Beyond Predictive Powers

A Comprehensive Comparison of T1D Risk Models

Lauric Ferrat
Introduction

Simulation of T1D screening by Gonçalo Lieria
What is a good predictive model?

1. Accurate
2. Cost is not prohibitive
3. Patient burden is low
4. Accessible
Method

- Generate a large set of predictive models;
- Assess them not only in term of predictive power but in term of cost and patient time.
Variables linked to T1D risk

1. Clinical
   - Age, logarithm of age
   - BMI, Z-BMI
   - Sex

2. Immunology
   - IA2A (Positive or Negative)
   - Autoantibody group (GAD, IAA, IA2A, GAD-IAA, GAD-IA2A, IAA-IA2A, GAD-IAA-IA2A)

3. Genetic
   - GRS2

4. Metabolic
   - derived measured from Oral Glucose Tolerance Test: AUC glucose, AUC C-peptide, C peptide index 30, β2-score, Index_{60}^{1}, DPTRS^{2}, and DPTRS_{60}^{2}, M_{120}^{3}, CPH^{4} and LR^{4}
   - HbA1c

Model fitting – generation of combinations of variables

> 2 millions different combinations of variables
Some rules when added to only generate plausible models such as

- **BMI** and **z-BMI** would not appear in the same model;
- A model would not have two OGTT derived variables simultaneously.

Here a list of the 100 first ones (out of 2154 models):

```r
formula
1  GRS2
2  age
3  BMI
4  log(age)
5  Index69
6  AUC-glucose
7  AUC-crepeptide
8  z-BMI
9  C-peptide-index-30
10 beta2-score
11 Hba1c
12 GRS2 + IA2
13 GRS2 + AB-group
14 GRS2 + Sex
15 GRS2 + age
16 IA2 + age
17 AB-group + age
18 Sex + age
19 GRS2 + BMI
20 IA2 + BMI
21 AB-group + BMI
22 Sex + BMI
23 age + BMI
24 GRS2 + log(age)
25 IA2 + log(age)
26 AB-group + log(age)
27 Sex + log(age)
28 BMI + log(age)
29 GRS2 + Index69
30 IA2 + Index69
31 AB-group + Index69
32 Sex + Index69
33 age + Index69
34 BMI + Index69
```
**Model fitting**

The TrialNet Pathway to Prevention dataset was split in two to train and test the predictive model.

<table>
<thead>
<tr>
<th></th>
<th>Training dataset (N=1524)</th>
<th>Testing dataset (N=1551)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>Female</td>
<td>728 (47.8%)</td>
<td>663 (42.7%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>796 (52.2%)</td>
<td>888 (57.3%)</td>
<td></td>
</tr>
<tr>
<td><strong>age</strong></td>
<td>9.22 (4.3)</td>
<td>9.15 (4.2)</td>
<td>0.628</td>
</tr>
<tr>
<td><strong>GRS2</strong></td>
<td>13.24 (2.1)</td>
<td>13.44 (2.1)</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>IA2</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>IA2</td>
<td>621 (40.7%)</td>
<td>731 (47.1%)</td>
<td></td>
</tr>
<tr>
<td>IA2 free</td>
<td>903 (59.3%)</td>
<td>820 (52.9%)</td>
<td></td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td>18.96 (4.5)</td>
<td>18.61 (4.6)</td>
<td>0.032</td>
</tr>
<tr>
<td><strong>AUC Cpeptide</strong></td>
<td>20.11 (9.5)</td>
<td>20.80 (10.2)</td>
<td>0.052</td>
</tr>
<tr>
<td><strong>AUC glucose</strong></td>
<td>522.83 (103.4)</td>
<td>542.21 (107.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Hba1c</strong></td>
<td>5.09 (0.3)</td>
<td>5.11 (0.3)</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>Stage</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0</td>
<td>760 (49.9%)</td>
<td>575 (37.1%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>543 (35.6%)</td>
<td>669 (43.1%)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>221 (14.5%)</td>
<td>307 (19.8%)</td>
<td></td>
</tr>
<tr>
<td><strong>T1D</strong></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>T1D</td>
<td>631 (41.4%)</td>
<td>515 (33.2%)</td>
<td></td>
</tr>
<tr>
<td>T1D free</td>
<td>893 (58.6%)</td>
<td>1036 (66.8%)</td>
<td></td>
</tr>
</tbody>
</table>


Results
Model performance in single AB

Variables included in the model
- Base (clinical + AB)
- Base + GRS
- Base + Metabolic
- Base + GRS + Metabolic
- Classic models

Brier score vs. Time dependent ROC AUC
Variable importance in single AB

- **z-BMI**
- **GRS2**
- **Index 60**
- **AUC glucose**
Model performances by presence of AUC Glucose

Presence of AUC_glucose in model
- Missing
- Present
Model Cost at single AB

Performance and cost models at 3 years horizon

Brier score vs. AUC

cost in $ to predict risk

AUC

0.8

0.7

0.6

0.5

0.11

0.10

0.09

0.08

0.275

0.250

0.225

0.200

0.175

0.150

0.125

0.100

0.075

0.050

0.025
Model patient time at single AB

Performance and patient time at 3 years horizon

Time needed to acquire data

Visit
- can be done from home
- clinicians needed

AUC

Brier score
Variables included in the model
- Base (clinical + AB)
- Base + GRS
- Base + Metabolic
- Base + GRS + Metabolic
- Classic models

Stage 1 and 2
A good model is an accessible model

Model available at https://t1dpredictor.diabetesgenes.org/
A good model is an accessible model
Easy to do a prototype
Hard to be legally compliant FDA, GDPR

<table>
<thead>
<tr>
<th>Standard</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEC 62304:2006</td>
<td>Medical device software — Software life cycle processes</td>
</tr>
<tr>
<td>ISO 13485</td>
<td>MEDICAL DEVICES</td>
</tr>
<tr>
<td>ISO/IEC 27001</td>
<td>INFORMATION SECURITY MANAGEMENT</td>
</tr>
</tbody>
</table>
**Conclusion**

- Despite its cost and its burden for patient OGTT related measures appear as critical for a good T1D prediction
- Cost and patient time can change drastically for similar predictive performance
- A good model is an accessible model