods [20, 21] requires numerous numerical integrations within a gradient-based method [22] or random search method (e.g. Section 4.5.2 in [27]), which is extremely computationally costly and potentially unstable. An attractive alternative is a two-stage approach [24–26], which results in considerable gain in computation time [23] and allows the use of robust estimation approaches to estimate the second stage model. In the first step, $y(x)$ is approximated nonparametrically using splines or kernel smoothing. Penalized splines of order 6 are recommended [28] to ensure sufficiently smooth derivative estimates with the penalty weight selected using generalized cross validation (GCV). Let us denote by $\tilde{y}(x)$ the resulting penalized spline and by $\tilde{y}'(x)$ the corresponding derivative function estimated by differentiating the spline $\tilde{y}(x)$. Then the estimates of θ may be computed as a least squares solution for the nonlinear regression,

$$\tilde{y}'(x_i) = f(\tilde{y}(x_i), \theta, x_i) + \varepsilon_i, \quad (14)$$

where $\{x_i\}$ is some grid on the interval $[1, M]$. The resulting estimator $\hat{\theta}$ has been used for a while for estimating ODE models [24, 26], but its consistency and asymptotic normality was only recently formally established by [25].

For estimating parameters in the modified EMAX model (11), we adapt the two-stage approach with $y(x) = \ln F(x)$, so that model (13) reduces to model (12). In our numerical studies, we found that more accurate PCR efficiency estimates result from fitted model (12) as compared to PCR efficiency estimates derived from fitted model (11). However, smoothing the log-transformed fluorescence measures would require exclusion of the linear ground phase with $F(x)$ values that are small in absolute value and possibly negative, and corresponding values of $\ln F(x)$ that are noisy or not even defined. To avoid the need to identify the end of the ground phase and beginning of the exponential phase, we smooth the background subtracted fluorescence $F(x)$ rather than $y(x) = \ln F(x)$. The smoothing is performed using the R package ‘fda’ developed to support functional data analysis as described in Ramsay and Silverman [28]. The resulting penalized spline $\tilde{F}(x)$ is then

differentiated to obtain spline representation for $F'(x)$ and $y'(x) = [F(x)]^{-1}F'(x)$. Then we tabulate $\tilde{y}(x_i) = \tilde{F}(x_i)$ and $\tilde{y}'(x_i) = [\tilde{F}(x_i)]^{-1}\tilde{F}'(x_i)$ on a grid $\{x_i\}$ in $[x_0, M]$, where x_0 is selected so that $\tilde{F}(x_i) > t_0$ for any $x_i > x_0$. The threshold t_0 is selected as the smallest level that allows removing the noise in $\tilde{y}'(x_i)$ estimates for small values of $\tilde{F}(x_i)$. In order to minimize the impact of any remaining noise in $\tilde{y}'(x_i)$ estimates, we employ MM-estimators [29] for nonlinear regression [30] to estimate parameter θ in (14). Such MM-estimators are highly robust with a minimal loss in efficiency [31]. They are implemented as part of the R package 'robustbase' (<http://r-forge.r-project.org/projects/robustbase/>).

Figure 4 shows examples of qRT-PCR reactions from the GUCY2C study data including (a) a GUCY2C standard with 10 templates; (b) a blood sample analyzed for GUCY2C mRNA; (c) a blood sample analyzed for TFRC mRNA; and (d) a blood sample analyzed for GAPDH mRNA. The dots represent tabulated empirical $\tilde{y}'(x_i) = [\tilde{F}(x_i)]^{-1}\tilde{F}'(x_i)$ estimates obtained from optimally smoothed fluorescence $\tilde{F}(x)$. The solid lines represent $e(F(x))$ vs. $F(x)$ corresponding to fitted EMAX, 4-parameter logistic, 5-parameter logistic, and Michaelis-Menten models. Only the modified EMAX model provides a good fit across the entire range of empirical $\tilde{y}'(x_i)$ estimates for all 4 sample reactions. The 5-parameter logistic model provides a good fit for higher values of $F(x)$, but it does not fit well $\tilde{y}'(x_i)$ when $F(x)$ is close to zero for the GUCY2C standard and TFRC. Meanwhile the range of the smallest values of $F(x)$, where $F(x) < t_C$ for some threshold t_C , is the most important for the purpose of relative quantification.

3.3 Approximation of efficiency for relative quantification

The fitted model (12) is used to compute the constant efficiency estimate necessary for relative quantification. Let t_C be the threshold used to compute the threshold cycle numbers ($t_C = F(C_t)$). Recall that efficiency adjusted relative quantification assumes approximately constant efficiency in cycles up to C_t . Although the instantaneous efficiency is not the same as the discrete efficiency, (9) implies that for the early exponential phase with approximately constant efficiency E , $e = \ln E$. Therefore, using the smoothed fluorescence $\tilde{F}(x_i)$ evaluated on a grid $\{x_i\}$ in $[x_0, M]$ and estimated parameter vector $\hat{\theta} = [\hat{a}, \hat{b}, \hat{c}, \hat{\lambda}]$ in model (12), we compute an average log-transformed instantaneous efficiency $\bar{e}$ in the range of cycle numbers x such that $t_0 < F(x) < t_C$ and then compute the average PCR efficiency for relative quantification as $\bar{e} = \exp(\bar{e})$. That is,

$$\bar{E}(t_C) = \exp\left\{\sum_{t_0 < \tilde{F}(x_i) < t_C} \bar{e}(x_i)\right\} = \exp\left\{\sum_{t_0 < \tilde{F}(x_i) < t_C} \left[\frac{\hat{a}\tilde{F}(x_i)^{\hat{\lambda}} - 1}{1 + \hat{b}\tilde{F}(x_i)^{\hat{\lambda}}} + \hat{c}\tilde{F}(x_i)\right]\right\}. \quad (15)$$

This PCR efficiency estimate is further referred to as EMAX efficiency. Formula (15) emphasizes that the PCR efficiency estimate needed for relative quantification generally depends on the threshold t_C . This is meaningful in the context of generally decreasing efficiency through the PCR cycles. The higher the threshold t_C , the more the need to take into account the decrease in efficiency between $x = 0$ and $x = C_t$. The choice of t_0 is also important because too low t_0 would allow including data points that may introduce excess

noise. Thus, t_0 and t_C are selected so that t_0 is sufficiently large to be above the background noise and the interval $[t_0, t_C]$ includes at least 3–5 grid points x_i used to compute .

4 Evaluation of accuracy of the PCR efficiency and dilution ratio estimates

Serial dilution experiments provide data that allow evaluating accuracy of efficiency estimation in comparison to the “gold standard” slope efficiency. Here, we used the qRT-PCR data from 48 serial dilutions of synthetic GUCY2C standard quantified using the ABI 7900 Sequence Detection System (Applied Biosystems) with Taqman probes in 48 different runs of the GUCY2C assay. GUCY2C standards were serially diluted to obtain samples with 10, 100, 1000, and 10,000 copies of synthetic GUCY2C templates, and three replicates of each dilution were amplified. The proposed EMAX efficiency estimates were compared to the most commonly used methods of individual reaction efficiency estimation based on the (i) exponential growth model; (ii) 4-parameter logistic model (L4); and (iii) 5-parameter logistic model (L5). The standard efficiency estimates (i)–(iii) were computed using the corresponding functions in R package qpcR [18]. The Michaelis-Menten model was not included in the comparison because this model fits poorly to the entire range of kinetic data (see Figure 4), and there is no existing algorithm for identifying the end of the log-linear phase (up to which the Michaelis-Menten model is appropriate) because this model has not been considered before for practical estimation of PCR efficiency. Meanwhile, the proposed modified EMAX model (11) includes Michaelis-Menten model as a particular case with $c = 0$ and $\lambda = 0$. Instead of Michaelis-Menten model, the exponential growth model (1) was included into comparison of methods because it is the most widely used model in applications. It also does not fit the entire range of kinetic data, and it was fitted to the subsets of the data points in the exponential range estimated as described in [8]. The nonparametric smoothed estimate of the log-transformed instantaneous efficiency in the left side of (14) was also investigated for the purpose of computing the PCR efficiency in (15), but results were inferior to any other considered method (data not shown since this approach has not been proposed or studied before).

The accuracy of different individual reaction efficiency estimation methods was compared in terms of the mean relative bias (%Bias) and root mean squared error (RMSE) of the estimated individual reaction efficiency with the “gold standard” slope efficiency (6) for the corresponding serial dilution treated as the true value of efficiency. Figure 5(a) provides the boxplots of all efficiency estimates, including the slope efficiency. The errors of individual efficiency estimates as compared to the “gold standard” slope efficiency are shown in Figure 5(b). Table 1 reports the mean relative bias (%Bias) and root mean squared error (RMSE) using each method of individual efficiency estimation. The hEMAX efficiency has the smallest absolute value of relative bias and the smallest RMSE. The estimates based on exponential and 4-parameter logistic model have negative bias, while the estimates based on 5-parameter logistic model overestimate the PCR efficiency (Figure 5). The estimates based on 4-parameter logistic model have smaller variability comparable to the variability in the slope efficiency estimates, but they also result in the largest absolute value of the mean relative bias (27.2%).

The accuracy of individual efficiency-adjusted relative qRT-PCR quantification was evaluated by estimating the known ratio between 10 and 10,000 copies of GUCY2C standard using the efficiency estimates from EMAX, exponential, 4-parameter and 5-parameter logistic models. Similar, or even smaller, ratios may be expected for quantification of the cancer-related biomarkers in blood. Figure 6 provides the boxplots of the relative errors of estimating ratios of 10 GUCY2C standard copies to 10,000 GUCY2C standard copies (the true log-ratio is $\ln(10/10,000)$). The mean relative % error and RMSE are summarized in Table 2. The log-ratio estimates based on the EMAX model yield the smallest relative error (−7.8%), while the RMSE is comparable for the log-ratio estimates based on EMAX and logistic models. Using exponential and 4-parameter logistic models for efficiency estimation resulted in high % bias (66% and 125%, respectively) of log-ratio estimates. Thus, the efficiency estimates based exponential and 4-parameter logistic models are not appropriate for computing relative GUCY2C expression, and only the modified EMAX and 5-parameter logistic models should be considered for analysis of the GUCY2C expression in patients blood samples.

5 Joint model of GUCY2C mRNA expression in the blood and colorectal cancer recurrence

In the GUCY2C study, blood samples were collected longitudinally after surgery from colorectal cancer patients. For the present analyses, 577 blood samples from 172 patients with two or more longitudinal samples per patient were assessed using the GUCY2C qRT-PCR assay. The data set included 32 patients with Stage 0/I (19%), 52 patients with Stage II (29%), and 90 patients (52%) with Stage III colorectal cancer. A Cox model for time to recurrence suggested that only chemotherapy was a significant predictor of longer time to recurrence (hazard ratio 0.26, 95% CI: 0.11–0.64, $p = 0.003$). Next, time to recurrence and GUCY2C expression in blood were analyzed in a joint model [32] using the R package JM [33, 34]. Since the efficiency estimates based exponential and 4-parameter logistic models resulted in highly biased estimates of the relative expression of synthetic GUCY2C, the GUCY2C mRNA expression in the blood samples was quantified using efficiency estimates based only on 5-parameter logistic and hEMAX models.

Let $g_{ij} = g_i(t_{ij})$ be the GUCY2C expression ratio for the i^{th} patient at post-surgery time t_{ij} . The *longitudinal process model* for changes in GUCY2C expression ratio over time was a random intercept and random slope model

$$g_{ij} = m_i(t_{ij}) + e_{ij}, \quad m_i(t_{ij}) = \beta_0 + \beta_1 t + b_{0i} + b_{1i} t, \quad (16)$$

where $e_{ij} \sim \mathcal{N}(0, \sigma^2)$ and vector of random effects $\mathbf{b}_i = [b_{0i}, b_{1i}] \sim \mathcal{N}(0, \Sigma)$ is assumed independent from e_{ij} . Let us also denote by $m'_i(t)$ the derivative of $m_i(t)$ with respect to time, $m'_i(t) = \beta_1 + b_{1i}$.

Further, for the i^{th} patient, let us denote by T_i^* the true recurrence event, by C_i the censoring time, by $T_i = \min(T_i^*, C_i)$ the observed event time, and by $\delta_i = I(T_i^* \leq C_i)$ the event indicator, where $I()$ is the indicator function. The standard approach to joint modeling of survival and longitudinal outcome is to assume that survival process for the i^{th} patient is described by hazard function

$$h_i(t | M_i(t), \mathbf{x}_i) = \lim_{s \rightarrow 0} s^{-1} Pr\{t \leq T_i^* < t + s | T_i^* \geq t, M_i, \mathbf{x}_i\} \quad (17)$$

where $M_i(t) = \{m_i(s), 0 \leq s \leq t\}$ is the history of the longitudinal process $m_i(s)$ up to time t and $\mathbf{x}_i$ is a vector of baseline covariates associated with the i^{th} patient. The following *relative risk models* [34] were considered:

$$\text{Model I: } h(t | \mathbf{b}_i) = h_0(t) \exp \{ \gamma' \mathbf{x}_i + \alpha_0 m_i(t) \} \quad (18)$$

$$\text{Model II: } h(t | \mathbf{b}_i) = h_0(t) \exp \{ \gamma' \mathbf{x}_i + \alpha_0 m_i(t) + \alpha_1 m_i'(t) \} \quad (19)$$

$$\text{Model III: } h(t | \mathbf{b}_i) = h_0(t) \exp \{ \gamma' \mathbf{x}_i + \alpha_1 m_i'(t) \}. \quad (20)$$

Here $h_0(t)$ is the baseline hazard function which is not specified but approximated by a B-spline or piecewise-constant function [33]. Vector γ is a regression coefficient vector for baseline covariates $\mathbf{x}_i$. For analysis of GUCY2C expression and time to recurrence, the baseline covariates include the indicator of chemotherapy (I_C), the indicator of Stage 0/I (I_{S0I}), and the indicator of Stage II (I_{SII}), so that $\mathbf{x}_i = [1, I_C, I_{S0I}, I_{SII}]$ and $\gamma = [\gamma_0, \gamma_1, \gamma_2, \gamma_3]$. Stage is included as a known predictor of time to recurrence, even though the difference between Stage II and Stage III patients only achieved borderline significance (hazard ratio 0.44, 95% CI: 0.19–1.03, $p = 0.058$) in the Cox proportional hazards model for time to recurrence.

For GUCY2C expression estimates based on 5-parameter logistic and EMAX models, the optimal joint model was selected as the one minimizing the Bayesian Information Criteria (BIC) among candidate models (18)–(20) with B-spline or piecewise-constant estimates of function $h_0(t)$, with 3 to 12 knots for B-spline or 3 to 9 pieces for piecewise-constant representation.

Table 3 presents results from the optimal joint model utilizing the GUCY2C ratios based on 5-parameter logistic models. This is model I given in (18), which implies that higher GUCY2C ratios are associated with lower values of the hazard function and longer time to

recurrence. This actually contradicts the plausible scientific hypothesis that higher GUCY2C ratios correspond to a larger number of tumor cells circulating in the blood.

The results from the optimal joint model utilizing the GUCY2C ratios based on EMAX models are presented in Table 4. This is model III given in (20), which implies that higher *slopes* of longitudinal trends in GUCY2C ratios are associated with higher values of the hazard function and shorter time to recurrence. It is estimated that 0.1 increase in the *slope* of the linear trend in GUCY2C ratios correspond to about 4-fold increase in the hazard ratio (Table 4).

Efficiency estimates based on 4-parameter logistic models were initially considered. The results from the optimal joint model (selected using BIC as described above) utilizing the GUCY2C ratios based on 4-parameter logistic models were essentially the same as the results based on 5-parameter logistic models (Table 3). The efficiency estimates based on the exponential model were not even feasible to consider because kinetic data for many GUCY2C reactions yielded estimated exponential phases with the number of cycles insufficient for reliable estimates of efficiency.

It should be noted that the uncertainty of fitting 5-parameter logistic and EMAX models was not taken into account for estimating the joint model, and the standard errors reported in Tables 3 and 4 do not adjust for this uncertainty. However, we should emphasize that this is a standard approach when relative RT-PCR expression levels are computed and then used as predictors of clinical outcomes. Any current method of estimating PCR efficiency is based on parameter estimates from a linear or nonlinear regression model (as described in section 2.2) but uncertainty of these parameter estimates is not taken into account when relative expression ratios are computed and then used as predictors of clinical outcomes.

6 Concluding remarks

The development of the new relative RT-PCR quantification approach was motivated by the unmet need for accurate quantification of low abundance cancer biomarkers. We have proposed a new approach for estimation of PCR efficiency needed for accurate efficiency-adjusted relative quantification. The new approach includes a new modified EMAX model for the rate of change in relative qRT-PCR efficiency and a suitably averaged estimated dynamic PCR efficiency, which may be used with any standard formula for efficiency-adjusted relative quantification. In our previous work [9], we reported improved quantification results using the 4-parameter logistic model for modeling kinetic qRT-PCR data as compared to using the exponential model. The 4-parameter logistic model works well when the kinetic qRT-PCR data are symmetric about the inflection point, which was the case for the data generated by ABI7000 platform used in [9]. However, the kinetic qRT-PCR data from the GUCY2C study were generated using the newer ABI7900 platform and were not symmetric about the inflection point. With a growing number of available qRT-PCR hardware platforms, it is unlikely that previously proposed models for fluorescence intensity as a function of cycle will accommodate the resulting variety of kinetic qRT-PCR data. In contrast, the proposed modified EMAX model describes the dynamics of qRT-PCR efficiency as a function of product accumulation and is expected to be invariant with respect

to implementation of qRT-PCR reactions. Thus, modeling the dynamics of qRT-PCR efficiency has potential to provide more accurate estimates of efficiency for qRT-PCR data generated by various qRT-PCR platforms.

Acknowledgments

7 Funding

This work was supported by the grant from the US National Institutes of Health (NCI R21 CA159816). The GUCY2C expression data were generated as part of the NIH funded study (NCI R01 CA095026).

References

1. Waldman SA, Hyslop T, Schulz S, Barkun A, Nielsen K, Haaf J, Bonaccorso C, Li Y, Weinberg DS. Association of GUCY2C expression in lymph nodes with time to recurrence and disease-free survival in pN0 colorectal cancer. *Journal of the American Medical Association*. 2009; 301(7):745–752. [PubMed: 19224751]
2. Chervoneva I, Li Y, Schulz S, Croker S, Wilson C, Waldman SA, Hyslop T. Selection of optimal reference genes for normalization in quantitative RT-PCR. *BMC bioinformatics*. 2010; 11(1):253. [PubMed: 20470420]
3. Peirson SN, Butler JN, Foster RG. Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Res*. 2003; 31:e73. [PubMed: 12853650]
4. Gentle A, Anastasopoulos F, McBrien NA. High-resolution semi-quantitative real-time PCR without the use of a standard curve. *Biotechniques*. 2001; 31(3):502–508. [PubMed: 11570493]
5. Liu W, Saint DA. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem*. 2002; 302:52–9. [PubMed: 11846375]
6. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001; 29:e45. [PubMed: 11328886]
7. Stahlberg A, Aman P, Ridell B, Mostad P, Kubista M. A quantitative real-time PCR method for detection of B-Lymphocyte Monoclonality by comparison of kappa and lambda Immunoglobulin Light Chain Expression. *Clinical Chemistry*. 2003; 49(1):51–59. [PubMed: 12507960]
8. Tichopad A, Dilger M, Schwarz G, Pfaffl MW. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res*. 2003; 31:e122. [PubMed: 14530455]
9. Chervoneva I, Li Y, Iglewicz B, Waldman S, Hyslop T. Relative quantification based on logistic models for individual polymerase chain reactions. *Statistics in medicine*. 2007; 26(30):5596–5611. [PubMed: 17968873]
10. Spiess AN, Feig C, Ritz C. Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. *BMC Bioinformatics*. 2008; 9:221. [PubMed: 18445269]
11. Schnell S, Mendoza C. Theoretical description of the polymerase chain reaction. *J Theor Biol*. 1997; 188:313–8. [PubMed: 9344735]
12. Dette H, Melas VB, Pepelyshev A, Strigul N. Robust and efficient design of experiments for the Monod model. *J Theor Biol*. 2005; 234:537–50. [PubMed: 15808874]
13. Brauer F, Castillo-Chavez C. *Mathematical models in population biology and epidemiology*. Springer Verlag; 2001.
14. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2^{-C_T} Method. *Methods*. 2001; 25(4):402–408. [PubMed: 11846609]
15. Higgins KM, Davidian M, Chew G, Burge H. The effect of serial dilution error on calibration inference in immunoassay. *Biometrics*. 1998; 54:19–32. [PubMed: 9544505]
16. Racine-Poon A, Weihs C, Smith AFM. Estimation of relative potency with sequential dilution errors in radioimmunoassay. *Biometrics*. 1991; 47:1235–1246. [PubMed: 1786316]
17. Demidenko E. *Mixed Models: Theory and Applications with R*. John Wiley and Sons; 2004.

18. Ritz C, Spiess AN. qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*. 2008; 24(13):1549–1551. [PubMed: 18482995]
19. Jagers P, Klebaner F. Random variation and concentration effects in PCR. *J Theor Biol*. 2003; 224:299–304. [PubMed: 12941588]
20. Hemker PW. Numerical methods for differential equations in system simulation and in parameter estimation. *Analysis and Simulation of biochemical systems*. 1972:59–80.
21. Stortelder WJ. Parameter estimation in dynamic systems. *Mathematics and Computers in Simulation*. 1996; 42(2):135–142.
22. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*. 1963; 11(2):431–441.
23. Voit EO, Almeida J. Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*. 2004; 20(11):1670–1681. [PubMed: 14988125]
24. Bellman R, Roth RS. The use of splines with unknown end points in the identification of systems. *Journal of Mathematical Analysis and Applications*. 1971; 34(1):26–33.
25. Brunel NJ. Parameter estimation of ODEs via nonparametric estimators. *Electronic Journal of Statistics*. 2008; 2:1242–1267.
26. Varah JM. A spline least squares method for numerical parameter estimation in differential equations. *SIAM Journal on Scientific and Statistical Computing*. 1982; 3(1):28–46.
27. Voit EO. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press; 2000.
28. Ramsay JO, Silverman BW. *Functional Data Analysis*. Spring; New York: 1997.
29. Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*. 1987; 15:642–656.
30. Stromberg AJ. Computation of high breakdown nonlinear regression parameters. *Journal of the American Statistical Association*. 1993; 88(421):237–244.
31. Maronna R, Martin D, Yohai V. *Robust statistics*. John Wiley and Sons; Chichester: 2006.
32. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330–339. [PubMed: 9147598]
33. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*. 2010; 35(9):1–33. [PubMed: 21603108]
34. Rizopoulos D. *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press; 2012.

Figure 1.
Kinetic qRT-PCR data for some blood samples in the GUCY2C study.

Figure 2.

Kinetic data from a single qRT-PCR reaction (solid circles) with fitted logistic curve (solid line), exponential growth curve (dotted line) and illustration of determination of the cycle number C_t corresponding to the threshold t (red solid line).

Figure 3.

Models for qRT-PCR instantaneous log-efficiency as a non-constant function of fluorescence F (red line for 4-parameter logistic model with $L = 0$, $U = 5\ln(2)$, $k = \ln(2)$, green line for 5-parameter logistic model with $L = 0$, $U = 4$, $k = -0.5$, $x_0 = 20$, $b = 1.6$, purple line for Michaelis-Menten model with $E_m = 2$, blue line for modified EMAX model with $a = 2.5$, $b = 8$, $c = -0.025$, $\lambda = 0.5$).

Figure 4.

Tabulated log-transformed instantaneous efficiency (empty black dots) and fitted non-constant functions of fluorescence F (red line for 4-parameter logistic (L4) model, green line for 5-parameter logistic (L5) model, magenta line for Michaelis-Menten (MM) model, blue line for modified EMAX (EMAX) model).

Figure 5.

(a) Individual reaction efficiency estimates in 465 reactions from 48 serial dilution experiments with copies of GUCY2C standard and (b) Errors of individual efficiency estimates as compared to the “gold standard” slope efficiency.

Figure 6.
Relative errors of $\log(10/10,000)$ estimates for GUCY2C standards.

Table 1

Mean relative % error (%Bias) and root mean squared error (RMSE) of individual reaction efficiency estimates from 48 serial dilution experiments with copies of GUCY2C standard.

Model	%Bias	RMSE
Exponential	-5.3	0.48
Logistic L4	-27.5	0.56
Logistic L5	13.9	0.34
EMAX	3.0	0.17

Table 2

Mean relative % error (%Bias) and root mean squared error (RMSE) of $\log(10/10,000)$ estimates for reactions from 48 serial dilution experiments with copies of GUCY2C standard.

Model	%Bias	RMSE
Exponential	66.2	1.98
Logistic L4	125.0	1.31
Logistic L5	-9.6	1.45
EMAX	-7.8	1.58

Table 3

Optimal joint model with GUCY2C ratios based on 5-parameter logistic model.

Effect	Parameter	Estimate	Standard Error	Hazard Ratio	p-value
Chemotherapy (I_C)	γ_1	-1.30	0.46	0.27	0.005
Stage 0/I vs. III (I_{S0I})	γ_2	-0.25	0.57	0.78	0.658
Stage II vs. III (I_{SII})	γ_2	0.63	0.48	1.87	0.188
GUCY2C association	α_0	-0.16	0.08	0.85	0.041

Table 4

Optimal joint model with GUCY2C ratios based on EMAX models. For GUCY2C association parameter, the hazard ratio corresponds to an average 0.1 increase in the longitudinal slope of GUCY2C expression.

Effect	Parameter	Estimate	Standard Error	Hazard Ratio	p-value
Chemotherapy (I_C)	γ_1	-1.33	0.46	0.26	0.005
Stage 0/1 vs. III (I_{S0I})	γ_2	-0.48	0.47	0.62	0.658
Stage II vs. III (I_{SI})	γ_2	-0.75	0.43	0.47	0.086
GUCY2C association	α_1	13.88	6.27	4.01	0.027