Research Report 238

Ambient Air Pollution and COVID-19 in California

Michael Kleeman et al.

Appendix D. Supplemental Information for Chapter 6: Air Pollution and Progression of COVID-19: A Multistate Analysis

---

Appendix D was reviewed by the HEI Review Committee and has been lightly edited for spelling, grammar, punctuation, and cross-references to the main report.

Correspondence may be addressed to Dr. Michael Kleeman, University of California, Davis, Department of Civil and Environmental Engineering, 1 Shields Avenue, Davis, CA 95616; email: *mjkleeman@ucdavis.edu*.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83998101 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not necessarily reflect the views of the Agency; thus, no official endorsement by it should be inferred. It also has not been reviewed by private-party institutions, including those that support the Health Effects Institute, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

CONTENTS

Variable Definitions for the Multistate Analysis

We prepared the data for the R mstate model.[1,2] Below, we explain how the multistate and Cox model variables were created for the analysis.

Table D1. Primary Variables for mstate Analysis

det	Whether or not a patient ever* deteriorated (0 or 1)
det.t	Time at which a patient first deteriorated (0.5 to 90 days), if they ever deteriorated; otherwise, time of last follow-up**
died	Whether or not a patient ever* died
died.t	Time at which a patient died
recov	Whether or not a patient was discharged alive (i.e., not a "discharge" by virtue of being dead)
recov.t	Time at which patient was *last* discharged (last hospitalization within 90 days)***

*In this context, "ever" is defined as within 90 days AND before any claims-related hospitalizations outside the Kaiser system. Anything after 90 days or after a claims-related hospitalization is censored. A deterioration event among hospitalized patients is defined as an event requiring high-flow oxygen therapy, bilevel positive airway pressure (BiPaP) ventilation, mechanical ventilation, intensive care unit (ICU) admission, or intensive-level care. Intensive-level care is intensive care that is provided outside the physical ICU and is defined by a higher staffing ratio than usual care.

**Last follow-up is either time of death, time of entrance into claims hospital, or end of study follow-up (90 days).

***Only hospitalizations that occurred before any claims-related hospitalizations are relevant.

Times *(.t*) are in days. For events that happen on the same day, give a value +0.5 (i.e., if the patient deteriorated on the same day they are hospitalized, *det.t* = 0.5; if they deteriorated on day 2 and died on day 2, then assign *died.t* = 2.5).

All patients must have a first hospitalization within KPSC within 21 days of a positive COVID-19 test. If their first hospitalization is claims-related, they are excluded from the cohort.

If patients have a subsequent hospitalization (second, third, etc.) at a non-KPSC facility, then we exclude any information that happens after the non-KPSC hospitalization and then:

The time for an event that did not happen (i.e., definitely death and possibly deterioration) is the time from the index date to this non-KPSC hospitalization, and the event indicator variable for the event is 0. *T0:* index time starts at the first hospitalization.

All members of the cohort have a 0/1 value for each of *det*, *died*, and *recov*, and all have a time between 0.5-90 for the variables that end in *.t*

Clarifying Examples

If the patient dies on day 10 without ever deteriorating or being discharged, then the patient's values are: *det* = 0, *det.t* = 10, *died* = 1, *died.t* = 10, *recov* = 0, *recov.t* = 10.

If the patient deteriorates on day 2, is discharged on day 5, and dies on day 10, then the patient's values are: *det* = 1, *det.t* = 2, *died* = 1, *died.t* = 10, *recov* = 1, *recov.t* = 5.

If the patient is discharged alive at day 5 and never dies, then the patient's values are: *det* = 0, *det.t* = 90, *died* = 0, *died.t* = 90, *recov* = 1*, recov.t* = 5.

If the patient is discharged alive at day 5 and is then hospitalized in a claims-related hospitalization at day 20, then the patient's values are: *det* = 0, *det.t* = 20, *died* = 0, *died.t* = 20, *recov* = 1, *recov.t* = 5.

Cox Analysis of Transitions Between Events

There are 6 Cox models (and thus 6 separate indicator/time variables).

The indicator variables are: *det*, *recov*, and *died*, as used in the mstate model described above. No additional indicator variables are needed for the Cox model, and the same variable definitions are used for the indicators in the mstate and Cox models.

The time variables are labeled to represent the boxes between which the arrows are drawn in Figure 15 of the main Investigators' Report. The boxes represent the following:

1. Hospitalization
2. Deterioration
3. Recovery
4. Death

So, we have arrows from

1 -> 2 (*hosp -> det*) (*time_hosptodet*)

1 -> 3 (*hosp -> recovery*) (*time_hosptorecov*)

1 -> 4 (*hosp -> died*) (*time_hosptodied*)

2 -> 3 (*det -> recovery*) (*time_dettorecov*)

2 -> 4 (*det -> died*) (*time_dettodied*)

3 -> 4 (recovery -> died) (*time_recovtodied*)

These have the same rules as described above for the multistate model (i.e., the definitions of what happens with Claims data are the same, etc.), but the times are slightly different. In the multistate variables above (e.g., *died.t*), the time is from time 0 (index date = first hospitalization date). In the Cox time variables, we only measure the relative time from the previous event. Thus, if someone was hospitalized on day 0, intubated on day 5, and died on day 7, the multistate variable *died.t* = 7. But in the Cox variable *time_dettodied*, the interval is only 2 days (7 days – 5 days).

These differences in censoring criteria and times between the multistate and Cox models that were run solely for the death outcome result in two different sample sizes for these two analyses. Specifically, the first full cohort analysis focused on the mortality outcome among patients hospitalized with COVID-19 (N = 21,415) who proceeded to death (N = 4,815), allowing patients to be hospitalized anywhere (KPSC hospital or another hospital with a claim sent to KPSC). For the multistate cohort, however, we had to censor/remove patients who were hospitalized outside of KPSC, because we did not have accurate records on their treatment regimens in those hospitals and, therefore, could not assess the severity of deterioration. This is reflected in the reduced number of COVID-19–related hospitalizations (N = 15,978) analyzed in the multistate analysis.

Confounder Analysis

See Tables D2 through D4 on the next pages.

Table D2. Confounders Identified for $PM_{2.5}$ Mass

	Transition 1: *hosp -> det*	Transition 2: *hosp -> recov*	Transition 3: *hosp -> death*	Transition 4: *det -> recov*	Transition 5: *det -> death*	Transition 6: *recov -> death*
Smoking status	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
BMI (linear and squared terms)	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE
Medicaid	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
Exercise Vital Sign	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
Proportion of housing units with >1 occupants per room (Census)	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE
Neighborhood Deprivation Index	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
Proportion of workers age 16+ commuting by public transportation (Census)	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
Relative humidity (%)	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
Temperature (°C)	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE
NDVI	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE

Table D3. Confounders Identified for Ozone

	Transition 1: *hosp -> det*	Transition 2: *hosp -> recov*	Transition 3: *hosp -> death*	Transition 4: *det -> recov*	Transition 5: *det -> death*	Transition 6: *recov -> death*
Smoking status	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
BMI (linear and squared terms)	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE
Medicaid	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
Exercise Vital Sign	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
Proportion of housing units with >1 occupants per room (Census)	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
Neighborhood Deprivation Index	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
Proportion of workers age 16+ commuting by public transportation (Census)	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
Relative humidity (%)	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Temperature (°C)	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE
NDVI	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE

Table D4. Confounders Identified for NO_2

	Transition 1: *hosp -> det*	Transition 2: *hosp -> recov*	Transition 3: *hosp -> death*	Transition 4: *det -> recov*	Transition 5: *det -> death*	Transition 6: *recov -> death*
Smoking status	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
BMI (linear and squared terms)	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
Medicaid	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
Exercise Vital Sign	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE
Proportion of housing units with >1 occupants per room (Census)	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
Neighborhood Deprivation Index	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
Proportion of workers age 16+ commuting by public transportation (Census)	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
Relative humidity (%)	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE
Temperature (°C)	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE
NDVI	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE

Table D5. Descriptive Characteristics of Patients Excluded from Mortality Cohort in Final Multistate Cohort

		Excluded from multistate	
Characteristic	Overall, *N* = 21,415[1]	0, *N* = 15,9778[1]	1, *N* = 5,438[1]
Age at diagnosis (years)	64 (52, 75)	63 (51, 74)	68 (56, 78)
Race/ethnicity			
White	4,861 (23%)	3,142 (20%)	1,719 (32%)
Asian/Pacific Islander	2,281 (11%)	1,810 (11%)	471 (8.7%)
Black	1,851 (8.6%)	1,342 (8.4%)	509 (9.4%)
Hispanic	12,077 (56%)	9,538 (60%)	2,539 (47%)
Other/Multiple/Unknown	345 (1.6%)	145 (0.9%)	200 (3.7%)
Gender			
Female	9,067 (42%)	6,700 (42%)	2,367 (44%)
Male	12,348 (58%)	9,277 (58%)	3,071 (56%)
Medicaid			
No	18,722 (87%)	13,960 (87%)	4,762 (88%)
Yes	2,693 (13%)	2,017 (13%)	676 (12%)
Smoking status			
Never-Smoker	13,392 (63%)	10,246 (64%)	3,146 (58%)
Ever-Smoker	7,738 (36%)	5,542 (35%)	2,196 (40%)
Unknown	285	189	96
$PM_{2.5}$ mass (μ/m^3)	12.30 (10.50, 14.00)	12.60 (11.00, 14.20)	11.50 (9.56, 13.20)
NO_2 (ppb)	21 (13, 25)	22 (14, 25)	17 (11, 25)
O_3 (ppb)	66 (60, 72)	66 (60, 73)	66 (59, 71)
$PM_{2.5}$ mass (μ/m^3)/IQR	3.51 (3.00, 4.00)	3.60 (3.14, 4.06)	3.29 (2.73, 3.77)
NO_2 (ppb)/IQR	1.76 (1.14, 2.14)	1.84 (1.23, 2.14)	1.47 (0.92, 2.10)
O_3 (ppb)/IQR	5.40 (4.93, 5.93)	5.40 (4.94, 5.96)	5.42 (4.86, 5.84)

[1]Median (IQR); n (%), proportion or native units as applicable

Table D6. Frequency of Patients by Transition Event

	Hospitalized	Deterioration	Recovery	Death	No event	Total
Hospitalized	0	6390	9256	62	270	15978
Deterioration	0	0	3800	2414	176	6390
Recovery	0	0	0	470	12586	13066
Death	0	0	0	0	2946	2946

Table D7. Proportion of Patients by Transition Event

	Hospitalized	Deterioration	Recovery	Death	No event
Hospitalized	0	0.400	0.579	0.004	0.017
Deterioration	0	0	0.595	0.378	0.028
Recovery	0	0	0	0.036	0.964
Death	0	0	0	0	1

Table D8. Cox Model Results for $PM_{2.5}$,NO_2, O_3 and Confounders, Standardized by IQR

	1. *hosp -> det*, Nobs = 15978[1], Nevents = 6390[2]		**2. *hosp -> recovered***, Nobs = 15978[1], Nevents = 9256[2]		**3. *hosp -> died***, Nobs = 15978[1] Nevents = 62[2]		**4. *det -> recovered***, Nobs = 6390[1], Nevents = 3800[2]		**5. *det -> died***, Nobs = 6390[1], Nevents = 2414[2]		**6. *recovered -> died***, Nobs = 130666 [1], Nevents = 470[2]	
Characteristic	**HR[3]**	**95% CI[3]**	**HR[3]**	**95% CI[3]**	**HR[3]**	**95% CI[3]**	**HR[3]**	**95% CI[1]**	**HR[3]**	**95% CI[3]**	**HR[1]**	**95% CI[1]**
$PM_{2.5}$ mass (μ/m^3)	1.17**	1.13, 1.21	1.00	0.97, 1.03	0.76	0.53, 1.09	0.96	0.92, 1.01	1.09**	1.03, 1.15	1.10	0.98, 1.24
NO_2 (ppb)	1.20**	1.14, 1.25	1.01	0.97, 1.04	0.63*	0.43, 0.92	1.04	0.98, 1.10	1.06	0.98, 1.15	1.03	0.86, 1.22
O_3 (ppb)	1.22**	1.14, 1.30	0.94*	0.89, 0.99	1.53	0.90, 2.60	0.98	0.92, 1.06	1.08	0.99, 1.19	1.25*	1.04, 1.50

[1] Nobs are total cases by event (hospitalization, deterioration, discharge, or death). See Figure 15 in the Investigators' Report for a graphic illustration of states and transitions.

[2] Nevents is the number of patients who deteriorated to another status. See Figure 15 in the Investigators' Report for a graphic illustration of states and transitions.
[3] HR = Hazard Ratio, CI = Confidence Interval, Stratified by age group (5-year), sex, and race/ethnicity.

* $P \leq 0.05$ (95% CI).

** $P \leq 0.01$ (99% CI).

Table D9. Pollutants Fully Adjusted with All Confounders Included A Priori

	1. *hosp -> det*, Nobs = 15978[1], Nevents = 6390[2]		**2. *hosp -> recovered***, Nobs = 15978[1], Nevents = 9256[2]		**3. *hosp -> died***, Nobs = 15978[1] Nevents = 62[2]		**4. *det -> recovered***, Nobs = 6390[1], Nevents = 3800[2]		**5. *det -> died***, Nobs = 6390[1], Nevents = 2414[2]		**6. *recovered -> died***, Nobs = 13066[1], Nevents = 470[2]	
Characteristic	**HR[3]**	**95% CI[3]**	**HR[3]**	**95% CI[3]**	**HR[3]**	**95% CI[3]**	**HR[3]**	**95% CI[1]**	**HR[3]**	**95% CI[3]**	**HR[1]**	**95% CI[1]**
$PM_{2.5}$ mass (μ/m^3)	1.14**	1.10, 1.18	1.0	0.97, 1.03	0.72	0.49, 1.07	0.97	0.92, 1.02	1.10**	1.04, 1.17	1.10	0.97, 1.24
NO_2 (ppb)	1.18**	1.13, 1.23	1.01	0.98, 1.04	0.59*	0.38, 0.91	1.04	0.98, 1.11	1.07	1.00, 1.15	1.03	0.87, 1.22
O_3 (ppb)	1.17**	1.10, 1.24	0.96	0.92, 1.00	1.54	0.92, 2.58	0.97	0.90, 1.05	1.09	0.99, 1.20	1.26*	1.01, 1.57

[1] Nobs are total cases by event (hospitalization, deterioration, discharge, or death). See Figure 15 in the Investigators' Report for a graphic illustration of states and transitions.

[2] Nevents is the number of patients who deteriorated to another status. See Figure 15 in the Investigators' Report for a graphic illustration of states and transitions.
[3] HR = Hazard Ratio, CI = Confidence Interval, Stratified by age group (5-year), sex, and race/ethnicity.

* $P \leq 0.05$ (95% CI).

** $P \leq 0.01$ (99% CI).

Analysis of Patients Who Transitioned from Hospital to Death

We followed established protocols and defined COVID-19 hospitalizations as hospitalizations occurring within 21 days of a positive COVID-19 test or diagnosis.[3–6] Thus, patients who are hospitalized for other reasons and happen to test positive for COVID-19 are included in the cohort of hospitalized patients. We found that these patients had a higher median age (85 years compared to 72 years) and a higher comorbidity score (median Elixhauser score of 6 compared to 4); all but 4 patients had a do-not-resuscitate order on file, compared to patients who died after a deterioration event. These patients died much more quickly, with a median time of only 5 days compared to 15 days in other statuses. We further conducted chart reviews of 12% of these cases (7 charts) and confirmed that patients in this group died of non-COVID-19-related causes, such as stage 4 breast cancer or renal failure. Thus, the unexpected finding may have resulted from some misclassification of COVID-19 deaths in frail older adult patients who happened, by chance, to live in areas with lower $PM_{2.5}$ and NO_2 and who were likely to die from other causes in the near future.

Table D10. Summary of Patient Health Record Data by Key Event Transitions

		Transition 3: *hosp -> death*	**Transition 4: *det -> recovered***	**Transition 5: *det -> death***	**All Deteriorations**	**Transition 2: *hosp -> recovered***
	N	62	3800	2414	6214	9256
Age	Q1	77	50	63	55	49
	Median	85	60	72	65	61
	Q3	89	70	80	75	73
Time from Admission to Death	Q1	4	NA	9	NA	NA
	Median	5	NA	15	NA	NA
	Q3	8	NA	22	NA	NA
Elixhauser Index	Q1	4	1	2	1	1
	Median	6	2	4	3	2
	Q3	10	4	6	5	4
BMI	Q1	20.7	28.31	26.53	27.5	26.77
	Median	24.245	32.28	30.48	31.54	30.76
	Q3	26.34	37.74	35.5	36.88	35.9
DNI Percentage	DNR with OK to intubate	1.5%	2.9%	33.3%	15.2%	0.6%
	DNR/DNI/comfort measures only	62.7%	1.2%	41.0%	17.3%	1.0%
	DNR/DNI/selective treatment	29.9%	6.6%	14.4%	9.7%	8.1%
	Full code/all procedures ok	6.0%	89.3%	11.3%	57.8%	90.2%

DNI = do not intubate; DNR = do not resuscitate; Q1 = first quartile; Q3 = third quartile.

References

1. de Wreede LC, Fiocco M, Putter H. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine*. 2010;99(3):261-274. doi:10.1016/j.cmpb.2010.01.001

2. Mortensen VH, Søgaard M, Mygind LH, Wolkewitz M, Kristensen B, Schønheyder HC. Incidence and mortality of hospital-acquired bacteraemia: a population-based cohort study applying a multi-state model approach. *Clinical Microbiology and Infection*. 2021;28(6):879-e9. doi:10.1016/j.cmi.2021.12.011

3. Nau C, Bruxvoort K, Navarro RA, et al. COVID-19 Inequities Across Multiple Racial and Ethnic Groups: Results From an Integrated Health Care Organization. *Annals of internal medicine*. Published online April 20, 2021. doi:10.7326/m20-8283

4. Tartof SY, Qian L, Hong V, et al. Obesity and mortality among patients diagnosed with COVID-19: results from an integrated health care organization. 2020;173(10):773-781.

5. Escobar GJ, Adams AS, Liu VX, et al. Racial Disparities in COVID-19 Testing and Outcomes. *Ann Intern Med*. 2021;174(6):786-793. doi:10.7326/M20-6979

6. Sharp AL, Huang BZ, Broder B, et al. Identifying patients with symptoms suspicious for COVID-19 at elevated risk of adverse events: The COVAS score. *The American Journal of Emergency Medicine*. 2021;46:489-494. doi:10.1016/j.ajem.2020.10.068