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]

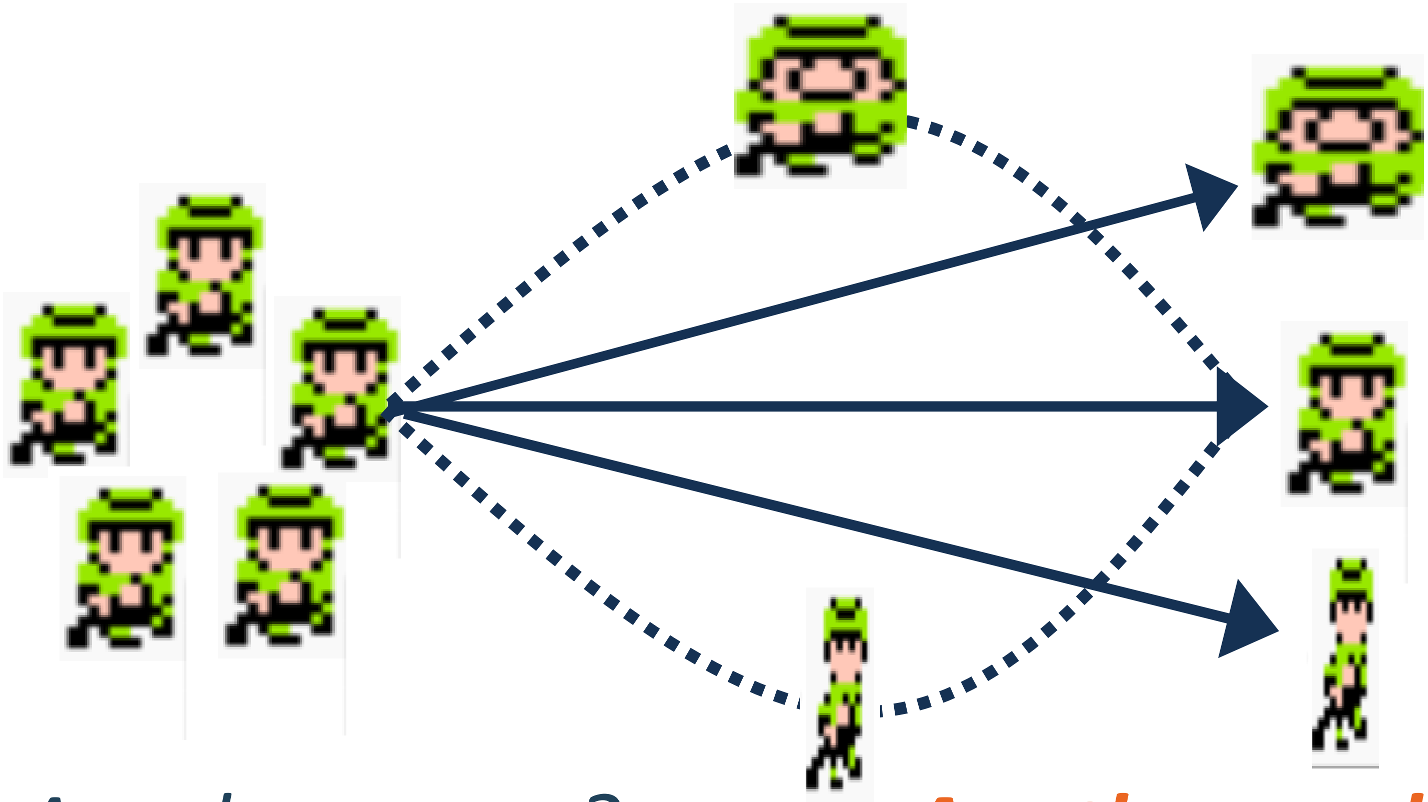
[1]*Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea*
[2]*Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea*
[3]*Department of Allergy and Clinical Immunology, Ajou University School of Medicine, Suwon, Korea*
[4]*Department 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	**Linear mixed model**
Express for each patient probability to belong to class	Find the latent profile of trajectories in each classes

Trajectory clustering

Epidemiology/Population

Blood pressure

BMI

Plasma Glu

ORIGINAL RESEARCH

ORIGINAL RESEARCH

Trajectories of Long-Term Normal Fasting Plasma Glucose and Risk of Coronary Heart Disease: A Prospective Cohort Study

Zhongshang Yuan, PhD; Yang Yang, MS; Chunxia Wang, MS; Jing Liu, PhD; Xiubin Sun, PhD; Yi Liu, PhD; Shengxu Li, MD, PhD; Fuzhong Xue, 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=0.038). 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 1.58 (95% confidence interval, 1.23–2.02) 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

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

Event (death etc.) → **Not necessary**

LCMM can operate with OHDSI ecosystem!

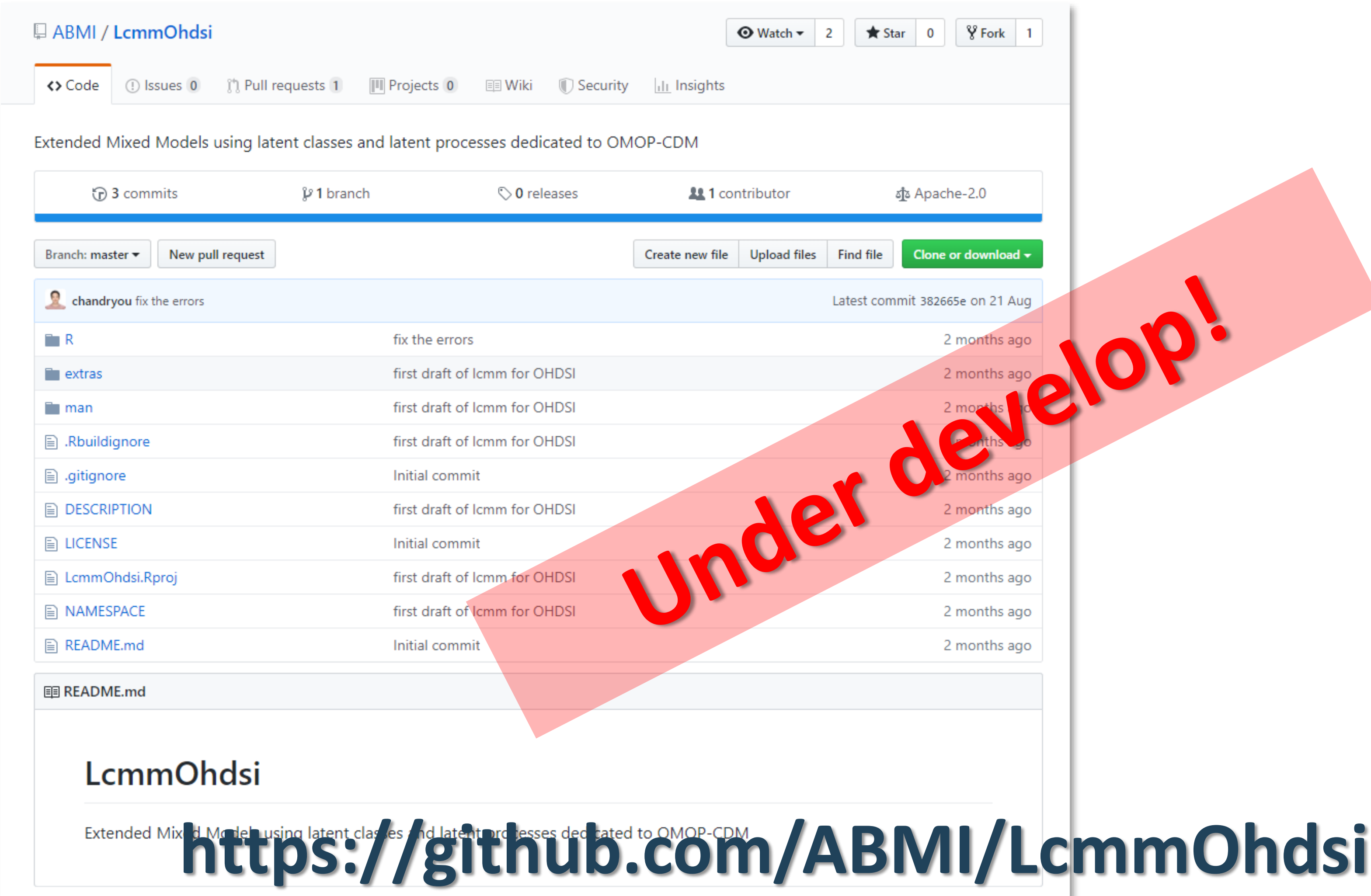
LCMM in OHDSI ecosystem

- 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

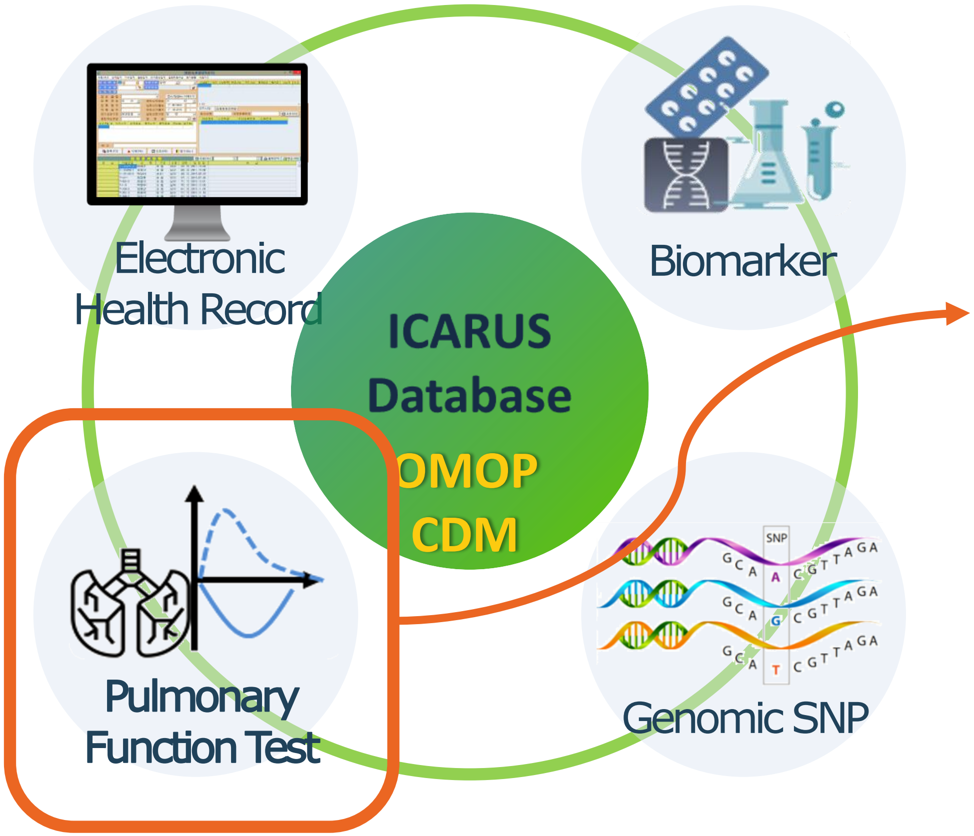


https://github.com/ABMI/LcmmOhdsi

Trajectory clustering to asthmatics

ICARUS Project

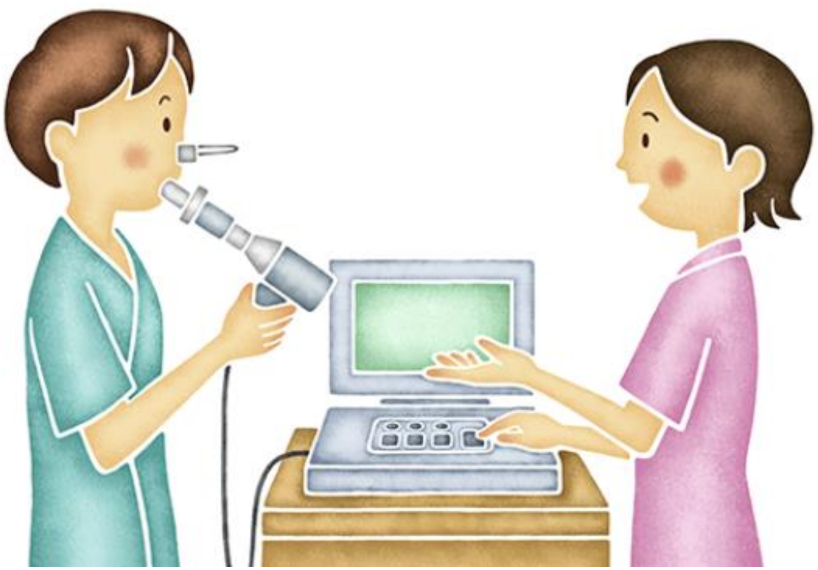
(Immune/Inflammatory Disease CDM Augmentation for Research Union System)

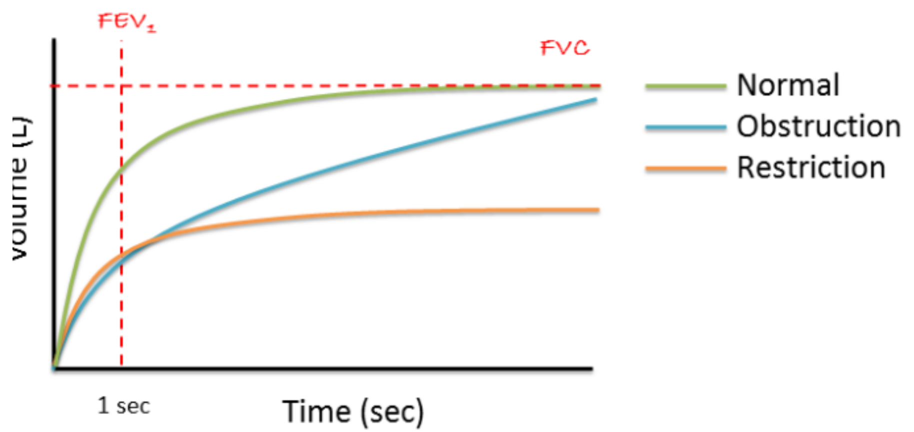


Lung function test data over 15 years were converted into OMOP CDM

Trajectory clustering to asthmatics

- **Forced expiratory volume in one second (FEV_1)** is mainly used as a lung function index when rating respiratory disorders





ıre 1: Representative volume-time curves for healthy (normal), obstructed (e.g. in asthma) and ricted (e.g. in fibrosis) lungs.

Trajectory clustering to asthmatics

- **Study population : Severe asthmatics**

ICARUS Database
N = 4,893

↓

Severe asthmatics
N = 1,388

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

↓

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

ATLAS

[11] Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. European respiratory journal. 2014;43(2):343-73.

Trajectory clustering to asthmatics

- **Trajectory clustering** using LCMM R package
  - **Measured variable** : **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 %

Trajectory clustering to asthmatics

- **Outcome : Asthma exacerbation**
  - Definition of asthma exacerbation :
    1. **Oral corticosteroid prescription for ≧ 3days**
       (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

	# of class	Shape of trajectory	AIC	BIC	%Class1	%Class2	%Class3	%Class4	%Class5
Model 1	2	Linear	73752.78	73803.25	6.26	93.74			
Model 2	2	Quadratic	73741.75	73802.32	4.61	95.39			
Model 3	2	Cubic	73711.28	73781.94	4.43	95.57			
Model 4	**3**	**Linear**	**73749.14**	**73819.80**	**12.61**	**34.75**	**52.64**		
Model 5	3	Quadratic	73696.37	73782.18	2.61	4.61	92.78		
Model 6	3	Cubic	73634.20	73735.15	3.57	3.83	92.61		
Model 7	4	Linear	73743.37	73834.23	31.57	9.30	7.83	51.30	
Model 8	4	Quadratic	73679.05	73790.10	2.17	2.09	2.52	93.22	
Model 9	4	Cubic	73623.74	73754.97	3.65	0.61	3.65	92.09	
Model 10	5	Linear	73739.72	73850.76	10.17	24.00	52.52	5.04	8.26
Model 11	5	Quadratic	73684.59	73820.88	1.83	3.04	9.22	63.65	22.26
Model 12	5	Cubic	73595.67	73757.19	3.13	0.70	5.74	0.96	89.48

The model which was 'linear' and '3 classes' was chosen

Result : Estimated trajectory of lung function from LCMM

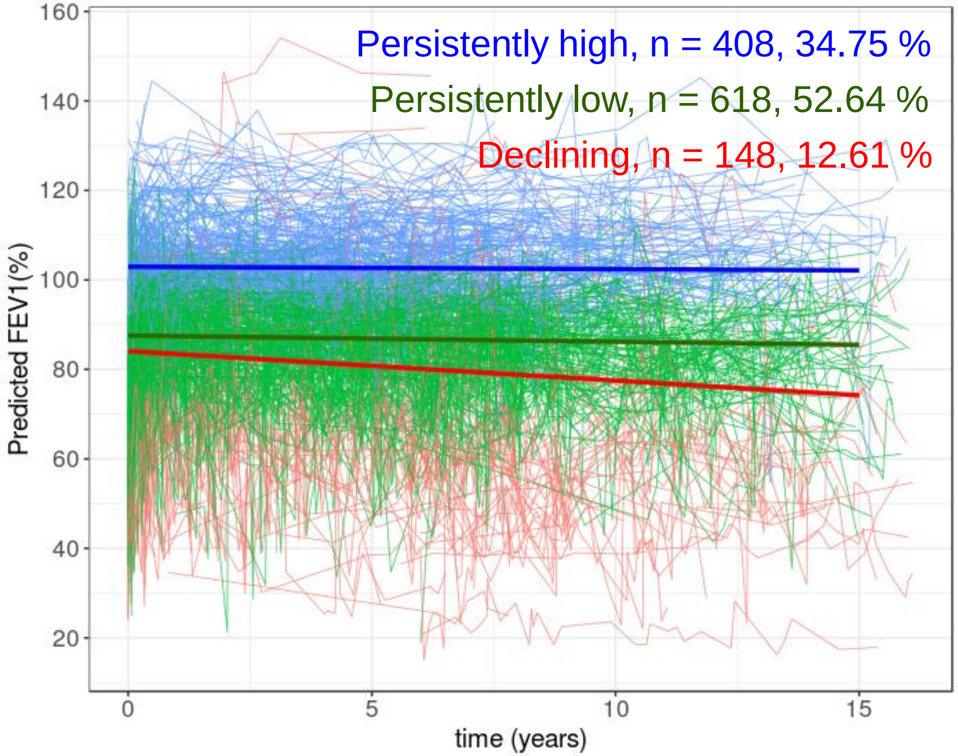


Figure 1-a. observed individual lung function trajectories and three estimated representative trajectories

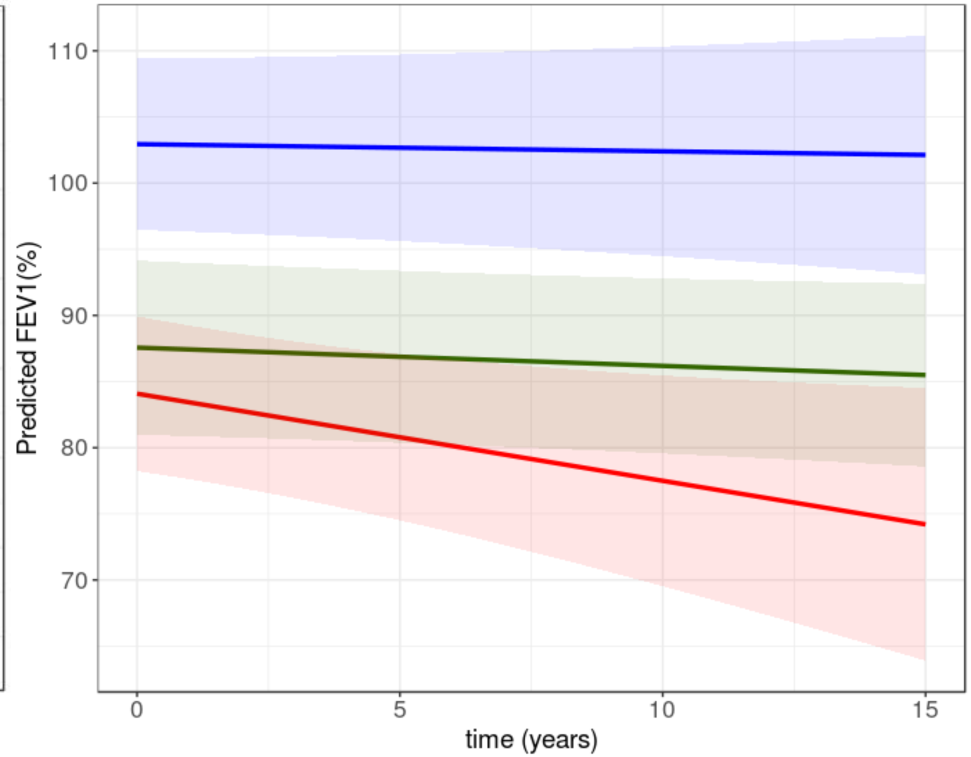
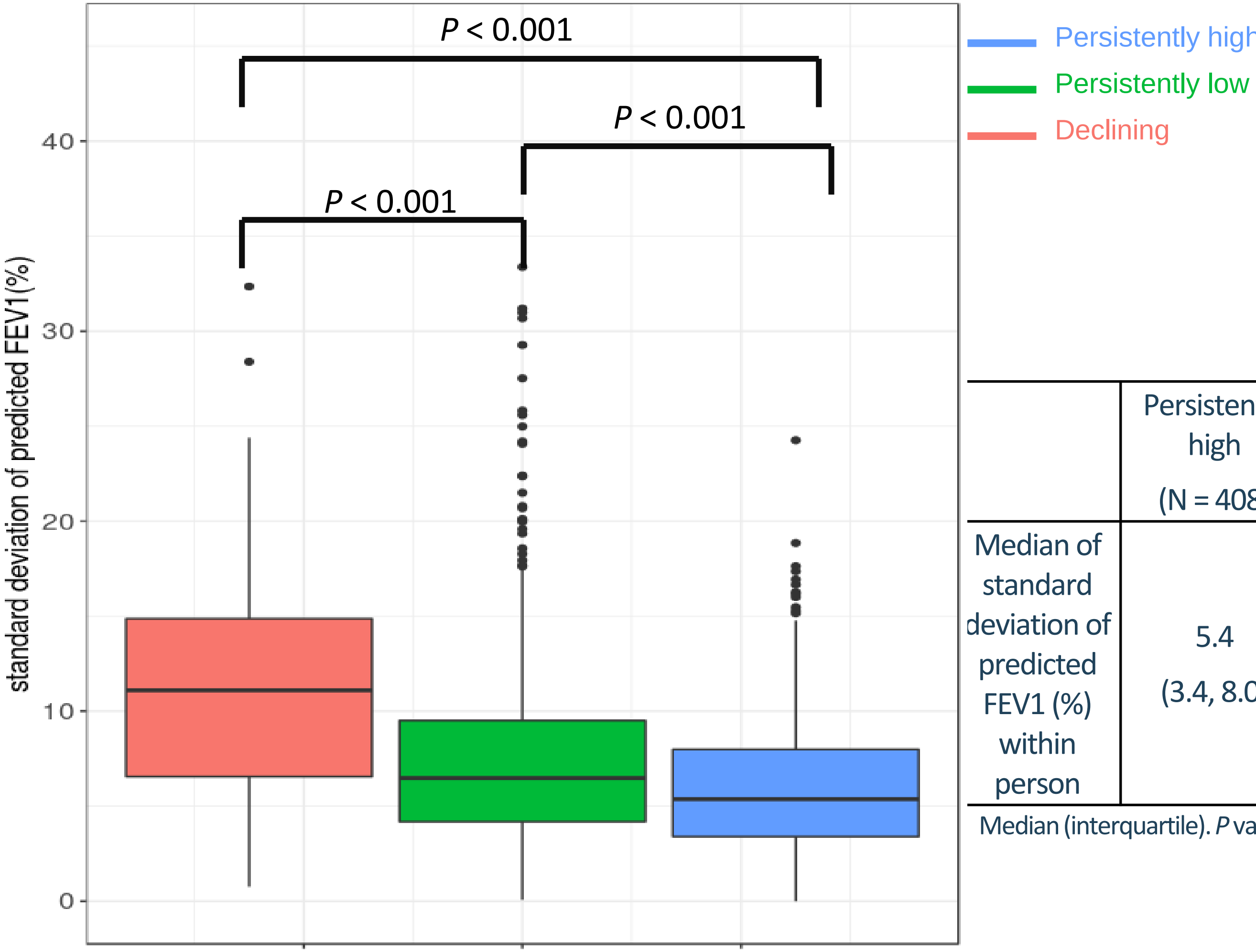


Figure 1-b. Tree estimated representative trajectories. The shaded areas indicate estimated 95% confidential intervals

	Persistently high (N = 408)	Persistently low (N = 618)	Declining (N = 148)
Intercept (95% CI)	102.947 (96.47, 109.42)	87.57 (80.98, 94.15)	84.08 (78.23, 89.93)
Slope (95% CI)	-0.05 (-0.38, 0.27)	-0.14 (-0.36, 0.09)	**-0.66 (-1.27, -0.05)**

Result : variability of lung function



	Persistently high (N = 408)	Persistently low (N = 618)	Declining (N = 148)	*P* value
Median of standard deviation of predicted FEV1 (%) within person	5.4 (3.4, 8.0)	6.5 (4.2, 9.5)	**11.1 (6.6, 14.9)**	<0.001

Median (interquartile). *P* value was calculated according to Kruskal-wallis test

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

Result : Baseline characteristics

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

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

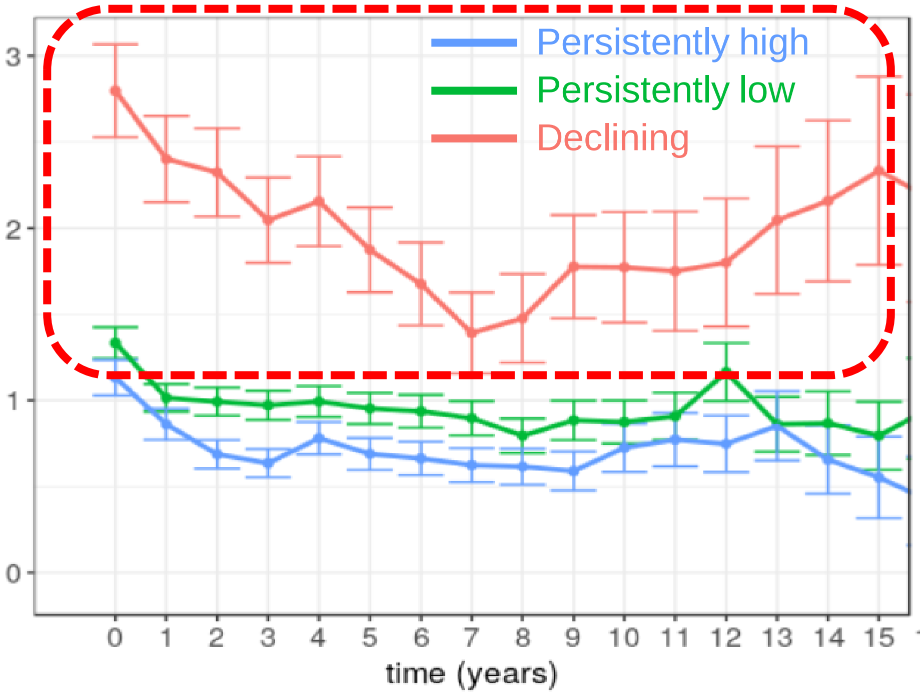


Figure 3-(a) Yearly count of **asthma exacerbation requiring oral corticosteroid**

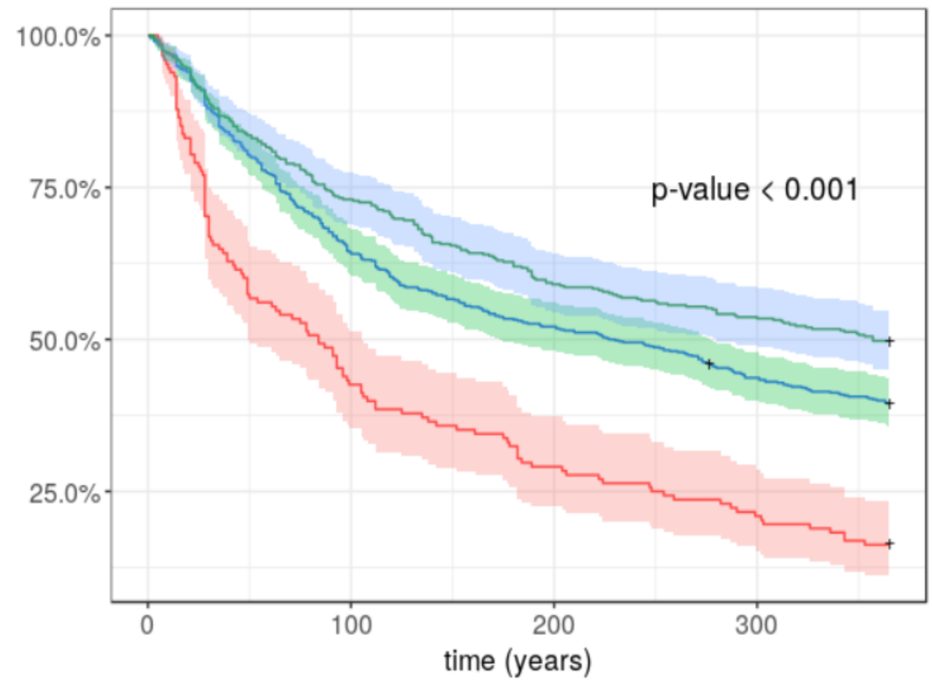


Figure 3-(b) Kaplan-Meier survival curve for the time to the **asthma exacerbation requiring oral corticosteroid**

Result : Yearly counts of event cohort and 1 year event free survival

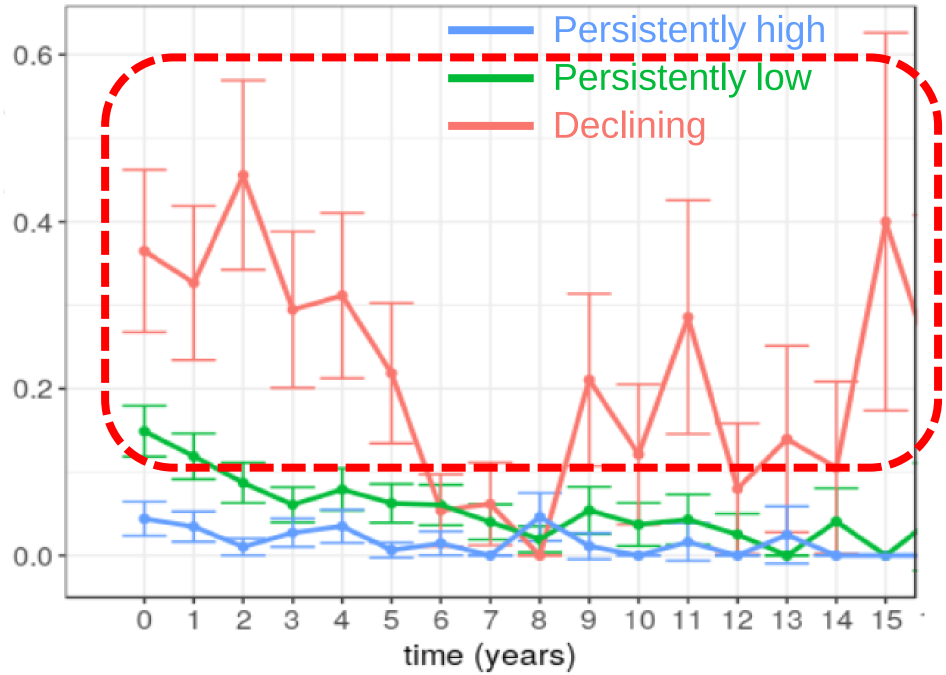


Figure 3-(c) Yearly count of **emergency room visit** due to asthma exacerbation

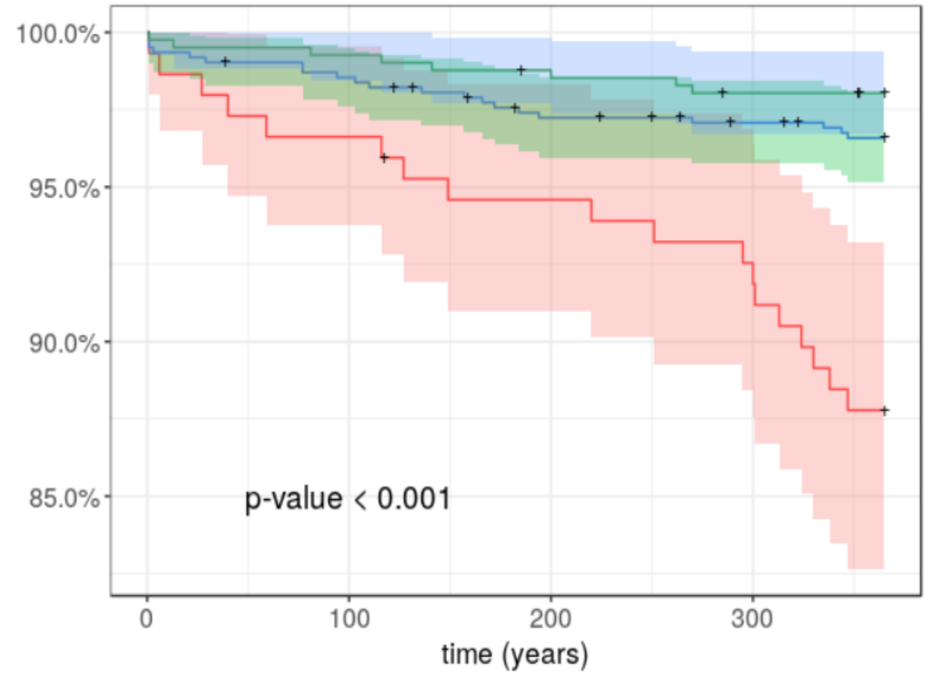
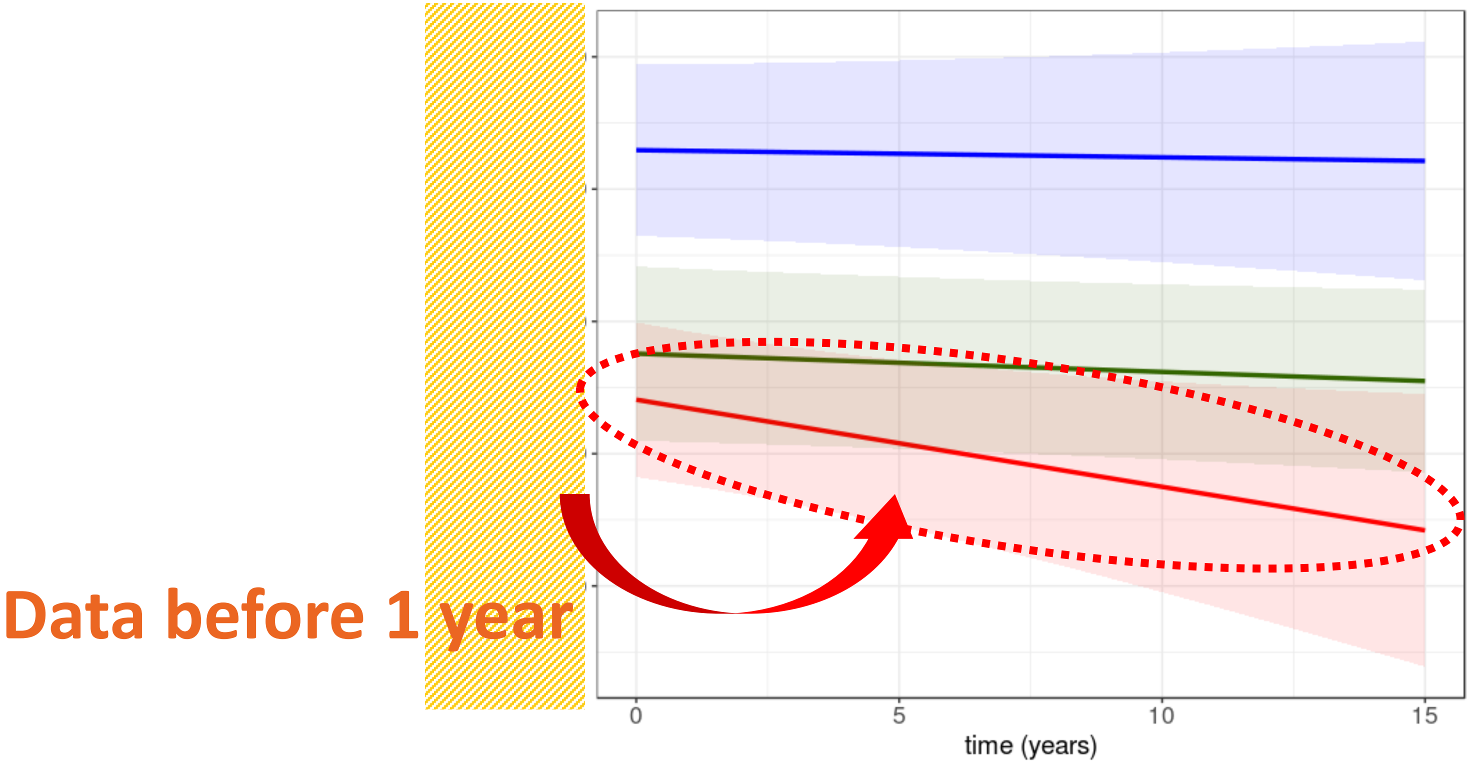


Figure 3-(d) Kaplan-Meier survival curve for t he time to **emergency room visit** due to asth ma exacerbation

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

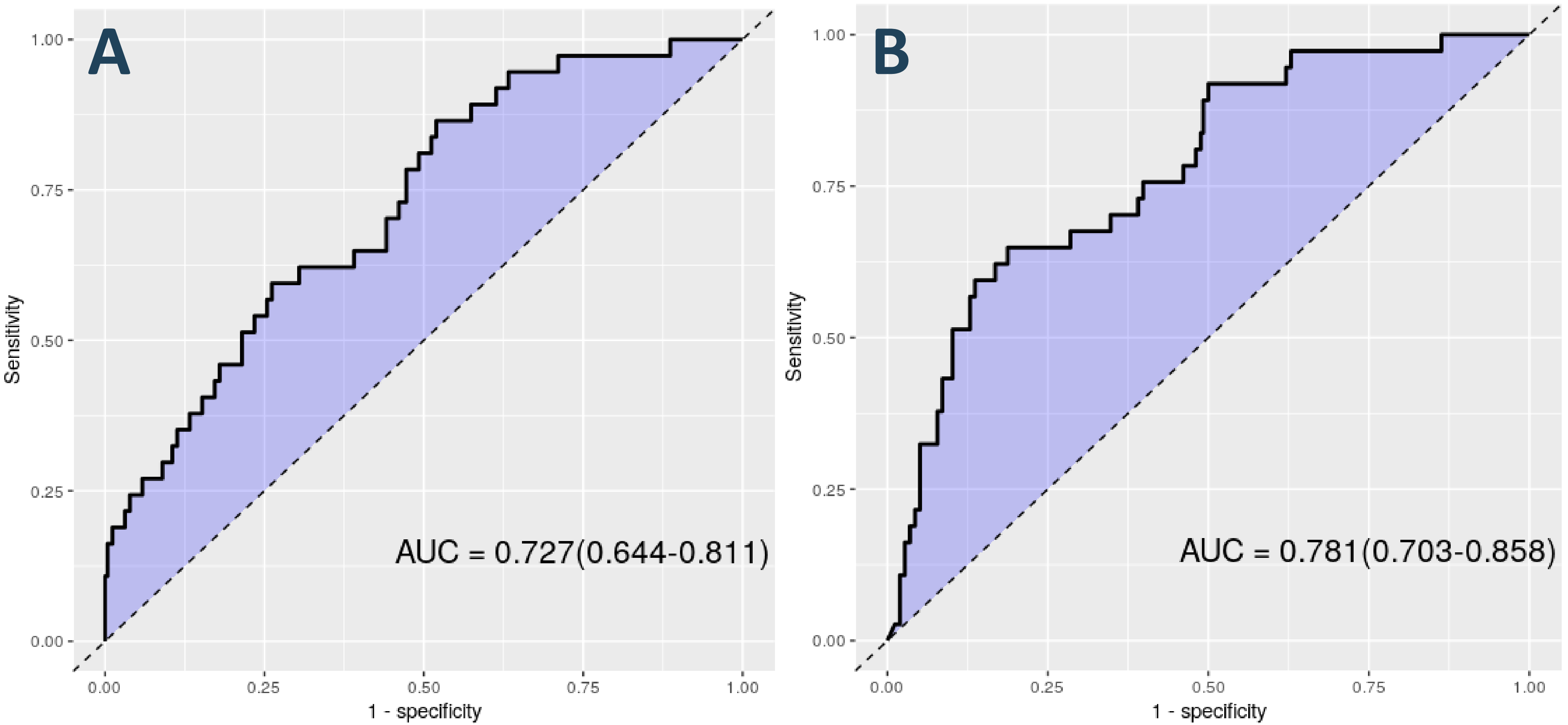
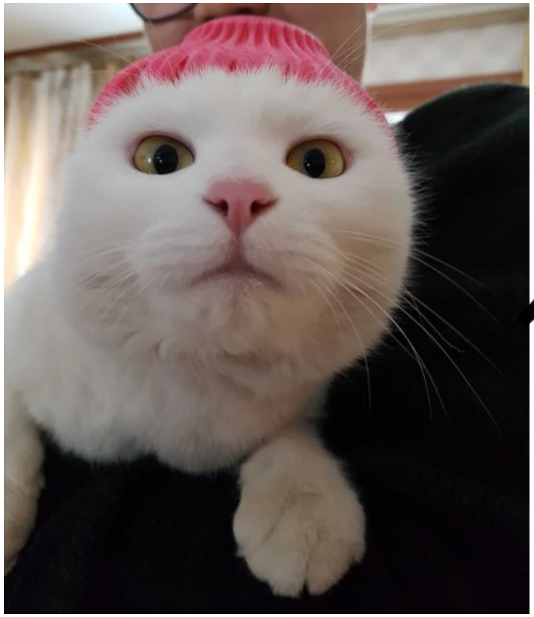
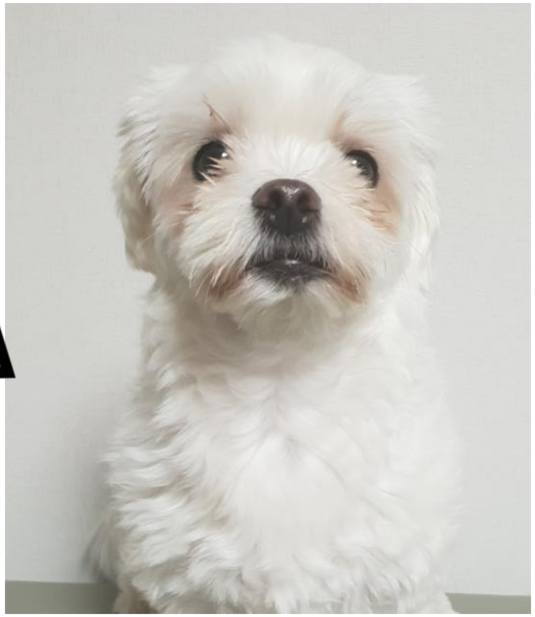


Figure 4. **The AUROC curves** for the declining group predicting model.
A : lasso logistic regression, B : gradient boosting model

Extension of PatientLevelPrediction

Apply model identifying cat for identifying dog



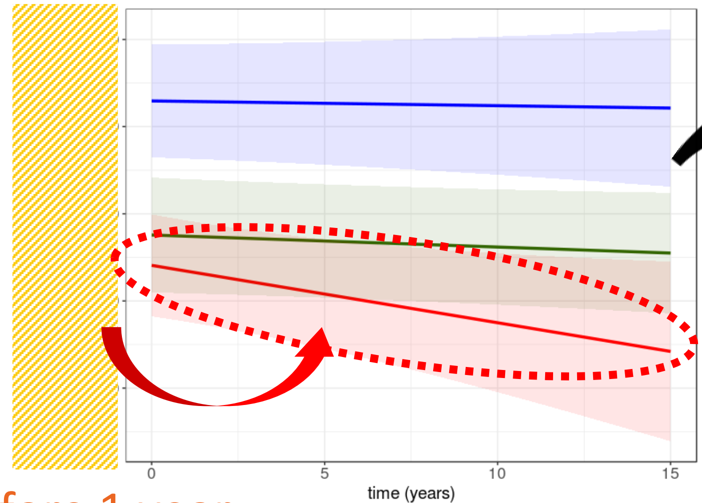


Is it cat? **Is it dog?**

Through the PatientLevelPrediction package, we can do similar analysis

Extension of PatientLevelPrediction

Apply model predicting declining group for predicting asthma exacerbation



Data before 1 year

Data before 1 year

Who will be include the declining group?

Who will be visit emergency room for asthma within 365 days from index date

Result : patient level prediction

	Target cohort	Outcome cohort	Machine learning model	AUROC
AUMC	8873	303	Lasso logistic	0.786
			Gradient boosting	0.642

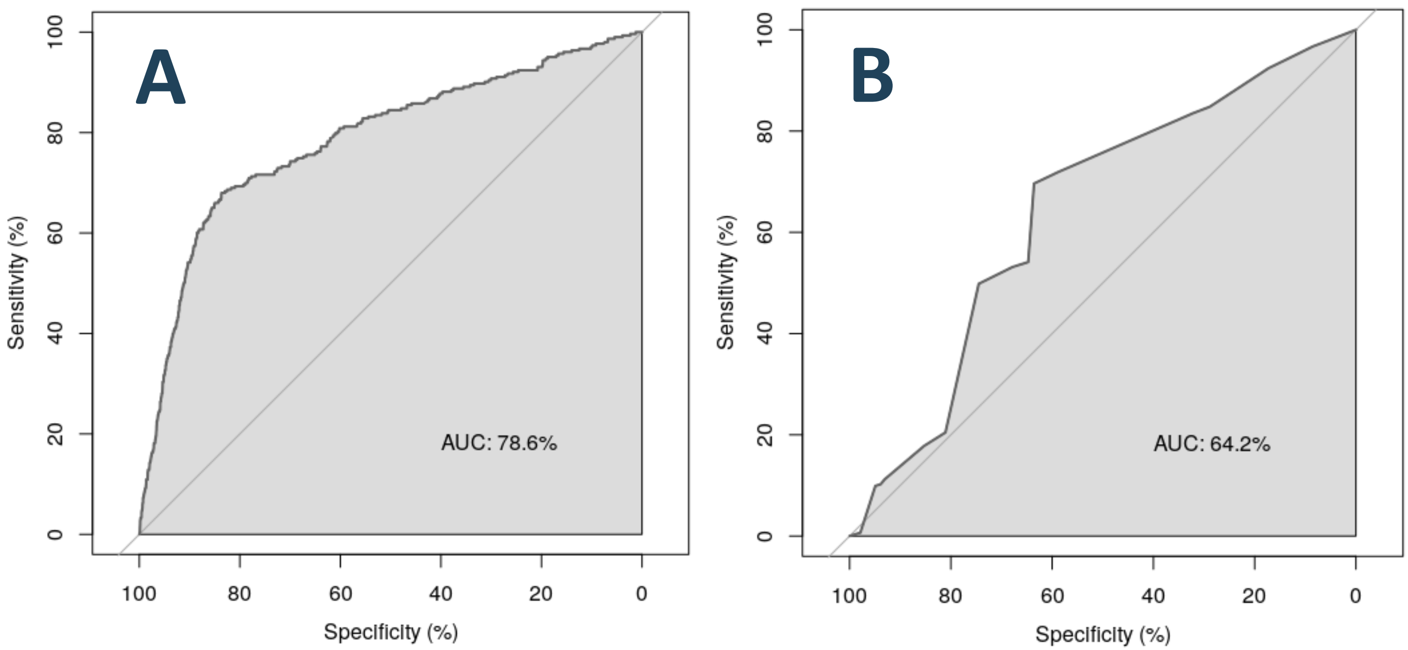


Figure 5. The AUROC curves of AUMC validation groups. A : lasso logistic regression, B : gradient boosting model

Result : patient level prediction

	Target cohort	Outcome cohort	Machine learning model	AUROC
NHIS-NSC	17240	771	Lasso logistic	0.685
			Gradient boosting	0.541

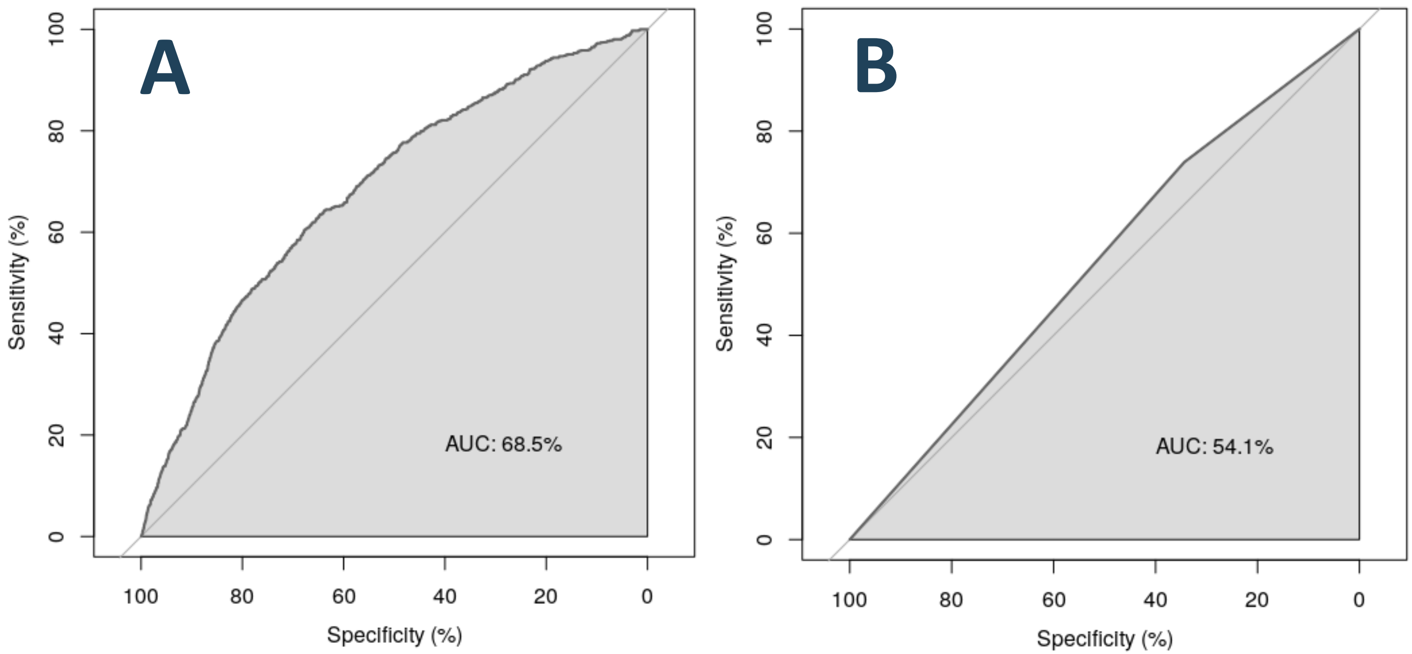


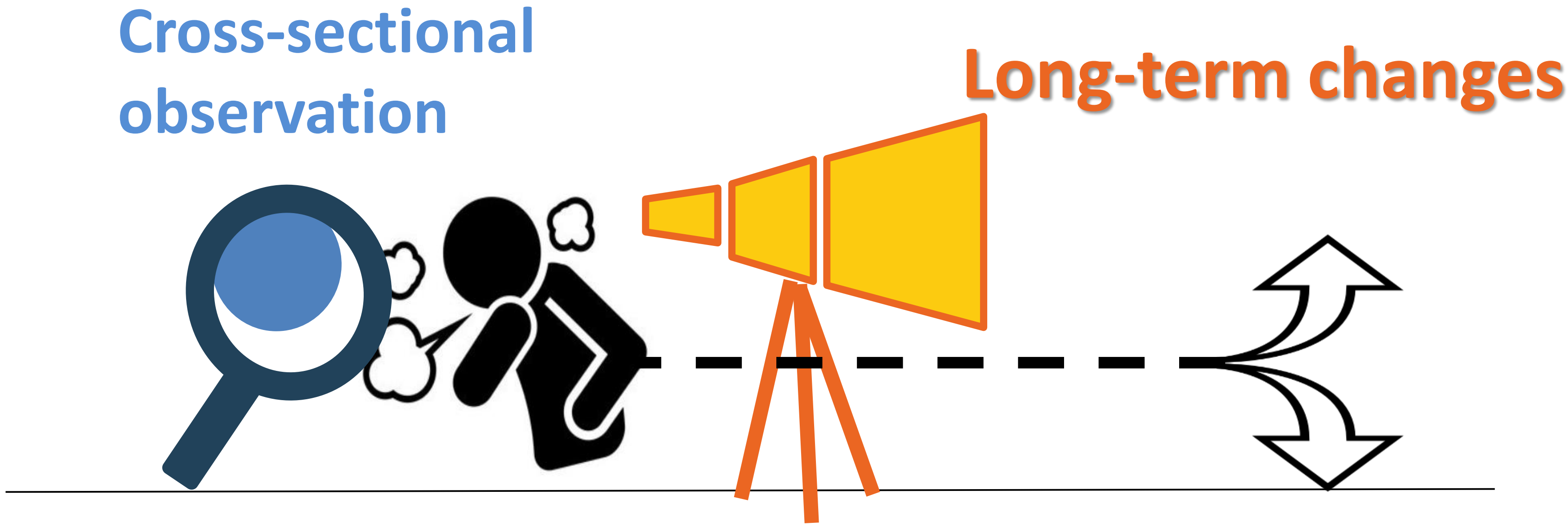
Figure 6. The AUROC curves of NHIS-NSC validation groups. A : lasso logistic regression, B : gradient boosting model

Conclusion

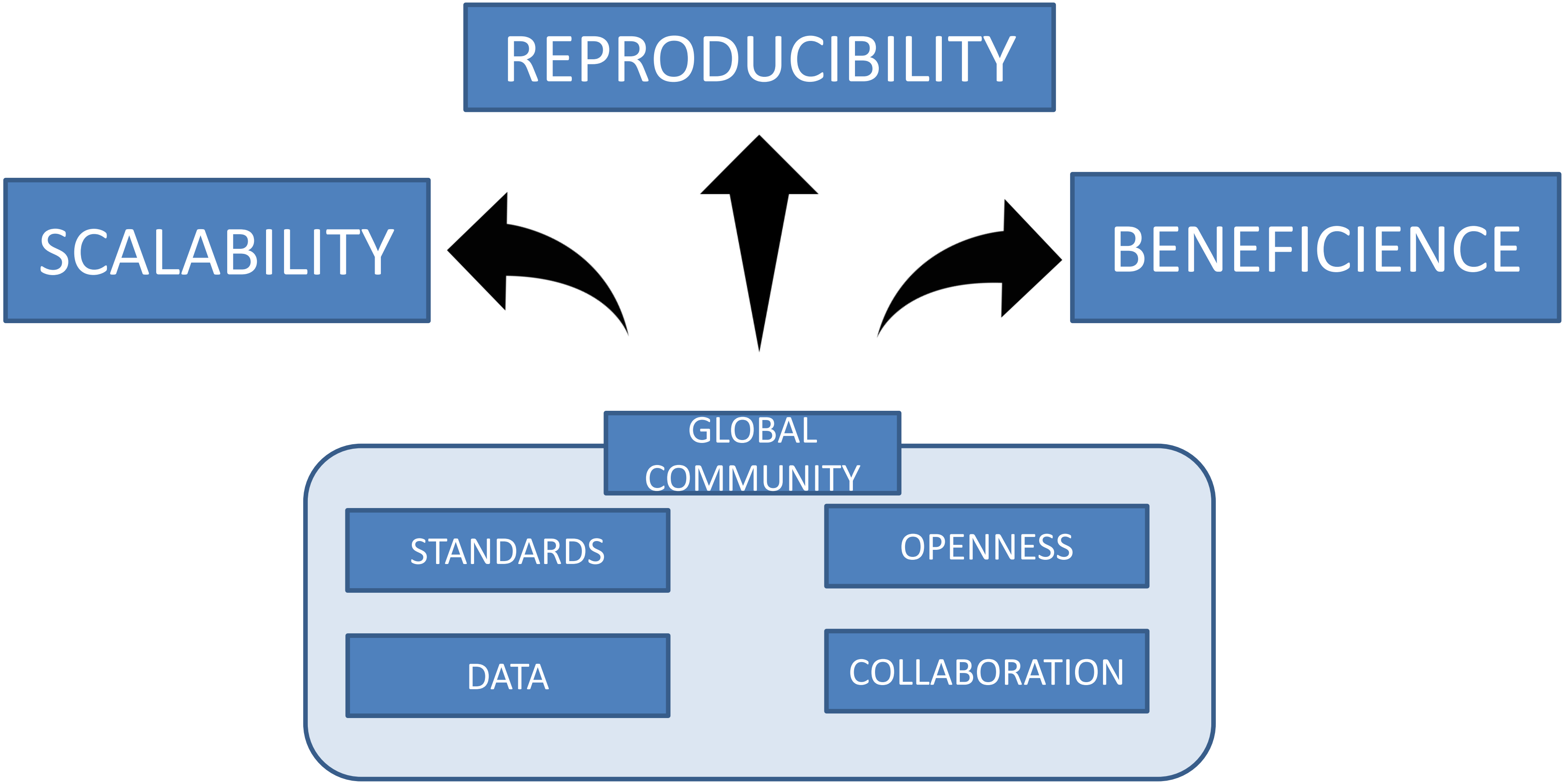
- **LCMM** can operate **with OHDSI ecosystem**
- Through the LCMM and OHDSI tools, **severe asthmatics can be classified** according to long-term change of lung function
- For prediction declining lung function, the **PatientLevelPrediction can be used**

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



Mission, Vision, and Values of OHDSI

- 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].