

The result of this exploratory analysis is a report that consists of the following:

- 1) One paragraph overview of background and some small number of primary research questions of interest.

The 2020 COVID-19 global lockdowns reportedly led to decreased ambient particulate matter (PM2.5) concentrations¹. The lockdowns were generally associated with changes in human mobility patterns, including increased time spent in residential locations and a decreased concentration of some individual's visiting workplaces.² However, the extent of these changes varied geographically³, turning the lockdowns into a natural experiment to assess how distinct community mobility patterns influence PM2.5 concentrations. Primary sources of PM2.5 vary geographically and are associated with country development status. Traffic is a primary contributor to PM2.5 in Western European countries, while industrial sources are a main contributor in lower-middle income countries, primary because of the reliance on heavily polluting industries, a hallmark feature of early stage economic development.⁴ Further, in low-middle-income countries, a substantial portion of PM2.5 concentration arises from household cooking with solid fuels—a source that is largely absent in high-income settings. This underscores a crucial distinction in pollution sources between stages of economic development.

Our goal is to evaluate how changes in mobility patterns impact PM2.5 concentrations, particularly changes in attendance at workplaces and time spent at home. We hypothesize that there will be differential effects of mobility changes on PM2.5 concentrations according to country development status. Specifically, we predict that decreases in workplace attendance will be associated with larger decreases in PM2.5 in lower-middle income countries as compared to high income countries, reflecting the differences in prevalence and reliance on heavily polluting industries. We also predict that increased time in residential locations will be associated with either an increase, or less dramatic decrease in PM2.5 in lower-middle income countries as compared to high income countries, reflecting increased frequency and/or quantity of cooking with solid fuels. These findings may help inform how human mobility patterns lead to disproportionate exposure to PM2.5 in low-middle income settings.

- 2) One paragraph describing your data sources: What is the structure of your data files (e.g., student level, school-level, teacher-level, county-level)? Do the variables come from surveys, administrative records, classroom observations, etc.? Please feel free to discuss any data decisions you've made or are considering (e.g., you're dropping certain observations, creating new measures, or are unsure of which measures to use). Make sure you include the sample sizes at each level (e.g., number of time points per unit, average number of students per school, etc.).

The PM2.5 data is available for 55 global monitoring sites through the EPA's AirNow program⁵, some countries have multiple monitoring sites, but the majority have a single monitoring site. The concentrations are available as daily averages, spanning back to 2017 for some sites, through October 2020 (the PM2.5 data is continuously updated but a compiled dataset must be formally requested and our most recent request spans only through October 2020). The mobility data is available through Google's COVID-19 Community Mobility Reports⁶, and is available from February 2020 through October 2022. Per Google, the data charts "movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential" during the global COVID-19 pandemic⁶. The data is presented as a percent change from baseline, where "A baseline day represents a normal value for that day of the week" prior to the COVID-

Commented [JR1]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9117801/>

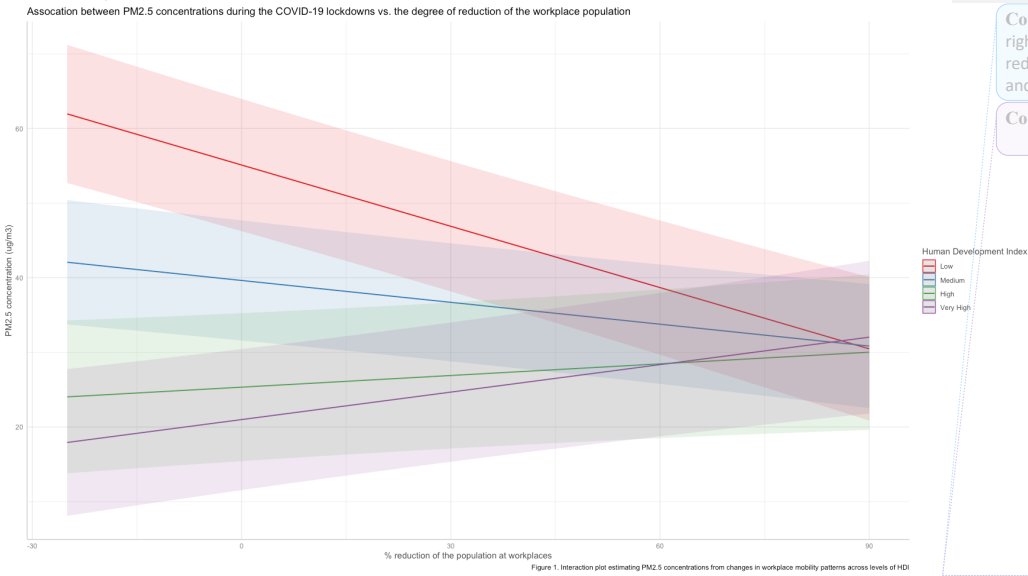
Commented [JR2]: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022GH000767>

19 pandemic⁷. For change in residential patterns, the value is presented as a percent change in time spent at residential locations, whereas for workplaces the value represents a percent change in concentration of people. The mobility data is available at either the city or country level. We were able to link mobility data to 37 PM2.5 monitoring sites, for the time period February 2020-October 2020. We linked data at the highest resolution as was possible for the individual monitoring site. For example, if mobility data was available at the city level it was utilized, if not we utilized the country level values (Table 1), we recognize this limitation and will discuss further in the limitation section. We also controlled for population density and the interaction of Human Development Index (HDI) with mobility pattern changes (both available from the World Bank^{7,8} at the country level).

| PM2.5 Monitoring Site | Mobility Site |
|-----------------------|-----------------------------|
| Abu Dhabi | Abu Dhabi |
| Accra | Accra Metropolitan Area |
| Almaty | Kazakhstan |
| Amman | Amman Metropolitan Area |
| Baghdad | Iraq |
| Bamako | Bamako Metropolitan Area |
| Bogota | Bogota |
| Buenos Aires | Buenos Aires |
| Chennai | Chennai |
| Colombo | Sri Lanka |
| Conakry | Guinea-Bissau |
| Dhahran | Nepal |
| Dhaka | Bangladesh |
| Dubai | Dubai |
| Dushanbe | Dushanbe Metropolitan Area |
| Embassy Kathmandu | Nepal |
| Guatemala City | Guatemala City |
| Hyderabad | Hyderabad |
| Islamabad | Islamabad Metropolitan Area |
| Jakarta Central | Jakarta |
| Jakarta South | Jakarta |
| Kabul | Kabul Metropolitan Area |
| Kampala | Kampala |
| Karachi | Karachi Metropolitan Area |

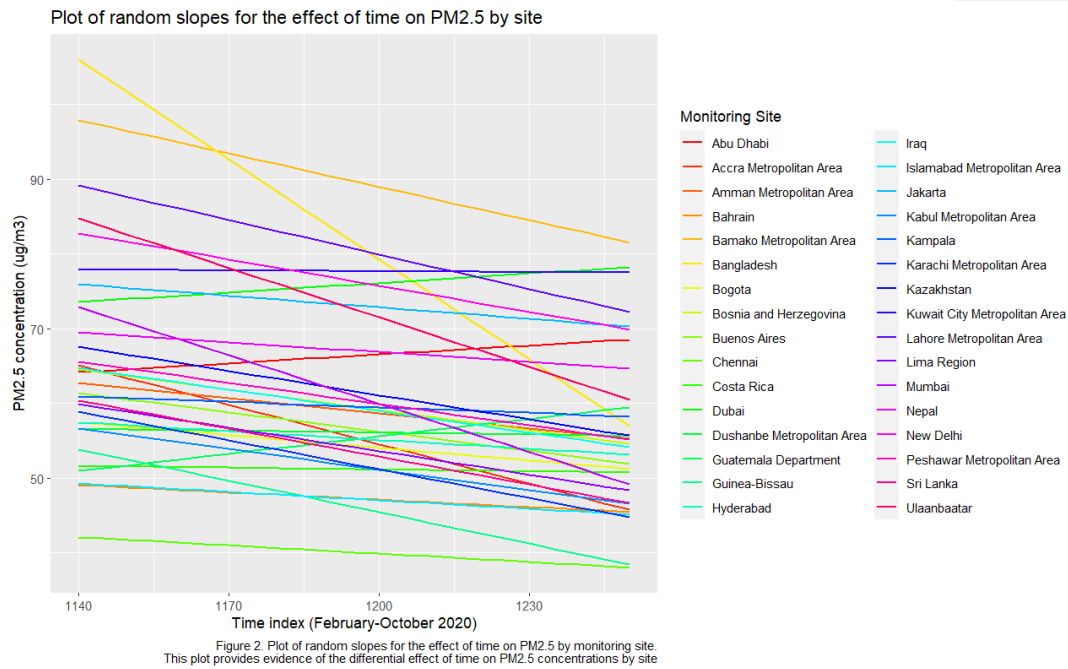
| | |
|-------------|-------------------------------|
| Kolkata | Kolkata |
| Kuwait City | Kuwait City Metropolitan Area |
| Lahore | Lahore Metropolitan Area |
| Lima | Lima Region |
| Manama | Bahrain |
| Mumbai | Mumbai |
| New Delhi | New Delhi |
| Nur-Sultan | Kazakhstan |
| Peshawar | Peshawar Metropolitan Area |
| San Jose | Costa Rica |
| Sarajevo | Bosnia and Herzegovina |
| Ulaanbaatar | Ulaanbaatar |
| Kabul | Kabul Metropolitan Area |

3) Two publication ready plots of your data. Publication ready means having things like nicely labeled axes and captions.



Commented [JR3]: I reversed the x axis so as you go to the right it actually shows a DECREASE in people at work (ie 90% reduction in people at work) I wanted to update the title and legend and stuff but rstudio keeps throwing me errors

Commented [XF4R3]: Changed the title and legend



4) An analysis of the variability in your data (primarily your outcome). You can do this by fitting an unconditional model (one without any extraneous covariates) to obtain a variance decomposition (allowing for ICC calculations and the like). These unconditional models are good starting point for getting a handle on where variation is, and what your data structure is.

| <i>Predictors</i> | <i>Estimates</i> | a_mean | |
|------------------------------------|------------------|---------------|----------|
| | | <i>CI</i> | <i>p</i> |
| (Intercept) | 47.73 | 39.33 – 56.14 | <0.001 |
| Random Effects | | | |
| σ^2 | 1721.86 | | |
| τ_{00} Mobility_SiteName | 602.56 | | |
| ICC | 0.26 | | |
| N Mobility_SiteName | 33 | | |
| Observations | 25832 | | |
| Marginal R^2 / Conditional R^2 | 0.000 / 0.259 | | |

5) One paragraph describing trends in your data, referring to your two plots and your variability analysis.

ICC: 26% of the unexplained variation is due to the variation across sites. The extent of site variation around the mean PM2.5 is estimated to be 602.56.

Descriptions of the plots and trends are included in part 7.

6) An initial mathematical model describing the primary model you are planning on fitting (or have fit). Try to focus attention on your primary covariates and outcome; keep these initial models simple and straightforward. As you write your model, be sure to define your subscripts! I.e., at the beginning of your model write something like "For time t for student i in school k we have..." Also, to be correct you would then subscript as tik , keeping your levels in order.

Commented [XF5]: Not too sure

The mathematical model (multilevel form)

For time t in site i

$$PM2.5_{ti} = \pi_{0i} + \pi_{1i}date_{ti} + \pi_{2i}month_{ti} + \pi_{3i}workplace_{ti} + \epsilon_{ti}$$

$$\epsilon_{ti} \sim N(0, \sigma^2)$$

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HDI_i + \gamma_{02}Population_i + u_{0i}$$

$$\pi_{1i} = \gamma_{10} + u_{1i}$$

$$\pi_{2i} = \gamma_{20} + \gamma_{21}hemisphere_i$$

$$\pi_{3i} = \gamma_{30} + \gamma_{31}HDI_i$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

At level 1 (time-varying) covariates: time, measured in days

$mobility_{ti}$: mobility data for each day

$month_{ti}$: month of the day

$workplace_{ti}$: workplace percentage change for each day

At level 2 (site specific invariant covariates):

HDI_i : HDI for each site

$population_i$: population density for each site

7) One paragraph of initial findings. If you have not yet fit a model, you can still describe preliminary findings in terms of trends, etc., in your plots and initial exploratory analysis.

We found a significant interaction between decreased concentrations of people at workplaces and human development index (HDI) on PM2.5 concentrations (Figure 1). We found that at sites with a low or medium HDI score, decreasing concentrations of people at workplaces was associated with drastic decreases in PM2.5 concentrations (Figure 1, red and blue lines). This likely reflects a decrease in emissions from highly polluting industry sources (e.g. manufacturing, agriculture) that were closed or reduced in activity during the COVID-19 lockdowns. In contrast, we see that in high and very high HDI settings, decreased time at work is associated with slight increases in PM2.5 concentrations. We were not expecting a predicted increase in PM2.5, hypothesizing instead that there would be a less dramatic decrease. We predicted that decreased workplace attendance would be associated with decreased PM2.5 concentrations in high income settings through a reduction in traffic. According to the literature, traffic generally did decrease in European settings, but at varying magnitudes, ranging from 25-75% reduction.⁹ It is possible that the reduction in traffic may not have been enough to offset the increased use of energy at residential locations (e.g. through heating, cooling, refrigeration, and lighting).

We found that the effect of time (February – October 2020) on PM2.5 concentrations varied widely by site (Figure 2). We expected this finding and attribute it in part to the differences in time to lockdown.

8) One paragraph describing next steps, blocks, barriers, concerns, or other things you would like to discuss and get feedback on.

Commented [JR6]: Need to write at the mathematical model for this

Commented [JR7]: <https://wjeb.biomedcentral.com/article/s/10.1186/s13017-021-00395-8>

Final Project EDA
Group 4

We still plan to run the analysis of change in time spent at residential locations on PM2.5 concentrations. We are also considering performing a principal component analysis of all the mobility variables available from google (change in concentration of people at retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and time at residential locations). We would then use the identified principal components as individual predictors in our model to represent the different categories of human mobility. We are still considering de-trending out PM2.5 data, as we have data spanning back to 2017, it would be great to utilize this data to remove the longer time trend which would allow us to isolate the effects during the period of interest (February – October 2020).

1. do Nascimento, C. M., de Oliveira, S. A., Santana, O. A., & Carvalho, H. (n.d.). Changes in air pollution due to COVID-19 lockdowns in 2020: Limited effect on NO₂, PM_{2.5}, and PM₁₀ annual means compared to the new WHO Air Quality Guidelines. *Journal of Global Health*, 12, 05043. <https://doi.org/10.7189/jogh.12.05043>
2. Venkatesh, U., Gandhi P, A., Ara, T., Rahman, M. M., & Kishore, J. (2022). Lockdowns, Community Mobility Patterns, and COVID-19: A Retrospective Analysis of Data from 16 Countries. *Healthcare Informatics Research*, 28(2), 160–169. <https://doi.org/10.4258/hir.2022.28.2.160>
3. *Infographic: Strict or Lenient? COVID-19 Lockdowns Compared.* (2020, June 18). Statista Daily Data. <https://www.statista.com/chart/22048/university-of-oxford-coronavirus-containment-and-health-index-selected-countries>
4. *Sources of PM_{2.5}-Associated Health Risks in Europe and Corresponding Emission-Induced Changes During 2005–2015—Gu—2023—GeoHealth—Wiley Online Library.* (n.d.). Retrieved November 22, 2023, from <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022GH000767>
5. *About AirNow | AirNow.gov.* (n.d.). AirNow.gov, U.S. EPA. Retrieved November 22, 2023, from <https://www.airnow.gov/about-airnow>
6. *COVID-19 Community Mobility Report.* (n.d.). COVID-19 Community Mobility Report. Retrieved November 22, 2023, from <https://www.google.com/covid19/mobility?hl=en>
7. *Overview—Community Mobility Reports Help.* (n.d.). Retrieved November 22, 2023, from <https://support.google.com/covid19-mobility/answer/9824897?hl=en>
8. *World Bank Open Data.* (n.d.). World Bank Open Data. Retrieved November 22, 2023, from <https://data.worldbank.org>
9. Yasin, Y. J., Grivna, M., & Abu-Zidan, F. M. (2021). Global impact of COVID-19 pandemic on road traffic collisions. *World Journal of Emergency Surgery*, 16(1), 51. <https://doi.org/10.1186/s13017-021-00395-8>

Final Project EDA
Group 4

