EHR-based Machine Learning Model Predicts Drug-induced QT Prolongation With Superior Performance Compared To Clinical Risk Predictors

Linyuan Jing, PhD,¹,² Thomas Morland, MD,¹ Christopher Kelsey,¹ Sushravya Raghunath, PhD,² John Pfeifer, MD,² Jeffrey Ruhl,¹ Brandon Fornwalt, MD, PhD,² Christopher Haggerty, PhD¹

1. Geisinger, Danville, PA, USA
2. Tempus Labs Inc, Chicago, IL, USA

tmorland@geisinger.edu
Disclosures

• Funding
  • Geisinger receives funding from Tempus for ongoing development of predictive modeling technology and commercialization
  • None of the Geisinger authors have ownership interests in any of the intellectual property resulting from the partnership. However, Dr. Fornwalt, Dr. Raghunath, and Dr. Pfeifer are now Tempus employees

• Individual Disclosures
  • L. Jing: Employee of Tempus Labs, Inc. T. Morland: Research Grant; Significant; Tempus Labs, Inc. Stock Shareholder; Significant; United Therapeutics Corporation. Other; Significant; United Therapeutics Corporation. C.R. Kelsey: None. S. Raghunath: Employee of Tempus Labs, Inc. J. Pfeifer: Employee of Tempus Labs, Inc. J. Rühl: None. B. Fornwalt: Other; Significant; Tempus. C.M. Haggerty: Stock Shareholder; Modest; Merck & Co. Other; Significant; Tempus.
What is Drug-induced Long QTc?

- Drug-induced long QTc is an iatrogenic delay in ventricular repolarization which can lead to the life-threatening arrhythmia, *torsades de pointes* (TDP)

- >200 drugs are QTc prolongers
  - Antiarrhythmics
  - Antibiotics
  - Antipsychotics
  - Antidepressants
  - Cancer therapies
Current Solutions

Tisdale Score
• Predicts QTc >500 ms or >60 ms compared to baseline
• Prospective, observational study of 1,200 patients in cardiac critical care unit
• Good performance: AUC=0.832

• Limitation: lack of generalizability due to inpatient focus

Optimized RISQ-PATH Score
• Broader application
  • Applicable to outpatient
  • Moderate performance: AUC=0.772

• Limitation: data collected at target ECG, not reflecting baseline status at drug start

Tisdale et al. (2013): Circulation: Cardio Quality/Outcomes
Vandael et al. (2018): British Journal of Clinical Pharmacology
Potential for a Machine Learning Approach

Machine learning has demonstrated high accuracy (e.g., AUROC >0.900) on targets related to QTc but has not been used to produce a general clinical model with high accuracy.
Design Criteria

• Highly accurate across a variety of clinical settings and scenarios

• Derived and validated utilizing data commonly available in the electronic health record (EHR) *at the time of medical decision-making*

• Assessment of drug-induced long QTc risk in general population as opposed to a congenital LQTS population
Study Design

Structure EHR features
- Demographics
- Vital Signs
- Lab tests
- Medications
- ECG measures
- ECG patterns

12-lead, 10-second voltage data

Deep Neural Network
- Clean
- Scale
- Impute
- XGBoost Classifier

Risk Score
- High risk of long QTc
- Low risk of long QTc

MUSE

Geisinger
Study Population

-3 year  -1 year  1 year
Pre-drug start  QTc drug start  QTc drug end

XGB set
Epic EHR features
Baseline ECG features
Baseline ECG traces

ECG DNN set
Target ECG
Prediction window

Endpoint: QTc > 500 ms from target on-drug ECG

<table>
<thead>
<tr>
<th></th>
<th>XGB dataset</th>
<th>ECG DNN dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>N patients</td>
<td>345,371</td>
<td>182,448</td>
</tr>
<tr>
<td>% Long QTc</td>
<td>5.7%</td>
<td>7.7%</td>
</tr>
<tr>
<td>Median Age, yrs</td>
<td>62 [48, 74]</td>
<td>66 [54, 78]</td>
</tr>
<tr>
<td>% Male sex</td>
<td>45%</td>
<td>47%</td>
</tr>
<tr>
<td>Total med count</td>
<td>2 [1, 3]</td>
<td>2 [1, 3]</td>
</tr>
</tbody>
</table>
Both XGBoost and DNN Trace Models Show Comparable Performance

- Dataset:
  - N = 182,448,
  - 7.7% long QTc (>500 ms)
- No significant improvement in composite model

All models were evaluated by 5-fold cross-validation (reported as mean[95% CI])
Machine Learning Shows Superior Predicting Power to Tisdale and RISQ-PATH

N=110,588, 8.8% long QTc

AUC: 0.86

AUC: 0.77

N=345,371, 5.7% long QTc

AUC: 0.86

AUC: 0.70

Tisdale scores were only available in patients with ECGs within 24h of drug start
Machine Learning Shows Superior Performance in All Clinical Settings

<table>
<thead>
<tr>
<th>Setting</th>
<th>Data size</th>
<th>% Long QTc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>110,588</td>
<td>8.8</td>
</tr>
<tr>
<td>Office Visit</td>
<td>32,271</td>
<td>3.3</td>
</tr>
<tr>
<td>Admission</td>
<td>62,528</td>
<td>13</td>
</tr>
<tr>
<td>ED</td>
<td>15,759</td>
<td>3.6</td>
</tr>
</tbody>
</table>

- Tisdale scores were predominantly calculated in hospital admissions
- In contrast, 75% of the overall dataset were from outpatient records
Machine Learning Model Has Better PPV

• When matching number of patients predicted as high-risk for long QTc, XGBoost model shows superior PPV (>50%) as compared to Tisdale and RISQ-PATH

<table>
<thead>
<tr>
<th></th>
<th>% Long QTc</th>
<th>% Predicted High-risk</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>XGBoost</td>
<td>3.1 [2.4, 3.8]</td>
<td>28.9 [22.2, 35.5]</td>
<td>98.5 [98.0, 99.0]</td>
<td>54.0 [49.8, 58.3]</td>
<td>95.8 [95.1, 96.5]</td>
<td></td>
</tr>
<tr>
<td>RISQ-PATH</td>
<td>5.7 [5.2, 6.2]</td>
<td>2.0 [1.3, 2.7]</td>
<td>9.8 [6.7, 12.9]</td>
<td>98.5 [97.9, 99.0]</td>
<td>28.3 [25.8, 30.7]</td>
<td>94.7 [94.1, 95.3]</td>
</tr>
<tr>
<td>XGBoost</td>
<td>2.0 [1.5, 2.5]</td>
<td>21.4 [16.5, 26.3]</td>
<td>99.2 [98.9, 99.5]</td>
<td>61.5 [56.8, 66.2]</td>
<td>95.4 [94.8, 96.0]</td>
<td></td>
</tr>
</tbody>
</table>

Results are reported as mean [95% confidence interval] across 5 test folds
Conclusions

• Machine learning models, either using structured EHR features or using ECG traces alone, can predict drug-induced QTc prolongation at medication initiation in a general clinical population with high accuracy

➢ Composite model using both EHR features and DNN risk score shows marginal improvement in model performance

• Machine learning model is superior to clinical risk predictors, Tisdale and RISQ-PATH at different clinical settings and operating points
Limitations

• Model trained on all QTc drugs: variable QTc risks and different mechanisms of action
  • Model was evaluated on commonly used, individual drugs but performance was unreliable due to small sample size for single medication

• Model endpoint QTc >500 ms
  • Rational:
    ➢ QTc > 500 ms can be harmful with or without TDP, e.g., interrupted or adjusted pharmacotherapy
    ➢ QTc > 500ms is a reasonable surrogate for risk of death from TDP
  • Limitation: TDP events may not be captured
    ➢ Death from TDP rarely captured on ECG

• Retrospective data from a single health system
  • Could validate initial findings with external data and/or prospective study
Acknowledgements

• Geisinger and Tempus teams
• Tempus Laboratories funding
• Thank you!

tmorland@geisinger.edu
Appendix
Machine Learning Model Shows Better Metrics than Original Tisdale/RISQ-PATH Studies

When matching sensitivities originally reported by Tisdale and RISQ-PATH, XGBoost model shows better specificity.

<table>
<thead>
<tr>
<th></th>
<th>% Long QTc</th>
<th>% Predicted High-risk</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original Tisdale Cohort</strong></td>
<td>30.7</td>
<td></td>
<td>74</td>
<td>77</td>
</tr>
<tr>
<td><strong>XGBoost</strong></td>
<td>5.7</td>
<td>21.6</td>
<td>74.3</td>
<td>81.6</td>
</tr>
<tr>
<td>[5.2, 6.2]</td>
<td>[16.4, 26.7]</td>
<td></td>
<td>[65.7, 82.8]</td>
<td>[76.5, 86.7]</td>
</tr>
<tr>
<td><strong>Original Optimized RISQ-PATH Cohort</strong></td>
<td>5.9</td>
<td></td>
<td>87.4</td>
<td>46.2</td>
</tr>
<tr>
<td>[5.2, 6.2]</td>
<td>[86.2, 88.5]</td>
<td></td>
<td>[45.8, 46.6]</td>
<td></td>
</tr>
<tr>
<td><strong>XGBoost</strong></td>
<td>5.7</td>
<td>52.8</td>
<td>92.4</td>
<td>49.7</td>
</tr>
<tr>
<td>[5.2, 6.2]</td>
<td>[41.8, 63.7]</td>
<td></td>
<td>[38.3, 61.0]</td>
<td></td>
</tr>
</tbody>
</table>

Results are reported as mean [95% confidence interval] across 5 test folds
Patient Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median[IQR] or %</th>
<th>Variable</th>
<th>Median[IQR] or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>N patients</td>
<td>345,371</td>
<td>Heart failure</td>
<td>13%</td>
</tr>
<tr>
<td>Long QTc</td>
<td>5.7%</td>
<td>Diabetes</td>
<td>21%</td>
</tr>
<tr>
<td>Age, yrs</td>
<td>62.2 [48.4, 74.4]</td>
<td>Hypertension</td>
<td>49%</td>
</tr>
<tr>
<td>Male sex</td>
<td>45.1%</td>
<td>Myocardial Infarction</td>
<td>11%</td>
</tr>
<tr>
<td>White race</td>
<td>95.7%</td>
<td>QTc, ms</td>
<td>440 [421, 462]</td>
</tr>
<tr>
<td>BMI, kg/cm²</td>
<td>29.5 [25.1, 35]</td>
<td>Max QTc ever</td>
<td>456 [434, 484]</td>
</tr>
<tr>
<td>Ever smoker</td>
<td>51.7%</td>
<td>QTc med count</td>
<td>2 [1, 3]</td>
</tr>
<tr>
<td>Potassium</td>
<td>4.2 [3.9, 4.5]</td>
<td>Sr. Creatinine</td>
<td>0.9 [0.8, 1.1]</td>
</tr>
</tbody>
</table>
Machine Learning Shows Superior Performance in All Clinical Settings

<table>
<thead>
<tr>
<th>Setting</th>
<th>Data size</th>
<th>% Long QTc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tisdale subset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Visit</td>
<td>32,271</td>
<td>3.3</td>
</tr>
<tr>
<td>Admission</td>
<td>62,528</td>
<td>13</td>
</tr>
<tr>
<td>ED</td>
<td>15,759</td>
<td>3.6</td>
</tr>
<tr>
<td>Overall dataset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office Visit</td>
<td>262,294</td>
<td>4.1</td>
</tr>
<tr>
<td>Admission</td>
<td>66,137</td>
<td>12.9</td>
</tr>
<tr>
<td>ED</td>
<td>16,940</td>
<td>3.7</td>
</tr>
</tbody>
</table>