APPENDIX A: Recreation dataset construction

This section discusses the construction of the recreation data in greater depth. In the raw Recreation.gov data, each record is a transaction. Transactions are grouped into orders, each of which with one or more transactions. For example, a single order might contain the following transactions, in order of transaction time: Registration/Walk-in, Make Payment, Change Number of Vehicles, Extend Stay Leave Later, Change Number of People, Checkout. Each transaction includes the date and time, campground or facility, unique user identifier (retained across orders), user's zip code of origin, arrival and departure dates for the order, group size, and campsite type. If the order contains a "Cancellation" transaction, then it is known that the order was cancelled.

For each date, we are able to determine the number of parties and the number of people present at each campground using information on the orders' arrival and departure dates. If the order was cancelled, voided, or listed as a no-show, it is not added to the number of occupied sites at a campground. Figure A1 provides a visualization of the data. We plot the average number of campers present at Glacier along with the proportion of days with observed smoke conditions in the sample; smoke conditions in Glacier overlap with times of greater visitation.

One of our primary variables of interest is the occupancy rate of a campground *i* on a given day *t*, which we define as (occupied campsites_{it})/(total number of campsites_{it}). The Recreation.gov data do not report the total number of campsites at each campground on a given date. While the data provide a list of campsites at each campground for 2017–18, the actual number of available campsites at some campgrounds varies from year to year. Some campgrounds, for example, were not yet open during the early years of the sample; others added or removed campsites over time. In some cases, campgrounds have shut down for entire seasons. To obtain the best possible estimate of the available campsites for each campground, we create an algorithm that predicts the number of campsites by year for each campground based on a combination of (i) the listed campsites in 2017–18, (ii) the maximum number of sites reserved on any given day in a given year, and (iii) the individual identification numbers for each site, to ensure that we capture as many of the available sites as possible. For each campground for each year, the algorithm proceeds in the following way:

1. If the maximum number of reserved sites in a year (item ii) matches the number of campsites listed in 2017–18 (item i), the algorithm applies that number.

2. If the maximum number of reserved sites does not match the number of campsites listed in 2017–18, the algorithm counts the number of times the within-year maximum number of occupants (item ii) was obtained. If it occurred three times or more, the algorithm applies that number for the yearly number of available campsites.

3. If step 2 fails (the within-year maximum number of occupants was not obtained at least three times), the algorithm checks how often the number of occupants matched the listed number of campsites in 2017–18 (item i). If it was more than three times, the algorithm applies that number for the yearly available campsites.

4. If both steps 2 and 3 fail, the algorithm checks if the maximum number of occupants in the preceding year and the following year matched, and if so it applies that number.

5. If none of these criteria are satisfied, the algorithm selects the number of sites available in 2017–18 (item i).

This algorithm accounts for many scenarios. If a campground had more available sites than was reported in 2017–18 (criterion i), then the yearly maximum would be achieved fairly frequently (item ii), providing a more accurate measure of campground size. If a campground was closed for an entire season, then the maximum number of sites reserved in a year (criterion ii) is 0, which occurs 365 times, so the number of available sites for that year would be set to 0. We manually assessed and corrected the results of this algorithm by examining a time series of the number of occupied sites for each campground and comparing against items (i), (ii), and (iii). Some campgrounds do not fill up, but by examining the individual identification numbers of each site (item iii), we can determine the number of available sites for each year.

Two other variables are of interest in regressions on campground use: the pre- and post-arrival cancellation rates. For the pre-arrival cancellation rate, for day *t*, we add the transactions of type "Cancellation," "Cancellation (Waive Penalty)," and "No-Show" for arrival date *t* if the cancellation was transacted within seven days (i.e., greater than or equal to *t* – 7). We divide this sum by the total number of reservations scheduled to arrive on *t*. Formally, for campground *i*, this is $\frac{\text{cancellations}_{it} + \text{cancellations (waived penalty)}_{it} + \text{no shows)}_{it}}{\text{reservations}_{it}}$. Intuitively, this measures the share of reservations for date *t* that were cancelled prior to arrival.

For post-arrival cancellations, we add transactions of type "Cancellation," "Cancellation (Waive Penalty)," and "Shorten Stay Leave Early" on day *t* if the date *t* falls between the scheduled arrival and departure date. We divide that sum by the number of occupants present at the campground on day *t*. Formally, for campground *i*, this is ($\text{cancellations}_{it}$ + cancellations (waived penalty_{it}) + shorten stay leave early_{it})/(occupants_{it}), for midstay cancellations only.

Appendix B: Results with Alternative Fire and Smoke Variables

Campground and campground visitor-days affected by wildfire and smoke

The measurement of campground-days near actively burning wildfires or impacted by smoke varies depending on how we define affected days. In the main text, we define "near to an active fire" as being within 20 km of a burning wildfire. The upper panel of Table B1 summarizes the number of campground-days and visitor-days affected when we instead use a 30 km bandwidth. The average number of days on which campgrounds experience a nearby fire increases from 1.5 to 2.8, and the percent of total visitor-days affected by a fire increases from 1.4 to 2.5. The distribution of fire days across regions is similar for both bandwidths.

The lower panel of Table B1 shows how the number of campground-days and visitor-days affected by smoke changes when we define smoky days using only the NOAA HMS smoke plume data, without restricting impacted days to be those with on-the-ground air quality above the 95th percentile on nonsmoky days (our definition of adverse smoke conditions in our baseline results). Contrasting Table B1 with Table 1, only approximately 26 percent of the days in which campgrounds were covered by smoke plumes had $PM_{2.5}$ levels above the 95th percentile.

Figure B2 shows trends over time in the number of campground-days and visitor-days affected by fire and smoke. In the upper panel, campground smoke days are defined as days in which a campground was covered by a smoke plume and PM2.5 was more than 1.64 SD above the seasonal mean; campground fire days are defined as days in which a fire burned within 20 km. In the lower panel, definitions of adverse smoke conditions are varied, with standard deviations above the seasonal mean that PM2.5 must be for the campground to be considered to have impacted air quality given in parentheses. We also plot the number of days campgrounds were under a smoke plume, irrespective of PM2.5. Finally the lower right panel shows differences in the number of camper-days near fire by fire distance thresholds.

Though the frequency of large wildfires in the western United States has increased over the past several decades (Westerling 2016), we observe no clear trends in exposure to fire or smoke over the 10 years of our data set. It may be that year-to-year variation in the numbers and locations of wildfire events masks long-term trends, especially over the relatively short span of our data set.

Behavioral responses to smoke and fire

In our regressions on campground use, we explore behavioral responses to smoke and wildfire. Equation (1) shows the main specification, where the dependent variable is a function of indicators for smoke, fire, and a series of location and time fixed effects. We test the effects of alternative definitions of the fire indicator and alternative sets of location and time fixed effects specifications in figures B3 through B5.

Our preferred model sets the fire variable equal to 1 when an active fire burns within 20 km of a campground. In figures B3–B5, we test distance bandwidths of 10 km and 30 km. The coefficient grows in magnitude as we narrow the bandwidth, indicating that campground use is affected more when fire is closer to the campground.

Figures B3–B5 also illustrate effects of our choice of fixed effect specifications. For each combination of smoke and fire variable, we show results of four specifications: (i) no fixed effects; (ii) campground and month × year fixed effects; (iii) campground, recreation area × month-of-year, and recreation area × year fixed effects; (iv) the same fixed effects as in (iii), but adding controls for holidays, week of year, and day of week; and (v) the same fixed effects as in (iv) but adding a control for the upcoming week's total precipitation.

In specification (i), standard errors are quite large and coefficients frequently do not have the expected sign. For example, the coefficient on smoke in the percent occupancy regression

(Figure B3) is positive, likely because recreation activity coincides with times of year with greater fire activity (see, for instance, Figure 1), emphasizing the importance of the fixed effects.

Specification (ii) greatly reduces standard errors. However, by including only campground and month × year fixed effects, the specification assumes seasonal variation in campground use is the same across campgrounds. The results of specification (ii) may be biased if time-varying, location-specific unobservables exist that are correlated with the independent variable of interest. In most cases, coefficients estimated from specification (ii) have the expected signs; however, we observe sign reversal in the smoke coefficient in the percent occupancy regressions.

Models (iii) and (iv) allow for different temporal effects by recreation area. The recreation area × month fixed effects allow for control of seasonality at the recreation area level, and the recreation area × year fixed effects control for differential trends across time for different recreation areas. These fixed effects take into account, for example, that different recreation areas peak at different times of year. For instance, the Grand Canyon in Arizona has different seasonal peaks than North Cascades National Park in northern Washington. Model (iv) additionally controls for seasonality, adding holiday indicators, day-of-week fixed effects, and week-of-year fixed effects. These controls distinguish the effects of weekdays from weekends and also account for popular times of the year, such as July 4 or Memorial Day. Including precipitation controls in model (v) does not have a substantial effect on coefficient estimates.

In summary, these sensitivity analyses reveal that results vary sensibly as definitions of the fire and smoke variables are altered. Fire and smoke coefficient estimates depend somewhat on the set of fixed effects we include in the regression, but results are consistent across specifications that account for recreation area-specific seasonal variation in visitation.

We present a final specification in Table B2. This table presents a specification similar to that in Table 4, but also includes an indicator for whether PM2.5 is more than 1.64 standard deviations above the seasonal mean. Since *Smoke* is defined as an interaction between the presence of a smoke plume and this indicator, we can interpret the coefficient on *Smoke* in this specification as the differential effect of smoke when there is poor air quality. Changes in recreation behavior appear to be driven primarily by the combination of smoke plumes and poor air quality, and not by poor air quality alone.

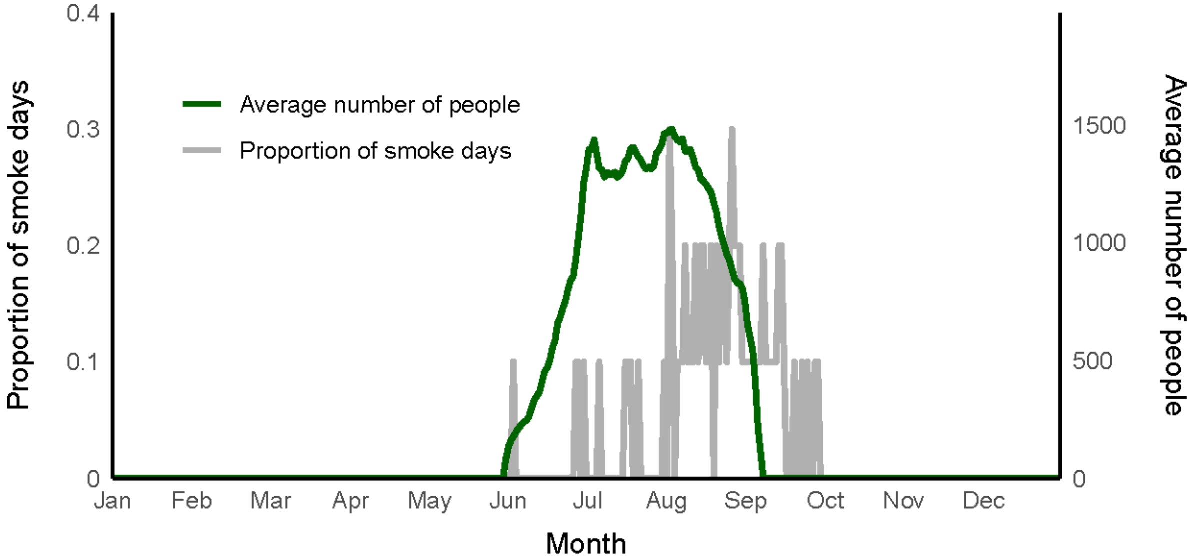

Figure B1. Occupancy and the proportion of smoke days at Glacier National Park, 2008-2017.

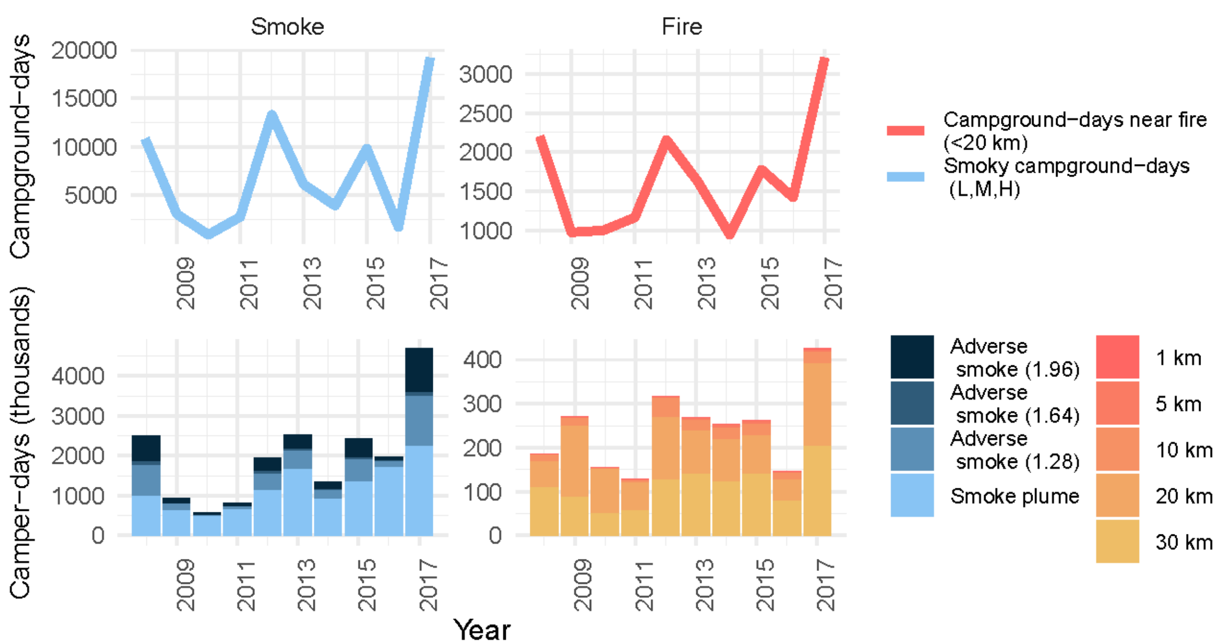

Figure B2. Prevalence of days near fire and with adverse smoke conditions, 2008-2017.

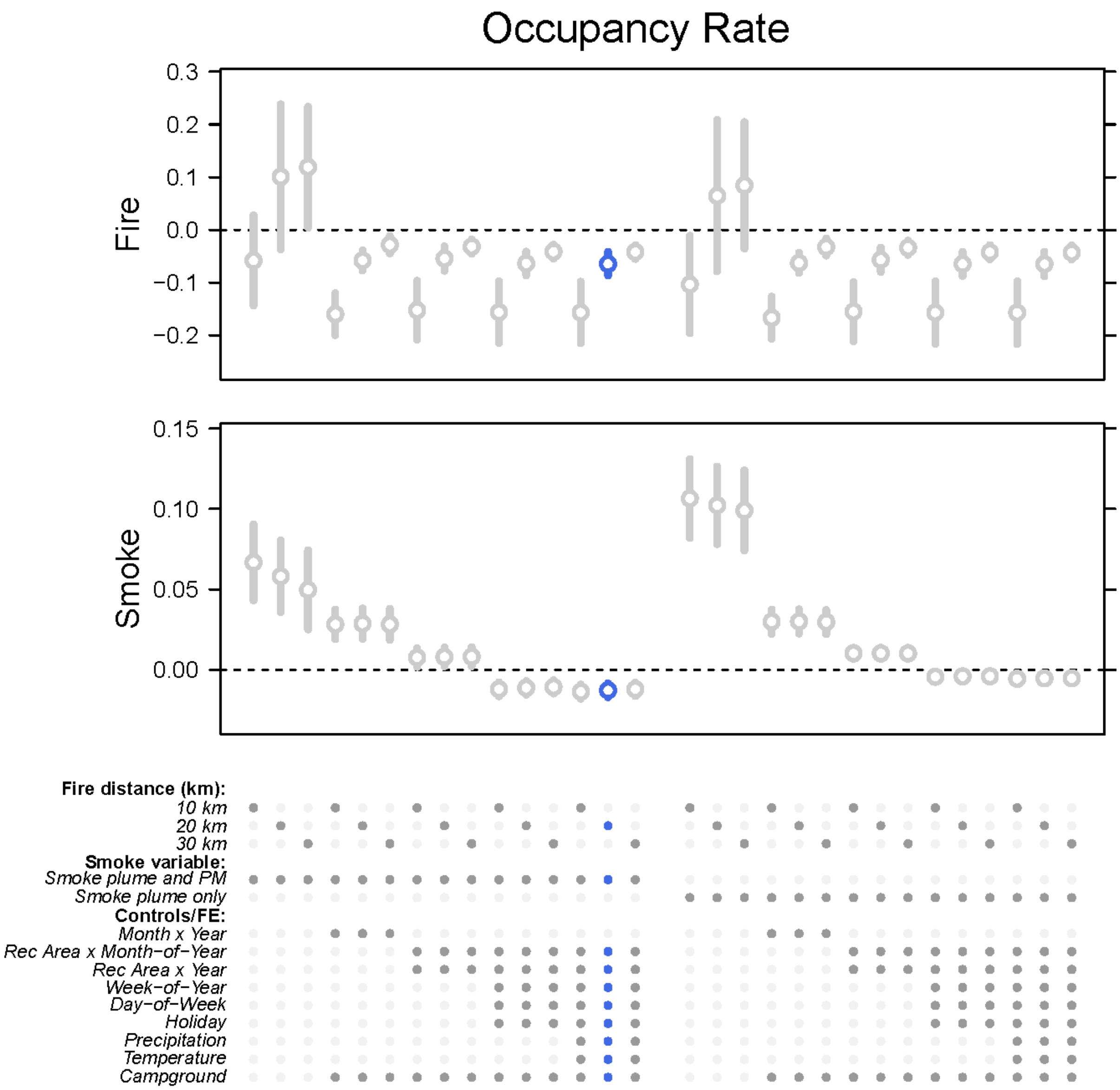

Figure B3. Specification chart for regression of campground occupancy rate on fire and smoke. The coefficients of interest are on the y-axis. The baseline model is shown in blue.

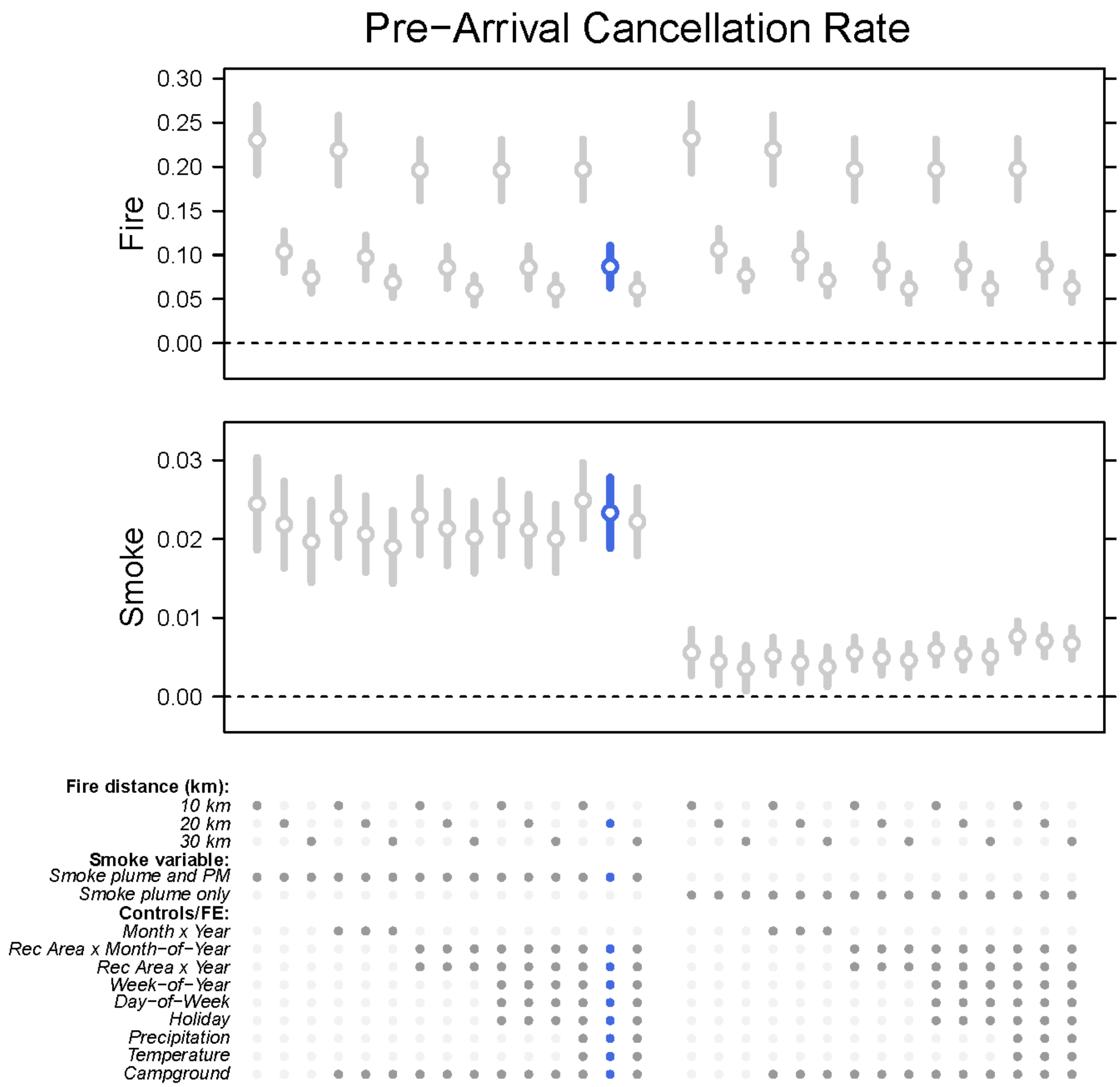

Figure B4. Specification chart for regression of pre-arrival cancellation rate on fire and smoke. The coefficients of interest are on the y-axis. The baseline model is shown in blue.

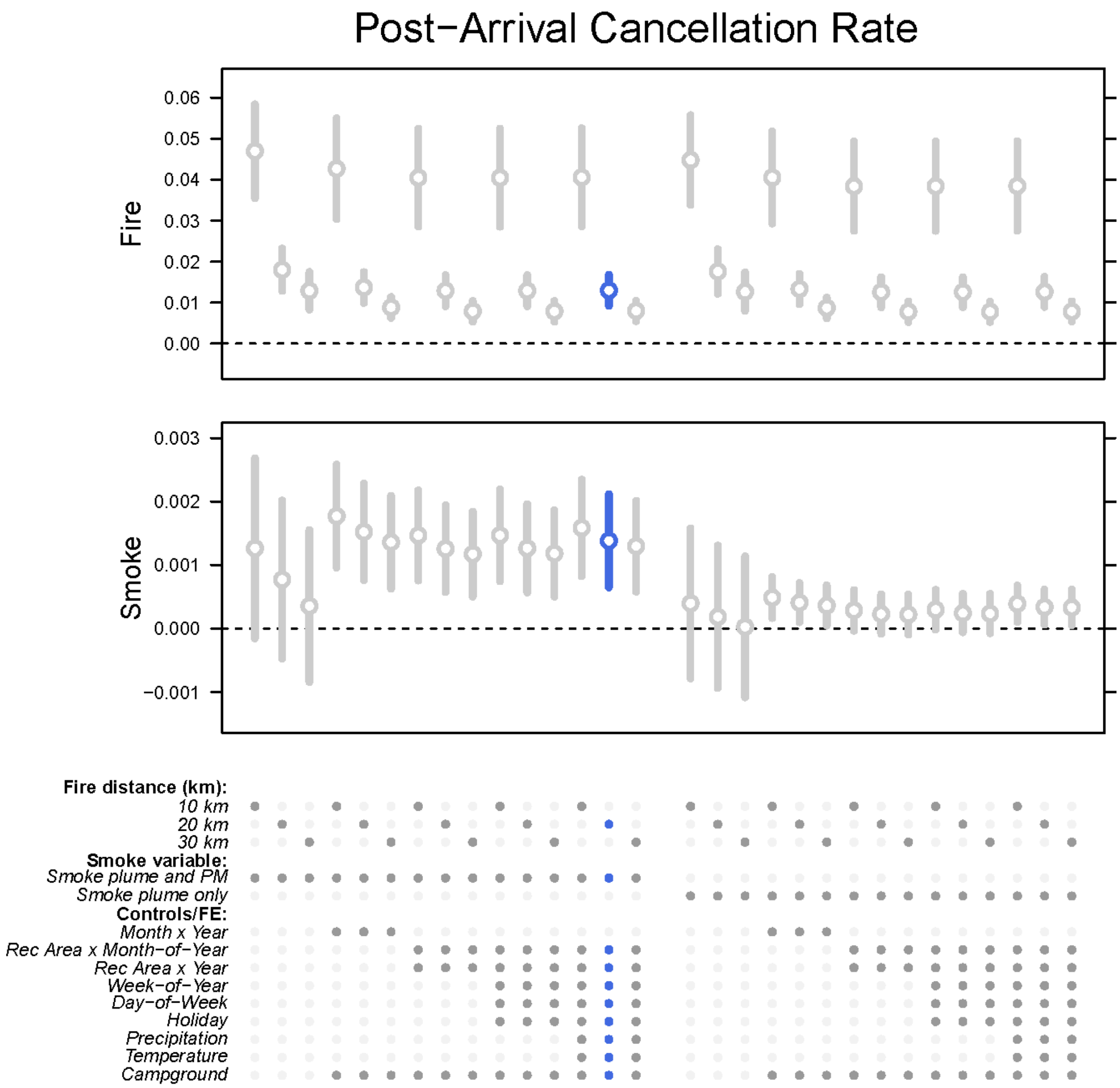

Figure B5. Specification chart for regression of post-arrival cancellation rate on fire and smoke. The coefficients of interest are on the y-axis. The baseline model is shown in blue.

Table B1. Annual campground- and camper-days near wildfires (within 30 km) and under smoke plumes, by region.

	Campground-days		Camper-days	
	Avg. annual days per campground	Percent of total available campground-days	Avg. annual camper-days (thousands)	Percent of total camper-days
I. Fire				
California	4.3	3.4	139	3.4
Pacific Northwest	3.1	4.3	26	1.8
Rocky Mountains	0.8	0.9	4	0.4
Great Basin	1.0	1.2	5	0.5
Southwest	4.1	3.8	29	3.8
Northern Rockies	3.0	3.7	15	2.2
Total	2.8	3.0	218	2.5
II. Smoke				
California	28	22	707	17
Pacific Northwest	31	44	345	24
Rocky Mountains	20	24	163	16
Great Basin	16	19	107	12
Southwest	14	13	54	7
Northern Rockies	34	43	211	32
Total	26	28	1,588	18

Notes: Fire days are days in which a campground is 30 km or less from an active wildfire. Days under smoke plumes are days in which campgrounds intersected a NOAA HMS smoke plume. Each campground's available campground-days are calculated as the number of days each year that the campground had at least one occupant.

Table B2. Estimated effects of wildfire and smoke on campground use, including $PM_{2.5}$

	Occupancy Rate	Pre-arrival Cancellation Rate	Post-arrival Cancellation Rate
Fire	-.064**	.087**	.013**
	[.011]	[.012]	[.0019]
Smoke	-.013**	.023**	.0014**
	[.0022]	[.0023]	[.00037]
$PM_{2.5}$	.001	-.001	-.0003
	[.003]	[.001]	[.0002]
Mean of dep. Var.	.31	.076	.0024
No. of obs.	1,349,460	688,653	842,240
R^2	0.72	0.048	0.13

Notes: $PM_{2.5}$ is an indicator variable for whether $PM_{2.5}$ was more than 1.64 SDs above the location-specific seasonal mean. All columns include campground, recreation area by month-of-year, recreation area by year, week-of-year, and day-of-week fixed effects, as well as indicators for holidays and days before holidays. In addition, regressions control for the upcoming week's total precipitation. Campground observations are weighted by the number of campsites, and standard errors, shown in brackets, are clustered by recreation area. The observations are restricted to May through September. ** $p < 0.01$, * $p < 0.05$.