Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Thirty-day hospital readmission prediction model based on common data model with weather and air quality data

Mr. Fookty Phengphakeo, MSc student
16 May 2025

Journal club
Commentator

Outline

- Introduction
- Objective
- Data sources
- Methods, model develop & study design
- Weather and air quality predictors in w-score
- Results
- Summary

Introduction

- Hospital **readmission** within 30 days is a key indicator of quality of care and a significant cost burden
- Predicting unplanned readmission is difficult and hospital readmission shortly after discharge indicates potential ***issues with the quality of care;***
- Existing ML models often lack sufficient verification and multi-center applicability;
- This study develops a prediction model that is
  - Based on OMOP CDM using standardized clinical data combined with weather and air quality factors.
  - weather *forecast* data for real-time prediction at discharge

Objective

- To develop and validate a model predicting unplanned hospital readmission within 30 days of discharge;
- To assess model transportability using **external** validation
- To investigate the impact of weather and air quality factors.

Data sources

- **Data**
  - Retrospective EHR data from two hospitals (SNUH, SNUBH) in Seoul, South Korea (2017-2018)
  - Using OMOP-CDM converted EHR data
  - Weather and air quality data from KMA and Air Korea (matched by postal code).
- **Population:** Adults >65 years discharged alive, with hospital/ER visits during the study period, living in the Seoul metropolitan area;
- N=61,922 visits included; 5,794 resulted in 30-day readmission
- **External Validation Cohort:** SNUBH dataset

SNUH : the Seoul National University Hospital

SNUBH : the Seoul National University Bundang Hospital.

KMA: the Korea Meteorological Administration

Methods (w-score and features)

- **W-score concept**: a patient-level score based on weather and air quality data for their residence locality;
- **W-score calculation:** derived using KMA's special weather report standards;
- **Clinical features:**
  - gender, age,
  - diagnosis conditions (CONDITION_ERA),
  - drug exposures (DRUG_ERA),
  - Charlson comorbidity index

W-score: weather and air quality scores for individual visits
KMA: Korea Meteorological Administration's standards

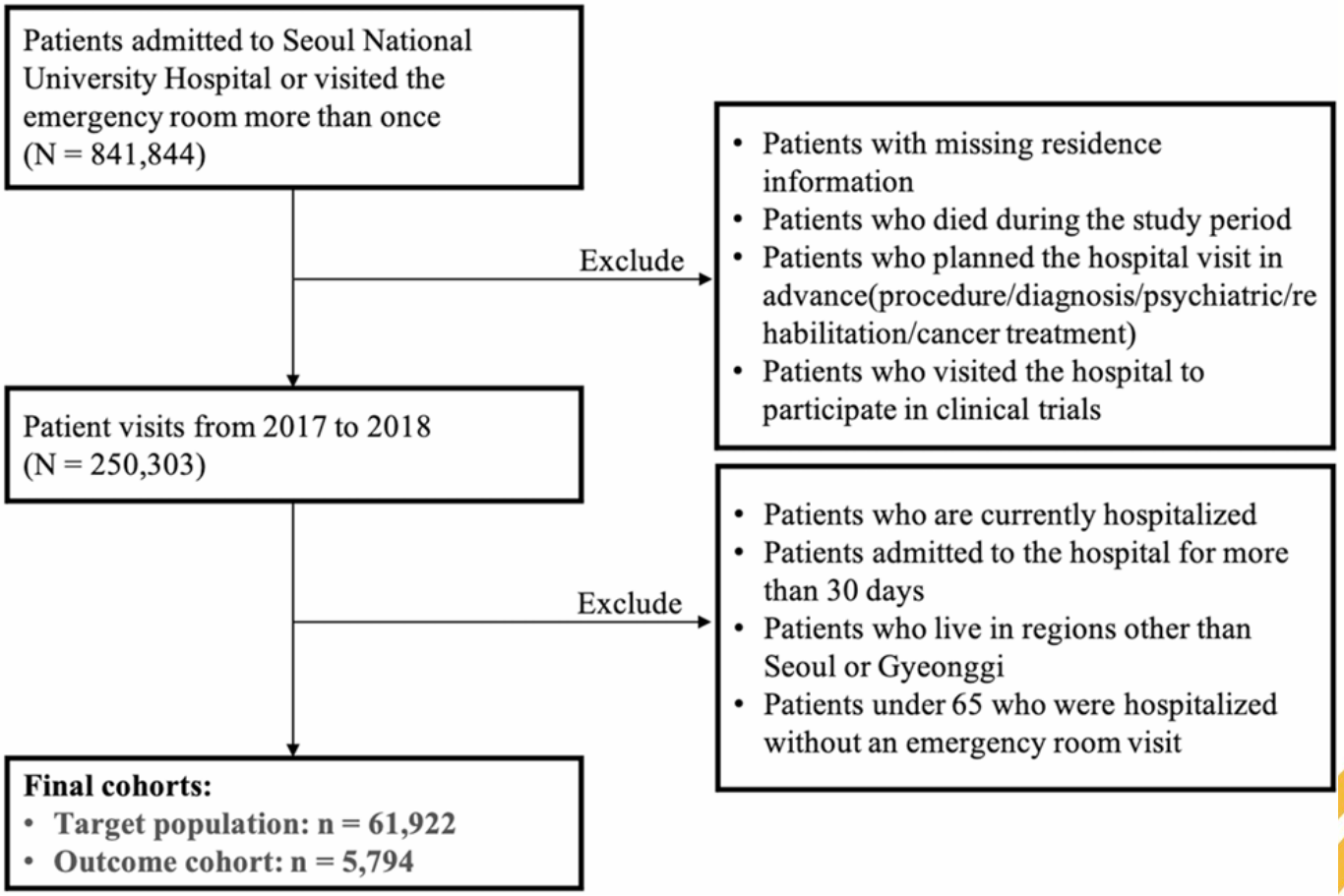
Methods (w-score and features)

- The W-score is calculated by **summing up weather and air quality forecast values for the 7 days** after a patient is discharged;
- **W-score based on forecast data** using forecast data (7 days from discharge) for the W-score is a practical design choice for potential **real-time clinical use** at the point of discharge;
- These values are determined based on the KMA **standards** for special weather reports.

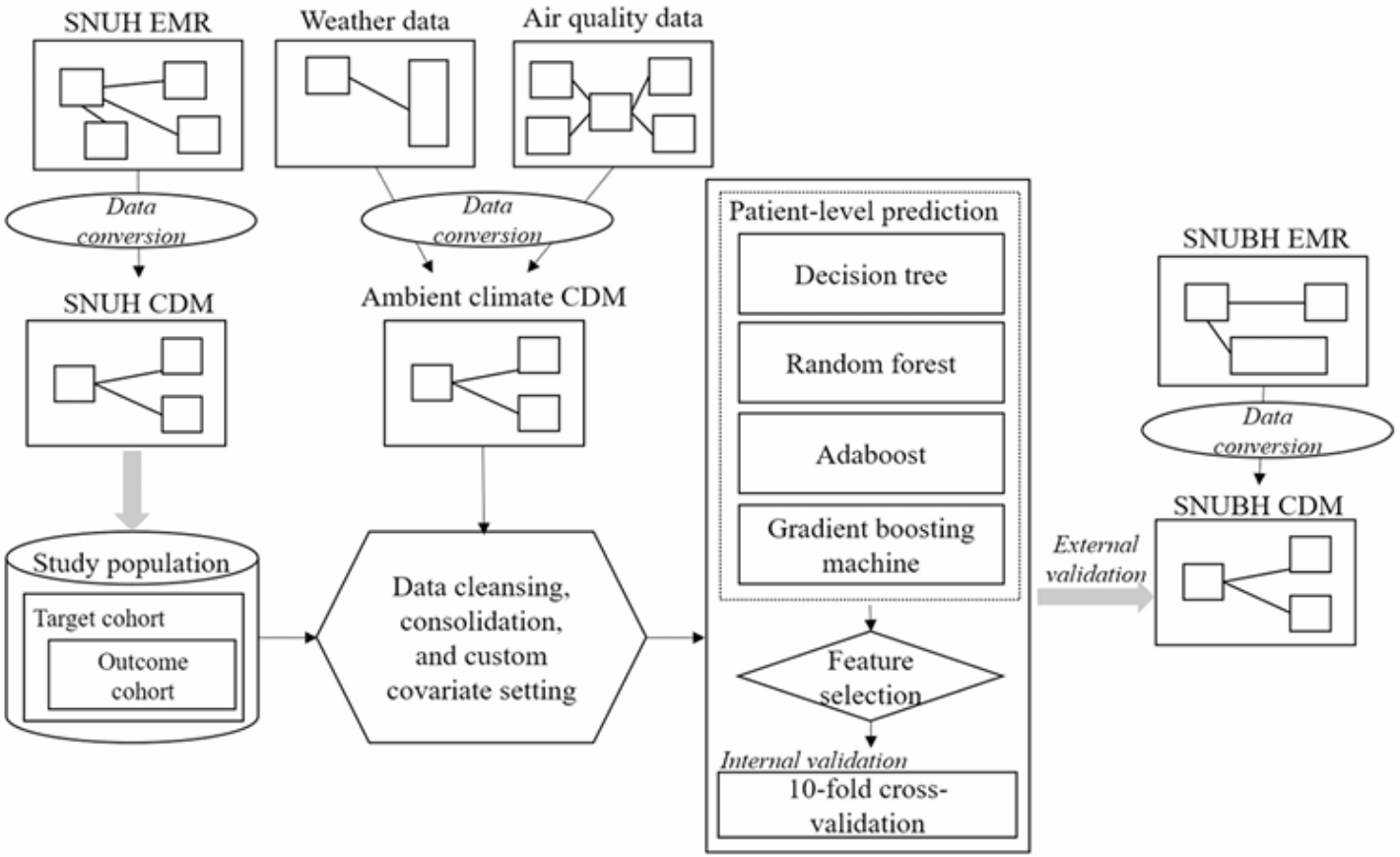
Model development & validation

- Four tree-based ML models compared:
  - Decision Tree;
  - Random Forest;
  - AdaBoost;
  - Gradient Boosting Machine (GBM).
- Internal validation 10-fold cross-validation (internal).
- **External validation** on separate hospital data (SNUBH).
- Conducting **external validation** on a separate dataset from a different hospital (even within the same university system) is a crucial step to assess model transportability and is a key strength compared to studies relying only on internal validation
- Performance metric AUC reported.

Study design



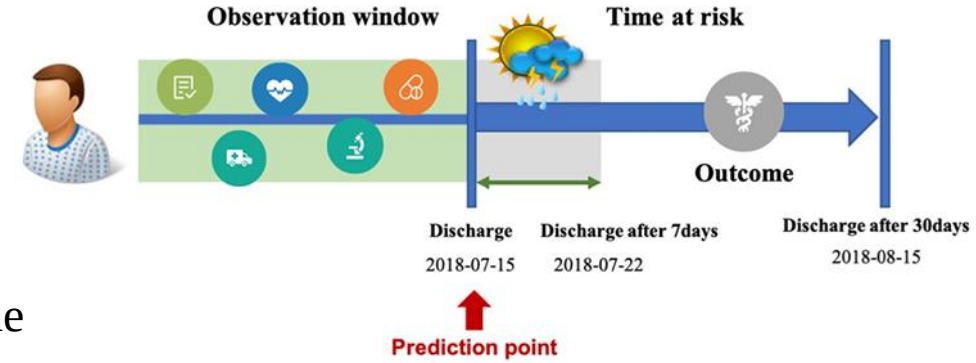
Overall methodology of the study



Model development (Prediction window)

The prediction model for readmission within 30 days was developed to reflect variables such as

- Clinical diagnosis
- Drug prescription prior to patient discharge date as well as to predict the occurrence of readmission of the patient by considering the W-score for the weather forecast at the patient's residence location after the discharge date



Basic characteristics of study data for each visit type

Characteristics		Derived cohorts		P value
		Readmitted (N = 5794)	Non-readmitted (N = 56,128)	
Age, year, mean (SD)		75.2 (6.8)	74.7 (6.7)	0.000
Gender	Male, n (%)	54.8	49.7	
	Female, n (%)	45.2	50.3	
Age during hospital visit	60s	23.8	26.0	
	70s	49.7	50.2	
	80s	23.7	21.4	
	90s	2.7	2.4	
Season during admission	Spring	25.4	24.1	0.049
	Summer	25.5	26.8	
	Fall	24.4	24.8	
	Winter	24.6	24.3	
Average length of stay, mean (SD)		2.5 (4.3)	0.2 (0.4)	
Charlson comorbidity index, mean		1.11	0.52	

Number of visits in each disease group and outcome incidence rate in our research cohorts

Disease groups	Train/test population (internal)		Valid population (external)	
	Target size (N)	% incidence	Target size (N)	% incidence
Diseases of the circulatory system (I00–I99)	9357	14.0	87,063	10.3
Mental and behavioral disorders (F00–F99)	3174	16.3	7228	17.5
Diseases of the musculoskeletal system and connective tissue (M00–M99)	13,564	11.8	41,015	11.7
Diseases of the respiratory system (J00–J99)	10,310	15.7	87,604	15.1

- Analyzes performance and feature importance across different disease subgroups

Weather and air quality predictors in w-score

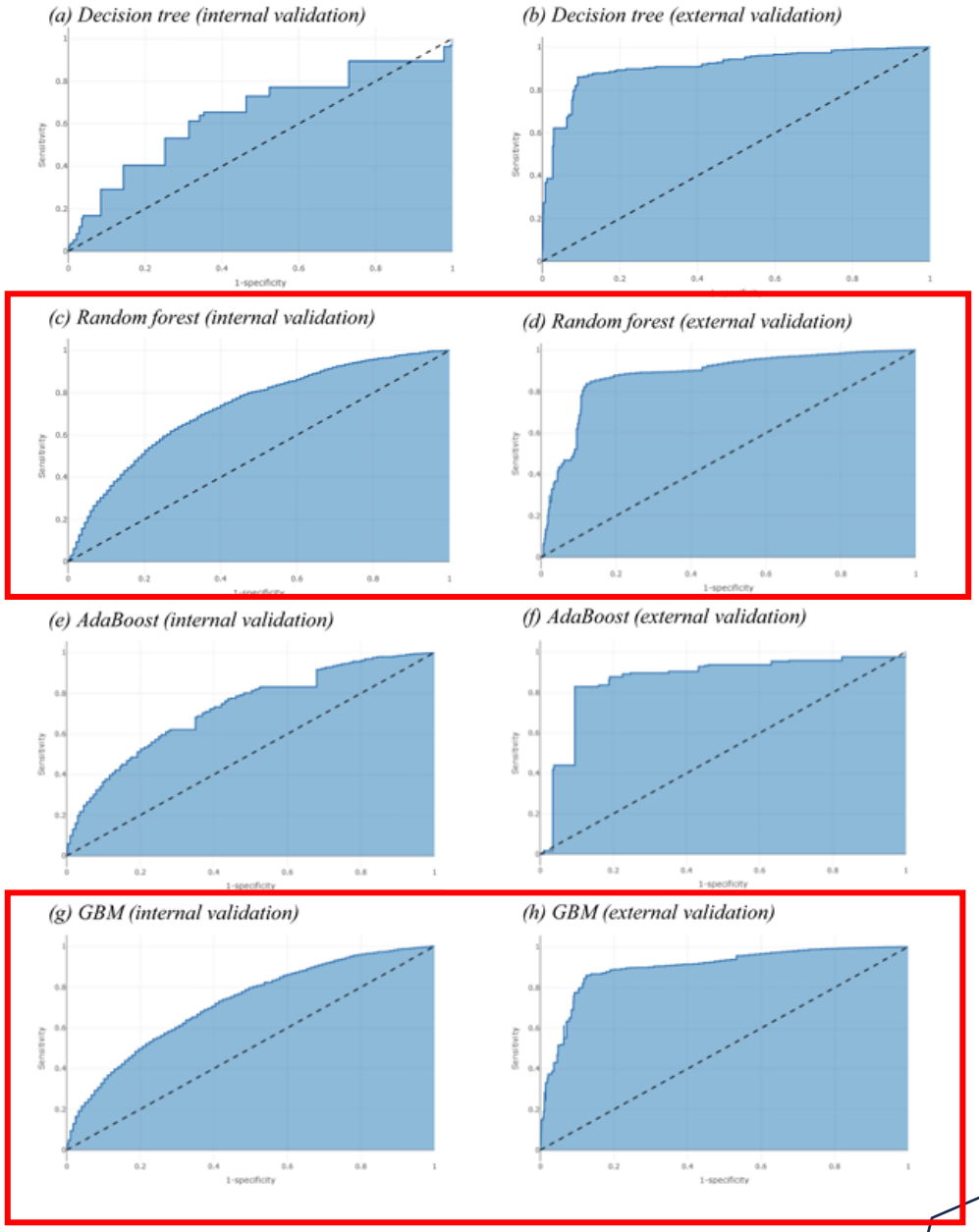
Disease groups	covariateName	covariateValue	CovariateMean	
			CovariateMean WithOutcome	CovariateMean WithNoOutcome
Diseases of the circulatory system	PM10	0.0016	12.59	13.13
	Rainfall	0.0011	2.22	2.32
	Humidity	0.0005	0.29	0.28
	Min Temperature	0.0006	0.65	0.59
	Max Temperature	0.0005	0.85	0.83
Mental and behavioral disorders	PM10	0.0016	12.36	13.24
	Rainfall	0.0012	1.95	2.30
	Humidity	0.0014	0.41	0.31
	Min Temperature	0.0008	0.67	0.58
	Max Temperature	0.0004	0.71	0.81
Diseases of the musculoskeletal system and connective tissue	PM10	0.0015	12.85	13.18
	Rainfall	0.0012	2.24	2.32
	Humidity	0.0008	0.31	0.30
	Min Temperature	0.0007	0.63	0.55
	Max Temperature	0.0007	0.86	0.84
Diseases of the respiratory system	PM10	0.0038	12.86	13.01
	Rainfall	0.0032	2.30	2.29
	Humidity	0.0005	0.31	0.29
	Min Temperature	0.0012	0.60	0.54
	Max Temperature	0.0036	1.01	0.97

The parameter values in each model

Models	Parameters	Values	Parameter mean
DT	classWeight		"Balance" or "None"
	maxDepth	10	The maximum depth of the tree
	minImpuritySplit	10^{-7}	Threshold for early stopping in tree growth. A node will split if its impurity is above the threshold, otherwise it is a leaf
	minSamplesLeaf	10	The minimum number of samples per leaf
	minSamplesSplit	2	The minimum samples per split
RF	Max depth	4, 10, 17	Max levels in a tree
	mtries	−1 = square root of total features, 5, 20	Number of features in each tree
	ntrees	500	Number of trees
ADA	Learning rate	1	Learning rate shrinks the contribution of each classifier by learning_rate. There is a trade-off between learningRate and nEstimators
	n estimators	4	The maximum number of estimators at which boosting is terminated
GBM	Learning rate	0.005, 0.01, 0.1	The boosting learn rate
	earlyStopRound	25	Stopping after rounds without improvement
	Max depth	4, 6, 17	Max levels in a tree
	minRows	2	Min data points in a node
	ntrees	100, 1000	Number of trees

ROC curves for the validation of the Adaboost and decision tree models.

- **RF (c, d)** and **GBM (g, h)** show higher sensitivity and area under the curve (AUC), particularly external validation



Comparison of disease-specific performance in each model based on the area under the receiver operating characteristic curve.

Disease groups	Prediction models	Internal validation		External validation	
		Clinical covariates only	Clinical covariates and W-scores	Clinical covariates only	Clinical covariates and W-scores
Diseases of the circulatory system	DT	0.653	0.674	0.664	0.679
	RF	0.693	0.686	0.688	0.681
	ADA	0.698	0.708	0.672	0.67
	GBM	0.726[a]	0.717[a]	0.704[a]	0.696[a]
Mental and behavioral disorders	DT	0.612	0.691	0.706	0.737
	RF	0.703	0.692[a]	0.743	0.686[a]
	ADA	0.716	0.654	0.747	0.728
	GBM	0.747[a]	0.676	0.751[a]	0.727
Diseases of the musculoskeletal system and connective tissue	DT	0.68	0.690[b]	0.856	0.889[b]
	RF	0.719	0.734	0.909	0.882
	ADA	0.726[b]	0.739[a]	0.917[b]	0.915[a]
	GBM	0.751[a]	0.725	0.883[a]	0.9
Diseases of the respiratory system	DT	0.634	0.607	0.651	0.622
	RF	0.653	0.643	0.658	0.638
	ADA	0.663	0.667	0.639	0.655
	GBM	0.672[a]	0.675[a]	0.669[a]	0.667[a]

External validation: Confirmed model by the strong musculoskeletal results (GBM: 0.883 clinical, 0.9 clinical + W-score).

- GBM showed the highest AUC performance of **75.1** in the clinical model
- Clinical and W-score model showed the best performance of **73.9** for musculoskeletal diseases

[a] Best performances for each disease
[b] Major improvements in external validation

Summary

- External validation showed transportability, especially for musculoskeletal diseases
- OMOP CMD facilitates multi-center expansion and standardization
- Weather and air quality (PM10, rainfall, max temp) were identified and support real-time, data-driven clinical decision-making across different hospital systems relevant predictors
- A predictive model that combines weather and environmental data with a patient's residence information is expected to enhance clinical decision making at the individual patient level.

Supplementary

C6 | 0.0209239350658136

ng machine model with clinical covariates and W-score. (Mental and behavioral disorders)

Train/Test by SNUH)			External (Valid by SNUBH)			
covariateValue	CovariateMeanWithNoOutcome	CovariateMeanWithOutcome	covariateName	covariateValue	CovariateMeanWithNoOutcome	CovariateMeanWithOutcome
0.026181862	1.394779003	5.822451855	visit_occurrence concept count during day -365 through -1 concept_count relative to index: Emergency Room Visit	0.026181862	2.160348876	35.74407583
0.020923935	0.913674489	3.820103335	visit_occurrence concept count during day -180 through -1 concept_count relative to index: Emergency Room Visit	0.020923935	1.277088225	15.50157978
0.013706259	0.385166337	1.124941287	visit_occurrence concept count during day -30 through -1 concept_count relative to index: Emergency Room Visit	0.013706259	0.492787655	3.5
0.009223429	35.90066912	63.68905589	visit_occurrence concept count during day -365 through -1 concept_count relative to index: Outpatient Visit	0.009223429	13.3302583	19.75829384
0.008816443	18.9411931	33.79614843	visit_occurrence concept count during day -180 through -1 concept_count relative to index: Outpatient Visit	0.008816443	7.317678631	10.31990521
0.00813775	3.865893884	6.62987318	visit_occurrence concept count during day -30 through -1 concept_count relative to index: Outpatient Visit	0.00813775	1.539248574	2.030015798
0.006072039	0.005277542	0.053546266	drug_era distinct concept count during day -365 through -1 concept_count relative to index	0.005818327	22.90992955	32.0450237
0.005818327	24.76618603	32.8506341	visit_occurrence concept count during day -30 through -1 concept_count relative to index: Inpatient Visit	0.005688532	0.496142234	0.355450237
0.005688532	0.613985487	0.44480977	drug_era distinct concept count during day -180 through -1 concept_count relative to index	0.005583638	18.55820195	24.40916272
0.005583638	19.5123928	26.13668389	drug_era distinct concept count during day -30 through -1 concept_count relative to index	0.004946665	13.02331432	14.89020537
0.005536614	0.001884836	0.048849225	drug_exposure during day -180 through -1 days relative to index: 50 ML sodium chloride 9 MG/ML Injection	0.004905898	0.425863804	0.665876777
0.004946665	12.54377533	15.48144669	visit_occurrence concept count during day -180 through -1 concept_count relative to index: Inpatient Visit	0.00473511	0.864978195	1.084518167
0.004905898	0.411365564	0.623767027	condition_era distinct concept count during day -180 through -1 concept_count relative to index	0.004715981	6.810466286	10.04265403
0.00473511	1.300442937	1.109441052	visit_occurrence concept count during day -365 through -1 concept_count relative to index: Inpatient Visit	0.004644802	1.18215364	1.695892575

Supplementary Table S3 | Supplementary Table S4 | Supplementary Table S5 | Supplementary Table S6

Supplementary

Supplementary Table S1. Weather advisory and warning issuance standards in Korea

Category	Criteria for issuing an weather advisory	Criteria for issuing an weather warning	W-score processing rule in this study
Drying	When the effective humidity of 35% or less is expected to last for more than 2 days	When the effective humidity of 25% or less is expected to continue for more than 2 days	Days with humidity below 35%
Cold wave	In October-April, when one of the following is true: 1) When the morning minimum temperature falls by more than 10°C than the previous day and is 3°C or less and is expected to be 3°C lower than the normal value 2) The morning minimum temperature is -12°C or less for 2 days When it is expected to continue abnormally 3) When significant damage is expected due to a sudden low temperature phenomenon	In October-April, when one of the following is true: 1) When the morning minimum temperature falls by more than 15°C than the previous day and is 3°C or less and is expected to be 3°C lower than the average value 2) The morning minimum temperature is -15°C or less. When it is expected to last longer than a day 3) When significant damage is expected in a wide area due to a sudden low temperature phenomenon	1) The day when the temperature difference between the minimum temperature and the previous day fell by more than 9 °C 2) The minimum daily temperature is less than -10 °C
Heat wave	When the maximum daily temperature is 33°C or higher is expected to last for 2 days or longer	When the maximum daily temperature is expected to last for more than 2 days when the maximum temperature is 35°C or higher	Days when the maximum daily temperature is 33 °C or higher
Downpour	When 3 hours of rainfall is expected to be more than 60mm or 12 hours of rainfall is expected to be more than 110mm	When 3 hours of rainfall is expected more than 90mm, or 12 hours of rainfall is expected more than 180mm	It rained, or it didn't rain

Q&A

THANK YOU