Journal of Pharmacokinetics and Pharmacodynamics (2019) 46:605–616
https://doi.org/10.1007/s10928-019-09661-4

ORIGINAL PAPER

Impact of Phase 1 study design on estimation of QT interval prolongation risk using exposure–response analysis

Nikolaos Tsamandouras[1] · Sridhar Duvvuri[1] · Steve Riley[2]

Received: 26 August 2019 / Accepted: 17 October 2019 / Published online: 29 October 2019

Abstract

The International Council for Harmonisation (ICH) guidelines have been revised allowing for modeling of concentration-QT (C-QT) data from Phase I dose-escalation studies to be used as primary analysis for QT prolongation risk assessment of new drugs. This work compares three commonly used Phase I dose-escalation study designs regarding their efficiency to accurately identify drug effects on QT interval through C-QT modeling. Parallel group design and 4-period crossover designs with sequential or interleaving cohorts were evaluated. Clinical trial simulations were performed for each design and across different scenarios (e.g. different magnitudes of drug effect, QT variability), assuming a pre-specified linear mixed effect (LME) model for the relationship between drug concentration and change from baseline QT (ΔQT). Analyses suggest no systematic bias in either the predictions of placebo-adjusted ΔQT (ΔΔQT) or the LME model parameter estimates across all evaluated designs. Additionally, false negative rates remained similar and adequately controlled across all evaluated designs. However, compared to the crossover designs, the parallel design had significantly less power to correctly exclude a clinically significant QT effect, especially in the presence of substantial intercept inter-individual variability. In such cases, parallel design is associated with increased uncertainty around ΔΔQT prediction, mainly attributed to the uncertainty around the estimation of the treatment-specific intercept in the model. Throughout all the evaluated scenarios, the crossover design with interleaving cohorts had consistently the best performance characteristics. The results from this investigation will further facilitate informed decision-making during Phase I study design and the interpretation of the associated C-QT modeling output.

Keywords QT prolongation · Concentration-QT modeling · Exposure–response modeling · Phase I study design · Power · Clinical trial simulations

Introduction

Early evaluation of QT prolongation risk of a drug under development is crucial to understand its benefit-risk relationship, facilitate associated decision-making and inform the intensity of electrocardiogram (ECG) monitoring during later stages of development [1]. Traditionally, QT prolongation risk assessment in the clinic was done with thorough QT (TQT) studies [1]. At the end of 2014, the consortium for Innovation and Quality in Pharmaceutical Development and the Cardiac Safety Research Consortium (IQ-CSRC) provided evidence from a prospective study [2] supporting the replacement of TQT studies by exposure–response analysis of early phase clinical data. Consequently, the International Council for Harmonisation (ICH) E14 Questions and Answers document was revised allowing for concentration (C)-QT modeling to be used as an

Sridhar Duvvuri was affiliated to Pfizer when the work was conducted.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10928-019-09661-4) contains supplementary material, which is available to authorized users.

✉ Nikolaos Tsamandouras
nikolaos.tsamandouras@pfizer.com

[1] Clinical Pharmacology, Early Clinical Development, Worldwide Research & Development, Pfizer, 1 Portland Street, Cambridge, MA 02139, USA

[2] Clinical Pharmacology, Global Product Development, Pfizer, Groton, CT, USA

alternative approach for assessing the QT prolongation risk of new drugs [3]. Based on these developments, a C-QT analysis using high quality ECG data and the associated concentration measurements from Phase I dose-escalation studies can be used in lieu of a dedicated TQT study. A scientific white paper was recently published providing recommendations and aligning industry and regulatory with regard to planning, conducting and reporting such C-QT modeling analyses [4].

Several different study designs are commonly used for Phase I dose-escalation studies (e.g. parallel designs, crossover designs with sequential or interleaving cohorts). The choice for a specific Phase I study design is based on several considerations, including the desired study duration, the pharmacokinetic (PK) and pharmacodynamic (PD) properties of the molecule and any potential for PK or PD carryover, safety issues, sampling limitations, the risk for subject dropouts and several other practical and operational concerns. However, based on the latest developments discussed above, an additional consideration when selecting a Phase I study design should now be its efficiency towards C-QT modeling analysis.

Clinical trial simulation methodologies have been employed to provide further insight regarding different aspects associated with C-QT modeling for QT prolongation risk assessment including: the operational characteristics of different C-QT model structures [5, 6]; the impact of study design parameters such as dose range, exposure margin and ECG variability [7]; and the impact of different computational platforms and confidence interval calculation methods [6]. However, the relative efficiency of the different types of Phase I study designs has never been investigated before, resulting in a significant knowledge gap.

Therefore, the aim of this work is to assess and compare, through simulations, commonly used Phase I study designs with regard to their efficiency/power to correctly identify drug effects on QT interval by a C-QT modeling approach as described in the recent white paper [4]. The results from this investigation will further facilitate informed decision-making during Phase I study design and the interpretation of the associated C-QT modeling output.

Methods

Evaluated study designs

Three different study designs that are commonly used in Phase I single ascending dose (SAD) studies were evaluated in this work using clinical trial simulations: (1) the parallel group design (Parallel); (2) the 4-period crossover, sequential cohorts, placebo substitution design (SEQ Crossover); and (3) the 4-period crossover, interleaving cohorts, placebo substitution design (ILV Crossover). A schematic illustrating these study designs is provided in Fig. 1. Across all study designs, single ascending doses between 1 and 800 mg were assumed to be evaluated (1, 3, 10, 30, 100, 200, 400 and 800 mg) in accordance with standard dose escalation practices.

The parallel group design (Fig. 1a) had a total sample size of 64 subjects equally randomized in 8 cohorts. Dose escalation is performed sequentially in this design with only one dose level evaluated per cohort. Within each cohort, 6 subjects are receiving active treatment and 2 subjects are receiving placebo.

The 4-period crossover, sequential cohorts, placebo substitution design (Fig. 1b) had a total sample size of 16 subjects, equally randomized in 2 cohorts. Dose escalation is initiated within the first cohort by escalating the dose across the 4 different periods (1, 3, 10 and 30 mg in period I, II, III and IV, respectively) and subsequently continued in the second cohort (100, 200, 400 and 800 mg in period I, II, III and IV, respectively). Each subject within a cohort receives 3 of the 4 allocated active doses and placebo. At any given period, 6 subjects within a cohort are receiving active treatment and 2 are receiving placebo.

Finally, the 4-period crossover, interleaving cohorts, placebo substitution design (Fig. 1c) had the same characteristics as the sequential cohorts crossover design described above with the difference that in the interleaving cohorts design, dose escalation is alternating between the two cohorts (e.g. starting dose level in Cohort 1, next dose level in Cohort 2, subsequent dose level back in Cohort 1, and so on). This alteration in dose escalation between the cohorts continues until the maximum dose level (800 mg) is achieved. Therefore, this design enables the evaluation of a wider dose range within the same subject compared to the sequential cohorts, crossover design.

Across all the designs the following post-dose time-points were assumed for coupled PK-ECG measurement collection: 0.5, 1, 2, 4, 8, 12, 16 and 24 h. Note that all the evaluated designs provide exactly the same total number of coupled post-dose PK-ECG measurements (512) for C-QT modeling (parallel design: 8 cohorts × 8 subjects × 8 measurements; crossover designs: 2 cohorts × 8 subjects × 4 periods × 8 measurements). Also, the distribution of these measurements across the different dose levels or placebo is identical in all designs (48 measurements at each of the 8 dose levels and 128 placebo measurements). Therefore, any observed differences across the designs regarding the C-QT modeling efficiency for QT prolongation risk assessment can be directly attributed to the fundamental properties of each design.

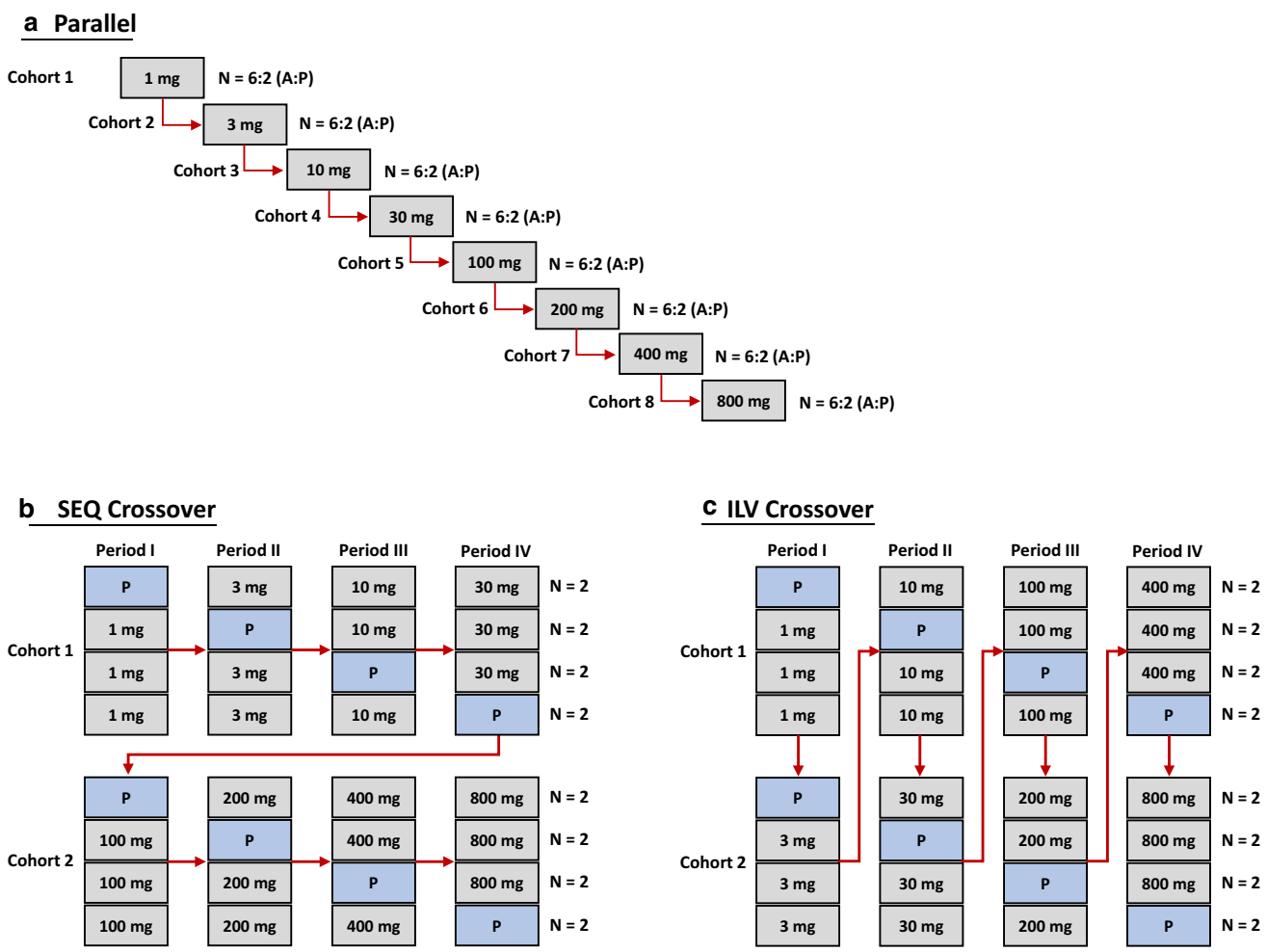

Fig. 1 Schematic of the evaluated study designs. **a** Parallel: parallel group design. **b** SEQ Crossover: 4-period crossover, sequential cohorts, placebo substitution design. **c** ILV Crossover: 4-period crossover, interleaving cohorts, placebo substitution design. *A* active drug, *P* placebo, *N* number of subjects

Population PK model used for the simulations

A theoretical 2-compartment PK model with first-order oral absorption and linear elimination was used to generate concentration–time profiles in all simulated studies. The associated fixed-effect PK parameters were: absorption rate constant (k_a) of 2 h^{-1}, apparent elimination clearance (CL/F) of 40 L/h, apparent inter-compartmental clearance (Q/F) of 3 L/h, apparent central volume of distribution (V_C/F) of 150 L and apparent peripheral volume of distribution (V_P/F) of 30 L. Inter-individual variability was assigned to all PK parameters assuming a log-normal distribution with a coefficient of variation (CV) of 50% for k_a and 30% for all other PK parameters (assuming no covariance structure between random effect parameters). Residual unexplained variability was also incorporated into the data using a proportional error model with a CV of 10%.

Exposure–response model used for the simulations

A pre-specified linear mixed effects (LME) model, as proposed in the recent white paper [4], was used to simulate the relationship between plasma concentrations and ΔQTc (change from baseline QTc, where QTc refers to QT measurements after heart rate correction) values (Eq. 1).

$$\Delta QTc_{ijk} = \left(\theta_0 + \eta_{0,i}\right) + \left(\theta_1 + \eta_{1,i}\right)C_{ijk} + \theta_2 TRT + \theta_3\left(QTc_{ij0} - \overline{QTc_0}\right) + \theta_k + \varepsilon_{ijk} \quad (1)$$

In Eq. 1, ΔQTc_{ijk} is the change from baseline QTc for subject i, in treatment j, at time k; θ_0 is the population mean intercept in the absence of any covariate effects (thus referring to placebo, at the first post-dose sampling time (0.5 h), for a subject with the population average baseline QTc); θ_1 is the population mean slope of the assumed linear association between concentration and ΔQTc; C_{ijk} is the drug concentration for subject i, in treatment j, at time k

(C_{ijk} is 0 for placebo); θ_2 is the covariate effect (treatment-specific intercept) associated with *TRT*, where *TRT* is a categorical covariate that takes the value of 0 for placebo and 1 for active treatment; θ_3 is the covariate effect associated with each individual's baseline QTc value (QTc_{ij0}) centered around the overall population mean of all the baseline QTc values ($\overline{QTc_0}$); θ_k is the covariate effect associated with each (apart from the first) sampling time k (i.e. for each sampling time a different covariate effect θ_k is estimated). Random effect terms referring to inter-individual variability were assigned on both the intercept θ_0 and slope θ_1 terms ($\eta_{0,i}$ and $\eta_{1,i}$ respectively), assuming normal distribution with mean [0,0] and an unstructured covariance matrix Ω; ε_{ijk} is the random effect term referring to residual variability in the ΔQTc data, assuming normal distribution with mean [0] and variance σ^2.

The parameter values of the Concentration-ΔQTc LME model used in this work for clinical trial simulations are presented in Table 1. For illustrative purposes, generated PK profiles, ΔQTc-time profiles and ΔQTc vs Concentration plots from one simulated study following each of the evaluated study designs are provided in Fig. 2.

Clinical trial simulation methodology

Using the models described above, 1000 studies were simulated for each of the 3 evaluated designs and across 11 different magnitudes (from 0 to 10 ms, in 1 ms increments) of drug effect on ΔΔQTc (time-matched placebo-corrected change from baseline QTc), resulting in a total of 33,000 simulated Phase I studies. For sensitivity analysis purposes, this simulation procedure was also repeated with 8 different perturbations on selected input parameters of the concentration-ΔQTc model (see Table 1), resulting in 264,000 additional simulated studies.

All study datasets were simulated and then fitted in R (version 3.4.1) using the *lme* function of the *nlme* package. Subsequently, the modeling outputs were used to calculate the model-predicted effect on ΔΔQTc and the associated

Table 1 Parameter values assigned to the concentration-ΔQTc LME model

Parameter	Default value	Sensitivity analysis[e]
Intercept (θ_0)[a]	− 5.76 ms	–
Slope (θ_1)	11 different slope values were evaluated generating drug effects (ΔΔQTc) of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 ms at the assumed supra-therapeutic concentration (geometric mean of Cmax at the 400 mg dose, i.e. 1892 ng/mL)[d]	–
Treatment-specific intercept (θ_2)	0.5 ms	0, 2.5 ms
Baseline effect (θ_3)	− 0.58 ms per unit difference of individual baseline QTc from the population mean	0
Time effects (θ_k)	1.93, 2.29, 3.53, − 3.15, − 0.28, 5.17 and 3.98 ms referring to 1, 2, 4, 8, 12, 16 and 24 h post-dose respectively	–
IIV on Intercept (η_0)[b, c]	8 ms	4 ms
IIV on Slope (η_1)[c]	75% CV	50% CV, 100% CV
RV (ε)[b]	5 ms	2.5 ms, 10 ms
Baseline QTc	Parallel design: individual values were sampled from a normal distribution with mean of 408 ms and standard deviation of 16 ms; 4-period crossover designs: individual values were sampled from the 4-dimensional multivariate normal with mean of 408 ms and standard deviation of 16 ms at each period and a within-subject correlation of 0.85 across periods	–

Input parameter values used in this work were inspired from internal C-QT modeling analyses (e.g. time effects) and literature data

Cmax maximum plasma concentration, *CV* coefficient of variation, *IIV* inter-individual variability, *RV* residual variability, *ΔΔQTc* time-matched placebo-corrected change from baseline QTc

[a]Refers to placebo, at the first post-dose sampling time (0.5 h), for a subject with the population average baseline QTc

[b]Reported as standard deviation

[c]Data were simulated assuming no correlation between Intercept and Slope random effects (η_0–η_1) and were fitted with an unstructured covariance matrix

[d]Slope values were calculated by re-arranging Eq. 2 to find what slope is needed to achieve a given average drug effect at the assumed supra-therapeutic concentration

[e]Sensitivity analysis was performed for selected input parameters by changing them to the values reported in the table (while keeping all other parameters to their default values) to assess their impact on conclusions

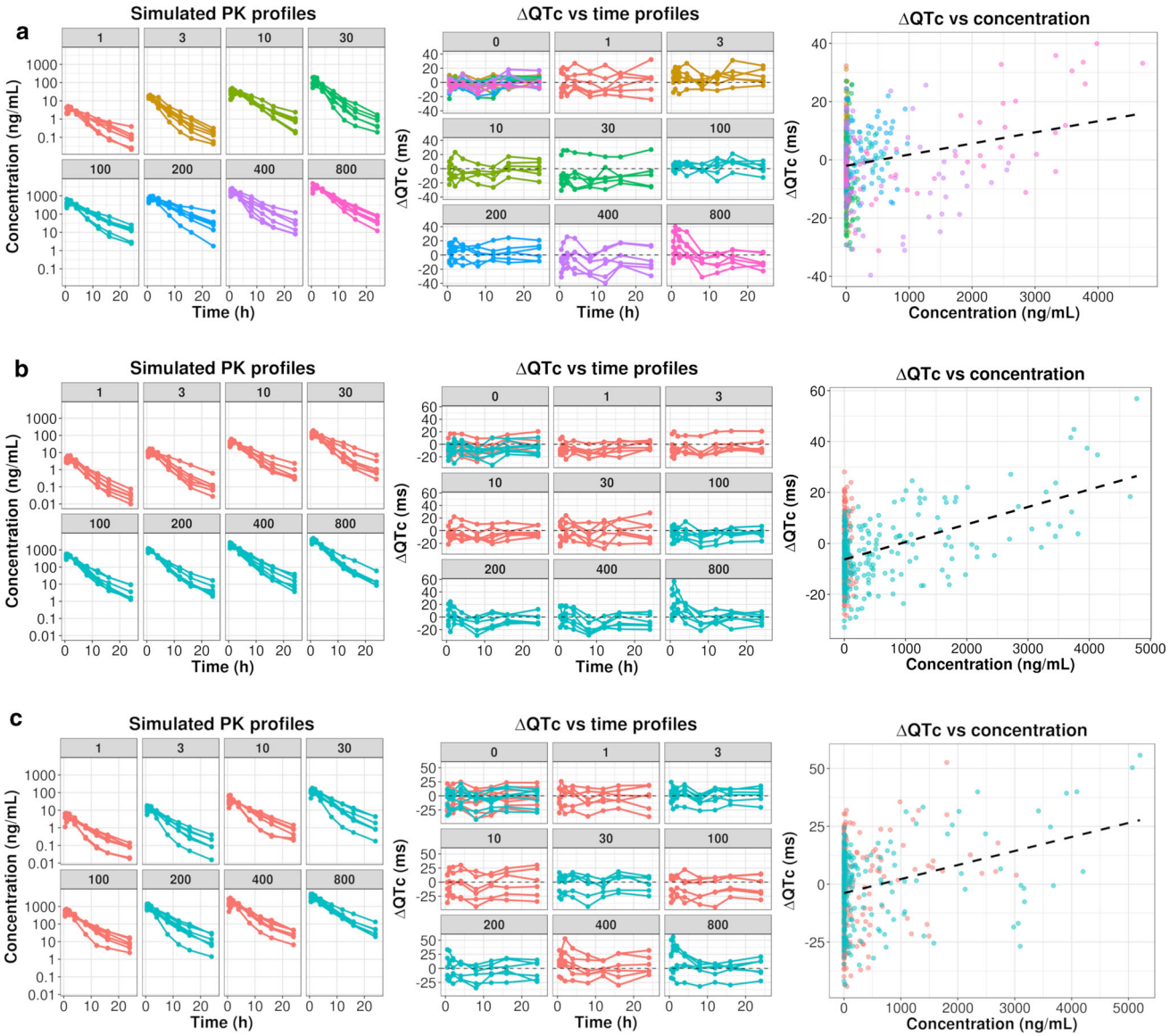

Fig. 2 Illustrative PK/QTc data from 1 simulated study with each of the 3 evaluated designs: **a)** Parallel (parallel group design); **b)** SEQ Crossover (4-period crossover, sequential cohorts, placebo substitution design); and **c)** ILV Crossover (4-period crossover, interleaving cohorts, placebo substitution design). Data from different cohorts in each design are highlighted with different colors. Subplot headers in the PK and the ΔQTc vs time plots indicate the administrated dose in mg (0 for placebo). The black dashed line in the ΔQTc vs concentration plots represents a linear regression trend-line. A slope that causes a drug effect on ΔΔQTc of 10 ms at the supra-therapeutic concentration (geometric mean Cmax at the 400 mg dose) has been employed to generate the simulated data above. All other input parameter values are reported in Methods for the population PK model and in Table 1 for the Concentration-ΔQTc LME model (default values)

two-sided 90% confidence intervals (CIs) at the assumed supra-therapeutic concentration, using the analytical approach suggested in [4] (Eqs. 2–4). The assumed supra-therapeutic concentration of interest for QT prolongation risk-assessment was set as the geometric mean of Cmax (maximum plasma concentration) at the 400 mg dose, i.e. 1892 ng/mL. Since doses up to 800 mg were evaluated in the simulated Phase I studies, a twofold exposure margin over the assumed supra-therapeutic concentration of interest was provided.

$$\Delta\Delta QTc(Conc) = \theta_{1,est} \cdot Conc + \theta_{2,est} \tag{2}$$

$$SE\Delta\Delta QTc(Conc) = \sqrt{Conc^2 \cdot \mathrm{var}(\theta_{1,est}) + \mathrm{var}(\theta_{2,est}) + 2\mathrm{cov}(\theta_{1,est}, \theta_{2,est}) \cdot Conc} \tag{3}$$

$$90\%\, CI\, \Delta\Delta QTc(Conc) = \Delta\Delta QTc(Conc) \pm t(0.95, df) \cdot SE\Delta\Delta QTc(Conc) \tag{4}$$

In Eqs. (2–4), *Conc* is the supra-therapeutic concentration of interest (see above); $\theta_{1,est}$ is the estimate of slope; $\theta_{2,est}$ is the estimate of treatment-specific intercept; $var(\theta_{1,est})$ is the variance referring to the uncertainty around the slope estimate; $var(\theta_{2,est})$ is the variance referring to the uncertainty around the treatment-specific intercept estimate; $cov(\theta_{1,est}, \theta_{2,est})$ is the covariance between the slope and the treatment-specific intercept with regard to the uncertainty around their estimates; t is the critical value determined from the t-distribution; *df* refers to the associated degrees of freedom; SE and 90% CI is the standard error and the two-sided 90% confidence interval respectively associated to the ΔΔQTc prediction. Note that in the current work, t was assigned the value of 1.645, assuming large enough degrees of freedom for normality approximation (the degrees of freedom reported in the *lme* output regarding the slope and the treatment-specific intercept was 486 for the two crossover designs; and 440 and 61 with respect to the slope and the treatment-specific intercept, respectively for the parallel design).

The probability to exclude a clinically important threshold of 10 ms prolongation on ΔΔQTc was calculated across the different designs and magnitudes of effect as the proportion of the simulated studies in which the upper bound of the two-sided 90% CI for ΔΔQTc prediction is less than 10 ms. This calculated probability was interpreted as the study power when the true magnitude of effect was indeed less than 10 ms and as the false negative rate when the true magnitude of effect was 10 ms. Additional outputs evaluated to aid the comparison between the different designs were bias, absolute error and 90% CI width with respect to ΔΔQTc predictions; and bias and precision with respect to the LME model parameter estimates. Finally, in the case that a substantial difference in power or false negative rate between the different designs would be detected, the change needed in the sample size of a given design to approximate the efficiency of the best performing design(s) was investigated (by performing the clinical trial simulation methodology described above with the appropriate sample size modifications).

Results

Comparison of the different designs regarding their efficiency for C-QT modeling

The model-predicted drug effect on ΔΔQTc, together with the associated 90% CIs for 100 randomly selected studies simulated with each of the 3 evaluated designs and assuming a true effect of 5 ms are illustrated in Fig. 3. No systematic bias (over- or under-prediction) can be observed across all 3 evaluated designs, as the model predictions appear to be randomly scattered around the true value of 5 ms. However, it was also evident that the parallel design was associated with higher uncertainty (wider CIs) in ΔΔQTc predictions compared to the crossover designs (see Fig. 3). Specifically, many of the studies with parallel design had the upper bound of the 90% CI crossing the clinical significance threshold of 10 ms, even for a true effect of 5 ms. On the contrary, crossover designs (and especially the one with interleaving cohorts) appear to clearly provide a more precise estimate around the true value of the drug effect on ΔΔQTc.

The probability to exclude a clinically significant effect of 10 ms on ΔΔQTc, across different magnitudes of true drug effect and across the three evaluated designs is presented in Fig. 4a. It was apparent that the parallel design had significantly less power to correctly exclude a 10 ms effect compared to crossover designs. Additionally, the interleaving crossover design appears to consistently outperform sequential crossover design in terms of power to correctly exclude a 10 ms effect. Specifically, parallel, sequential crossover and interleaving crossover designs had $> 80\%$ power for a drug effect of 3, 5 and 6 ms, respectively. On the other hand, the false negative rate (probability to exclude a 10 ms effect when the true drug effect was indeed 10 ms) was similar and well controlled across all designs (5.7% for the parallel design, 7% for the sequential crossover design and 5.6% for the interleaving crossover design).

Since a substantial difference in power between the different designs was detected, the change needed in the sample size of a given design to approximate the power of the best performing design(s) was investigated to further understand their relative efficiency. Specifically, it was found that to approximate the power curve achieved with the crossover designs, an increase in the sample size of the parallel design from 6:2 to 18:6 (active drug:placebo) subjects per cohort would be needed (Fig. 4b). This translates to a total sample size of 192 subjects in the Phase I study to get similar efficiency (in terms of power) with that achieved with just 16 subjects over 4 periods in the crossover designs. Regarding the relative efficiency of the two crossover designs, it was found that to approximate the power curve achieved with the interleaving crossover design, an increase in the sample size of the sequential crossover design from 16 to 24 subjects would be needed (Fig. 4c). This required 50% increase in sample size highlights the efficiency of interleaving rather than sequential cohorts for C-QT exposure–response analysis.

Additional outputs evaluated to aid the comparison between the different designs were bias, absolute error, and 90% CI width with respect to ΔΔQTc predictions (Fig. 5). Simulations illustrate (Fig. 5a) that no matter the study

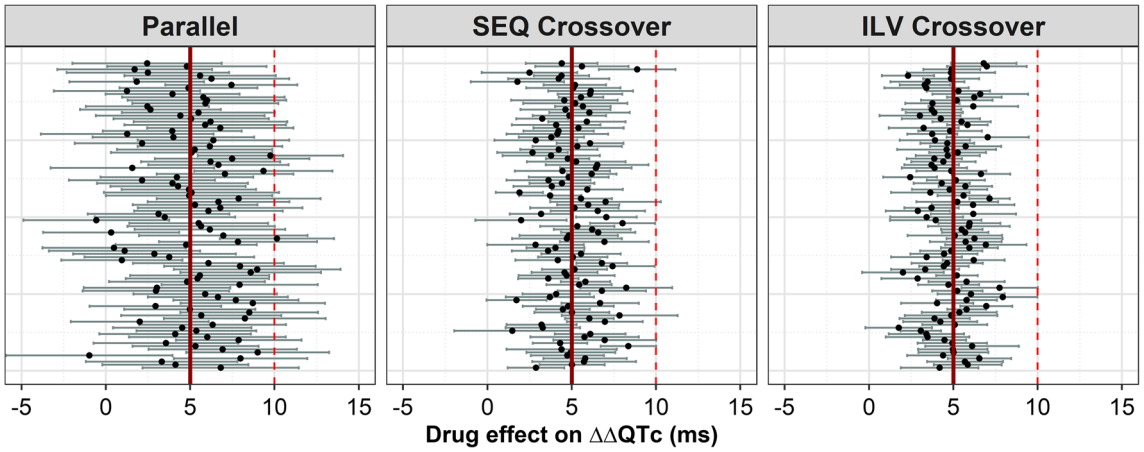

Fig. 3 Model-predicted drug effect on ΔΔQTc (black circles) together with the associated 90% CIs (grey bars) for 100 randomly picked studies simulated with each of the 3 evaluated designs and assuming a true effect of 5 ms (solid red line). The dotted red line corresponds to the clinical significance threshold of 10 ms. Parallel: parallel group design; SEQ Crossover: 4-period crossover, sequential cohorts, placebo substitution design; ILV Crossover: 4-period crossover, interleaving cohorts, placebo substitution design

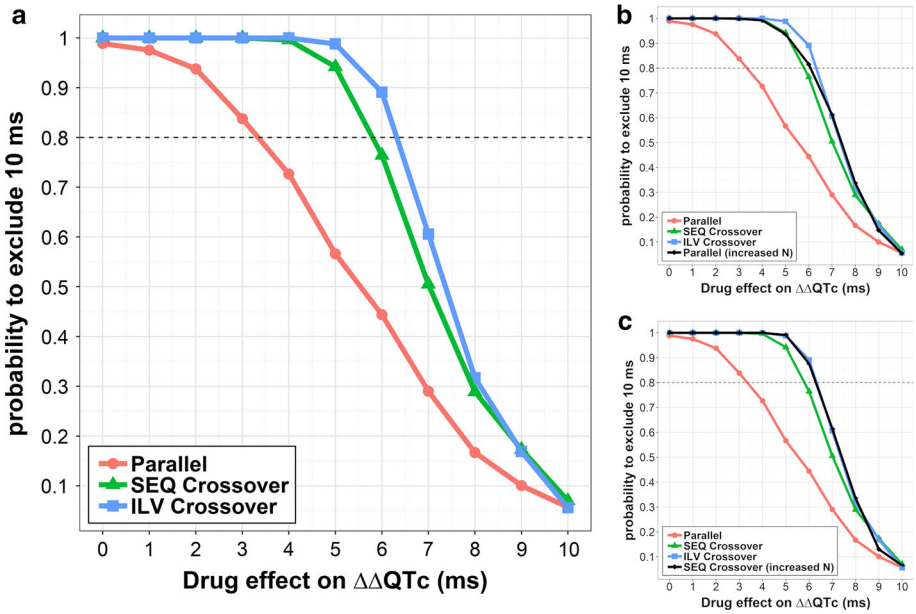

Fig. 4 **a** Probability to exclude a clinically significant effect of 10 ms on ΔΔQTc across different magnitudes of true drug effect (x-axis) and across the different evaluated designs (colored lines). The horizontal dashed line represents the nominal power of 80%. **b** As plot (a) with the addition of the power curve obtained with a parallel design of increased sample size [18:6 (active drug:placebo) subjects per cohort instead of 6:2]. **c** As plot (a) with the addition of the power curve obtained with a sequential crossover design of increased sample size (24 subjects instead of 16). Parallel: parallel group design; SEQ Crossover: 4-period crossover, sequential cohorts, placebo substitution design; ILV Crossover: 4-period crossover, interleaving cohorts, placebo substitution design

design or the magnitude of drug effect, C-QT modeling analysis does not yield any systematic over- or under-prediction of the true effect and bias is minimal (less than 0.21 ms across all evaluated designs and magnitudes of effect). However, it was also clear that the parallel design was associated with higher absolute error (a measure of how far away are C-QT model predictions from the true value) and larger 90% CI width (Fig. 5b, c, respectively) compared to the crossover designs. Additionally, it is apparent that the crossover design with interleaving cohorts outperformed sequential crossover design with respect to these characteristics, as it consistently yielded smaller absolute errors and tighter confidence intervals (Fig. 5b, c). Indicatively, for a true drug effect of 5 ms, a median absolute error of 1.81, 1.11 and 0.84 ms and a median 90% CI width of 8.68, 4.45 and 3.97 ms was obtained for the parallel, sequential crossover and interleaving crossover design, respectively. As expected, across all designs, a

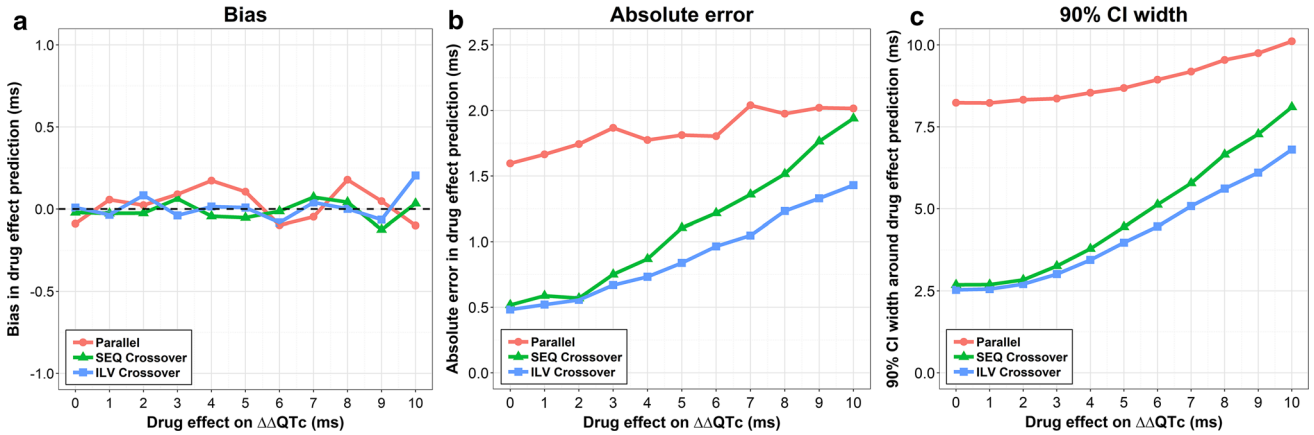

Fig. 5 Bias (**a**), absolute error (**b**) and 90% CI width (**c**) regarding ΔΔQTc prediction across different magnitudes of true drug effect (x-axis) and across the different evaluated designs (colored lines). For each simulated dataset, bias was calculated as the model predicted minus the true drug effect on ΔΔQTc; absolute error was calculated as the absolute value of [model predicted minus the true drug effect on ΔΔQTc]; 90% CI width was calculated as the upper minus the lower bound of the 90% CI (see methods for CI calculation). For all these three metrics, the median obtained across the 1000 simulated datasets for each design and magnitude of effect is presented. Parallel: parallel group design; SEQ Crossover: 4-period crossover, sequential cohorts, placebo substitution design; ILV Crossover: 4-period crossover, interleaving cohorts, placebo substitution design

trend of increase in absolute error and 90% CI width was observed along with the increase in the magnitude of the drug effect (Fig. 5b, c).

The bias and precision in the parameter estimates of the pre-specified LME model was also investigated across the different magnitudes of effect and study designs (Fig. 6). No significant bias was observed across all estimated LME model parameters (both fixed and random effects), irrespective of study design or magnitude of drug effect on ΔΔQTc. Specifically, the median of the parameter estimates obtained from the fitted datasets (1000 for each design and magnitude of effect) accurately matched the true parameters used for simulation, providing further confidence in the C-QT modeling approach and the employed estimation routine. Although all the LME model parameters were unbiased across all the evaluated designs, differences between the designs can be observed with respect to the precision of the estimates and particularly the treatment-specific intercept. Specifically, it is obvious that the parallel design is associated with substantial uncertainty around the estimation of the treatment-specific intercept, while this estimate is much more precise with crossover designs (Fig. 6). Since the estimate of the treatment-specific intercept together with its precision directly influence the prediction of drug effect on ΔΔQTc and the associated 90% CIs (Eqs. 2–4), uncertainty around the treatment-specific intercept is a major factor contributing to the wide 90% CI and the decreased power observed in parallel compared to crossover designs. Regarding the slope estimate (the other parameter directly influencing the prediction of drug effect on ΔΔQTc and the associated 90% CI), differences in precision between the designs were smaller, with the parallel design not performing worse than crossover designs. Additionally, it was observed that the crossover design with interleaving cohorts consistently yields less uncertainty around the estimate of slope compared to the sequential cohort crossover design (Fig. 6), explaining the tighter 90% CI and higher power observed with the former versus the latter design.

Sensitivity analysis

A sensitivity analysis was performed to assess the impact of selected input parameters of the concentration-ΔQTc LME model on the conclusions regarding the relative performance of the evaluated designs (Fig. 7). Although it was evident that power curves can be impacted (especially by assigning different values for the variability terms), the analysis showed that across all the evaluated input parameter perturbations, the interleaving crossover design consistently yielded the higher power, followed by the sequential crossover design and lastly the parallel design. However, across all the evaluated perturbations, the false negative rate remains very similar between the different designs and always below 10% (ranged from 4.8% to 9.6% across all perturbations and designs).

More specifically, assuming a smaller residual variability (standard deviation of 2.5 ms instead of 5 ms) results in an increase in power across all designs (e.g. from 44%, 76% and 89% to 46%, 84% and 96% for the parallel, sequential crossover and interleaving crossover design, respectively and a true effect of 6 ms). Assuming a larger residual variability (standard deviation of 10 ms instead of 5 ms) results, as expected, in a decrease in power across all

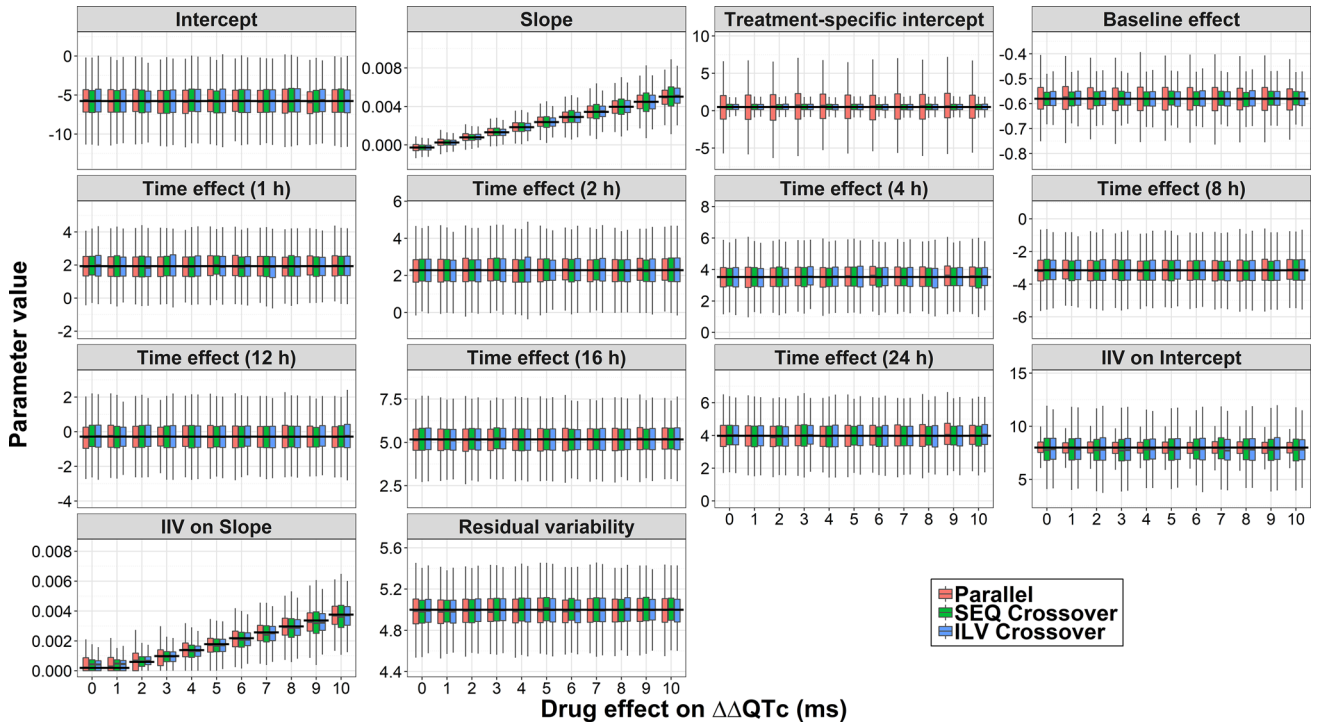

Fig. 6 Bias and precision in LME model parameter estimates. Horizontal black lines correspond for each parameter to the true value used in simulations (see Table 1) and boxplots correspond to the model estimates from the 1000 fitted datasets across the different magnitudes of drug effect (x-axis) and different designs (color stratification). All fixed effect parameters are reported in "ms" except of slope that is reported in "ms per ng/mL" and baseline effect that is reported in "ms per unit difference of individual baseline QTc from the population mean". Inter-individual variability (IIV) and residual variability are reported as standard deviations. Note that different slopes were used to generate the different magnitudes of effect (first row, second plot). Consequently, different slope IIV standard deviations were also used to achieve a constant coefficient of variation (CV) of 75% across the different magnitudes of effect (bottom left plot). Parallel: parallel group design; SEQ Crossover: 4-period crossover, sequential cohorts, placebo substitution design; ILV Crossover: 4-period crossover, interleaving cohorts, placebo substitution design

designs (e.g. from 44%, 76% and 89% to 36%, 56% and 64% for the parallel, sequential crossover and interleaving crossover design, respectively and a true effect of 6 ms).

Decreasing inter-individual variability on the slope (coefficient of variation of 50% instead of 75%) results in an increase in power across all designs (e.g. from 44%, 76% and 89% to 47%, 93% and 97% for the parallel, sequential crossover and interleaving crossover design, respectively and a true effect of 6 ms). Conversely, increasing inter-individual variability on the slope (coefficient of variation of 100% instead of 75%) results in a decrease in power across all designs (e.g. from 44%, 76% and 89% to 41%, 62% and 75% for the parallel, sequential crossover and interleaving crossover design, respectively and a true effect of 6 ms).

With regard to the inter-individual variability assigned on the intercept it was illustrated that its decrease (standard deviation of 4 ms instead of 8 ms) substantially increases the power of the parallel design (e.g. from 44 to 69% for a true effect of 6 ms), while crossover designs are relatively insensitive to this change (e.g. from 76% and 89% to 79% and 88% for the sequential crossover and interleaving crossover design, respectively and a true effect of 6 ms). As a result, under the assumption of a small inter-individual variability on the intercept, parallel design still yields inferior power compared to crossover designs, but the gap between them is much smaller (Fig. 7). The reason for this is that since parallel designs do not provide data after both treatment and placebo administration within the same subject (in contrast to crossover designs), uncertainty around the estimation of the treatment-specific intercept increases along with the increase in intercept's inter-individual variability. Characteristically, the bias and precision in the parameter estimates of the LME model for this scenario of reduced inter-individual variability on intercept can be found in Online Resource (Figure S1), which showed that the precision of the treatment-specific intercept in the parallel design is improved compared to the default scenario (Fig. 6). However, even in this case, the precision around the estimation of the treatment-specific intercept is still inferior to the crossover designs.

Additionally, different values of the treatment-specific intercept were also evaluated to assess the impact on our conclusions. Specifically, it was illustrated that by either

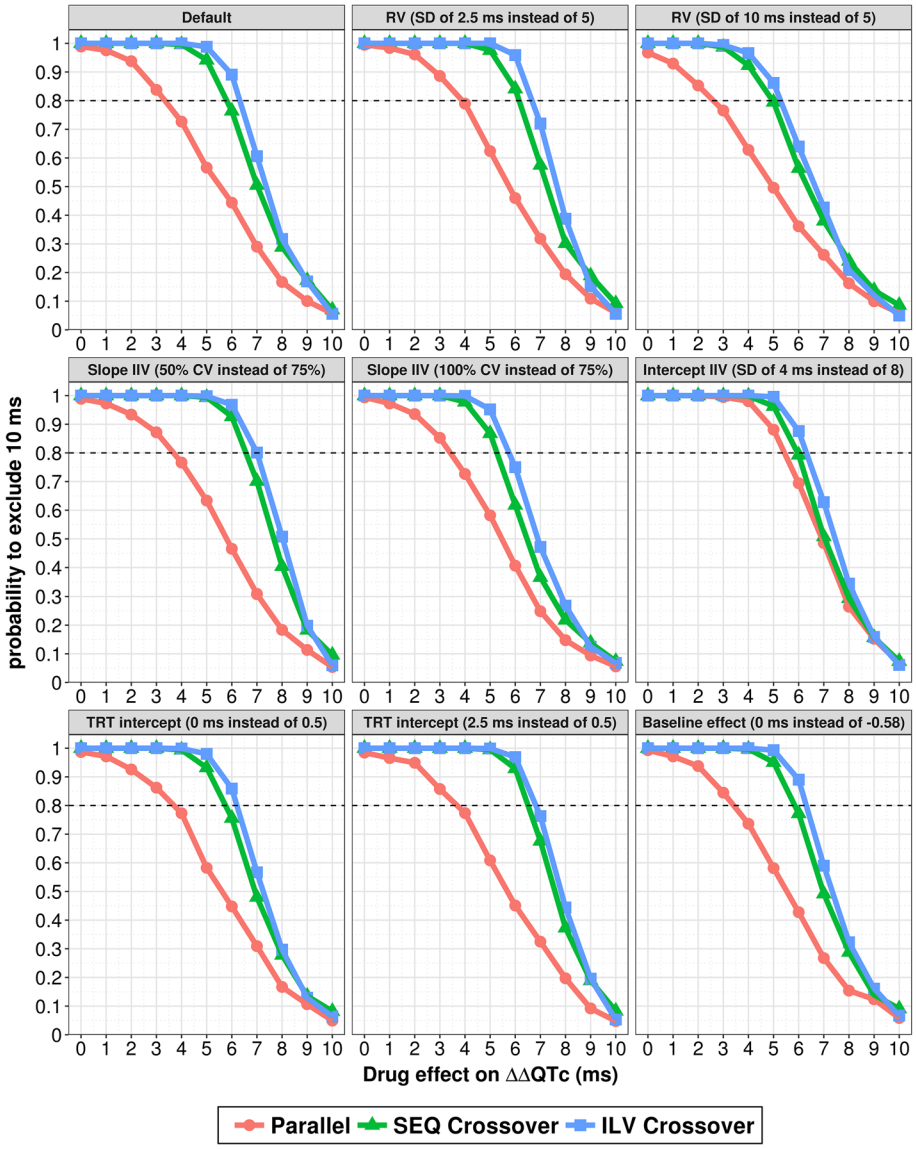

Fig. 7 Sensitivity analysis addressing the impact of perturbations in selected input parameters of the concentration-ΔQTc LME model on the performance of the evaluated designs. Top-left subplot corresponds to the default model parameters used in this work (see Table 1), while all other subplots correspond to perturbations from the default model parameters as described in each subplot's header. Plotted in each subplot is the probability to exclude a clinically significant effect of 10 ms on ΔΔQTc across different magnitudes of true drug effect (x-axis) and across the different evaluated designs (colored lines). The horizontal dashed line represents the nominal power of 80%. *RV* residual variability, *IIV* inter-individual variability, *SD* standard deviation, *CV* coefficient of variation, *TRT intercept* treatment-specific intercept, *Parallel* parallel group design, *SEQ Crossover* 4-period crossover, sequential cohorts, placebo substitution design, *ILV Crossover* 4-period crossover, interleaving cohorts, placebo substitution design

reducing (0 ms instead of 0.5 ms) or substantially increasing (2.5 ms instead of 0.5 ms) the assumed treatment-specific intercept used to simulate the data, the relative performance of the evaluated designs remained largely unaffected (Fig. 7). Finally, assuming the absence of a baseline effect (0 instead of -0.58 ms per unit difference of individual baseline QTc from the population mean) also did not have an impact on the performance of the different designs (Fig. 7).

Discussion

Subsequent to the ICH E14 guidance revision [3], it is expected that C-QT modeling analyses of PK and ECG data from Phase I dose-escalation studies will be increasingly used in the future in lieu of dedicated TQT studies [4]. Therefore, when designing a Phase I study it is important to also prospectively think of its efficiency towards C-QT analysis, amongst all other considerations. The current work employed clinical trial simulations to assess and compare 3 commonly used Phase I study designs with regard to their efficiency to correctly identify drug effects on QTc interval under the C-QT modeling framework that was described in the recent white paper [4].

Firstly, this work provided further evidence that exposure–response modeling of Phase I data using the prespecified LME model is very robust in detecting the true drug effect on QT interval (given that the underlying model assumptions hold [4]). Specifically, it was illustrated that no systematic bias (over- or under-prediction) was observed in either the LME model parameter estimates (Fig. 6) or the prediction of drug effect on ΔΔQTc (Fig. 5), across all the evaluated Phase I designs and regardless of

the magnitude of drug effect. Additionally, the false-negative rate (the probability to falsely conclude that the drug does not have a QT effect which exceeds the threshold of regulatory concern) was reasonably controlled and was very similar across all the evaluated Phase I designs (ranged from 5.6 to 7% for the default simulations and from 4.8 to 9.6% across all sensitivity analyses). This finding is important since it provides evidence that selection of each of these designs is not particularly critical from a safety/regulatory perspective as they all yield a low probability to miss a true and clinically significant QT signal.

However, this work clearly highlights that selection of a Phase I study design is particularly crucial in order to achieve adequate power to correctly exclude the possibility of a clinically significant QT signal and potentially waive the requirement for a dedicated TQT study. Particularly, it was found that the crossover design with interleaving cohorts has overall the best operational characteristics as it consistently provided (across all the evaluated magnitudes of drug effect and sensitivity analyses scenarios) the highest probability to correctly exclude a clinically significant effect on QT interval (Figs. 4 and 7). This increased power was also accompanied with the smallest absolute error and CI width in comparison to the other designs, while bias was sustained at minimal levels (Fig. 5). The superior performance of the crossover design with interleaving cohorts for C-QT population modeling analysis can be attributed to two main reasons: (1) in contrast to the parallel design, placebo and treatment data are obtained within the same subject, thus facilitating the estimation of the treatment-specific intercept; and (2) in contrast to the crossover design with sequential cohorts, a wider concentration range is evaluated within the same subject (Fig. 1c), thus facilitating the estimation of the slope. It can be speculated that the benefits arising from obtaining a wider concentration range within the same subject may also be generalized to the population exposure–response analysis of other non-ECG related endpoints of Phase I dose-escalation studies (e.g. vital signs).

Additionally, the results of this work provide evidence that the parallel design can potentially be inefficient compared to crossover designs for C-QT modeling analysis. It was specifically illustrated that it may lead to a substantial loss of power to correctly exclude a clinically significant effect and high uncertainty around the prediction of drug effect on ΔΔQTc. This can be mainly attributed to the increased uncertainty around the estimation of the treatment-specific intercept when using a parallel design (Fig. 6). The treatment-specific intercept is a parameter lacking a clear physiological interpretation. However, its use has been recommended and it is included in the pre-specified LME model as it gives flexibility to the model and reduces bias under the scenario of model misspecification (e.g. ignoring nonlinearity) [4, 5]. Apart from this indisputable advantage it brings in a pre-specified modeling framework, we illustrate here that it is also associated with the disadvantage that it may substantially decrease the power of a parallel design due the uncertainty around its estimation and the propagation of this uncertainty in the prediction of drug effect. Additionally, it was shown that the power loss in a parallel design through the difficulty in treatment-specific intercept estimation is increased along with the magnitude of the intercept's inter-individual variability (Fig. 7). Based on all the above it is clear that having the treatment-specific intercept as a default parameter in the LME model is not without limitations when using a parallel design and its impact may need to be carefully evaluated on a case-by-case basis.

Since the focus of this work was the comparison of different Phase I designs in the framework of the pre-specified white paper LME model [4], clinical trial simulations were performed under the premise that the basic assumptions of this model hold (linear C-QT relationship with no time delay). It is not expected that the different designs can exhibit different efficiency in alleviating the fundamental problems that will arise from model misspecification in case of assumption violation (e.g. nonlinear data fitted with a linear model).

In addition, it should be noted that as with any clinical trial simulation work, results are conditional on the input parameters used for model simulations. For this reason, several magnitudes of drug effect (by assigning different slopes) were evaluated and performed additional sensitivity analyses to examine the impact of other key input parameters of the LME model on our conclusions. Specifically, the evaluated ranges of residual variability of the model (standard deviation of 2.5-10 ms) and the slope's inter-individual variability (50–100% CV) are expected to sufficiently cover the clinically observed range based on literature and internally-obtained data. The intercept's inter-individual variability term is a parameter that can vary significantly across different studies (based on the authors' experience). Specifically, based on internal Phase 1 datasets this term has been observed to range anywhere between 3.6 and 9.7 ms. In this work a value of 8 ms (standard deviation) was assigned, however, a scenario with a reduced intercept's inter-individual variability (4 ms) was also evaluated to cover the lower end of the observed range. As discussed previously, this was identified to be a particularly sensitive parameter. In the presence of limited intercept variability, although the parallel design still yields numerically inferior power compared to crossover designs, the difference between them is substantially reduced and is potentially of less clinical importance. With regard to the treatment-specific intercept, an extensive range was also evaluated, since we assessed scenarios with no (0 ms),

small (0.5 ms) and large (2.5 ms) values assigned to this parameter. Finally, the sensitivity of our conclusions to the presence or absence of a baseline covariate effect was evaluated. Overall, this work illustrated that across all the evaluated input parameter perturbations described above, the relative performance of the different designs remains unchanged, with the interleaving crossover design and parallel design consistently yielding the higher and lower power, respectively to correctly exclude a clinically significant QT effect (Fig. 7). Values assigned to the remaining parameters of the LME model that were not assessed in sensitivity analysis (e.g. time effects) were not expected to affect the efficiency of the different designs towards ΔΔQTc prediction.

Finally, it should be highlighted, that Phase I study design depends on several considerations (Introduction) and efficiency of C-QT modeling analysis is only one of them. Therefore, although this study supports the use of a crossover design with interleaving cohorts to maximize power for C-QT analysis, there may be several cases that an alternative design may be a better option when considering all factors. For example, a parallel design might be more suitable for a compound with a very long half-life, or when there are safety concerns about exposing subjects to more than one dose of active compound. Also, a crossover design with sequential cohorts might be preferential to interleaving cohorts when there is concern regarding the subjects' study participation time and potential for drop-outs. Thus, this work does not aim to declare one design superior over the other, but rather provide the necessary context to informatively weigh advantages and disadvantages of each design and allow decision-making on a case-by-case basis.

Conclusions

This work highlights the competency of interleaving cohort, cross-over, Phase I study designs for QT prolongation risk assessment using C-QT modelling. In the event that no other considerations are limiting Phase I study design, the latter should be favored over other designs as they maximize the power to correctly exclude a clinically significant effect on QT interval, while retaining a low false negative rate. The results from this investigation will further facilitate informed decision-making during Phase I study design and the interpretation of the associated C-QT modeling output.

Acknowledgments The authors would like to thank the members of Pfizer's Early Clinical Development Clinical Pharmacology team for their comments and discussions.

Compliance with ethical standards

Conflict of interest All authors are employees and stockholders of Pfizer Inc.

References

1. FDA (2005) Guidance for Industry. E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. Silver Spring
2. Darpo B, Benson C, Dota C, Ferber G, Garnett C, Green C, Jarugula V, Johannesen L, Keirns J, Krudys K, Liu J, Ortemann-Renon C, Riley S, Sarapa N, Smith B, Stoltz R, Zhou M, Stockbridge N (2015) Results from the IQ-CSRC prospective study support replacement of the thorough QT study by QT assessment in the early clinical phase. Clin Pharmacol Ther 97(4):326–335. https://doi.org/10.1002/cpt.60
3. FDA (2017) E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs—questions and answers (R3). Guidance for Industry, SIlver Spring
4. Garnett C, Bonate PL, Dang Q, Ferber G, Huang D, Liu J, Mehrotra D, Riley S, Sager P, Tornoe C, Wang Y (2018) Scientific white paper on concentration-QTc modeling. J Pharmacokinet Pharmacodyn 45(3):383–397. https://doi.org/10.1007/s10928-017-9558-5
5. Garnett C, Needleman K, Liu J, Brundage R, Wang Y (2016) Operational characteristics of linear concentration-QT models for assessing QTc interval in the thorough QT and phase I clinical studies. Clin Pharmacol Ther 100(2):170–178. https://doi.org/10.1002/cpt.361
6. Lu J, Li J, Helmlinger G, Al-Huniti N (2018) Assessing QT/QTc interval prolongation with concentration-QT modeling for Phase I studies: impact of computational platforms, model structures and confidence interval calculation methods. J Pharmacokinet Pharmacodyn 45(3):469–482. https://doi.org/10.1007/s10928-018-9582-0
7. Ferber G, Sun Y, Darpo B, Garnett C, Liu J (2018) Study design parameters affecting exposure response analysis of QT data: results from simulation studies. J Clin Pharmacol 58(5):674–685. https://doi.org/10.1002/jcph.1065