How does regional air pollution affect pedestrian volume, automobile traffic volume, and transit ridership?

Prachanda Tiwari
EBP US, Boston, MA, 02110
Email: prachanda.tiwari@ebp-us.com
ORCID: 0009-0009-6624-8854

Patrick A. Singleton
Department of Civil and Environmental Engineering
Utah State University, Logan, UT, 84322-4110
Email: patrick.singleton@usu.edu
ORCID: 0000-0002-9319-2333

Keunhyun Park
Department of Forest Resources Management
University of British Columbia, 2424 Main Mall, Vancouver, BC Canada V6T 1Z4
Email: keun.park@ubc.ca
ORCID: 0000-0001-5055-7833

Supplemental Information

Data and variables

Multimodal traffic volumes

Automobile traffic volume counts on various streets and highways were taken from continuous count stations (CCSs) maintained by the Utah Department of Transportation (UDOT). The stations record the number of cars, trucks, etc. crossing a given station by using sensor devices such as inductive loops and overhead microwave radar sensors. The UDOT counts provide the number of vehicles crossing each location per day for CCSs distributed throughout Utah. There were only six count stations located in Cache County which could be selected for the analysis. The data cover a two-year period from January 2018 through December 2019. To avoid complicating the analysis, this work does not consider the impacts of the COVID-19 pandemic.

Pedestrian traffic volumes came from a novel big data source: pedestrian push-button data obtained from high-resolution traffic signal controller logs. In the US, many if not most intersection traffic signals include push-button activated pedestrian detectors. In Utah, such real-time and archived data on pedestrian push-button presses are available from nearly all (2,000+) traffic signals throughout the state (UDOT, 2023). While not a perfect measurement of pedestrian activity—one person can press multiple times, or one person can press for a group of people—a recent research project found that push-button data could be successfully used to estimate pedestrian crossing volumes (Singleton et al., 2020). Those authors compared push-button data with ground-truth pedestrian volumes collected from over 10,000 hours of video at 90 signalized intersections throughout Utah, and developed a set of simple regression models to convert push-button data to estimated pedestrian crossing volumes. One successful technique was to ignore button-presses that occurred within 15 seconds of a previous button-press, to avoid over-counting multiple-press behaviors. Details of these prediction methods are provided elsewhere (Runa & Singleton, 2021), but the methods had good accuracy (correlation of 0.84 between observed and predicted volumes, mean error of ±3.0 pedestrians per hour). For this project, we used daily estimates of pedestrian volumes at 39 signals in/near Logan, the major city in Cache County, for the same two-year time period.

Bus ridership data were obtained from the public transportation service provider operating in the study area. The Cache Valley Transit District (CVTD) provided the total daily bus ridership (boardings) across all of their bus routes for each day throughout the study period. There is no rail-based public transportation service in Cache County, and no service on Sundays. Unfortunately, stop-level boarding data were not available from the agency. Therefore, it is important to note that the public transportation data has a different structure than the automobile/pedestrian data, as it captures the area-wide bus ridership rather than any stop- or route-specific ridership.

While the decision to use system-level transit ridership was largely the result of data limitations, we also did this because route-level or even stop-level public transportation data is conceptually slightly different than the automobile/pedestrian data used to analyze those modes. Automobile traffic volumes count every automobile trip passing that location, and pedestrian traffic volumes estimate (approximately) every person crossing the street at that intersection. In contrast, stop-level boarding and alighting data would only capture the trips starting or ending near that particular stop. Route-level ridership data would count trips that started/ended anywhere along that route, complicating and potentially obscuring spatial details. Thus, to maintain clarity and avoid any inconsistency in data used for analysis across each mode (and due to data limitations), we opted to use system-wide bus ridership data. We acknowledge that doing this prevented us from studying locational variations in the effect of air pollution on bus use, and thus answering our second research objective for this mode.

Air pollution, weather, and temporal control variables

Daily air quality information (air quality index, based on concentrations of particulate matter) was collected from sensors and was obtained from the US Environmental Protection Agency (EPA). In 2012, the Utah Division of Air Quality (UDAQ) revamped its air quality categorization in line with the EPA standard and created six color-coded categories. The Air Quality Index (AQI) is representative of the pollution due to ozone, particulate matter, and oxides of nitrogen, sulfur, and carbon. The categories are described in Table SI-1. In Utah, news reports, air quality apps, and recommended government actions all utilize these AQI categories and colors (US EPA, n.d.; Utah DEQ, 2022; UDOT, 2022; Williams, 2023). For our study, $PM_{2.5}$ data from a monitoring station in Smithfield (a suburb of Logan) was used. During the study period, the highest daily AQI value was 140, so only three-color categories (green, yellow, and orange) were considered in our analysis.

Table SI-1: Air Quality Index (AQI) (US EPA, n.d.)

Color	*AQI range*	*Health concern*	*Description*
Green	0–50	Good	Air quality is satisfactory, and air pollution poses little or no risk.
Yellow	51–100	Moderate	Air quality is acceptable. However, there may be a risk for some people, particularly those who are unusually sensitive to air pollution.
Orange	101–150	Unhealthy for Sensitive Groups	Members of sensitive groups may experience health effects. The general public is less likely to be affected.
Red	151–200	Unhealthy	Some members of the general public may experience health effects; members of sensitive groups may experience more serious health effects.
Purple	201–300	Very Unhealthy	Health alert: The risk of health effects is increased for everyone.
Maroon	301–500	Hazardous	Health warning of emergency conditions: everyone is more likely to be affected.

To control for atmospheric environmental impacts on multimodal traffic volumes (thus helping to isolate the unique influence of air pollution), daily weather data about precipitation, snow, temperature, etc. were obtained from the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NCEI). Specifically, weather data for all signals were obtained from the weather station (USC00425186) located at Utah State University in Logan. The station reported daily precipitation (in mm), snowfall (in mm), and maximum and minimum temperature (in °C). Since precipitation included a mix of all kinds of precipitation (rain and snow), a combined precipitation variable was created with the following categories: no rain and no snow, light rain (1–25mm), light snow (1–50mm), heavy rain (>25mm), and heavy snow (>50mm). Also, from a nearby weather station (USW00094128) located at the Logan–Cache Airport, a dataset containing historical temperature for the last 30 years was obtained. A maximum temperature difference variable was created as a measure of how hot a day was compared to the 30-year average temperature on the same day.

Besides the weather controls, three additional control variables were introduced to account for temporal variations in multimodal traffic volumes. A seasonal categorical variable was created which distributed the 12 months into four seasons. Days of the week were categorized into Saturday, Sunday, and weekdays to control for the effects of weekends on traffic. Also, holidays in the state of Utah during the study period were identified (Office Holidays, n.d.). We checked the day-of-week distributions of multimodal traffic volumes, and volumes tended to be much more consistent within weekdays than they were for Saturdays, Sundays, and holidays; therefore, we did not split out, for example, Monday and Friday from Tuesday, Wednesday, and Thursday.

Count station-level variables

In order to measure variations in the air quality–traffic volume relationship across locations, we also collected built and sociodemographic environment variables at each pedestrian and automobile traffic volume location. A quarter-mile buffer was created around each location. This buffer radius was a subjective decision by the authors, but was informed by research on the built environment influences of walking (Ewing & Clemente, 2013); a quarter-mile buffer was also used in two recent pedestrian-related studies in Utah (Park et al., 2023; Singleton et al., 2021). Information regarding population and employment density, commercial and residential land uses, bus stops, park coverage, and schools were calculated from the EPA's Smart Location Database. Similarly, sociodemographic attributes, including median household income and mean car ownership, were obtained from the American Community Survey (ACS) 2013-2017 and processed using the quarter-mile buffers.

Descriptive statistics of the variables in the compiled data are shown in Table SI-2. Note that the table only reports the built and sociodemographic environment for the 39 pedestrian count locations, since the subsequent analysis (see Results) found no locational variations in relationships for automobile traffic count stations. Remember, bus ridership was measured system-wide, not for particular locations.

Table SI-2: Explanation of the variables and descriptive statistics

Variable	*Mean*	*SD*	*#*	*%*
Multimodal traffic volumes				
Pedestrian traffic volumes	379	1,033		
(N = 27,157 = 39 locations × 730 days – missing data)				
Automobile traffic volumes	12,489	8,410		
(N = 3,987 = 6 stations × 730 days – missing data)				
Bus ridership	4,708	1,991		
(N = 608 = 1 system × 730 days – Sundays – missing data)				
Temporal variables (730 days)				
Day of Week: Weekday			522	71.5
Saturday			104	14.2
Sunday			104	14.2
Holiday: False			706	96.7
True			24	3.3
Season: Winter			180	24.7
Spring			184	25.2
Summer			184	25.2
Fall			182	24.9
Precipitation: No rain / no snow			532	73.0
Light rain (1–25mm)			117	16.0
Heavy rain (>25mm)			2	0.3
Light snow (1–50mm)			57	7.8
Heavy snow (>50mm)			21	2.9
Max temperature (°C) difference from average	0.04	4.73		
Air quality index: Green (AQI = 0–50)			626	85.7
Yellow (AQI = 51–100)			88	12.1
Orange (AQI = 101–150)			16	2.2
Built and social environment variables (39 pedestrian volume locations, ¼-mile buffer)				
Percentage of residential parcels	20.0	13.3		
Percentage of commercial parcels	33.1	17.1		
Percentage of vacant land	6.6	4.0		
Population density (1,000 people/mi^2)	5.0	2.1		
Employment density (1,000 jobs/mi^2)	9.5	6.3		
Intersection density (#/mi^2)	88.0	37.1		
% 4-way intersections	44.6	21.4		
Number of bus stops	6.1	3.7		
Number of schools	0.2	0.5		
Park acreage	1.1	2.8		
Household income (median, $1,000)	37.4	9.0		
Car ownership (mean)	1.6	0.3		

Analysis methods

Since the dataset for multimodal traffic volumes covered multiple locations and across a span of two years, multilevel modeling was an appropriate approach for our analysis. Multilevel models can adequately represent the two-level nature of our data: daily traffic volumes Y_{ij} (level one), nested within locations (level two). Such models also allow clear specifications of variations in model coefficients at level one (across level two units j), including fixed and random intercepts (β_{0j}), slopes (β_{hj}) for h level-one variables (x_{ij}), and cross-level interactions in which level-two variables (z_j) affect level-one slopes. In other words, multilevel models can represent variations in the air quality–traffic volume relationship (slope) across locations and due to locational characteristics. A simple multilevel model with level-one residuals R_{ij} is represented in the following Eq. 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + R_{ij} \quad (1)$$

In line with the first study objective—to examine the relationship between air quality and traffic volume for each mode—we estimated separate multilevel models for automobile traffic volumes and for pedestrian volumes. Dependent variables (Y_{ij}) were the natural log of the daily traffic volume, and independent (level one) variables (x_{hij}) were daily air quality, weather, and temporal controls. Different specifications for air quality were considered, but the best-fitting and most intuitive results were found for dummy variables representing the green, yellow, and orange AQI categories (Table SI-1). For both modes, we allowed the intercept (but not the slopes) to vary across locations. For pedestrian volumes (39 locations), we used a random effects intercept model (Eq. 2), in which the intercept coefficient varied randomly following a normal distribution for level-two residuals U_{0j}. For automobile traffic volumes (6 locations), the few sites meant we used a fixed effects intercept model (Eq. 3), in which a different intercept coefficient was estimated for each station k.

$$Y_{ij} = \beta_{0j} + \textstyle\sum_h \beta_h x_{hij} + R_{ij} \quad \text{(2a), where}$$
$$\beta_{0j} = \gamma_{00} + U_{0j} \quad \text{(2b).}$$

$$Y_{ij} = \beta_{0j} + \textstyle\sum_h \beta_h x_{hij} + R_{ij} \quad \text{(3a), where}$$
$$\beta_{0j} = \textstyle\sum_k \gamma_{0k} D_k \quad \text{(3b), and}$$

D_k is a dummy variable equal to 1 for station k and 0 otherwise.

To address the study's second objective—exploring variations across locations in the effect of area-wide air pollution on multimodal traffic volumes—we first modified the first objective models and allowed slopes for the air quality dummy variables to vary across count stations. Again, for pedestrian volumes, this was a random effects slope model (Eq. 4), in which the random coefficients were normally distributed; for automobile traffic volumes, this was a fixed effects slope model (Eq. 5), in which different coefficients were estimated for each station. If the slopes were found to vary across locations—measured using likelihood-ratio tests versus the models for the first objective—we then tested whether g level-two location characteristics (z_{gj}) were significant in predicting the intercept and air quality slope variations across locations. In the terminology of multilevel modeling, these effects are called cross-level interactions (γ_{gh}), because they result in an interaction of a level-two variable (built or social environment) with a level-one variable (air quality). Only variables with significant interaction coefficients were retained in the final models.

$$Y_{ij} = \beta_{0j} + \sum_h \beta_{hj} x_{hij} + R_{ij} \quad \text{(4a), where}$$
$$\beta_{0j} = \gamma_{00} + \sum_g \gamma_{g0}\, z_{gj} + U_{0j} \quad \text{(4b), and}$$
$$\beta_{hj} = \gamma_{h0} + \sum_g \gamma_{gh}\, z_{gj} + U_{hj} \quad \text{(4c).}$$

$$Y_{ij} = \beta_{0j} + \sum_h \beta_{hj} x_{hij} + R_{ij} \quad \text{(5a), where}$$
$$\beta_{0j} = \sum_k \gamma_{0k} D_k \quad \text{(5b),}$$
$$\beta_{hj} = \sum_k \gamma_{hk} D_k \quad \text{(5c), and}$$

D_k is a dummy variable equal to 1 for station k and 0 otherwise.

For bus ridership, in line with the first objective to examine the relationship of air quality and traffic volumes, we estimated a simple linear regression model as represented by Eq. 6. The dependent variable (Y_{ij}) was the natural log of the daily total system-wide bus ridership, and the independent variables (x_i) were air quality, weather, and temporal controls. Because of the nature of the public transportation data (system-wide, not location-specific), we could not address the study's second objective for bus ridership.

$$Y_i = \beta_0 + \beta_1 x_i + R_i \quad \text{(6)}$$

Model estimation was performed using the "lme4" package (Bates et al., 2015) in R (The R Foundation, n.d.).

Results

Pedestrian volumes

Table SI-3 reports the results of the random intercept model for pedestrian volumes. The coefficient estimates for both the yellow (β = -0.053, *SE* = 0.011, *t* = -4.916, *p* = <0.001) and orange air quality days (β = -0.136, *SE* = 0.023, *t* = -5.929, *p* = <0.001) were negative and significant. This implies that pedestrian volumes decreased during episodes of poor air quality (compared to green days), especially on orange days (unhealthy for sensitive groups).

Table SI-3: Random intercept model for pedestrian volumes

Coefficients	*Estimate*	*SE*	*df*	*t-statistic*	*p-value*
Intercept	5.092	0.157	38.23	32.356	<0.001
Day of week (ref. = Weekday)					
Saturday	-0.366	0.009	27110	-38.513	<0.001
Sunday	-1.020	0.009	27110	-107.919	<0.001
Holiday (ref. = No holiday)	-0.914	0.019	27110	-49.317	<0.001
Season (ref. = Winter)					
Spring	0.266	0.011	27110	24.874	<0.001
Summer	0.373	0.010	27110	36.192	<0.001
Fall	0.361	0.011	27110	34.074	<0.001
Precipitation (ref. = No rain / no snow)					
Light rain	-0.060	0.009	27110	-6.293	<0.001
Heavy rain	-0.157	0.062	27110	-2.521	0.012
Light snow	-0.259	0.013	27110	-19.657	<0.001
Heavy snow	-0.341	0.020	27110	-16.968	<0.001
Max temperature difference from average	0.007	0.001	27110	9.411	<0.001
Air quality index (ref. = Green)					
Yellow (AQI = 51–100)	-0.053	0.011	27110	-4.916	<0.001
Orange (AQI = 101–150)	-0.136	0.023	27110	-5.929	<0.001

Notes: N = 27,157; # groups = 39; log-likelihood = -21,661; between-group variance = 0.963; residual variance = 0.286.

Table SI-4 reports the results of the random intercept and random slope model for pedestrian volumes. By estimating an earlier model (not shown), we found that there were significant random slopes for the air quality variables: a likelihood-ratio test found that the random intercept and slope model (log-likelihood = -21,656) was (marginally) significantly ($\chi^2 = 9.924$, $df = 5$, $p = 0.077$) better fitting than the random intercept only model (log-likelihood = -21,661). Therefore, we estimated several models, each testing cross-level interactions with air quality involving built and social environment variables. As shown in Table SI-4, there were significant interaction effects for three variables: the percentage of commercial parcels, the percentage of 4-way intersections and average car ownership. For the commercial land use variable, there was a positive and significant interaction term with yellow days ($\beta = 0.001$, $SE = 0.001$, $t = 2.072$, $p = 0.042$) but not orange days. This implies that the negative effect of yellow air quality days on pedestrian volumes was attenuated in places with more commercial land uses. For the intersection variable, there was a positive and significant interaction term with orange days ($\beta = 0.003$, $SE = 0.001$, $t = 2.004$, $p = 0.050$) but not yellow days. This implies that the negative effect of orange air quality days on pedestrian volumes (see Table SI-3) was attenuated in places with a greater share of 4-way intersections. For the car ownership variable, there was a negative and marginally significant interaction term with yellow days ($\beta = -0.076$, $SE = 0.040$, $t = -1,935$, $p = 0.057$). This implies that the negative effect of yellow air quality days on pedestrian volumes (see Table SI-3) was enhanced in places with greater average household car ownership.

Table SI-4: Random intercept and random slope model for pedestrian volumes

Coefficients	*Estimate*	*SE*	*df*	*t-statistic*	*p-value*
Intercept	-0.105	1.207	33.02	-0.087	0.931
Day of week (ref. = Weekday)					
Saturday	-0.366	0.009	27070	-38.534	<0.001
Sunday	-1.020	0.009	27060	-107.983	<0.001
Holiday (ref. = No holiday)	-0.914	0.019	27060	-49.341	<0.001
Season (ref. = Winter)					
Spring	0.266	0.011	27070	24.897	<0.001
Summer	0.373	0.010	27070	36.218	<0.001
Fall	0.361	0.011	27070	34.095	<0.001
Precipitation (ref. = No rain / no snow)					
Light rain	-0.060	0.009	27060	-6.304	<0.001
Heavy rain	-0.157	0.062	27060	-2.525	0.012
Light snow	-0.260	0.013	27060	-19.678	<0.001
Heavy snow	-0.341	0.020	27060	-16.976	<0.001
Max temperature difference from average	0.007	0.001	27070	9.417	<0.001
Air quality index (ref. = Green)					
Yellow (AQI = 51–100)	0.008	0.087	69.73	0.096	0.924
Orange (AQI = 101–150)	-0.170	0.195	54.95	-0.871	0.387
Built and social environment variables					
Percentage of commercial parcels	0.007	0.007	32.15	0.877	0.387
Population density (1,000 people/mi^2)	0.322	0.071	29.76	4.532	<0.001
Intersection density (#/mi^2)	0.008	0.004	30.03	2.029	0.051
% 4-way intersections	0.003	0.007	33.00	0.388	0.700
Number of bus stops	0.080	0.035	29.65	2.253	0.032
Number of schools	-0.496	0.229	29.92	-2.172	0.038
Household income (median, $1,000)	0.049	0.016	30.00	3.037	0.005
Car ownership (mean)	0.239	0.411	31.42	0.582	0.564
Cross-level interactions					
Yellow AQI with % commercial parcels	0.001	0.001	71.71	2.072	0.042
Orange AQI with % commercial parcels	0.002	0.002	55.14	1.311	0.195
Yellow AQI with % 4-way intersections	0.000	0.001	69.94	0.606	0.546
Orange AQI with % 4-way intersections	0.003	0.001	54.63	2.004	0.050
Yellow AQI with Car ownership	-0.076	0.040	70.24	-1.935	0.057
Orange AQI with Car ownership	-0.091	0.089	56.87	-1.020	0.312

Notes: N = 27,157; # groups = 39; log-likelihood = -21,622; between-group variance = 0.408; residual variance = 0.285; random coefficient variance for yellow AQI = 0.001; random coefficient variance for orange AQI = 0.005.

Because cross-level interaction terms are difficult to interpret in any type of regression model and even more difficult when they affect random slope coefficients, we also calculated what are called "posterior slopes" (Snijders & Bosker, 2015). Since the random air quality coefficients are not estimated by the model (just their mean and standard deviation), we used empirical Bayes estimation to let the model and data give us the "best" estimate of each location's slope coefficients. See a multilevel modeling textbook (Snijders & Bosker, 2015) for details on this calculation. Since the air quality coefficients were also interacted with built and social environment variables, we then multiplied each location's values for these level-two variables with their respective coefficients, and added them to the random portion obtained through empirical Bayes estimation to get the total value of the posterior slopes for yellow and orange air quality days (vs. green days).

Figure SI-1 plots these posterior slopes, first in a scatterplot (yellow vs. orange) and second in a combined plot vs. AQI. The left portion of the figure shows how most locations had a more negative orange coefficient than yellow coefficient (below the diagonal in the lower left quadrant), and how the posterior slopes were positively correlated (which is expected, since they are both conditional on the same data at each location). The right portion of the figure shows how air quality coefficients in the orange range (AQI = 101–150) are typically more extreme (mostly more negative) than coefficients in the yellow range (AQI = 51–100). In both portions of Figure SI-1, it appears that only a couple of locations had positive coefficients for yellow or orange AQI.

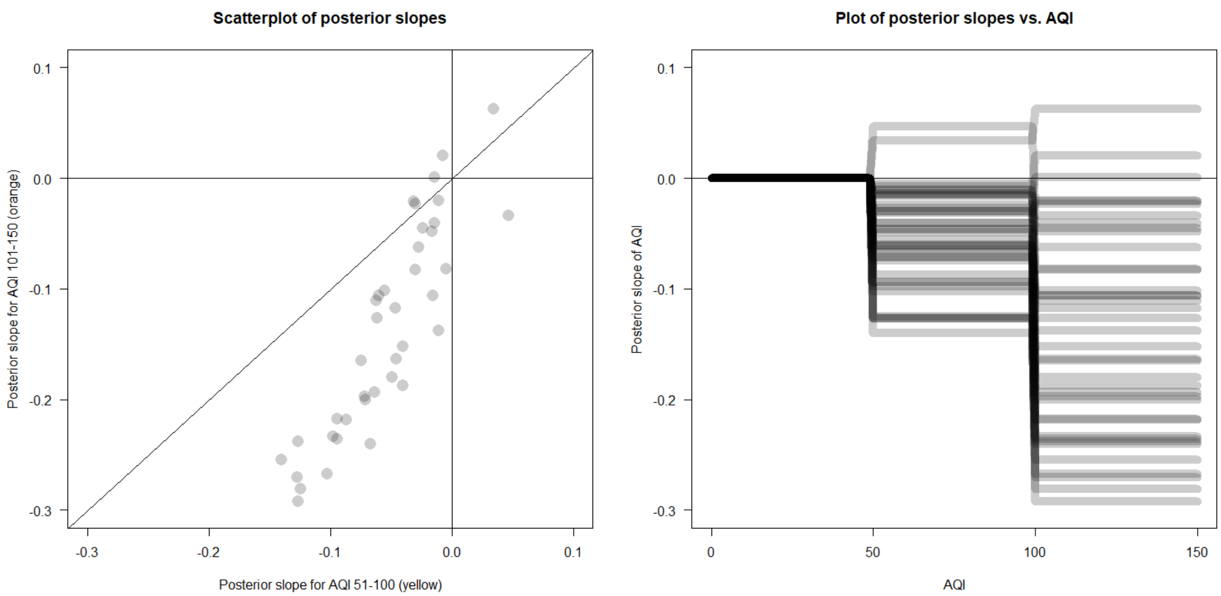

Figure SI-1: Figures showing posterior slopes for pedestrian volumes for yellow (AQI = 51–100) and orange (AQI = 101–150) air quality levels (left: scatterplot; right: plot vs. AQI).

Automobile traffic volumes

Table SI-5 reports the results of the fixed intercept model for automobile traffic volumes. One of the air quality variables (orange) was positively and significantly associated with automobile traffic volumes ($\beta = 0.049$, $SE = 0.015$, $t = 3.333$, $p = 0.001$). The positive association implies that driving increased during unhealthy (orange) air quality days when compared to days with good (green) air quality. The coefficient for yellow air quality was not significantly different from zero, implying no detectable difference in automobile traffic volumes on yellow (moderate) versus green air quality days.

Table SI-5: Fixed intercept model for automobile traffic volumes

Coefficient	*Estimate*	*SE*	*t-statistic*	*p-value*
Intercept (station 301)	9.060	0.008	1207.460	<0.001
Difference for station 363	1.084	0.007	147.750	<0.001
Difference for station 510	-0.663	0.007	-91.996	<0.001
Difference for station 511	-0.400	0.007	-55.207	<0.001
Difference for station 620	0.219	0.007	29.726	<0.001
Difference for station 622	0.946	0.007	129.548	<0.001
Day of week (ref. = Weekday)				
Saturday	-0.122	0.006	-19.753	<0.001
Sunday	-0.614	0.006	-98.842	<0.001
Holiday (ref. = No holiday)	-0.320	0.012	-27.073	<0.001
Season (ref. = Winter)				
Spring	0.097	0.007	14.267	<0.001
Summer	0.135	0.007	19.693	<0.001
Fall	0.109	0.007	16.172	<0.001
Precipitation (ref. = No rain / no snow)				
Light rain	-0.021	0.006	-3.367	0.001
Heavy rain	-0.024	0.039	-0.620	0.535
Light snow	-0.062	0.008	-7.399	<0.001
Heavy snow	-0.123	0.013	-9.742	<0.001
Max temperature difference from average	0.000	0.000	0.348	0.728
Air quality index (ref. = Green)				
Yellow (AQI = 51–100)	-0.003	0.007	-0.474	0.636
Orange (AQI = 101–150)	0.049	0.015	3.333	0.001

Notes: N = 3,987; adjusted R-squared = 0.963.

It is possible that some of the positive association between air quality category and automobile traffic volumes that we found in Table SI-5 could be the result of a different cause-and-effect relationship. Specifically, more driving could cause worse air pollution. (Thanks to a reviewer for emphasizing this possibility.) Although our analysis did not test this opposite direction of causality, future work should examine these inter-dependent relationships using more temporally fine-grained data and/or dynamic models with time lags. Nevertheless, we suspect that the share of the measured association (Table SI-5) due to this explanation is likely small, because air pollution is affected by many factors beyond transportation emissions (including wildfire smoke, agricultural emissions, and atmospheric conditions like temperature inversions).

Table SI-6 reports the results of the fixed intercept and fixed slope model for automobile traffic volumes, which involved interaction terms included between the air quality categories and each station. None of the air quality–station interaction terms were significant ($p > 0.10$), which implies that there was no significant difference in the relationship between air quality and automobile traffic volumes across the six count stations. Because no significant slope variation was detected, we did not estimate a subsequent model to predict this variation from built and social environment variables.

Table SI-6: Fixed intercept and fixed slope model for automobile traffic volumes

Coefficients	*Estimate*	*SE*	*t-statistic*	*p-value*
Intercept (Station 301)	9.060	0.008	1170.678	<0.001
Difference for Station 363	1.084	0.008	136.704	<0.001
Difference for Station 510	-0.660	0.008	-84.793	<0.001
Difference for Station 511	-0.399	0.008	-51.036	<0.001
Difference for Station 620	0.217	0.008	27.182	<0.001
Difference for Station 622	0.943	0.008	120.087	<0.001
Day of week (ref. = Weekday)				
Saturday	-0.122	0.006	-19.745	<0.001
Sunday	-0.614	0.006	-98.815	<0.001
Holiday (ref. = No holiday)	-0.320	0.012	-27.074	<0.001
Season (ref. = Winter)				
Spring	0.097	0.007	14.302	<0.001
Summer	0.135	0.007	19.721	<0.001
Fall	0.109	0.007	16.205	<0.001
Precipitation (ref. = No rain / no snow)				
Light rain	-0.021	0.006	-3.355	0.001
Heavy rain	-0.024	0.039	-0.618	0.536
Light snow	-0.062	0.008	-7.364	<0.001
Heavy snow	-0.123	0.013	-9.733	<0.001
Max temperature difference from average	0.000	0.000	0.360	0.719
Air quality index (ref. = Green)				
Yellow (AQI = 51–100) (Station 301)	-0.011	0.016	-0.704	0.482
Difference for Station 363	0.004	0.022	0.167	0.868
Difference for Station 510	-0.008	0.022	-0.352	0.725
Difference for Station 511	0.004	0.022	0.183	0.855
Difference for Station 620	0.019	0.022	0.861	0.390
Difference for Station 622	0.033	0.023	1.412	0.158
Orange (AQI = 101–150) (Station 301)	0.079	0.036	2.174	0.030
Difference for Station 363	-0.031	0.051	-0.604	0.546
Difference for Station 510	-0.076	0.050	-1.534	0.125
Difference for Station 511	-0.058	0.050	-1.153	0.249
Difference for Station 620	-0.018	0.050	-0.367	0.714
Difference for Station 622	0.009	0.050	0.184	0.854

Notes: N = 3,987; adjusted R-squared = 0.963.

Bus ridership

Table SI-7 reports the results of the linear regression model for bus ridership. It should be noted that we did not run a multilevel model for our public transportation data because we did not have location-specific data, only system-level bus ridership. Since the transit service provider (CVTD) did not operate during Sundays, this variable's estimates are missing from the model. The estimates for both the yellow and orange air quality days were found to be negative but were not statistically significant.

Table SI-7: Linear regression model for bus ridership

Coefficients	*Estimate*	*SE*	*t-statistic*	*p-value*
Intercept	8.686	0.026	332.388	<0.001
Day of week (ref. = Weekday)				
Saturday	-1.245	0.025	-49.883	<0.001
Holiday (ref. = No holiday)	-1.238	0.066	-18.829	<0.001
Season (ref. = Winter)				
Spring	-0.078	0.031	-2.512	0.012
Summer	-0.394	0.030	-13.292	<0.001
Fall	0.057	0.031	1.839	0.066
Precipitation (ref. = No rain / no snow)				
Light rain	-0.032	0.027	-1.185	0.237
Heavy rain	0.144	0.231	0.626	0.532
Light snow	-0.069	0.037	-1.843	0.066
Heavy snow	-0.041	0.058	-0.697	0.486
Max temperature difference from average	0.000	0.002	0.065	0.949
Air quality index (ref. = Green)				
Yellow (AQI = 51–100)	-0.017	0.031	-0.556	0.578
Orange (AQI = 101–150)	-0.075	0.063	-1.186	0.236

Notes: N = 608; adjusted R-squared = 0.836.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Ewing, R., & Clemente, O. (2013). *Measuring urban design: Metrics for livable places*. Springer.

Office Holidays. (n.d.). *Calendars of public holidays and bank holidays*. Office Holidays. Retrieved October 7, 2020, from https://www.officeholidays.com

Park, K., Singleton, P. A., Brewer, S., & Zuban, J. (2023). Pedestrians and the built environment during the COVID-19 pandemic: changing relationships by the pandemic phases in Salt Lake County, Utah, USA. *Transportation Research Record: Journal of the Transportation Research Board*, *2677*(4), 448-462. https://doi.org/10.1177/03611981221083606

Runa, F., & Singleton, P. A. (2021). Assessing the impacts of weather on pedestrian signal activity at 49 signalized intersections in northern Utah. *Transportation Research Record: Journal of the Transportation Research Board, 2675*(6), 0361198121994111. https://doi.org/10.1177/0361198121994111

Singleton, P. A., Park, K., & Lee, D. H. (2021). Varying influences of the built environment on daily and hourly pedestrian crossing volumes at signalized intersections estimated from traffic signal controller event data. *Journal of Transport Geography, 93*, 103067. https://doi.org/10.1016/j.jtrangeo.2021.103067

Singleton, P. A., Runa, F., & Humagain, P. (2020). *Utilizing archived traffic signal performance measures for pedestrian planning and analysis* (UT-20.17). Utah Department of Transportation. https://rosap.ntl.bts.gov/view/dot/54924

The R Foundation. (n.d.). *The R project for statistical computing*. https://www.r-project.org/

United States Environmental Protection Agency (US EPA). (n.d.). AirNow. https://www.airnow.gov

US EPA. (2023). Research on health effects from air pollution. https://www.epa.gov/air-research/research-health-effects-air-pollution

Utah Department of Environmental Quality (Utah DEQ). (2022). *Air Quality Forecast: Forecast Legend*. https://air.utah.gov/forecastLegendAQI.html

Utah Department of Transportation (UDOT). (2022). *TravelWise: Rethink Your Trip*. https://travelwise.utah.gov/

UDOT. (2023). Automated traffic signal performance measures (ATSPM). https://udottraffic.utah.gov/atspm/

Williams, C. (2023, November 29). Tired of Utah's latest inversion? Relief is in sight as more snow arrives. *KSL*. https://www.ksl.com/article/50803860/tired-of-utahs-latest-inversion-relief-is-in-sight-as-more-snow-arrives