The PhysioNet/Computers in Cardiology Challenge 2006: QT Interval Measurement

George B. Moody, Hans Koch, Uwe Steinhoff
MIT, Cambridge, MA USA and PTB, Berlin, Germany

Can the QT interval be measured by fully automated methods with an accuracy acceptable for clinical evaluations?

The EU's European Medicines Agency, the US FDA, Japan's National Institutes of Health Services, and their counterparts in other nations, have adopted common guidelines (ICH E14) that require "throrough QT/QTc studies" for all new drugs. These guidelines require manual measurement of QT intervals, but leave open the possibility that automated methods might be accepted in the future:

> *"If well-characterized data validating the use of fully-automated technologies become available, the recommendations in the guidance for measurement of ECG intervals could be modified."*

A major motivation of this Challenge is to provide such data.

Methods

Participants in the Challenge developed and evaluated methods for measuring the QT interval, using the 549 records of the PTB Diagnostic ECG Database, which is freely available at

http://physionet.org/physiobank/database/ptbdb/

For each record, entrants selected a typical (representative) beat and measured the times of the beginning (PQ) and end (T-end) of the QT interval for that beat in lead II. (Participants were allowed to make use of any of the other 14 leads in order to obtain estimates of the QT interval in lead II.)

The challenge was organized in three divisions:

Division 1 (manual methods acceptable under ICH E14):

Fifteen entrants produced sets of manually reviewed QT measurements, and the record-by-record medians of these defined the "gold standard" reference. Division 1 participants were required to measure at least half of the records; most measured 95% or more.

Divisions 2 and 3 (fully automated methods):

Twenty-five entrants submitted sets of automatically-derived measurements. Division 2 and 3 participants were required to measure at least 95% of the records; the selection of which records to omit, if any, had to be made algorithmically. Participants in **division 3** (the open source division) also submitted source code for their methods, to be posted on PhysioNet following the conclusion of the Challenge.

Awards were offered for the best results in each division, and for the best results overall.

PhysioNet/Computers in Cardiology Challenges are conducted using the facilities of PhysioNet, a public service of the Research Resource for Complex Physiologic Signals, which is supported by a grant from the National Center for Research Resources of the US National Institutes of Health (P41 RR13622). The Challenge awards are funded by Computers in Cardiology.

Visit http://physionet.org/challenge/2006 for access to the QT measurements, the sources for the division 3 entries, and additional information about the Challenge.

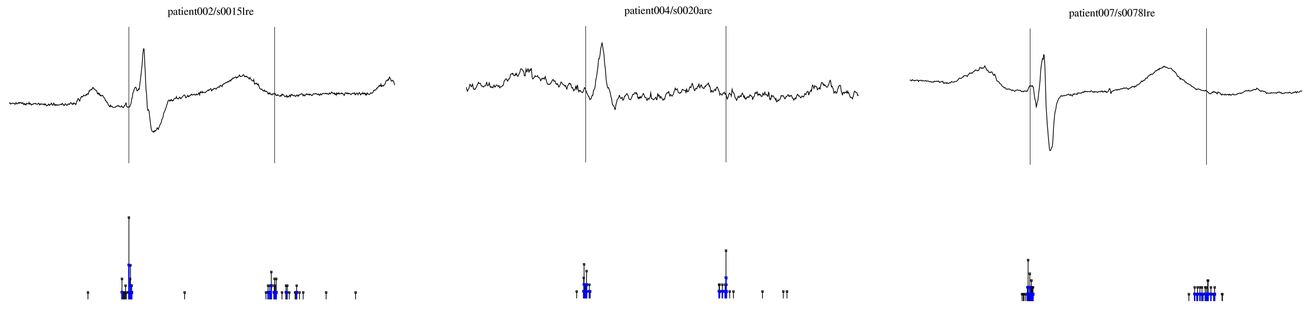
Results

Entries were scored by comparison with the "gold standard" median of manually-reviewed measurements. The score used for ranking was the RMS difference from the "gold standard", divided by the fraction of records measured. In most cases, the fraction of records measured was between 0.95 and 1, so this factor had little effect on the ranking.

	Best score	Typical scores
Division 1	**6.67 ms**	**10-20 ms**
Division 2	**16.34 ms**	**20-30 ms**
Division 3	**17.33 ms**	**20-30 ms**
Meta-6	**10.93 ms**	

Distributions of PQ and T-end measurements (examples):

Shown are three of the 549 records in the Challenge data set. At the top of each plot is the waveform chosen most often as "typical", with the medians of the manually-reviewed PQ and T-end times overlaid. The lower sections of each plot show the distribution of measurements of that waveform across all entries (in black); the manually-reviewed measurements are overlaid in blue.



Distributions of QT measurement errors (examples):

Shown at left below are QT measurement errors for six entries in divisions 2 and 3, plotted as functions of the "gold standard" QT measurements.

The "Meta-6" algorithm exploits the variations in these error distributions by using the median of the measurements in these six entries to derive its QT estimate (error distribution at right). Meta-6 acheived a score of 10.93 ms, better than all but four of the manual entries, and significantly better than any of the original six entries upon which it is based.