

The Cardiac Safety Research Consortium electrocardiogram warehouse: Thorough QT database specifications and principles of use for algorithm development and testing

Paul Kligfield, MD,^a Cynthia L. Green, PhD,^b Justin Mortara, PhD,^c Philip Sager, MD,^d
Norman Stockbridge, MD, PhD,^e Michael Li, MS,^e Joanne Zhang, PhD,^e Samuel George, MD,^f
Ignacio Rodriguez, MD,^g Daniel Bloomfield, MD,^h and Mitchell W. Krucoff, MD^b *New York, NY; Durham, NC;
Milwaukee, WI; San Francisco and Santa Clara, CA; Silver Spring, MD; Rahway and Nutley, NJ*

This document examines the formation, structure, and principles guiding the use of electrocardiogram (ECG) data sets obtained during thorough QT studies that have been derived from the ECG Warehouse of the Cardiac Safety Research Consortium (CSRC). These principles are designed to preserve the fairness and public interest of access to these data, commensurate with the mission of the CSRC. The data sets comprise anonymized XML formatted digitized ECGs and descriptive variables from placebo and positive control arms of individual studies previously submitted on a proprietary basis to the US Food and Drug Administration by pharmaceutical sponsors. Sponsors permit the release of these studies into the public domain through the CSRC on behalf of the Food and Drug Administration's Critical Path Initiative and public health interest. For algorithm research protocols submitted to and approved by CSRC, unblinded "training" ECG data sets are provided for algorithm development and for initial evaluation, whereas separate blinded "testing" data sets are used for formal algorithm evaluation in cooperation with the CSRC according to methods detailed in this document. (Am Heart J 2010;160:1023-8.)

This document details the derivation and structure of CSRC ECG data sets from thorough QT (TQT) Studies originating in the Food and Drug Administration (FDA) ECG Warehouse and describes the principles guiding the use of these data sets by the Cardiac Safety Research Consortium (CSRC). These principles are designed to ensure the scientific credibility of new algorithm assessment testing and preserve transparently the fairness and public interest of access to these data, commensurate with the mission of the CSRC.^{1,2}

Background of the CSRC ECG warehouse

The CSRC is a public-private partnership, developed within the Critical Path Initiative of the US FDA through a memo of understanding between the FDA and the Duke

Clinical Research Institute.^{1,3} CSRC brings together stakeholder constituencies from the FDA, academia, and industry to focus on cardiac safety issues pertinent to new drug and device development. One of the original goals of the CSRC was the development and scientific use of a large, centralized ECG repository (the ECG Warehouse, developed by FDA with hardware and software support from Mortara Instrument, Milwaukee, WI). Data in the ECG warehouse include proprietary digital ECG recordings from TQT and related studies conducted in the course of drug development and post-market surveillance by industry sponsors. The FDA's ECG Warehouse currently holds more than 4,000,000 ECGs that can be linked with descriptive study variables such as age, sex, and drug treatment group.

The ECG Warehouse data are owned by the entities, generally pharmaceutical company manufacturers, that submit these data to the FDA. Release of the ECG data to the CSRC for public domain access by investigators, equipment manufacturers and algorithm developers was obtained with permission from the data owners as a collaborative effort of scientific goodwill on behalf of the public health. In conjunction with the release of the ECG data sets to the CSRC, a Scientific Oversight Committee (SOC) within CSRC developed processes supporting the submission and evaluation of proposals for use of the released ECG data, also intended to foster collaboration

From the ^aDivision of Cardiology, Department of Medicine, Weill Cornell Medical College, New York, NY, ^bDuke University Medical Center and Duke Clinical Research Institute, Durham, NC, ^cMortara Instrument, Milwaukee, WI, ^dSager Consulting Partners, San Francisco, CA, ^eCenter for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, ^fNewCardio, Inc., Santa Clara, CA, ^gMerck Research Laboratories, Merck and Co, Inc, Rahway, NJ, and ^hHoffmann-LaRoche, Inc, Nutley, NJ. Submitted September 3, 2010; accepted September 3, 2010.

Reprint requests: Paul Kligfield, MD, Weill Cornell Medical Center, 525 East 68th Street, New York, NY 10065.

E-mail: pkligfi@med.cornell.edu

0002-8703/\$ - see front matter

© 2010, Mosby, Inc. All rights reserved.

doi:10.1016/j.ahj.2010.09.002

within the scientific research community and across stakeholders. The initial focus of these protocols was to foster efforts to develop better surrogate markers for evaluation of electrophysiological safety, in particular algorithms that could more accurately, efficiently, or automatically detect QT prolongation signals from an active compound (such as moxifloxacin) compared to placebo exposure. To maintain an equitable and scientifically responsible governance of the ECG Warehouse, individual research protocols cleared by the SOC are approved by the CSRC Executive Committee before ECG data are released.

ECG data sets from thorough QT studies

Most of the data in the ECG Warehouse are waveforms collected from TQT studies performed in the early clinical course of new drug evaluation in healthy volunteers.⁴⁻⁷ Currently available CSRC ECG data sets are comprised of anonymized digital XML waveforms and key descriptive data from complete baseline, placebo and moxifloxacin periods of representative TQT studies. These data do not include information from any sponsor's proprietary drug.

The CSRC intends that these anonymized ECG data sets will be used for research that includes ECG algorithm testing and new ECG biomarker development for risk assessment, using changes associated with moxifloxacin (or other positive controls) compared to placebo as the surrogate marker. It is recognized that the effects of positive control drugs on the ECG may differ from other markers of arrhythmogenicity or other risk.⁸⁻¹⁰ In the future, other data sets, including populations with diseases under study and with alternative drug effects, are expected to be made available for evaluation, using access and governance processes similar to those described here for the TQT data sets.

A CSRC Data Management Committee (DMC) was established to develop standards for quality control and for performance scoring and publication of data derived from the ECG Warehouse. Like all CSRC committees, the DMC includes representatives from academia, industry and the FDA. For quality control, the DMC was mandated to assemble the ECG data sets, to verify that the digitized ECG waveforms can be used for research, and to match the clinical data provided with waveforms using the ECG Warehouse unique identifiers. The CSRC DMC also establishes the integrity of each individual ECG data set with vetting methods that are described below. The statistical basis of performance evaluation and guidelines for dissemination of results to the public through peer review reporting are also described below.

To protect patients' rights and sponsors' proprietary rights, only deidentified, fully anonymized data are provided to investigators for CSRC SOC approved

projects. Subject identifiers are stripped from each data set as are true dates, sponsor identification, site, ECG core laboratory, and ECG device information from the original FDA submission. Unique identifiers (UIDs) are assigned to each ECG in the warehouse. The technique used to generate the UIDs does not allow determination of how the UIDs were generated but does provide a link between the annotated ECG waveform in the ECG Warehouse and available descriptive data.

To date, data from eight TQT studies have been released by their sponsors to the CSRC for assembly and vetting; two learning and one testing data set are available for use from CSRC as of August 2010. Datasets range in size from 2,425 to 14,128 digitized ECGs per study, in populations that range in size from 31 to 181 subjects. For larger studies the CSRC may provide, along with the full data sets, randomized smaller subsets of data comprised of 20 to 50 subjects based on study population size and study design in each of the placebo and positive control study arms of the learning data sets; these subsets will reflect the sex and age distribution of the original population. These subsets will allow more robust testing of measurement algorithms using simulated populations with smaller numbers of subjects as an additional estimate of algorithm performance over a range of population variations, as well as providing a modality for enhanced "head to head" comparison of various algorithm methods.

Dataset formation, integrity, and vetting

The ECG Warehouse provides support for the evaluation and testing of ECG algorithms that relate to cardiac safety, using placebo and positive control digitized ECG waveforms obtained in the course of TQT studies that have been released to the CSRC. Structurally the CSRC will maintain a distinct partition between publicly released ECG waveforms and descriptor data ("unblinded" or "training" data sets) and waveforms released without descriptor data ("blinded" or "testing" data sets). This allows investigators of CSRC SOC approved proposals to pursue further algorithm development and refinement within specified training data sets, while undergoing more scientifically rigorous and valid performance testing in the unique testing data sets.

When the data are used to develop a new algorithm or to modify an existing algorithm, development will be performed using unblinded "training data sets." All reporting of algorithm performance derived from the unblinded data will be clearly identified as unblinded training results. When an algorithm is already developed, it may be evaluated in a separate blinded "testing data set" process for performance validation. Algorithm performance will only be considered definitive when reports are based on independent assessments from the blinded testing data sets provided by the CSRC.

Before any public release for SOC approved proposals, the CSRC DMC provides a structure for vetting the data quality. Once a sponsor has agreed to the release study information by FDA to the CSRC, the FDA links the ECG waveforms to the designated clinical data, all of which are electronically transferred to the CSRC. The CSRC DMC is responsible for assembling the data sets, which are then passed to the Duke Clinical Research Institute for uniquely identified secure, password-protected storage. The assembled data from each released study are then validated according to the following processes:

- ECG waveform checking:
 1. Random XML waveforms are reviewed by a physician to confirm realistic signal analyzability
 2. The number of XML waveforms are verified to match the number of ECGs in the sponsor-submitted clinical data set
 3. Waveform UIDs are verified to match the original clinical data set UIDs
- Clinical data set checking:
 1. File size and number of records are confirmed
 2. Missing data are identified and verified as missing
 3. Frequencies and ranges of selected variables are reviewed
 4. Anonymity of data is confirmed
- Waveform and clinical cross data checking:
 1. The Fridericia-corrected QT interval (QTcF) from sponsor-submitted moxifloxacin or other positive control measurements are quantitatively compared with measurements from a fully automated commercial algorithm using the double-delta (dd) approach. The placebo- and baseline-adjusted QTcF (ddQTcF) are compared for each analysis method at each collection time point.

The purpose of the comparison between sponsor-submitted data and the measurements from a fully automated algorithm is to confirm similar directional and timing trends of the expected positive effect as a quality control measure for assembly of the data set. Neither the sponsor measurements nor those of the automated algorithm used for this confirmation are considered to be gold standards for QT measurements per se.

Once vetted, the CSRC DMC will distribute the assembled data sets as either training or testing data sets. The training data sets are comprised of the XML digital ECG waveforms from the ECG Warehouse along with key descriptive sponsor data that can be merged and matched according to the UIDs. The descriptive data formats are archived in both SAS and Excel formats. The anonymized sponsor data include variables such as treatment (baseline, placebo or positive control), sex, age, race, ECG date, study design, study day, nominal and actual ECG time, replicate number, sponsor-reported ECG parameters (QT, RR, PR, QRS, etc), and other available common elements. In

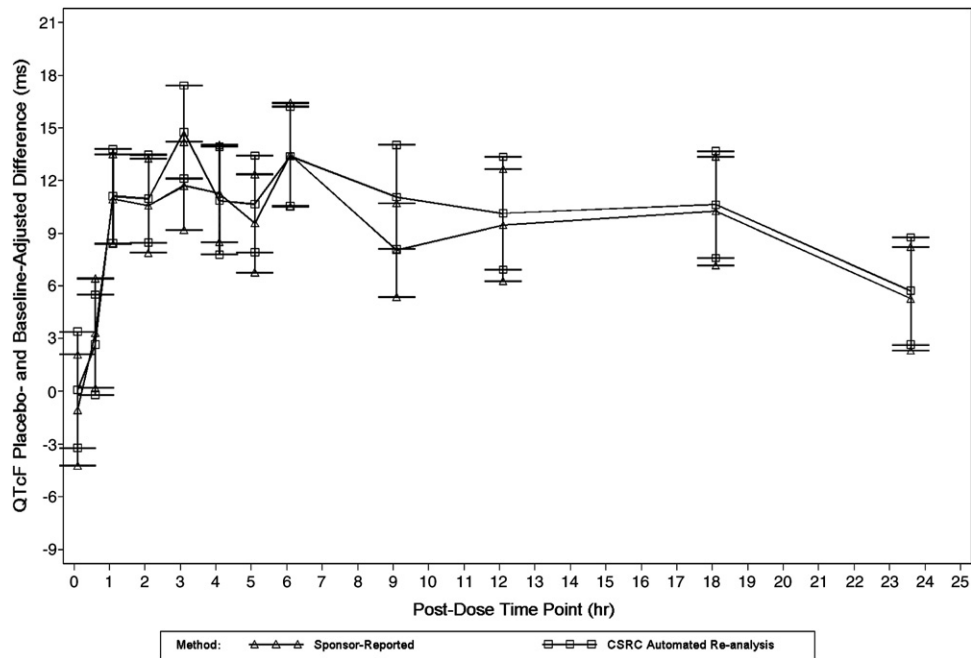
addition to removing all study and subject identifiers, the ECG dates are further de-identified by redefining all studies to start on January 1, 1960. The testing data sets supplied to public investigators will include only the XML digital ECG waveforms, which can be matched with the original descriptor data for analysis of algorithm performance only by the CSRC DMC. Investigators will submit their algorithm measurements to the CSRC DMC to assess the “blinded” algorithm performance in the testing data sets.

Algorithm development and testing: further detail on training data sets

For algorithm development, the outcome of the examination of the ECG warehouse training data sets will be new or more refined criteria. These training data sets are provided to public investigators completely “open label,” including waveforms descriptors and specification of the treatment group (baseline, placebo, moxifloxacin). It is expected that if unblinded data sets are made available to public investigators by the CSRC, the investigators will agree to submit the derived algorithm measurements to the CSRC DMC using the independent blinded “algorithm testing” process described below, which will then represent the true “tested performance” of the algorithm for public reporting. As a key corollary condition of using data from the ECG Warehouse, in any presentation or publication relating to performance of the derived algorithm from the unblinded data sets, public investigators must agree to define clearly the training nature of the study.

Another critical aspect relating to optimal use of the ECG Warehouse is the recognition that the standard made accessible through these data is that of an ECG-based assay sensitivity for moxifloxacin or other positive QT signal detection. This is not necessarily a gold standard for the actual ECG interval measurements. Thus, while core laboratory measurements of QT and other intervals from the original FDA submission are made available to public investigators as part of the open label training data sets, direct comparisons of QT or other interval measurements between core laboratory measurements and a new algorithm under testing are scientifically unjustified, since there is no gold standard for individual waveform measurements per se. An example of sponsor-submitted TQT data as reanalyzed by the CSRC from one of the “learning” data sets by an automated algorithm for global measurement of the QTcF is shown in [Figure 1](#). Note that while neither set of measurements is proposed to be a “gold standard” for the underlying measurements, both methods identify similar time and magnitude related changes in the double-delta findings for placebo- and baseline-adjusted QTcF response to moxifloxacin. This same principle is incorporated into the blinded “test set” reporting, as detailed below.

It is hoped and encouraged that investigators will publish the details of the algorithm and results of the

Figure 1

Proof of concept: ddQTcF response to moxifloxacin in CSRC learning data set study 4, by sponsor-submitted measurements and by CSRC re-analysis of ECG Warehouse waveforms using a fully automated algorithm for global QT measurement. The new measurements are similar in direction and timing with the results of the initial sponsor study.

training process in such a way that the methodology can be understood or learned from by others. The training ECG data sets may also be used by public investigators who have already derived established algorithms, but in this situation, it will still be agreed that testing of algorithm performance solely in an unblinded series will not be considered or publicly represented as definitive.

Algorithm performance in blinded testing data sets

Prior to formal algorithm testing, the algorithm and test criteria ordinarily will have been established using one or more learning data sets, or separately in previous studies by its developer. One or more testing data sets will be released to the investigator after SOC approval of the protocol and disclosure of the algorithm criteria with the understanding that the results of the analysis will be made available to the public, either through publication in the peer review literature, through tabulation on the CSRC Web site, or both. The investigator may choose whether or not to reveal the specific methodology of their algorithm in their publication.

To test either established or novel algorithms for QT or related interval measurements, public investigators will be supplied only ECG waveforms remaining blinded to all

descriptors including time of ECG and drug treatment (placebo or positive control). ECG findings of the tested algorithm will be processed and tabulated by the algorithm developer and returned to the CSRC for unblinding, including compilation of the investigators waveform measurements with drug treatment group, time sequence and relevant descriptors. Once compiled, analysis of “assay sensitivity” for detection of QT prolongation associated with positive control exposure is conducted by the CSRC statistician according to standard double-delta methodology. Analysis up to this point thus determines whether the public investigator’s algorithm performance meets assay sensitivity standards at least to the level of ICH E14 guidelines.⁴

Comparison of assay sensitivity between the original FDA data submission (generally from a core laboratory methodology) and the novel algorithm developed by the public investigator will be defined exclusively using the blinded testing database outcome. Such comparison will be performed by the CSRC statistician, both within the total TQT data set and also within smaller randomized subsets or “simulation” populations within each TQT study.

With respect to TQT, primary measurements of algorithm performance will include baseline-adjusted delta QT (dQT), delta RR (dRR), delta QTcF (dQTcF) and baseline- and placebo-adjusted double-delta QTcF

(ddQTcF). Analyses will include directional and timing trends, as well as the mean and standard deviations of the dQTcF and ddQTcF differences between the FDA-submitted (core laboratory) data and measurements generated by the tested algorithm of the investigator. Non-QT interval data extracted by developers from the testing data sets may also be analyzed by CSRC in a similar manner. The formal statistical comparisons of FDA-submitted and public investigator algorithm-generated blinded measurements will include:

- Bland-Altman plots of dQT/dRR/dQTcF, including mean difference and SD of differences
- 90% CI for dQTcF by treatment group at each time point, using raw means and standard deviations plotted over time
- 90% CI for ddQTcF at each treatment time point, using raw means and standard deviations plotted over time
- dQTcF variability, including the within-subject and between-subject variances obtained from a mixed repeated measures statistical model adjusted for no less than the baseline QTcF, time, treatment, and time by treatment interaction
- Within-subject average SD of QTcF by treatment group and combined based on replicates at each time point over the whole time course for each subject. A paired *t* test will be used to test the null hypothesis that the mean algorithm SD of QTcF is the same as the average core lab SD of QTcF.
- Theoretical statistical power and sample size using bootstrap simulations. Random selections with replacement over a range of sample sizes will be used to observe the average statistics within smaller populations, such as average time of the maximum moxifloxacin effect, average maximum effect, average within-subject and between-subject variances. The number of times that assay sensitivity is successfully detected across all bootstrapping procedures for a given sample size (power) will also be evaluated. The power curves will show the minimum theoretical number of subjects that would be required to achieve 80% power for each analysis method.

Rules of engagement

As a public-private partnership, the CSRC operates primarily along the lines of collaborative, transparent programs across stakeholders, in good faith and on behalf of the public health. Along those lines, the CSRC and its DMC committee are charged to manage the ECG Warehouse fairly, responsibly and in a manner that supports good science to advance the accuracy and efficiency of QT measurement for cardiac safety reporting in drug development. Thus, on behalf of

SOC approved research projects using the ECG Warehouse, the “rules of engagement” document supplied to all public investigators specifies that the distinctly different training and testing processes and results will be clearly distinguished in any presentation of algorithm performance. Specifically, the SOC intends that criteria developed by adapting algorithm test criteria to the training database not be considered to represent “tested performance” of the final algorithm, and that algorithms developed from the training database be rigorously and prospectively evaluated in the independent, blinded testing database. Standardized use of the blinded testing data sets will allow meaningful comparisons of assay sensitivity and other key elements of test performance between unique algorithms.

To support the principles above, the CSRC will maintain a partition between publicly released ECG waveforms with complete descriptor data (“unblinded training” data sets) and waveforms released without descriptor data (“blinded testing” data sets). Copying or supplying these data for other manufacturers or projects not previously approved by the CSRC is strictly prohibited. All reporting of findings using unblinded data from the training data sets will be clearly identified by investigators as unblinded training results. Investigators will agree only to define and report final, validated algorithm performance based on results from blinded data analyses processed from measurements returned to the CSRC for formal performance evaluation.

Access to the ECG warehouse

Proposals for use of data from the ECG Warehouse can be submitted to the SOC using the proposal format on the CSRC website (www.cardiac-safety.org). ECG waveforms from the warehouse will be made available for public investigators of SOC approved projects. For both unblinded training and blinded testing TQT data sets, waveform release has been designed so that algorithm developers can execute their own algorithms without the necessity to release proprietary measurement strategies to the CSRC or any other outside agency. For blinded data set performance evaluation (scoring) of endpoints, such as moxifloxacin signal detection, algorithm developer's measurements from blinded data ECG waveforms will be submitted to the CSRC statistician where they will be matched to descriptors including treatment assignments (placebo vs moxifloxacin), and analysis results will be returned to sponsors. Statistical analysis of both assay sensitivity and for comparisons between measurements submitted to FDA by the original sponsor and measurements provided by the SOC-approved investigator will be reported as detailed above. A nominal fee will be charged for access and use of the blinded data set to cover the CSRC costs associated with the maintenance of the ECG data set and the algorithm scoring of the blinded data set.

Publication/dissemination of algorithm performance

Publication or dissemination of algorithm performance will be expected from all investigators and algorithm developers using the blinded data set (s). Commitments to public domain information release will be obtained at the time of SOC research project approval for use of these data sets, and resulting reports will be expected to be delivered by database users in a format that is publicly available. Database users will also agree that any publication or dissemination of results from unblinded study cohorts will clearly indicate that the performance results represent only unblinded "training set" data and hence does not constitute a validated performance assessment of the algorithm under study.

Summary

Placebo and active control ECG waveforms in XML format that have been submitted to the FDA as part of TQT studies and subsequently released by their sponsors are available for research purposes from the CSRC. Further information regarding access to the ECG data sets may be obtained from the CSRC at www.cardiac-safety.org.

Acknowledgement

CLG was supported in part by a Clinical and Translational Sciences Award Grant to Duke University (UL1 RR024128).

References

1. CSRC scientific rationale and mission statement. Available at: <http://www.cardiac-safety.org/about-us>. Last accessed June 10, 2010.
2. CSRC ECG Warehouse database use statement of principles. Available at: <http://www.cardiac-safety.org/ecg-database-tool-kit/Data%20Agreement.pdf/view>. Last accessed June 10, 2010.
3. Available at: <http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm>. Last accessed 08/08/10.
4. Food and Drug Administration. International Conference on Harmonisation; guidance on E14 clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs; availability. Notice. Fed Regist 2005;70:61134-5.
5. Shah RR. Drugs, QTc interval prolongation and final ICH E14 guideline: an important milestone with challenges ahead. Drug Saf 2005;28:1009-28.
6. Zhang J, Machado SG. Statistical issues including design and sample size calculation in thorough QT/QTc studies. J Biopharm Stat 2008; 18:451-67.
7. Darpo B. The thorough QT/QTc study 4 years after the implementation of the ICH E14 guidance. Br J Pharmacol 2010; 159:49-57.
8. Couderc JP, Zareba W, Moss AJ, et al. Identification of sotalol-induced changes in repolarization with T wave area-based repolarization duration parameters. J Electrocardiol 2003;36: 115-20.
9. Sarapa N, Morganroth J, Couderc JP, et al. Electrocardiographic identification of drug-induced QT prolongation: assessment by different recording and measurement methods. Ann Noninvasive Electrocardiol 2004;9:48-57.
10. Xue J, Gao W, Chen Y, et al. Identify drug-induced T wave morphology changes by a cell-to-electrocardiogram model and validation with clinical trial data. J Electrocardiol 2009;42: 534-42.