Improving the Reproducibility of QT Dispersion Measures

Kaspar Lund, M.S.S.E.,* Juha S. Perkiömäki, M.D., Ph.D.,†
Christian Brohet, M.D., Ph.D.,‡ Mohammed Zaïdi, M.D.,‡
Hanne Elming, M.D., Ph.D., Christian Torp Pedersen, M.D., Ph.D.,#
Heikki V. Huikuri, M.D., D.M.S.CI.,†
Hans Nygaard, M.S.S.E., D.M.S.CI.,*
and Anders Kirstein Pedersen, M.D., D.M.S.CI.*

From the *Department of Cardiology, Skejby University Hospital, Aarhus N, Denmark;

†Department of Cardiology, Oulu University Hospital, Oulu, Finland; ‡Department of Cardiology, Saint-Luc Hospital Catholic University of Louvain, School of Medicine, Brussels; and

#Department of Cardiology, Gentofte Amtssygehus, Hellerup, Denmark

Background: The low reproducibility of the QT dispersion (QTD) method is a major reason why it is not used in clinics. The purpose of this study was to develop QT dispersion parameters with better reproducibility and identification of patients with a high risk of ventricular arrhythmia or death.

Methods and Results: Three institutions using different methods for measuring QT intervals provided QT databases, which included more than 3500 twelve-lead surface ECGs. The data represented low and high risk subjects from the following groups: the normal population EpiSet (survivors vs dead from cardiovascular causes), acute myocardial infarction patients AmiSet (survivors vs dead) and remote myocardial infarction patients ArrSet (with vs without a history of ventricular arrhythmia). The EpiSet, AmiSet, and the ArrSet contributed with N = 122, 0, and 110 ECGs for reproducibility analysis, and 3244, 446, and 100 ECGs for the analysis of prognostic accuracy. The prognostic accuracy was measured as the area under the Receiver Operator Curve. The QT intervals were divided into six QT pairs; the longest pair consisted of the longest and the shortest QT intervals etc. The QT dispersion trend (QTDT) was defined as the slope of the linear regression of the N longest QT pairs after estimation of missing QT intervals by interpolation of measured QT intervals. The QTMAD and the QTSTD methods were defined as twice the mean absolute deviation and the standard deviation of the N longest QT pairs. The reproducibility was improved by 27% and 19% in the EpiSet and the ArrSet relative to the reproducibility of QTD. The accuracy improved for the EpiSet and the ArrSet and was maintained for the AmiSet.

Conclusions: By using the three longest and the three shortest QT intervals in QTDT, QTMAD, or QTSTD, the reproducibility improved significantly while maintaining or improving the prognostic accuracy compared to QTD.

A.N.E. 2001;6(2):143–152

QT; QT dispersion; electrophysiology; sudden death; accuracy; precision

The QT dispersion is a measure of the heterogeneity of ventricular refractory periods, which is an important factor in the development of ventricular arrhythmia. The classical QT dispersion (QTD) is defined as the difference between the longest and the shortest QT interval in the 12-lead surface ECG

(ECG12). Prolonged QTD has been observed in individuals with a variety of heart diseases, including hypertrophic cardiomyopathy, heart failure, myocardial infarction, diabetes mellitus, and angina.^{1–6}

However, the clinical value of QTD is limited.^{7,8} Biphasic T waves, the gradual transition from

mono- to biphasic waves, and flat T waves contribute to varying "standards" for the rejection of leads and a low reproducibility of QTD. The standard deviations (SD) for intra- and interobserver reproducibility are about 12 ms and 16 ms, respectively.^{1,9,10} Random measurement errors of QT create a bias that increases QTD; in particular for the patients with the lowest physiological dispersion. That is, the bias reduces the separation of patients with low versus high dispersion.¹¹ Rejection of leads reduces QTD substantially so correction of missing QT intervals is a relevant subject.^{12,13} Day and coworkers proposed a correction for missing QT intervals, where QTD is divided by the square root of the number of measured QT intervals.¹⁴ No consensus has been reached for this method.^{12,15}

QTD is based on only two of the up to twelve available QT intervals. The SD of the QT intervals, proposed by Priori and coworkers,² uses all the QT intervals, but there is no evidence in the literature that it is systematically more reproducible and accurate. Nevertheless, our hypothesis is that using more than the two extreme QT intervals can improve the reproducibility and possibly the prognostic accuracy. We also believe that the benefit of such a method is to a large extent independent of the way the QT interval is measured.

The purpose of this study was to improve primarily the reproducibility and secondarily the prognostic accuracy by using more than the two extreme QT intervals to calculate the QT dispersion. The research was done in two stages. In the first stage, we developed methods to identify abnormal QT intervals and a new dispersion method, the QT Dispersion Trend (QTDT). QTDT enabled a very detailed analysis of the change in reproducibility and prognostic accuracy versus the gradual inclusion of more and more QT intervals in QTDT. The purpose of the second stage was to obtain the same performance as the QTDT method with a method that was so simple that doctors would use it. This resulted in the second parameter the Mean Absolute Deviation (QTMAD) and a modification of the SD of the QT intervals (QTSTD). We also evaluated the compensation for missing QT intervals proposed by Day and coworkers.¹⁴

The study was designed to have a high external validity by (1) including three different populations covering the spectrum from the normal population, to remote and acute myocardial infarction patients, and (2) using different methods for measuring the QT interval. (Table 1)

Table 1. Terms

Accuracy	Area under the ROC curve multiplied by 100
ROC	Receiver operator curve
QTDT	QT dispersion trend
QTMAD	Mean absolute deviation of the QT interval from the mean QT interval
QTSTD	Standard deviation of QT intervals

METHODS

Study Population

The study was based on existing databases, each one employing a different manual method for measuring QT. Subjects with less than three QT intervals in the precordial or the limb leads were excluded.

The arrhythmia database (ArrSet)¹⁶ included 110 subjects with a remote myocardial infarction in a case-control design with matching for age and sex. The leads were recorded at a paper speed of 50 mm/s at a gain of 10 mm/mV and a paper format of 2*6 leads. QT was measured in two successive beats as the intersection between the tangent of the terminal part of the T wave and the baseline. Subjects were included twice, once for each beat. Leads were rejected if the amplitude of the T wave was too low, if the baseline was too unstable or if the start of QRS was ill defined. Beats with too few QT intervals (N = 21) were excluded. Some subjects did therefore only contribute with one beat so there is not necessarily a one to two relation between subjects and beats in the following counts. The 199 beats accepted for this study came from a low-risk group including 98 beats (i.e., about 49 subjects) with no history of arrhythmia, and a high risk group including 101 beats (i.e., about 51 subjects) who had a recent history of ventricular tachycardia or ventricular fibrillation or had been resuscitated from cardiac arrest. The subjects age were mean + SD = 54 + 9 years. The reproducibility was analyzed by considering the measurement of QT in two successive beats as two repeated measurements.

The Acute Myocardial Infarction database (AmiSet)¹⁷ included 247 subjects that were hospitalized for acute myocardial infarction and followed for 14 + 6 months. Twenty-three subjects died from all causes of death in this period, 134 survived, and the status was unknown for 90 subjects. It is overly likely that the patients with unknown status survived because 9.3% (23 dead of

the 247 included subjects) is the expected mortality of an unselected population of acute myocardial infarction patients. We did therefore classify patients with unknown status as survivors. An ECG12 was taken early after hospitalization ($N = 201$), during the second week in the cardiac ward ($N = 175$), and at 4 months after MI ($N = 71$). Of the 447 ECG12s, $N = 37$ came from subjects who died, $N = 198$ from subjects who survived, and $N = 212$ from subjects with unknown status classified as survivors. QT was measured using a semiautomatic method applied to the averaged QRS-T complex of 10 seconds of ECG12 plus Frank's orthogonal XYZ leads. A startpoint for QRS and an endpoint for the T wave were determined automatically from Frank's orthogonal leads and displayed on a computer screen as common start- and endpoints for QT in all leads of ECG12. The operator then adjusted the endpoints visually, if necessary. All ECG12s for each subject were included so a subject could be included from one to three times. ECG12s with too few QT intervals ($N = 1$) were excluded. The 446 ECG12s accepted for this study came from the high risk group including ($N = 36$) and the low risk group ($N = 410$) with subjects who were alive or for whom the status was unknown at the end of the follow-up. The age was 64 ± 12 years. No data was available for reproducibility analyses.

The EpiSet database consisted of a random sample of 3781 individuals (data were collected by the Center of Preventive Medicine in Denmark) from a general Danish population recruited in 1982-1984 and followed for 11 years.⁹ The medical history was obtained by interview. The medical staff performed a physical examination and an ECG12 was recorded at 25 mm/s at a gain of 10 mV and a paper format of 3*4 leads. QT was measured in one beat using a digitizer table connected to a personal computer. The operator clicked with the digitizer on the endpoint of the T wave and could not control the measurement afterwards. The endpoint of the T wave was defined visually as the point of return to the isoelectric line. In the presence of a U wave, the end of the T wave was defined visually as the nadir between the two waves. Leads were rejected if the T wave was too flat or in case of a TU-merge without nadir. Subjects with bundle branch block ($N = 28$), atrial-fibrillation ($N = 10$), death from noncardiovascular causes ($N = 167$), missing or too low quality ECG12s ($N = 78$), or too few QT intervals ($N = 4$) were excluded. The 3244 subjects accepted for this study consisted of the high risk group ($N = 92$) with subjects that died from

cardiovascular causes during follow-up and the low risk group ($N = 3152$) with subjects who were alive at the end of the follow-up. The age was 44 ± 10 years. Reproducibility was analyzed for 122 subjects for whom two repeated measurements were available.

Accuracy

The lack of a true result, such as an intercardiac dispersion measure, makes it much more difficult to develop new dispersion measures. However, a clinical meaningful benchmark is the prognostic accuracy of identifying patients of high risk of cardiac death. A higher area under the Receiver Operator Curve (ROC) implies a more prognostic accurate method. The ROC shows how the tradeoff between sensitivity and specificity changes according to the classification threshold. The optimal threshold depends on many factors: purpose of the test, sensitivity, specificity, other available clinical tests, possible treatment, and economics.¹⁸ The advantage of the ROC area is that we avoid dealing with specific classification thresholds. We calculate the ROC area by observing that it is proportional to the standardized normal deviate of Mann-Whitney for a given number of patients.¹⁹ The ROC area was shown in percentage from 50% to 100% corresponding to no and total discriminative power, respectively.

We are only interested in the circumstance under which the ROC area improves. That is, it doesn't matter whether the ROC area changes from 50% to 55% or from 80% to 85%.

Overview of the Signal Processing

The flowchart in Figure 1 gives an overview of the signal processing, which is described in detail in the following paragraphs and illustrated in the appendix using an example.

Rejection of QT Outliers

The abnormal QT intervals are either the shortest or the longest QT intervals. Up to two erroneous QT intervals were detected by measuring whether the difference between the shortest and next shortest or the longest and the next longest QT interval exceeded the threshold in Equation 1. Erroneous QT intervals were rejected and subsequently treated as if they had not been measured.

$$th = m * \alpha + \beta$$

Equation 1: Calculation of the threshold for QT outliers. Symbols: th: the calculated threshold; m :

Figure 1. Flowchart of the signal processing. Notice that QTDT includes an extra step for estimation of missing QT intervals. QTMAAD and QTSTD, however, are only calculated from the available QT intervals. The flowchart does not show that the rejection of QT outliers is a little more complex for QTDT than for QTMAAD and QTSTD.

multiplier ; α : the slope of the linear regression line of the sorted QT intervals not including the shortest and the longest QT intervals. All values are weighted equally β : offset.

Estimation of Missing QT Intervals

First QT outliers were rejected. Then missing QT intervals were substituted by estimated values in order to prevent "uncontrolled" compensation for missing QT intervals in the QT Dispersion Trend method. Missing QT intervals were estimated as either the average of all the measured QT intervals or by linear interpolation after arranging QT according to the anatomical position of the leads. That is I, aVR, II, aVF, III, aVL, I, aVR, II, aVF, III, aVL for the frontal leads and V1, V2, V3, V4, V5, V6, V1, V2, V3, V4, V5, V6 for the precordial leads (see Fig. 1). The sign of the leads was omitted since it does not influence QT. This reorganization reveals a trend in the QT intervals with a period of six leads.²⁰ This trend is presumably the reason why linear interpolation is better than substitution by the mean of the measured QT intervals.

QT Pairs

The sorted QT intervals were treated in pairs (QT pairs) throughout the study. The first pair con-

sisted of the longest and the shortest QT intervals, the second pair the second longest and the second shortest QT intervals etc. The twelve QT intervals can be organized in a total of six QT pairs.

QT Dispersion Parameters

The longest QT dispersion (QTD1) is the range of the first QT pair, the second longest QT dispersion (QTD2) is the range of the second QT pair etc. Up to six QTD values can be calculated this way.

QT Dispersion Trend (QTDT)

QTDT was the slope of the linear regression line of the sorted QT intervals after estimation of missing QT intervals. The linear regression of the line minimized the classical chi-square merit function in Equation 2. Notice that the QT intervals were included in QT pairs starting with the longest pair.

$$\chi^2(a, b) = \sum_{i=1}^N w_i^2 (QT_i - a - bx_i)^2 + w_{13-i}^2 (QT_{13-i} - a - bx_{13-i})^2$$

and

$$w_i = w_{13-i} = 2 + \text{step} * (i - 1),$$

if $w_i < 0$, then set $w_i = 0$

Equation 2: Chi-square merit function for the linear regression. Symbols: a and b: offset and slope to be estimated, x_i is the order of the sorted QT intervals $x_1=1, x_2=2 \dots x_{12}=12$; QT_i are the sorted QT intervals; w_i : The QT intervals in a pair are weighted equally i.e. $w_1 == w_{12}, w_2 == w_{11}$ etc. . . - step: change in weight from a QT pair to the next I: $I=1 \dots N - N: N = 1 \dots 6$ number of included QT pairs.

Mean Absolute Deviation of QT Intervals

The QTMAD was calculated as twice the mean of the absolute deviation of the N longest and the N shortest QT intervals from their mean value after rejection of QT outliers. The factor twice makes it identical to the difference between the mean of the N longest and the mean of the N shortest QT intervals in the very frequent case where there is $2*N$ QT intervals available. The rejection of QT outliers is simplified as follows. Calculate the slope in Equation 1 as the difference between the second longest and the second shortest QT interval divided by $M - 3$ where M is the number of measured QT intervals.

Standard Deviation of QT Pairs

The rejection of QT outliers and missing QT intervals were treated as for QTMAD. QTSTD is the SD of the N longest QT, whereas the original formula is defined as the SD of all the QT intervals.² QTSTD was calculated from the available QT intervals when their number was less than $2*N$. We did not divide QTSTD with the mean of the QT intervals as Priori et al.,² because a patient with a long QT and a long QT dispersion might then obtain the same score as a patient with a shorter QT and a shorter QT dispersion. This is inconsistent since prolonged QT and prolonged QT dispersion are both associated with higher risk of arrhythmia.^{1,16,21} Furthermore the division with the mean QT interval introduces, indirectly, a correction for heart rate that may not be optimal.^{9,22}

Relative Reproducibility

The relative reproducibility was here defined as the SD of repeated measurements divided by the mean value of the measurements. The relative reproducibility enabled the comparison of methods that scale the dispersion differently (see the example in

the appendix). Even the adjustment of a method's parameters changed the scale of the dispersion.

Statistical Methods

Variances were compared by the two-sided F-test. Statistical significance was defined as $P < 0.05$.

RESULTS

Number of Different QT Intervals in a Beat

The median number of measured QT intervals was 11, 12 and 12 for ArrSet, EpiSet, and AMISet. Each ECG12 in ArrSet and EpiSet had on average only one identical shortest and one longest QT interval. For the AMISet, however, we found one shortest QT interval but five identical longest QT intervals per ECG12.

Rejection of QT Outliers

The parameters m and β in Equation 1 were identified by optimization as follows. The prognostic accuracy of QTD was calculated for all combinations of m and β in the range: $1 \leq m \leq 8$ and $40 \leq \beta \leq 90$. Different combinations result in different rejections of QT intervals and consequently different prognostic accuracies for the three databases. The optimal combination, $m = 5$ and $\beta = 60$, maximizes the prognostic accuracy of all the databases. For this combination of m and β , no QT intervals were rejected from ArrSet. The rejection of two QT intervals from the AMISet improved the accuracy from 75.7 to 75.8. The rejection of 49 QT intervals from the EpiSet improved the accuracy from 61.6 to 63.7.

Substitution of Missing QT Intervals and Number of Included QT Intervals

The number of QT pairs to be included in QTD was initially analyzed using Table 2. There was a big decrease in accuracy from the dispersion of the second QT pair (QTD2) to the third QT pair (QTD3) in the AMISet, whereas it remained at the same level from QTD1 to QTD5 in the ArrSet and from QTD1 to QTD4 in the EpiSet. The prognostic accuracy for the substitution of missing QT intervals with the mean of the measured QT intervals was almost as good as interpolation for the first three to four QT pairs. The inclusion of three to six QT pairs showed, however, that substitution by interpolation was superior. The relative reproducibility deteriorated rapidly from the third QT pair and on

Table 2. Accuracy and Reproducibility for the Dispersion of the QT Pairs

Accuracy for QTD of QT Pairs						
Estimation of missing QT intervals	QT Pair 1 (QTD1)	QT Pair 2 (QTD2)	QT Pair 3 (QTD3)	QT Pair 4 (QTD4)	QT Pair 5 (QTD5)	QT Pair 6 (QTD6)
AMISet (N = 446)						
Average of QT intervals	75.8	73.8	68.2	61.2	50.2	46.9
Linear interpolation	75.8	73.8	68.1	61.0	51.5	50.0
ArrSet (N = 199)						
Average of QT intervals	71.8	70.0	70.7	72.1	69.3	60.3
Linear interpolation	71.8	71.2	72.0	74.8	74.5	66.9
EpiSet (N = 3248)						
Average of QT intervals	63.7	61.0	63.7	62.0	56.0	52.9
Linear interpolation	63.7	61.3	63.6	62.1	57.0	53.0
Reproducibility						
ArrSet (N = 95)						
Mean QTD (ms)	82.7	60.9	44.5	29.5	16.9	6.2
Reproducibility SD (ms)	11.0	8.2	9.6	8.0	7.8	6.1
Relative reproducibility (%)	13.3	13.5	21.5	27.2	46.0	99.3
EpiSet (N = 122)						
Mean QTD (ms)	65.0	41.5	28.4	18.7	11.0	3.4
Reproducibility SD (ms)	21.5	14.6	10.1	8.8	6.2	4.1
Relative reproducibility (%)	33.1	35.3	35.4	47.01	56.88	119.4

QT outliers have been rejected. The numbers in bold indicate row-by-row QT-pairs with similar accuracy. The row-label indicates how the missing Q intervals were substituted. The reproducibility section of the table used substitution by interpolation. The numbers in bold indicate relative reproducibility at similar levels.

QT pair i = QT pair 1 is the longest QT pair; QT pair 2 is the second longest QT pair etc.; N = number of subjects.

in the ArrSet and from the fourth QT pair and on in the EpiSet. On this background, we included the first three QT pairs and reevaluated this choice in Table 3 after optimization of QTDT.

Optimizing the Reproducibility of the QTDT

Table 4 shows that inclusion of the first three QT pairs improved the reproducibility of QTDT significantly compared to QTD for both the ArrSet and the EpiSet. The reproducibility was not improved by inclusion of more than three QT pairs. The reproducibility of QTDT, QTMAD and QTSTD were practically the same.

Optimizing the Reproducibility and Accuracy of the QT Dispersion Trend

The reproducibility and accuracy for QTDT of the longest QT pair is seen in Figure 2 at step = -2.

Two QT pairs are included when step is between -2 and -1 and three QT pairs are included when step is above -1. The best reproducibility and accuracy for all databases are obtained at step ≥ -0.5 . The AMISet reaches maximal accuracy at step = -0.5 and its accuracy drops below its initial level at step = 0.25. We selected step = 0 as the best compromise between the different improvements of accuracy and reproducibility for the databases. Furthermore, when step = 0, then the three QT pairs are weighted equally, thus simplifying the linear regression.

Day's Correction for the Number of Measured QT Intervals

Day's correction for the number of measured QT intervals¹⁴ was applied to QTD of the longest QT pair (subheading in Table 4) and QTDT of the first three QT pairs (column "pair 1-3" in Table 4). The

Table 3. Accuracy of Dispersion Variables vs Number of Included QT Pairs

Method	Accuracy of QT Pairs					
	Pair 1	Pair 1-2	Pair 1-3	Pair 1-4	Pair 1-5	Pair 1-6
AMISet (N = 446, accuracy of QTD = 75.7)						
QTDT	75.8	76.9	76.4	75.7	75.2	75.1
QTMAD	75.7	76.8	75.6	74.3	72.6	71.4
QTSTD	75.7	76.8	76.4	75.7	74.8	74.1
ArrSet (N = 199, accuracy of QTD = 71.8)						
QTDT	71.8	72.4	72.8	73.3	73.7	73.8
QTMAD	71.8	72.2	72.8	73.1	73.2	73.4
QTSTD	71.8	72.1	72.7	73.0	73.1	72.9
EpiSet (N = 3248, accuracy of QTD = 61.6)						
QTDT	63.7	63.7	64.4	64.7	64.7	64.6
QTMAD	63.7	63.5	64.3	64.5	64.5	64.4
QTSTD	63.7	63.5	63.9	63.9	63.9	63.9

QTDT was calculated with step = 0 (see eq. 2) and no correction for the number of missing QT intervals. The number in bold indicates the maximum accuracy in each row. Erroneous Q intervals were removed for calculation of QTDT, QTMAD, and QTSTD but not for QTD.

Pair i to j = pair 1 include the first QT pair, pair 1-2 include the two longest QT pairs etc.; N = number of subjects.

correction reduced the accuracy of (QTD and QTDT) in the ArrSet from (71.8, 72.8) to (70.8, 72.1). The accuracy for QTD and QTDT in the AMISet increased from (75.7, 76.4) to (76.1, 76.6). The accuracy for QTD and QTDT in the EpiSet increased from (61.6, 64.4) to (62.9, 64.4). That is, the accuracy decreased substantially in one database and a minor increase was observed in two databases. We concluded that Day's correction should in general not be applied.

Putting It All Together

Table 3 confirms that inclusion of the first three QT pairs maintained or increased the accuracy. The optimal number of QT pairs was two for the AMISet, five to six for the ArrSet, and four to five for the EpiSet. QTMAD and QTSTD had nearly the same accuracy as QTDT when only three QT pairs were included.

Averaging QT measured in successive beats be-

Table 4. Comparison of the Reproducibility of QTD, QTDT, QTMAD, and QTSTD

	Relative Reproducibility (%)					
	QT Pair 1	QT Pair 1-2	QT Pair 1-3	QT Pair 1-4	QT Pair 1-5	QT Pair 1-6
ArrSet (N = 95, relative reproducibility for QTD = 13.3)						
QTDT	13.3	11.1	10.8*	10.4	10.3	10.3
QTMAD	13.3	10.6	10.9	10.6	10.5	10.6
QTSTD	13.3	10.7	10.5	10.1	10.0	9.9
EpiSet (N = 122, relative reproducibility for QTD = 39.5)						
QTDT	33.1	30.6	28.9**	28.4	28.3	28.3
QTMAD	32.8	30.2	29.0	28.6	29.0	29.4
QTSTD	32.8	30.7	29.9	29.6	29.6	29.7

QT outliers were rejected for the calculation of QTDT, QTMAD, and QTSTD but not for QTD. None of the variables were corrected for the number of missing Q intervals. All Q pairs are weighted equally in QTDT (step = 0) and missing Q intervals were estimated by linear interpolation.

* P = 0.02. The reproducibility increased 19% relative to the reproducibility of QTD.

** P = 0.0003. The reproducibility increased 27% relative to the reproducibility of QTD.

Symbols: - step: change in weight from a QT-pair to the next as defined in equation 2.

Figure 2. Relative reproducibility and accuracy of the QTDT method versus the weighing of the QT pairs. QTDT included up to the first three QT pairs. The horizontal lines indicate the accuracy for the inclusion of only the longest QT pair.

fore calculation of QTD also increases the reproducibility. QT was measured in two successive beats in the ArrSet. After averaging the QT intervals measured in successive beats for 95 patients, the accuracy of QTD was 73.0% versus 74.0%, 74.2% and 74.2% for QTDT, QTMAD and QTSTD, respectively.

DISCUSSION AND CONCLUSION

This study showed that using the three longest and the three shortest QT intervals in QTDT, QTMAD, and QTSTD improve the reproducibility significantly by 19%–27% relative to the classical QTD method. The prognostic accuracy of the new methods was maintained or improved compared with the accuracy of QTD. The external validity of the study is high because the dispersion methods improved the reproducibility for very different methods of measuring QT in widely different populations. The improved accuracy is probably due to a small reduction in the bias of QTD caused by measurement errors of QT¹¹ and probably also a reduction of the consequences of inconsistent evaluations of the T wave morphology.

The accuracy and reproducibility of QTDT, QTMAD, and QTSTD were approximately the same if only the three longest and shortest QT intervals were used. It is acceptable to exclude the subjects with less than six measurable QT intervals. In this case QTMAD is simply the mean of the three longest minus the mean of the three shortest QT intervals. No compensation for missing QT intervals or rejection of QT outliers is necessary. If QT outliers are present then the accuracy can be further improved by rejecting them as described in this study.

The results are modulated by important differences between the databases. The rejection of erroneous QT intervals ($N = 48$) from the EpiSet improves the accuracy from 61.6 (subheading "EpiSet . . ." in Table 3) to 63.7 (column: "pair 1" in Table 3) and the relative reproducibility from 40% to 33% (Table 4). The EpiSet was measured using a digitizer board with no visual feedback. The operator can therefore not see whether he accidentally moved the digitizer while digitizing a point. This is probably one of the reasons for the few very large measurement errors in this database (typically more than 100 ms in 49 of 3248 ECGs). The AMISet and ArrSet were measured with methods including visual feedback preventing such errors. The accuracy increases for the inclusion of one to five of the longest and shortest QT intervals for both the ArrSet and the EpiSet, whereas the accuracy only increases for the inclusion of one to two of the longest and shortest QT intervals in the AMISet. The latter had five identical longest QT intervals per ECG. It is conceivable that several identical longest QT interval does not contribute with more information than the inclusion of a single of these values. The presence of five identical longest QT intervals must be a feature, specific to the measurement method used for the AmiSet as follows. After averaging the QRS complexes the computer generated a QT interval that was common to all leads. Then the operator could adjust the QT intervals visually on the screen but was apparently satisfied with the computer-generated interval in five leads per ECG.

The accuracy (ROC area) of the QT dispersion varies from extremely high 76% to low 64% in the AMISet and the EpiSet. Though tempting, it is not possible to conclude anything about the relative merits of the QT measurement methods, because the populations are very different.

This study confirms like other studies,^{12,15} that day's correction for the number of measured QT

intervals should not be applied. This correction is, however, part of the standard deviation of the QT intervals proposed by Priori and coworkers, and could therefore be one reason why this measure does not consistently outperform the accuracy of QTD in the literature. The other reason is probably that the inclusion of more than the three longest and shortest QT intervals does not increase the accuracy consistently as shown in this study. When only the three longest and shortest QT intervals are used, as done in QTSTD, then Day's correction is the same for all ECG12s and therefore without any influence. The best way to handle missing QT intervals is probably by interpolation from the measured values (Table 2). The QTDT method uses the latter method and performed slightly better than the QTMAD and the QTSTD methods using no substitution.

The reproducibility can also be improved efficiently by averaging QT measured in successive beats before calculation of QTD. This will also reduce problems due to asynchronous recording of leads.¹¹ It will, however, not eliminate the problem that different operators have different opinions about whether a wave is a U wave or the second

phase of a biphasic T wave or whether the lead should be rejected. The inclusion of six QT intervals in QTDT, QTMAD, and QTSTD will reduce the impact of these inconsistencies because it is unlikely that the operators disagree systematically on the rejection of leads and the location of the endpoint of the T wave in all six leads. Indeed, analysis of the ArrSet showed that QTDT, QTMAD, and QTSTD were more accurate than QTD when applied to the averaged QT intervals. We believe for these reasons that the new methods will improve the interobserver reproducibility and the man to machine reproducibility efficiently.

The main limitation of this study is the general problem that the optimization of a method often has a low external validity. We have limited this risk by working with three databases including more than 3500 patients representing widely different populations. Nevertheless, more studies of existing databases are needed to validate the result of this study, to find the interobserver and the man to machine reproducibility and to analyze how the methods improve the risk stratification of patients.

APPENDIX

1: Example of the calculation of QTD, QTMAD and QTSTD

The table below includes the data for the example. All units are in milliseconds. The erroneous QT interval is shown in *italic*. The symbols have been introduced in the article.

	I	II	III	aVR	aVL	aVF	V1	V2	V3	V4	V5	V6
QT	451	540	454	451	424	418	466	442	421	445	454	400

2: Sort the data

QT values selected for the calculation of the dispersion are shown in **bold**.

Sorted QT	400	418	421	424	442	445	451	451	454	454	466	556
-----------	------------	------------	------------	-----	-----	-----	-----	-----	------------	------------	------------	-----

3: Simplified detection of erroneous QT intervals as described for the QTMAD method.

Slope: $\alpha = (\text{second longest} - \text{second shortest QT interval}) / (N-3)$: $(466-418)/(12-3) = 5.33$

Outlier threshold: $th = (m \cdot \alpha + \beta) = 5 \cdot 5.33 + 60 = 86.6$

Second shortest minus shortest QT interval: $418-400 = 18 < 86.6 = > \text{QT is okay}$

Longest minus second longest QT interval: $556-466 = 90 > 86.6 = > \text{erroneous QT interval}$

4: Calculation of the dispersion after rejection of erroneous QT intervals:

QTD = $466-400 = 66 \text{ ms}$

QTMAD* = $(400 + 418 + 421)/3 - (454 + 454 + 466)/3 = 45 \text{ ms}$

QTSTD = standard deviation of: 400, 418, 421, 454, 454, 466 = 26 ms

* When there are six or more available QT intervals, then QTMAD can be simplified to the mean of the three longest minus the mean of the three shortest QT intervals.

5: Remarks

Notice that the weight of the extreme QT intervals is reduced gradually from QTD to QTSTD to QTMAD. This can be seen from the relation: $QTD \geq 2 \cdot QTSTD \geq QTMAD$. When the shortest QT intervals are identical and the longest QT intervals are identical then $QTD = 2 \cdot QTSTD = QTMAD$.

REFERENCES

- Perkiomaki JS, Koistinen MJ, Yli-Mayry S, et al. Dispersion of QT interval in patients with and without susceptibility to ventricular tachyarrhythmias after previous myocardial infarction. *J Am Coll Cardiol* 1995;26:174-179.
- Priori SG, Napolitano C, Diehl L, et al. Dispersion of the QT interval. A marker of therapeutic efficacy in the idiopathic long QT syndrome. *Circulation* 1994;89:1681-1689.
- Pye M, Quinn AC, Cobbe SM. QT interval dispersion: A noninvasive marker of susceptibility to arrhythmia in patients with sustained ventricular arrhythmias? *Br Heart J* 1994;71:511-514.
- Buja G, Miorelli M, Turrini P, et al. Comparison of QT dispersion in hypertrophic cardiomyopathy between patients with and without ventricular arrhythmias and sudden death. *Am J Cardiol* 1993;72:973-976.
- Wei K, Dorian P, Newman D, et al. Association between QT dispersion and autonomic dysfunction in patients with diabetes mellitus. *J Am Coll Cardiol* 1995;26:859-863.
- Higham PD, Furniss SS, Campbell RW. QT dispersion and components of the QT interval in ischemia and infarction. *Br Heart J* 1996;73:32-36.
- Kautzner J, Malik M. QT interval dispersion and its clinical utility. *Pacing Clin Electrophysiol* 1997;20:2625-2640.
- Statters DJ, Malik M, Ward DE, et al. QT dispersion: Problems of methodology and clinical significance. *J Cardiovasc Electrophysiol* 1994;5:672-685.
- Elming H, Holm E, Jun L, et al. The prognostic value of the QT interval and QT interval dispersion in all-cause and cardiac mortality and morbidity in a population of Danish citizens [see comments]. *Eur Heart J* 1998;19:1391-1400.
- Tran HT, Fan C, Tu QW, et al. QT measurement: A comparison of three simple methods. *Ann Noninvas Electrocardiol* 1998;3:228-231.
- Lund K, Arildsen H, Perkiomaki JS, et al. Bias of QT dispersion. *Ann Noninvas Electrocardiol* 2001;5:38-42.
- Hnatkova K, Malik M, Kautzner J, et al. Adjustment of QT dispersion assessed from 12 lead electrocardiograms for different numbers of analyzed electrocardiographic leads: Comparison of stability of different methods. *Br Heart J* 1995;72:390-396.
- Murray A, McLaughlin NB, Campbell RW. Measuring QT dispersion: Man versus machine. *Heart* 1997;77:539-542.
- Day CP, McComb JM, Matthews J, et al. Reduction in QT dispersion by sotalol following myocardial infarction. *Eur Heart J* 1991;12:423-427.
- Glancy JM, Garratt CJ, Woods KL, et al. Use of lead adjustment formulas for QT dispersion after myocardial infarction. *Br Heart J* 1997;74:676-679.
- Perkiomaki JS, Huikuri HV, Koistinen JM, et al. Heart rate variability and dispersion of QT interval in patients with vulnerability to ventricular tachycardia and ventricular fibrillation after previous myocardial infarction. *J Am Coll Cardiol* 1999;30:1331-1338.
- Brohet C, Zaidi M, Robert A, et al. Risk stratification after myocardial infarction: Pre-eminence of the QT Dispersion. 1998. XIII World Congress of Cardiology - Rio de Janeiro, Brazil, Monduzzi Editore - Bologna (Italy).
- Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561-577.
- Henderson AR. Assessing test accuracy and its clinical consequences: A primer for receiver operating characteristic curve analysis. *Ann Clin Biochem* 30:521-539.
- Lund K, Lund B, Arildsen H, et al. Spatial properties of QT and the T wave morphology. *Ann Noninvas Electrocardiol* 2000;5:214-221.
- Day CP, McComb JM, Campbell RW. QT dispersion: An indication of arrhythmia risk in patients with long QT intervals. *Br Heart J* 1990;63:342-344.
- Zabel M, Woosley RL, Franz MR. Is dispersion of ventricular repolarization rate dependent? *Pacing Clin Electrophysiol* 1997;20:2405-2411.