Online Appendix
(Not for Publication)

Seasonal allergies and accidents

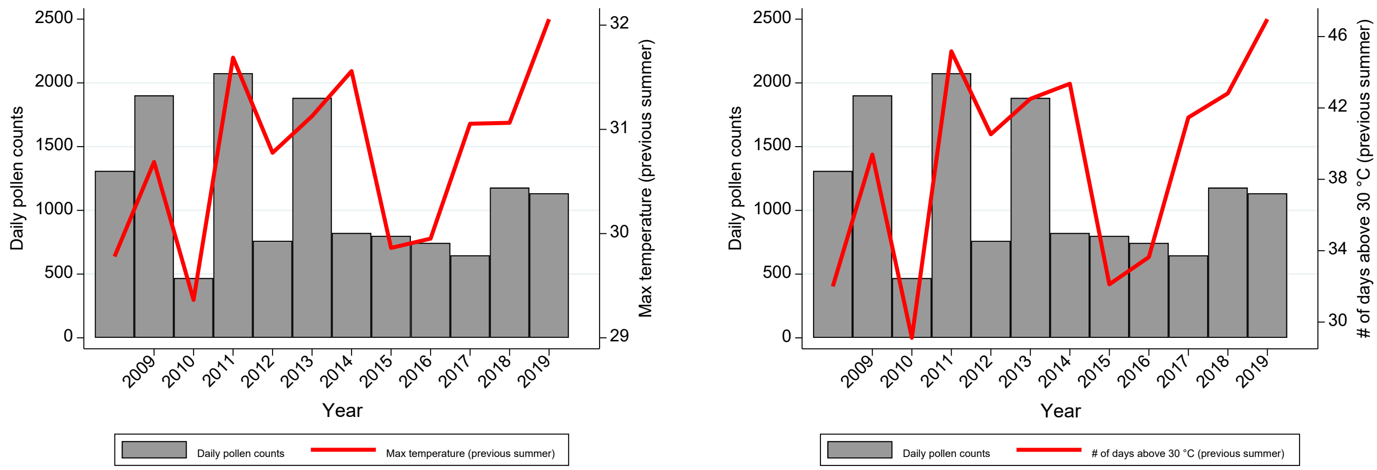
By Mika Akesaka and Hitoshi Shigeoka

Table of Contents

Appendix A: Additional figures and tables

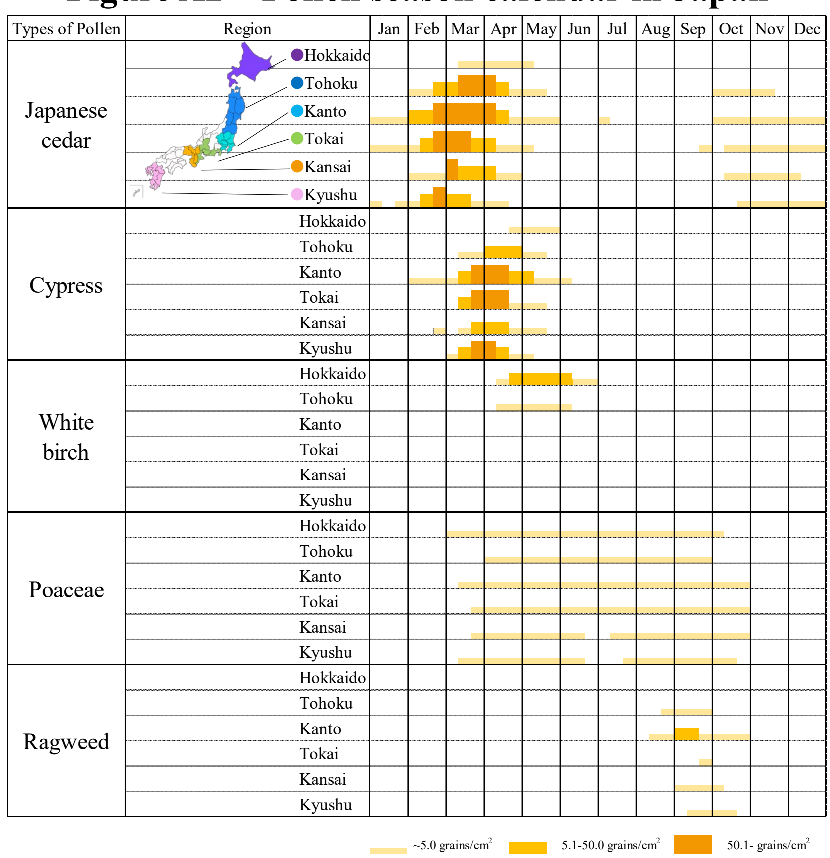
Figure A1—Time series of pollen counts and maximum temperature

A. Maximum temperature of the previous year **B. Number of days above 30℃ in the previous year**



Notes: The graphs illustrate the relationship between the average daily pollen counts (grains/m^3) from February to May and the average maximum temperature (in ℃) in panel A and the number of days the temperature exceeded 30℃ in panel B during July and August of the preceding summer. Data were collected from 120 pollen monitoring stations for the period 2008 to 2019.

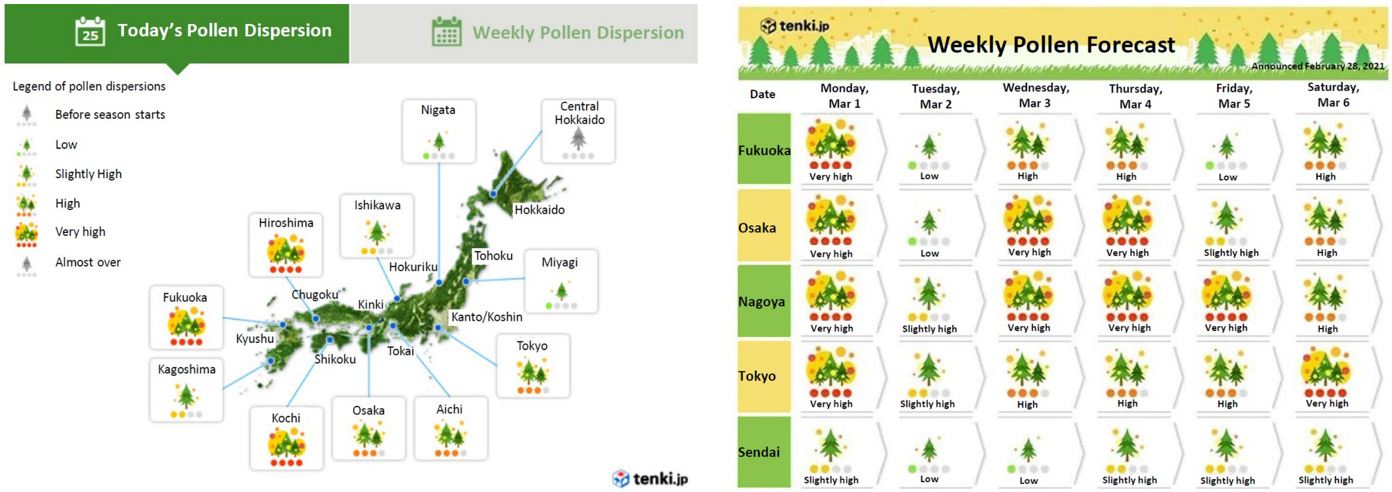
Figure A2—Pollen season calendar in Japan



Notes: The figure displays the pollen dispersal season in Japan, for five selected pollen types: Japanese cedar, cypress, white birch, Poaceae, and ragweed—across six regions (Hokkaido, Tohoku, Kanto, Tokai, Kansai, and Kyushu). The height of the bars signifies the average pollen quantity.
Source: Kishikawa, R., E. Koto, C. Oshikawa, N. So, A. Sugiyama, A. Saito, et al. 2020. “Pollen Calendar of Important Allergenic Airborne Pollen in Japan.” *Japanese Journal of Palynology*, 65(2): 55–66.

Figure A3—Pollen forecast for the current day and week

A. Current day's pollen levels **B. Forecast of the week's pollen levels**

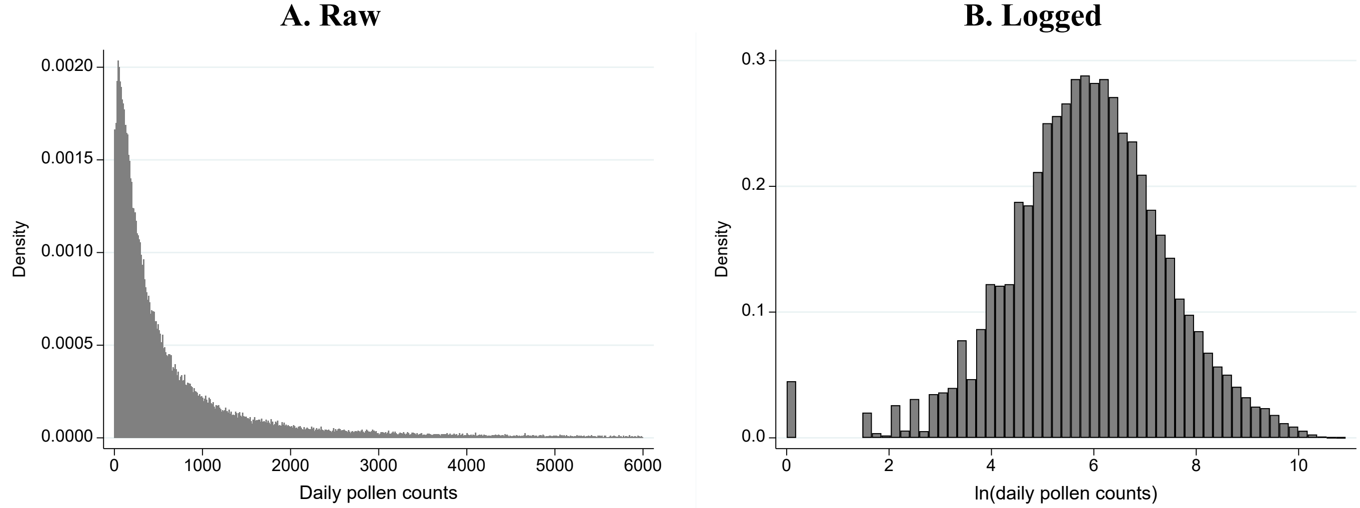


Notes: The graphs display pollen levels as reported on television and a website (https://tenki.jp/) on a typical day during Japan's pollen season. Panel A illustrates the current day's pollen levels at various locations, while panel B presents the forecasted pollen levels for the week (March 1 to March 6, as of February 28, 2021) at different locations from south to north Japan, indicating varying magnitudes. We have obtained permission to translate the original Japanese content into English.

Sources: Japan Weather Association (2022). https://tenki.jp/ (in Japanese)

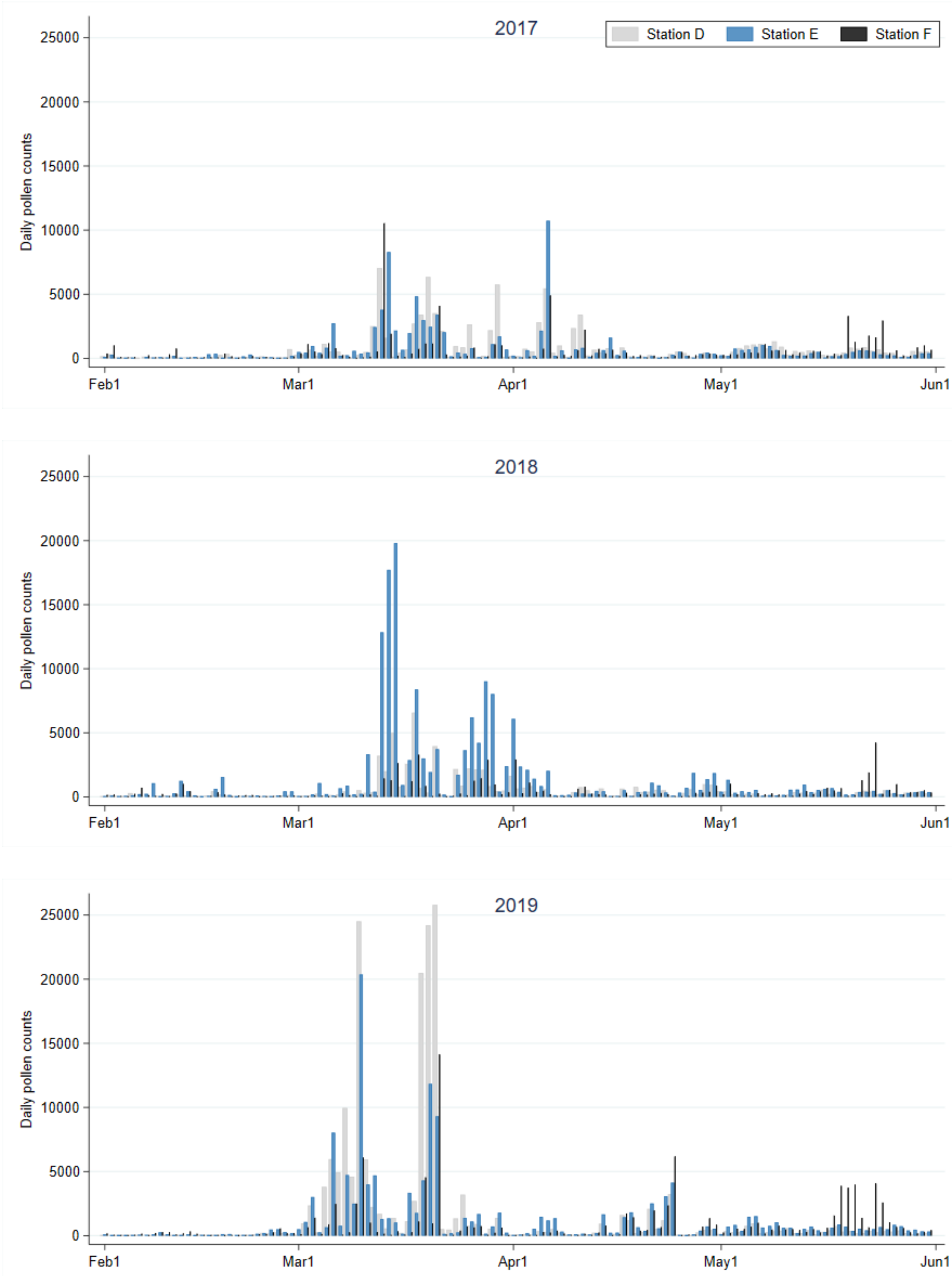
Figure A4—Distribution of the daily pollen counts

A. Raw **B. Logged**



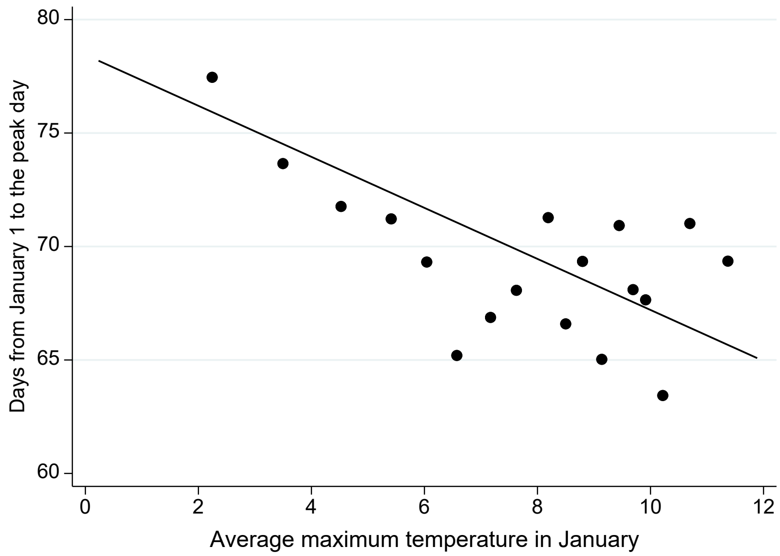
Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N=970,309). A total of 705 emergency response units are available. Panel A displays the daily pollen counts, while panel B illustrates their logged values (grains/m^3). To accommodate zero pollen counts, we add one (0.83%) before taking the logarithm in panel B. In panel A, any daily pollen count exceeding 6,000 (3.06%) is excluded from the graph.

Figure A5—Temporal variation in pollen counts from Niigata Prefecture



Notes: The graphs display the daily variation in pollen counts (grains/m^3) for the period 2017 to 2019 at three monitoring stations in Niigata Prefecture, situated on the northern coast of Honshu, the main island of Japan.

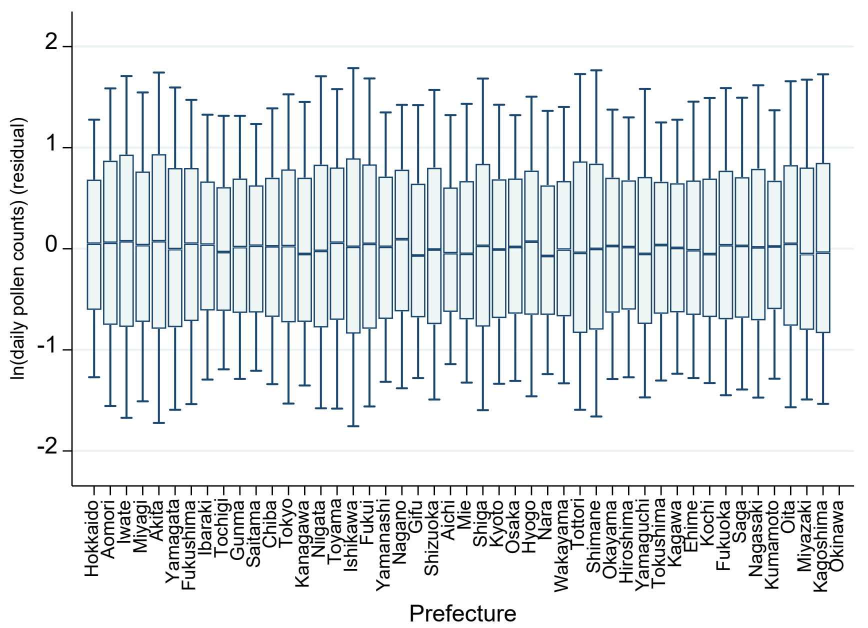
Figure A6—Winter temperature and the peak of pollen seasons



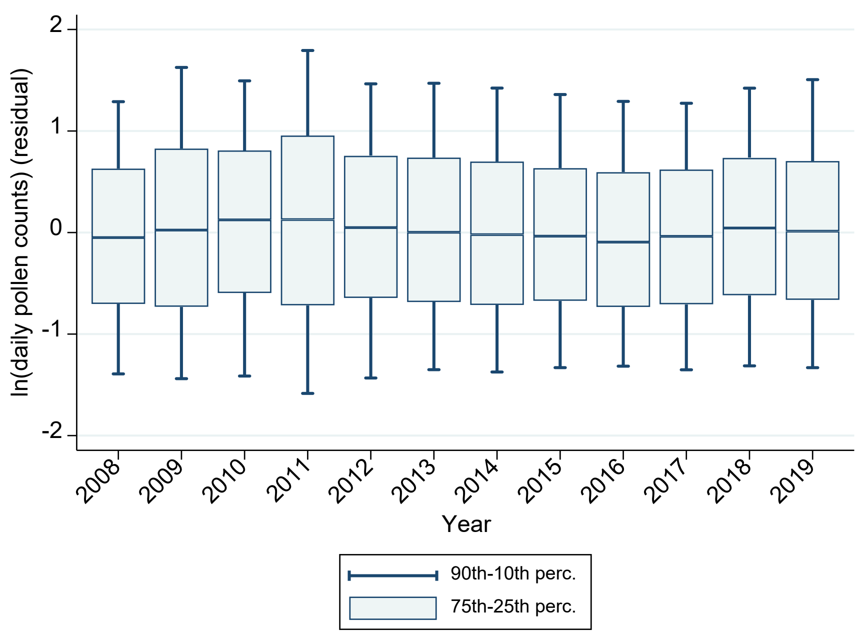
Notes: The sample encompasses all pollen monitoring stations from 2008 to 2019. The figure displays binscatter plots illustrating the relationship between the number of days from January 1st to the peak of the pollen season (on the *y*-axis) and the average maximum temperature (in ℃) in January of each year (on the *x*-axis), after controlling for the pollen monitoring station fixed effects. The peak is determined as the initial day of the year when pollen counts exceed 5,000 grains/m^3, roughly corresponding to the 96th percentile of pollen counts.

Figure A7—Identifying variation in pollen counts

A. Residual variation in logged daily pollen counts, by prefecture

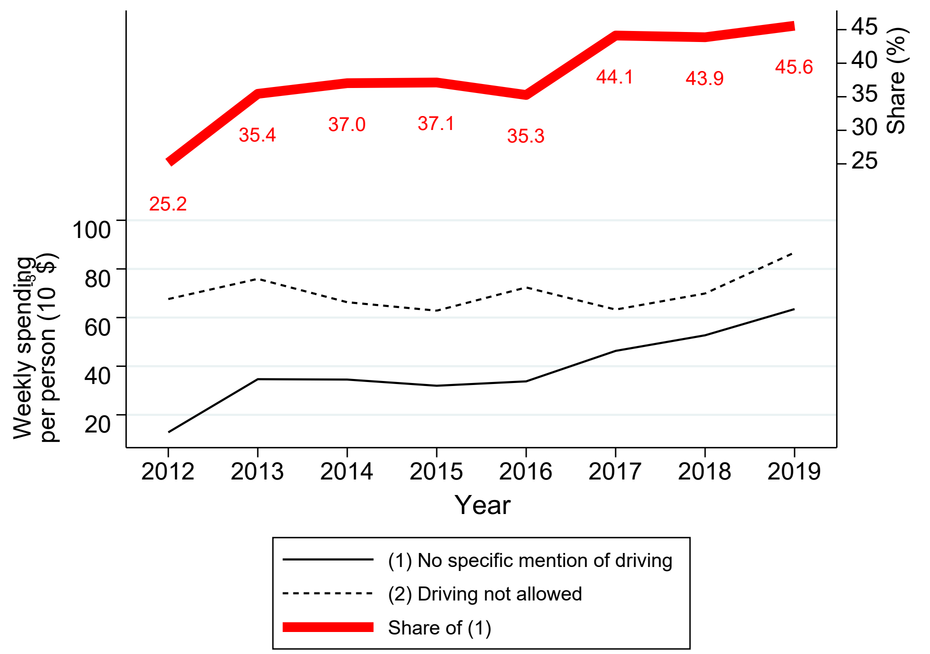


B. Residual variation in logged daily pollen counts, by year



Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N=970,309). A total of 705 emergency response units are available. The graphs display the interquartile range and interdecile range of residual variation in logged daily pollen counts (grains/m^3) by prefecture (panel A) and by year (panel B). With a total of 46 prefectures (excluding Okinawa Prefecture, lacking a pollen monitoring station), residuals are computed after regressing logged daily pollen counts on all controls in equation (1), including unit, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population in each unit.

Figure A8—Time series of expenditures for seasonal allergy medications (by type)



Notes: The sample is derived from retail scanner records from February to May for the period 2012 to 2019. The figure displays the weekly expenditure (in 10^{-3} $) per individual for two categories of seasonal allergy medications: "No specific mention of driving" and "Driving not allowed," showcased on the left y-axis. Concurrently, the proportion of the former type is illustrated on the right y-axis throughout the period from 2012 to 2019. See Appendix Table A1 for the subsets corresponding to the two drug types. An exchange rate of 100 yen/$ is applied.

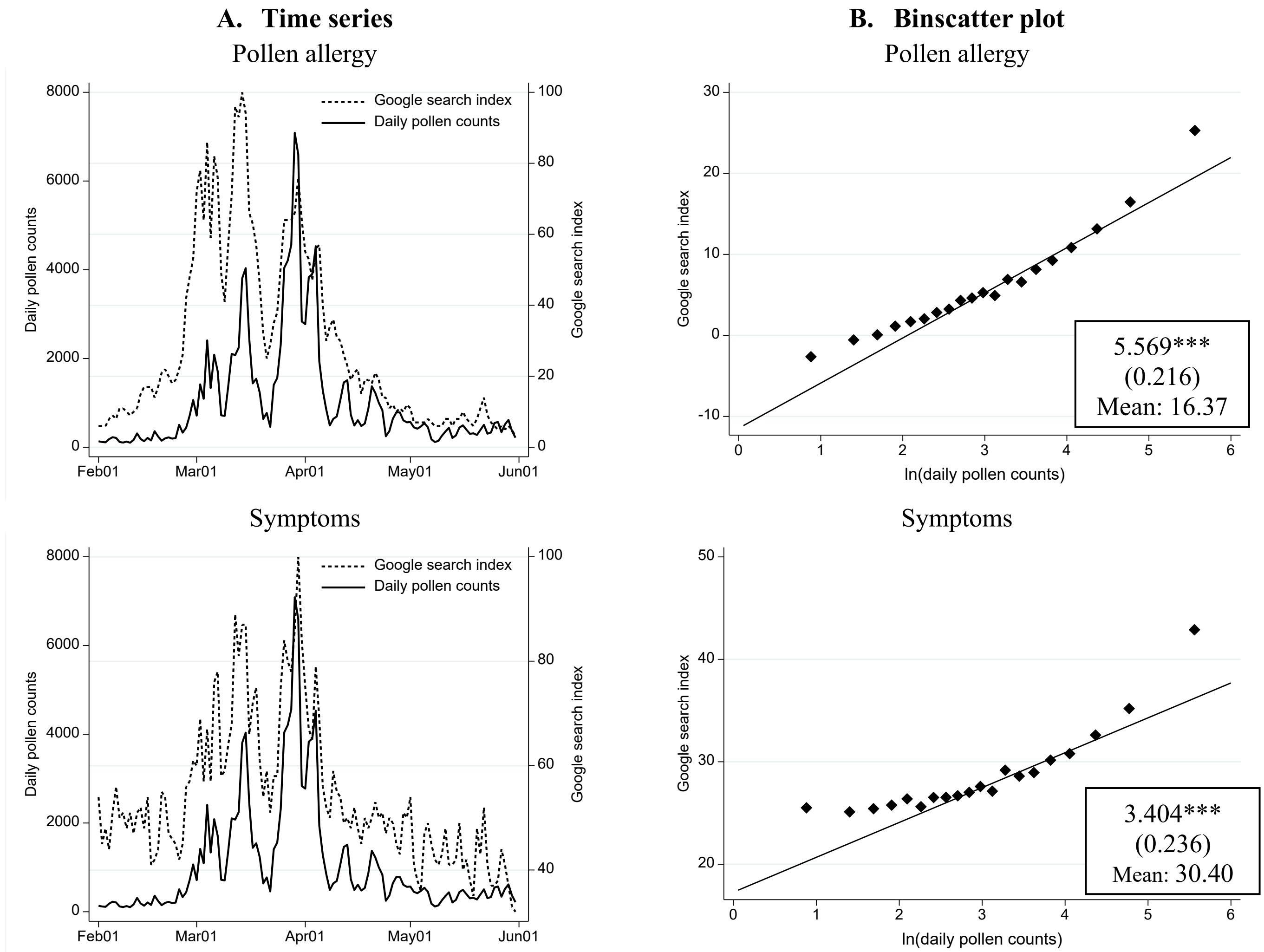
Table A1—List of medications for seasonal allergies

Brand name in Japanese	Brand name in English	Year of release	Mention of driving
アレジオン	Alesion	1994	Driving not allowed
エバステル	Evastel	1996	Driving not allowed
ジルテック	Zyrtec	1998	Driving not allowed
タリオン	Talion	2000	Driving not allowed
アレグラ	Allegra	2001	No specific mention of driving
アレロック	Allelock	2001	Driving not allowed
クラリチン	Claritin	2002	No specific mention of driving
ザイザル	Xyzal	2010	Driving not allowed
ディレグラ	Dellegra	2013	No specific mention of driving
ビラノア	Bilanoa	2016	No specific mention of driving
デザレックス	Desalex	2016	No specific mention of driving
ルパフィン	Rupafin	2017	Driving not allowed

Notes: The table presents brand names of allergy medications utilized for seasonal allergy treatment, indicating the year of release and any special advisories regarding driving post-medication consumption. Medications shaded in gray lack explicit instructions regarding driving. Note that our accident data primarily covers the period from 2008 to 2019.

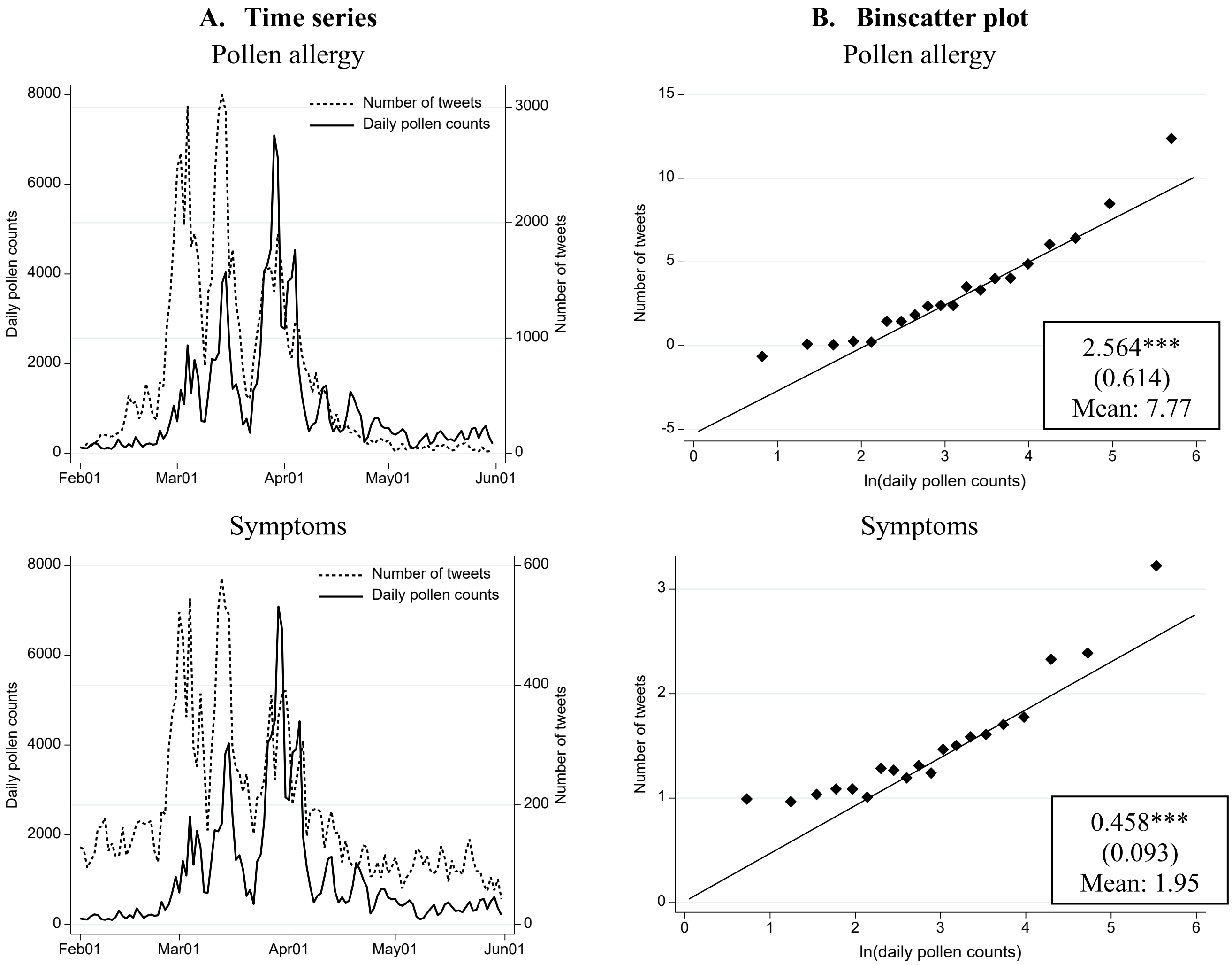
Appendix B: Symptoms of seasonal allergies

Figure B1—Daily pollen counts and Google Trends



Notes: The sample is derived from Google Trends data, with observations at the prefecture-per-day level. Panel A illustrates the time-series patterns of average daily pollen counts (grains/m^3) and Google search index for pollen allergy-related and symptom-related keywords in 2018 on a national scale. See Appendix Table B1 for the list of search terms within each category. June is omitted because only four stations in Hokkaido (the northernmost island of Japan) were still active in June. Panel B presents the binscatter plots illustrating the relationship between logged daily pollen counts (grains/m^3, on the x-axis) and the Google search index for the same keywords (on the y-axis), after controlling for month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population for the period 2016 to 2019 (N= 21,551). Estimates from the variants of equation [1], where unit fixed effects are replaced by prefecture fixed effects, are reported in the box. Standard errors clustered at the prefecture level are reported in parentheses. Estimates are weighted by the population in each prefecture per year. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

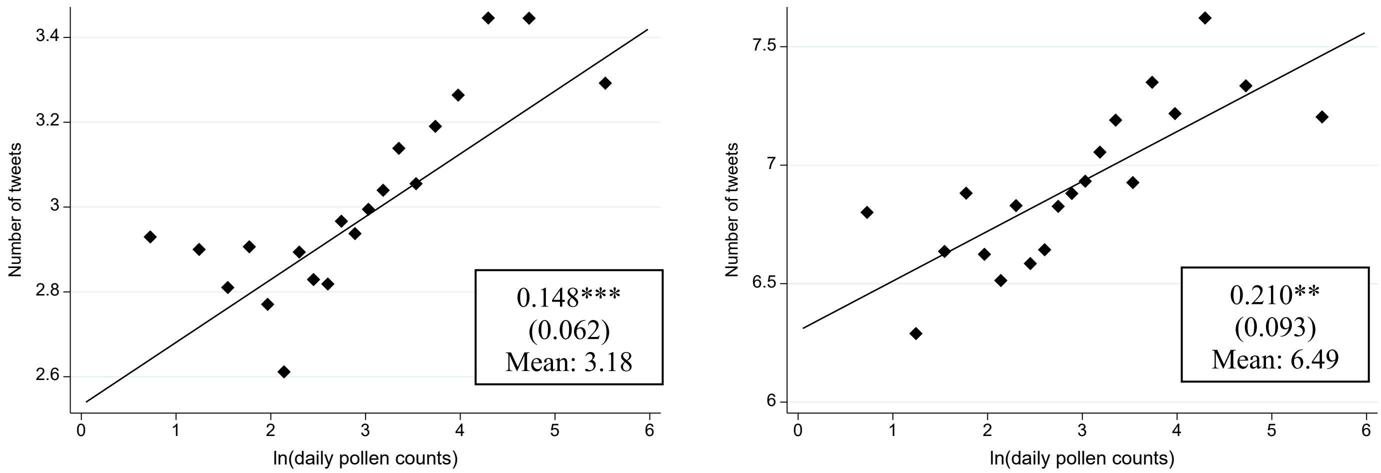
Figure B2—Daily pollen counts and tweets



Notes: The sample is derived from Twitter data, with observations at the prefecture-per-day level. Panel A presents the time series patterns of the average daily pollen counts (grains/m^3) and the number of tweets for pollen allergy-related and symptom-related keywords in 2018 on a national scale. See Appendix Table B1 for the list of search terms within each category. June is omitted because only four stations in Hokkaido (the northernmost island of Japan) were still active in June. Panel B displays the binscatter plots illustrating the relationship between logged daily pollen counts (grains/m^3, on the *x*-axis) and the number of tweets for the same keywords (on the *y*-axis), after controlling for month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population for the period 2016 to 2019 (N= 21,551). Estimates from the variants of equation [1], where unit fixed effects are replaced by prefecture fixed effects, are provided in the box. Standard errors clustered at the prefecture level are enclosed in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Figure B3—Binscatter plot of pollen counts and tweets: Sleep-related

Sleeplessness (“Having a hard time falling asleep”) | **Sleepiness** (“Feeling sleepy”)



Notes: The sample is derived from Twitter data for the period 2016 to 2019, with observations at the prefecture per day level (N=21,551). The graphs display binscatter plots illustrating the relationship between logged daily pollen counts (grains/m^3, on the x-axis) and the frequency of tweets containing the terms “Hard time falling asleep” and “Feeling sleepy” (on the y-axis), after controlling for month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. See Appendix Table B1 for the categorized list of search terms. Estimates from the variants of equation [1], where unit fixed effects are replaced by prefecture fixed effects, are provided in the box. Standard errors clustered at the prefecture level are enclosed in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

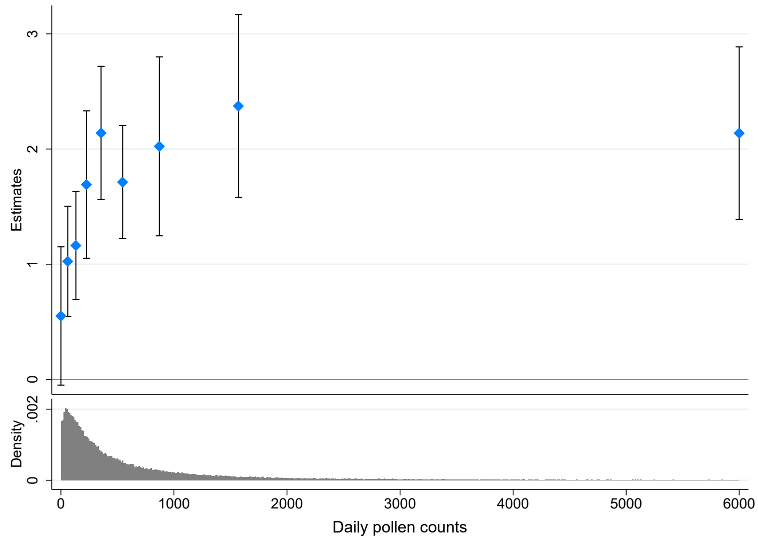
Table B1—List of search terms for symptoms

Category #	Categories	Japanese	English
Common for Google Trends and Tweets			
1	**Pollen allergy**	花粉	Pollen
		花粉症	Pollen allergy
		スギ花粉	Japanese cedar pollen
2	**Symptoms**	鼻水	Runny nose
		鼻づまり	Nasal congestion
		くしゃみ	Sneezing
		目のかゆみ	Itchy eyes
Only for Tweets (sleep-related)			
3	**Sleeplessness**	寝付けない，ねつけない	Having a hard time falling asleep
		寝れない，ねれない	
		眠れない，ねむれない	
4	**Sleepiness**	眠い，ねむい	Feeling sleepy
		眠たい，ねむたい	
		眠すぎる，ねむすぎる	

Notes: The table lists the keywords for each category in both Japanese and English (for reference).

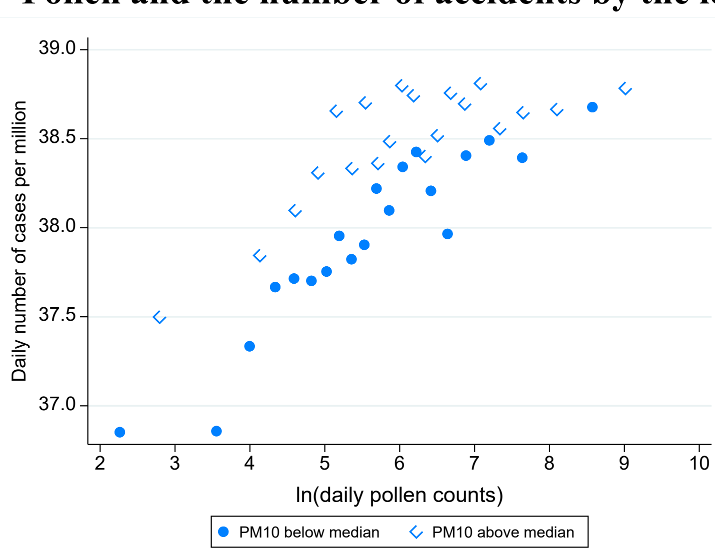
Appendix C: Ambulance records

Figure C1—Dose responses



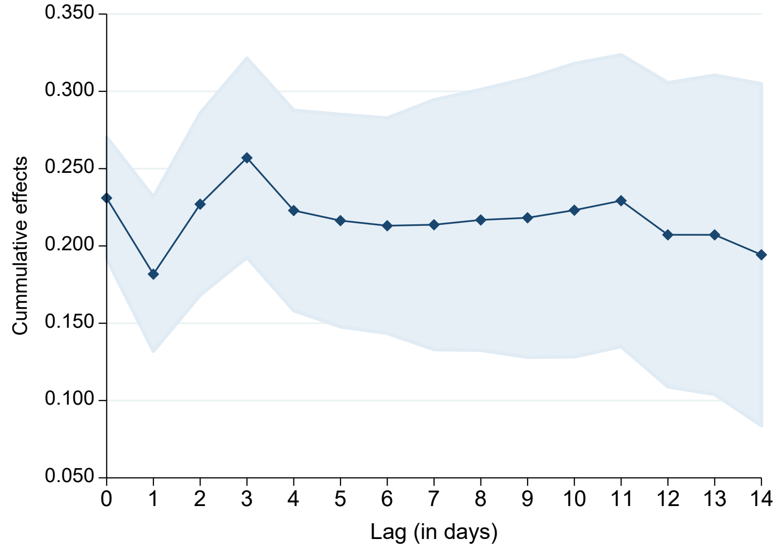
Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N= 970,309). A total of 705 emergency response units are available. The plots exhibit estimates and 95% confidence intervals of the treatment effects of daily pollen counts (in levels), using a variant of equation [1] where the logged daily pollen is replaced by dummies for each decile of daily pollen levels (grains/m^3). The dependent variable is the number of daily cases per million people. Standard errors are clustered at the pollen monitoring station levels. All specifications include unit, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population of each unit. The histogram at the bottom displays the distribution of daily pollen counts (grains/m^3).

Figure C2—Pollen and the number of accidents by the level of PM_{10}



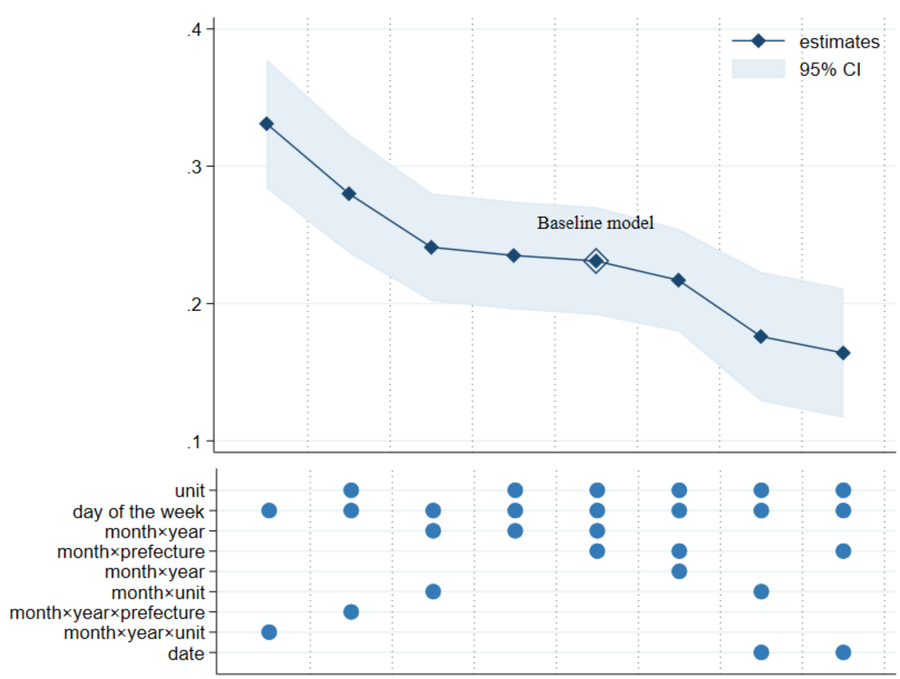
Notes: The sample is derived from ambulance records for the period 2009 to 2019, with observations at the unit per day level (N= 806,839). A total of 705 emergency response units are available. The figure displays binscatter plots illustrating the relationship between logged daily pollen counts (grains/m^3, on the *x*-axis) and the number of daily cases per million people for all accidents (on the *y*-axis), categorized as below and above the daily median of PM_{10}, after controlling for unit, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population of each unit.

Figure C3—Varying window for the sum of coefficients in the distributed lag model



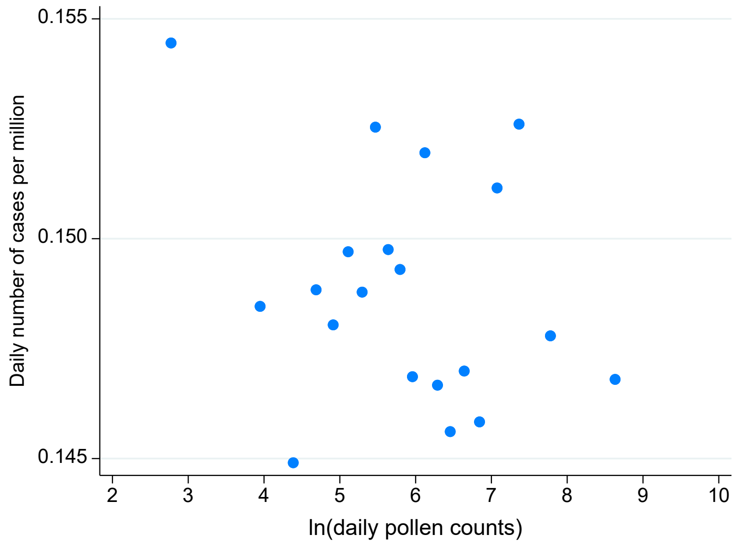
Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N= 970,309). A total of 705 emergency response units are available. The plots depict the estimates and 95% confidence intervals of the sum of coefficients ($=\sum_{k\in K}\beta_k$) in the distributed lag model derived from equation [2], with varying windows of up to 14 days. The dependent variable is the daily number of cases per million people. All specifications include fixed effects for unit, month-by-year, month-by-prefecture, and day-of-week. Additionally, logged pollen counts and weather covariates (precipitation, temperature, wind speed) for the days preceding and following the observation date within the specified time horizon are included. Estimates are weighted by the population of each unit.

Figure C4—Different time FE



Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N= 970,309). A total of 705 emergency response units are available. The plots display estimates and 95% confidence intervals of treatment effects on logged daily pollen counts from equation [1] under various specifications with high-dimensional unit and time fixed effects. Standard errors are clustered at the pollen monitoring station level. The dependent variables are the number of daily cases per million people. The baseline model includes fixed effects for unit, month-by-year, month-by-prefecture, and day-of-week. All specifications additionally include weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population of each unit.

Figure C5—Pollen counts and emergency ambulance transports due to cancer



Notes: The sample is derived from ambulance records for the period 2015 to 2019, with observations at the unit per day level (N= 407,463). A total of 705 emergency response units are available. These ambulance records encompass detailed diagnosis information, equivalent to ICD10, starting from 2015. The figure displays binscatter plots illustrating the relationship between the logged daily pollen counts (grains/m^3, on the *x*-axis) and the frequency of daily emergency ambulance transports due to cancer per million people (on the *y*-axis), after controlling for unit, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population of each unit.

Table C1—Pollen and pollution interactions

	(1)	(2)
ln(pollen counts)	0.230***	0.267***
	(0.021)	(0.027)
$1(PM_{10} \geq$ median)		0.068
		(0.222)
ln(pollen) × $1(PM_{10} \geq$ median)		-0.053
		(0.036)
R-squared	0.46	0.46
N	806,839	806,839
Unit FE	X	X
Day-of-week FE	X	X
Month-by-year FE	X	X
Prefecture-by-month FE	X	X

Notes: The sample is derived from ambulance records for the period 2009 to 2019, with observations at the unit per day level (N= 806,839). Pollution data are available starting from 2009. A total of 705 emergency response units are available. The dependent variable is the daily number of cases per million people for all accidents. Column (1) replicates the baseline estimates presented in Table 2 for the subset with available pollution data. Column (2) presents estimates derived from the variation of equation [1], where a dummy variable, which takes a value of 1 when the daily PM_{10} level exceeds the median, and 0 otherwise, and its interaction with logged pollen counts is also included. Standard errors clustered at the pollen monitoring station level are reported in parentheses. In addition to the fixed effects listed in the table, weather covariates (precipitation, temperature, wind speed), darkness, and logged population are included in the estimation. Estimates are weighted by the population of each unit. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table C2—Different levels of clustering

Clustering variables	N of clusters	SE	
		0.231	
Monitoring stations (baseline)	120	(0.020)	***
Monitoring stations and date	120 + 1,796	(0.028)	***
Monitoring stations and month-year	120 + 60	(0.030)	***
Unit	705	(0.019)	***
Unit and date	705 + 1,796	(0.028)	***
Unit and month-year	705 + 60	(0.031)	***
Prefecture	47	(0.016)	***
Prefecture and date	47 + 1,796	(0.025)	***
Prefecture and month-year	47 + 60	(0.028)	***
Conley (50 km)	-	(0.018)	***
Conley (100 km)	-	(0.019)	***
Conley (150 km)	-	(0.020)	***

Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N= 970,309). A total of 705 emergency response units are available. The dependent variable is the number of daily cases per million people. Estimates from equation [1] are reported. All specifications include unit, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. The numbers of pollen monitoring stations, units, and prefectures are 120, 705, and 46, respectively. The last three rows present spatially clustered standard errors following the methodology outlined in Conley (1999), employing distance cutoffs of 50, 100, and 150 km, respectively. Estimates are weighted by the population in each unit. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table C3—Alternative specifications

	(1)	(2)	(3)
	level-log OLS (Baseline)	log-log OLS	Poisson pseudo-maximum likelihood (PPML)
ln(pollen counts)	0.231***	0.0054***	0.030***
	(0.020)	(0.0006)	(0.003)
R-squared	0.46	0.90	-
N	970,309	970,309	970,309
Unit FE	X	X	X
Day-of-week FE	X	X	X
Month-by-year FE	X	X	X
Prefecture-by-month FE	X	X	X

Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level (N=970,309). A total of 705 emergency response units are available. Column (1) reports the results from Table 2 (baseline) for ease of comparison. Column (2) reports estimates from the variant of equation [1], wherein the dependent variable takes the logarithm of the number of daily cases per million people. Estimates are weighted by the population in each unit in columns (1) and (2). Column (3) reports the marginal effect of Poisson pseudo-maximum likelihood (PPML) using the *dydx* command in Stata. The estimate in column (3) can be converted to 0.222 cases per million people (= 0.030/0.135), where 0.135 is the average population in a million, which is comparable to the baseline estimate in column (1). The standard errors clustered at the pollen station level are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table C4—Placebos

	(1)	(2)	(3)
	Baseline	Assigning last year's pollen counts	Assigning next year's pollen counts
ln(pollen counts)	0.231***	0.007	0.021
	(0.020)	(0.022)	(0.022)
R-squared	0.46	0.47	0.46
N	970,309	873,307	874,359
Mean of dep. var	33.03	33.21	32.90
Unit FE	X	X	X
Day-of-week FE	X	X	X
Month-by-year FE	X	X	X
Prefecture-by-month FE	X	X	X

Notes: The sample is derived from ambulance records for the period 2008 to 2019, with observations at the unit per day level. A total of 705 emergency response units are available. The dependent variable is the number of daily cases per million people. The estimates from equation [1] are reported. All specifications include fixed effects for unit, month-by-year, month-by-prefecture, and day-of-week, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Column (1) replicates the estimates from Table 2 (baseline) for ease of comparison. Columns (2) and (3) falsely assign the pollen levels of the corresponding day from the previous and subsequent years, respectively (for instance, for March 3, 2018, in unit X, columns (2) and (3) assign the pollen levels of March 3, 2017, and March 3, 2019, within the same unit X). Estimates are weighted by the population in each unit. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Appendix D: Police records

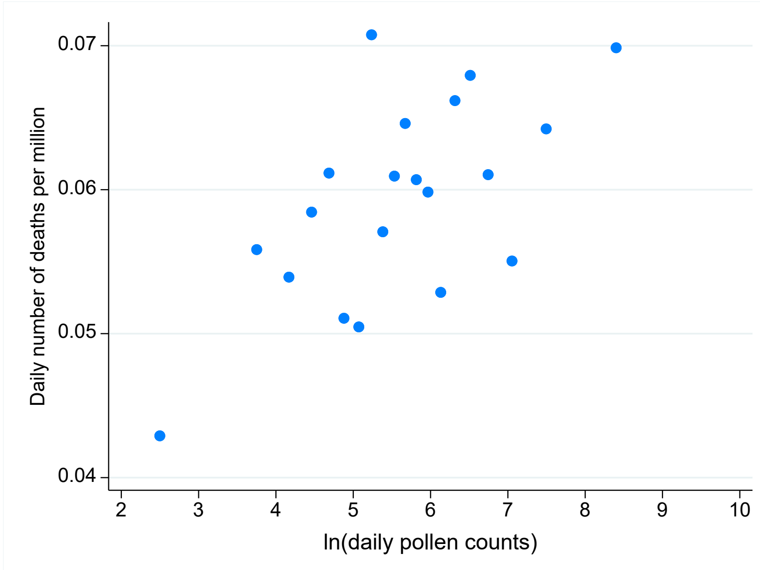
The police records encompass all 690,415 traffic accidents (or 106,533 during the pollen season) occurring between 2019 and 2020. These data are compiled at the individual accident level, detailing information such as location, date, and time of occurrence. Unlike the ambulance service, which operates at a unit level (N=705), the police service is administered at the municipal level (N=1,700). Consequently, we aggregate casualty figures to the municipal-day level by consolidating hourly observations within municipalities.

Figure D1 illustrates the binscatter plots depicting the relationship between the logged average daily pollen count (grains/m^3) and the number of traffic fatalities per million people, while controlling for municipality, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. The figure clearly demonstrates the positive relationship between these variables.

Table D1 presents the estimates from equation [1], wherein the unit fixed effect is replaced by the municipality fixed effect. For ease of comparison, column (1) shows the death/fatal estimates for traffic accidents recorded between 2008 and 2019, sourced from ambulance records. Column (2) displays the mortality estimates derived from the 2019 to 2020 police records. The estimate of 0.0040 (p-value < 0.01) in column (2) surpasses the 0.0026 estimate from column (1), suggesting a potential underestimation of the impact of pollen exposure on traffic accident fatalities.[1] However, at conventional levels, the two estimates are not statistically distinguishable.

[1] One plausible explanation for this observation is that police records encompass all deaths resulting from traffic accidents within 24 hours, unlike ambulance records, which solely encompass deaths occurring upon hospital admission. Notably, the count of traffic accidents resulting in death recorded in our ambulance records for 2019 is 1,771, whereas the corresponding figure reported to the National Police Agency (NPA) is 3,215 (NPA 2022).

Figure D1—Pollen and mortality due to traffic accidents (police records)



Notes: The sample is derived from police records for the period 2019 to 2020, with observations at the municipality per day level (N= 399,749). A total of 1,700 municipalities exist. The figure displays binscatter plots illustrating the relationship between logged daily pollen counts (grains/m^3, on the x-axis) and the number of deaths per million people within 24 hours resulting from traffic accidents (on the y-axis), after controlling for municipality, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population.

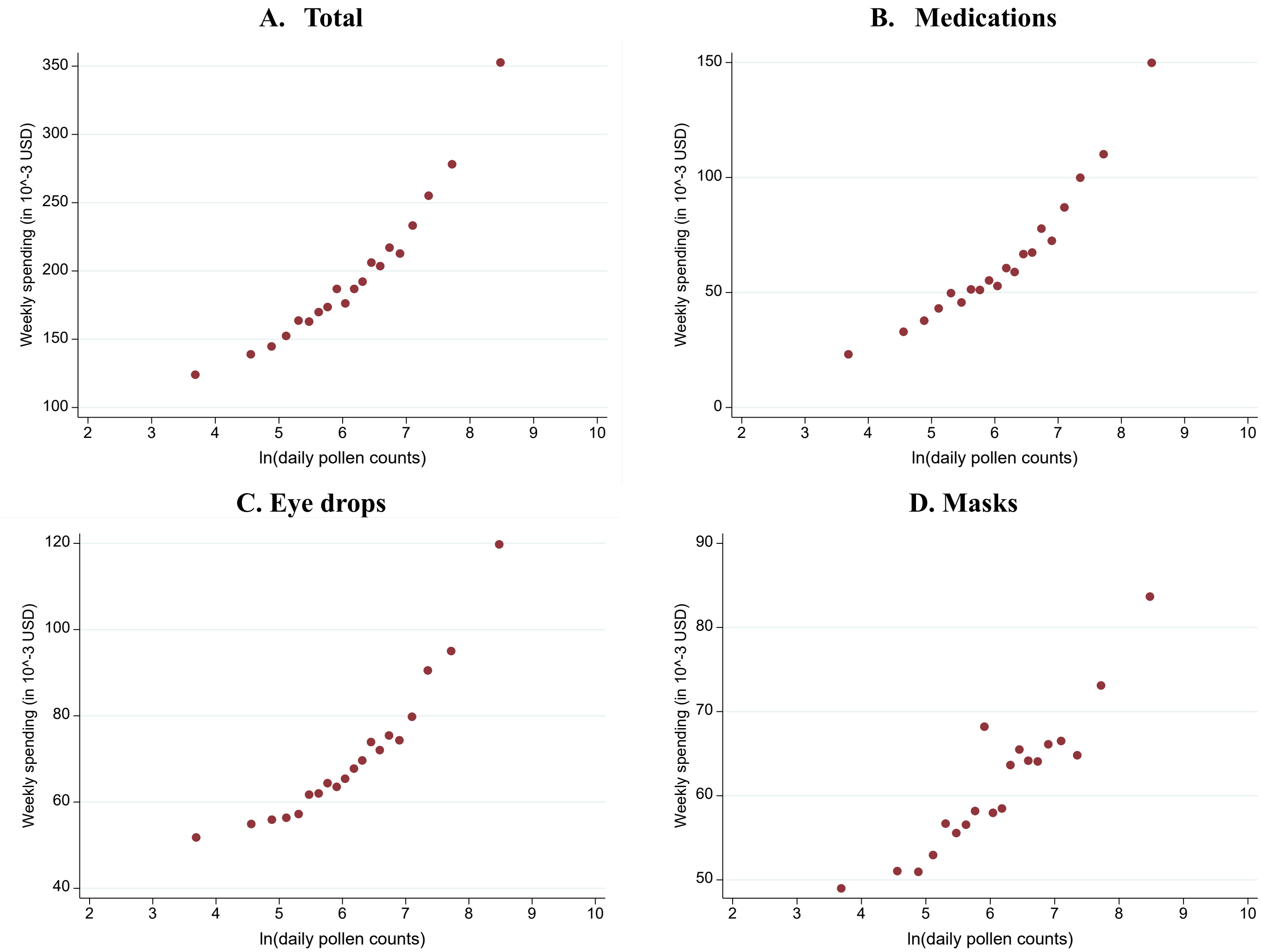
Table D1—Mortality due to traffic accidents (ambulance records vs. police records)

	Ambulance records	Police records
	(1)	(2)
ln(pollen counts)	0.0026***	0.0037**
	(0.0007)	(0.0015)
R-squared	0.01	0.01
N	970,309	399,749
N of unit/municipality	705	1,700
N of clusters	120	120
Mean of dep. var	0.077	0.125
Unit/municipality FE	X	X
Day-of-week FE	X	X
Month-by-year FE	X	X
Prefecture-by-month FE	X	X

Notes: The sample for column (1) is derived from ambulance records for the period 2008 to 2019, while the sample for column (2) is derived from police records for the period 2019 to 2020. The level of observation is units per day for column (1) and municipality per day for column (2). There are a total of 705 emergency response units and 1,700 municipalities. The dependent variable is the number of deaths due to traffic accidents per million people. Estimates from equation [1] are reported along with standard errors clustered at the pollen monitoring station level in parentheses. In addition to the fixed effects listed in the table, we include weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population in each unit in column (1) and by the population in each municipality in column (2). Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Appendix E: Avoidance behavior from retail scanner data

Figure E1—Pollen count and weekly expenditure on allergy products



Notes: The sample is derived from retail scanner data from February to May for the period 2012 to 2019, with observations at a weekly per-person level (N= 4,303,417). The graphs display binscatter plots illustrating the relationship between logged daily mean pollen counts (grains/m^3, on the *x*-axis) and weekly expenditure per person (in 10^{-3}\$, on the *y*-axis) for all allergy products in panel A, and individually for each product in panels B to D: medications (panel B), eye drops (panel C), and masks (panel D), after controlling for municipality, month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. The shares of medications (panel B), eye drops (panel C), and masks (panel D) are 35%, 37%, and 28%, respectively. An exchange rate of 100 yen/$ is applied.

Table E1—Nielsen Panel Data vs. Quick Purchase Report

Feature	**Nielsen Homescan Panel (U.S.)**	**Quick Purchase Report (QPR) by Macromill**
Country	United States	Japan
Unit of Observation	Household	Household
Method	Homescan + manual report (partly)	Homescan
Target Product	UPC-coded packaged goods + some non-barcode items (manual report)	JAN-coded packaged goods
Coverage Period	2004-	2011-
Sample Size	40,000–60,000	Approximately 35,000
Panel Structure	Yes	Yes
Nationally Representative	Yes	Yes
Information Included:		
Purchase Date	Yes	Yes
Retailer/Channel Type	Yes	Yes
Online Purchase	Yes	Yes
Price Paid	Yes	Yes
Coupon/Deal Flag	Yes	No

Sources:
Nielsen: https://www.chicagobooth.edu/research/kilts/research-data/nielseniq?utm_source=chatgpt.com
Macromill: https://www.e-stat.go.jp/bigdataportal/dataintro/253

Table E2—Summary statistics of retail scanner data

Variables	Obs	Mean	Std. dev.	Min	Max
A. Outcomes (weekly per person in $\$10^{-3}$)					
Spending: Total	4,303,417	295.78	2343.95	0	2,141,620
Spending: Medications	4,303,417	109.94	1528.34	0	227,200
Spending: Eye drops	4,303,417	103.41	1018.20	0	104,000
Spending: Masks	4,303,417	82.52	1324.56	0	2,141,620
B. Individual characteristics					
Ages: 0–24	4,303,417	0.13	0.33	0	1
Ages: 25–44	4,303,417	0.38	0.48	0	1
Ages: 45–64	4,303,417	0.38	0.49	0	1
Ages: 65 years and older	4,303,417	0.11	0.32	0	1
Female	4,303,417	0.50	0.50	0	1
Student	4,303,417	0.11	0.31	0	1
Household head	4,303,417	0.46	0.50	0	1
Have kids	4,303,417	0.55	0.50	0	1
Own house	4,303,417	0.58	0.49	0	1
Annual salary category	4,303,417	2.61	2.22	1	13
C. Regressors (daily average in a week)					
Pollen counts (grains/m^3)	4,303,417	843.2	1337.4	0	22,812
Logged (Pollen counts)	4,303,417	0.1	0.2	0	3.9
Precipitation (mm)	4,303,417	12.0	6.0	-8	26
Average temperature (℃)	4,303,417	2.8	0.9	0	11.6
Average wind speed (m/s)	4,303,417	10.5	1.3	7	13

Notes: The sample is derived from retail scanner data from February to May for the period 2012 to 2019, with observations at a weekly per-person level. Expenditure is measured in units of 10^{-3} \$. An exchange rate of 100 yen/\$ is applied.

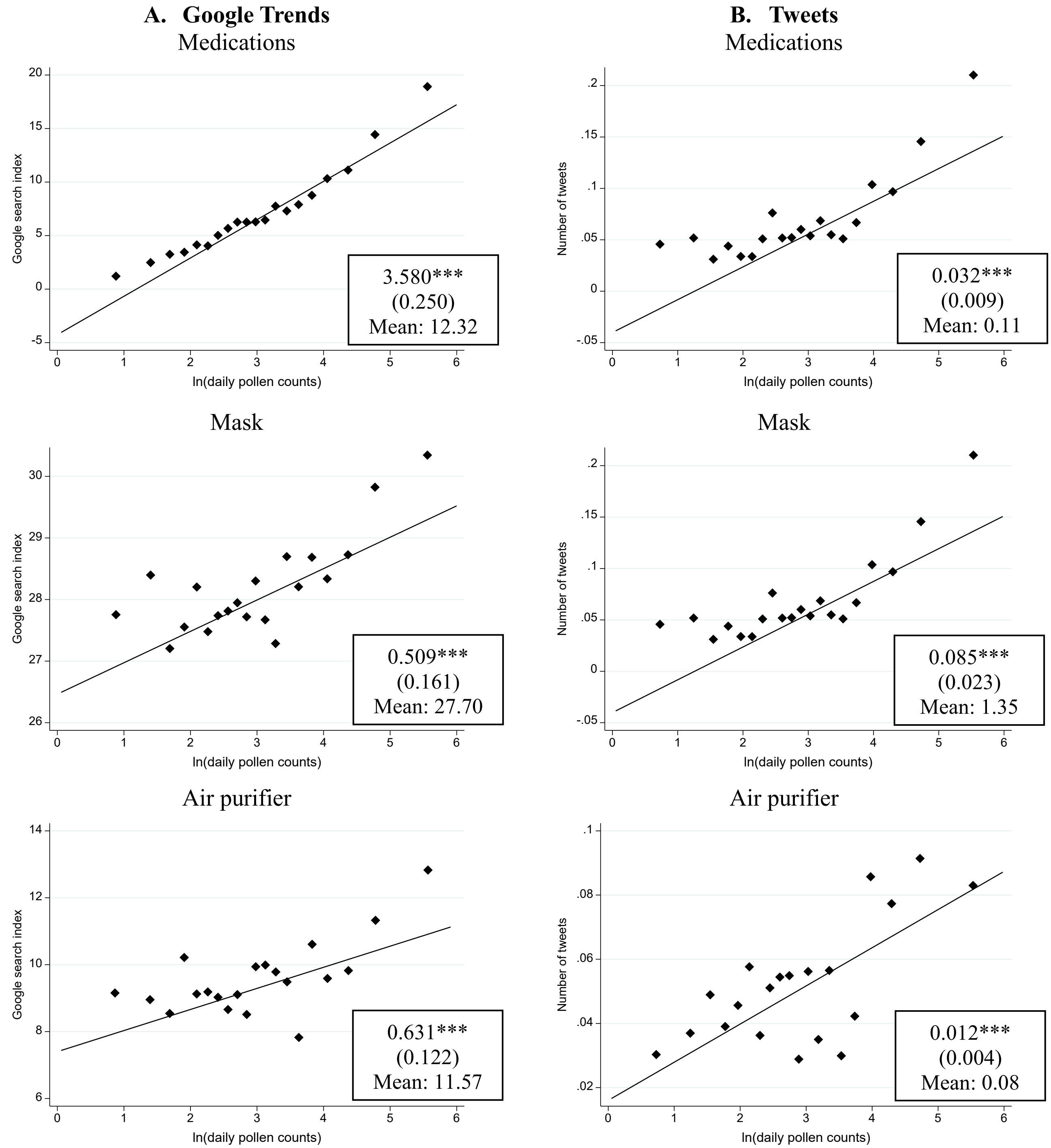
Table E3—Purchasing allergy products: Heterogeneity

	A. By age groups		B. By gender	
	Non-elderly	Elderly	Male	Female
	(1)	(2)	(3)	(4)
ln(pollen counts)	46.686***	43.916***	38.994***	49.415***
	(3.122)	(3.417)	(2.325)	(3.604)
R-squared	0.008	0.005	0.008	0.007
N	3,147,182	934,660	2,137,760	2,165,657
Mean of dep. var (in 10^{-3} \$)	306.88	316.50	248.98	341.97
Municipality FE	X	X	X	X
Year-prefecture FE	X	X	X	X
Week FE	X	X	X	X

Notes: The sample is derived from retail scanner data from February to May for the period 2012 to 2019, with observations at a weekly per-person level. The dependent variable is the weekly expenditure (in 10^{-3} \$) on all allergy products, using an exchange rate of 100 yen/\$. Estimates from the variant of equation [1] are reported along with standard errors clustered at the pollen monitoring station level in parentheses. In addition to the fixed effects listed in the table, weather covariates (precipitation, temperature, wind speed), and darkness are included. The elderly category comprises individuals aged 60 years and older, while the non-elderly category encompasses everyone else. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Appendix F: Avoidance behavior from Google Trends/Tweets

Figure F1—Pollen and Google Trends/Tweets



Notes: The samples are derived from Google Trends data for panel A and Twitter data for panel B for the period 2016 to 2019. The level of observation is at the prefecture level per day. The graphs display binscatter plots illustrating the relationship between logged daily pollen counts (grains/m^3, on the *x*-axis) and the Google search index in panel A, and the number of tweets in panel B for keywords related to medications (on the *y*-axis), masks, and air purifiers, after controlling for month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. See Appendix Table F1 for the list of search terms within each category. Estimates from the variant of equation [1], wherein unit fixed effects are replaced by prefecture fixed effects, are presented in the box. Standard errors, clustered at the prefecture level, are reported in parentheses. Estimates are weighted by the population in each prefecture per year. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table F1—List of search keywords

Category #	Categories	Japanese	English
1	**Medications** (product name of popular allergy medications)	アレジオン	Alesion
		アレグラ	Allegra
		クラリチン	Claritin
2	**Mask**	マスク	Mask
3	**Air purifier**	空気清浄機	Air purifier

Notes: The table lists the keywords for each category in both Japanese and English (for reference).

Appendix G: Avoidance Behavior from Cellphone Mobility Records

Our geolocation data, referred to as "Mobile Spatial Statistics" (MSS), is provided by NTT DOCOMO, Inc., Japan's largest mobile phone carrier. MSS utilizes the location information of 85 million NTT DOCOMO users (as of March 2022) to provide population estimates at a 500×500 meter mesh on an hourly basis across Japan. For detailed procedures on constructing these population estimates, see Terada et al. (2013).

Our dataset on mobility measures is structured as follows: Firstly, for each municipality, we select a 500×500 meter mesh with the highest number of establishments in the customer service industry (e.g., accommodations, restaurants, and entertainment) based on information from the 2016 Economic Census (MIC 2019).[2] This choice of the service industry aims to capture bustling areas such as business districts, shopping, and dining areas, which are more likely to represent the population engaged in outdoor activities. Secondly, we provide the list of these meshes to NTT DOCOMO, Inc., which returns the estimated population at each mesh for the period from February 2014 to May 2019. Thirdly, we aggregate the estimated population at the unit level by calculating the average across all municipalities within the unit. We specifically use the estimated population at 2 p.m. as the daily population in commercial areas tends to peak around this time (Seike et al. 2015). This measure serves as a proxy for engaging in outdoor activities, hereafter referred to as the "outdoor population," which we analyze to examine avoidance behavior.

Finally, we briefly discuss the advantages and disadvantages of this dataset. There are two types of geolocation data in Japan: the first originates from the leading smartphone mapping application in Japan, "Docomo Chizu NAVI," which collects GPS coordinates of each smartphone device whenever the device is turned on. A notable feature of this application is its ability to effectively track individuals over time (albeit only for recent years). Researchers can identify individuals' "home" locations as the most frequent locations of geographically contiguous stays (Miyauchi et al. 2021) and measure whether individuals leave their homes. However, the drawback is that the sample is limited to individuals who have granted permission to share their location information, leading to selection biases in both application users and those who consent, resulting in a relatively small sample size (545,000 users as of 2019). The second type of data, including ours, is based on the transmission of information from each mobile terminal (in our case, 85 million users) to base stations when mobile devices are turned on. This type of data offers a more nationally representative sample with broad spatial coverage of the entire country, albeit providing only hourly estimated population data for each area.

Given that the primary objective of this study is to provide nationally representative estimates of the effects of pollen exposure on accidents and corresponding avoidance behaviors over an extended period, we have chosen to analyze the latter dataset. Due to its representativeness and the extensive time span covered by the sample, this dataset has been widely utilized, particularly for measuring people's mobility during the COVID-19 pandemic (e.g., Kondo 2021; Kuroda et al. 2022).

---

[2] Due to budgetary constraints, our dataset comprises one mesh per municipality. Nevertheless, we affirm that our mobility measure effectively encompasses overall daytime outdoor activity, utilizing the 2019 data, wherein we possess outdoor population statistics for all meshes. Notably, our mobility measure derived from the representative mesh exhibits a correlation as high as 0.889 with the summary measure, which aggregates data from all meshes hosting at least one service establishment.

References

Kondo, K. 2021. "Simulating the Impacts of Interregional Mobility Restriction on the Spatial Spread of COVID-19 in Japan." *Scientific Reports*, 11: 18951.

Kuroda, Y., T. Sato, Y. Matsuda. 2022. "How Long do Voluntary Lockdowns Keep People at Home? The Role of Social Capital during the COVID-19 Pandemic." *Data Science and Service Research Discussion Paper No. 125.*

MIC (Ministry of Internal Affairs and Communications). 2019. "Economic Census." (in Japanese) https://www.e-stat.go.jp/gis/statmap-search?page=1&type=1&toukeiCode=00200553&toukeiYear=2016&aggregateUnit=H&serveyId=H002005112016&statsId=T000918 (accessed March 23, 2022).

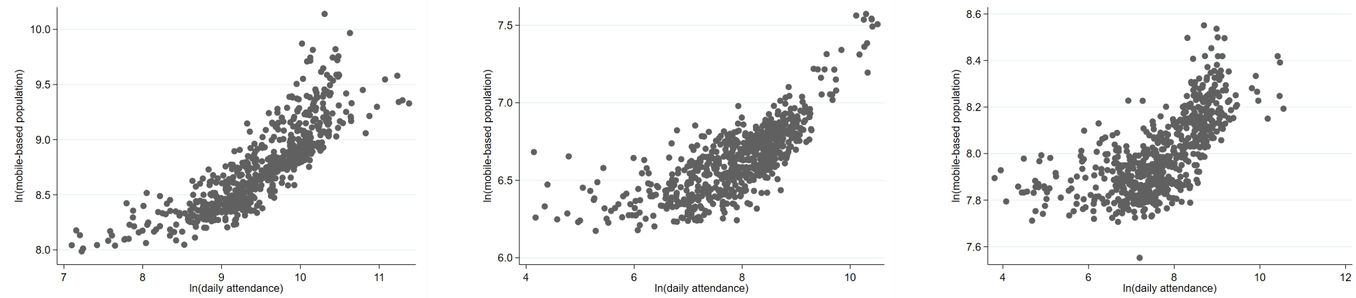
Miyauchi, Y., K. Nakajima, S.J. Redding. 2021. "The Economics of Spatial Mobility: Theory and Evidence Using Smartphone Data." *NBER Working Paper No. 28497.*

Seike, T., H. Mimaki, S. Morita. 2015. "Study on the Population Characteristics in a City Center District Utilizing Mobile Spatial Statistics" (in Japanese). *Journal of Architecture and Planning*, 80(713): 1625–1633.

Terada, M., T. Nagata, M. Kobayashi. 2013. "Mobile Spatial Statistics' Supporting Development of Society and Industry—Population Estimation Technology for Mobile Spatial Statistics." *NTT DOCOMO Technical Journal*, 14: 16–23.

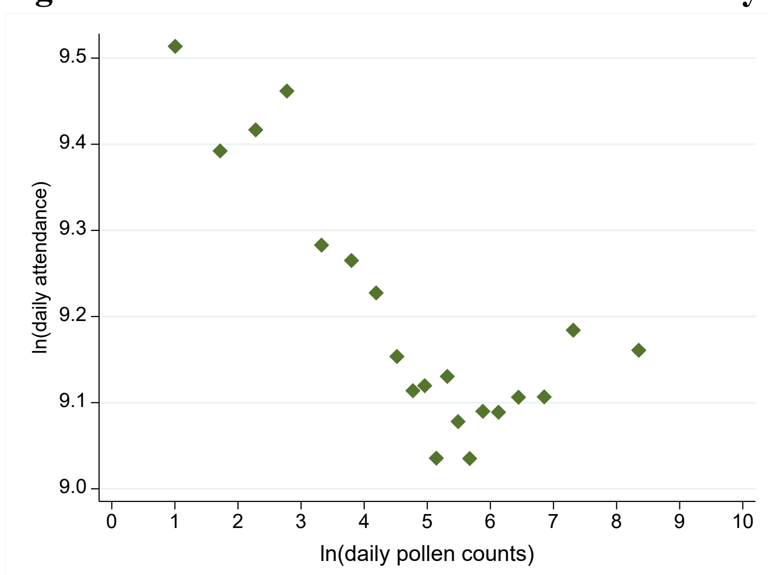
Figure G1—Actual Zoo attendance and population estimate from cellphone mobility records

A. Ueno Zoo B. Tama Zoological Park C. Inokashira Park Zoo



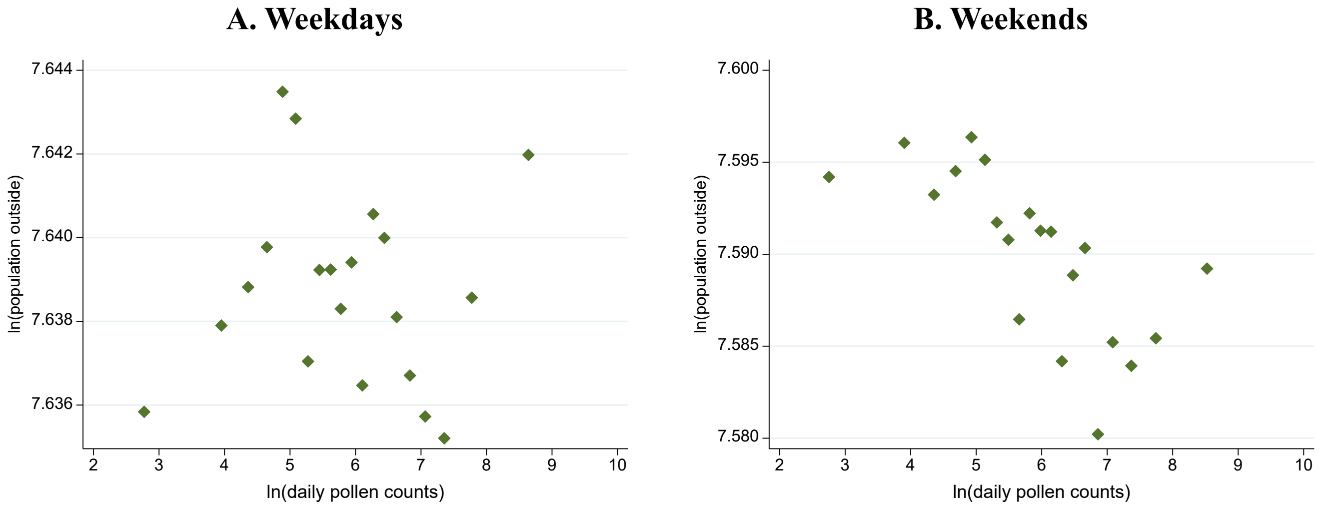
Notes: The sample is derived from attendance records from the three major zoos operated by the Tokyo Metropolitan Government and corresponding population estimates from cellphone mobility records for the period 2014 to 2019, with observations at the zoo per day level. The graphs display scatter plots illustrating the relationship between the logged daily attendance from admission records (on the *x*-axis) and the logged daily number of cell-phone–based population estimates at 2 p.m from location data (on the *y*-axis). Panels A–C present results for Ueno Zoo, Tama Zoological Park, and Inokashira Park Zoo, respectively. The correlations between the two variables in Panels A–C are 0.839, 0.756, and 0.673, respectively.

Figure G2—Pollen and zoo attendances in Tokyo



Notes: The sample is derived from admission records from the three major zoos operated by the Tokyo Metropolitan Government (Ueno Zoo, Tama Zoological Park, and Inokashira Park Zoo) for the period 2008 to 2019, with observations at the zoo per day level. The graphs display binscatter plots illustrating the relationship between the logged daily pollen counts (grains/m^3, on the *x*-axis) and the logged daily attendance (on the *y*-axis). The specification control for zoo and date fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and the logged average daily admissions for each zoo. Estimates are weighted by the average daily admissions in each zoo.

Figure G3—Pollen and avoiding going out



Notes: The sample is derived from cellphone mobility records for the period 2014 to 2019, with observations at the unit per day level (N= 343,454 for panel A and 135,399 for panel B). There are a total of 705 emergency response units. The graphs display binscatter plots illustrating the relationship between the logged daily pollen counts (grains/m^3, on the *x*-axis) and the logged daily number of people outdoors at 2 p.m. (on the *y*-axis). Panels A and B examine weekdays and weekends, respectively, controlling for month-by-year, month-by-prefecture, and day-of-week fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and logged population. Estimates are weighted by the population in each unit.

Table G1—Avoid attending zoos

	(1) All	(2) Weekdays	(3) Weekends	(4) Paid	(5) Free
ln(pollen counts)	-0.0332***	-0.0222*	-0.0473***	-0.0175	-0.0705***
	(0.0095)	(0.0120)	(0.0130)	(0.0161)	(0.0124)
R-squared	0.95	0.95	0.96	0.90	0.93
N	3,474	2,333	1,141	3,425	3,474
Zoo FE	X	X	X	X	X
Date FE	X	X	X	X	X

Notes: The sample is derived from admission records from the three major zoos operated by the Tokyo Metropolitan Government (Ueno Zoo, Tama Zoological Park, and Inokashira Park Zoo) for the period 2008 to 2019, with observations at the zoo per day level. The dependent variable is the logged daily attendance. Column (1) encompasses all days, while Columns (2) and (3) separately report estimates for weekdays and weekends. Columns (4) and (5) distinguish between paid and free admissions. The specification control for zoo and date fixed effects, alongside weather covariates (precipitation, temperature, wind speed), darkness, and the logged average daily admissions for each zoo. The average daily admissions in each zoo weight estimate. Robust standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table G2—Avoiding going out and accidents

	All (1)	Weekdays (2)	Weekends (3)
ln(outdoor population)	0.956**	0.403	2.011***
	(0.415)	(0.447)	(0.756)
R-squared	0.48	0.48	0.51
N	478,853	343,454	135,399
Unit FE	X	X	X
Day-of-week FE	X	X	X
Month-by-year FE	X	X	X
Prefecture-by-month FE	X	X	X

Notes: The sample is derived from the ambulance records for the period 2014 to 2019, which are matched with "Mobile Spatial Statistics" data provided by NTT DOCOMO, Inc. at the unit-day level (N= 478,853). A total of 705 emergency response units are available. We present estimates from regressing daily accidents per million people (our primary outcome) on the logged outdoor population, employing the same sets of fixed effects and controls as in equation [1], except for the logged number of pollen counts. Standard errors clustered at the pollen monitoring station level are reported in parentheses. Columns (2) and (3) restrict the samples to weekdays and weekends, respectively. Estimates are weighted by the population in each unit. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table G3—Avoiding going out with full covariates

Outcome: logged outdoor population

	A. All	B. By type of day	
		Weekdays	Weekends
	(1)	(2)	(3)
ln(pollen counts)	-0.0005	0.0000	-0.0021***
	(0.0006)	(0.0006)	(0.0007)
Rainfall (base: no rainfall)			
<1 mm	-0.0061***	-0.0059***	-0.0047**
	(0.0013)	(0.0011)	(0.0019)
1 mm≤ & <2 mm	-0.0167***	-0.0144***	-0.0208***
	(0.0035)	(0.0024)	(0.0080)
≥2 mm	-0.0167***	-0.0150***	-0.0249***
	(0.0045)	(0.0041)	(0.0082)
Mean temperature (base: <0 ℃)			
[0, 5) ℃	0.0018	0.0068**	0.0096
	(0.0033)	(0.0034)	(0.0064)
[5, 10) ℃	0.0111***	0.0139***	0.0189***
	(0.0036)	(0.0037)	(0.0067)
[10, 15) ℃	0.0089**	0.0120***	0.0160**
	(0.0043)	(0.0043)	(0.0070)
[15, 20) ℃	0.0061	0.0094*	0.0172**
	(0.0045)	(0.0049)	(0.0068)
[20, 25) ℃	0.0120***	0.0165***	0.0200***
	(0.0043)	(0.0046)	(0.0069)
≥ 25 ℃	0.0273***	0.0264***	0.0319***
	(0.0057)	(0.0066)	(0.0079)
Mean wind speed	-0.0022***	-0.0021***	-0.0023***
	(0.0005)	(0.0005)	(0.0007)
Darkness	-0.0084***	-0.0090***	-0.0073***
	(0.0014)	(0.0012)	(0.0023)
ln(population)	1.3108***	1.2630***	1.4497***
	(0.1975)	(0.1997)	(0.2074)
R-squared	0.98	0.99	0.99
N	478,853	343,454	135,399
Unit FE	X	X	X
Day-of-week FE	X	X	X
Month-by-year FE	X	X	X
Prefecture-by-month FE	X	X	X

Notes: The sample is derived from cellphone mobility records from February to May for the period 2014 to 2019, with observations at the unit per day level. There are a total of 705 emergency response units. Estimates from the variant of equation [1] are reported along with standard errors clustered at the pollen monitoring station level in parentheses. The dependent variable is the logged daily outdoor population at 2 p.m. Estimates are weighted by the population in each unit. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Table G4—Avoiding going out: Heterogeneity

Outcome: logged outdoor population

	A. Weekdays				B. Weekends			
	Non-elderly	Elderly	Male	Female	Non-elderly	Elderly	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(pollen counts)	-0.0007	0.0005	-0.0005	-0.0000	-0.0029***	-0.0013	-0.0025***	-0.0025***
	(0.0007)	(0.0006)	(0.0007)	(0.0006)	(0.0009)	(0.0009)	(0.0008)	(0.0009)
Rain (base: no rain)								
<1 mm	-0.0050***	-0.0121***	-0.0026**	-0.0104***	-0.0032	-0.0111***	-0.0019	-0.0086***
	(0.0011)	(0.0016)	(0.0011)	(0.0013)	(0.0020)	(0.0027)	(0.0019)	(0.0023)
1 mm≤ & < 2 mm	-0.0089***	-0.0283***	-0.0051**	-0.0220***	-0.0170**	-0.0365***	-0.0115	-0.0317***
	(0.0020)	(0.0040)	(0.0023)	(0.0027)	(0.0077)	(0.0122)	(0.0073)	(0.0094)
≥ 2 mm	-0.0095**	-0.0334***	-0.0043	-0.0271***	-0.0162*	-0.0466***	-0.0144**	-0.0332***
	(0.0042)	(0.0057)	(0.0041)	(0.0048)	(0.0086)	(0.0095)	(0.0072)	(0.0100)
R-squared	0.97	0.97	0.98	0.98	0.97	0.97	0.98	0.98
N	341,433	340,495	341,093	341,343	134,450	134,140	134,409	134,387
Unit FE	X	X	X	X	X	X	X	X
Day-of-week FE	X	X	X	X	X	X	X	X
Month-by-year FE	X	X	X	X	X	X	X	X
Prefecture-by-month FE	X	X	X	X	X	X	X	X

Notes: The sample is derived from cellphone mobility records from February to May for the period 2014 to 2019, with observations at the unit per day level. There are a total of 705 emergency response units. Estimates from equation [1] are reported along with standard errors clustered at the level of pollen monitoring station in parentheses. The dependent variable is the logged daily outdoor population at 2 p.m. In addition to the FEs and weather covariates outlined in the table, average wind speed, darkness, and logged population are also included. The elderly category comprises individuals aged 60 years and older, while the non-elderly category encompasses everyone else. Estimates are weighted by the population in each unit. Significance levels: *** p<0.01, ** p<0.05, * p<0.10.

Appendix H: Data Appendix

Data	Source
Ambulance records	Years: 2008–2019 (detailed diagnosis information is available from 2015 onwards) Data description: ambulance records archive Source: Fire and Disaster Management Agency (FDMA) of the Ministry of Internal Affairs and Communications https://www.fdma.go.jp/en/post1.html (last accessed March 19, 2026)
Police records	Years: 2019–2020 Data description: traffic accident records of accidents involving injuries or deaths Source: National Policy Agency (NPA) https://www.npa.go.jp/publications/statistics/koutsuu/opendata/index_opendata.html (last accessed March 19, 2026)
Pollen	Years: 2008–2020 Data description: hourly pollen counts from 120 stations, as well as hourly rainfall, temperature, wind speed, and wind direction from nearby weather stations during the pollen season (February to May except for Hokkaido, where the pollen season is March to June). Source: Ministry of the Environment (MOE), Pollen Monitoring System "Hanako-san" https://tenki.jp/pollen/ (last accessed March 19, 2026) *Note*: MOE terminated data collection of pollen counts in 2021.
Temperature	Years: 2007–2020 Data description: hourly temperature (outside of the pollen season) Source: Japan Automated Meteorological Data Acquisition System (AMeDAS) operated by the Japan Meteorological Agency (JMA) https://www.data.jma.go.jp/obd/stats/etrn/ (last accessed March 19, 2026)
Pollution	Years: 2009 April–2019 March Data description: hourly SO_2, NO, NO_2, CO, OX, PM_{10} Source: National Institute for Environmental Studies https://www.nies.go.jp/igreen/index.html (last accessed March 19, 2026)
Population	Years: 2008-2020 Data description: municipality-level population records from the Basic Resident Register Source: Ministry of Internal Affairs and Communications https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200241&tstat=000001039591&cycle=7&tclass1=000001039601&tclass2val=0 (in Japanese) (last accessed March 19, 2026)
Google Trends data	Years: 2016–2019 Data description: Google search index reflecting search term popularity for selected keywords, ranging from 0 to 100 in a given prefecture and day, proportional to total searches within the period.
Twitter data	Years: 2016–2019 Data description: the number of tweets that contain the selected keywords
Retail scanner data	Years: 2012–2019 Data description: called "Quick Purchase Report," which is the daily panel of purchase records from roughly 30,000 monitors

	Source: Macromill, Inc https://www.macromill.com/service/digital-data/consumer-purchase-history-data/ (in Japanese)
Cellphone mobility records	Years: 2014–2019 Data description: called "Mobile Spatial Statistics" data, which are estimates based on the location information of 85 million NTT DOCOMO cellphone users (as of March 2022) Source: NTT DOCOMO, Inc https://mobaku.jp/ (in Japanese) (last accessed March 19, 2026)
Zoo attendances	Years: 2008–2019 Data description: daily admission data from three major zoos (Ueno Zoo, Tama Zoological Park, and Inokashira Park Zoo) provided by the Tokyo Metropolitan Government