ORIGINAL RESEARCH ARTICLE

Baseline Correction in Parallel Thorough QT Studies

Joanne Zhang · Qianyu Dang · Marek Malik

Abstract

Background In parallel thorough QT (TQT) studies, it has been speculated that either baseline correction should be omitted, under the assumption that it only adds noise to the data, or a time-averaged baseline instead of a time-matched baseline correction should be considered in order to reduce the study variability.

Objective This study characterized the assumptions and implications of different baseline correction approaches in parallel TQT studies submitted for regulatory review.

Data and methods 57 parallel TQT studies conducted between 2002 and 2009 in 5591 healthy volunteers were evaluated. Only moxifloxacin and placebo arms, including their baselines, were considered. The options of using no baseline correction, time-averaged baseline correction, and time-matched baseline correction were examined and compared.

Results QTc values exhibited a diurnal pattern, with longer QTc intervals during sleep preserved when correcting for a time-averaged baseline. Post-dose and baseline QTc values were highly correlated (mean $\rho = 0.80$, range 0.56–0.98 and mean $\rho = 0.79$, range 0.50–0.96 in the placebo and moxifloxacin groups, respectively). The variability of raw QTc values was substantially larger than that of baseline-adjusted QTc values. The difference in the point estimate of QTc differences between moxifloxacin and placebo differed by up to ± 4 ms between the time-averaged and the time-matched baseline corrections. Statistical tests indicate that assumptions of time-averaged baseline and no baseline correction are not appropriate.

Conclusions Baseline correction in parallel TQT studies leads to more precise QTc estimates. Because of possible inaccuracy introduced by time-averaged baseline correction, the time-matched baseline correction appears to be preferable for a parallel TQT study to both reduce the intrinsic variability due to circadian patterns and obtain more accurate point estimates.

J. Zhang (✉) · Q. Dang
Division of Biometrics VI, Office of Biostatistics, Center of Drug Evaluation and Research, Food and Drug Administration, 10903 New Hampshire Avenue, Bldg. #21 Room 4668, Silver Spring, MD 20993-0002, USA
e-mail: Joanne.Zhang@fda.hhs.gov

M. Malik (✉)
St. Paul's Cardiac Electrophysiology and St. George's, University of London, Verulam Avenue, Purley, Surrey, London CR8 3NQ, England
e-mail: Marek.Malik@btinternet.com

1 Introduction

The diurnal variation of the QT/QTc interval has been the subject of many investigations. Longer QTc intervals during sleep compared to the awake state have been repeatedly reported [1]. Substantial changes in autonomic balance during sleep have also been observed, with an increase in parasympathetic tone or a decrease in sympathetic activity or both, depending on the sleep stage [2]. Consequently, variations in the ECG parameters, including the QT intervals, over the course of 24 h were also reported [3]. Some

researchers reported that the coherence between the heart rate and the QT interval fluctuations is also age-related [4]. The ICH E14 [5] guidance recommends that practically all new drugs undergo a dedicated clinical study to exclude the possibility of drug-induced QT/QTc prolongation. Since most of these thorough QT (TQT) studies have multiple QT measurements over 24 h, the circadian pattern contributes to intrinsic QT interval variability, which needs to be addressed in the conduct and analysis of the studies. As ICH E14 indicates, "This can be accomplished in several ways, including the collection of multiple ECGs at baseline and during the study." In crossover studies, the same subject receives each of the treatments at different study periods, with ECG recordings at the same time points. Thus, under the assumption of no period effect, the diurnal variability of the QT/QTc interval cancels out when a time-matched difference of the two treatments (e.g., an active drug and a placebo) is considered. The situation is different in parallel studies, as each subject receives only one of the studied drugs, and the investigation of QTc drug effects is based on comparisons of data from different study subgroups. In this text, we therefore discuss the problems of baseline correction in parallel TQT studies.

While the current practice is to collect a whole day time-matched baseline on the day (or multiple days) prior to dosing in each subject, baseline correction in parallel TQT studies is still a matter of debate. In principle, there are three possibilities [see part I of the Electronic supplementary material (ESM) for a formal description of the assumptions]:

1. No baseline correction, with the on-treatment QTc data compared between different treatment groups. To our knowledge, this approach is not being used in TQT practice. However, suggestions that no baseline correction should be used in parallel studies have been made [6]. When the no baseline correction approach is considered, it needs to be assumed for each time point that when a study population is randomized into different subpopulations, the baseline QTc values averaged over each of the subpopulations will be the same. That is, this approach assumes that the QTc-averaged differences in different randomly selected parts of the study population are not statistically significantly different. In practice, the omission of baseline correction is based on the assumption that baseline values only add noise to the data. However, as we discuss later in this text, baseline correction reduces the variability in most situations.

2. Time-averaged baseline correction, with an averaged baseline QTc value obtained in each subject and subtracted from the on-treatment data in the same subject. (This also includes the cases where a full baseline day is not used, and where only a pre-dose QTc value is subtracted from the on-treatment data in the same subject.) The resulting within-subject differences are subsequently compared between different parallel study groups. This approach is frequently advocated in parallel TQT studies [7] because it is believed that it leads to a substantial reduction in data variability. This method makes a similar assumption to no baseline correction: when randomizing a study population into different subpopulations, the baseline QTc values averaged over each of the subpopulations will exhibit the same diurnal variation (but not necessarily the same QTc values).

3. Time-matched baseline correction, with the differences between baseline and on-treatment data obtained at each time point separately in each subject, and these differences compared between different parallel study groups. This approach is frequently dismissed in parallel TQT studies because it is assumed that, compared to time-averaged baseline correction, it only adds noise to the data. This correction only assumes that baseline QTc values and their diurnal pattern are the same in repeated recordings of the same subject.

To compare these baseline correction possibilities, we have evaluated 57 parallel TQT studies conducted in healthy volunteers. Our aim was to (1) characterize the diurnal QTc profile and to examine the differences between the baseline corrections in parallel TQT studies; (2) evaluate the necessity of collecting and correcting for baseline values; and (3) discuss how to optimally correct for baseline values in parallel TQT studies.

2 Methods

2.1 Data Sets

The data were obtained from 57 TQT parallel studies conducted between 2002 and 2009. All the studies were submissions from pharmaceutical sponsors to the FDA, and all were conducted in healthy volunteers, were placebo-controlled, used a single 400 mg dose of moxifloxacin as a positive control, and all recorded a whole-day time-matched pre-dose baseline. The analyzed studies are fully representative of parallel TQT investigations.

Only moxifloxacin and placebo data and their corresponding baselines were evaluated. The average sample size per study per arm was 49 subjects (22 females). In total, the 57 studies investigated 5,591 subjects on placebo or moxifloxacin (2,488 females; 44.5 %). The time interval between the baseline and post-dose varied from 1 to 29 days. The basic demographic distributions between the moxifloxacin and placebo groups were similar (Table 1).

Table 1 Characteristics of the 57 parallel TQT studies investigated

Characteristic	Placebo arm	Moxifloxacin arm
Mean N (total)	49	49
Mean N (female)	22	22
Minimum N (total)	22	21
Maximum N (total)	82	101
Total N	2,795	2,796
Mean age (y)	32.0	31.7
Minimum age (y)	18	18
Maximum age (y)	80	78
Minimum number of days between baseline and post-dose	1	1
Maximum number of days between baseline and post-dose	29	29

N number of subjects, *TQT* thorough QT

On-treatment and time-matched baseline QT measurements were available at 0.5, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, and 24 h relative to dosing. The number of studies using each time point was not evenly distributed because of different data collection schedules (Table 2). However, the majority of the studies included QT measurements at 1, 2, 3, 4, 6, 8, and 12 h post-dose.

All studies used 12-lead Holter recordings, and all had at least triplicate measurements at each time point. For the analysis, the mean QTc values of the triplicates (or of more replicates where available) were used at each time point. All studies also followed the standard practice of randomly assigning subjects who passed the screening tests to different study arms, including the placebo and moxifloxacin arms used in this investigation. In all studies, the randomization aimed at assigning the same number of subjects to the placebo and moxifloxacin arms. Each study also had active treatment arm/arms of investigated drugs, but the data of these arms are not used in this investigation.

2.2 Data Analysis

Fridericia's heart rate correction [8], which yields QTcF values, was utilized for all studies. This was adequate for the present purposes since neither placebo nor moxifloxacin lead to systematic heart rate changes [9]. However, if studying drugs that lead to systematic heart rate changes, individualized heart rate correction or other methods might need to be considered [9].

The diurnal profile of the QTcF intervals was evaluated in the baseline data for both the placebo and the moxifloxacin arms of the studies. For each subject, the oscillations around mean QTcF values were obtained for each time point and averaged in each study (see part I of the ESM). The averaged oscillations in each study were displayed graphically.

To judge the validity of the assumptions made in the no baseline correction and time-averaged baseline correction methods, the differences between the averaged baseline QTcF values in the placebo and moxifloxacin arms and the differences between the averaged baseline oscillations around the mean QTcF values in the placebo and moxifloxacin arms were calculated in each study and displayed graphically. The so-called global test [10] was applied to compare the averaged baseline QTcF values of the placebo and moxifloxacin arms. Studies were identified in which the difference between the averaged baseline QTcF values in the placebo and moxifloxacin arms exceeded 5 ms (see part I of the ESM). It is also possible to use so-called local tests [10] to perform the statistical test at each time point after multiple endpoint adjustment (results not reported here).

To evaluate the differences in the approaches to baseline correction, three differences between the moxifloxacin and placebo arms were calculated for each study, corresponding to the three correction possibilities. That is, in each study and at each time point, the three differences between the moxifloxacin and placebo arms were obtained. For comparison, the values corresponding to no baseline correction and time-averaged baseline correction were plotted against the time-matched correction values. The statistical test after adjusting for the number of time points was applied to compare the averaged circadian patterns of the baseline QTcF values of the placebo and moxifloxacin arms. Studies were identified in which the difference between the averaged circadian patterns of baseline QTcF values in the placebo and moxifloxacin arms exceeded 5 ms (see part I of the ESM).

In practice, when baseline QTc is included as a covariate in an ANCOVA model, the same results will be obtained using post-dose QTc values or ΔQTc. However, it is still necessary to decide on the kind of baseline that should be included as a covariate. Hence, the same considerations still apply.

For each study, Pearson's correlation coefficients between the post-dose and the time-matched baseline were

Table 2 Number of studies utilizing individual time points

Time point (hours after dosing)	0.5	1	2	3	4	5	6	8	10	12	14	16	18	24
Frequency (number of studies)	34	55	54	49	52	24	51	48	37	53	20	28	18	33

calculated at each time point and subsequently averaged for each study. To investigate the influence of the time between the baseline and the post-dose, the analysis was grouped according to the duration of this interval.

The final conclusion of TQT studies depends not only on the mean estimate of the difference between the treatment arm and placebo but also on the variability of the statistic expressed in practical terms as the width of the one-sided 95 % confidence interval (or the two-sided 90 % confidence interval) of the mean difference between the two treatments. Therefore, the widths of these confidence intervals were calculated at each time point of each study for all three correction possibilities (see part II of the ESM). The widths of the confidence intervals obtained with no baseline correction and time-averaged baseline correction were plotted against the widths of the confidence intervals obtained with time-matched correction. Comparisons of the widths of the confidence intervals were performed using Bland–Altman statistics [11].

The number of subjects influences the width of the confidence intervals of the mean difference for the sample. Therefore, we also compared the mean standard deviations of post-dose raw QTcF with the mean standard deviations of ΔQTcF (baseline-adjusted post-dose QTcF values). For each study, we first calculated the standard deviations at each time point for the raw QTc and ΔQTcF, and subsequently obtained the averages of these standard deviations across all time points for the raw QTcF and ΔQTcF.

3 Results

The profiles of diurnal QTcF oscillations in the baseline data are shown in Fig. 1. A clear diurnal pattern is seen, with larger QTcF values occurring during the night.

The top panel of Fig. 2 shows the differences between the mean baseline QTcF values in the placebo and moxifloxacin arms of the same studies across all time points. As can be seen, the differences are far from 0 for quite a number of studies. The figure shows that the differences between the mean baseline values of the placebo and moxifloxacin arms range from 9 to +10 ms. The figure shows that baseline QTcF values averaged in different randomized subgroups of a study population may differ substantially, so they cannot cancel each other out. This does not mean that the studies were conducted poorly, since variability in this range is to be expected from between-subject differences. The example in the bottom panel of Fig. 2 shows a study in which the averaged baseline QTcF values were statistically significantly different in the placebo and moxifloxacin arms. In 10 studies (17.5 %), the observed difference between the baseline

QTcF values averaged in the placebo and moxifloxacin is greater than 5 ms.

The top panel of Fig. 3 shows the differences between the mean baseline diurnal QTcF oscillations in the placebo and moxifloxacin arms of the same studies across all time points. While the ranges of the baseline diurnal QTcF variations are highly correlated between the placebo and moxifloxacin arms of the same studies, differences of up to ±6 ms were found between the baseline oscillations in the placebo and moxifloxacin arms of the same study. The figure shows that the baseline QTcF diurnal oscillations averaged in different randomized subgroups of a study population may differ substantially; therefore, after correcting for the time-averaged baseline, the diurnal variability is still preserved. (The diurnal oscillations in the figures show per-study averages of the oscillations around the mean value obtained separately in each subject.) Again, this does not imply any problem with how the studies were conducted, since variability in this range is to be expected. The example in the bottom panel of Fig. 3 shows a study in which the circadian patterns of QTcF values were statistically significantly different in the placebo and moxifloxacin arms. The observed difference between circadian patterns of baseline QTcF values averaged in the placebo and moxifloxacin arms exceeded 5 ms in 3 studies (5.3 %).

Figure 4 shows scatter diagrams of QTcF, the point estimates of the difference between the moxifloxacin and placebo arms found at different time points of the individual studies using the different baseline corrections. The differences between the time-matched baseline corrections and the other two baseline correction methods show the extent to which the assumptions of the no correction (top panel) and time-matched correction (bottom panel) methods were violated in the analyzed data (see part II of the ESM).

Figure 5 shows examples of two studies. In the study shown in the top panel, the difference between time-matched baseline correction and time-averaged baseline correction is negligible, but the difference from no baseline correction is much larger. In the study shown in the lower panel, all three baseline corrections yield different results at some time points.

Figure 6 shows Pearson's correlation coefficients between baseline and on-treatment QTcF values. These were similar for placebo and moxifloxacin and appeared to be independent of study duration (see part III of the ESM).

Figure 7 shows Bland–Altman plots of the width of the one-sided 95 % confidence intervals of the differences between the moxifloxacin and placebo arms obtained with different baseline correction methods (see part II of the ESM). As expected, no baseline correction resulted in a substantial increase in the variability of the mean difference between the moxifloxacin and placebo arms compared

Fig. 1 Mean diurnal variations of QTcF in the baseline data of placebo arms (*top panel*) and moxifloxacin arms (*bottom panel*). In each subject, diurnal variations around the mean baseline QTcF value were calculated and averaged in each study. The *panels* show the resulting averages. *Lines of different colors* correspond to different studies, with the same color coding used in both panels
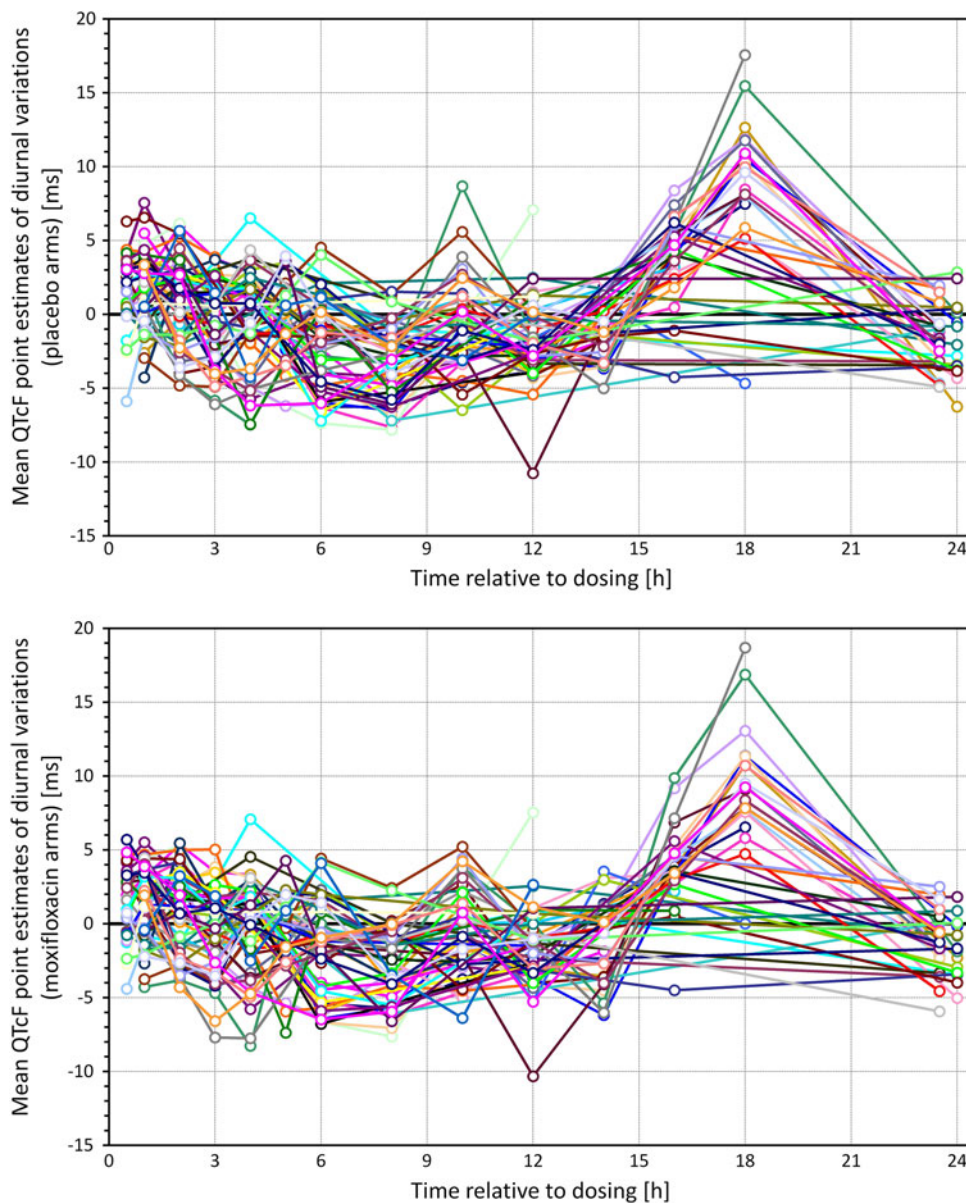
to the time-matched correction. Also, as expected, the time-averaged baseline correction led to a reduction of the variability of this difference compared to the time-matched correction. However, this reduction was numerically only tiny, with an average reduction of the width of the confidence interval of 0.34 ms.

Finally, Fig. 8 shows the scatter diagrams of the standard deviations of ΔQTcF versus the standard deviation of the post-dose QTcF for both the placebo and moxifloxacin groups. The findings from both placebo and moxifloxacin data are consistent: the standard deviation after the baseline correction (time-matched correction used in the figure) was almost always substantially smaller than that without the baseline adjustment. Figure 8 also shows scatter plots of the standard deviations of ΔQTcF, QTcF for moxifloxacin

versus ΔQTcF, QTcF for placebo. There was no systematic departure of the standard deviation for either ΔQTcF or QTcF between moxifloxacin and placebo.

4 Discussion

4.1 Variability Among Randomized Subgroups

Even though a parallel TQT study is a randomized clinical trial, there is still a difference between subgroups. As shown in Fig. 2, baseline QTcF values in different randomized subgroups of a study population may differ substantially (by up to ±10 ms in our data). As shown in Fig. 3, baseline QTcF diurnal oscillations in different

Fig. 2 Differences between the averaged baseline QTcF values in the placebo and moxifloxacin arms of the same study. In each study, mean baseline QTcF values were obtained in the placebo and moxifloxacin arms at each time point. The *top panel* shows the differences between these averages. *Lines of different colors* correspond to different studies. The *bottom panel* shows an example of a study within which the averaged baseline QTcF values in the placebo arm (60 subjects, *green graph*) and in the moxifloxacin arm (60 subjects, *red graph*) were statistically significantly different from each other (*p* = 0.0203 from the global test [10])
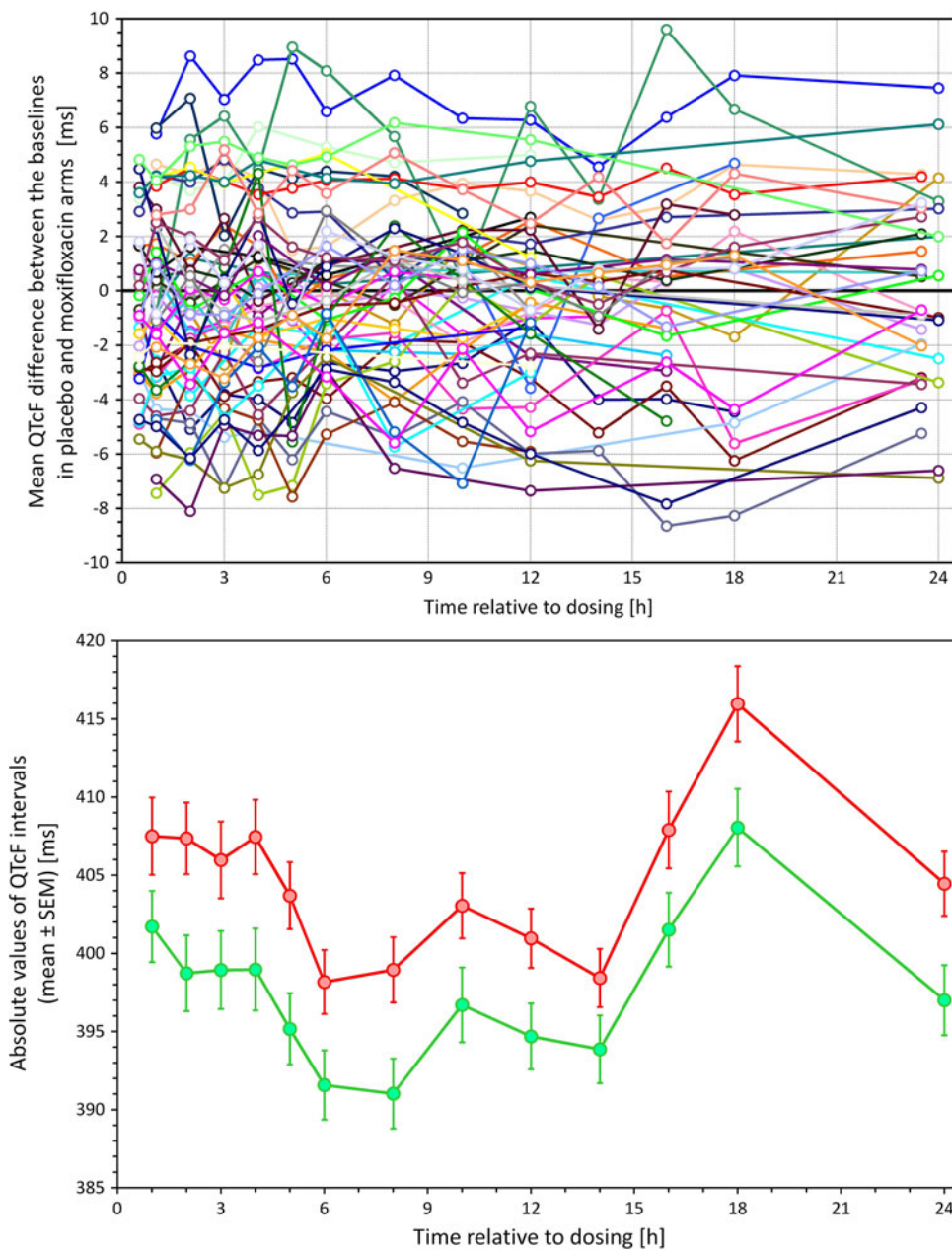

randomized subgroups of a study population may also differ substantially (by up to ±6 ms in our data). Thus, the speculation that "the expected mean diurnal effect patterns are the same for the study drug and placebo groups due to randomization" [6] is not entirely justified. Consequently, the argument that diurnal variability can be adjusted for by subtracting the placebo group mean from the treatment group mean does not appear to hold.

As already stated, the observed differences from the assumptions of the no baseline and time-averaged baseline correction methods do not mean that these studies were conducted improperly. The examples presented in Figs. 2 and 3 show studies in which the assumptions of the no

baseline and the time-averaged baseline corrections were formally rejected by statistical tests. However, the statistical rejection of the hypotheses does not solve the principal problem.

From a statistical point of view, the assumptions of the correction methods refer to the expected values, while the top panels in Figs. 2 and 3 show their estimates based on the observed data. A difference between expected values and their observed estimates is to be expected. However, the magnitude of the difference between the expected and observed values is important in this situation because it far too frequently exceeds the regulatory threshold used when judging TQT studies [5].

Fig. 3 Differences between the averaged circadian patterns of baseline QTcF values in the placebo and moxifloxacin arms of the same study. The *top panel* shows the differences in the baseline diurnal variations in the placebo and moxifloxacin arms of the same study (that is, the differences between the two panels in Fig. 1). *Lines of different colors* correspond to different studies, with the same color coding used as in Fig. 1. The *bottom panel* shows an example of a study in which the averaged circadian patterns of baseline QTcF values in the placebo arm (40 subjects, *green graph*) and in the moxifloxacin arm (40 subjects, *red graph*) were statistically significantly different from each other ($p = 0.0276$ from the local test [10] adjusted for the number of time points: 12). Note the substantial difference in the circadian patterns of the placebo and moxifloxacin subjects between 4 and 8 h post-dosing in the *bottom panel*
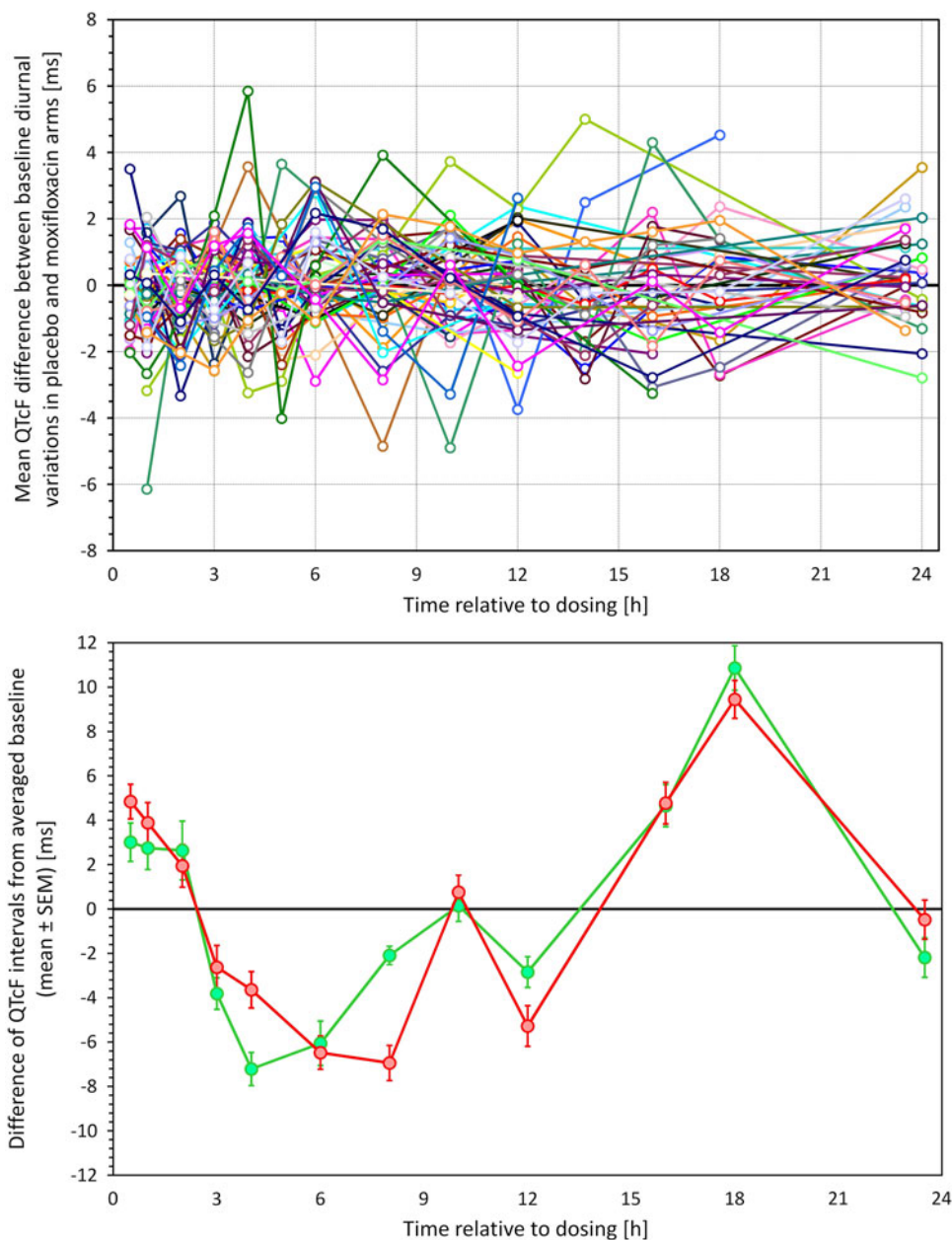

We were not able to reject the null hypothesis of no difference between the two groups (placebo and moxifloxacin arms) for a number of the studies. This is because a typical TQT study is not designed to test for a zero difference. The current sample size, which was designed to exclude a 10 ms difference between the study drug and placebo arms [5], is not appropriate for testing the null hypothesis of no difference between the two groups. Even though we cannot reject the null hypothesis of no difference, this does not mean that the two groups are the same. Based on the variability of current studies, much larger sample sizes are needed to test for a zero difference. Nevertheless, as already explained, point estimates still provide valuable information.

4.2 Correlation Between Post-Dose and Baseline

The data from the investigated studies convincingly show that the post-dose and pre-dose QTc measurements are highly correlated. As shown by Hills and Armitage [12] (see also part IV of the ESM), the baseline-corrected measure is superior to the uncorrected measure if the correlation coefficient between the baseline and the post-dose value exceeds 0.5. As seen in Fig. 6, this was the case in all of the investigated studies (in one particular

Fig. 4 Scatter diagrams of the point estimates of the difference between QTcF in the moxifloxacin and placebo arms found at individual time points of separate studies using different baseline corrections. The *top panel* shows a comparison of no baseline correction with time-matched correction. The *bottom panel* shows a comparison of time-averaged correction with time-matched correction. In both panels, the *bold solid line* shows the mean value while the *dashed lines* show the band of ±2 SD around the mean
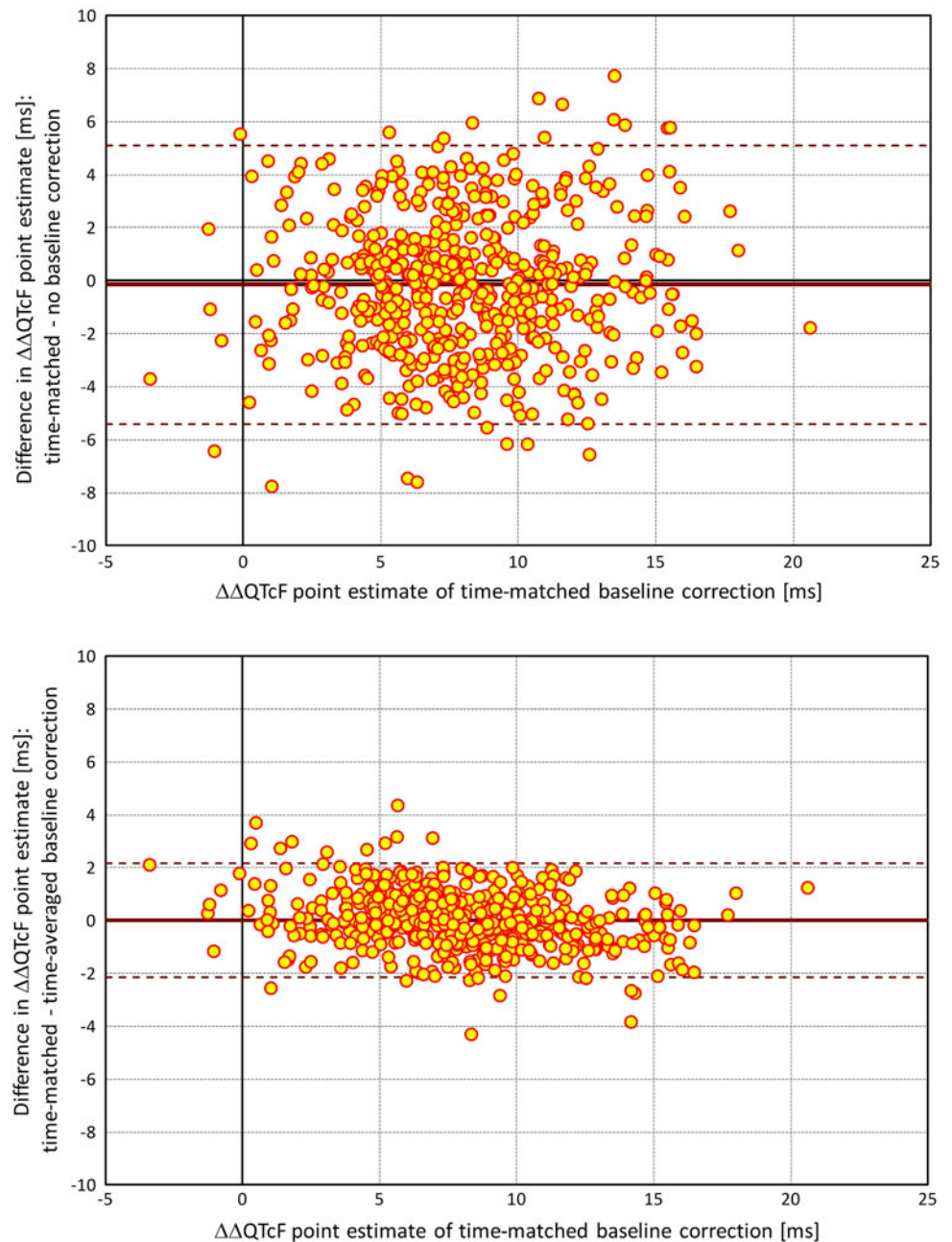

study, this correlation coefficient in the moxifloxacin arm was equal to 0.5). However, as explained in part IV of the ESM, even for correlation coefficients that are slightly below 0.5, the principle can still hold under certain circumstances.

Contrary to our expectations, the correlation coefficients between the post-dose and baseline did not weaken with increased study duration (i.e., with an increased time interval between the baseline and post-treatment data). It appears that the impact of the duration between the pre- and post-dose measurements is, if anything, fairly small.

4.3 Time-Matched Baseline and Time-Averaged Baseline

Since circadian patterns appear to differ among individuals, the concept of a mean population pattern lacks physiological justification. The individuality of QTc circadian patterns is not surprising. It fits well with the intra-individual stability and inter-individual variability of the QT/RR relationship [13, 14]. In earlier studies that suffered from inaccurate heart rate correction [15], the reported magnitude of the individual circadian variations of QTc was substantially exaggerated. When accurate corrections

Fig. 5 Examples of two studies in which different baseline corrections led to different results. The *top panel* shows a study in which time-averaged baseline correction led to a successful moxifloxacin test of assay sensitivity (with some lower bands of the one-sided 95 % confidence intervals above 5 ms), while the other baseline corrections did not. The *bottom panel* shows a study in which the point estimates of the moxifloxacin effect were also substantially different between time-averaged and time-matched baseline corrections at specific time points (differences of this magnitude can easily make the difference between a positive or a negative TQT study)
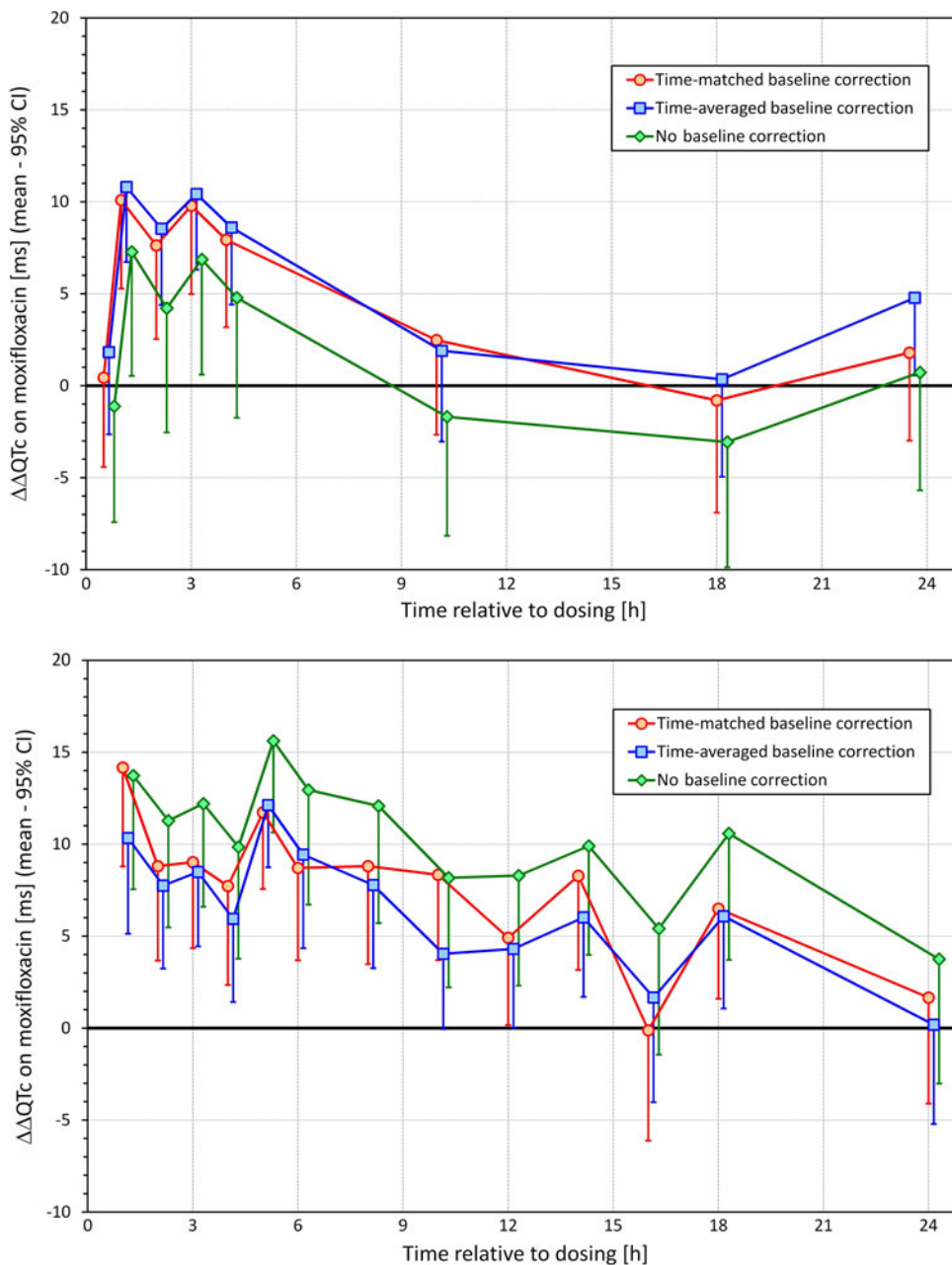

for heart rate and QT/RR hysteresis are made, the range of individual circadian variations was reported to be single milliseconds [16], which is in broad agreement with the differences observed in our data. Nevertheless, even inaccuracies of single milliseconds may lead to false-positive or false-negative conclusions of TQT studies. Therefore, time-matched baseline correction is physiologically more justified than time-averaged baseline correction, which, as we have already discussed, is based on assumptions that are not supported by our data.

Mathematically, it is not difficult to prove that a time-averaged baseline correction results in reduced variability (part V of the ESM). However, as we showed, differences

in diurnal patterns may still be present after adjusting for a time-averaged baseline. This could result in a false conclusion. When powering a TQT study, the main aim is to estimate the largest one-sided upper 95 % confidence bound for the baseline-adjusted mean difference between the study drug and placebo [5, 17]. While reducing the data variability is important, the central estimate must not depart from the truth. Both precision and accuracy are needed.

As shown in the bottom panel of Fig. 4, even though the difference in the point estimates of $\Delta\Delta QTcF$ between the time-averaged baseline and the time-matched baseline for moxifloxacin is evenly distributed around 0 across time, the

Fig. 6 Pearson correlation coefficients between baseline and post-dose QTcF values as a function of study duration
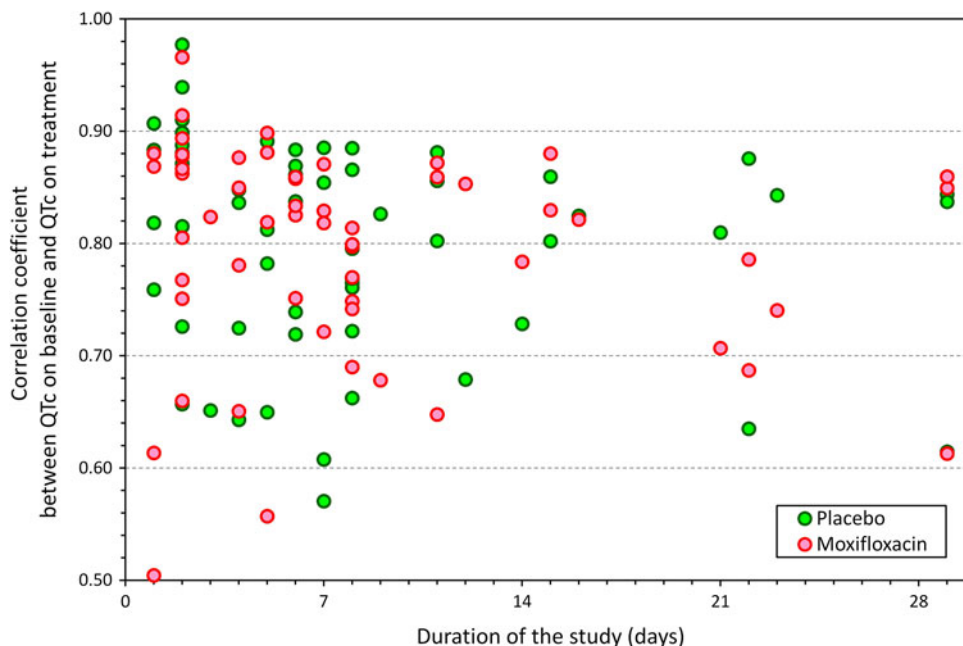
largest observed difference was as much as 4 ms. If this is projected into the outcome of the study drug, a negative TQT study may become positive or vice versa solely if the baseline correction method is changed. Thus, when a time-averaged baseline is used, the estimate will have a smaller variability, but the accuracy may be compromised (see part VI of the ESM for an example).

In a large number of the cases shown in Fig. 4, the differences between time-averaged and time-matched baseline corrections were within the ±2 ms band. This could be perceived as insignificantly small as well as comparable to other possible sources of error in the data from TQT studies. However, even differences of this magnitude can lead to false-positive and false-negative conclusions of TQT studies. While other sources of data inaccuracy (e.g., imprecision due to signal quality) may be difficult to control, inaccuracy resulting from the baseline correction employed can easily be eliminated by modifying the design of the TQT study and its statistical analysis plan. The evidence that significant differences can occur between different baseline-correction methods suggests that the physiologically more appropriate time-matched baseline correction should be used universally.

It might be argued that the time-matched method is susceptible to faulty readings at a particular time point, which would result in a biased estimate. However, this is a measurement error issue rather than a baseline correction issue; measurement errors can lead to wrong conclusions of TQT studies irrespective of their design and/or statistical evaluation. Moreover, if substantial measurement errors occurred for some study subjects, the standard deviation of the baseline-corrected QTc values would increase, which

then would result in a wider confidence interval of the mean estimates and QTc changes. Evaluating the confidence interval of the mean estimates of QTc changes is an integral part of any TQT study evaluation [5]. (Note that the differences between the time-matched and time-averaged baseline corrections shown in the bottom panel of Fig. 5 are not accompanied by changes in the widths of the confidence interval.)

For a well-controlled study, the diurnal variability can be controlled through a time-matched baseline adjustment. Moreover, as we have shown in Fig. 7, the differences in the widths of the confidence bounds obtained with time-averaged and time-matched baseline correction are not large. In practical terms, the increase in the power of the study achieved by using time-averaged baseline correction rather than time-matched baseline correction is only fairly modest.

Thus, to summarize the distinction between time-matched and time-averaged baseline corrections: the data from the analyzed studies show that these baseline-correction techniques lead to different results which may, in individual studies, be substantial enough to lead to different conclusions. While it is true that time-averaged baseline correction (and likewise correction for multiple pre-dose measurements, which has practically the same properties and is based on the same assumptions) leads to lower variability of point estimates, the reduction of the variability is small in comparison to the time-matched baseline correction. The individual-specific properties of ventricular repolarization and its dynamics are the physiologic basis for the time-matched baseline approach. We also believe that the individual-specific properties are why the averaged

Fig. 7 Bland–Altman plots comparing the width of one-sided 95 % confidence intervals (CI) of the difference between the moxifloxacin and placebo arms using no-baseline correction versus time-matched baseline correction (*top panel*) and using time-averaged baseline correction versus time-matched baseline correction (*bottom panel*). In both panels, the *bold solid line* shows the mean value, while the *dashed lines* show the band of ±2 SD around the mean
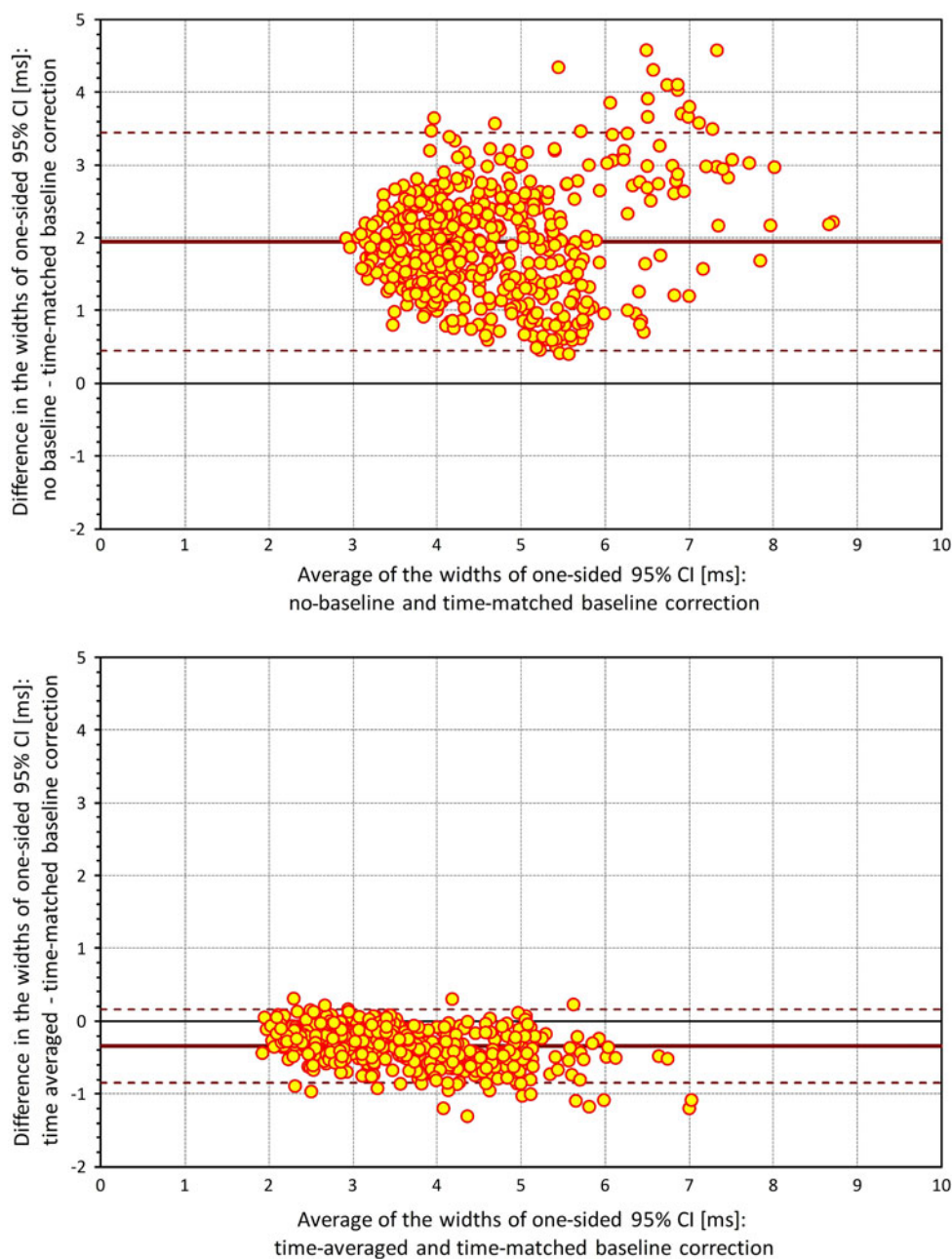
diurnal QTc patterns in two randomized subgroup populations in the same study are not necessarily the same. Therefore, time-matched baseline correction is a safer way to obtain valid estimates of drug-induced QTc changes.

4.4 Limitations

The limitations of our analyses also need to be considered. We used Fridericia correction, which is not meant to suggest that it is the best correction method for all TQT studies. However, since neither moxifloxacin nor placebo leads to systematic heart rate changes, the differences in the analyzed data resulting from using different heart rate correction methods are of little importance [9]. We have not presented sex-specific analyses, although these have also been performed. The results separated by sex are practically identical to those reported here. The lack of influence of the study duration on the correlation between baseline and post-treatment data might be due to the small number of studies with substantial durations (see Fig. 6). We have also not considered the quality of the QTc data [18, 19] in the evaluated studies, which might have contributed to some of the lower correlation coefficients shown in Fig. 6. External factors are known to influence the QTc interval, so if some external factors (such as responses to weather changes) differed between the baseline and on-treatment recordings,
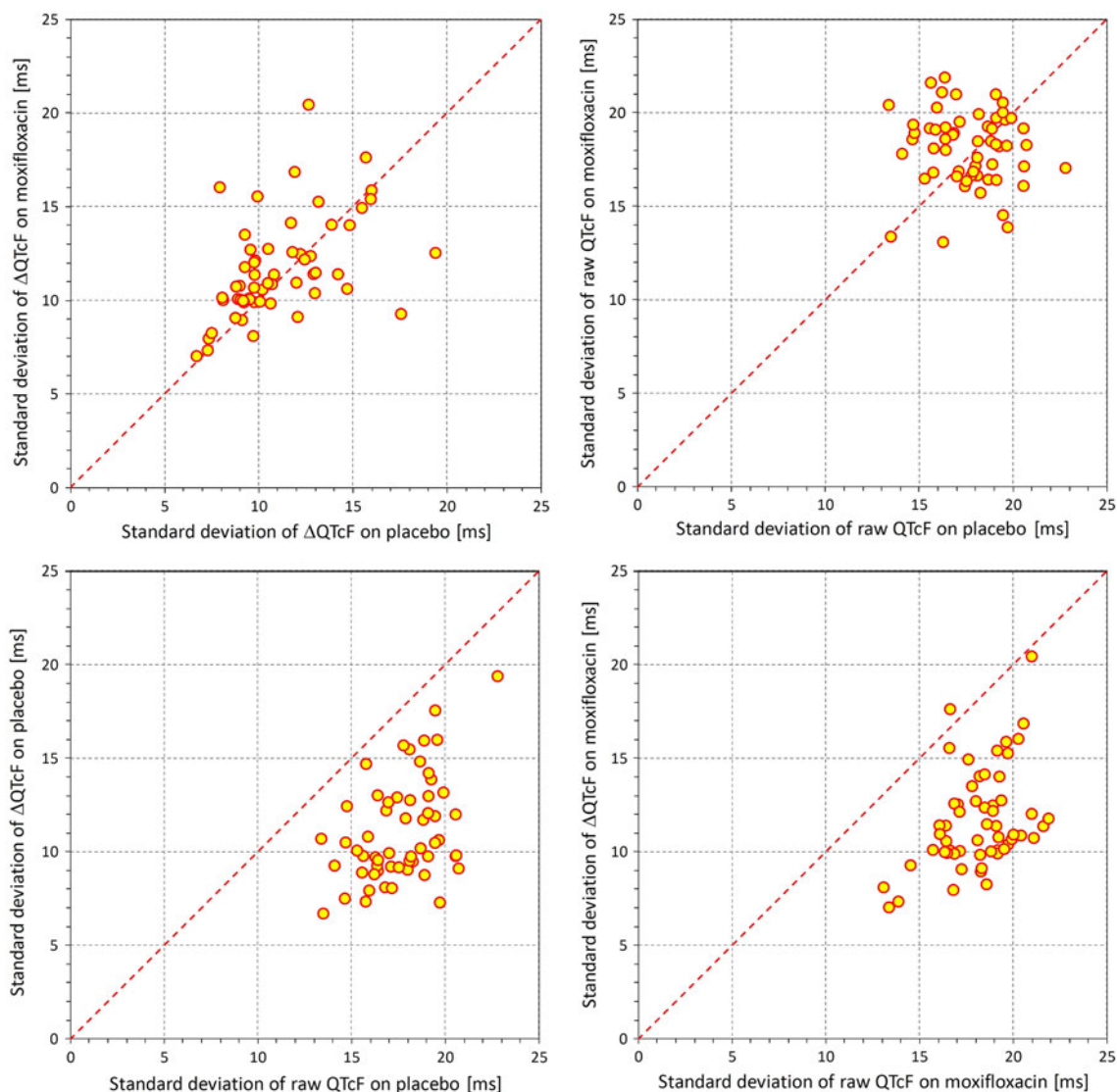
Fig. 8 The *top panel* shows scatter diagrams of the standard deviations of ΔQTcF (correction for time-matched baseline) and QTcF (post-dose) for moxifloxacin versus placebo, respectively, in individual studies. The *bottom panel* shows scatter plots of ΔQTcF (correction for time-matched baseline) and QTcF (post-dose) for placebo and moxifloxacin, respectively, in individual studies. In each plot, each *mark* corresponds to a separate study. Specifically, in each study, the standard deviations at each time point for the raw QTc and ΔQTcF were calculated. The averages of these standard deviations across all time points for the raw QTcF and ΔQTcF are shown

even the intra-individual diurnal pattern might change. Finally, the 57 TQT studies were not conducted at the same time, on the same site, or by the same investigators. Per-protocol time points might be different. The central laboratories that handled the ECGs also differed, meaning that the ECG patterns may have been interpreted differently. Differences in moxifloxacin treatment blinding might also have contributed some inconsistency. Moxifloxacin treatment was double-blinded in some studies and open-label in others. Similarly, there may have been some other inconsistencies between the analyzed studies.

Although there may have been deviations among different studies, intra-study consistency between both

baseline and on-treatment data is still expected. It is unlikely that discrepancies among the studies had a significant influence on our findings. The differences in baseline-correction approaches were likely independent of the measurement technologies used.

5 Conclusion

In spite of its limitations, this investigation allowed us to conclude that correcting for the baseline in parallel TQT studies leads to a more precise QTc estimate. Because of the possible inaccuracy introduced by time-averaged

baseline correction, time-matched baseline correction appears to be preferable for a parallel TQT study, in order to both reduce the intrinsic variability due to individual-specific circadian patterns and to obtain more accurate point estimates.

References

1. Ode E, Aizawa Y, Arai Y, Shibata A. Diurnal variation of QT interval in patients with VVI pacemaker. Tohoku J Exp Med. 1985;145:419–26.

2. Bexton RS, Vallin HO, Camm AJ. Diurnal variation of the QT interval—influence of the autonomic nervous system. Brit Heart J. 1986;55:253–8.

3. Koch HJ, Raschka C, Banzer W. Diurnal variation of ECG intervals and R or T amplitudes in healthy male subjects assessed by means of spectral and consignor analysis. Jpn Heart J. 1999;40:45–53.

4. Yeragani VK, Berger R, Pohl R, Balon R. Effect of age on diurnal changes of 24-hour QT interval variability. Pediatr Cardiol. 2005;26:39–66.

5. Center for Biologics Evaluation and Research. E14 Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs. Guidance to industry. Fed Regist. 2005;70:61134–5.

6. Meng Z, Quan H, Fan L, Kringle R, Sun G. Use of the average baseline versus the time-matched baseline in parallel group thorough QT/QTc studies. J Biopharmaceut Stat. 2010;20:665–82.

7. Sun GG, Quan H, Kringle R, Meng Z. Comparison of statistical models adjusting for baseline in the analysis of parallel-group thorough QT/QTc studies. J Biopharm Stat. 2012;22:438–62.

8. Fridericia LS. Die Systolendauer im Elekrokardiogramm bei normalen Menschen und bei Herzkranken. Acta Med Scand. 1920;53:469–86.

9. Garnett CE, Zhu H, Malik M, Fossa AA, Zhang J, Badilini F, Li J, Darpö B, Sager P, Rodriguez I. Methodologies to characterize the QT/corrected QT interval in the presence of drug-induced heart rate changes or other autonomic effects. Am Heart J. 2012;163:912–30.

10. Zhang J. Testing for positive control activity in a thorough QTc study. J Biopharmaceut Stat. 2008;18:517–28.

11. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;8476:307–10.

12. Hills M, Armitage P. The two-period cross-over clinical trial. Brit J Clin Pharmacol. 1979;8:7–20.

13. Malik M, Färbom P, Batchvarov V, Hnatkova K, Camm AJ. Relation between QT and RR intervals is highly individual among healthy subjects: implications for heart rate correction of the QT interval. Heart. 2002;87:220–8.

14. Batchvarov VN, Ghuran A, Smetana P, Hnatkova K, Harries M, Dilaveris P, Camm AJ, Malik M. QT-RR relationship in healthy subjects exhibits substantial intersubject variability and high intrasubject stability. Am J Physiol Heart Circ Physiol. 2002;282:H2356–63.

15. Molnar J, Zhang F, Weiss J, Ehlert FA, Rosenthal JE. Diurnal pattern of QTc interval: how long is prolonged? Possible relation to circadian triggers of cardiovascular events. J Am Coll Cardiol. 1996;27:76–83.

16. Malik M, Hnatkova K, Schmidt A, Smetana P. Accurately measured and properly heart-rate corrected QTc intervals show little daytime variability. Heart Rhythm. 2008;5:1424–31.

17. Malik M, Garnett C, Zhang J. Thorough QT studies—questions and quandaries. Drug Safety. 2010;33:1–14.

18. Malik M. Errors and misconceptions in ECG measurement used for the detection of drug induced QT interval prolongation. J Electrocardiol. 2004;37(Suppl):25–33.

19. Malik M, Zhang J, Johannesen L, Hnatkova K, Garnett C. Assessing electrocardiographic data quality and possible replacement of pharmacologic positive control in thorough QT/QTc studies by investigations of drug-free QTc stability. Heart Rhythm. 2011;8:1777–85.