ORIGINAL ARTICLE
https://doi.org/10.30773/pi.2022.0352

Print ISSN 1738-3684 / On-line ISSN 1976-3026
OPEN ACCESS

Relationships of Antidepressant Medication With Its Various Factors Including Nitrogen Dioxides Seasonality: Machine Learning Analysis Using National Health Insurance Data

Kwang-Sig Lee[1], Hae-In Kim[2], and Byung-Joo Ham[3]

[1]AI Center, Korea University College of Medicine, Seoul, Republic of Korea
[2]Industrial Management Engineering, Korea University School of Industrial Management Engineering, Seoul, Republic of Korea
[3]Department of Mental Health, Korea University College of Medicine, Seoul, Republic of Korea

Objective This study employs machine learning and population-based data to examine major factors of antidepressant medication including nitrogen dioxides (NO_2) seasonality.

Methods Retrospective cohort data came from Korea National Health Insurance Service claims data for 43,251 participants with the age of 15–79 years, residence in the same districts of Seoul and no history of antidepressant medication during 2002–2012. The dependent variable was antidepressant-free months during 2013–2015 and the 103 independent variables for 2012 or 2015 were considered, e.g., particulate matter less than 2.5 micrometer in diameter ($PM_{2.5}$), PM_{10}, NO_2, ozone (O_3), sulphur dioxide (SO_2) and carbon monoxide (CO) in each of 12 months in 2015.

Results It was found that the Cox hazard ratios of NO_2 were statistically significant and registered values larger than 10 for every three months: March, June–July, October, and December. Based on random forest variable importance and Cox hazard ratios in brackets, indeed, the top 20 factors of antidepressant medication included age (0.0041 [1.69–2.25]), migraine and sleep disorder (0.0029 [1.82]), liver disease (0.0017 [1.33–1.34]), exercise (0.0014), thyroid disease (0.0013), cardiovascular disease (0.0013 [1.20]), asthma (0.0008 [1.19–1.20]), September NO_2 (0.0008 [0.01]), alcohol consumption (0.0008 [1.31–1.32]), gender - woman (0.0007 [1.80–1.81]), July NO_2 (0.0007 [14.93]), July PM_{10} (0.0007), the proportion of the married (0.0005), January $PM_{2.5}$ (0.0004), September $PM_{2.5}$ (0.0004), chronic obstructive pulmonary disease (0.0004), economic satisfaction (0.0004), January PM_{10} (0.0003), residents in welfare facilities per 1,000 (0.0003 [0.97]), and October NO_2 (0.0003).

Conclusion Antidepressant medication has strong associations with neighborhood conditions including NO_2 seasonality and welfare support.
Psychiatry Investig 2023;20(6):515-523

Keywords Antidepressive agents; Particulate matter; Nitrogen; Machine learning.

INTRODUCTION

Major depressive disorder (MDD) is a major contributor for disease burden on the globe.[1-4] MDD, "a mood disorder that causes a persistent feeling of sadness and loss of interest," registers a variety of severity (mild to severe) and duration (months to years).[1] Its incidence grew rapidly in the world, i.e., by 47.86% from 172 million to 258 million during 1990 to 2017.[2] Its years-lost-to-disability ranking was third in the world for 2017 and its disability-adjusted-life-years standing was fifth in Korea for 2015.[3,4] It is reported to have various predictors such as demographic conditions (age, gender), socioeconomic status (education, employment, income), neighborhood conditions (crowding, housing, pollution, violence), and health-related predictors (drinking, exercise, smoking, diseases, genetics).[5-7]

Specifically, numerous studies and reviews report a positive association between MDD and particulate matter.[8-13] These cohort and cross-sectional studies were characterized by varying numbers of participants (4,008–71,271) and diverse origins including Africa, America, Asia, and Europe. Neuroinflammation was the central component of a causal pathway

Received: December 6, 2022 **Revised:** February 6, 2023
Accepted: March 5, 2023

✉ **Correspondence:** Byung-Joo Ham, MD, PhD
Department of Mental Health, Korea University Anam Hospital, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Republic of Korea
Tel: +82-2-2286-6843, **Fax:** +82-2-6280-5810, **E-mail:** byungjoo.ham@gmail.com

between particulate matter and MDD in these studies. For example, a recent study[8] employed machine learning and population-based data to examine major predictors of antidepressant medication including the concentration of particulate matter under 2.5 μm ($PM_{2.5}$). Based on random forest variable importance, the top 15 predictors of antidepressant medication during 2013–2015 included cardiovascular disease, age, household income, gender, the district-level proportion of recipients of national basic living security program benefits, district-level social satisfaction, diabetes mellitus, January 2012 $PM_{2.5}$, district-level street ratio, drinker, chronic obstructive pulmonary disease, district-level economic satisfaction, exercise, March 2012 $PM_{2.5}$, and November 2012 $PM_{2.5}$. This recent study concluded that antidepressant medication had strong associations with neighborhood conditions including socioeconomic satisfaction and the seasonality of particulate matter.

However, the previous study covered a limited set of disease history and air pollution, i.e., diabetes mellitus, cardiovascular disease, chronic obstructive pulmonary disease, and $PM_{2.5}$. Given the significant associations of MDD with these diseases and $PM_{2.5}$, it is important to investigate the association of antidepressant medication with other disease histories and air pollutions as well, that is, PM_{10}, nitrogen dioxides (NO_2), ozone (O_3), sulphur dioxide (SO_2), and carbon monoxide (CO). Specifically, one study used time series analysis and 84,207 health insurance records in China's 57 cities, reporting NO_2 to be a risk factor for MDD in terms of hospitalization.[14] However, no machine learning investigation has been done on this topic. In this context, this study uses machine learning and population-based data to analyze major factors of antidepressant medication including 16 disease predictors and 72 air pollutions including NO_2. To our best knowledge, this study presents the most comprehensive analysis for the determinants of antidepressant medication, using a population-based cohort of 43,251 participants and the richest collection of 103 predictors such as 6 demographic/socioeconomic factors, 16 disease predictors, 9 district-level factors, and 72 district-level air pollutions including $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO in each of 12 months (e.g, January $PM_{2.5}$, …, December $PM_{2.5}$, January NO_2, …, December NO_2).

METHODS

Participants

The source of retrospective cohort data for this study was Korea National Health Insurance Service sample research data for 1 million subscribers in Korea (For more description, see https://nhiss.nhis.or.kr/bd/ab/bdaba022eng.do). The final data for this study included 43,251 participants with the age of 15–79 years, residence in the same districts of Seoul and no history of antidepressant medication during 2002–2012. This study was approved by the Institutional Review Board (IRB) of Korea University Anam Hospital on August 19, 2019 (2019AN0354). Informed consent was waived by the IRB.

Variables

The dependent variable was antidepressant-free months during 2013–2015. The following antidepressant medications for all psychiatric disorders (F01–F99) were included: selective serotonin reuptake inhibitor, serotonin-norepinephrine reuptake inhibitor, tricyclic antidepressant, monoamine oxidase inhibitor, and others (i.e., bupropion, trazodone, mirtazapine, and vortioxetine).[13] The 103 independent variables for 2012 were considered: 6 demographic/socioeconomic conditions including gender, age, household income (an insurance fee with the range of 1 [the lowest group] to 10 [the highest group]), smoker (never, former, current), alcohol consumption (0, 1–2, 3–4, ≥5 times per week), and exercise (0 vs. ≥1 times per week); 16 disease factors (each being coded as no vs. yes), i.e., diabetes mellitus, cardiovascular disease, chronic obstructive pulmonary disease, thyroid disease, fat and hyper nutrition, malnutrition, other disorders of glycoprotein metabolism, metabolism, asthma, renal failure, heart failure, migraine and sleep disorder, cerebral palsy and other paralysis syndrome, skin disease, liver disease, and malignant neoplasm; 9 district-level conditions such as population, proportion of the married, economic satisfaction (0–10), social satisfaction (0–10), residents in welfare facilities per 1,000, deprivation index, crude birth rate, recipients of national basic living security program benefits per 1,000, and street ratio; 72 district-level air pollutions including $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO in each of 12 months in 2015.[15-19]

Analysis

A comparison was made between the random forest for survival analysis (the random forest) and the Cox proportional-hazards model (the Cox model)[8,20,21] regarding the prediction of antidepressant-free months. The training and validation sets consisted of 32,438 observations (75%) and 10,813 observations (25%), respectively. The number of trees was 1,000, the maximum depth of the tree was not predetermined and the logrank splitting rule was used for the random forest. Random forest permutation variable importance, the contribution of a variable for the performance of the model, was adopted to identify major predictors of antidepressant-free months. RStudio 1.1.453 (RStudio, Boston, MA, USA) was employed for the analysis.

RESULTS

The descriptive statistics of the 43,251 participants are presented in Tables 1 and 2. The proportion of those taking antidepressants during 2013−2015 was 2.26% (978). The respective statistics of those with male status, the age of 60 years or

Table 1. Descriptive statistics on demographic and socioeconomic variables: antidepressant used vs. not used between 2013 and 2015

Variable	Total (N=43,251)	Antidepressant used (N=978)	Not used (N=42,273)
Gender			
Man	21,014 (49)	377 (39)	20,637 (49)
Woman	22,237 (51)	601 (61)	21,636 (51)
Age			
15−29 yr	3,022 (7)	33 (3)	2,989 (7)
30−39 yr	5,081 (12)	60 (6)	5,021 (12)
40−49 yr	9,902 (23)	182 (19)	9,720 (23)
50−59 yr	12,548 (29)	261 (27)	12,287 (29)
60−69 yr	8,576 (20)	262 (27)	8,314 (20)
70−79 yr	4,122 (10)	180 (18)	3,942 (9)
Household income (decile 00−10)			
00−03 Low	10,062 (23)	238 (24)	9,824 (23)
04−06 Middle	11,018 (25)	205 (21)	10,813 (26)
07−10 High	22,035 (51)	531 (54)	21,504 (51)
Blank*	136 (0)	4 (0)	132 (0)
Smoker			
Non	27,021 (62)	638 (65)	26,383 (62)
Former	6,434 (15)	141 (14)	6,293 (15)
Current	9,288 (21)	188 (19)	9,100 (22)
Did not answer	508 (1)	11 (1)	497 (1)
Drinker			
Less often than once per week	22,449 (52)	572 (58)	21,877 (52)
Once or twice per week	14,077 (33)	250 (26)	13,827 (33)
3−4 Times per week	4,395 (10)	105 (11)	4,290 (10)
5 or more times per week	1,860 (4)	44 (4)	1,816 (4)
Did not answer	470 (1)	7 (1)	463 (1)
Exercise			
Less often than once per week	9,167 (21)	234 (24)	8,933 (21)
Once or more often per week	33,490 (77)	734 (75)	32,756 (77)
Did not answer	594 (1)	10 (1)	584 (1)
District-level information			
Population	454,478 (164,837)	454,500 (164,800)	454,500 (164,800)
Proportion of the married	0.55 (0.04)	0.55 (0.04)	0.55 (0.04)
Economic satisfaction	5.54 (0.45)	5.49 (0.51)	5.54 (0.45)
Social satisfaction	5.98 (0.59)	5.98 (0.59)	5.98 (0.59)
Residents in welfare facilities per 1,000	19.28 (7.58)	19.28 (7.58)	19.28 (7.58)
Deprivation index	0.03 (2.52)	0.08 (2.52)	0.03 (2.52)
Crude birth rate	9.10 (1.40)	8.90 (1.00)	9.10 (1.40)
Recipients of NBLSPB per 1,000	181.90 (104.40)	187.10 (104.40)	181.90 (104.40)
Street ratio	23.02 (3.29)	22.99 (3.29)	23.02 (3.29)

Values are presented number (%) or median (interquartile range). *no insurance fee payment (e.g., professional soldier). NBLSPB, National Basic Living Security Program Benefits

Table 2. Descriptive statistics on disease history between 2002 and 2012

Variable	Total (N=43,251)	Prescribed antidepressant (N=978)	NOT prescribed (N=42,273)
Diabetes miletus			
Yes	8,711 (20)	305 (31)	8,406 (20)
No	34,540 (80)	673 (69)	33,867 (80)
Cardiovascular disease			
Yes	16,584 (38)	542 (55)	16,042 (38)
No	26,667 (62)	436 (45)	26,231 (62)
Chronic obstructive pulmonary disease			
Yes	16,525 (38)	472 (48)	16,053 (38)
No	26,726 (62)	506 (52)	26,220 (62)
Thyroid disease			
Yes	7,494 (17)	249 (25)	7,245 (17)
No	35,757 (83)	729 (75)	35,028 (83)
Fat and hyper nutrition			
Yes	15,334 (35)	513 (52)	14,821 (35)
No	27,917 (65)	465 (48)	27,452 (65)
Malnutrition			
Yes	1,083 (3)	40 (4)	1,043 (2)
No	42,168 (97)	938 (96)	41,230 (98)
Other disorders of glycoprotein metabolism			
Yes	14,685 (34)	493 (50)	14,192 (34)
No	28,566 (66)	485 (50)	28,081 (66)
Metabolism			
Yes	15,457 (36)	516 (53)	14,941 (35)
No	27,794 (64)	462 (47)	27,332 (65)
Asthma			
Yes	10,654 (25)	339 (35)	10,315 (24)
No	32,597 (75)	639 (65)	31,958 (76)
Renal failure			
Yes	432 (1)	16 (2)	416 (1)
No	42,819 (99)	962 (98)	41,857 (99)
Heart Failure			
Yes	607 (1)	35 (4)	572 (1)
No	42,644 (99)	943 (96)	41,701 (99)
Migraine and sleep disorder			
Yes	10,688 (25)	427 (44)	10,261 (24)
No	32,563 (75)	551 (56)	32,012 (76)
Cerebral palsy and other paralysis syndrome			
Yes	349 (1)	16 (2)	333 (1)
No	42,902 (99)	962 (98)	41,940 (99)
Skin disease			
Yes	18,452 (43)	476 (49)	17,976 (43)
No	24,799 (57)	502 (51)	24,297 (57)
Liver disease			
Yes	14,694 (34)	468 (48)	14,226 (34)
No	28,557 (66)	510 (52)	28,047 (66)
Malignant neoplasm			
Yes	2,387 (6)	89 (9)	2,298 (5)
No	40,864 (94)	889 (91)	39,975 (95)

Values are presented number (%)

higher, and the household income of the 7th decile or higher were 49% (21,014), 30% (12,698), and 51% (22,035) as of 2012. The 21%, 48%, and 79% of the participants were current smokers, current drinkers, and those with exercise as of 2012, respectively. Likewise, the corresponding statistics of those with disease history were 25% (10,688) for migraine and sleep disorder, 34% (14,694) for liver disease, 17% (7,494) for thyroid disease, 38% (16,584) for cardiovascular disease, 25% (10,654) for asthma, and 38% (16,525) for chronic obstructive pulmonary disease as of 2012. The yearly averages of air pollutions over 25 districts in Seoul were 26 μg/m^3 for $PM_{2.5}$, 48 μg/m^3 for PM_{10}, 0.023 ppm for NO_2, 0.027 ppm for O_3, 0.005 ppm for SO_2, and 0.517 ppm for CO as of 2015. Finally, the district averages for the proportion of the married, economic satisfaction, and residents in welfare facilities were 0.55, 5.54 (over 10), and 19 per 1,000 as of 2012, correspondingly. The performance measures of the random forest and the Cox model are shown in Table 3. The accuracy of the random forest was around 59% and the C-Index (accuracy) of the Cox models were 60% across board. Based on random forest variable importance (Table 4) and Cox hazard ratios in brackets (Table 5), the top 20 factors of antidepressant medication during 2013–2015 included age (0.0041 [1.69–2.25]), migraine and sleep disorder (0.0029 [1.82]), liver disease (0.0017 [1.33–1.34]), exercise (0.0014), thyroid disease (0.0013), cardiovascular disease (0.0013 [1.20]), asthma (0.0008 [1.19–1.20]), 2015 September NO_2 (0.0008 [0.01]), alcohol consumption (0.0008 [1.31–1.32]), gender-woman (0.0007 [1.80–1.81]), 2015 July NO_2 (0.0007 [14.93]), 2015 July PM_{10} (0.0007), the proportion of the married (0.0005), 2015 January $PM_{2.5}$ (0.0004), 2015 September $PM_{2.5}$ (0.0004), chronic obstructive pulmonary disease (0.0004), economic satisfaction (0.0004), 2015 January PM_{10} (0.0003), residents in welfare facilities per 1,000 (0.0003 [0.97]), and 2015 October NO_2 (0.0003). It was also found in Table 5 that the hazard ratios of NO_2 were statistically significant and registered values larger than 10 for every three months: 2015 March, June–July, October, and December.

DISCUSSION

As a unique contribution of this study, the Cox hazard ratios of NO_2 were found to be statistically significant and registered values larger than 10 for every three months: March, June–July, October, and December. Based on random forest variable importance, in addition, the top 20 factors of antidepressant medication during 2013–2015 included age, migraine and sleep disorder, liver disease, exercise, thyroid disease, cardiovascular disease, asthma, 2015 September NO_2, alcohol consumption, gender - woman, 2015 July NO_2, 2015 July PM_{10},

Table 3. Model performance: random forest vs. Cox model

Air pollution variable	100-error rate (%)	C-index (%)
Each month of 2012	Random forest	Cox model
$PM_{2.5}$, PM_{10}, NO_2	59.1	
$PM_{2.5}$		68.7
PM_{10}		68.6
NO_2		68.6
O_3		68.7
SO_2		68.6
CO		68.6

PM, particulate matter; NO_2, nitrogen dioxides; O_3, ozone; SO_2, sulphur dioxide; CO, carbon monoxide

Table 4. Random forest variable importance: top-30 variables

Rank	Variable	Importance
1	Age	0.0041
2	Migraine and sleep disorder	0.0029
3	Liver disease	0.0017
4	Exercise	0.0014
5	Thyroid disease	0.0013
6	Cardiovascular disease	0.0013
7	Asthma	0.0008
8	NO_2_201509	0.0008
9	Alcohol consumption	0.0008
10	Gender, woman	0.0007
11	NO_2_201507	0.0007
12	PM_{10}_201507	0.0007
13	Proportion of the married	0.0005
14	PM_201501	0.0004
15	PM_201509	0.0004
16	Chronic obstructive pulmonary disease	0.0004
17	Economic satisfaction	0.0004
18	PM_{10}_201501	0.0003
19	Residents in welfare facilities per 1,000	0.0003
20	NO_2_201510	0.0003
21	PM_{10}_201505	0.0003
22	PM_201508	0.0003
23	PM_201510	0.0003
24	Diabetes miletus	0.0003
25	Crude birth rate	0.0003
26	NO_2_201505	0.0003
27	PM_201504	0.0002
28	Heart failure	0.0002
29	Malnutrition	0.0002
30	NO_2_201512	0.0002

PM, particulate matter; NO_2, nitrogen dioxides

Table 5. Cox model hazard ratios and p-values

Predictor/air pollution	$PM_{2.5}$		PM_{10}		NO_2		O_3		SO_2		CO	
	HR	p	HR	p	HR	p	HR	p	HR	p	HR	p
Air pollution 201501	0.97	0.32	0.97	0.61	0.00	0.47	12.91	0.38	0.79	0.03*	1.37	0.65
Air pollution 201502	0.93	0.04*	0.99	0.76	0.00	0.01**	0.00	0.18	1.45	0.26	0.18	0.22
Air pollution 201503	0.99	0.88	1.07	0.35	11.52	0.02*	0.07	0.96	1.39	0.01**	0.22	0.44
Air pollution 201504	1.07	0.27	1.03	0.66	0.00	0.02*	16.34	0.50	0.51	0.13	2.21	0.71
Air pollution 201505	0.93	0.17	0.91	0.45	0.00	0.01*	0.00	0.07†	1.10	0.76	2.56	0.74
Air pollution 201506	1.07	0.13	1.04	0.74	11.56	0.04*	13.81	0.03*	0.94	0.84	11.66	0.48
Air pollution 201507	0.98	0.59	0.99	0.78	14.93	0.02*	0.00	0.65	0.93	0.75	0.31	0.49
Air pollution 201508	1.04	0.41	1.02	0.75	0.00	0.01*	14.56	0.70	1.09	0.78	17.62	0.01**
Air pollution 201509	1.06	0.27	1.00	0.91	0.00	0.01**	0.00	0.31	1.05	0.77	0.05	0.01**
Air pollution 201510	0.94	0.31	1.09	0.02*	15.20	0.04*	14.33	0.18	0.88	0.15	13.06	0.09†
Air pollution 201511	0.98	0.72	0.89	0.01**	0.00	0.58	0.00	0.28	0.82	0.25	0.09	0.42
Air pollution 201512	0.97	0.17	1.00	0.85	18.66	0.76	17.19	0.25	1.22	0.12	2.09	0.68
Gender, man (ref.)												
Woman	1.80	0.01***	1.81	0.01***	1.80	0.01***	1.81	0.01***	1.81	0.01***	1.81	0.01***
Age, 15–29 yr (ref.)												
30–39 yr	1.03	0.91	1.03	0.90	1.03	0.90	1.03	0.90	1.03	0.89	1.03	0.90
40–49 yr	1.37	0.10	1.38	0.10†	1.38	0.10†	1.38	0.10†	1.38	0.09†	1.38	0.10†
50–59 yr	1.30	0.17	1.30	0.17	1.31	0.17	1.31	0.17	1.31	0.16	1.30	0.17
60–69 yr	1.69	0.01**	1.70	0.01**	1.70	0.01**	1.70	0.01**	1.70	0.01**	1.70	0.01**
70–79 yr	2.23	0.01***	2.24	0.01***	2.24	0.01***	2.24	0.01***	2.25	0.01***	2.24	0.01***
Household income, high (ref.)												
Low	1.01	0.91	1.01	0.91	1.01	0.92	1.01	0.90	1.01	0.89	1.01	0.91
Middle	0.89	0.19	0.90	0.19	0.89	0.17	0.89	0.18	0.89	0.19	0.90	0.19
Not available	1.54	0.39	1.53	0.40	1.55	0.39	1.54	0.39	1.53	0.40	1.54	0.39
Smoker, non (ref.)												
Former	1.38	0.01**	1.38	0.01**	1.38	0.01**	1.38	0.01**	1.38	0.01**	1.38	0.01**
Current	1.59	0.01***	1.59	0.01***	1.59	0.01***	1.59	0.01***	1.59	0.01***	1.60	0.01***
Did not answer	2.14	0.09†	2.13	0.09†	2.15	0.09†	2.16	0.08†	2.16	0.08†	2.12	0.09†

Table 5. Cox model hazard ratios and p-values (continued)

Predictor/air pollution	$PM_{2.5}$		PM_{10}		NO_2		O_3		SO_2		CO	
	HR	p	HR	p	HR	p	HR	p	HR	p	HR	p
Alcohol consumption, 0 per week (ref.)												
1–2 per week	1.02	0.84	1.02	0.84	1.02	0.81	1.02	0.84	1.02	0.81	1.02	0.83
3–4 per week	1.31	0.02*	1.31	0.02*	1.32	0.02*	1.32	0.02*	1.32	0.02*	1.31	0.02*
5 or more per week	1.15	0.39	1.15	0.39	1.16	0.39	1.15	0.40	1.16	0.39	1.16	0.38
Did not answer	0.39	0.13	0.39	0.13	0.39	0.13	0.39	0.13	0.38	0.12	0.39	0.13
Exercise, 0 per week (ref.)												
1 or more per week	0.88	0.09†	0.88	0.09†	0.88	0.09†	0.88	0.10†	0.88	0.10†	0.88	0.10†
Did not answer	0.82	0.65	0.82	0.65	0.81	0.62	0.82	0.64	0.81	0.63	0.81	0.64
Population	1.00	0.30	1.00	0.60	1.00	0.02*	1.00	0.54	1.00	0.84	1.00	0.80
Proportion of the married	0.26	0.68	0.10	0.72	15.13	0.01*	15.48	0.31	16.04	0.15	10.25	0.16
Economic satisfaction	0.91	0.76	0.79	0.42	1.40	0.27	1.10	0.73	0.93	0.89	0.90	0.72
Social satisfaction	0.67	0.08†	1.03	0.91	0.17	0.01**	0.72	0.08†	1.28	0.67	1.09	0.75
Residents in welfare facilities per 1,000	0.99	0.18	0.99	0.30	0.97	0.02*	0.99	0.43	1.00	0.85	1.00	0.82
Deprivation index	0.95	0.03*	0.99	0.86	0.86	0.01*	0.98	0.38	0.97	0.31	1.04	0.29
Crude birth rate	0.96	0.60	1.04	0.57	0.91	0.38	0.94	0.33	0.97	0.64	0.96	0.45
Street ratio	1.00	0.89	0.98	0.71	0.89	0.01*	0.99	0.65	0.98	0.63	0.99	0.64
Recipients of NBLSPB per 1,000	1.00	0.38	1.00	0.20	1.01	0.01*	1.00	0.96	1.00	0.74	1.00	0.02*
Diabetes miletus	1.12	0.14	1.12	0.14	1.12	0.15	1.12	0.14	1.12	0.14	1.12	0.13
Cardiovascular disease	1.20	0.02*	1.20	0.02*	1.20	0.02*	1.20	0.02*	1.20	0.02*	1.20	0.03*
Chronic obstructive pulmonary disease	1.10	0.14	1.10	0.14	1.10	0.15	1.11	0.13	1.11	0.13	1.11	0.14
Thyroid disease	1.13	0.12	1.13	0.12	1.13	0.13	1.13	0.13	1.13	0.12	1.13	0.12
Fat amd hyper nutrition	1.37	0.37	1.37	0.36	1.36	0.38	1.37	0.36	1.37	0.36	1.36	0.38
Malnutrition	1.20	0.27	1.20	0.26	1.20	0.26	1.19	0.29	1.20	0.28	1.19	0.28
Other disorders of glycoprotein metabolism	0.84	0.46	0.84	0.46	0.84	0.45	0.84	0.46	0.84	0.46	0.85	0.49
Metabolism	1.08	0.82	1.08	0.82	1.09	0.80	1.08	0.82	1.07	0.83	1.07	0.83
Asthma	1.20	0.01**	1.20	0.01**	1.19	0.01	1.20	0.01*	1.20	0.01*	1.20	0.01**
Renal failure	1.08	0.76	1.07	0.78	1.08	0.77	1.09	0.73	1.07	0.78	1.08	0.77
Heart failure	1.55	0.01*	1.55	0.01*	1.55	0.01*	1.55	0.01*	1.54	0.01*	1.55	0.01*
Migraine and sleep disorder	1.82	0.01***	1.82	0.01***	1.82	0.01***	1.82	0.01***	1.82	0.01***	1.82	0.01***
Cerebral palsy and other paralysis syndrome	1.61	0.06†	1.61	0.06†	1.62	0.06†	1.61	0.06†	1.61	0.06†	1.61	0.06
Skin disease	1.08	0.24	1.08	0.25	1.08	0.24	1.08	0.23	1.08	0.24	1.08	0.23
Liver disease	1.34	0.01***	1.34	0.01***	1.33	0.01***	1.33	0.01***	1.34	0.01***	1.34	0.01***
Malignant neoplasm	1.21	0.09†	1.21	0.09†	1.22	0.09†	1.21	0.10†	1.21	0.09†	1.22	0.09†

*$p<0.050$; **$p<0.010$; ***$p<0.001$; †$p<0.100$. HR, hazard ratio; PM, particulate matter; NO_2, nitrogen dioxides; O_3, ozone; SO_2 sulphur dioxide; CO, carbon monoxide

the proportion of the married, 2015 January $PM_{2.5}$, 2015 September $PM_{2.5}$, chronic obstructive pulmonary disease, economic satisfaction, 2015 January PM_{10}, residents in welfare facilities per 1,000, and 2015 October NO_2.

To our best knowledge, this study presents the most comprehensive analysis for the determinants of antidepressant medication, using a population-based cohort of 43,251 participants and the richest collection of 103 predictors such as 6 demographic/socioeconomic factors, 16 disease predictors, 9 district-level factors, and 72 district-level air pollutions including $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO in each of 12 months. Moreover, this study draws four important clinical and policy implications below. Firstly, the findings of this study agree with those of existing literature on the positive associations of antidepressant medication with age, cardiovascular disease, alcohol consumption, and gender-woman.[8] These four predictors ranked within the top 20 in random forest variable importance and were statistically significant in the Cox model hazard ratios in both studies. Secondly, the results of this study are consistent with those of previous reviews on the positive relationships of depression with sleep disorder,[22] liver disease,[23] and asthma.[24] The importance rankings of these three predictors were within the top 20 and their hazard ratios were statistically significant in this study, i.e., migraine and sleep disorder (1.82), liver disease (1.33–1.34) and asthma (1.19–1.20). Little population-based studies have been done and no machine learning literature has been available on these important issues. In this vein, this study makes a unique contribution in this direction.

Thirdly, this study brings new sights on the seasonality of NO_2 and antidepressant medication. A recent meta-analysis[25] examined 39 original studies on the positive associations of depression with various air pollutions in terms of the relative risk for short-term exposure (less than one month): $PM_{2.5}$ (1.009), PM_{10} (1.009), NO_2 (1.022), SO_2 (1.024), O_3 (1.011), and CO (1.062). These studies reviewed were published during 2007–2021 and covered some of the six air pollutions included in this study (i.e., $PM_{2.5}$, PM_{10}, NO_2, O_3, SO_2, and CO). However, most of these previous studies were cross-sectional and none of them addressed the issue of seasonality based on machine learning. In this study, 2015 July NO_2 ranked within the top 20 in variable importance and its hazard ratio was statistically significant, i.e., 14.93. It was also found in this study (Table 5) that the hazard ratios of NO_2 were statistically significant and registered values larger than 10 for every three months: March, June–July, October, and December. This unique contribution has not been available in the existing literature. Finally, the variable importance of residents in welfare facilities per 1,000 was within the top 20 and its hazard ratio was statistically significant in this study (0.97). These findings affirm the importance of social determinants in the prediction of antidepressant medication.[8]

However, this study had some limitations. Firstly, machine learning is a data-driven approach and this study did not consider pathways among antidepressant medication and its various predictors including NO_2 seasonality. Little analysis has been done and more examination is needed in this direction. Secondly, it was not the scope of this study to evaluate the effects of different subsampling methods on random forest variable importance.[26] Thirdly, it needs to be noted that there would exist room for improvement in the performance measures of the random forest and the Cox model within the ranges of 59%–69%. Increasing the sample size would be an effective solution. Fourthly, the hyper-parameters of the random forest came from existing literature[8,20,21] and hyper-parameter tuning is expected to strengthen its performance. Fifthly, data on PM_{10}, NO_2, O_3, SO_2, and CO before 2015 were not available at the time of data collection. The model performance and validity would have been higher with these data as of 2012. Sixthly, the binary categories of psychiatric disorders were defined based on the International Classification of Diseases 10th Revision code code and this could be a source of potential bias. Finally, uniting various kinds of deep learning approaches for various kinds of MDD data would bring new innovations and deeper insights in this line of research.

In conclusion, antidepressant medication has strong associations with neighborhood conditions including NO_2 seasonality and welfare support. Strong interventions for these factors are really needed for the effective management of MDD.

Availability of Data and Material

The code and data presented in this study are not publicly available. But the code and data are available from the corresponding author upon reasonable request and under the permission of Korea National Health Insurance Service.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Kwang-Sig Lee, Byung0-Joo Ham. Data curation: all authors. Formal analysis: all authors. Funding acquisition: Kwang-Sig Lee, Byung-Joo Ham. Investigation: all authors. Methodology: all authors. Project administration: Kwang-Sig Lee, Byung-Joo Ham. Resources: Kwang-Sig Lee, Byung-Joo Ham. Software: Kwang-Sig Lee, Byung-Joo Ham. Supervision: Kwang-Sig Lee, Byung-Joo Ham. Validation: all authors. Visualization: all authors. Writing—original draft: Kwang-Sig Lee, Byung-Joo Ham. Writing—review & editing: Kwang-Sig Lee, Byung-Joo Ham.

ORCID iDs

Kwang-Sig Lee	https://orcid.org/0000-0002-0576-0098
Hae-In Kim	https://orcid.org/0009-0008-4844-3587
Byung-Joo Ham	https://orcid.org/0000-0002-0108-2058

Funding Statement

This work was supported by (1) National Research Foundation of Korea (NRF) grant funded by the Ministry of Education, Science and Technology of South Korea (No.NRF-2020M3E5D9080792) and (2) Korea Health Industry Development Institute grant funded by the Ministry of Health and Welfare of South Korea (No.HI22C1302 [Korea Health Technology R&D Project]). The funders had no role in the design of the study, the collection, analysis and interpretation of the data and the writing of the manuscript.

REFERENCES

1. Mayo Clinic. Patient care & health information: diseases & conditions: depression. Available at: https://www.mayoclinic.org/diseases-conditions/depression/symptoms-causes/syc-20356007. Accessed February 28, 2021.
2. Liu Q, He H, Yang J, Feng X, Zhao F, Lyu J. Changes in the global burden of depression from 1990 to 2017: findings from the global burden of disease study. J Psychiatr Res 2020;126:134-140.
3. Kim YE, Park H, Jo MW, Oh IH, Go DS, Jung J, et al. Trends and patterns of burden of disease and injuries in Korea using disability-adjusted life years. J Korean Med Sci 2019;34(Suppl 1):e75.
4. Institute for Health Metrics and Evaluation. Findings from the global burden of disease study 2017. Seattle, WA: IHME; 2018.
5. Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, et al. Addressing the burden of mental, neurological, and substance use disorders: key messages from Disease Control Priorities, 3rd edition. Lancet 2016;387:1672-1685.
6. Won E, Ham BJ. Imaging genetics studies on monoaminergic genes in major depressive disorder. Prog Neuropsychopharmacol Biol Psychiatry 2016;64:311-319.
7. Lee KS, Ham BJ. Machine learning on early diagnosis of depression. Psychiatry Investig 2022;19:597-605.
8. Lee KS, Kim G, Ham BJ. Associations of antidepressant medication with its various predictors including particulate matter: machine learning analysis using national health insurance data. J Psychiatr Res 2022; 147:67-78.
9. Liu Q, Wang W, Gu X, Deng F, Wang X, Lin H, et al. Association between particulate matter air pollution and risk of depression and suicide: a systematic review and meta-analysis. Environ Sci Pollut Res 2021; 28:9029-9049.
10. Xue T, Guan T, Zheng Y, Geng G, Zhang Q, Yao Y, et al. Long-term $PM_{2.5}$ exposure and depressive symptoms in China: a quasi-experimental study. Lancet Reg Health West Pac 2020;6:100079.
11. Braithwaite I, Zhang S, Kirkbride JB, Osborn DPJ, Hayes JF. Air pollution (particulate matter) exposure and associations with depression, anxiety, bipolar, psychosis and suicide risk: a systematic review and meta-analysis. Environ Health Perspect 2019;127:126002.
12. Xue T, Zhu T, Zheng Y, Zhang Q. Declines in mental health associated with air pollution and temperature variability in China. Nat Commun 2019;10:2165.
13. Kim KN, Lim YH, Bae HJ, Kim M, Jung K, Hong YC. Long-term fine particulate matter exposure and major depressive disorder in a community-based urban cohort. Environ Health Perspect 2016;124:1547-1553.
14. Ma Y, Jiang Y, Guo T, Wang J, Chen L, Wei C, et al. Short-term exposure to ambient nitrogen dioxide and increased hospitalization burden for depression in China: a multicity analysis. Int J Environ Health Res 2022 Sep 25. [Epub]. Available at: https://doi.org/10.1080/09603123.2022.2126828.
15. Statistics Korea. Korean statistical information service. Available at: https://kosis.kr/index/index.do. Accessed December 1, 2021-July 31, 2022.
16. Seoul Metropolitan Government. 2012 Seoul air quality evaluation report. Seoul: Air Quality Policy Division; 2013.
17. Seoul Metropolitan Government. 2013 Seoul air quality evaluation report. Seoul: Air Quality Policy Division; 2014.
18. Seoul Metropolitan Government. 2014 Seoul air quality evaluation report. Seoul: Air Quality Policy Division; 2015.
19. Seoul Metropolitan Government.. 2015 Seoul air quality evaluation report. Seoul: Air Quality Policy Division; 2016.
20. Lee KS, Jang JY, Yu YD, Heo JS, Han HS, Yoon YS, et al. Usefulness of artificial intelligence for predicting recurrence following surgery for pancreatic cancer: retrospective cohort study. Int J Surg 2021;93:106050.
21. Yu YD, Lee KS, Kim JM, Ryu JH, Lee JG, Lee KW, et al. Artificial intelligence for predicting survival following deceased donor liver transplantation: retrospective multi-center study. Int J Surg 2022;105:106838.
22. Steiger A, Pawlowski M. Depression and sleep. Int J Mol Sci 2019;20: 607.
23. Xiao J, Lim LKE, Ng CH, Tan DJH, Lim WH, Ho CSH, et al. Is fatty liver associated with depression? A meta-analysis and systematic review on the prevalence, risk factors, and outcomes of depression and non-alcoholic fatty liver disease. Front Med (Lausanne) 2021;8:691696.
24. Gao YH, Zhao HS, Zhang FR, Gao Y, Shen P, Chen RC, et al. The relationship between depression and asthma: a meta-analysis of prospective studies. PLoS One 2015;10:e0132424.
25. Borroni E, Pesatori AC, Bollati V, Buoli M, Carugno M. Air pollution exposure and depression: a comprehensive updated systematic review and meta-analysis. Environ Pollut 2022;292(Pt A):118245.
26. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 2007;8:25.