Supporting Information for

Emergency department visits respond non-linearly to wildfire smoke

Sam Heft-Neal, Carlos F. Gould, Marissa Childs, Mathew V. Kiang, Kari Nadeau, Mark Duggan, Eran Bendavid, Marshall Burke

Corresponding Author: Sam Heft-Neal
E-mail: sheftneal@stanford.edu

This PDF file includes:

Supporting text
Figs. S1 to S18

Supporting Information Text

Choice of Main Specification. There are a number of researcher degrees of freedom in choosing our main specification including, but not limited to, how we model nonlinearities, choice of fixed effects, lag structure, choice of controls, and how we calculate confidence intervals. In this section we discuss each of these choices in detail.

Modeling Nonlinearities. We find strong evidence that the ED visit response to increasing ambient wildfire smoke $PM_{2.5}$ concentration is nonlinear. We therefore consider a linear model inappropriate. When we estimate polynomial responses of degrees 2-6 we find that 4th-6th degree polynomials produce similar responses to each other (and to cubic splines) and that these responses suggest sharper nonlinearities than either the 2nd or 3rd degree polynomials can capture. We choose a 4th degree polynomial as our main parametric specification as it is the simplest specification whereby further increasing flexibility does not meaningfully change the shape of the estimated response.

For our nonparametric models, where we bin days into different ranges of smoke $PM_{2.5}$, we consider several different bin cutoffs. We do not have reason to believe there are specific cutoffs that would be most appropriate in our setting and so we consider cutoffs that correspond to 'round' numbers and ultimately present results for cause-specific ED visits with nonparametric models that have bin cutoffs chosen to reflect the 50th, 75th, 90th, 95th, and 99th percentile of observed smoke $PM_{2.5}$ concentrations. We also consider fixed bin widths of 10 and 25 $\mu g m^{-3}$, respectively (Fig S5), but opt for the quantile based cutoffs as our main nonparametric specification because they better characterize the full distribution of smoke $PM_{2.5}$ concentrations, including extreme days which are relatively uncommon but may generate outsized impacts.

Fixed Effects. As discussed in Methods our model includes several fixed effects. All specifications we consider include zipcode fixed effects because we want to limit our model to within-location comparisons over time. In addition, we include day-of-week fixed effects in all specifications, not to remove potential biases, but because ED visits have a strong weekly cycle and absorbing this variation (which is uncorrelated with wildfire smoke) reduces noise in our outcome variable. Our choice of model fixed effects therefore comes down to the appropriate ways to account for seasonality (both smoke and many ED outcomes exhibit strong seasonality) and for trends over time (both smoke and ED visits trended up over our period).

To model seasonality, we consider month-of-year fixed effects (which assume a consistent seasonality across the state) and more flexible specifications that allow seasonality to vary across counties or zipcodes. In our data, we see variation in the seasonality of both ED visits and in smoke $PM_{2.5}$ across the state. In particular, patterns of seasonality appear to differ between Northern and Southern California and between coastal and inland regions. Given these within-state differences in seasonality, we therefore consider county by month-of-year or zipcode by month-of-year fixed effects to be more appropriate for modeling seasonality in California. Because they produce similar responses (Fig S4b) to each other we elect to include the simpler version (county by month-of-year) in our main specification.

Similarly, to control for time trends we could include a year fixed effect, which would assume constant trends across the state and across months of the year. However, we find that ED visits are trending differently by season: over the study period, ED visits have been trending upward in every month but visits in the winter have trended upward 10-15% faster than visits in the summer. Similarly, wildfire smoke is also trending differently by season with monthly average smoke levels increasing fastest in the late fall and early winter. To account for these differences in trends across the year, we consider season-by-year, month-by-year, and week-by-year fixed effects. Because we see similar responses for the three approaches that model time trends sub-annually (Fig S4b), we again opt to include the simplest option, season-by-year fixed effects (which we define it its most aggregate form to be wildfire season (i.e., June–October) vs. non-wildfire season (i.e., remaining months).

Lag Structure. To account for potential delayed impacts of ambient wildfire smoke on ED visit rates, we include daily lags in our model. For our main specification, we include 7 lags (day-of exposure plus an additional week of daily lags). However, we consider 7, 14, and 28 lags as possible choices for our main specification. We ultimately opt for 7 days of lags because it produces the simplest model (i.e., the smallest number of coefficients to estimate) and because it is the most conservative approach in the sense that it produces smaller estimates for the number of ED visits attributable to wildfire smoke than analogous models with 14 or 28 lags (Fig S18).

Controls. Our main specification includes controls for temperature, rainfall, and distance to the nearest active fire. We include temperature and rainfall because both could be correlated with fire activity (and therefore smoke levels) and could also directly influence ED visits. Temperature, in particular, is a potentially important omitted variable in our model because there is a well documented relationship between ambient temperatures and ED visits and temperature is an important driver of contemporaneous fire activity. However, temperature could also be a 'bad control' in the sense that wildfire smoke could be one of the channels through which temperature impacts ED visits. Thus, by including temperature, we may be absorbing some of the effect of wildfire smoke on ED visits, which would lead us to underestimate the magnitude of wildfire smoke's impact on ED visits. The 'true' effect of wildfire smoke on ED visits is likely somewhere in the middle of the smaller effect estimated when we include temperature and the larger effect estimated when we omit it from our model. We also include distance to fire to try and help distinguish between fire and smoke effects.

We choose to include these covariates in our main specification both because of the potential omitted variable issue noted above, but also because including them is the more 'conservative' choice in the sense that the resulting model estimates a smaller overall impact on ED visits from wildfire smoke.

Confidence Intervals. For all-cause ED visits responses estimated with a polynomial functional form (Fig 2) we calculate bootstrapped 95% confidence intervals at each concentration across the smoke $PM_{2.5}$ distribution clustered at the zipcode level.

For nonparametric specifications we calculate analytical standard errors for the cumulative effect (i.e., sum of lags) that account for correlation across lags. We also cluster our standard errors at the zipcode level. When we estimate responses for cause-specific ED visits we further Bonferroni correct confidence intervals to account for the large number of regressions we are running (Fig S9). The Bonferroni correction is an aggressive approach to accounting for multiple hypotheses and likely results in wider confidence intervals than other corrections would (e.g., Holm-Bonferroni, Benjamini-Hochberg). We opt to apply it because it is a conservative correction. However, this choice should be taken into consideration when evaluating the cause-specific responses shown.

Robustness of estimated all-cause ED visit response to specification choices

For all-cause ED visits, all functional form choices produce response functions that increase at low to moderate smoke $PM_{2.5}$ concentrations and decline at high levels (Fig S4-S5). However, there is some variability in the magnitude of estimated changes depending on choice of polynomial or bin cutoff.

When we evaluate different combinations of possible fixed effects, because of the strong sub-state differences in seasonality, we find that estimated responses do differ somewhat depending on whether or not we allow seasonality to vary across parts of the state. Similarly, we find that allowing time trends to vary sub-annually produces a response curve with different magnitudes of changes than a model that includes year fixed effects (which assume constant trends across the year). In both cases, we find similar results when we increase the spatial and temporal resolution of the sub-state and sub-annual fixed effects, respectively.

To assess robustness of our main results, we also include as much as a month of lags. While our main results are qualitatively similar regardless of how many lags we include (Fig S4c), the magnitude of the cumulative response does vary.

Finally, when we estimate an unadjusted model the estimated response function retains the concave increasing then decreasing shape we see in other specification but has a much higher positive peak and a sharper decline implying a larger increase in ED visits for low and moderate smoke $PM_{2.5}$ and a sharper decline in ED visits at high smoke $PM_{2.5}$ (Fig S4d). These differences appear to be driven primarily by whether we control for temperature in the model.

Robustness of attribution estimates to specification choices. We find that 10 of 14 alternative model specifications considered produce attribution estimates within our main confidence interval (Fig S18a). Two specification produce lower estimates. However, these models were not selected as the main specification because as discussed above they were assessed to not be sufficiently flexible in how they model seasonality or time trends. Two other models produce estimates higher than the upper end of our main confidence interval. These higher attribution estimates come from models that omit temperature as a control. Finally, we find that models that include fewer lags tend to produce lower attribution estimates (Fig S18b).

ICD Groupings. Excel file with comprehensive list of outcomes included in the study and the corresponding ICD-9 and ICD-10 codes can be downloaded here.

Supplemental Figures

Fig. S1. Trend in ED visit rates over time. Population-weighted monthly average of daily zipcode level ED visit rate over time.

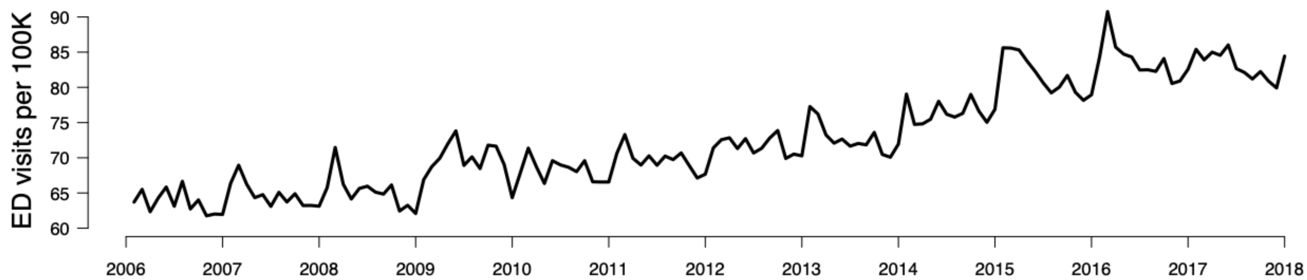

Fig. S2. Spatial patterns of ED visits. **a.** Spatial distribution of average zipcode level rates for all-cause ED visits. **b** Average rates by primary diagnosis. Areas not covered by zip codes are filled in with county-level averages. All maps are colored according to visit rate quantiles in order to highlight the spatial distribution.

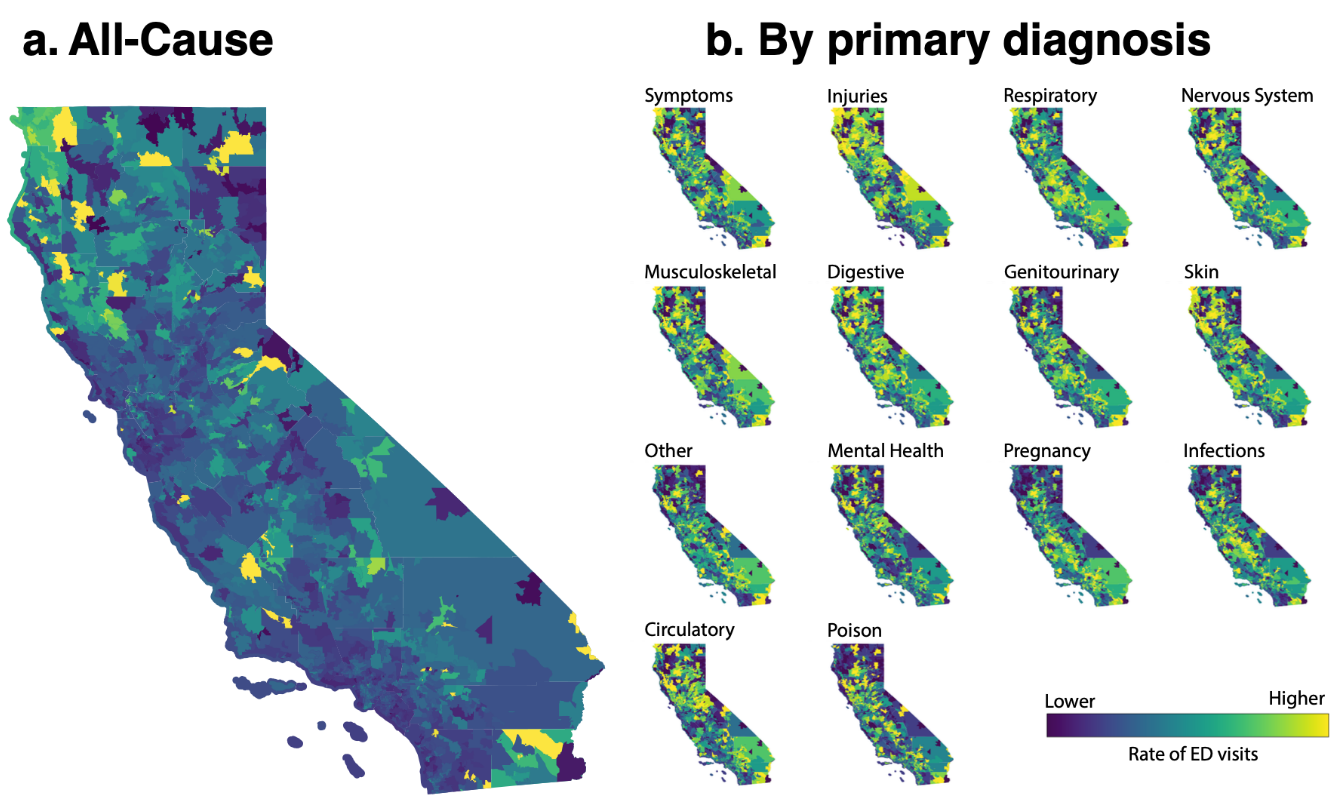

Fig. S3. Seasonality in ED visits by primary diagnosis. Each panel shows the population weighted month-of-year average daily zipcode level ED visit rate for a principal diagnosis grouping. Panels are sorted from most (symptoms) to least (poison) frequent diagnosis. To highlight the seasonality in diagnoses with widely varying rates, the vertical axes vary across panels. To quantify the importance of seasonality the maximum-minimum ratio (mmr) is calculated for each diagnosis group as the ratio of the rate in the highest month to the rate in the lowest month and shown in the panel text. Seasonality is largest in ED visits for respiratory conditions with the maximum-minimum ratio (2.5) nearly twice as large as for the next most seasonal condition (poisonings = 1.4). All remaining diagnosis groupings exhibit smaller month-to-month variation.

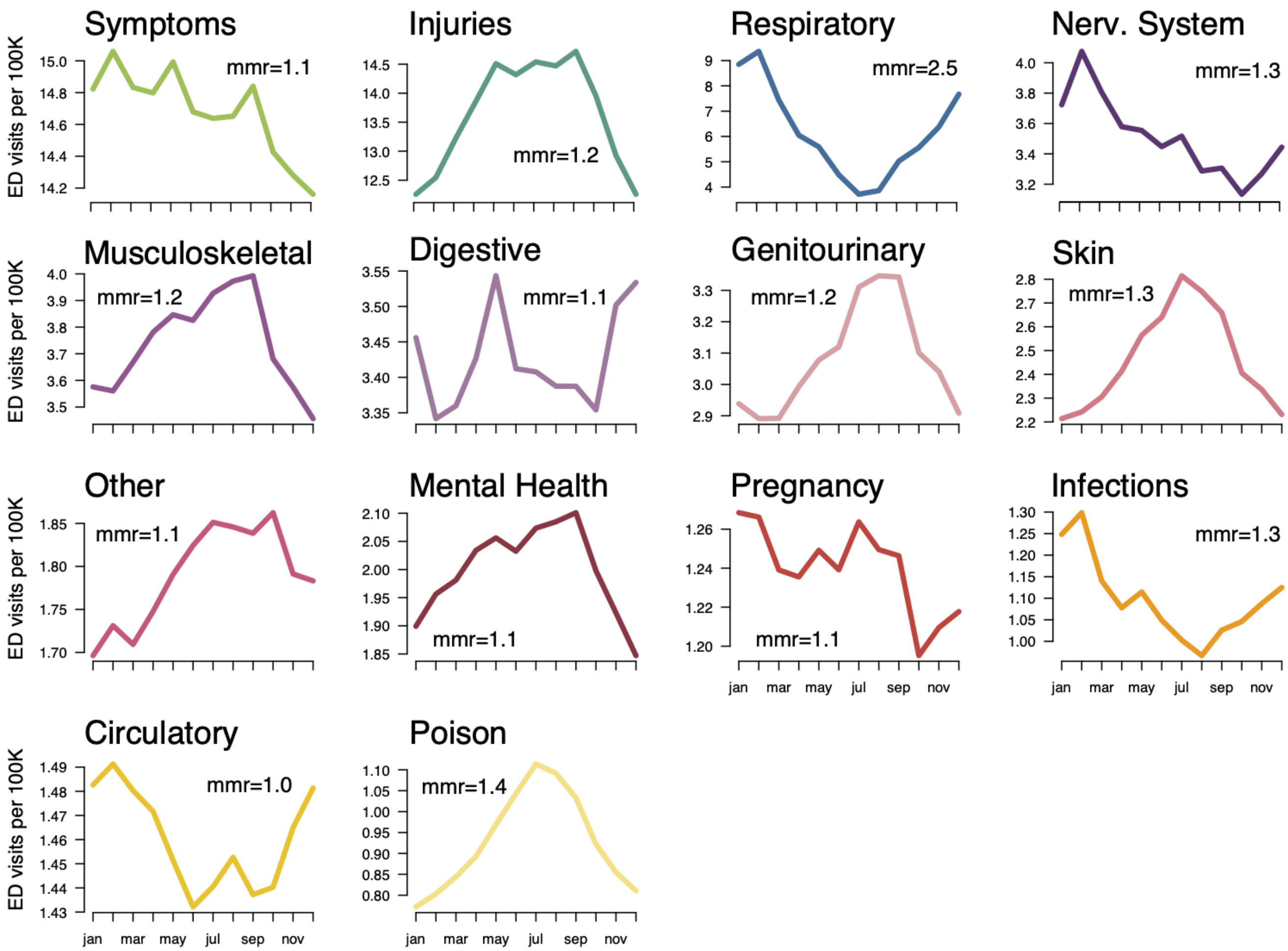

Fig. S4. General shape of estimated cumulative response is robust to modeling choices. The estimated response of total ED visits to wildfire smoke is shown across **a.** different smoke intensity functional forms, **b.** choice of fixed effects, **c.** lag structures, and **d.** with inclusion of different controls. Additional estimated responses are shown separately for nonparametric binned specifications (Fig S5), poisson models of ED visit counts (Fig S6), at a weekly aggregation (Fig S7), and for the subset of the sample far from active fires (Fig S11).

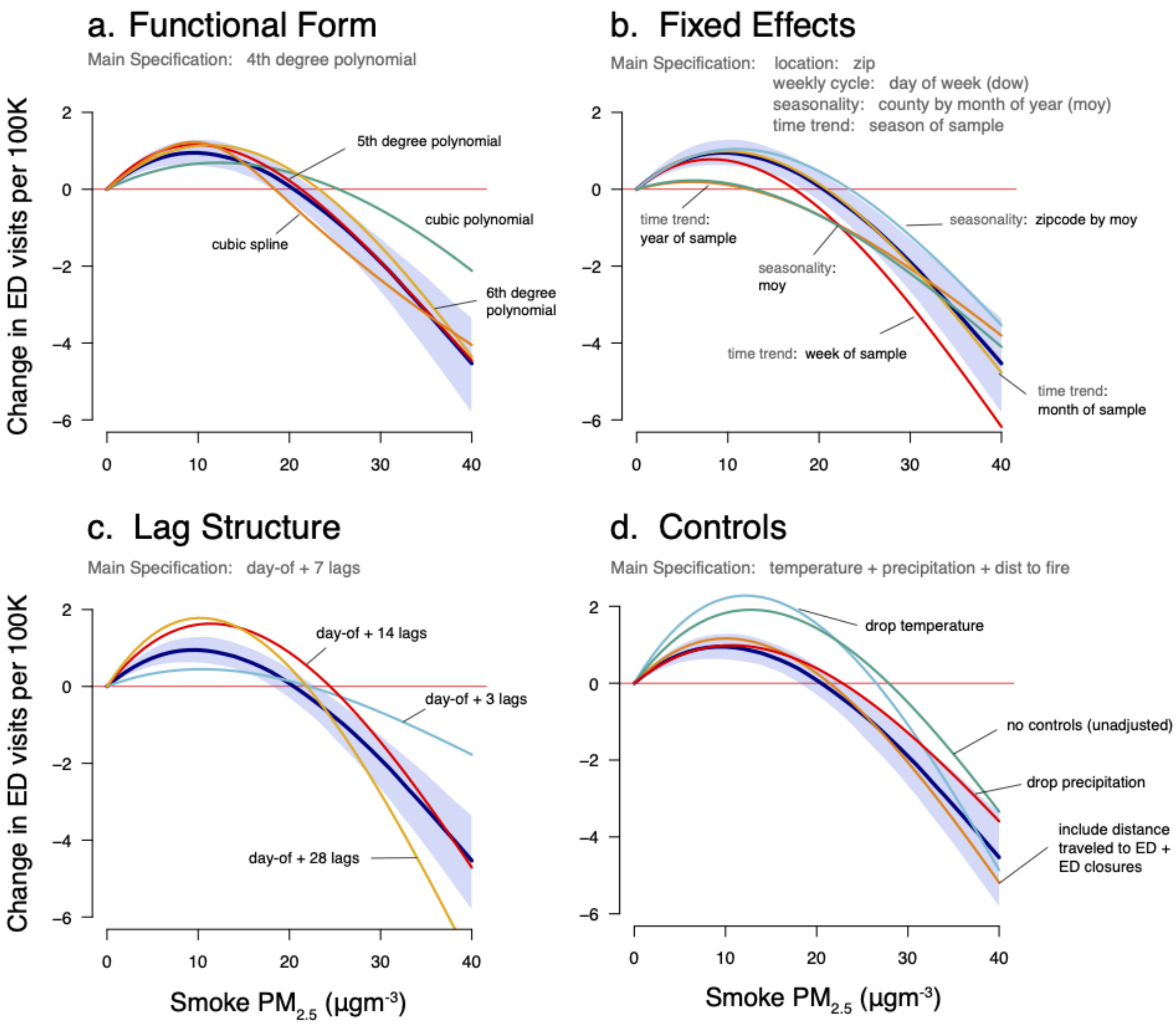

Fig. S5. General shape of estimated cumulative response is robust to binning choices. Total ED visit responses estimated from nonparametric binned models of smoke $PM_{2.5}$ concentration with different bin cutoffs. Responses estimated from binned models of smoke $PM_{2.5}$ **a.** with bin cutoffs corresponding to select quantiles **b.** with evenly divided 10 μgm^{-3} smoke $PM_{2.5}$ bins, **c.** with evenly divided 25 μgm^{-3} smoke $PM_{2.5}$ bins. All models include the fixed effects and controls included in our main specification (Methods).

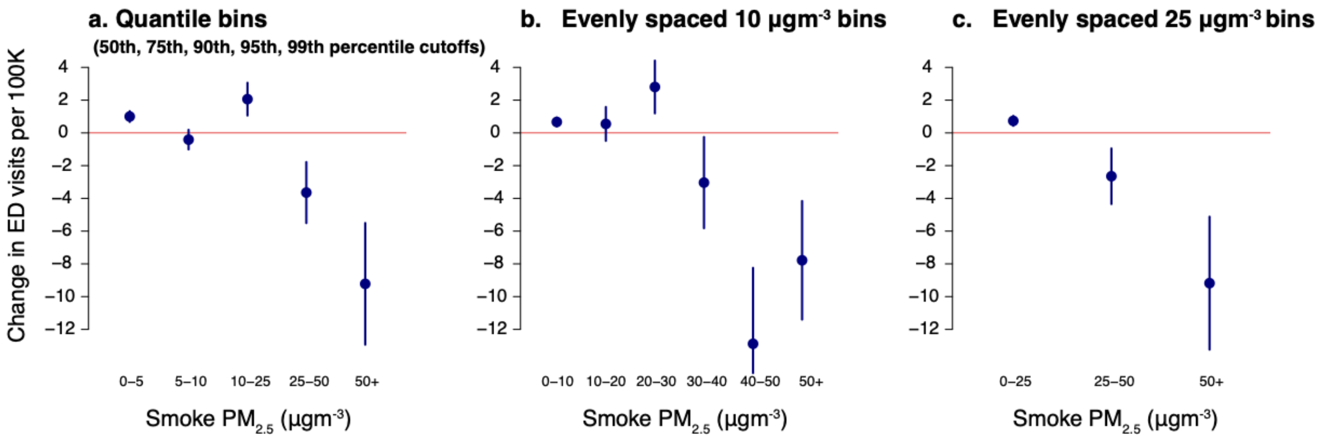

Fig. S6. General shape of estimated cumulative response is robust to alternative regression methods. Total ED visit responses estimated from Poisson model. Responses estimated from poisson model of daily ED visit counts **a.** with main binned specification and **b.** with main 4th degree polynomial specification. Both models include the same controls and fixed effects as our main specifications (Methods).

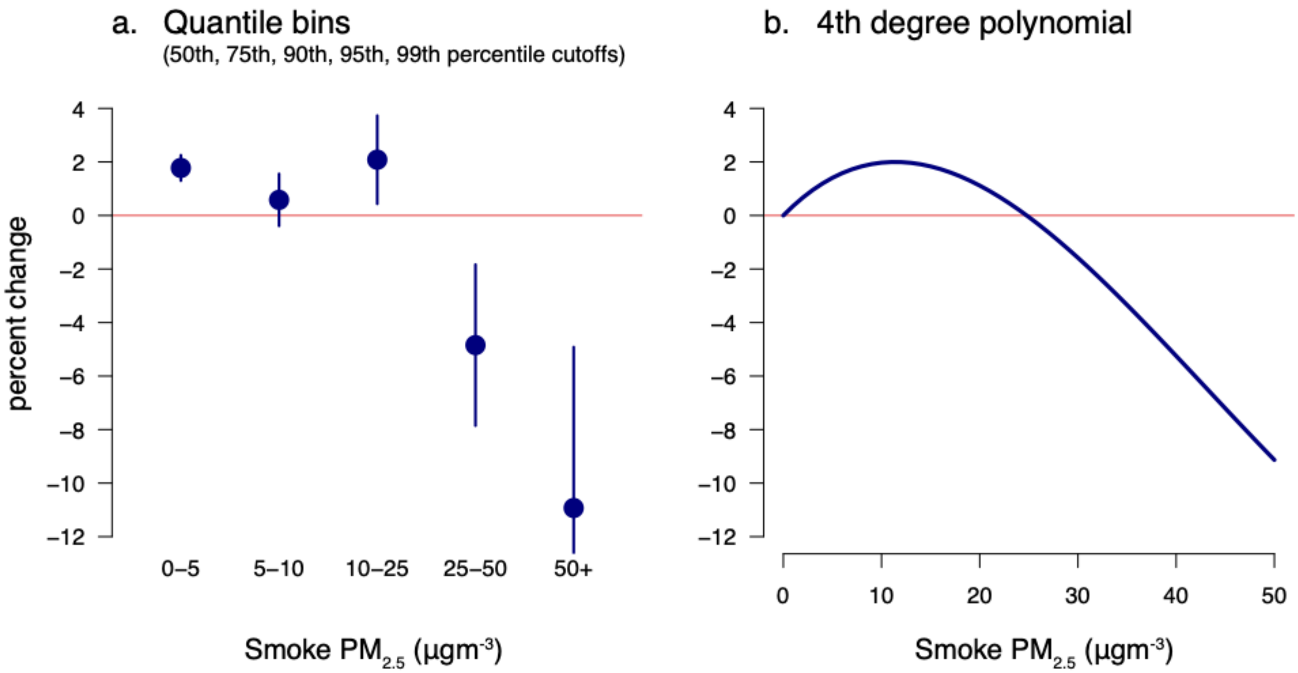

Fig. S7. General shape of estimated cumulative response is robust to different temporal aggregations. Our main binned model (shown in Fig S5a) re-estimated after aggregating from daily to weekly level. Model includes the same controls and analogous fixed effects as our main specifications (Methods).

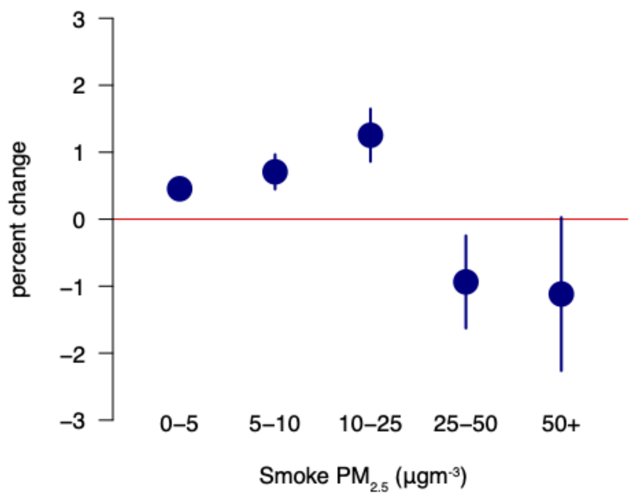

Fig. S8. Proportion of ED visits in California by primary diagnosis. Figure shows the breakdown of primary diagnoses by group and sub-group (analogous to Figure 1b) for all visits in California 2006-2017. Black labels indicate first-level categories and white labels indicate sub-categories.

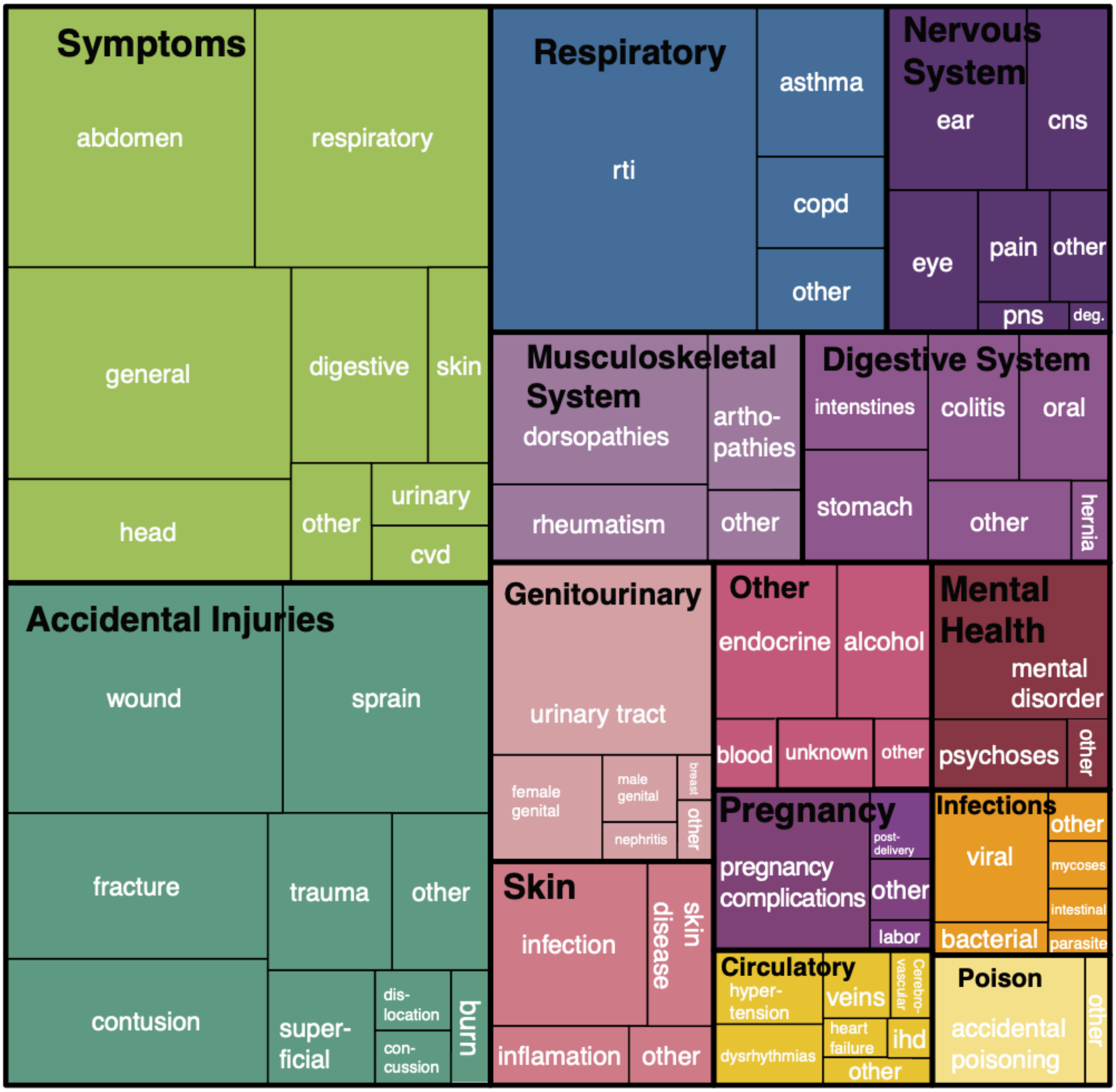

Fig. S9. Average cumulative effects of wildfire smoke on ED visit rates by principal diagnosis for the 14 primary ICD groupings. Each response comes from a different regression estimated where the number of visits with a given principal diagnosis is the outcome. All models are specified as binned models with the same controls used in our main specification (Methods). To account for multiple hypothesis testing we apply the Bonferroni correction. Bonferroni 95% confidence intervals therefore reflect $\alpha = 0.05/14$. Base rates shown are zipcode level daily population weighted average ED visits per 100,000 people with that outcome as the primary diagnosis.

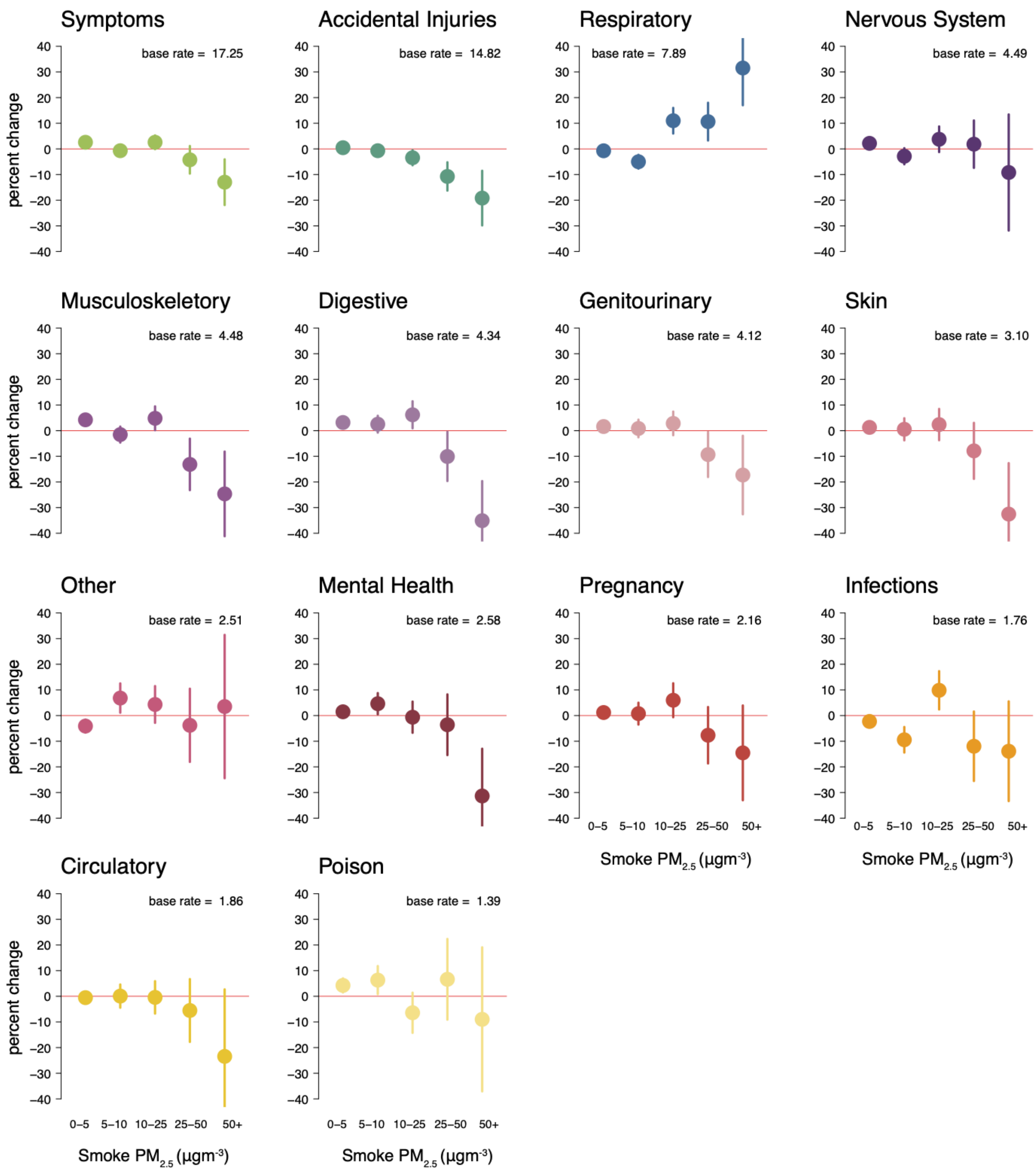

Fig. S10. Cumulative effects of wildfire smoke on ED visits for different types of respiratory conditions. The response in each panel comes from a different regression where ED visits with a given principal diagnosis are the outcome and we estimate the binned model described in Methods. Vertical lines indicate Bonferroni-corrected 95% confidence intervals. ICD groupings correspond to Figs 1b and S8 and all ICD codes corresponding to each grouping are listed in the supplement. Base rates shown are daily population weighted average ED visits per 100,000 people with that outcome as the primary diagnosis. Acronyms in labels: RTI = respiratory tract infection. URI = upper respiratory infection. Note, as illustrated in Fig 1b, that URI, bronchitis, pneumonia, and influenza are sub-types of RTIs.

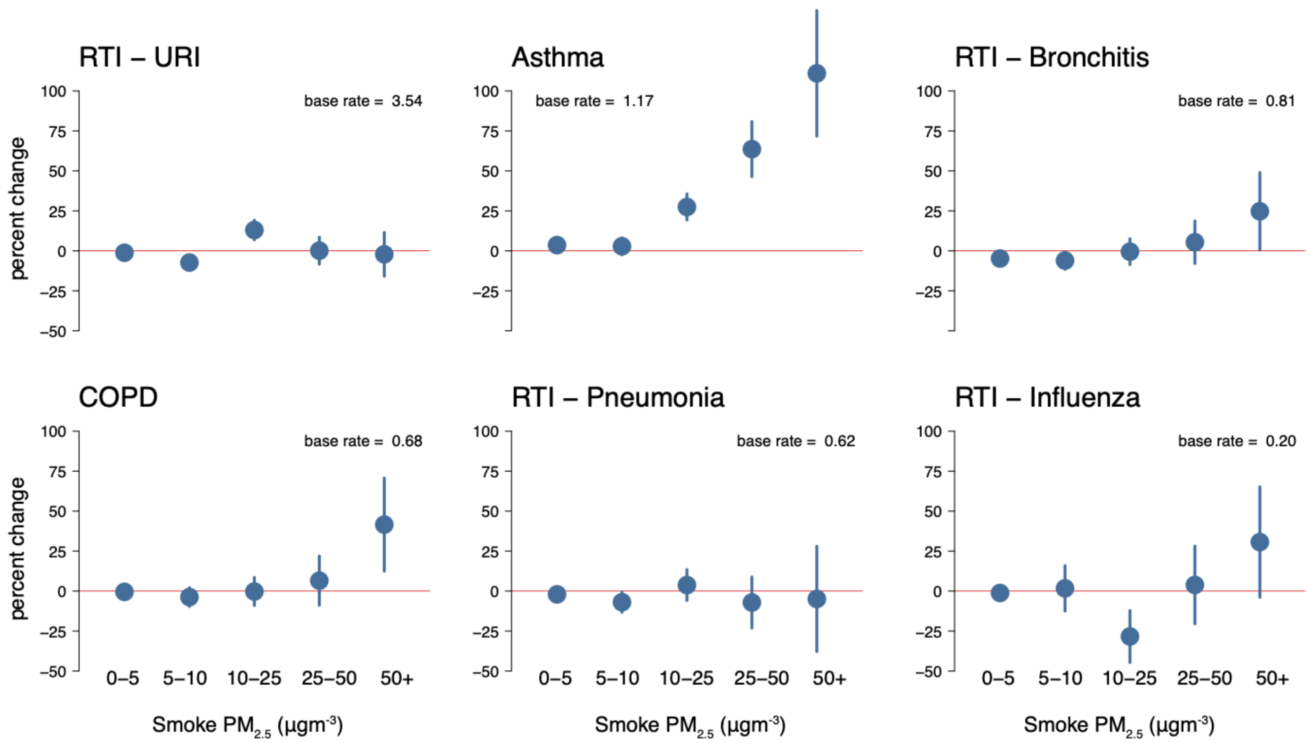

Fig. S11. Responses estimated from subsample of observations far from active fires. Responses estimated using our main specification on subsamples of the data limiting to zipcode-days when there were either no fires or the nearest fire was more than 25km (or 50km) from the zipcode (Methods).

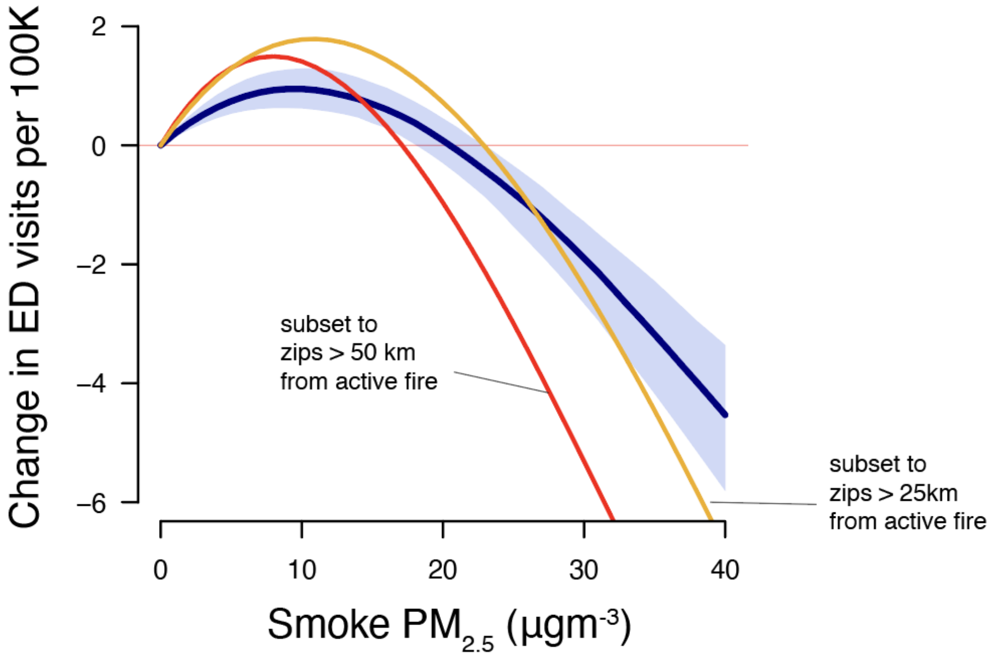

Fig. S12. Contributions of different diagnoses to estimated changes in ED visits in response to wildfire smoke. a. Diagnosis-specific responses were aggregated based on whether they reflected increases or decreases in ED visits for a given wildfire smoke concentration. Cumulative increases (blue) and decreases (red) are plotted along with the net response (black). **b.** Each barplot highlights which diagnoses the estimated increases and decreases in ED visits in response to wildfire smoke come from that are shown in the left panel. Top right panel shows when smoke $PM_{2.5}$ is 10 μgm^{-3} ED visits for most diagnoses increase, with 28% of the increase coming for visits for undiagnosed symptoms. At this smoke concentration only accidental injuries and infections show declines in ED visits with most of the decline coming from injuries. At high smoke $PM_{2.5}$=50μgm^{-3} (bottom panel) only ED visits for respiratory conditions increase. Visits for all other diagnoses decline with the largest decline coming from fewer visits for accidental injuries.

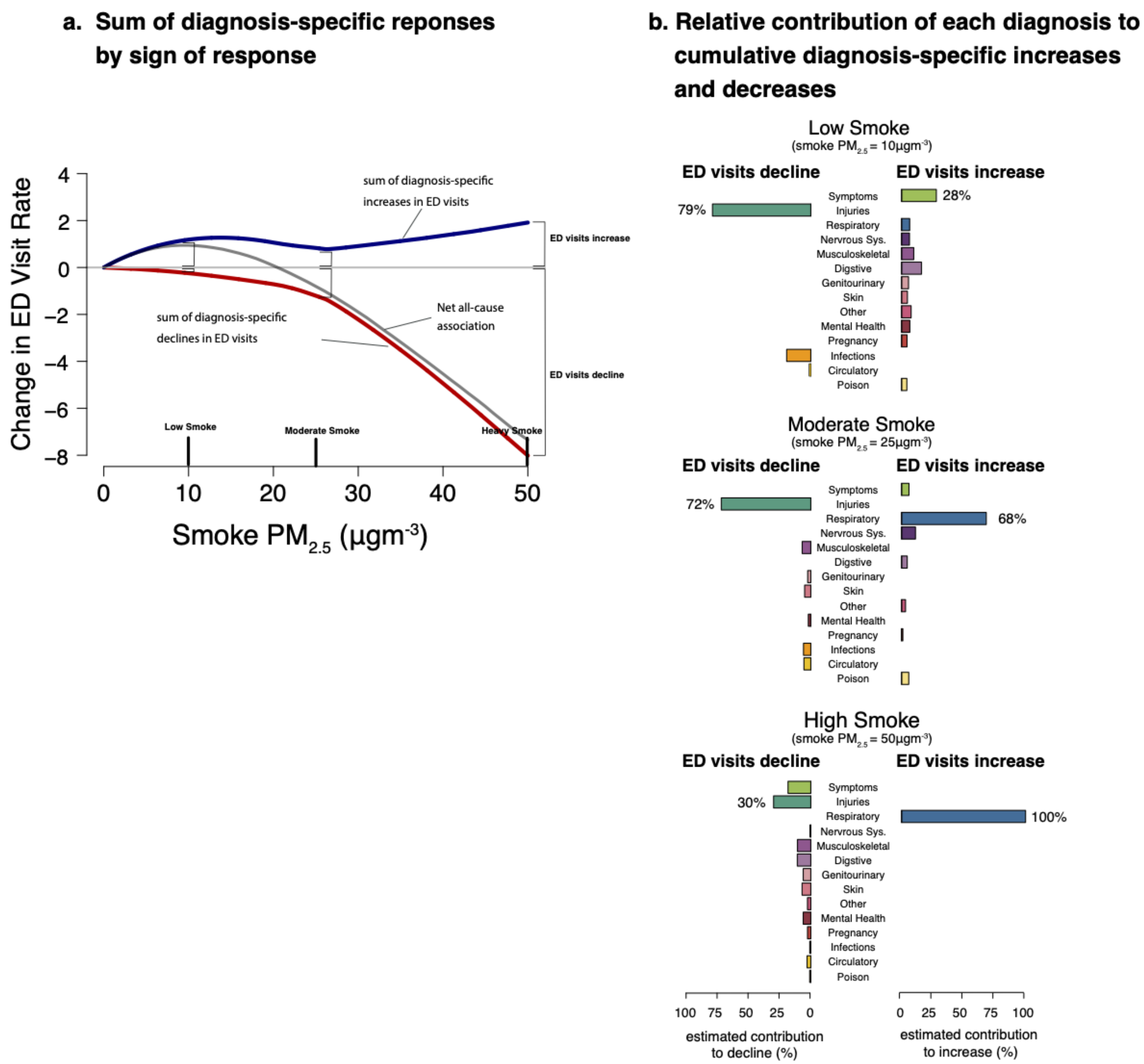

Fig. S13. Cumulative effects of wildfire smoke on ED visits for different symptoms. The response in each panel comes from a different regression where ED visits with a given principal diagnosis are the outcome and we estimate the binned model described in Methods. Vertical lines indicate Bonferroni-corrected 95% confidence intervals. ICD groupings correspond to Figs 1b and S8 and all ICD code groupings are listed in the supplement. Base rates shown are daily population weighted average ED visits per 100,000 people in a zipcode with that outcome as the primary diagnosis. Acronyms in labels: CVD = cardiovascular system.

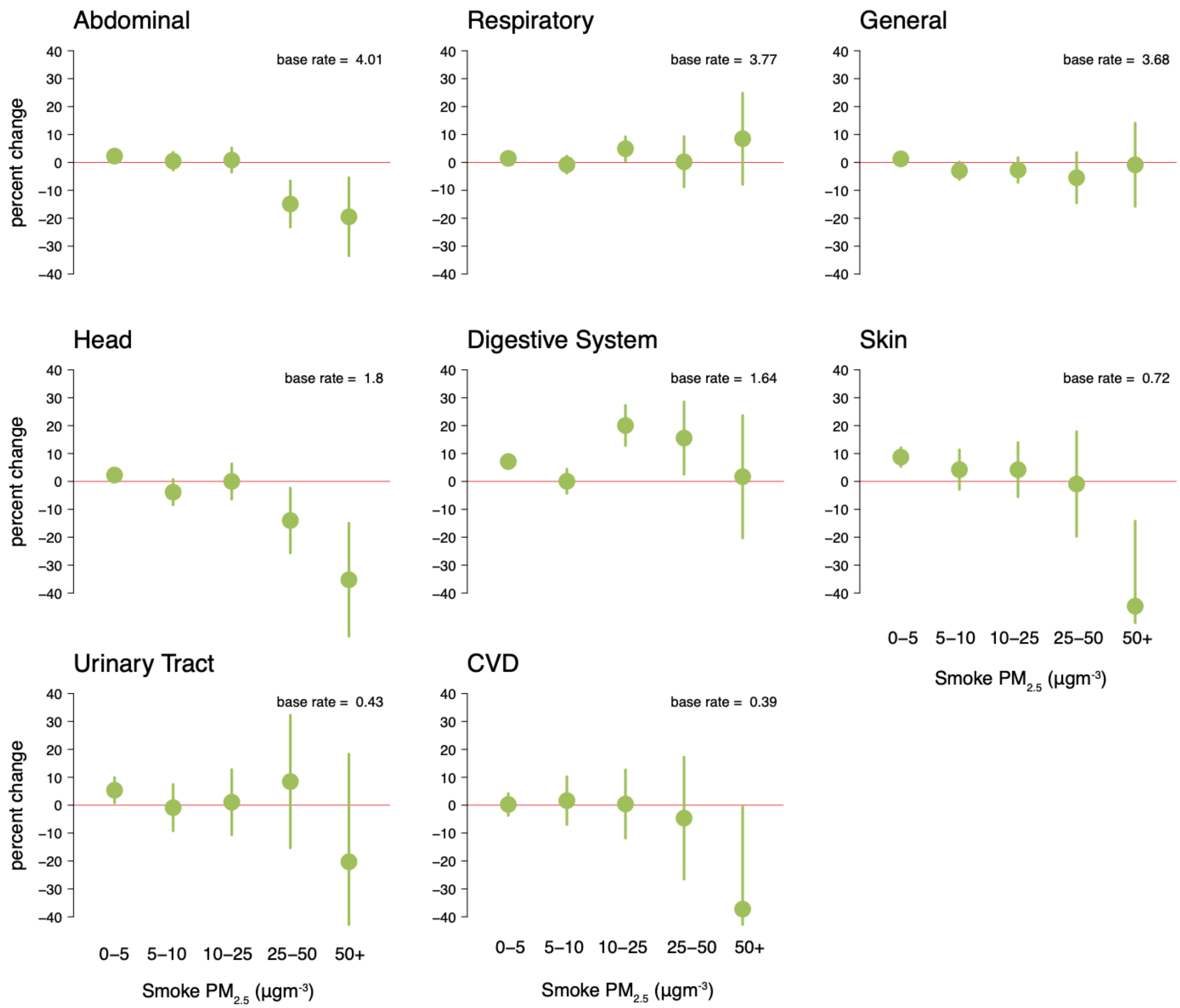

Fig. S14. Cumulative effects of wildfire smoke on ED visits for different types of accidental injuries. The response in each panel comes from a different regression where ED visits with a given principal diagnosis are the outcome and we estimate the binned model described in Methods. Vertical lines indicate Bonferroni-corrected 95% confidence intervals. ICD groupings correspond to Figs 1b and S8 and all ICD code groupings are listed in the supplement. Base rates shown are daily population weighted average ED visits per 100,000 people in a zipcode with that outcome as the primary diagnosis.

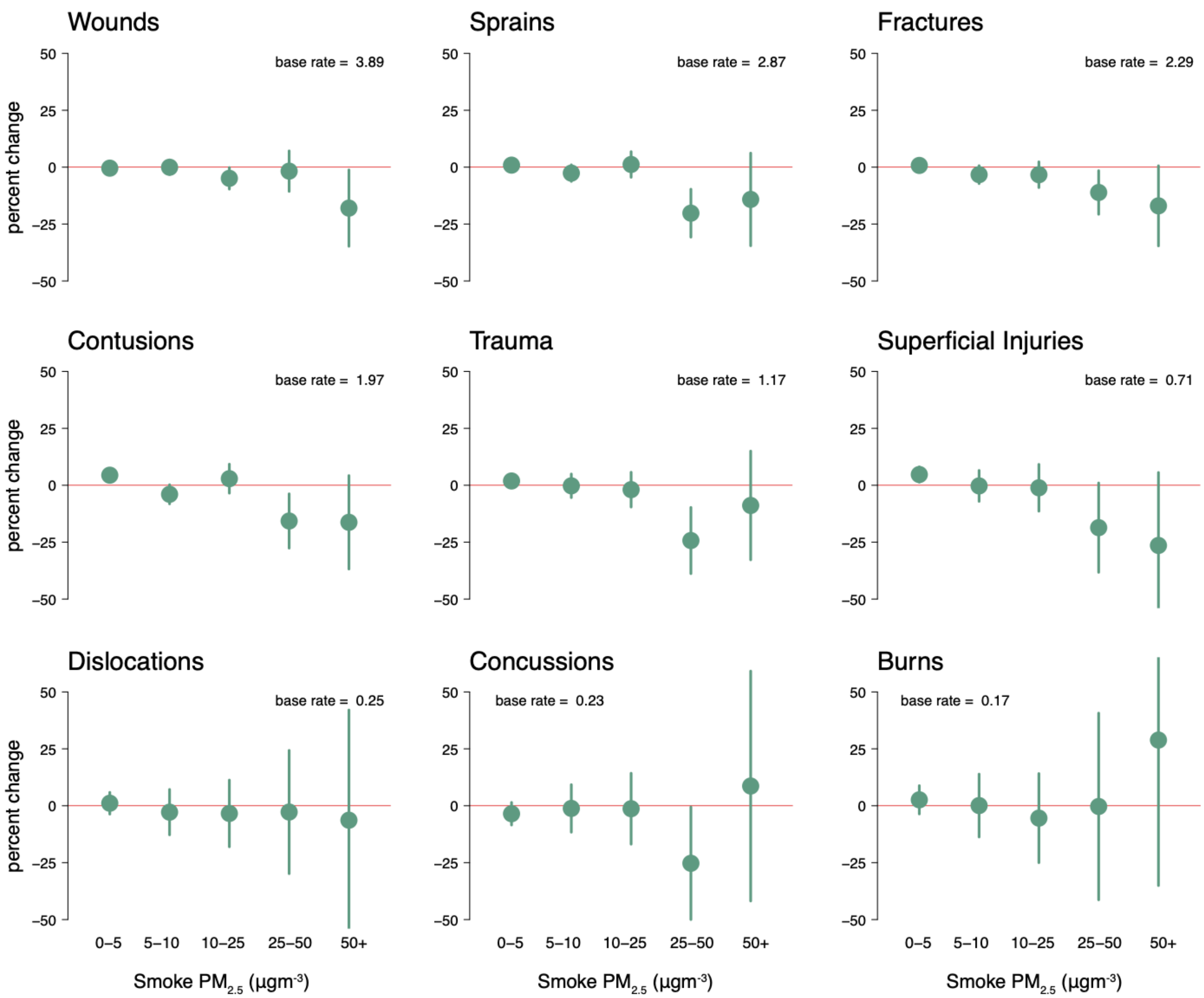

Fig. S15. Cumulative all-cause ED visit response to wildfire smoke by age group. The cumulative effects of wildfire smoke on all-cause ED visits for different age groups. Results come from five different regressions, one for each age group, with age specific zip by day ED visit rates as the outcome. Regressions are weighted by the group-specific zip code population. Cumulative responses show the impact of wildfire smoke $PM_{2.5}$ in the week following smoke exposure.

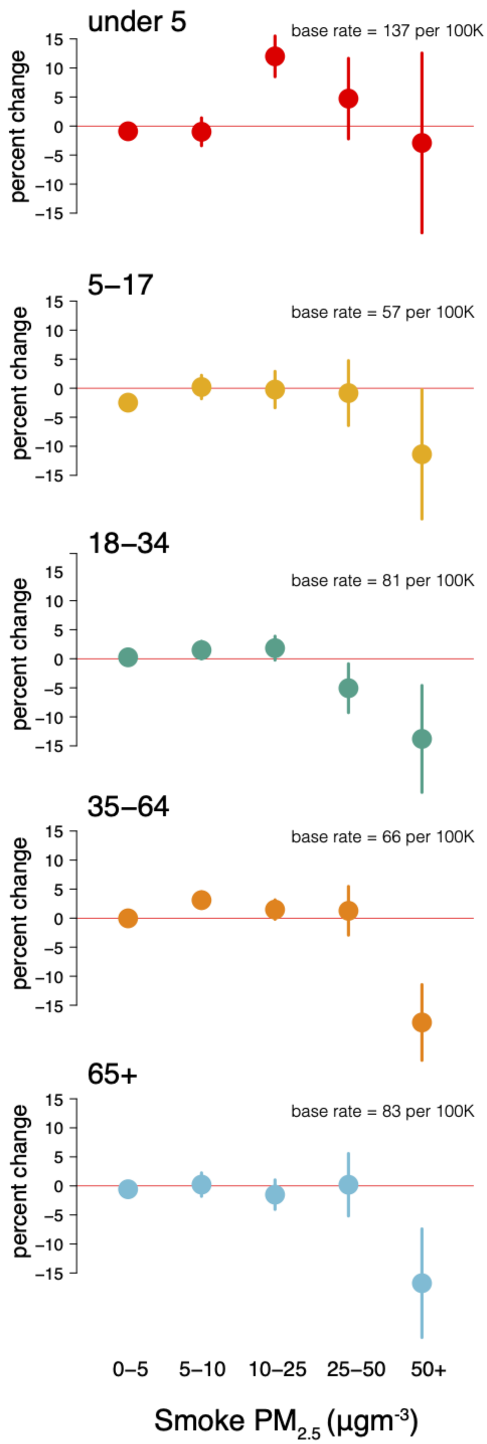

Fig. S16. Insurance coverage is correlated with other measures of vulnerability. The distribution of different factors by tercile of zipcode level insurance coverage. **a.** shows the insurance coverage sample split used to generate the splits in all subsequent panels (and used in Fig 4). The population that is least insured is more likely to receive insurance from a public provider (**b**), has higher rates of ED visits (**c**), has lower income (**e**) and higher rates of non-English speakers (**f**). The distance from home zip code to the nearest ED is similar across groups (**d**). Colored lines on the x-axis indicate median values for each tercile of insurance coverage.

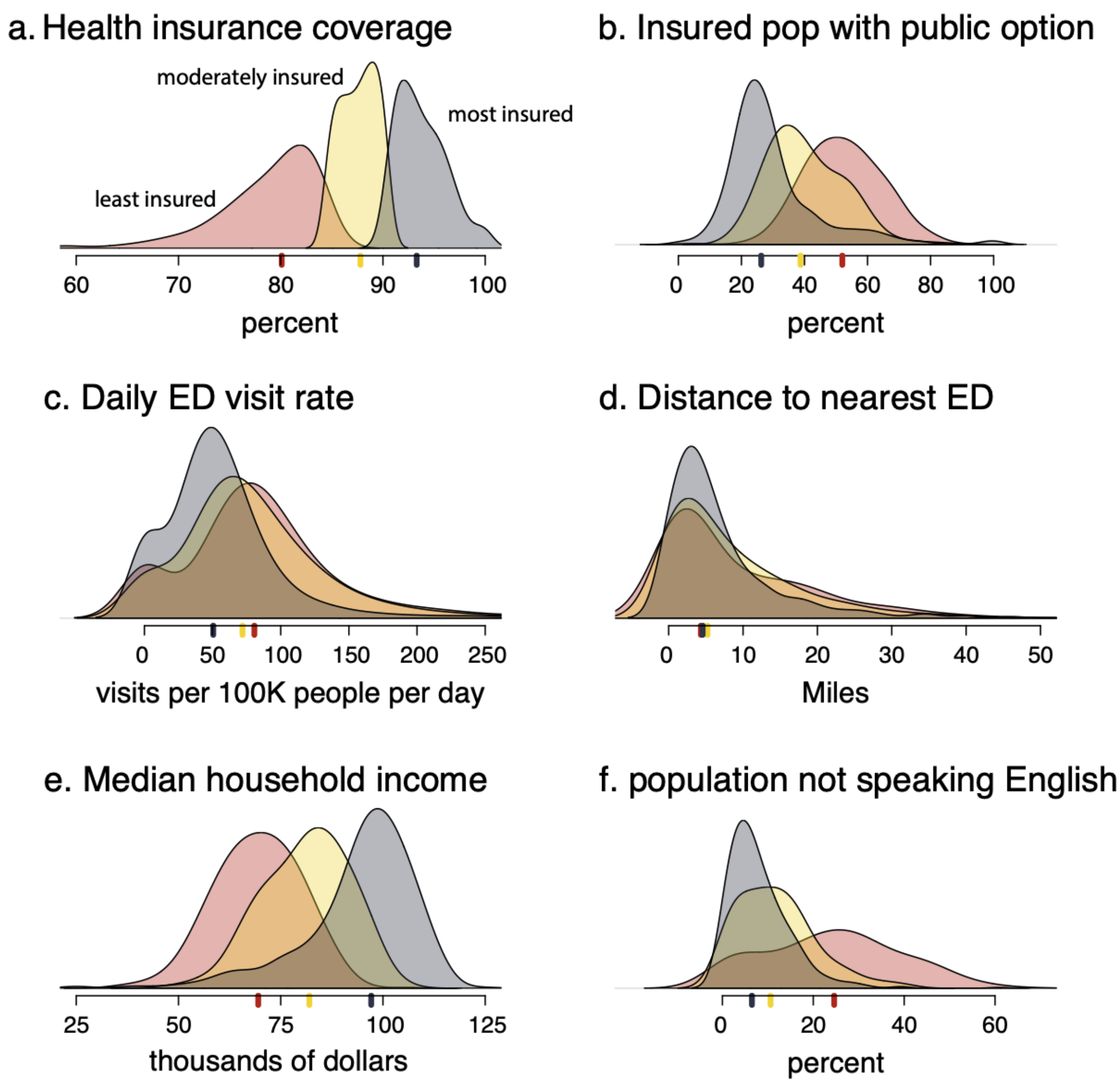

Fig. S17. Annual all-cause ED visits attributable to wildfire smoke. a. We apply the response curve shown in Fig 2 to past smoke $PM_{2.5}$ concentrations in order to estimate the annual number of excess ED visits attributable to smoke during our sample period (estimated attributions using alternative model specifications are shown in Fig S18). The height of the bar indicates the main estimate and the vertical lines show the 95% CI. While our estimated response curve indicates sharp declines in total ED visits at high smoke intensities, most smoke-days are low to medium intensity which leads to an overall increase in ED visits attributable to smoke. **b.** Population-weighted average exposure by smoke intensity. While 2008 and 2016 had a similar number of total days with smoke exposure, there were far more high intensity smoke-days in 2008 leading to that year having less than half the estimated attributable ED visits as 2016.

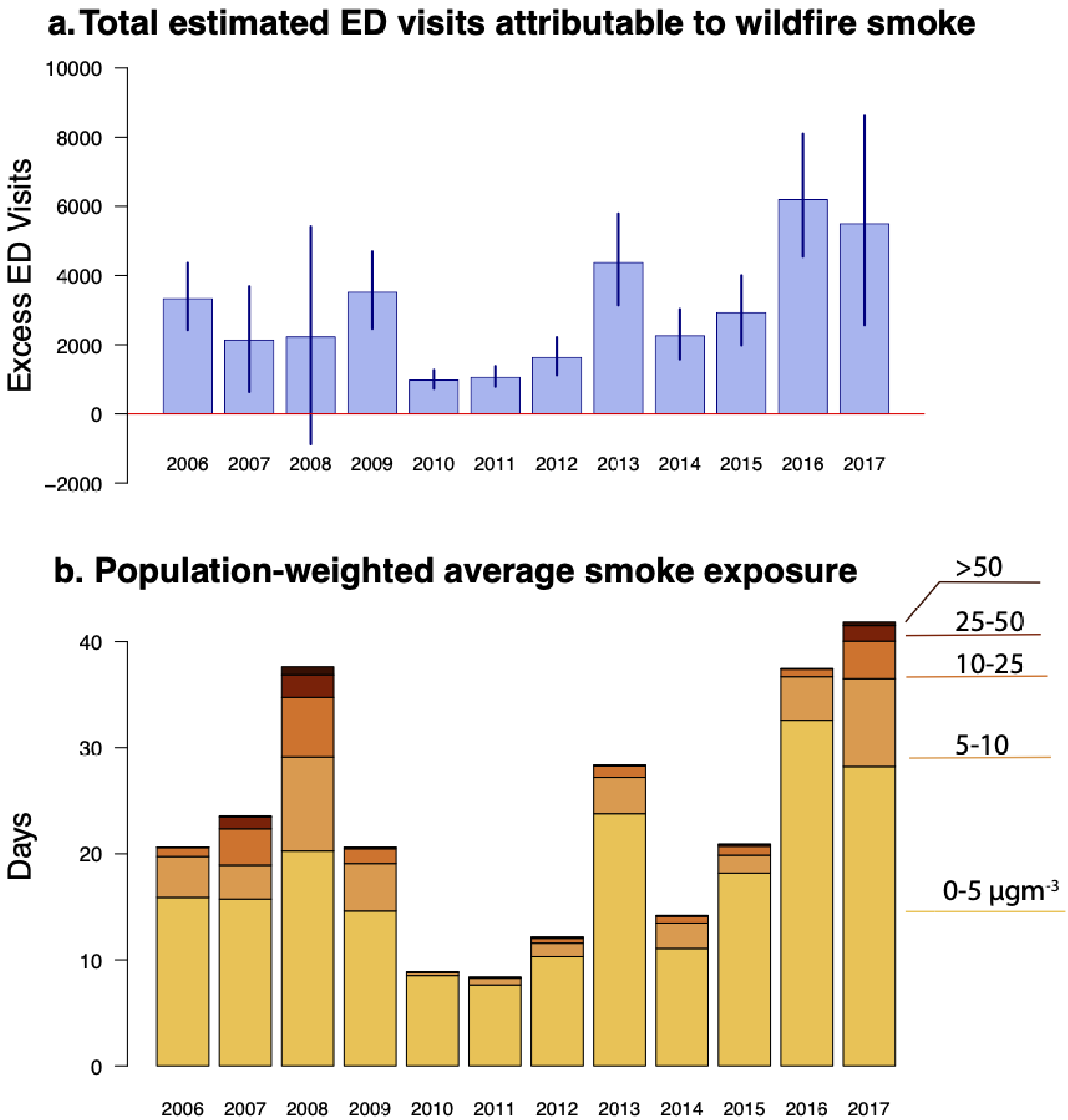

Fig. S18. Sensitivity of all-cause ED visit attribution estimates to model specification and number of lags included. a. Average annual attribution estimates were re-calculated utilizing response curves estimated from models with different specifications (Fig S4) and compared to our main estimate and associated 95% confidence interval (shown in blue). Elements of model specification varied include functional form, choice of seasonal and time fixed effects, and choice of controls. 10 of 14 model specifications produce estimates within the confidence interval of our main model. Specifications that less flexibly model seasonality or time trends produce lower estimates and specifications that omit temperature as a control produce higher estimates. **b.** Average annual attribution estimated from models with different numbers of lags included indicate models with small number of lags may underestimate total effects.