Classification of asthma using longitudinal lung function test data, and development of prognosis prediction model

Youjin Park, Pharm.D ¹, Seng Chan You, MD ², Youngsoo Lee, MD ³, Hyun Young Lee, Ph.D ⁴, Eun Young Lee, MS ², Hae-Sim Park, M.D, Ph.D ³, Rae Woong Park, MD, Ph.D ¹, ²

¹Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea
²Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea
³Department of Allergy and Clinical Immunology, Ajou University School of Medicine, Suwon, Korea
⁴Department of Statistics, Clinical Trial Center, Ajou University Medical Center, Suwon, Korea
• Introduction of **trajectory clustering (LCMM)**
• Application of **LCMM in OHDSI ecosystem**
• **Use case**: *New phenotyping of asthmatics using long-term followed lung function data over 15 years*
• Extension of **PatientLevelPrediction**
Trajectory clustering

Trajectory means the course of measured variable over age or time
Trajectory clustering

Are they same?

Are they really same?
Trajectory clustering

**Latent class mixed model (LCMM)**

- innovative statistical method used to **identify subgroups** of participants **with heterogeneous trajectories**

**Multinomial logistic model**

Express for each patient probability to belong to class

**Linear mixed model**

Find the latent profile of trajectories in each classes
Trajectories of Long-Term Normal Fasting Plasma Glucose and Risk of Coronary Heart Disease: A Prospective Cohort Study

Zhongheng Yuan, PhD; Yang Yang, MS; Chunxia Wang, MS; Jing Liu, PhD; Xiubin Sun, PhD; Yi Liu, PhD; Shengou Li, MD, PhD; Fudong Xu, PhD

Background—Fasting plasma glucose (FPG) levels can vary over time and its longitudinal changing patterns may predict cardiometabolic risk. We aim to identify different trajectories of FPG in those who remained normoglycemic and investigate the association between trajectory groups and coronary heart disease risk in a large prospective cohort study.

Methods and Results—A total of 20,514 subjects between ages 20 and 80 years were included at baseline. All participants had maintained normal FPG throughout an average follow-up period of 5.8 years. We identified 3 distinct trajectories using a group-based trajectory model, labeled by initial value and changing pattern: low-increasing (n=12,694), high-increasing-decreasing (n=5330), and high-decreasing-increasing (n=2490). The coronary heart disease incidence density among these 3 groups (3.00, 4.05, and 3.26 per 1000 person-years, respectively) was significantly different (P<0.0038). The high-increasing-decreasing group was characterized by a starting FPG of 4.80 mmol/L, and increased up to 5.42 mmol/L at age 55, then decreased thereafter. Treating the low-increasing group as the reference, the age- and sex-adjusted hazard ratio was 1.08 (95% confidence interval, 1.02–1.02) for the high-increasing-decreasing group by Cox proportional hazard regression. After adjustment for other potential confounding factors, the hazard ratio is 1.40 (95% confidence interval, 1.08–1.81). The association persisted after adjustment for baseline FPG, mean, or SD of FPG.

Conclusions—Distinct trajectories of long-term normal FPG are associated with the development of coronary heart disease, which is independent of other metabolic factors including FPG levels. These findings have implications for intervention and prevention of coronary heart disease among individuals who are normoglycemic. (J Am Heart Assoc. 2018;7:e007607. DOI: 10.1161/JAHA.117.007607.)

Key Words: epidemiology • fasting plasma glucose • group-based trajectory model • proportional hazard regression • cardiovascular disease risk factors
LCMM in OHDSI ecosystem

- Setting for trajectory clustering using LCMM package

<table>
<thead>
<tr>
<th>Needed information</th>
<th>In OMOP-CDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient number</td>
<td>Person ID (or subject ID)</td>
</tr>
<tr>
<td>Measured value</td>
<td>Value as number</td>
</tr>
<tr>
<td>Measurement date or time to measure</td>
<td>Index date and measurement date</td>
</tr>
<tr>
<td>Age</td>
<td>Year of birth and measurement date</td>
</tr>
<tr>
<td>Gender</td>
<td>Gender concept Id</td>
</tr>
<tr>
<td>Event (death etc.)</td>
<td>Event cohort created by ATLAS</td>
</tr>
</tbody>
</table>

Not necessary

**LCMM can operate with OHDSI ecosystem!**
LCMM in OHDSI ecosystem

• If the results of trajectory clustering are inserted into cohort table (result table), all OHDSI tools can be used such as PatientLevelPrediction.

Cohort Definition ID
Subject ID
Cohort start date
Cohort end date

All we have to do is defined cohort!
Develop LCMMohdsi

https://github.com/ABMI/LcmmOhdsi

Under develop!
Trajectory clustering to asthmatics

ICARUS Project
(Immune/Inflammatory Disease CDM Augmentation for Research Union System)

Lung function test data over 15 years were converted into OMOP CDM
Trajectory clustering to asthmatics

- Forced expiratory volume in one second ($\text{FEV}_1$) is mainly used as a lung function index when rating respiratory disorders.
Trajectory clustering to asthmatics

• Study population: **Severe asthmatics**

  - ICARUS Database
    - N = 4,893

  - Severe asthmatics
    - N = 1,388

  - At least 1 measurement of FEV1(%) value
    - N = 1,174

  - GINA guideline step 4-5 medication (MD/HD ICS + LABA)

Trajectory clustering to asthmatics

- **Trajectory clustering** using LCMM R package
  - Measured variable: $\text{FEV}_1$ (%)
  - ‘Time = 0’ is index date which represented HD/MD ICS-LABA combination starting date
  - ‘Time = t’ is time to measure (years) which was calculated as ‘measurement date – index date’
  - The shape and optimal number of classes were determined by following criteria:
    1. The lower Bayesian information criterion (BIC) score
    2. Proportion of each classes were more than 5 %
• Outcome: **Asthma exacerbation**

  – Definition of asthma exacerbation:
    1. Oral corticosteroid prescription for \( \geq 3 \) days
       (15mg/day of prednisolone or its equivalent dose)
    2. Emergency room visit
**Result**: Estimated trajectory of lung function from LCMM

Table 1: Latent class mixed model (LCMM) results of model fitting process

<table>
<thead>
<tr>
<th># of class</th>
<th>Shape of trajectory</th>
<th>AIC</th>
<th>BIC</th>
<th>%Class1</th>
<th>%Class2</th>
<th>%Class3</th>
<th>%Class4</th>
<th>%Class5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>2</td>
<td>Linear</td>
<td>73752.78</td>
<td>73803.25</td>
<td>6.26</td>
<td>93.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>2</td>
<td>Quadratic</td>
<td>73741.75</td>
<td>73802.32</td>
<td>4.61</td>
<td>95.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>2</td>
<td>Cubic</td>
<td>73711.28</td>
<td>73781.94</td>
<td>4.43</td>
<td>95.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model 4</strong></td>
<td>3</td>
<td>Linear</td>
<td>73749.14</td>
<td>73819.80</td>
<td>12.61</td>
<td>34.75</td>
<td>52.64</td>
<td></td>
</tr>
<tr>
<td>Model 5</td>
<td>3</td>
<td>Quadratic</td>
<td>73696.37</td>
<td>73828.12</td>
<td>2.61</td>
<td>4.61</td>
<td>92.78</td>
<td></td>
</tr>
<tr>
<td>Model 6</td>
<td>3</td>
<td>Cubic</td>
<td>73634.20</td>
<td>73735.15</td>
<td>3.57</td>
<td>3.83</td>
<td>92.61</td>
<td></td>
</tr>
<tr>
<td>Model 7</td>
<td>4</td>
<td>Linear</td>
<td>73743.37</td>
<td>73834.23</td>
<td>31.57</td>
<td>9.30</td>
<td>7.83</td>
<td>51.30</td>
</tr>
<tr>
<td>Model 8</td>
<td>4</td>
<td>Quadratic</td>
<td>73679.05</td>
<td>73790.10</td>
<td>2.17</td>
<td>2.09</td>
<td>2.52</td>
<td>93.22</td>
</tr>
<tr>
<td>Model 9</td>
<td>4</td>
<td>Cubic</td>
<td>73623.74</td>
<td>73754.97</td>
<td>3.65</td>
<td>0.61</td>
<td>3.65</td>
<td>92.09</td>
</tr>
<tr>
<td>Model 10</td>
<td>5</td>
<td>Linear</td>
<td>73739.72</td>
<td>73850.76</td>
<td>10.17</td>
<td>24.00</td>
<td>52.52</td>
<td>5.04</td>
</tr>
<tr>
<td>Model 11</td>
<td>5</td>
<td>Quadratic</td>
<td>73684.59</td>
<td>73820.88</td>
<td>1.83</td>
<td>3.04</td>
<td>9.22</td>
<td>63.65</td>
</tr>
<tr>
<td>Model 12</td>
<td>5</td>
<td>Cubic</td>
<td>73595.67</td>
<td>73757.19</td>
<td>3.13</td>
<td>0.70</td>
<td>5.74</td>
<td>0.96</td>
</tr>
</tbody>
</table>

The model which was 'linear' and '3 classes' was chosen
Result: Estimated trajectory of lung function from LCMM

Figure 1-a. observed individual lung function trajectories and three estimated representative trajectories

- Persistently high, n = 408, 34.75%
- Persistently low, n = 618, 52.64%
- Declining, n = 148, 12.61%

<table>
<thead>
<tr>
<th></th>
<th>Persistently high</th>
<th>Persistently low</th>
<th>Declining</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N = 408)</td>
<td>(N = 618)</td>
<td>(N = 148)</td>
</tr>
<tr>
<td>Intercept (95% CI)</td>
<td>102.947 (96.47, 109.42)</td>
<td>87.57 (80.98, 94.15)</td>
<td>84.08 (78.23, 89.93)</td>
</tr>
<tr>
<td>Slope (95% CI)</td>
<td>-0.05 (-0.38, 0.27)</td>
<td>-0.14 (-0.36, 0.09)</td>
<td>-0.66 (-1.27, -0.05)</td>
</tr>
</tbody>
</table>

Figure 1-b. Tree estimated representative trajectories. The shaded areas indicate estimated 95% confidential intervals
Result: variability of lung function

Figure 2. Distribution of each person’s standard deviation of predicted FEV1 (%). P values were calculated according to Wilcoxon test.

<table>
<thead>
<tr>
<th></th>
<th>Persistently high (N = 408)</th>
<th>Persistently low (N = 618)</th>
<th>Declining (N = 148)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median of standard deviation of predicted FEV1 (%) within person</td>
<td>5.4 (3.4, 8.0)</td>
<td>6.5 (4.2, 9.5)</td>
<td><strong>11.1</strong> (6.6, 14.9)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Median (interquartile). P value was calculated according to Kruskal-wallis test.
# Result: Baseline characteristics

<table>
<thead>
<tr>
<th></th>
<th>Persistently high (N = 408)</th>
<th>Persistently low (N = 618)</th>
<th>Declining (N = 148)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, years</strong></td>
<td>38.00 ± 15.03</td>
<td>38.82 ± 13.8</td>
<td>46.42 ± 13.55</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Follow-up duration, years</strong></td>
<td>7.05 (4.15, 10.74)</td>
<td>6.59 (3.37, 10.72)</td>
<td>8.15 (4.11, 11.57)</td>
<td>0.105</td>
</tr>
<tr>
<td><strong>Female, n (%)</strong></td>
<td>252 (61.76)</td>
<td>337 (54.53)</td>
<td>89 (60.14)</td>
<td>0.059</td>
</tr>
<tr>
<td><strong>BMI, kg/m² (N)</strong></td>
<td>24.12 ± 13.17 (167)</td>
<td>25.71 ± 33.78 (373)</td>
<td>24.24 ± 4.81 (110)</td>
<td>0.705</td>
</tr>
<tr>
<td><strong>Total IgE (KU/L)</strong></td>
<td>213.5 (79.3, 512.8) (346)</td>
<td>241 (92, 477) (501)</td>
<td>199 (46.5, 535) (103)</td>
<td>0.688</td>
</tr>
<tr>
<td><strong>Blood eosinophil (%)</strong></td>
<td>3.7 (1.9, 6.8) (375)</td>
<td>4.2 (1.8, 7.5) (562)</td>
<td>4.4 (2.7, 8.5) (127)</td>
<td>0.034</td>
</tr>
<tr>
<td><strong>Blood neutrophil (%)</strong></td>
<td>55.4 (49.4, 61.9) (374)</td>
<td>54.9 (48.4, 62.7) (560)</td>
<td>55.7 (48.8, 62.1) (126)</td>
<td>0.764</td>
</tr>
<tr>
<td><strong>Serum EDN (ng/mL)</strong></td>
<td>46.5 (30.5, 63.3) (61)</td>
<td>50.1 (31.3, 83.1) (102)</td>
<td>65.8 (45.9, 100.2) (24)</td>
<td>0.009</td>
</tr>
<tr>
<td><strong>Serum periostin (ng/mL)</strong></td>
<td>75.0 (54.0, 102.0) (97)</td>
<td>73.2 (51.6, 95.0) (185)</td>
<td>74.0 (52.9, 115.3) (48)</td>
<td>0.676</td>
</tr>
<tr>
<td><strong>Rhinosinusitis, N (%)</strong></td>
<td>284 (69.6)</td>
<td>383 (62.0)</td>
<td>83 (56.1)</td>
<td>0.005</td>
</tr>
<tr>
<td><strong>Urticaria/angioedema, N (%)</strong></td>
<td>36 (8.8)</td>
<td>47 (7.6)</td>
<td>2 (1.4)</td>
<td>0.010</td>
</tr>
<tr>
<td><strong>Anaphylaxis, N (%)</strong></td>
<td>17 (4.2)</td>
<td>20 (3.2)</td>
<td>3 (2.0)</td>
<td>0.443</td>
</tr>
<tr>
<td><strong>Hypertension, N (%)</strong></td>
<td>13 (3.2)</td>
<td>28 (4.5)</td>
<td>9 (6.1)</td>
<td>0.291</td>
</tr>
<tr>
<td><strong>Diabetes Mellitus, N (%)</strong></td>
<td>7 (1.7)</td>
<td>12 (1.9)</td>
<td>3 (2.0)</td>
<td>0.956</td>
</tr>
<tr>
<td><strong>Osteoporosis, N (%)</strong></td>
<td>10 (2.5)</td>
<td>9 (1.5)</td>
<td>5 (3.4)</td>
<td>0.257</td>
</tr>
<tr>
<td><strong>GERD, N (%)</strong></td>
<td>12 (2.9)</td>
<td>21 (3.4)</td>
<td>6 (4.1)</td>
<td>0.802</td>
</tr>
<tr>
<td><strong>Ischemic heart disease, N (%)</strong></td>
<td>2 (0.5)</td>
<td>9 (1.5)</td>
<td>7 (4.7)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Mean ± Standard; Median (interquartile range); BMI, Body mass index; GERD, Gastroesophageal reflux disease.
Result: Yearly counts of event cohort and 1 year event free survival

**Figure 3-(a)** Yearly count of asthma exacerbation requiring oral corticosteroid

**Figure 3-(b)** Kaplan-Meier survival curve for the time to the asthma exacerbation requiring oral corticosteroid
**Result**: Yearly counts of event cohort and 1 year event free survival

**Figure 3-(c)** Yearly count of **emergency room visit** due to asthma exacerbation

**Figure 3-(d)** Kaplan-Meier survival curve for the time to **emergency room visit** due to asthma exacerbation
Who will be included in declining group?

With `patientLevelPrediction`, the most negative group was predicted using the data 1 year before the index date.
Result: patient level prediction

Figure 4. The AUROC curves for the declining group predicting model. A: lasso logistic regression, B: gradient boosting model.
Extension of PatientLevelPrediction

Apply model identifying cat for identifying dog

Is it cat?  Is it dog?

Through the PatientLevelPrediction package, we can do similar analysis
Extension of PatientLevelPrediction

Apply model predicting declining group for predicting asthma exacerbation

Who will be included in the declining group?

Who will be visit emergency room for asthma within 365 days from index date
## Result: patient level prediction

<table>
<thead>
<tr>
<th>Target cohort</th>
<th>Outcome cohort</th>
<th>Machine learning model</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUMC</td>
<td>8873</td>
<td>Lasso logistic</td>
<td>0.786</td>
</tr>
<tr>
<td></td>
<td>303</td>
<td>Gradient boosting</td>
<td>0.642</td>
</tr>
</tbody>
</table>

Figure 5. The AUROC curves of AUMC validation groups. A: lasso logistic regression, B: gradient boosting model.
Result: patient level prediction

<table>
<thead>
<tr>
<th>Target cohort</th>
<th>Outcome cohort</th>
<th>Machine learning model</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHIS-NSC</td>
<td>17240</td>
<td>Lasso logistic</td>
<td>0.685</td>
</tr>
<tr>
<td></td>
<td>771</td>
<td>Gradient boosting</td>
<td>0.541</td>
</tr>
</tbody>
</table>

Figure 6. The AUROC curves of NHIS-NSC validation groups. A: lasso logistic regression, B: gradient boosting model
Conclusion

• **LCMM** can operate with OHDSI ecosystem

• Through the LCMM and OHDSI tools, **severe asthmatics can be classified** according to long-term change of lung function

• For prediction declining lung function, the **PatientLevelPrediction can be used**
Conclusion

• Not only cross-sectional phenotyping, it is also important to define phenotype by long-term changes over time
OHDSI: Open Innovation based on the open community

- Reproducibility
- Scalability
- Beneficience

Global Community

- Standards
- Openness
- Data
- Collaboration
• Our Mission
To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

• Our Vision
A world in which observational research produces a comprehensive understanding of health and disease.
Mission, Vision, and Values of OHDSI

• Innovation: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

• Reproducibility: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.

• Community: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.

• Collaboration: We work collectively to prioritize and address the real world needs of our community’s participants.

• Openness: We strive to make all our community’s proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.

• Beneficence: We seek to protect the rights of individuals and organizations within our community at all times.
Thank You!

Youjin Park, Pharm.D
Department of Biomedical Sciences, Ajou University
Graduate School of Medicine, Suwon, Korea
H.P. : 010-9942-7796
E-mail : dbwls5223@ajou.ac.kr

Acknowledgement: This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [grant number: HI16C0992].