AIR POLLUTION EXPOSURE ASSESSMENT: APPLICATIONS IN PERSONAL AIR MONITORING STUDIES

A Dissertation

by

INYANG UWAK

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PUBLIC HEALTH

Chair of Committee,	Natalie Johnson
Committee Members,	Jennifer Horney
	Xiaohui Xu
	Antonio Rene
Head of Department,	Mark Benden

December 2020

Major Subject: Epidemiology and Environmental Health

ABSTRACT

Exposure to air pollution is a global public health burden that continues to plague humans. This project evaluated the evidence and levels of exposure to air pollutants in different populations. We conducted a systematic review and meta-analysis using the Navigation Guide systematic review methodology to evaluate prenatal exposure to particulate matter (PM) air pollution and birth weight. Results from this study led us to explore various methods of exposure assessment in two personal air monitoring studies. The first was an evaluation of the maternal exposure to polycyclic aromatic hydrocarbons in pregnant women in their third trimester in Mc Allen, Texas, using three methods for the exposure assessment: silicone wristbands, XAD resin and filters. The next study focused on personal exposure to polycyclic aromatic hydrocarbons (PAHs) in a group of teachers in El Paso, Texas applying similar methods of assessment.

From the systematic review, our results confirmed previous evaluations showing an inverse association between prenatal PM exposure and low birth weight. Variation across individual study outcomes largely reflects methodological differences across studies. Our application of the Navigation Guide methodology highlights the utility of quality and strength of evidence ratings when evaluating environmental health studies. It also illustrated the effect the exposure assessment methods had on the quality and strength of evidence.

The McAllen study results showed that all the pregnant women enrolled, were exposed to PAHs at different magnitudes and compositions. Wristbands captured a

significant number of PAHs when compared with XAD and filters. Similarly, results from the El Paso study revealed that the teachers were exposed to various PAHs both from traffic and wood/coal combustion sources. These results can inform future regulation of traffic emissions around school environments. Finally, these studies provide confirmation that wristbands are an effective technology that can be easily utilized in personal exposure assessment studies in environmental health.

DEDICATION

This dissertation is dedicated to the memory of my father.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Natalie Johnson, who in the past 4 years, has provided me with excellent academic mentorship, encouraged me on developing research projects, and allowed me the flexibility which was important for me to balance work and parenting two children. I would never have accomplished this without her support and encouragement, for this and more, I am truly grateful.

I want to extend my gratitude to my committee members, Dr. Jennifer Horney, Dr. Xiaohui Xu, and Dr. Antonio Rene for their guidance and support throughout my academic pursuits at Texas A&M. I am very thankful to my head of department Dr. Benden for his exceptional leadership as well as his support of graduate students and to Fawne Toler for her kind assistance and tremendous support behind the scenes. Special thanks to Margaret Foster, Dr. Weihsueh Chiu, Dr. Juleen Lam, Dr. Xiaohui Xu, Dr. Brandie Taylor, Dr. Kirsten Koehler, Megan Moriarty, Samuel Taiwo, Natalie Olson and Angelica Fuentes for being a great research team, contributing and co-authoring the Navigation Guide systematic review and meta-analysis. My sincere appreciation goes to the Center for Advancing Research in Transportation Emissions, Energy, and Health (CARTEEH) at Texas A&M Transportation Institute for offering me an assistantship and for the opportunity to lead the project in El Paso. The lessons learned during that experience has shaped my research career. I also wish to thank the Dr. Liu and Dr. Shields from TAMU Geochemical and Environmental Research Group (GERG) for

granting me access to their lab and for their guidance throughout the lab analysis. Thanks also go to my writing accountability group members Drs. Melissa Noel, Edidiong Mendie, and Erica Tyler for their encouragement throughout the COVID19 lockdown and thereafter. Special thanks to the Johnson lab team for their love, care and support.

My utmost appreciation goes to my Creator, without whom I would not exist and to my husband and children for their love, patience and understanding throughout this academic journey.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Natalie Johnson, Dr. Jennifer Horney, Dr. Xiaohui Xu, Dr. Antonio Rene.

The Navigation Guide systematic review and meta-analysis was in collaboration with Margaret Foster, Dr. Weihsueh Chiu, Dr. Natalie Johnson, Dr. Juleen Lam, Dr. Xiaohui Xu, Dr. Brandie Taylor, Dr. Kirsten Koehler, Megan Moriarty, Samuel Taiwo, Natalie Olson and Angelica Fuentes.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by the U.S. Department of Transportation funded University Transportation Center (UTC), the Center for Advancing Research in Transportation Emissions, Energy, and Health (CARTEEH) at the Texas A&M Transportation Institute (TTI).

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

According to the World Health Organization, 4.2 million people die each year as a result of exposure to outdoor air pollution, and 91% of the world's population reside in places with poor air quality (WHO, 2020). In the United States, as of 2017, 107,500 deaths were attributable to exposure to outdoor air pollution (Health Effects Institute, 2019). Outdoor air pollution, specifically, is the 5th leading risk factor for deaths and disability globally and 12th in the US (Health Effects Institute, 2019). Though there have been reports of a temporary reduction in air pollution following the COVID-19 pandemic, however, as businesses and industries reopen, the amount of air pollutants is returning to pre-pandemic levels (M.A. Zambrano-Monserrate et al., 2020).

1.1. Sources of Exposure

Humans can be exposed to air pollution indoors or outdoors. Indoor air pollutants are those found inside our homes/offices and include fumes from cooking and cleaning agents, as well as off-gassing of chemicals from flooring and furniture. Outdoor or ambient air pollutants are those found in the air outdoors. Outdoor air pollution comes from two primary sources, natural and anthropogenic. Man-made sources, such as emissions from motor vehicles, industry, agricultural waste, heat, and power generation, contribute the most to outdoor air pollution (WHO, 2019). Natural sources of pollution include forest fires, volcanic eruptions, and windblown dust (e.g., Sahara Desert dust). Most of these air pollutants often contain particulate matter at varying concentrations.

1.2. Particulate Matter (PM)

Particulate matter air pollution is a complex mixture of particles. PM can be found in combustion particles, metals, organic compounds, roadside dust, and in combination with other toxic compounds (Park et al., 2018). Particulate matter air pollution can be broadly classified into three based on their aerodynamic equivalent diameter (AED). Particulate matter between 2.5 and 10 μm are called coarse PM or PM_{10}, those less than 2.5 μm are called fine PM or $PM_{2.5}$, while ultrafine particles or $PM_{0.1}$ are less than 0.1 μm in diameter (Anderson et al., 2012). These diameters determine how far down the respiratory tract the particles can reach. Ultrafine and fine PM pose a significant health risk when inhaled because they can make it through the bronchial tree into the alveoli and ultimately end up in the bloodstream while coarse PM can be removed by mucociliary clearance in the upper airway (Tsuda et al., 2013).

1.2.1. Health Effects associated with PM

Numerous studies have been published on the deleterious health effects of air pollution exposure, specifically PM. In the most recent joint American Thoracic and European Respiratory Societies' policy statement on what constitutes adverse health effects of air pollution, several conclusions were drawn, namely that exposure to air pollution is associated with respiratory and cardiovascular disease mortality (Thurston et al., 2017). Moreover, an increased incidence of preterm birth and low birth weight/growth restriction was associated with maternal air pollution exposure (Thurston et al., 2017). Furthermore, the US Environmental Protection Agency (USEPA) concluded in their 2019 Integrated Science Assessment for Particulate Matter that human exposure to fine particulate pollution causes early death from both short-term and long-

term exposure, cardiovascular morbidity, and adverse respiratory and nervous system effects (EPA 2019). Based on the overwhelming consensus that exposure to air pollution is a significant environmental risk factor, it is necessary to accurately measure human exposure to air pollutants to support human health risk assessments for regulatory decision-making, increase community awareness, and guide preventive intervention strategies, to protect the most vulnerable in the population, from the harmful effects of air pollution.

1.3. Polycyclic Aromatic Hydrocarbons (PAH)

PAHs are environmentally persistent organic compounds typically occurring in mixtures or combined with other compounds. PAHs are made up of carbon and hydrogen atoms bonded in two or more benzene rings in angular or linear arrangements (Abdel Shafy et al., 2015). Those with three or fewer rings are called low molecular weight (LMW) PAHs, and those with more than three ring structures are high molecular weight (HMW) PAHs (Kanaly et al., 2000). The LMW PAHs are water soluble and more volatile; therefore, exist in the gaseous phase (Zhang et al., 2008). The HMW PAHs, on the other hand, is less soluble in water (hydrophobic) with lower volatility and, as such, exist in the particle phase predominantly bound to particulates (Zhang et al., 2008). PAHs can occur naturally in the environment like volcanic eruptions and forest fires, as well as from man-made sources like vehicular exhaust and cigarette smoke. PAHs can also be produced from industrial processes. The sources of PAHs can be grouped into three, petrogenic, pyrogenic, and biological (Abdel Shafy et al., 2015). Petrogenic PAHs are those formed from moderate temperature processes and are associated with coal and petroleum products or fossil fuels (Abdel Shafy et al., 2015). Pyrogenic PAHs are formed from organic compounds subjected

to higher temperatures and low oxygen processes and are associated with wood and biomass combustion (Abdel Shafy et al., 2015). Humans are commonly exposed to a mixture of PAHs through inhalation of polluted air from vehicle exhaust, incinerator emissions, wood smoke, and cigarette smoke, as well as dermal exposure to creosote-treated wood (McCormick et al., 2011). Other sources include the ingestion of smoked, grilled, and barbecued foods.

1.3.1. Health Effects associated with PAHs

The US-Environmental Protection Agency (EPA) has identified 16 PAHs as those with the greatest potential for adverse health effects on humans. These 16 PAHs (see figure 1) include naphthalene (NAP), acenaphthylene (ACY), acenaphthene (ACE), fluorene (FLU), phenanthrene (PHEN), anthracene (ANTH), fluoranthene (FLTH), pyrene (PYR), benzo[a]anthracene (B[a]A), chrysene (CHRY), benzo[b]fluoranthene (B[b]F), benzo[k]fluoranthene (B[k]F), benzo[a]pyrene (B[a]P), benzo[g,h,i]perylene (B[ghi]P), indeno[1,2,3-c,d]pyrene (IND), and dibenz[a,h]anthracene (D[ah]A) (EPA). Several of them are irritants, while others are probable or known carcinogens (IARC 2006). The harmful health effects of exposure to PAHs depends on the route of exposure (inhalational, oral or dermal), duration of exposure, presence of pre-existing conditions, concentration, and toxicity of the PAH (Kim et al., 2013). Acute exposure to PAHs has been shown to cause irritation of the eyes, airway or skin surface, nausea/vomiting (Kim et al., 2013). Long term exposures have been studied in occupational settings and found to cause skin, lung, and bladder cancers as well as DNA damage (Kim et al., 2013, Armstrong et al., 2004). A review study on coke oven workers found lung cancer to be associated with Benzo[a]pyrene exposure (Armstrong et al., 2004, Armstrong et al., 2002). A 16-fold higher risk

of lung cancer was found in these workers (Armstrong et al., 2004, Armstrong et al., 2002). The following PAHs are classified as probable carcinogens by the US EPA benzo(a)anthracene, benzo(a)pyrene, benzo(b)fluoranthene, benzo(k)fluoranthene, chrysene, dibenz(ah)anthracene, and indenol(1,2,3-cd)pyrene (USEPA 2008). Studies show that the risk of cancer increases when humans exposed to a mixture of individual PAHs (Kim et al., 2013). Further, PAHs have been shown to cross the placenta and negatively impact fetal health. Zhang et al measured 15 PAHs in 95 paired maternal and umbilical cord serum in Shanghai confirming the transplacental transfer of PAHs (Zhang et al., 2017). Their findings revealed higher concentrations of PAHs in maternal serum when compared to the placental and fetal tissues also, low molecular weight PAHs were predominantly in placental tissues (Zhang et al., 2017). In a New York City prospective cohort, prenatal exposure to PAHs was linked to an increased risk of cognitive developmental delay and lower mental development index by the third year of life (Perera et al. 2006). Rundle et al in a study of prenatal exposure to PAHs and BMI trajectories identified an association with higher BMI in boys at age 5 but the trajectories converged at age 11 (Rundle et al 2019).
Overall, most of the epidemiological studies conducted have been in occupational settings and animal populations, it is imperative to replicate more these studies in susceptible human populations with evidence of exposure to PAHs to identify and examine more health effects.

1.4. Human Exposure Assessment

Human exposure to air pollutants can be measured in a variety of ways. However, there are two basic types of human exposure assessment of air pollutants, air monitoring and

biomonitoring. Air monitoring can be conducted by direct measurements and indirect measurements.

1.4.1. Ambient Monitoring

Indirect methods involve the use of pollutant concentrations collected from fixed locations. This typically takes the form of continuous ambient monitoring stations (CAMS) and may be accompanied by time-activity logs/questionnaires to provide more information on the location of the participants and time spent in those locations. Several studies have shown that ambient monitors play an essential role in revealing current air pollutant levels in various locations; however, when used alone for human exposure assessment, they are often prone to misclassification (Askariyeh et al., 2019).

1.4.2. Personal Monitoring

Direct measurements involve the use of personal monitors, which are typically lightweight devices that pull air from the participants breathing zone using a pump through a collection device, such as filter or diffusion tube worn, strapped on, or carried by an individual as they go about their daily activities. Personal monitors can be divided into passive monitors/samplers and active samplers. Passive monitors rely on diffusion or adsorption of the pollutant molecules into a collection media over a period, and samples are then analyzed in a laboratory to evaluate the concentration of the pollutants.

Active samplers use a pump to pull air continuously through a collector connected to a battery. Direct personal monitoring is a superior method for gaining spatially refined measures

for air pollution exposure assessment. However, limitations include the cost, burden to participants, and limited ability to apply in long-term exposure studies, which significantly hinders the temporal resolution.

1.4.3. Biomonitoring

Bio-monitoring or biological monitoring is the measurement of pollutant concentrations or their metabolites present in human tissues or biological fluids using biomarkers. For example, cotinine, a metabolite of nicotine, is a biomarker of exposure to environmental tobacco smoke and can be found in urine, saliva, blood, or hair samples (Benowitz et al., 2009). When conducting biomonitoring, biomarkers are the compounds used to identify the presence of the pollutant. Biomarkers complete the link between exposure and internal dose; however, they incorporate all exposure routes, including ingestion, inhalation, and dermal (National Research Council.,1997). Biomarkers are very effective tools in human exposure assessment because of their ability to measure the internal dose of a pollutant. Biomarkers that confirm exposure to a specific compound are called biomarkers of exposure. Biomarkers of effect reflect the effect of the exposure and biomarkers of susceptibility reveals the risk/likelihood of the effect. Monohydroxy polycyclic aromatic hydrocarbons (OH-PAHs) metabolites of PAHs found in urine can be used to measure acute exposures to PAHs due to their short half-life (Li et al., 2008). An example of one such OH-PAHs is 1-hydroxy pyrene (1-PYR), a common urinary biomarker of exposure to polycyclic aromatic hydrocarbons (Li et al., 2008). Several studies have utilized this biomarker of inhalational exposure to PAHs. Urinary 8-hydroxydeoxyguanosine (8-OHdG) is a biomarker of effect of PAH exposure and oxidative

stress (Huang et al., 2012). Urinary 8-OHdG concentration was found to increase following elevated particulate PAH exposure in a longitudinal study of traffic conductors (Huang et al., 2012).

One example on the relevance and usefulness of biomarkers was a study on the pollution exposures associated with the 2008 Beijing Olympics which showed an approximately 50% decrease in ambient pollution levels during the Olympics and a 30-60% decrease in several biomarkers of systemic oxidative stress and pulmonary inflammation, with increases when the pollution controls were lifted. Though this study was in healthy participants, the risk would probably be higher in a susceptible subpopulation, like infants or the elderly (Huang et al., 2012, Thurston et al., 2017).

1.5. Disparities in Exposure

Populations at risk of exposure to air pollutants include vulnerable populations and biologically susceptible populations. Those vulnerable, based on disparities in exposure to air pollutants, include individuals that live in communities near pollution sources like power plants, busy highways, and industry. Unfortunately, most of these areas are communities of color, as well as individuals with low socio-economic status. Many studies have shown that race/ethnicity and socio-economic status play a significant role in who gets exposed to air pollution. Studies have found mobile source and traffic-related exposures to be higher in minority and low-income populations (Tian et al., 2013, Houston et al., 2014). Another study conducted in the Sacramento area of California found $PM_{2.5}$ emissions from road traffic correlated with schools with a higher percentage of Black, Hispanic, and multi-ethnic students and students eligible for reduced-price

meals (Gaffron & Niemeier., 2015). Further, Tessum et al. found that in the US, Black and Hispanic minorities were disproportionately exposed to levels of $PM_{2.5}$ from the activities of the White majority (Tessum et al.,2019). Blacks and Hispanics, on average, were found to bear a "pollution burden" of 56% and 63% excess exposure, respectively, relative to the exposure caused by their consumption (Tessum et al.,2019).

Biologically susceptible populations include pregnant women, children, and persons with pre-existing cardiorespiratory diseases that might put them at a higher risk of harm from exposure to air pollution. The developing fetus is more susceptible to toxic environmental insults due to their immature physiology and their reduced ability to mount an immune response (Barr et al., 2007). Various studies have confirmed the association between air pollution exposure and adverse birth outcomes. One of such is a study of nonsmoking pregnant women in Krakow by Jedrychowski et al which revealed that prenatal exposure to $PM_{2.5}$ and PAHs were associated with decreased birth weight and birth length (Jedrychowski et al., 2017).

Another susceptible population are people living or working near border regions or ports of entry. Examples of some of such areas in US include Laredo, El Paso, Brownsville, Nogales, and San Ysidro (San Diego) (Quintana et al., 2014). These communities are often low income, Hispanic, and are disproportionately exposed to air pollution associated with these ports (Quintana et al., 2014). A study by Galaviz et al of a group of pedestrian border commuters on the San Ysidro port of entry at the US-Mexico border revealed that they had a 6-fold increase in 1-nitropyrene a biomarker of exposure to diesel exhaust, a 3-fold increase in carbon monoxide, and a 2-fold increase in gravimetric $PM_{2.5}$ exposures when compared to the control group (non-border commuters) (Galaviz et al., 2014). Given the above precedence, it is necessary to consider

approaches to reducing air pollution exposure disparities, some of which include strengthening exposure assessment research that drives regulation for residential areas and schools to be located far from potential sources of exposure. Other practical approaches include the addition of green space, vegetative barriers, and promoting active modes of transportation.

1.6. Specific Aims

The main goal of the research in this dissertation was to investigate the application of different exposure assessment techniques to quantify $PM_{2.5}$ and PAHs.

Aim 1. To assess the state of the literature on the association between particulate matter air pollution and birthweight by conducting a systematic review and meta-analysis using the Navigation Guide methodology and identify how the risk of bias may impact estimated effect size.

Aim 2. To evaluate maternal exposure to polycyclic aromatic hydrocarbons using personal air sampling and biomonitoring in Mc Allen.

Aim 3. To characterize personal exposure to PAHs from traffic sources in El Paso and to evaluate the effectiveness of silicone wristbands in capturing PAHs.

We evaluated two types of personal monitoring, active and passive, in two different populations, and explored the potential sources of these PAHs.

1.7. Significance

The broad and significant impact of air pollution exposure on health makes it critical to identify efficient ways to measure human exposure accurately. Refined measurement techniques support causal associations gleaned through epidemiological studies, which continue to link air pollution exposure with a variety of health effects. Exposure assessment also aids in the identification of thresholds for these health outcomes, which supports policies and regulatory levels to reduce the burden of disease caused by air pollution exposure. Currently, there has been increased rollbacks of some clean air regulation, reduction of funding to regulatory agencies, and reduced government funding for environmental research, making this more pertinent than ever (Environmental Integrity Project, 2019). Last, exposure assessment at the community-level allows for more awareness and individual empowerment to lobby for ways to reduce these exposures and ensure communities can breathe clean air.

1.8. References

Abdel-Shafy, H. I., & Mansour, M. S. M. (2016). A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum, 25*(1), 107-123. doi:https://doi.org/10.1016/j.ejpe.2015.03.011

Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the air: a review of the effects of particulate matter air pollution on human health. *Journal of Medical Toxicology: official journal of the American College of Medical Toxicology*, *8*(2), 166–175. https://doi.org/10.1007/s13181-011-0203-1

Armstrong, B., Hutchinson, E., Unwin, J., & Fletcher, T. (2004). Lung cancer risk after exposure to polycyclic aromatic hydrocarbons: a review and meta-analysis. *Environmental Health Perspectives*, *112*(9), 970–978. https://doi.org/10.1289/ehp.6895

Armstrong BG, Hutchinson E, Fletcher T. 2002. Cancer Risk following Exposure to Polycyclic Aromatic Hydrocarbons (PAHs): A Meta-Analysis. Rep No 068. Sudbury, UK: This Health and Safety Executive. Available: http://www.hse.gov.uk/research/rrhtm/rr068.htm [accessed 19 May 2020].

Askariyeh, M. H., Vallamsundar, S., Zietsman, J., & Ramani, T. (2019). Assessment of traffic-related air pollution: Case study of pregnant women in South Texas. *International Journal of Environmental Research and Public Health*, *16*(13). https://doi.org/10.3390/ijerph16132433

Barr, D. B., Bishop, A., & Needham, L. L. (2007). Concentrations of xenobiotic chemicals in the maternal-fetal unit. *Reproductive Toxicology*, *23*(3), 260-266.doi:https://doi.org/10.1016/j.reprotox.2007.03.003

Benowitz, N. L., Hukkanen, J., & Jacob, P., 3rd (2009). Nicotine chemistry, metabolism, kinetics and biomarkers. *Handbook of Experimental Pharmacology*, (192), 29–60. https://doi.org/10.1007/978-3-540-69248-5_2

Environmental Integrity Project. (2019). *The thin green line cuts in state pollution control agencies threaten public health*. Retrieved from https://environmentalintegrity.org/wp-content/uploads/2019/12/The-Thin-Green-Line-report-12.5.19.pdf

Gaffron, P., & Niemeier, D. (2015). School locations and traffic emissions—environmental (in)justice findings using a new screening method. *International Journal of Environmental Research and Public Health*, *12*(2), 2009–2025. https://doi.org/10.3390/ijerph120202009

Galaviz, V. E., Yost, M. G., Simpson, C. D., Camp, J. E., Paulsen, M. H., Elder, J. P., Hoffman, L., Flores, D., Quintana, P. J. E. (2014). Traffic pollutant exposures experienced by pedestrians waiting to enter the U.S. at a major U.S.–Mexico border crossing. *Atmospheric Environment, 88*, 362-369. doi:https://doi.org/10.1016/j.atmosenv.2013.12.042

Health Effects Institute. 2019. State of Global Air 2019. Special Report. Boston, MA:Health Effects Institute, https://www.stateofglobalair.org/sites/default/files/soga_2019_fact_sheet.pdf

Houston, D., Li, W., & Wu, J. (2014). Disparities in exposure to automobile and truck traffic and vehicle emissions near the Los Angeles-Long Beach port complex. *American Journal of Public Health*, *104*(1), 156–164. https://doi.org/10.2105/AJPH.2012.301120

Huang, W., Wang, G., Lu, S. E., Kipen, H., Wang, Y., Hu, M., Lin, W., Rich, D., Ohman-Strickland, P., Diehl, S. R., Zhu, P., Tong, J., Gong, J., Zhu, T., & Zhang, J. (2012). Inflammatory and oxidative stress responses of healthy young adults to changes in air quality during the Beijing Olympics. *American Journal of Respiratory and Critical Care Medicine*, *186*(11), 1150–1159. https://doi.org/10.1164/rccm.201205-0850OC

International Agency for Research on Cancer [Accessed on Oct 10, 2006]; List of All Agents Evaluated to Date. 2006 < http://monographs.iarc.fr/ENG/Classification/index.php>.

Jedrychowski, W. A., Majewska, R., Spengler, J. D., Camann, D., Roen, E. L., & Perera, F. P. (2017). Prenatal exposure to fine particles and polycyclic aromatic hydrocarbons and birth

outcomes: a two-pollutant approach. *International Archives of Occupational and Environmental Health*, *90*(3), 255–264. https://doi.org/10.1007/s00420-016-1192-9

Kanaly, R. A., & Harayama, S. (2000). Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *Journal of Bacteriology*, *182*(8), 2059–2067. https://doi.org/10.1128/jb.182.8.2059-2067.2000

Kim, K., Jahan, S. A., Kabir, E., & Brown, R. J. C. (2013). A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environment International, 60,* 71-80. doi:https://doi.org/10.1016/j.envint.2013.07.019

Li, Z., Romanoff, L. C., Lewin, M. D., Porter, E. N., Trinidad, D. A., Needham, L. L., Patterson, D. G., Jr, & Sjödin, A. (2010). Variability of urinary concentrations of polycyclic aromatic hydrocarbon metabolite in general population and comparison of spot, first-morning, and 24-h void sampling. *Journal of Exposure Science & Environmental Epidemiology*, *20*(6), 526–535. https://doi.org/10.1038/jes.2009.41

National Research Council (US) Committee on Environmental Epidemiology, & National Research Council (US) Commission on Life Sciences. (1997). Environmental Epidemiology: Volume 2: Use of the Gray Literature and Other Data in Environmental Epidemiology. National Academies Press, Washington DC, US: 3, Exposure Assessment in Environmental Epidemiology. Available from: https://www.ncbi.nlm.nih.gov/books/NBK233635/ [accessed 9 April 2020].

Park, M., Joo, H. S., Lee, K., Jang, M., Kim, S. D., Kim, I., Borlaza, L.J, Lim, H., Shin, H., Chung, K.H., Choi, Y., Park, S.G., Bae, M., Lee, J., Song, H., Park, K. (2018). Differential

toxicities of fine particulate matters from various sources. *Scientific Reports, 8*(1), 17007-11. doi:10.1038/s41598-018-35398-0

Perera, F. P., Rauh, V., Whyatt, R. M., Tsai, W. Y., Tang, D., Diaz, D., Hoepner, L., Barr, D., Tu, Y. H., Camann, D., & Kinney, P. (2006). Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. *Environmental Health Perspectives*, *114*(8), 1287–1292. https://doi.org/10.1289/ehp.9084

Quintana, P., Khalighi, M., Castillo Quiñones, J. E., Patel, Z., Guerrero Garcia, J., Martinez Vergara, P., Bryden, M., & Mantz, A. (2018). Traffic pollutants measured inside vehicles waiting in line at a major US-Mexico Port of Entry. *The Science of the Total Environment*, *622-623*, 236–243. https://doi.org/10.1016/j.scitotenv.2017.11.319

Rogers, S. W., Ong, S. K., Kjartanson, B. H., Golchin, J., & Stenback, G. A. 2002. Natural Attenuation of Polycyclic Aromatic Hydrocarbon-Contaminated Sites: Review. *Practice Periodical of Hazardous, Toxic & Radioactive Waste Management*, 6(3), 141.

Rundle, A. G., Gallagher, D., Herbstman, J. B., Goldsmith, J., Holmes, D., Hassoun, A., Oberfield, S., Miller, R. L., Andrews, H., Widen, E. M., Hoepner, L. A., & Perera, F. (2019). Prenatal exposure to airborne polycyclic aromatic hydrocarbons and childhood growth trajectories from age 5-14 years. *Environmental Research*, *177*, 108595. https://doi.org/10.1016/j.envres.2019.108595

Sexton, K., Salinas, J. J., McDonald, T. J., Gowen, R. M., Miller, R. P., McCormick, J. B., & Fisher-Hoch, S. P. (2011). Polycyclic aromatic hydrocarbons in maternal and umbilical cord blood from pregnant Hispanic women living in Brownsville, Texas. *International Journal of*

Environmental Research and Public Health, *8*(8), 3365–3379. https://doi.org/10.3390/ijerph8083365

Tessum C.W., Apte J.S., Goodkind A.L., Muller N.Z., Mullins K.A., Paolella D.A., Polasky S., Springer N.P., Thakrar S.K., Marshall J.D., Hill J.D. (2019) *Proceedings of the National Academy of Sciences of the United States of America*, 116 (13) , pp. 6001-6006.

Thurston, G. D., Kipen, H., Annesi-Maesano, I., Balmes, J., Brook, R. D., Cromar, K., De Matteis, S., Forastiere, F., Forsberg, B., Frampton, M. W., Grigg, J., Heederik, D., Kelly, F. J., Kuenzli, N., Laumbach, R., Peters, A., Rajagopalan, S. T., Rich, D., Ritz, B., Samet, J. M., Sandstrom, T., Sigsgaard, T., Sunyer, J., Brunekreef, B. (2017). A joint ERS/ATS policy statement: what constitutes an adverse health effect of air pollution? An analytical framework. *The European Respiratory Journal*, *49*(1), 1600419. https://doi.org/10.1183/13993003.00419-2016

Tian, N., Xue, J., & Barzyk, T.M. (2013). Evaluating socioeconomic and racial differences in traffic-related metrics in the United States using a GIS approach. *Journal of Exposure Science and Environmental Epidemiology, 23*, 215-222. https://doi.org/10.1038/jes.2012.83

Tsuda, A., Henry, F. S., & Butler, J. P. (2013). Particle transport and deposition: basic physics of particle kinetics. *Comprehensive Physiology*, *3*(4), 1437–1471. https://doi.org/10.1002/cphy.c100085

World Health Organization, 2020. Air Pollution. https://www.who.int/health-topics/air-pollution#tab=tab_1

Zambrano-Monserrate, M. A., Ruano, M. A., & Sanchez-Alcalde, L. (2020). Indirect effects of COVID-19 on the environment. *The Science of the Total Environment, 728*, 138813. https://doi.org/10.1016/j.scitotenv.2020.138813

Zhang, X., Li, X., Jing, Y., Fang, X., Zhang, X., Lei, B., & Yu, Y. (2017). Transplacental transfer of polycyclic aromatic hydrocarbons in paired samples of maternal serum, umbilical cord serum, and placenta in shanghai, china. *Environmental Pollution, 222*, 267-275. doi:https://doi.org/10.1016/j.envpol.2016.12.046

2. A SYSTEMATIC REVIEW AND META-ANALYSIS TO EVALUATE PRENATAL EXPOSURE TO PARTICULATE MATTER AIR POLLUTION AND BIRTHWEIGHT

2.1. Introduction

Human exposure to PM ultimately begins in-utero, and these exposures can impact future health outcomes. Studies have shown that developmental exposures to PM can cause placental vascular function disruption and endothelial dysfunction as a result of oxidative stress (Glinianaia et al. 2004, Backes et al. 2012). These can impact the fetus in various ways leading to developmental abnormalities like intra-uterine growth restriction, low birth weight, and preterm birth (Rossner et al. 2011, Rudra et al. 2011).

Low birth weight (LBW) is defined as infants born weighing less than 2,500 grams (Cutland et al. 2017). A primary cause of LBW is preterm birth (PTB), which are infants born before 37 weeks of gestation (Cutland et al. 2017). Small for gestational age (SGA) is also used to describe infant weights below the 10th percentile in comparison to infants of the same gestational age (Cutland et al. 2017). A subset of LBW and SGA infants require intensive neonatal care for immediate health issues and may have chronic health outcomes later in life (Belbasis et al. 2016).

Many studies have investigated the association between prenatal PM exposure and low birth weight (LBW), which is mainly attributable to the large-scale availability of birth weight data through birth records. These studies showed that pooled estimates of effect sizes for LBW for a 10ug/m3 increase in PM exposure during pregnancy varied widely and were inconsistent with the different stages of pregnancy (Glinianaia et al. 2004, Sun et al. 2016, Lamichhane et al. 2015, Steib et al. 2012). Even with the substantial evidence on the association between

developmental PM exposure and adverse birth outcomes, including LBW, PTB, and SGA births, there have been inconsistencies in the conclusions about the association and magnitude of the effect. Due to this heterogeneity across studies and potential risk of bias in individual studies, we evaluated the human evidence on prenatal exposure to particulate matter and birth weight.

In this review, we assessed each study for risk of bias and conducted a meta-analysis on a subset of studies to determine the magnitude of effect. We focused on birth weight as a continuous outcome variable to determine the impact of bias on effect size estimates. Many studies treat birth weight as a categorical variable defined as "normal" or "low birth weight" (<2,500 g at birth), thus limiting the power to determine pooled effect estimates across studies due to differences in baseline values across populations. Treating birthweight as a continuous variable allows for the assessment of the effect on population distributions and increases the ability to generalize across populations. Case studies have illustrated the importance of considering a continuous scale to provide added information about the exposure-disease continuum, inform population variability, and increase the predictive power of risk assessment (Woodruff et al. 2008).

2.2. Methods

2.2.1. The Navigation Guide Systematic Review Methodology

This review was conducted applying the navigation guide systematic review methodology. This methodology was developed in 2011 as part of an interdisciplinary collaboration between clinicians, academicians, and practitioners to harmonize the approaches for assessing evidence in the clinical sciences with environmental health (Woodruff et al., 2011).

It is based on the Cochrane Collaboration and the Grading of Recommendations Assessment Development and Evaluation - GRADE methods but accounts for the differences in evidence and decision context in environmental health risk assessments given their reliance on animal and human observational studies versus randomized controlled trials (Cumpston et al. 2019, Guyatt et al. 2008, Woodruff and Sutton 2014). To date, the Navigation Guide methodology has been applied in numerous reviews of environmental exposures, including the human evidence for effects of airborne pollutants on the diagnosis of autism spectrum disorder (Lam et al. 2016) and both the human and non-human evidence for effects of perfluorooctanoic acids (PFOAs) on fetal growth (Johnson et al. 2014; Koustas et al. 2014; Lam et al. 2014). The results of these studies and others demonstrate the utility of this approach in applying rigor and transparency in support of evidence-based decisions to environmental health problems. As is standard for systematic reviews and the Navigation Guide methodology, we developed a protocol prior to initiating the study and registered it in PROSPERO (http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017058805). See Figure 2.1 for a summary of the Navigation Guide process (adapted from Johnson et al. 2014).

2.2.1.1. Specify the Study Question

The aim of this study was to answer the question: "Does developmental exposure to ambient particulate air pollution affect birth weight?" A "Participants," or "Population," "Exposure," "Comparator," "Outcomes" (PECO) statement was developed to answer the study question (Higgins & Green, 2011).

Population: Humans

Exposure: Developmental exposure to ambient particulate air pollution

Comparator: Humans exposed to lower levels of air pollution than the more highly exposed humans.

Outcomes: Birth weight measured as a continuous variable.

2.2.1.2. Select the Evidence

Search Methods: A search was performed using various online databases including, Ovid Medline, Embase, and Global Health on November 23, 2015, using the following search terms; particulate matter, birth weight and low birth weight (see Appendix B, Table B1). The search was limited to studies written in English but not limited by publication date. We updated the search on February 27, 2020, to identify any new studies. Medical Subject Headings (MESH) database was used to compile synonyms for ambient particulate air pollution and birth weight. For the exposure, we separated the search terms into three categories using the terms, particulate matter, pm 2* or pm 10 and pollut* then combined in a Boolean search using the "OR" statement. For the outcome, we combined birth weight and low birth weight and its synonyms in a Boolean search using the "OR" statement. Next, the exposure and outcome terms were combined using the "AND" statement. We used the "ti,ab." function to limit the search to titles and abstracts.

Study Selection Criteria: Studies were included if they were original, evaluated ambient particulate air pollution, and reported associations with birth weight. Studies were excluded if; 1) the article did not report birth weight; 2) the paper did not report ambient particulate air pollution

exposure; 3) the paper contained no original data; 4) the article did not involve human subjects, or 5) other reason (explanation required).

Data Collection: The data relating to study characteristics and outcome measures from all included studies were extracted into the Health Assessment Workspace Collaborative (HAWC) database. The characteristics extracted included study population, location, sample size, exposure period/duration, pollutant class, methods used to estimate exposures, and all relevant estimates of association relating particulate air pollution exposure with birth weight.

2.2.1.3. Reviewing the Evidence

Risk of Bias: The risk of bias was assessed for each included individual study based on the Navigation Guide systematic review methodology (Johnson et al., 2014). The domains addressed recruitment strategy, blinding, confounding, exposure assessment, incomplete outcome data, selective outcome reporting, and conflict of interest (Table 2.1). Each domain was rated as either "low," "probably low," "probably high," or "high" risk of bias, based on a specific criterion (Appendix B, Table B3).

Statistical Analyses: We performed a meta-analysis to summarize the effect of exposure to ambient particulate air pollution on birth weight and assess the impact of study design as well as other characteristics on findings. The data was categorized into three pollutant areas; $PM_{2.5}$, PM_{10} and $PM_{2.5-10}$ or coarse PM for four exposure windows. The exposure windows considered were, first trimester, second trimester, third trimester and the entire pregnancy. For $PM_{2.5}$ and PM_{10}, due to there being enough papers, only studies with "low" or "probably low" risk of bias was analyzed. We calculated pooled estimates of risk estimates (and their 95% confidence

intervals) associated with a common exposure unit (grams change in birth weight per 10 u/m^3 increase in pollutant exposure) from the included studies using random effect models accounting for within- and between-study variability using the Knapp-Hartung modification (Knapp & Hartung, 2003). This approach accounts for uncertainty in the estimate of τ^2 in the standard error estimates, generally resulting in wider confidence intervals. The pooled effect estimates, and the study-specific estimates were presented in a forest plot. Heterogeneity was assessed using the I^2 statistic. Sources of heterogeneity explored using subgrouping included the following: ethnicity (non-Hispanic White only, Hispanic only, Black only), risk of bias, geographic locale (Americas, Europe, Asia), and spatial scale of exposure assessment. A sensitivity analysis was also conducted by excluding individual studies one at a time.

Quality of Evidence across studies: The quality of the overall body of evidence was rated as either "high," "moderate," or "low." All studies were given an initial rating of "moderate" quality according to the previously described Navigation Guide approach, and then adjustments were considered ("downgrades" or "upgrades") based on the characteristics of the studies to reach a conclusion of the final quality rating (Johnson et al., 2014). We considered "downgrades" to the quality rating based on five categories: risk of bias, indirectness, inconsistency, imprecision, and potential for publication bias (Table 2.2 adapted from Johnson et al. 2014). We considered "upgrades" to the quality rating based on the following; large magnitude of effect, dose-response, and if residual confounding would minimize the overall effect estimate (Balshem et al. 2011). Possible downgrades or upgrades were: 0 (no change from initial quality rating), -1 (1 level downgrade) or – 2 (2 level downgrade), +1 (1 level upgrade) or +2 (2 level upgrade) then compared ratings to reach a final decision (see Appendix B, Table B4).

Strength of Evidence across studies: We rated the overall strength of evidence based on four consideration: (1) Quality of body of evidence (the rating from the previous step); (2) Direction of effect; (3) Confidence in effect (likelihood that a new study could change our conclusion); and (4) Other compelling attributes of the data that may influence certainty. Possible ratings were "sufficient evidence," "limited evidence," "inadequate evidence," or "evidence of lack of toxicity" (Appendix B, Table B4).

2.3. Results

2.3.1. Included Studies

A total of 733 records were retrieved from the database search. 103 met our inclusion criteria from title and abstract screen. After full text screen (n=32), and an updated search (n=21) in February 2020, a total of 53 studies were included in the final analysis (see Figure 2.2).

A summary of the characteristics of these studies is detailed in Table 2.3. A total of 20 studies measured $PM_{2.5}$ exposure, 17 studies measured PM_{10} exposure and 1 study measured coarse PM ($PM_{2.5-10}$). Many studies measured a combination of pollutants. 3 studies measured all three pollutants, and 12 studies measured a combination of two pollutants. Exposure assessment methods included ambient monitoring as the primary technique (30 studies), followed by modeling (20 studies), a combination of monitoring and modeling (2 studies) and in one case personal modeling for a 48h duration in the second trimester of pregnancy. Generally, studies reported effect estimates for trimester-specific and entire pregnancy exposure windows (28 studies). However, 15 studies reported estimates for only the entire pregnancy and 10 studies

reported just by trimester and not entire pregnancy. Study locations ranged globally, and geographic location was taken into consideration in the meta-analysis.

2.3.2. Risk of Bias Assessment for Individual Studies

The risk of bias designations generally was rated as "low" or "probably low" for most domains. Individual study determinations are summarized in Figure 2.3 and Table 2.3. In a few cases, recruitment across study groups was determined to be "high." For instance, Pedersen et al. 2013 investigated low birth weight in a large European cohort study, wherein study participants were recruited from different populations in varying proportions. Confounding was predominantly rated as "probably low" (58% of studies). In some cases, studies were rated as "high" or "probably high" risk in addressing confounding. In these cases, investigators only accounted for two or less of the pre-determined important potential confounders, which could have introduced bias into analyses. In a few cases reviewers determined a "probably high" risk of bias in the "other" category, defined as if the study appeared to be free of other problems that could put it at a risk of bias. For instance, regarding Mannes et al. 2005, reviewers determined the risk of residual confounding and over adjustment bias in the linear regression model, as authors adjusted for an intermediate on the pathway between exposure and outcome. In addition, authors also did not account for extreme values in birthweight of gestational age. In general, the domain with a considerable number of studies rated as "probably high" (43%) was related to the robustness of exposure assessment. This was mainly due to reliance on county-level monitoring data without adequate temporal coverage or spatial resolution. Overall, for $PM_{2.5,}$ 12 studies (out of a total of 30 studies measuring $PM_{2.5}$) were rated overall as "low" or "probably low" risk of

bias. For PM_{10}, 10 studies (out of a total of 29 studies measuring PM_{10}) were rated overall as "low" or "probably low" risk of bias and used for subsequent meta-analysis. For studies on coarse PM, none of the 5 studies were given an overall rating of "low" or "probably low." This was largely the result of risk of exposure misclassification based on county-level measurements employed in most of these studies (Darrow et al. 2011; Ebisu et al. 2016; Morello-Frosch et al. 2010; Parker and Woodruff 2008).

2.3.3. Meta-analysis

We conducted a meta-analysis on studies rated as "low" or "probably low" risk of bias for exposures to $PM_{2.5}$ and PM_{10}. This included 18 studies for $PM_{2.5}$ and 10 studies for PM_{10}. For $PM_{2.5\text{-}10}$, there were a limited number of studies overall that measured this pollutant class, and none were rated as "low" or "probably low." Thus, we used the existing five studies rated as "high" or "probably high" in our meta-analysis. A summary of the meta-analysis results using a random effects model is shown in Table 2.4, separated by pollutant class and exposure window (trimester or entire pregnancy). For $PM_{2.5}$, the overall random effects estimates ranged from 5.69g to 27.55g decrease in birth weight per 10 $\mu g/m^3$ increase in $PM_{2.5}$ (Appendix B, Figures B1.1-1.4). The meta-estimate for the 1^{st} trimester was not statistically significant, but those for the other exposure windows were. Substantial heterogeneity was evident in each exposure window (I^2 ranged from 68% to 94%). In each exposure window, at least one study reported a positive association (increase in birth weight). Subgrouping based on ethnicity, spatial scale, or geographic location did not explain the observed heterogeneity (Appendix B, Figures B6, B7, and B8). Including "high" and "probably high" risk of bias studies further increased

heterogeneity (Appendix B, Figure B4.1-4.2). Influence analysis showed that for the second trimester, heterogeneity is explained by a single study (Hyder et al. 2014) with a large effect size (Appendix B, Figure B9.2). Omitting this study reduced I^2 from 68% to 40% and reduced the meta-estimate from –5.69 (–10.58, –0.79) to –3.81 (–7.88, 0.25). For other exposure windows, heterogeneity could not be attributed to any single study (Appendix B, Figure B9.3-9.4).

For PM_{10}, the overall random effects estimates ranged from a 3.22g increase to an 8.65g decrease in birth weight per 10 $\mu g/m^3$ increase in PM_{10} (Appendix B, Figure B2). The meta-estimates for the 1st and 2nd trimesters were not statistically significant (effect estimate 3.22, 95% CI: -3.13, 9.58 and -3.37, 95% CI: -8.22, 1.48, respectively), but those for the other exposure windows were. Low heterogeneity was seen in the trimester-based exposure windows (I^2 ranging 0-14%). However, substantial heterogeneity was evident for the entire pregnancy (I^2 =84%). Subgrouping based on ethnicity, spatial scale, or geographic location did not explain the observed heterogeneity (Appendix B, Figures B6.2, B7.5-7.8, B8.5-8.8). Including "high" and "probably high" risk of bias studies increased heterogeneity in all cases (Appendix B, Figure B5). Influence analysis showed that for the entire pregnancy, heterogeneity was explained largely by a single study (Geer et al. 2012) that reported a positive association (increase in birthweight), whereas all the other studies consistently showed an inverse association (Appendix B, Figure B9.8). Omitting this study reduced the I^2 from 84% to 0% and changed the meta-estimate from –8.65 (–16.83, –0.48) to –11.22 (–13.17, –9.26). For the other exposure windows, similar results in terms of both heterogeneity and meta-estimates were obtained in the influence analyses (Appendix B, Figures B9.5-9.7).

A smaller number of studies examined "coarse" PM, defined as PM with aerodynamic diameter between 2.5 and 10 microns. None of these studies were rated as having "low" or "probably low" risk of bias, as discussed previously. Thus, when including all studies, overall random effects estimates ranged from a 2.70g to 8.81g decrease in birth weight per 10 $\mu g/m^3$ increase in $PM_{2.5}$-PM_{10} (Appendix B, Figure B3). The meta-estimates for the 2nd and 3rd trimesters were not statistically significant, -2.90 (-10.04, 4.23) and -4.93 (-10.82, 0.96) respectively, and each of these had high heterogeneity (I^2 ranging 70-76%). Subgrouping based on ethnicity, spatial scale, or geographic location did not explain this observed heterogeneity. Heterogeneity was reduced to 55% for the 2nd trimester when omitting the most influential study (Parker and Woodruff 2008), though this left only two studies remaining with a pooled estimate that remained statistically not significant (Appendix B, Figure B9.10). Similarly, in the 3rd trimester, omitting the most influential study (Ebisu et al. 2016) reduced heterogeneity to 64%, but the pooled estimate remained statistically not significant (Appendix B, Figure B9.11). For the 1st trimester and the entire pregnancy, the meta-estimates were statistically significant, -2.70 (-3.90, -1.49) and -8.81 (-10.32, -7.31) respectively, with no observed heterogeneity in both cases (I^2 0%). For the 1st trimester, omitting any of the studies lead to meta-estimates that were statistically not significant (Appendix B, Figure B9.9). For the entire pregnancy, meta-estimates remained statistically significant under influence analyses, with no heterogeneity (Appendix B, Figure B9.12).

2.3.4. Quality of the Body of Evidence

The initial rating for the quality of evidence was "moderate" based on the Navigation Guide methodology (Johnson et al. 2014). Using the factors for rating the quality of evidence (Table 2.2), we determined the following evaluations (Table 2.4). For $PM_{2.5}$ exposure in the first trimester, a downgrade of 2 levels was supported, based on "imprecision" due to the lack of a statistically significant meta-estimate, as well as a wide confidence interval indicating potential impact of random error. Moreover, a downgrade for "inconsistency" was due to the substantial heterogeneity that could not be explained. The resulting quality of evidence rating was "very low." For $PM_{2.5}$ exposure in the second trimester, a downgrade of 1 level was supported based on "imprecision". Heterogeneity was explained by a single study and omitting this study lead to an effect estimate no longer statistically significant. The resulting quality of evidence rating was "low." For $PM_{2.5}$ exposure in the third trimester, as well as exposure throughout entire pregnancy, a downgrade of 1 level was supported, based on "inconsistency" due to the substantial heterogeneity that could not be explained. Thus, the resulting quality of evidence rating was "low."

For PM_{10} exposure during the first trimester, a downgrade of 2 levels was supported based on "imprecision" due to a wide confidence interval and the lack of a statistically significant meta-estimate with low heterogeneity. In addition, a downgrade for "inconsistency" was determined based on one-point estimate that indicated an inverse relationship (increase in birth weight). The resulting quality of evidence rating was "very low." For PM_{10} exposure during the second trimester, a downgrade of 1 level was supported based on "imprecision" due to the lack of a statistically significant meta-estimate with low heterogeneity. The resulting quality of

evidence rating was "low." For PM_{10} exposure during the third trimester, no change in the quality of evidence was indicated, as the meta-estimate was statistically significant with low heterogeneity (Figure 2.4). The resulting quality of evidence rating was "moderate." Last, for PM_{10} exposure during the entire pregnancy, no change in the quality of evidence was indicated. Heterogeneity was explained by a single study, and omitting that study lead to a precise, statistically significant meta-estimate. The resulting quality of evidence rating was "moderate." Meta-analysis results with "moderate" quality of evidence ratings are displayed in Figure 2.5.

For exposure to coarse PM ($PM_{2.5-10}$) during the first trimester, a downgrade of 2 levels was supported based on "risk of bias" (all studies were rated "high" or "probably high"), "imprecision" due to few studies (n=3), and a high degree of influence of any one study had on statistical significance. The resulting quality of evidence rating was "very low." For $PM_{2.5-10}$ exposure during the second and third trimesters, downgrades of 3 levels are supported based on "risk of bias" (all studies were rated "high" or "probably high"), "imprecision" due to the lack of a statistically significant meta-estimate, and "inconsistency" due to high, unexplained heterogeneity. The resulting quality of evidence rating was "very low." Last, for $PM_{2.5-10}$ exposure throughout the entire pregnancy, a downgrade of 1 level was supported based on "risk of bias" (all studies were rated "high" or "probably high"). The resulting quality of evidence rating was "low."

2.3.5. Strength of the Body of Evidence

Based on the strength of evidence rating illustrated in Table 2.4, the following evaluations were made. For $PM_{2.5}$, there is "inadequate evidence" for all exposure windows due

to “low” or “very low” quality of evidence, based on either imprecision of the estimate or high and unexplained heterogeneity (none of the other considerations were influential in this evaluation). For PM_{10}, there is “limited evidence” that increasing exposure during the third trimester or during the entire pregnancy will lead to a reduction in birth weight. The quality of evidence for these exposure windows was rated as “moderate.” Although the direction of the effect estimate was in the “adverse” direction, confidence in the effect estimate is limited because chance, bias, and confounding cannot be ruled out with reasonable confidence, and additional data could alter this conclusion. No other compelling attributes of the data exist that would influence this evaluation. For other exposure windows, evidence for PM_{10} is “inadequate” due to “low” or “very low” quality of evidence, based on either imprecision of the estimate and/or the presence of a relationship in the opposite (non-adverse) direction (none of the other considerations were influential in this evaluation). For $PM_{2.5-10}$, there is “inadequate evidence” that increasing exposure is during any exposure window leads to a reduction in birth weight. The available evidence is insufficient to assess the effects of exposure, mainly due to the high risk of bias in individual studies and the reliance on a small set of often heterogeneous studies. None of the other considerations from Table 2.2 were influential to this evaluation.

2.4. Discussion

Numerous case-control and cohort studies previously demonstrated an association between developmental exposures to ambient air pollution and reduced fetal growth or infant birthweight. A previous systematic review found an association between $PM_{2.5}$ exposure and LBW and SGA births, as well as PM_{10} exposure and SGA (Shah et al. 2011). A recent systematic

review including cohort and cross-sectional studies in U.S. populations demonstrated a significant association of air pollutants and heat exposure with adverse birth outcomes, such as preterm birth and low birth weight (Bekkar et al. 2020). Despite these observed associations, there have been inconsistencies in the conclusions about the association and magnitude of the effect. Initial systematic reviews based on a relatively small number of studies (n=3), were not able to draw conclusions on effect size (Bonzini et al. 2010; Bosetti et al. 2010; Ghosh et al. 2007). More recent systematic reviews, which performed a meta-analysis on a larger number of studies (>30) showed that pooled estimates of effect size for LBW for a 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure during entire pregnancy ranged from −15.9g (−26.8, −5.0) (Sun et al. 2016) to −22.17g (−37.93, −6.41) (Lamichhane et al. 2015). Steib et al. (2012) also reported estimates per 10 $\mu g/m^3$ increase in $PM_{2.5}$ exposure to be −23.4g (−45.5, −1.4), all of which are consistent with our pooled estimate of −27.55g (−48.45, −6.65) per 10 $\mu g/m^3$. This agreement is likely due to several of the same studies used across these meta-analyses. For PM_{10}, Lamichhane et al. (2015) reported estimates for a 10 $\mu g/m^3$ increase at −10.31g (−13.57 to −3.13 g), whereas Stieb et al. (2012) published estimates for a 20 $\mu g/m^3$ increase at −16.8g (−20.2 to −13.3), both of which are also consistent with our pooled estimate of −8.65g (−16.83, −0.48) per 10 $\mu g/m^3$. These previous investigators cited that they were not able to rule out the consequences of specific biases that may be as a result of differences in study methodology, study design, population demographics, exposure period, characterization of confounding and data collection.

In our analysis, there was substantial heterogeneity across the different pollutant classes. Also, the spatial scale employed, large scale (at the city or county level or >/= 10km) in comparison to medium scale (census tract, zip code, postal code, nearest monitor, <10km and

>/=5km) or small scale (<5km) led to greater heterogeneity. These findings underscore the complexity of estimating exposure across gestation. While one study (Jedrychowski et al. 2009) employed personal monitoring during pregnancy, the cost of adequate temporal coverage is great since it is infeasible for participants to carry monitors over time. Despite the significant heterogeneity, we still observed a decrease in birthweight for every 10 $\mu g/m^3$ increase in $PM_{2.5}$ across all trimesters (except the 1^{st}) and entire pregnancy, as well as for every 10 $\mu g/m^3$ increase in PM_{10} across the third trimester and entire pregnancy. The "inadequate" evidence rating for $PM_{2.5}$ reflects the quality, which received downgrades for inconsistency, driven mainly by heterogeneity. Similar conclusions were drawn by Lam et al. (2016) for the association between early-life exposure to air pollution as a whole and diagnosis of autism spectrum disorder.

Some limitations that may be associated with our study include the reliance on expert evaluation in the process used for the risk of bias, quality and strength ratings. However, this limitation was overcome by creating a diverse team of experts from relevant fields to participate in this process. Moreover, by publishing a pre-specified protocol and employing two independent reviewers for each study, our analysis includes a degree of transparency and robustness that is absent when using less structured approaches. Additionally, the rating of the quality of evidence across studies was dependent on the available data, which may be limited, and secondary/co-exposures like ultrafine particulate matter or heat were not addressed. There is also the potential for additional unmeasured confounding.

A major strength of our study is the transparency and thoroughness of the Navigation Guide systematic review process, which incorporates the GRADE system for assessing the

quality of synthesized human evidence in environmental health research in the absence of randomized clinical trials (Woodruff et al. 2014).

2.5. Conclusion

Overall, we conclude that the existing evidence supports an association between developmental exposure to ambient particulate matter air pollution and a decrease in birth weight. However, our findings reveal the need to standardize and improve exposure assessment methods in air pollution research because the various forms of exposure measurement utilized in the studies contributed to the heterogeneity seen in the meta-analysis. Furthermore, some of the unexplained heterogeneity found in our study may be resolved with additional studies which will strengthen the evidence.

2.6. References

Backes, C. H., Nelin, T., Gorr, M. W., & Wold, L. E. (2013). Early life exposure to air pollution: how bad is it?. *Toxicology Letters*, *216*(1), 47–53. https://doi.org/10.1016/j.toxlet.2012.11.007

Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. 2011. Grade guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 64:401-406.

Basu R, Harris M, Sie L, Malig B, Broadwin R, Green R. 2014. Effects of fine particulate matter and its constituents on low birth weight among full-term infants in California. *Environmental Research* 128:42-51.

Bekkar B, Pacheco S, Basu R, DeNicola N. 2020. Association of air pollution and heat exposure with preterm birth, low birth weight, and stillbirth in the us: A systematic review. *JAMA* Netw Open 3:e208243.

Beland L-P, Oloomi S. 2019. Environmental disaster, pollution, and infant health: Evidence from the deepwater horizon oil spill. *Journal of Environmental Economics and Management* 98:102265.

Belbasis L, Savvidou MD, Kanu C, Evangelou E, Tzoulaki I. 2016. Birth weight in relation to health and disease in later life: An umbrella review of systematic reviews and meta-analyses. *BMC Medicine* 14:147.

Bell ML, Ebisu K, Belanger K. 2007. Ambient air pollution and low birth weight in connecticut and massachusetts. *Environmental Health Perspectives* 115:1118-1124.

Bell ML, Belanger K, Ebisu K, Gent JF, Lee HJ, Koutrakis P, et al. 2010. Prenatal exposure to fine particulate matter and birth weight: Variations by particulate constituents and sources. *Epidemiology* 21:884-891.

Bijnens EM, Derom C, Gielen M, Winckelmans E, Fierens F, Vlietinck R, et al. 2016. Small for gestational age and exposure to particulate air pollution in the early-life environment of twins. *Environmental Research* 148:39-45.

Bonzini M, Carugno M, Grillo P, Mensi C, Bertazzi PA, Pesatori AC. 2010. Impact of ambient air pollution on birth outcomes: Systematic review of the current evidences. *Med Lav* 101:341-363.

Bosetti C, Nieuwenhuijsen MJ, Gallus S, Cipriani S, La Vecchia C, Parazzini F. 2010. Ambient particulate matter and preterm birth or birth weight: A review of the literature. *Archives of Toxicology* 84:447-460.

Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. 2019. Updated guidance for trusted systematic reviews: A new edition of the cochrane handbook for systematic reviews of interventions. Cochrane Database Syst Rev 10:ED000142.

Cutland CL, Lackritz EM, Mallett-Moore T, Bardaji A, Chandrasekaran R, Lahariya C, et al. 2017. Low birth weight: Case definition & guidelines for data collection, analysis, and presentation of maternal immunization safety data. *Vaccine* 35:6492-6500.

Darrow LA, Klein M, Strickland MJ, Mulholland JA, Tolbert PE. 2011. Ambient air pollution and birth weight in full-term infants in Atlanta, 1994-2004. *Environmental Health Perspectives* 119:731-737.

Ebisu K, Berman JD, Bell ML. 2016. Exposure to coarse particulate matter during gestation and birth weight in the U.S. *Environment International* 94:519-524.

Erickson AC, Ostry A, Chan LH, Arbour L. 2016. The reduction of birth weight by fine particulate matter and its modification by maternal and neighbourhood-level factors: A multilevel analysis in British Columbia, Canada. *Environmental Health* 15:51.

Fong KC, Di Q, Kloog I, Laden F, Coull BA, Koutrakis P, et al. 2019. Relative toxicities of major particulate matter constituents on birthweight in Massachusetts. *Environmental Epidemiology* 3:e047.

Geer LA, Weedon J, Bell ML. 2012. Ambient air pollution and term birth weight in Texas from 1998 to 2004. *Journal of Air Waste Management Assoc* 62:1285-1295.

Ghosh R, Rankin J, Pless-Mulloli T, Glinianaia S. 2007. Does the effect of air pollution on pregnancy outcomes differ by gender? A systematic review. *Environmental Research* 105:400-408.

Giovannini N, Schwartz L, Cipriani S, Parazzini F, Baini I, Signorelli V, et al. 2018. Particulate matter (pm10) exposure, birth and fetal-placental weight and umbilical arterial ph: Results from a prospective study. *Journal of Maternal, Fetal & Neonatal Medicine* 31:651-655.

Glinianaia, S. V., Rankin, J., Bell, R., Pless-Mulloli, T., & Howel, D. (2004). Particulate air pollution and fetal health: a systematic review of the epidemiologic evidence. *Epidemiology (Cambridge, Mass.), 15*(1), 36–45. https://doi.org/10.1097/01.ede.0000101023.41844.ac

Gouveia N, Bremner SA, Novaes HM. 2004. Association between ambient air pollution and birth weight in sao paulo, brazil. *Journal of Epidemiology & Community Health* 58:11-17.

Gray SC, Edwards SE, Miranda ML. 2010. Assessing exposure metrics for pm and birth weight models. *Journal of Exposure Science & Environmental Epidemiology* 20:469-477.

Gray SC, Edwards SE, Schultz BD, Miranda ML. 2014. Assessing the impact of race, social factors and air pollution on birth outcomes: A population-based study. *Environmental Health* 13:4.

Guo P, Chen Y, Wu H, Zeng J, Zeng Z, Li W, et al. 2020. Ambient air pollution and markers of fetal growth: A retrospective population-based cohort study of 2.57 million term singleton births in China. *Environment International* 135:105410.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. 2008. Grade: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal* 336:924-926.

Han Y, Ji Y, Kang S, Dong T, Zhou Z, Zhang Y, et al. 2018. Effects of particulate matter exposure during pregnancy on birth weight: A retrospective cohort study in Suzhou, China. *Science of the Total Environment* 615:369-374.

Hannam K, McNamee R, Baker P, Sibley C, Agius R. 2014. Air pollution exposure and adverse pregnancy outcomes in a large uk birth cohort: Use of a novel spatio-temporal modelling technique. *Scandinavian Journal of Work and Environmental Health* 40:518-530.

He T, Zhu J, Wang J, Ren X, Cheng G, Liu X, et al. 2018. Ambient air pollution, h19/dmr methylation in cord blood and newborn size: A pilot study in Zhengzhou city, China. *Chemosphere* 212:863-871.

Huang C, Nichols C, Liu Y, Zhang Y, Liu X, Gao S, et al. 2015. Ambient air pollution and adverse birth outcomes: A natural experiment study. *Population Health Metrics* 13:17.

Hyder A, Lee HJ, Ebisu K, Koutrakis P, Belanger K, Bell ML. 2014. Pm2.5 Exposure and birth outcomes: Use of satellite- and monitor-based data. *Epidemiology* 25:58-67.

Jedrychowski W, Perera F, Mrozek-Budzyn D, Mroz E, Flak E, Spengler JD, et al. 2009. Gender differences in fetal growth of newborns exposed prenatally to airborne fine particulate matter. *Environmental Research* 109:447-456.

Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, et al. 2014. The navigation guide - evidence-based medicine meets environmental health: Systematic review of human evidence for PFOA effects on fetal growth. *Environmental Health Perspectives* 122:1028-1039.

Keller JP, Chang HH, Strickland MJ, Szpiro AA. 2017. Measurement error correction for predicted spatiotemporal air pollution exposures. *Epidemiology* 28:338-345.

Kim OJ, Ha EH, Kim BM, Seo JH, Park HS, Jung WJ, et al. 2007. Pm10 and pregnancy outcomes: A hospital-based cohort study of pregnant women in Seoul. *Journal of Occupational & Environmental Medicine* 49:1394-1402.

Kirwa K, McConnell-Rios R, Manjourides J, Cordero J, Alshawabekeh A, Suh HH. 2019. Low birth weight and PM2.5 in Puerto Rico. *Environmental Epidemiology* 3.

Knapp G, Hartung J. 2003. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* 22:2693-2710.

Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The navigation guide - evidence-based medicine meets environmental health: Systematic review of nonhuman evidence for pfoa effects on fetal growth. *Environmental Health Perspectives* 122:1015-1027.

Kumar N. 2012. Uncertainty in the relationship between criteria pollutants and low birth weight in Chicago. *Atmospheric Environment* (1994) 49:171-179.

Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The navigation guide - evidence-based medicine meets environmental health: Integration of animal and human evidence for pfoa effects on fetal growth. *Environmental Health Perspectives* 122:1040-1051.

Lam J, Sutton P, Kalkbrenner A, Windham G, Halladay A, Koustas E, et al. 2016. A systematic review and meta-analysis of multiple airborne pollutants and autism spectrum disorder. *PLoS One* 11:e0161851.

Lamichhane DK, Leem JH, Lee JY, Kim HC. 2015. A meta-analysis of exposure to particulate matter and adverse birth outcomes. *Environmental Health and Toxicology* 30:e2015011.

Lamichhane DK, Ryu J, Leem JH, Ha M, Hong YC, Park H, et al. 2018. Air pollution exposure during pregnancy and ultrasound and birth measures of fetal growth: A prospective cohort study in Korea. *Science of the Total Environment* 619-620:834-841.

Laurent O, Wu J, Li L, Chung J, Bartell S. 2013. Investigating the association between birth weight and complementary air pollution metrics: A cohort study. *Environmental Health* 12:18.

Lavigne E, Burnett RT, Stieb DM, Evans GJ, Godri Pollitt KJ, Chen H, et al. 2018. Fine particulate air pollution and adverse birth outcomes: Effect modification by regional nonvolatile oxidative potential. *Environmental Health Perspectives* 126:077012.

Li Z, Yuan X, Fu J, Zhang L, Hong L, Hu L, et al. 2019. Association of ambient air pollutants and birth weight in Ningbo, 2015-2017. *Environmental Pollution* 249:629-637.

Mannes T, Jalaludin B, Morgan G, Lincoln D, Sheppeard V, Corbett S. 2005. Impact of ambient air pollution on birth weight in Sydney, Australia. *Occupational & Environmental Medicine* 62:524-530.

Medeiros A, Gouveia N. 2005. [relationship between low birthweight and air pollution in the city of Sao Paulo, Brazil]. *Rev Saude Publica* 39:965-972.

Merklinger-Gruchala A, Kapiszewska M. 2015. Association between pm10 air pollution and birth weight after full-term pregnancy in Krakow city 1995-2009--trimester specificity. *Annals of Agricultural and Environmental Medicine* 22:265-270.

Morello-Frosch R, Jesdale BM, Sadd JL, Pastor M. 2010. Ambient air pollution exposure and full-term birth weight in California. *Environmental Health* 9:44.

Parker JD, Woodruff TJ, Basu R, Schoendorf KC. 2005. Air pollution and birth weight among term infants in California. *Pediatrics* 115:121-128.

Parker JD, Woodruff TJ. 2008. Influences of study design and location on the relationship between particulate matter air pollution and birthweight. *Pediatric Perinatal Epidemiology* 22:214-227.

Pedersen M, Giorgis-Allemand L, Bernard C, Aguilera I, Andersen AM, Ballester F, et al. 2013. Ambient air pollution and low birthweight: A European cohort study (Escape). *Lancet Respiratory Medicine* 1:695-704.

Rahmalia A, Giorgis-Allemand L, Lepeule J, Philippat C, Galineau J, Hulin A, et al. 2012. Pregnancy exposure to atmospheric pollutants and placental weight: An approach relying on a dispersion model. *Environment International* 48:47-55.

Rhee J, Fabian MP, Ettinger de Cuba S, Coleman S, Sandel M, Lane KJ, et al. 2019. Effects of maternal homelessness, supplemental nutrition programs, and prenatal PM2.5 on birthweight. *International Journal of Environmental Research Public Health* 16(21):4154.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental Health Perspectives* 122:711-718.

Rossner, P., Jr, Tabashidze, N., Dostal, M., Novakova, Z., Chvatalova, I., Spatova, M., & Sram, R. J. (2011). Genetic, biochemical, and environmental factors associated with pregnancy outcomes in newborns from the Czech Republic. *Environmental Health Perspectives*, *119*(2), 265–271. https://doi.org/10.1289/ehp.1002470

Rudra, C. B., Williams, M. A., Sheppard, L., Koenig, J. Q., & Schiff, M. A. (2011). Ambient carbon monoxide and fine particulate matter in relation to preeclampsia and preterm delivery in

western Washington State. *Environmental Health Perspectives, 119*(6), 886–892. https://doi.org/10.1289/ehp.1002947

Salam MT, Millstein J, Li YF, Lurmann FW, Margolis HG, Gilliland FD. 2005. Birth outcomes and prenatal exposure to ozone, carbon monoxide, and particulate matter: Results from the children's health study. *Environmental Health Perspectives* 113:1638-1644.

Santos Vde P, de Medeiros AP, de Lima TA, Nascimento LF. 2014. [the effect of air pollutants on birth weight in medium-sized towns in the state of sao paulo]. *Brazilian Journal of Epidemiology* 32:306-312.

Savitz DA, Bobb JF, Carr JL, Clougherty JE, Dominici F, Elston B, et al. 2014. Ambient fine particulate matter, nitrogen dioxide, and term birth weight in New York, New York. *American Journal of Epidemiology* 179:457-466.

Schembari A, de Hoogh K, Pedersen M, Dadvand P, Martinez D, Hoek G, et al. 2015. Ambient air pollution and newborn size and adiposity at birth: Differences by maternal ethnicity (the born in Bradford study cohort). *Environmental Health Perspectives* 123:1208-1215.

Schwarz L, Bruckner T, Ilango SD, Sheridan P, Basu R, Benmarhnia T. 2019. A quantile regression approach to examine fine particles, term low birth weight, and racial/ethnic disparities. *Environmental Epidemiology* 3:e060.

Sellier Y, Galineau J, Hulin A, Caini F, Marquis N, Navel V, et al. 2014. Health effects of ambient air pollution: Do different methods for estimating exposure lead to different results? *Environment International* 66:165-173.

Shah PS, Balkhair T, Knowledge Synthesis Group on Determinants of Preterm LBWb. 2011. Air pollution and birth outcomes: A systematic review. *Environment International* 37:498-516.

Stieb DM, Chen L, Eshoul M, Judek S. 2012. Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. *Environmental Research* 117:100-111.

Stieb DM, Chen L, Beckerman BS, Jerrett M, Crouse DL, Omariba DW, et al. 2016. Associations of pregnancy outcomes and PM2.5 in a national Canadian study. *Environmental Health Perspectives* 124:243-249.

Sun X, Luo X, Zhao C, Zhang B, Tao J, Yang Z, et al. 2016. The associations between birth weight and exposure to fine particulate matter (PM2.5) and its chemical constituents during pregnancy: A meta-analysis. *Environmental Pollution* 211:38-47.

van den Hooven EH, Pierik FH, de Kluizenaar Y, Willemsen SP, Hofman A, van Ratingen SW, et al. 2012. Air pollution exposure during pregnancy, ultrasound measures of fetal growth, and adverse birth outcomes: A prospective cohort study. *Environmental Health Perspectives* 120:150-156.

Vinikoor-Imler LC, Davis JA, Meyer RE, Messer LC, Luben TJ. 2014. Associations between prenatal exposure to air pollution, small for gestational age, and term low birthweight in a state-wide birth cohort. *Environmental Research* 132:132-139.

Winckelmans E, Cox B, Martens E, Fierens F, Nemery B, Nawrot TS. 2015. Fetal growth and maternal exposure to particulate air pollution--more marked effects at lower exposure and modification by gestational duration. *Environmental Research* 140:611-618.

Woodruff TJ, Zeise L, Axelrad DA, Guyton KZ, Janssen S, Miller M, et al. 2008. Meeting report: Moving upstream-evaluating adverse upstream end points for improved risk assessment and decision-making. *Environmental Health Perspectives* 116:1568-1575.

Woodruff TJ, Sutton P, Navigation Guide Work G. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. *Health Affairs* (Millwood) 30:931-937.

Woodruff TJ, Sutton P. 2014. The navigation guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. *Environmental Health Perspectives* 122:1007-1014.

Xiao Q, Chen H, Strickland MJ, Kan H, Chang HH, Klein M, et al. 2018. Associations between birth outcomes and maternal PM2.5 exposure in Shanghai: A comparison of three exposure assessment approaches. *Environment International* 117:226-236.

Xue T, Zhu T, Han Y. 2018. Association between birthweight and ambient PM2.5 in the United States: Individually-varied susceptibility and spatial heterogeneity. *Environment International* 119:388-397.

Yang CY, Tseng YT, Chang CC. 2003. Effects of air pollution on birth weight among children born between 1995 and 1997 in Kaohsiung, Taiwan. *Journal of Toxicology & Environmental Health* A 66:807-816.

Ye L, Ji Y, Lv W, Zhu Y, Lu C, Xu B, et al. 2018. Associations between maternal exposure to air pollution and birth outcomes: A retrospective cohort study in Taizhou, China. *Environmental Science & Pollution Research* 25:21927-21936.

Yuan L, Zhang Y, Wang W, Chen R, Liu Y, Liu C, et al. 2020. Critical windows for maternal fine particulate matter exposure and adverse birth outcomes: The Shanghai birth cohort study. *Chemosphere* 240:124904.

3. MATERNAL EXPOSURE TO POLYCYCLIC AROMATIC HYDROCARBONS IN SOUTH TEXAS, EVALUATION OF SILICONE WRISTBANDS AS PERSONAL PASSIVE SAMPLERS

3.1. Introduction

Polycyclic aromatic hydrocarbons are a group of complex organic compounds formed from incomplete combustion. They are environmentally persistent contaminants produced from various sources, including vehicular emissions, tobacco smoke, cooking and other combustion processes (Abdel-Shafy et al., 2015). PAHs are commonly found bound to fine particulate matter ($PM_{2.5}$) and ultrafine particulate matter ($PM_{0.1}$) (Mohanraj et al., 2012). PAH metabolites can be excreted in urine and used as biological markers of exposure (Strickland et al 1996). Biomarkers of exposure are effective tools for assessing exposure to specific compounds and determining health risks (Huang et al., 2012). One of the commonly measured biomarkers of exposure to PAHs is urinary 1-hydroxypyrene (1-OHP), a metabolite of pyrene which is often present in PAH mixtures (Tsai et al., 2003).

PAHs are known to be reproductive and developmental toxicants that can cross the placenta due to their lipophilic nature (Drwal et al., 2019). Maternal exposure to PAHs during pregnancy is associated with adverse effects on children's health. Several epidemiologic studies have linked exposure to PAHs with various adverse birth outcomes and respiratory diseases. These include, pre-term birth, low birth weight, asthma, and chronic obstructive pulmonary disease (COPD) (Huang et al., 2012). In a multi-ethnic longitudinal cohort in New York City, prenatal exposure to ambient PAHs was associated with fetal growth restriction, adverse

cognitive development, and obesity in childhood (Choi et al., 2008, Perera et al., 2006, Rundle et al.,2012). Results from a previous biomonitoring study in South Texas, in the city of Brownsville, indicate pregnant women and their fetuses are exposed to PAH compounds based on measurements in maternal and umbilical cord blood samples (McCormick et al. 2011). While the prevalence of childhood asthma in this region (11.4%) is higher than the state average (8%) and there is an elevated prevalence of pre-term birth (12.8% vs. state-wide 10.8%), studies have yet to investigate the role of early life air pollution exposure, including PAHs.

McAllen is the largest city in Hidalgo county, located along the U.S.-Mexico border in the southern part of Texas. McAllen is ranked 12th most polluted city in terms of year-round particle pollution (American Lung Association State of the Air). The annual $PM_{2.5}$ concentration during the sampling period measured at CAMS 43 (Continuous Ambient Monitoring Station) in Hidalgo county ranged from 0.2 ug/m^3 to 58 ug/m^3 (TCEQ, 2016). These levels of PM may be due to the presence of local factories (*maquiladoras)* which lack proper waste management facilities as well as certain agricultural activities like forest burning may play a contributory role (Carrillo et al. 2018). Previously, we carried out a pilot study in Hidalgo County, to characterize personal exposure to $PM_{2.5}$ in pregnant women (Zamora et al, 2018). Compared to an ambient monitoring station, the person specific $PM_{2.5}$ was frequently more than double the observed mass concentration from the stationary monitor, indicating individuals may be exposed to elevated $PM_{2.5}$ mass concentrations during a critical window of susceptibility (Zamora et al, 2018).

The main purpose of this study was to quantify personal exposure to PAHs in pregnant women using the same set of samples from the $PM_{2.5}$ study. The second aim was to compare the PAH levels measured using active samplers with concentrations in silicone wristbands deployed

as passive. While active air sampling yields refined quantitative data, is not feasible for use over a prolonged period such as pregnancy. Recent studies support silicone wristbands to be effective in measuring personal exposure to PAHs, as well as other volatile organic compounds (Anderson et al.,2017). In an exposure assessment study in rural Peru, Bregmann et al. observed higher levels of PAHs in wristbands worn by women compared to men, inferring that silicone wristbands may distinguish lifestyle factors in environmental health studies (Bregmann et al., 2017). Further, a study by Dixon et al, on 22 pregnant women in New York City, demonstrated that silicone wristbands were able to recover PAHs in 48-hour deployments (Dixon et al. 2018). They also found more significant correlations between PAH and OH-PAH pairs in silicone wristband and urine samples than between filters and urine samples. Pregnant women form part of a biologically susceptible population and as such studies like this are important in environmental health to demonstrate effective alternative exposure assessment tools that are useful in evaluating future developmental or reproductive health outcomes.

3.2. Materials and Methods

3.2.1. Study Cohort

A total of 17 pregnant women receiving prenatal care at the Rio Grande Women's Clinic located in the McAllen-Edinburgh-Mission region were recruited into the study. Eligibility criteria included healthy women with no history of chronic disease and no existing comorbidities (non-diabetic, non- asthmatic), between the ages of 21-35 years, in their third trimester with singleton pregnancies, and residing in a non-smoking household. Written informed consent was

obtained from the participants and all study procedures were approved by Texas A&M University Institutional Review Board.

3.2.2. Study Design

Participants were given light-weight backpacks containing air sampling equipment with a silicone wristband attached close to the inlet mounted on the shoulder strap (see Figure 3.1). The backpacks were delivered by local community workers to their homes one day prior to their prenatal care appointment, and participants were administered a brief questionnaire. The questionnaire included demographic information, as well as information about their home environment. Participants were told to carry the backpack for 24 hours, then return the backpack to the clinic the following day during their prenatal appointment. This process was repeated two more times over a window of six weeks, coinciding with bi-weekly appointments. Participants were instructed to always carry the bag with them and place it close to their breathing zone during sleep or rest, while driving or otherwise sitting. At the clinic appointment, participants provided a urine sample in sterile urine cups. The urine volume was recorded, and samples were diluted with deionized water to achieve a final volume of 100mL. 5 mL of the sample was sent to Baylor Scott & White Clinical laboratory for creatinine analysis. Clinical staff also collected a small hair sample from the base of the neck to test for cotinine a metabolite of nicotine, which was carried out at Johns Hopkins University. This was done to rule out exposure to environmental tobacco smoke (ETS), which is a common source of PAHs. Sample collection took place between June 2015 and April 2016.

3.2.3. Urine Sample Methodology

Urine samples were kept frozen and shipped to the Johnson laboratory at the Texas A&M University School of Public Health, where they were stored at -80°C until analysis. Urinary 1-hydroxypyrene (1-OHP) analysis was based on the method described in Ramsauer et al. and Toriba et al. Samples were treated with 1.0 M hydrochloric acid (HCL) and 5 mL of 4M acetate buffer to adjust the pH to 5.0. Next, 75uL β-glucuronidase and samples were incubated at 37°C with gentle agitation for 4h. Following incubation, 100 mg of blue rayon was added to increase the sorption of 1-OHP, and samples were incubated for 1h at room temperature on a benchtop orbital shaker set at 60 rpm. The blue rayon was removed from the liquid phase, washed with 5 mL of DI water and air dried. Next, the blue rayon was extracted with 20 mL of methanol/ammonia (50/1, v/v) with sonication for 30 minutes. The liquid phase was dried with nitrogen and re-suspended in 5 mL methanol/ethyl acetate (1/1, v/v). The samples were run through a Sep-Pak Alumina A cartridge, (Waters, Milford, MA) preconditioned with 50 mL methanol and ethyl acetate. The samples were then eluted with 10 mL methanol. The elutant was dried under nitrogen and reconstituted in 50 uL methanol. A 25 uL aliquot was used for the 1-OHP analysis, and the remainder was saved for future analysis of nitrated OHP metabolites. Samples were spiked with 20 uL of d_9-1-OHP internal standard (200 ng/mL), prepared using methods described by Toriba et al. (Toriba et al., 2007). A Waters Acquity H-Class UPLC system coupled with a Thermo triple quadruple mass spectrometer was used for 1-OHP analysis. A Zorbax SSHD Eclipse Plus-C18 column (3.0 x 50mm, 1.8 um; Agilent, Santa Clara, CA) with a guard column (2.1 x 5mm, 1.8 um; Agilent, Santa Clara, CA) at 25°C was used for separation. A gradient elution of 0.1% ammonium acetate in water and methanol was used at a flow rate of

0.4 mL/min. Sample volumes of 10 uL were injected for each analysis. Negative electroscopy ionization at 650°C was applied, and the mass spectrometer was run in the selective reaction-monitoring (SRM) mode. Mass transitions were monitored at m/z 217→189 and 226 →198 for the 1-OHP and d_9-1-OHP analytes, respectively. Urinary concentrations of 1-OHP were expressed as ug/g creatinine to correct for variations in urine dilution.

3.2.4. Active Sampling Methodology

Each backpack given to the participants contained a Personal Environmental Monitor (PEM, MSP Inc.) connected to an external pump (BGI 400, Mesa Labs, Inc), a personal DataRAM ^{TM} (pDR-1200, Thermo Scientific Corp., Waltham, Mass) and a 37-mm Teflon filter used for gravimetric calibration to measure $PM_{2.5}$. A second line from the external pump was used to draw air through a 2 μm pore size, 37-mm polytetrafluoroethylene (PTFE) filter (Pall Corporation, Ann Arbor, MI) followed by an XAD-2 sorbent tube (SKC, Inc., Eighty Four, PA) to measure PAH concentrations. At the end of each 24-hour sampling period, when the participants returned the backpack, the PTFE filters and XAD tubes were removed and placed in clean petri dishes and PTFE bags, respectively, and stored at -20°C in the Johnson laboratory at Texas A&M University until analysis.

The PTFE filters and XAD resin tubes were Soxhlet-extracted with dichloromethane, after being spiked with the internal standards, for 16 hours and concentrated to 1 mL.

3.2.5. Passive Sampling Methodology

Silicone wristbands were precleaned before deployment according to methods described by O'Connell et al. Wristbands were washed five times to remove any background contaminants. A solution of ethyl acetate and hexane (1:1) was used for the first three washes followed by ethyl acetate and methanol (1:1) for the last two washes. Wristbands were then dried under nitrogen and stored in airtight PTFE bags at 4°C until deployment.

Silicone wristband extraction: After deployment, wristbands were shipped in PTFE bags to Texas A&M University and stored at -20°C until analysis. Wristbands were rinsed briefly with deionized water to remove surface particulates then isopropanol to dry residual water. Silicone wristbands were extracted twice in 100 mL of ethyl acetate using an orbital shaker set at 60 rotations per minute for 2 hours and reduced to 1 mL.

PAH analysis: Filters, XAD tubes and wristbands were all analyzed for 16 PAHs, including acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benzo(a)anthracene, chrysene, benzo(k)fluoranthene, benzo(a)pyrene, dibenz(a,h)anthracene, benzo(b)fluoranthene, benzo(e)pyrene, indeno[1,2,3-c,d]pyrene, and benzo(g,h,i)perylene. The PAHs were analyzed using Hewlett-Packard 6890 gas chromatograph (GC) coupled with a Hewlett-Packard 5973 mass selective detector. Separation of PAHs was accomplished with a DB-5 MS fused silica capillary column (30 m × 0.25 mm i.d., 0.50 μm film thickness, Agilent Technologies). The GC oven was temperature programmed to increase from an initial temperature of 60°C to 150°C at 15°C per min, then at 5°C per min to 220°C, and ramped at 10°C per min to a final temperature of 300°C with a final holding time of 10 min. PAHs identification was based on the comparison of the retention time and mass spectra of selected

ions with the calibration standards. Quality control (QC) samples, including field blanks and instruments blanks, were included in daily analysis. The PAH method detection limit (ranged from 0.5 to 3.5 ng/g).

3.2.6. Statistical Analysis

PAH concentrations were summarized for wristbands, XAD tubes and filters for each participant (n=17) for the three 24-hour sampling periods. One of the participants only completed 2 rounds, giving a final total of N=50 rounds for all the participants. Summary statistics of the central tendency and spread of concentrations were determined for urinary 1-OHP and PAHs measured from active and passive sampling for each round. A nonparametric analysis was performed using Spearman's rank-order correlations (r) because the concentrations were not normally distributed. Missing data were assigned values half of the limit of detection to avoid a false positive correlation among pairs of non-detected values. The correlation was used to evaluate the relationship between PAH concentrations in the wristbands, filter and XAD tubes, to reveal any trends in the PAH detected with the wristbands, filter, xad and filter+xad (active). An r coefficient of 0.20-0.39 was considered weak, 0.40-0.59 moderate, 0.60-0.79 strong and >/= 0.81 very strong (Mukaka., 2012). Statistical significance was set at an alpha of 0.05 for the analyses. Statistical analysis was conducted using Prism, version 8.

3.3. Results

All the participants were of Hispanic origin and majority were not employed outside of their homes (Table 3.1). None of the women were exposed to environmental tobacco smoke (ETS) based on the results of the cotinine analysis (Zamora et al., 2018). A total of 16 PAHs were measured in the wristbands, filter and XAD samples.

The median PAH concentrations in the wristbands from all participants were highest for phenanthrene (64 ng/wristband), fluorene (10 ng/wristband) and chrysene (4 ng/wristband). The median PAH concentrations from the filter samples were highest for phenanthrene (23 ng/filter), benzo[g,h,i]perylene (8 ng/filter), and benzo(a)anthracene (6 ng/filter). For XAD samples, the median PAH concentrations were highest for fluorene (9 ng/tube), acenaphthene (8 ng/tube), and phenanthrene (5 ng/tube). In the three different media, phenanthrene was detected among the highest concentrations. Figure 3.2 shows the detection frequencies of each media (wristbands, XAD, and filters). Wristbands and filters seemed to capture PAHs with similar percentage contributions, as illustrated in Table 3.2. Further, filters captured more HMW PAHs, while WB and XAD captured more LMW PAHs. The XAD samples were found to have a high percentage of fluorene (29%), whereas the WB and Filter samples had high percentages of phenanthrene (69% and 27% respectively).

Table 3.3 illustrates the median concentrations of the 16 PAHs from wristbands, filter and XAD from all three rounds. Based on the central tendency and spread it was observed that PAHs with relatively low molecular weight were detected in WB and XAD at high and medium concentrations while the contrary was observed for filter. Also, the low molecular weight PAHs

(MW<202g) appeared to have been generally captured more than the high molecular weight (MW>228g) (Nethery et al. 2012).

Spearman correlation coefficients were calculated to compare concentrations of PAHs captured in the wristbands with those detected in XAD and filters for all 3 rounds (Table 3.4). The Spearman correlation coefficients for wristband and filter comparisons indicated no statistically significant correlations for any of the PAHs. There were weak to moderate correlations between wristbands and XAD for three of the 16 PAHs. For XAD and filter samples, there were mostly weak except one moderate correlation (anthracene).

Urinary 1-OHP was detected in 88% of the samples at a limit of detection of 0.01ng/mL urine. Creatinine values ranged from 5.6 to 78.6 mg/dL, with an average of 31.2 mg/dL. Concentrations of all pollutants and urinary 1-OHP did not vary significantly by rounds. Spearman correlation analysis were also performed to determine the relationship between PAHs, $PM_{2.5}$, black carbon and 1-OHP. There were no significant correlations between urinary 1-OHP and $PM_{2.5}$ as well as total active PAH concentrations. However, urinary 1-OHP levels were positively correlated with black carbon concentrations (r= 0.34, p-value=0.0271). The Spearman correlation coefficients for $PM_{2.5}$ and black carbon indicated a positive correlation (r= 0.48, p-value=0.004). Further, we compared pyrene levels measured by wristbands, XAD and filters with urinary 1-OHP and did not detect a significant correlation.

3.4. Discussion

This study describes personal exposure to polycyclic aromatic hydrocarbons in pregnant women in their third trimester living in an area with high rates of pre-term birth and childhood

asthma. We applied both active (filters, XAD) and passive sampling (wristbands) methods to determine exposure. Maternal exposure to PAHs in this study was comparable to levels measured in other cohorts in the US. The median level of exposure for the sum of 16 PAHs measured via active sampling over 24 hours was 5.54 ng/m^3 (filters) and 43.82 ng/m^3 (XADs) which is similar to levels in other US populations. In a maternal cohort living in Fresno, California, the median PAH exposure during pregnancy was predicted to be 3.6 ng/m^3, using a spatio-temporal model to assign daily estimated exposure to PAHs (Padula et al. 2014). Comparison of PAH levels between different media (filters, XAD, and wristband) revealed low variation among participants' filters between the three rounds and high variation in participants' profiles for XAD and WB was observed.

Phenanthrene was detected most frequently in the largest proportion in the three different media, which is comparable to findings from other studies (Wang et al., 2019, Dixon et., 2018). Recent exposure assessment studies, comparing wristband and active air monitoring obtained similar results applying the same methodology (Rohlman et al 2019; Dixon et al., 2018). In a group of 22 pregnant women living in New York, sampled over 48-hours by Dixon et al, the silicone wristbands detected phenanthrene as the highest median concentration (228 ng/band), similar to our study (Dixon et al., 2018). We detected a slightly higher median concentration (385.85 ng/band) in our study participant which is similar to the range detected in a Native American community located in Washington State, where phenanthrene levels ranged from 100-500 ng/band (Rohlman et al. 2019). Phenanthrene is an abundant component of indoor air pollution resulting from combustion during cooking and heating (Mumford et al., 1995). In our study, 88% of the participants were stay-at-home moms, and as such, cooking activities may

have contributed to the total individual concentrations. The study by Zamora et al, in the same population confirmed that the participants spent over 70% of their time in the home microenvironment and cooking emissions were found to be a significant $PM_{2.5}$ source for many of our study participants (Zamora et al., 2018).

Results from our study also demonstrate 1-OHP correlated with black carbon (BC) concentrations quantified from active sampling. Cooking can contribute to personal BC exposures particularly with the use of unclean stoves in poorly ventilated spaces. Further, a study by Gardiner et al in black carbon workers revealed that exposure to black carbon could result in an elevation of 1-OHP which, may be as a result of the deposition or adsorption of PAHs on the Black Carbon particles, however, there was no statistically significant correlation (Gardiner et al., 1992).

One limitation from our study may be the timing of urine collection and half-life of 1-OHP. 1-OHP excretion represents recent PAH exposure with half-life values reported at 9.8 hours,12 hours, and a range from 6 to 35 hours (Viau et al., 1995, Brzeznicki et al., 1997 & Jongeneelen et al., 1990). In our study, participants provided morning urine samples (first void) prior to 24 hours of personal monitoring. Personal monitors, including silicone wristbands, were typically returned in mid- to late-morning at scheduled prenatal exams. It is likely that 1-OHP levels reflected previous day exposure, thereby influencing correlation analyses. In addition, 1-OHP levels also reflect dietary exposure, and several of our study participants reported regular consumption of grilled and charbroiled foods. In this pilot, we did not restrict dietary intake during sampling, thus, dietary PAHs may represent a significant source of 1-OHP excretion. The main cooking methods reported from the questionnaire were gas (53%) and electric (29%).

Details on ventilation during cooking were not adequately reported. This study was limited by the small sample size and therefore a population-level exposure may not be inferred.

Strengths of this study include, the good compliance by the participants, as only one participant did not complete the study, careful control of major confounding factors of PAHs exposure like environmental tobacco smoke (ETS), as well as the use of a relatively homogenous population in an attempt at reducing inter-individual variability.

The prenatal period is an important window of exposure that can significantly affect health outcomes of the fetus after birth (Rundle et al.,2012). A study conducted by Miller et al, found that prenatal exposure to PAHs and environmental tobacco smoke (ETS) increased the risk of respiratory diseases including asthma in the first two years of life (Miller et al. 2004). Our study was able to exclude ETS exposure (nicotine levels in participant hair samples were negligible, indicating a lack of exposure to environmental tobacco smoke) a significant source of PAHs, therefore, it would be beneficial to follow up with the participants to assess the possible effects in their offspring.

3.5. Conclusion

This study showed that all the pregnant women enrolled in this study were exposed to PAHs at different magnitudes and compositions. Using wristbands, filters and XAD tubes we identified exposures to 16 PAHs. Out of these three methods, wristbands captured a significant number of PAHs when compared with XAD and filters. This study provides confirmation that wristbands are an effective technology that can be easily utilized in personal exposure assessment studies in environmental health. This approach is particularly relevant to periods of

pregnancy, a known window of susceptibility for a variety of long-term health outcomes in offspring.

3.6. References

Abdel-Shafy HI, Mansour MSM. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum*. **2016**;25(1):107-123.

Air Quality Index Report | Texas Commission on Environmental Quality. Web Site. https://www.tceq.texas.gov/cgi-bin/compliance/monops/yearly_summary.pl Accessed July 20, 2020.

Anderson, K. A.; Points, G. L., 3rd; Donald, C. E.; Dixon, H. M.; Scott, R. P.; Wilson, G.; Tidwell, L. G.; Hoffman, P. D.; Herbstman, J. B.; O'Connell, S. G., Preparation and performance features of wristband samplers and considerations for chemical exposure assessment. *Journal of Exposure Science & Environmental Epidemiology* **2017,** *27*, (6), 551-559.

Bergmann, A. J.; North, P. E.; Vasquez, L.; Bello, H.; Del Carmen Gastanaga Ruiz, M.; Anderson, K. A., Multi-class chemical exposure in rural Peru using silicone wristbands. *Journal of Exposure Science & Environmental Epidemiology* **2017,** *27*, (6), 560-568.

Brzeznicki, S.; Jakubowski, M.; Czerski, B., Elimination of 1-hydroxypyrene after human volunteer exposure to polycyclic aromatic hydrocarbons. *International Archives of Occupational & Environmental Health* **1997,** *70*, (4), 257-60.

Carrillo, G.; Perez Patron, M. J.; Johnson, N.; Zhong, Y.; Lucio, R.; Xu, X., Asthma prevalence and school-related hazardous air pollutants in the US-Mexico border area. *Environmental Research* **2018,** *162*, 41-48.

Choi, H.; Rauh, V.; Garfinkel, R.; Tu, Y.; Perera, F. P., Prenatal exposure to airborne polycyclic aromatic hydrocarbons and risk of intrauterine growth restriction. *Environmental Health Perspectives* **2008,** *116*, (5), 658-65.

Dixon, H. M.; Scott, R. P.; Holmes, D.; Calero, L.; Kincl, L. D.; Waters, K. M.; Camann, D. E.; Calafat, A. M.; Herbstman, J. B.; Anderson, K. A., Silicone wristbands compared with traditional polycyclic aromatic hydrocarbon exposure assessment methods. *Analytical & Bioanalytical Chemistry* **2018,** *410*, (13), 3059-3071.

Drwal, E.; Rak, A.; Gregoraszczuk, E. L., Review: Polycyclic aromatic hydrocarbons (PAHs)-Action on placental function and health risks in future life of newborns. *Toxicology* **2019,** *411*, 133-142.

Gardiner, K.; Hale, K.A.; Calvert, I.A.; Rice, C.; Harrington J.M., The suitability of the urinary metabolite 1-hydroxypyrene as an index of poly nuclear aromatic hydrocarbon bioavailability from workers exposed to carbon black. *Annals of Occupational Hygiene*. **1992** Dec;36(6):681-8.

Huang H-B, Lai C-H, Chen G-W, Lin Y-Y, Jaakkola JJK, Liou S-H, Wang S-L. Traffic-related air pollution and DNA damage: A longitudinal study in Taiwanese traffic conductors. *PLoS ONE* **2012**, 7, (5): e37412.

Jongeneelen, F. J.; van Leeuwen, F. E.; Oosterink, S.; Anzion, R. B.; van der Loop, F.; Bos, R. P.; van Veen, H. G., Ambient and biological monitoring of cokeoven workers: determinants of

the internal dose of polycyclic aromatic hydrocarbons. *British Journal Industrial Medicine* **1990,** *47*, (7), 454-61.

Levy Zamora, M.; Pulczinski, J. C.; Johnson, N.; Garcia-Hernandez, R.; Rule, A.; Carrillo, G.; Zietsman, J.; Sandragorsian, B.; Vallamsundar, S.; Askariyeh, M. H.; Koehler, K., Maternal exposure to PM2.5 in south Texas, a pilot study. *Science of the Total Environment* **2018,** *628-629*, 1497-1507.

McAllen-Edinburg – State of the Air 2020 | American Lung Association. State of the Air 2020 report, website https://www.stateoftheair.org/city-rankings/msas/mcallen-edinburg-tx.html#pmann Accessed July 20, 2020.

Miller RL, Garfinkel R, Horton M, Camann D, Perera F, Whyatt R, Kinney P, Polycyclic aromatic hydrocarbons, environmental tobacco smoke, and respiratory symptoms in an inner-city birth cohort. *Chest*. **2004**;126(4):1071-1078.

Mohanraj, R.; Dhanakumar, S.; Solaraj, G., Polycyclic aromatic hydrocarbons bound to PM 2.5 in urban Coimbatore, India with emphasis on source apportionment. *ScientificWorldJournal* **2012,** *2012*, 980843.

Mukaka, M. M. A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal* **2012**, *24*(3), 69-71.

Mumford, J. L.; Li, X.; Hu, F.; Lu, X. B.; Chuang, J. C., Human exposure and dosimetry of polycyclic aromatic hydrocarbons in urine from Xuan Wei, China with high lung cancer mortality associated with exposure to unvented coal smoke. *Carcinogenesis* **1995,** *16*, (12), 3031-6.

Nethery, E., Wheeler, A., Fisher, M., Sjodin, A., Zheng, L., Romanoff, L., Foster, W., Arbuckle, T., Urinary polycyclic aromatic hydrocarbons as a biomarker of exposure to PAHs in air: A pilot study among pregnant women. *Journal of Exposure Science & Environmental Epidemiology* **2012**, 22, 70–81.

Padula, A. M., Noth, E. M., Hammond, S. K., Lurmann, F. W., Yang, W., Tager, I. B., Shaw, G. M., Exposure to airborne polycyclic aromatic hydrocarbons during pregnancy and risk of preterm birth. *Environmental Research* **2014**, 135,221-226.

Perera, F. P.; Rauh, V.; Whyatt, R. M.; Tsai, W. Y.; Tang, D.; Diaz, D.; Hoepner, L.; Barr, D.; Tu, Y. H.; Camann, D.; Kinney, P., Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. *Environmental Health Perspectives* **2006,** *114*, (8), 1287-92.

Ramsauer, B.; Sterz, K.; Hagedorn, H. W.; Engl, J.; Scherer, G.; McEwan, M.; Errington, G.; Shepperd, J.; Cheung, F., A liquid chromatography/tandem mass spectrometry (LC-MS/MS) method for the determination of phenolic polycyclic aromatic hydrocarbons (OH-PAH) in urine of non-smokers and smokers. *Analytical & Bioanalytical Chemistry* **2011,** *399*, (2), 877-89.

Rundle, A.; Hoepner, L.; Hassoun, A.; Oberfield, S.; Freyer, G.; Holmes, D.; Reyes, M.; Quinn, J.; Camann, D.; Perera, F.; Whyatt, R., Association of childhood obesity with maternal exposure to ambient air polycyclic aromatic hydrocarbons during pregnancy. *American Journal of Epidemiology* **2012,** *175*, (11), 1163-72.

Sexton, K.; Salinas, J. J.; McDonald, T. J.; Gowen, R. M.; Miller, R. P.; McCormick, J. B.; Fisher-Hoch, S. P., Polycyclic aromatic hydrocarbons in maternal and umbilical cord blood from

pregnant Hispanic women living in Brownsville, Texas. *International Journal of Environmental Research and Public Health* **2011,** *8*, (8), 3365-79.

Strickland P, Kang D, Sithisarankul P. Polycyclic aromatic hydrocarbon metabolites in urine as biomarkers of exposure and effect. *Environmental Health Perspectives* **1996**; 104:927-932.

Toriba, A.; Kitaoka, H.; Dills, R. L.; Mizukami, S.; Tanabe, K.; Takeuchi, N.; Ueno, M.; Kameda, T.; Tang, N.; Hayakawa, K.; Simpson, C. D., Identification and quantification of 1-nitropyrene metabolites in human urine as a proposed biomarker for exposure to diesel exhaust. *Chemical Research & Toxicology* **2007,** *20*, (7), 999-1007.

Tsai SS, Yu HS, Liu CC, Yang CY. Increased incidence of preterm delivery in mothers residing in an industrialized area in Taiwan. *Journal Toxicology & Environmental Health A.* **2003**;66(11):987-994.

Viau, C.; Carrier, G.; Vyskocil, A.; Dodd, C., Urinary excretion kinetics of 1-hydroxypyrene in volunteers exposed to pyrene by the oral and dermal route. *Science of the Total Environment* **1995,** *163*, (1-3), 179-86.

4. PERSONAL EXPOSURE TO POLYCYCLIC AROMATIC HYDROCARBONS IN THE VICINITY OF US–MEXICO BORDER CROSSINGS, A PILOT STUDY OF SCHOOL TEACHERS IN EL PASO, TEXAS.

4.1. Introduction

Traffic is a significant source of air pollution in urban environments. Traffic-related air pollution (TRAP) is mainly derived from vehicular emissions from fuel combustion. Emissions include carbon monoxide (CO), nitrogen oxides (NOx), particulate matter (PM), polycyclic aromatic hydrocarbons (PAHs), and mobile source air toxics like benzene, formaldehyde, and acetaldehyde (HEI, 2010). Chronic exposure to TRAP is associated with all-cause mortality and is a significant cause of morbidity, particularly from cardiovascular and respiratory diseases. Moreover, there is accumulating evidence on the adverse impact on pregnancy, birth outcomes, and various neurological diseases. Mechanistic research demonstrates TRAP exposure can cause such diseases through various pathways, including oxidative stress (Kelly et al., 2019). Populations that may be more sensitive to TRAP exposures include pregnant women and the developing fetus, those with pre-existing health conditions, such as asthma, heart disease, diabetes, and low socioeconomic groups, at increased risk for pre-existing conditions.

Urban traffic is an important source of polycyclic aromatic hydrocarbons (Huang et al., 2012). PAHs are bound to particulate matter (PM), including the fine PM, particles with a diameter less than 2.5 μm ($PM_{2.5}$), and ultrafine PM, particles less than 0.1 μm in diameter (Mohanraj et al.,2012). PAHs are present as complex mixtures of volatile and semivolatile species ranging from low to high molecular weight (Abdel-Shafy & Mansour 2015). Numerous experimental studies demonstrate the toxicity and carcinogenicity of PAH mixtures. Human

epidemiologic studies have linked exposure to PAHs to cancer, adverse birth outcomes, neurological and respiratory effects, including childhood asthma (Morakinyo et al., 2016, USEPA.,2017). People living near major roadways are at high risk for PAH exposure, as well as those occupationally exposed (like roofers, road workers) or those exposed to tobacco smoke or biomass.

PAHs are not routinely measured at ambient monitoring stations, which mainly report only EPA criteria air pollutants, like ozone, $PM_{10,}$ and $PM_{2.5}$. Exposure and health effects studies have primarily relied on biomonitoring of urine or blood metabolites or personal sampling techniques, mainly active samplers, to measure individual-level exposure. More recently, passive sampling approaches, including silicone wristbands, have been applied to measure personal exposure to PAHs. Dixon et al reported significant correlations between PAHs adsorbed to wristbands with urinary metabolites of different hydroxylated PAHs, overall supporting the utility of these tools for exposure assessment (Dixon et al., 2018).

The city of El Paso, located in West Texas, is home to some of the busiest border crossings in the U.S. Large numbers of vehicles spend hours idling in traffic while waiting for border inspection. Traffic congestion, coupled with high traffic volumes, contributes to increased TRAP and air quality challenges in the region. Also, Interstate-10 is a major freeway running through the heart of El Paso. The city has four major vehicular bridges linking the U.S. and Mexico. In 2010, according to the Bureau of Transportation Statistics, an estimated 10 million passenger cars, and more than 700,000 trucks drove through El Paso into Mexico (Lee et al., 2019). El Paso experiences significant air quality issues and, in 2015, were in nonattainment of the National Ambient Air Quality Standards (NAAQS) for PM_{10}. The idling traffic from the

border crossings, as well as elevated sunlight intensities, arid climate, terrain, sandstorms, and sustained temperature inversions, are all contributory factors to the poor air quality in El Paso (Lee et al., 2019). It is estimated that 60% of lifetime cancer risk can be attributed to on-road sources in El Paso (Collins et al., 2011). Schools located near areas with high air pollution have been associated with lower attendance and a higher proportion of students failing to meet state educational testing standards (Mohai et al., 2011). The numerous health effects of prolonged exposure warrant a better understanding of personal exposure, especially in near-road communities. In the first part of this study, which is reported elsewhere, we conducted a personal exposure study characterizing $PM_{2.5}$ exposure in a group of school teachers living and working in El Paso. Average $PM_{2.5}$ mass concentrations measured at CAMS 12 (Continuous Ambient Monitoring Station) during the sampling period was 8.0 $\mu g/m^3$ (TCEQ 2018). When compared with the ambient data, personal monitoring data was higher for 88% of the sampling days. The personal monitoring data for $PM_{2.5}$ was also compared with the ambient data measured at a monitor placed outside the school for the same sampling period. Although the match between the personal and ambient monitoring improved with the outside school monitor measuring an average of 7.57$\mu g/m^3$, the personal monitoring data still measured higher for 75% of the sampling days. This is consistent with other studies (Askariyeh et al., 2019; Levy Zamora et al., 2018) that have found personal monitoring to reduce the exposure misclassification inherent with ambient monitoring.

In this pilot study, our primary objective was to evaluate personal exposure to PAHs, using active and passive sampling techniques, among the same population of teachers living and working in an area with high Traffic Related Air Pollution in El Paso. We recruited school

teachers because they represent a unique population less studied in air quality research but often spend long hours in the school environment when compared to students.

4.2. Methods

4.2.1. Study Location and Population

El Paso is part of the Paso Del Norte region along the U.S.-Mexico border. In 2018, the population of El Paso was 840,000, with 80% of Hispanic origin (U.S. Census Bureau). The American Lung Associations' State of the Air Report for 2014 ranked El Paso, the 8th most polluted U.S. city in annual particulate matter pollution (El Paso state of the air 2014).

Before the start of this study, we conducted a geospatial assessment using traffic density data to identify hotspot areas most affected by traffic emissions, then selected schools within the hot spot areas for the study (details reported in another paper). This study took place in a high school located 0.62 miles from the I-10 freeway and 2.48 miles from the Bridges of America border crossing (see Figure 4.1).

All study activities were reviewed and approved by the Texas A&M University Institutional Review Board (IRB#: IRB2018-0209D) and the El Paso ISD External Research Board before initiation. Eligibility criteria included 18 years of age or older, non-smoker, and no history of cardiopulmonary disease. Participants completed a written informed consent form before the onset of the study and a questionnaire at the end of the study. The study questionnaire covered demographic information, information on their commute, and questions on home heating and cooking sources. In total, 10 participants completed the study. Demographic information is displayed in Table 4.1.

4.2.2. Exposure Assessment

All the personal sampling took place during the winter season (December 2018). Participants received lightweight backpacks containing air monitoring devices and pre-cleaned silicone wristbands, to be worn on their wrists throughout the sampling period, to capture both inhalational and dermal exposures. Backpacks were deployed between 9:30 am and 10 am and retrieved the same time the following day. Backpacks were equipped with a personal environmental monitor (PEM, MSP Inc.) as a single-stage impactor $PM_{2.5}$ inlet with an external pump operated at a flow rate of 4 L min−1. (BGI 400, Mesa Labs. Inc.) downstream a personal DataRAM PDR-1200 (Thermo Scientific Corp., Waltham, MA). A 37-mm Teflon filter was used for subsequent analysis and gravimetric calibration of $PM_{2.5}$ measurements. PAH concentrations were measured using a second line from the pump used to draw air through a 2 μm pore size, 37-mm polytetrafluoroethylene (PTFE) filter (Pall Corporation, Ann Arbor, MI) followed by an XAD-2 sorbent tube (SKC, Inc., Eighty Four, PA) at a flow rate of 1 L min−1. Participants were asked to place the backpack at breathing level while sleeping, driving, or sitting and instructed to wear the wristband through normal daily activities. A questionnaire was administered to gather information about their home environment and work commute at the end of sampling. Returned PTFE filters, and wristbands were removed, individually placed in clean petri dishes (filters) or PTFE bags (wristbands) and shipped to the Johnson Laboratory at Texas A&M University for storage (-20°C) until extraction and PAH analysis.

4.2.3. Wristband Pre-cleaning and Post-deployment Extraction

Pre-cleaning and post-deployment extraction of wristbands were conducted based on methods described in O'Connell et al. Briefly, wristbands were pre-cleaned using 1:1 (v:v) ethyl acetate and hexane for three washes, followed by 1:1 (v:v) ethyl acetate and methanol for two washes over a total duration of 12.5 hours. Subsequently, the wristbands were dried under nitrogen and stored in airtight PTFE bags at 4°C until deployment.

Post-deployment, wristbands were first rinsed with deionized water to remove surface debris and rinsed with isopropanol to remove residual water. The wristbands were extracted with 100 mL ethyl acetate and ultrasonicated for 2 hours. Individual bands were transferred to a new vial and the process repeated. The two fractions were combined and reduced to 1 ml using a TurboVap LV evaporator with high purity nitrogen (99.99%) (Zymark Center, Hopkinton, MA). Samples were loaded on a multi-layer chromatography column to eliminate lipids from the skin or personal care products. The cleanup column was dry-packed with glass wool and 1% deactivated alumina, which was pre-conditioned with hexane. Samples were eluted with ethyl acetate and reduced to 1 mL. Samples were spiked with 100 uL of PAH internal standard and transferred to amber gas chromatography vials for analysis.

4.2.4. Filter Extraction

Filters were individually placed into 60 mL vials and 20 mL of dichloromethane was added, enough to cover the filters. A 100 uL of PAH surrogate was added to the samples. Vials were capped and sonicated for 30 minutes. The extraction process was repeated and both extracts

were combined. The filter was then removed with forceps, and the extracts were concentrated to 1 mL under nitrogen then spiked with 100 uL of PAH internal standard.

4.2.5. Analytical Analysis

Samples were analyzed for 25 PAHs (Table 4.2) using a Hewlett-Packard 6890 gas chromatograph coupled to a Hewlett Packard 5973 Mass Spectrometer following methods described by Bera et al. (Bera et al., 2019). Sample extracts were spiked with internal standards (d_{10}-fluorene and d_{12}-Benzo(a)pyrene) injected in splitless-mode into a 30 m x 0.25 mm i.d. (0.5 µm) Agilent DB-5ms fused silica capillary column. The inlet was operated at a constant temperature of 300°C. The oven temperature was initially at 60°C, ramped to 150°C at 12°C/min post-injection, then increased to 220°C at 5°C/min, and finally set to 300°C at 10°C/min and held for 10 minutes. Helium was used as the carrier gas at a constant flow rate of 2.5 mL/min, and the transfer line and ion source temperatures were 280°C and 230°C, respectively. A solvent delay of 5 minutes was used to allow for elution of the hexane solvent peak prior to detection. The target compounds were quantified using their relative response factors to the appropriate surrogate standards (d_{10}-naphthalene, d_{10}- acenaphthene, d_{10}-phenanthrene, and d_{12}-chrysene), which are calculated using a 5-point calibration analyzed at the beginning of each sequence. Recovery was calculated as the percent difference between the concentration of surrogates injected on the GC/MS and the expected concentration based on the surrogate spike volume and concentration.

4.2.6. Quality Control

Field blank samples were collected during the wristband conditioning and cleaning steps. PAH surrogate standards were added to all samples to quantify chemical recoveries during the extraction process. Injections of hexane were included in all analytical samples to account for instrument background responses. Calibration standards were analyzed before and after each analytical batch to monitor instrument performance for both wristband and filter samples.

4.2.7. Data Analysis

Concentrations of PAHs in the wristbands were reported as ng/wristband and for filters, ng/filter. To compare the concentration of PAHs detected from the wristbands with those from the filters, a Spearman correlation analysis was conducted. This method was used because the PAH concentrations were not normally distributed. An r_s coefficient of 0.20-0.39 was considered to be weak, 0.40-0.59 moderate, 0.60-0.79 strong and >/= 0.81 very strong (Mukaka et al., 2012). A p-value of <0.05 was statistically significant. Statistical analyses were performed using GraphPad Prism version 8.

4.2.8. Source Identification

PAH source identification was conducted using the diagnostic ratios (Zhang et al. 2008). The sum of all the low molecular weight PAHs was compared to the sum of all the high molecular weight PAHs (∑LMW/∑HMW). If the ratio was greater than 1, it indicates that the PAHs are of a petrogenic source or from petroleum products, and if the ratio was less than 1, it indicates that the PAHs are of a pyrogenic or combustion source (Zhang et al. 2008). Similarly,

the ratio of the concentration of fluoranthene and the sum of fluoranthene and phenanthrene (FLA/FLA+PY) was used to identify the source of the PAHs (De La Torre-Roche et al., 2009). If the ratio was <0.1, it indicates a petrogenic source and if >0.1, a pyrogenic source. Last, to determine if the source was from traffic emissions, the ratio of benzo [a]pyrene to benzo[ghi]pyrene (BaP/BghiP) was calculated and, if this ratio was <0.6, it indicates a non-traffic source, however, if >0.6, it indicates source predominantly from traffic emissions (Katsoyiannis et al.,2007).

4.3. Results

Out of the 25 PAHs measured, 8 compounds were present in 50% or more of the samples, and a total of 15 were found in at least one wristband and filter samples. Naphthalene, 2-methylnaphthalene, 1-methylnapthalene, 2,6-dimethylnapthalene, biphenyl, and phenanthrene were detected in all 10 samples (both wristband and filters). Dibenzo[a,h]anthracene was the only PAH not present in neither the wristband nor filter samples (Table 4.2). In the wristbands, a total of 16 out of the 25 PAHs analyzed were detected. Nine of the PAHs detected in the wristbands were priority PAHs as defined by the U.S. Environmental Protection Agency due to their potential toxicity to humans. Similarly, 23 PAHs were present in the filters, out of which 14 were priority PAHs.

Phenanthrene was the most abundant PAH measured in the wristbands, followed by 2-methylnaphthalene and naphthalene (Figure 4.2). The median concentrations were 20.34 ng wristband^{-1} (phenanthrene), 14.03 ng wristband^{-1} (2-methylnaphthalene), and 13.67 ng wristband^{-1} (naphthalene). For the filter samples, naphthalene, 2-methylnaphthalene, and 1-

methylnapthalene were the most abundant compounds, with median concentrations of 35.68, 20.05 and 11.33 ng wristband^{-1}, respectively. Of the 25 PAHs analyzed, 10 high molecular weight compounds were measured on the filters, whereas only 4 were detected in the wristbands. Overall, the filters captured more of the high molecular weight PAHs, but in low concentrations, while the wristbands captured more low molecular weight PAHs in high concentrations.

Spearman correlation analysis was restricted to 8 PAHs that were detected >50% in the filters and wristbands (Table 4.4). There was a strong and significant positive correlation between 1-methylnaphthalene (r_s = 0.73, p-value = 0.03) in the wristband and filter PAHs. Also, there was a strong significant negative correlation for fluoranthene (r_s =-0.70, p-value = 0.04). 2-methylnaphthalene had a moderate correlation. The rest of the PAH compounds had mostly weak correlations.

The diagnostic ratios calculated for filter samples using FLA/FLA+PYR were >0.5 (Figure 4.3) indicating the source is from grass, wood, or coal combustion (De La Torre-Roche et al. 2009). Furthermore, the diagnostic ratios calculated for the filters using ∑LMW/∑HMW were all greater than one indicating petrogenic source (Zhang et al. 2008). Similar results were found when the diagnostic ratios were calculated from the wristband samples (Figure 4.3). All the FLA/FLA+PYR ratios calculated from the wristband samples were also higher than 0.5. These diagnostic ratios confirm that the dominant source of the PAHs encountered by the participants were petrogenic. To confirm that the petrogenic PAHs were from traffic emissions, we calculated the ratio of benzo[a]pyrene to benzo[ghi]perylene (BaP/BghiP) for the filter samples only, since the wristbands did not capture those compounds. The results were greater

than 0.6 indicating that the source of the PAH may be from traffic emissions (Katsoyiannis et al.,2007).

4.4. Discussion

In this pilot study, we evaluated the personal exposure to PAHs in a small group of school teachers living and working in El Paso. Naphthalene, 2-methylnaphthalene, 1-methylnapthalene, 2,6-dimethylnapthalene, biphenyl, and phenanthrene were detected in all 10 wristband and filter samples. Phenanthrene (22%) was the most abundant compound in the wristband samples, which is consistent with findings from previous research where wristbands were deployed in a similar samples size (n=10) in a U.S. population (Wang et al., 2019, Romanak et al., 2019). Naphthalene (38%) was the most abundant compound in the filter samples. Dibenzo(a,h)anthracene was the only compound not detected by neither the wristbands nor the filters, Rohlman et al reported similar findings in a Native American population (Rohlman et al., 2019). Our study which took place in Winter also detected 9 priority PAHs from wristband samples and 14 from the filter samples. Rohlman et al study also observed seasonal differences with more EPA priority PAHs captured in the Winter when compared with the Spring, these changes were however not statistically significant and may have been due to more time spent indoors as well as the use of woodstoves for heat (Rohlman et al., 2019).

Findings from the Spearman's correlation analysis revealed that wristband samples weakly correlated with the filter samples except for 1-methylnaphthalene which had strong correlations, as well as 2-methylnaphthalene that had a moderate correlation. Dixon et al. performed a similar correlation analysis on PAHs detected in >50% of their samples and found

moderate correlations with 2-methylnaphthalene. One reason for the generally weak correlations between the wristbands and filter samples may be because the wristbands captured more low molecular weight (LMW) PAHs, which partition in the gaseous phase, while the filter samples captured more high molecular weight (HMW) compounds that partition in the particle phase. Similarly, the study by Rohlman et al. detected more LMW PAHS than in HMWs in their wristband samples and reported the abundance of the same compounds as our study (phenanthrene and 2-methylnaphthalene) (Rohlman et al.,2019). These findings were contrary to Wang et al. (Wang et al., 2019), who observed that wristbands captured a significant amount of high molecular weight PAHs. Also, Dixon et al. found higher detection frequencies for HMW PAHs in wristbands than polyurethane foam filters (PUF) (Dixon et al. 2018). This difference may have been due to rinsing the wristbands with water and isopropanol before analysis, inadvertently washing off some HMW PAHs.

Findings from the diagnostic ratio analysis revealed that the PAHs were of a petrogenic source (Tobiszewski & Namiesnik 2011). Petrogenic PAHs can be found in crude oil products, gasoline and motor oil releases as well as a wide range of transportation emissions (Abdel-Shafy & Mansour 2015). These occur more often in urban locations. A recent study by Balmer et al showed that motor vehicles and indoor firewood burning were the highest sources of PAHs in the U.S. (Balmer et al., 2019). The diagnostic ratio further indicated that these petrogenic PAHs may have come from traffic sources.

Some limitations associated with this study include the rinsing of the waistbands before analysis which may have led to a potential underestimation of the actual concentrations of HMW PAHs in the wristband samples. Another limitation is the small sample size, which reduced the

ability to track exposure levels across a wider population group and was primarily due to difficulties recruiting participants and the regulations associated with the school district on receiving compensation from external sources. This study was also conducted in the winter season and in a limited time-period (December); therefore, generalizing these results to the whole year may not reflect an accurate exposure. PAH levels have been found to be higher during the winter months due to heating, this may have influenced our results (Jung et al., 2010). It would be beneficial to see the effect of variation in seasons, temperatures, and humidity on PAH detection. This study was performed in one school located in a high traffic density area. Additional data from schools located in a low traffic density area (i.e., as a control school) could improve the assessment of the traffic impacts on school exposure. Despite these limitations, this study indicates the ability of wristbands to serve as personal samplers. Results from this study provides a baseline level of polycyclic aromatic hydrocarbons prevalent in this community for future studies to use for comparison.

Other studies (McCarthy et al., 2013; Morawska et al., 2017) have found the indoor air quality levels to be influenced by outdoor conditions and indoor sources. Outdoor conditions are related to proximity to major roadways, high traffic activities, including increased idling, prevailing wind conditions, and other emission sources. Indoor sources in schools are related to poor ventilation, open windows and doors letting outdoor air inside and high student occupancy. The school is in the downwind direction, approximately 1km from a major highway I-10 and 2.5 km from a major port of entry (BOTA) characterized by high levels of idling activities, especially from heavy-duty truck traffic. Built-in 1916, the school is one of the oldest operating schools in the region and the maintenance and upkeep of the school has been lacking especially

in terms of the heating, ventilation, and air-conditioning system (HVAC), electrical upgrades and building repairs (El Paso Times 2018). An effective HVAC system is essential to filter out the outside air pollutants and poor maintenance of the system could lead to increased pollution both from outside air and self-pollution from the HVAC system (National Research Council, 2007). Studies (Shendell et al., 2004) have found an association between increased student absenteeism, lower performance with poorly maintained and ventilated schools. These findings highlight a critical need to improve the indoor air quality in the school using targeted abatement techniques such as upgrading the filtration systems, use of mechanical ventilation units, timing the opening and closing of windows and doors to avoid inflow of outdoor pollutants during peak traffic activities, and conducting anti-idle campaigns to reduce idling of vehicles during student pick-up and drop-off periods (US EPA).

4.5. Conclusion

Wristbands and filters worn by teachers in El Paso were able to recover both dermal and inhalational PAH exposures in a 24-hour period. A total of 23 PAHs were detected by the filters and 16 PAHs were captured by silicone wristbands. Naphthalene and its related compounds appeared to be the most prevalent PAH of concern. Based on the diagnostic ratios, the PAHs were mostly from petrogenic and traffic sources. Currently, there are no regulations for traffic emissions around school environments. Similar studies should be conducted in a larger, more diverse population to increase generalizability and strengthen the evidence for driving regulations and policy change.

4.6. References

Abdel-Shafy, H. I., & Mansour, M. S. M. (2016). A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian Journal of Petroleum, 25*(1), 107-123. doi:https://doi.org/10.1016/j.ejpe.2015.03.011

Balmer, J. E., Hung, H., Yu, Y., Letcher, R. J., & Muir, D. C. G. (2019). Sources and environmental fate of pyrogenic polycyclic aromatic hydrocarbons (PAHs) in the arctic. *Emerging Contaminants, 5*, 128-142. doi:https://doi.org/10.1016/j.emcon.2019.04.002

Bera, G., Camargo, K., Sericano, J. L., Liu, Y., Sweet, S. T., Horney, J., Jun, M., Chiu, W., Rusyn, I., Wade, T. L., & Knap, A. H. (2019). Baseline data for distribution of contaminants by natural disasters: results from a residential Houston neighborhood during Hurricane Harvey flooding. *Heliyon*, *5*(11), e02860. https://doi.org/10.1016/j.heliyon.

Collins, T. W., Grineski, S. E., Chakraborty, J., & McDonald, Y. J. (2011). Understanding environmental health inequalities through comparative intracategorical analysis: racial/ethnic disparities in cancer risks from air toxics in El Paso County, Texas. *Health & place*, *17*(1), 335–344. https://doi.org/10.1016/j.healthplace.2010.11.011

District Information / District Information. (n.d.). Retrieved February 11, 2020, from https://www.episd.org/domain/224

Dixon, H. M., Scott, R. P., Holmes, D., Calero, L., Kincl, L. D., Waters, K. M., Camann, D. E., Calafat, A. M., Herbstman, J. B., & Anderson, K. A. (2018). Silicone wristbands compared with traditional polycyclic aromatic hydrocarbon exposure assessment methods. *Analytical and bioanalytical chemistry*, *410*(13), 3059–3071. https://doi.org/10.1007/s00216-018-0992-z

El Paso - state of the air 2014 | American lung association. State of the Air 2014 Web site. http://www.stateoftheair.org/2014/states/texas/el-paso-48141.html. Accessed July 20, 2019.

EPISD breaks bond promise to fix El Paso High School: Anna Lucia Mares. (n.d.). Retrieved February 11, 2020, from https://www.elpasotimes.com/story/opinion/2018/08/11/episd-breaks-bond-promise-fix-el-paso-high-school-anna-lucia-mares/948040002/

HEI (Health Effects Institute). Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects. HEI special report 17. 2010.

Huang H-B, Lai C-H, Chen G-W, Lin Y-Y, Jaakkola JJK, Liou S-H, et al. (2012) Traffic-Related Air Pollution and DNA Damage: A Longitudinal Study in Taiwanese Traffic Conductors. PLoS ONE 7(5): e37412. https://doi.org/10.1371/journal.pone.0037412

Jung, K. H., Patel, M. M., Moors, K., Kinney, P. L., Chillrud, S. N., Whyatt, R., Hoepner, L., Garfinkel, R., Yan, B., Ross, J., Camann, D., Perera, F. P., & Miller, R. L. (2010). Effects of Heating Season on Residential Indoor and Outdoor Polycyclic Aromatic Hydrocarbons, Black Carbon, and Particulate Matter in an Urban Birth Cohort. *Atmospheric environment (Oxford, England: 1994)*, *44*(36), 4545–4552. https://doi.org/10.1016/j.atmosenv.2010.08.024

Katsoyiannis, A., Terzi, E., & Cai, Q. Y. (2007). On the use of PAH molecular diagnostic ratios in sewage sludge for the understanding of the PAH sources. Is this use appropriate? *Chemosphere*, *69*(8), 1337-1339.

Kelly, F. J., & Fussell, J. C. (2015). Linking ambient particulate matter pollution effects with oxidative biology and immune responses. *Annals of the New York Academy of Sciences, 1340,* 84–94. https://doi.org/10.1111/nyas.12720

Lee, K., & Sener, I. N. (2019). Understanding Potential Exposure of Bicyclists on Roadways to Traffic-Related Air Pollution: Findings from El Paso, Texas, Using Strava Metro Data. *International journal of environmental research and public health*, *16*(3), 371. https://doi.org/10.3390/ijerph16030371

Levy Zamora, M., Pulczinski, J. C., Johnson, N., Garcia-Hernandez, R., Rule, A., Carrillo, G., Zietsman, J., Sandragorsian, B., Vallamsundar, S., Askariyeh, M. H., & Koehler, K. (2018). Maternal exposure to PM2.5 in south Texas, a pilot study. *Science of the Total Environment*, *628–629*, 1497–1507. https://doi.org/10.1016/j.scitotenv.2018.02.138

McCarthy, M. C., Ludwig, J. F., Brown, S. G., Vaughn, D. L., & Roberts, P. T. (2013). Filtration effectiveness of HVAC systems at near-roadway schools. *Indoor Air*, *23*(3), 196–207. https://doi.org/10.1111/ina.12015

Mohai, P., Kweon, B.-S., Lee, S., Ard, K. (2011). Air pollution around schools is linked to poorer student health and academic performance. *Health Affairs, 30*(5), 852-862. doi: 10.1377/hlthaff.2011.0077.

Mohanraj, R., Dhanakumar, S., & Solaraj, G. (2012). Polycyclic aromatic hydrocarbons bound to PM 2.5 in urban Coimbatore, India with emphasis on source apportionment. *The Scientific World Journal*, *2012*, 980843. https://doi.org/10.1100/2012/980843

Morakinyo, O. M., Mokgobu, M. I., Mukhola, M. S., & Hunter, R. P. (2016). Health Outcomes of Exposure to Biological and Chemical Components of Inhalable and Respirable Particulate Matter. *International journal of environmental research and public health*, *13*(6), 592. https://doi.org/10.3390/ijerph13060592

Morawska, L., Ayoko, G. A., Bae, G. N., Buonanno, G., Chao, C. Y. H., Clifford, S., Fu, S. C., Hänninen, O., He, C., Isaxon, C., Mazaheri, M., Salthammer, T., Waring, M. S., & Wierzbicka, A. (2017). Airborne particles in indoor environment of homes, schools, offices, and aged care facilities: The main routes of exposure. In *Environment International* (Vol. 108, pp. 75–83). Elsevier Ltd. https://doi.org/10.1016/j.envint.2017.07.025

Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, *24*(3), 69-71.

National Research Council. (2007). *Green Schools: Attributes for Health and Learning*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11756.

O'Connell, S. G., Kincl, L. D., & Anderson, K. A. (2014). Silicone wristbands as personal passive samplers. *Environmental science & technology*, *48*(6), 3327-3335.

Roberto, J., Lee, W. Y., & Campos-Díaz, S. I. (2009). Soil-borne polycyclic aromatic hydrocarbons in El Paso, Texas: analysis of a potential problem in the United States/Mexico border region. *Journal of Hazardous Materials*, *163*(2-3), 946-958.

Rohlman, D., Dixon, H. M., Kincl, L., Larkin, A., Evoy, R., Barton, M., Phillips, A., Peterson, E., Scaffidi, C., Herbstman, J. B., Waters, K. M., & Anderson, K. A. (2019). Development of an environmental health tool linking chemical exposures, physical location, and lung function. *BMC public health*, *19*(1), 854. https://doi.org/10.1186/s12889-019-7217-z

Romanak, K. A., Wang, S., Stubbings, W. A., Hendryx, M., Venier, M., & Salamova, A. (2019). Analysis of brominated and chlorinated flame retardants, organophosphate esters, and polycyclic aromatic hydrocarbons in silicone wristbands used as personal passive samplers. *Journal of Chromatography A*, *1588*, 41-47.

Shendell, D. G., Prill, R., Fisk, W. J., Apte, M. G., Blake, D., & Faulkner, D. (2004). Associations between classroom CO2 concentrations and student attendance in Washington and Idaho. *Indoor Air*, *14*(5), 333–341. https://doi.org/10.1111/j.16000668.2004.00251.x

TCEQ 2018. Data by Year by Site by Parameter https://www.tceq.texas.gov/cgi-bin/compliance/monops/yearly_summary.pl Accessed July 20, 2020.

Tobiszewski, M., & Namieśnik, J. (2012). PAH diagnostic ratios for the identification of pollution emission sources. *Environmental pollution (Barking, Essex : 1987)*, *162*, 110–119. https://doi.org/10.1016/j.envpol.2011.10.025

U.S. Census Bureau QuickFacts: El Paso County, Texas. https://www.census.gov/quickfacts/elpasocountytexas. Accessed July 20, 2019.

U.S. Environmental Protection Agency. (2017). NIEHS/EPA Children's Environmental Health and Disease Prevention Research Centers Impact Report: Protecting children's health where they live, learn, and play. EPA Publication No. EPA/600/R-17/407. Retrieved from https://www.epa.gov/sites/production/files/2017-10/documents/niehs_epa_childrens_centers_impact_report_2017_0.pdf?pdf=children's-center-report.

US Environmental Protection Agency, (2015). *Best Practices for Reducing Near-Road Pollution Exposure at Schools*. Retrieved from https://www.epa.gov/sites/production/files/2015-10/documents/ochp_2015_near_road_pollution_booklet_v16_508.pdf

Wang, S., Romanak, K. A., Stubbings, W. A., Arrandale, V. H., Hendryx, M., Diamond, M. L., Salamova, A., & Venier, M. (2019). Silicone wristbands integrate dermal and inhalation exposures to semi-volatile organic compounds (SVOCs). *Environment International*, *132*, 105104. https://doi.org/10.1016/j.envint.2019.105104

Zhang, W., Zhang, S., Yue, D., Wan, C., Ye, Y., & Wang, X. (2008). Characterization and loading estimation of polycyclic aromatic hydrocarbons in road runoff from urban regions of Beijing, China. *Environmental Toxicology and Chemistry: An International Journal*, *27*(1), 31-37.

5. CONCLUSION

5.1. Summary

This study provided an overview of air pollution exposure assessment with a focus on PAHs in two different populations using personal monitoring. In chapter 1, we gave a brief synopsis on ambient particulate matter air pollution and exposure assessment. The second chapter was a systematic review and meta-analysis on the association between prenatal ambient particulate matter exposure and birthweight. The third chapter was on an exposure assessment study in Mc Allen, Texas to quantify personal exposure of pregnant women to PAHs as well as compare inhalational exposures from silicone wristbands deployed as passive samplers and the traditional active sampling air monitoring. In the fourth chapter, we evaluated personal exposure to PAHs, using active and passive sampling, among a population of teachers living and working in a high traffic area in El Paso. In this final chapter, a summary of each of the findings will be presented and potential future directions will be proposed.

5.2. A Systematic Review and Meta-analysis to Evaluate Prenatal Exposure to Particulate Matter Air Pollution and Birthweight

This review was done using the Navigation Guide Systematic review methodology which is based on the Cochrane Collaboration and the Grading of Recommendations Assessment Development and Evaluation - GRADE methods accounting for the differences in evidence and decision context in environmental health studies (Higgins & Green., 2011, Guyatt et al. 2008). Despite the substantial evidence on the association between developmental PM exposure and

adverse birth outcomes, including low birth weight, pre-term birth, and small for gestational age births, there have been inconsistencies in the conclusions about the association and magnitude of the effect. Cognizant of this heterogeneity across studies and potential risk of bias in individual studies, we evaluated the human evidence on prenatal exposure to particulate matter and birth weight. Based off our results, we confirmed the inverse association between prenatal exposure to PM_{10} and $PM_{2.5}$ and low birth weight and highlighted the impact of PM exposure classification on effect estimates. However, our findings reveal the need to standardize and improve exposure assessment methods in air pollution research because the various forms of exposure measurement utilized in the studies contributed to the heterogeneity seen in the meta-analysis.

5.3. McAllen Study: Evaluating Maternal Exposure to Polycyclic Aromatic Hydrocarbons using Silicone Wristbands as Passive Samplers

In this study, seventeen pregnant women in their 3rd trimester carried personal monitoring devices for three 24-hour sampling periods at 2-week intervals. The personal monitors were encased in a backpack and included an XAD resin, PTFE filters as active samplers and a silicone wristband for passive sampling. Out of these three methods, wristbands captured a significant number of PAHs when compared to XAD and filters. Phenanthrene was detected among the highest PAH concentration in the three different media. This study provides confirmation that wristbands are biologically relevant, effective technology that can be easily utilized in personal exposure assessment studies in environmental health. Our approach is particularly important to periods of pregnancy, a known window of susceptibility for a variety of long-term health outcomes in offspring.

5.4. Personal Exposure to Polycyclic Aromatic Hydrocarbons in the Vicinity of U.S.-Mexico Border Crossings, a Pilot Study of School Teachers in El Paso, Texas

This study sort to evaluate exposure to PAHs from traffic sources in a population of teachers living in a high traffic area using filters and silicone wristbands. A total of 23 PAHs were detected by the filters and 16 PAHs were captured by silicone wristbands. The PAHs were mostly from petrogenic and traffic sources as hypothesized. Phenanthrene and Naphthalene both low molecular weight PAHs were the most prevalent compounds recovered from the wristband and filter samples respectively. Harmful effects from exposure to these compounds are dependent on dose, mode of exposure and duration of the exposure. Naphthalene in high concentrations is possibly cancer causing (Lu et al., 2005) and phenanthrene is a lung and skin irritant (National Center for Biotechnology Information., 2020).

Our study further confirmed that wristbands and filters can capture and recover PAHs in an acute 24-hour period and can be utilized for large scale studies. Such studies are needed in border communities with high traffic to strengthen policies of reduction of traffic emissions.

5.5. Future Directions

Wristbands have the added advantage of capturing both inhalational and dermal exposures. They are also economical and convenient to use when compared to the traditional methods. These wristbands have been proven effective for use in 6 continents (Donald et al., 2016, Dixon et al., 2019) and should be utilized more in exposure assessment studies. There is indeed room for more exposure assessment studies using wristband as thus far, only 15 papers have been published.

Also, currently, there are no regulations for traffic emissions around school environments. Therefore, similar exposure assessment studies should be conducted in larger, more diverse populations to increase generalizability and strengthen the evidence for driving regulations and policy change.

5.6. References

Dixon, H. M., Armstrong, G., Barton, M., Bergmann, A. J., Bondy, M., Halbleib, M. L., Hamilton, W., Haynes, E., Herbstman, J., Hoffman, P., Jepson, P., Kile, M. L., Kincl, L., Laurienti, P. J., North, P., Paulik, L. B., Petrosino, J., Points, G. L., 3rd, Poutasse, C. M., Rohlman, D., Scott, R.P., Smith, B., Tidwell, L.G., Walker, C., Waters, K.M., Anderson, K. A. (2019). Discovery of common chemical exposures across three continents using silicone wristbands. *Royal Society Open Science*, *6*(2), 181836. https://doi.org/10.1098/rsos.181836

Donald, C. E., Scott, R. P., Blaustein, K. L., Halbleib, M. L., Sarr, M., Jepson, P. C., & Anderson, K. A. (2016). Silicone wristbands detect individuals' pesticide exposures in West Africa. *Royal Society Open Science*, *3*(8), 160433. https://doi.org/10.1098/rsos.160433

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann, H. J., & GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

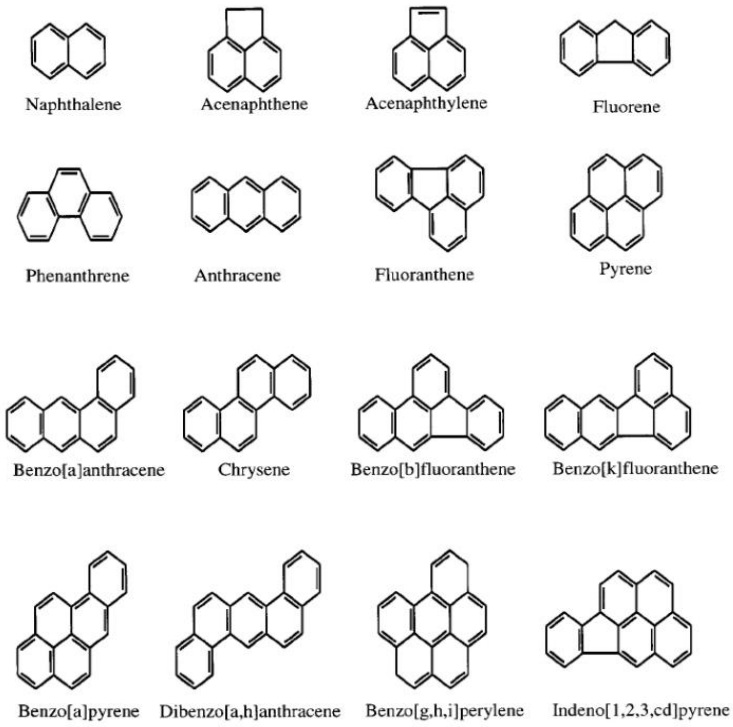
Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Lu, R., Wu, J., Turco, R. P., Winer, A. M., Atkinson, R., Arey, J., Paulson, S. E., Lurmann, F., Miguel, A.H., Eiguren-Fernandez, A. (2005). Naphthalene distributions and human exposure in Southern California. *Atmospheric Environment*, *39*(3), 489-507. doi:https://doi.org/10.1016/j.atmosenv.2004.09.045

National Center for Biotechnology Information (2020). PubChem Compound Summary for CID 995, Phenanthrene. Retrieved October 8, 2020 from https://pubchem.ncbi.nlm.nih.gov/compound/Phenanthrene.

APPENDIX A

TABLES AND FIGURES



Rogers et al.,2002

Figure 1.1 US-EPA 16 Criteria Polycyclic Aromatic Hydrocarbons (Rogers et al., 2002)

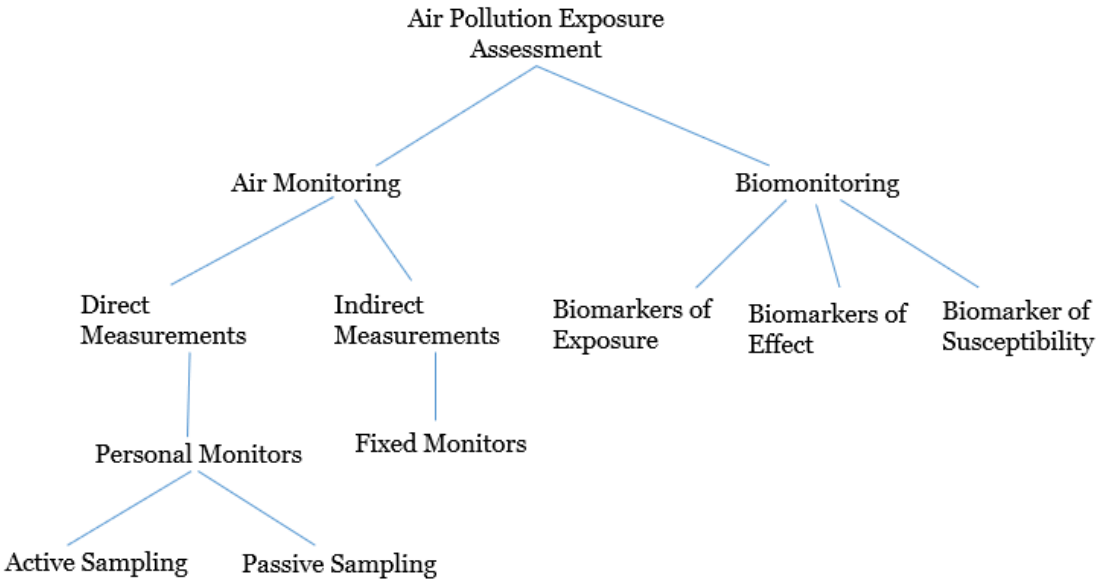


Figure 1.2 Schematic diagram of Air Pollution Exposure Assessment

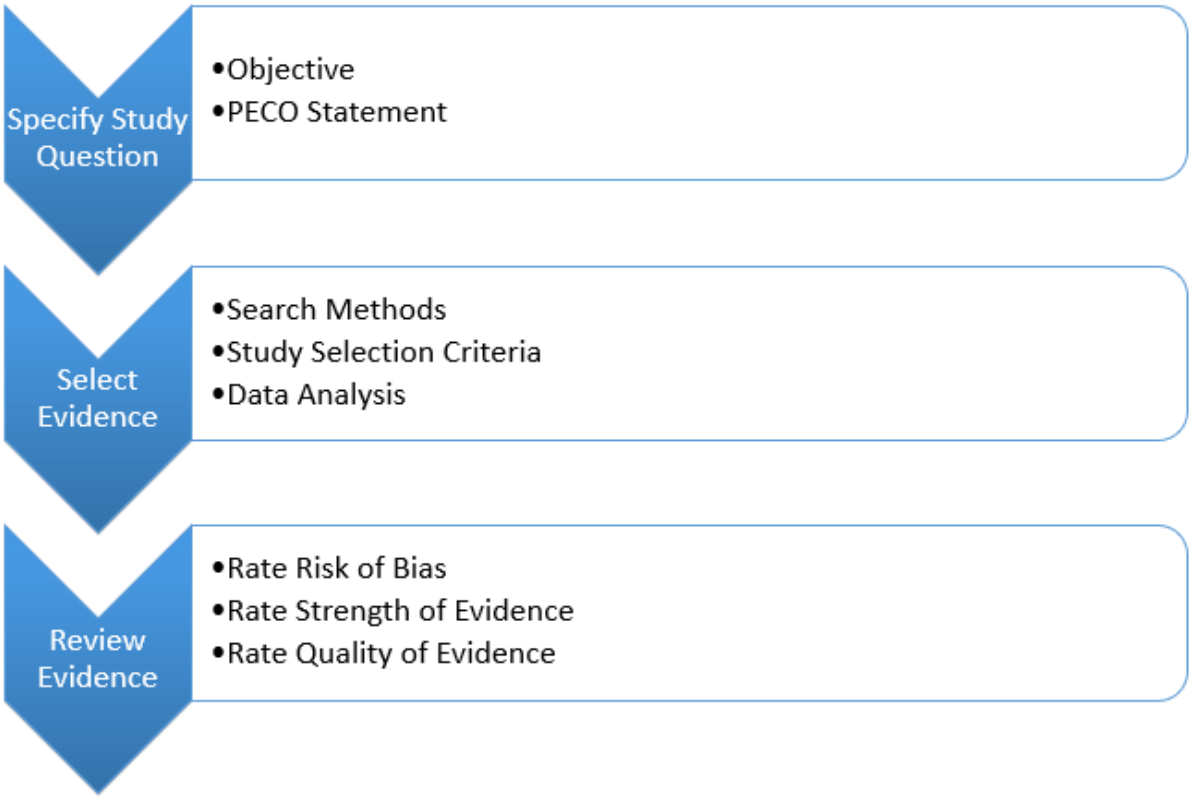


Figure 2.1 Navigation Guide Steps (adapted from Johnson et al. 2014)

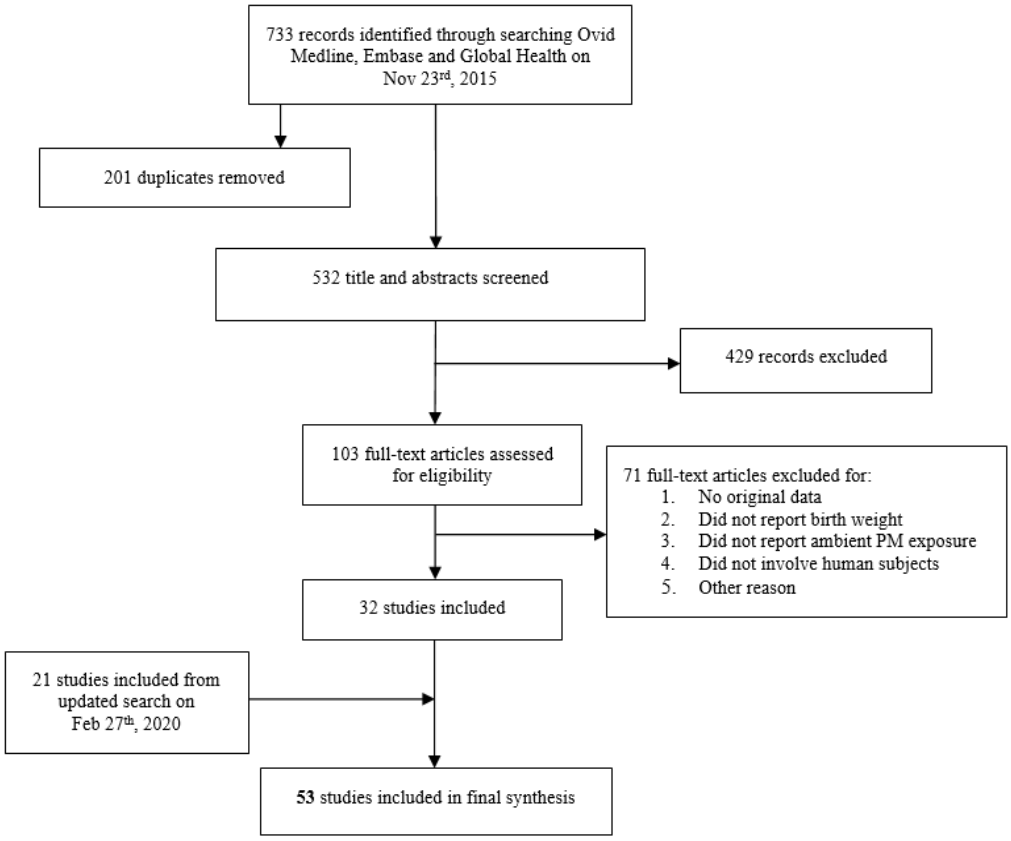


Figure 2.2 Flowchart of Study Selection process showing studies relevant to Prenatal Exposure to Ambient Particulate Matter and Birthweight.

++ Low
+ Probably low
- Probably high
-- High

1. Was the strategy for recruiting participants consistent across study groups?
2. Was confounding adequately addressed?
3. Were incomplete outcome data adequately addressed?
4. Does the study report appear to have been comprehensive in its outcome reporting?
5. Was knowledge of the exposure adequately prevented during the study?
6. Were exposure assessment methods robust?
7. Did the study appear to be free of other problems that could put it at a risk of bias?

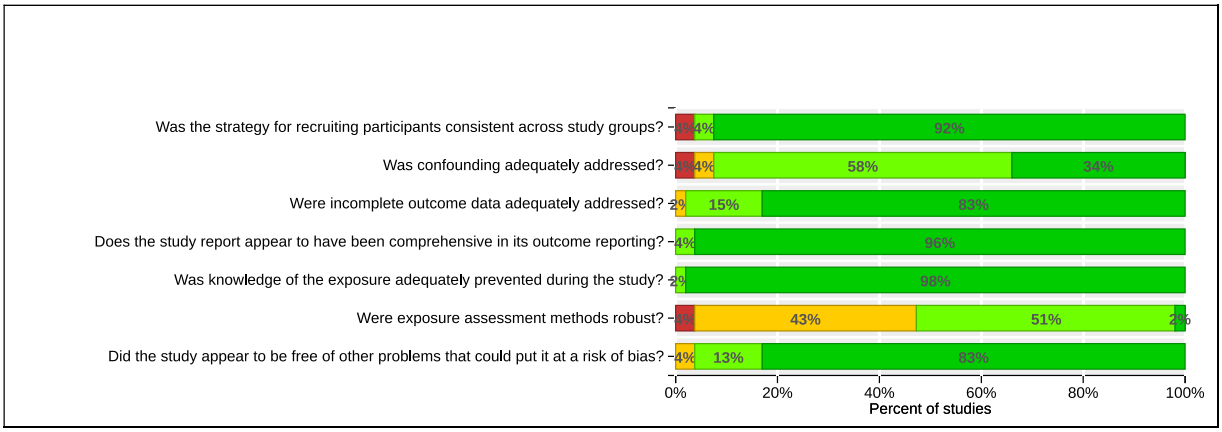


Figure 2.3 Risk of Bias Heat Map

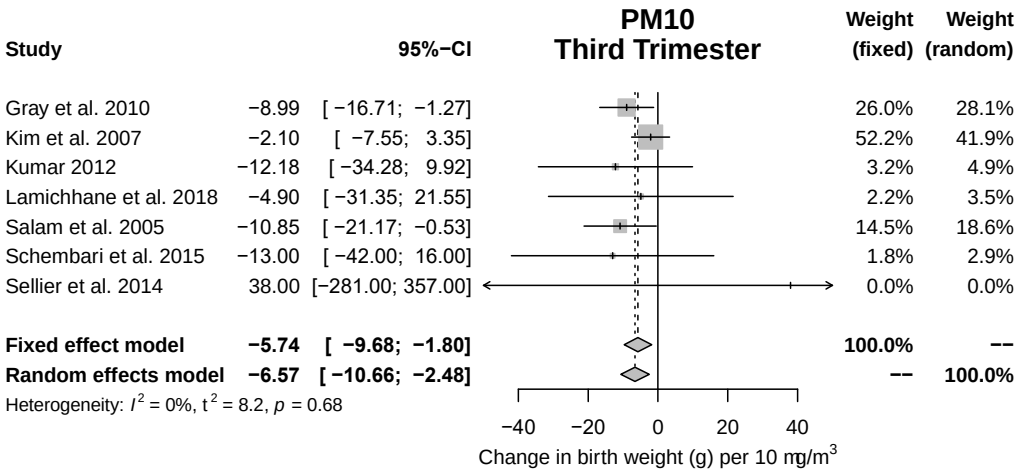


Figure 2.4 PM_{10} (Third Trimester) including Studies rated as "Low" or "Probably low" Risk of Bias

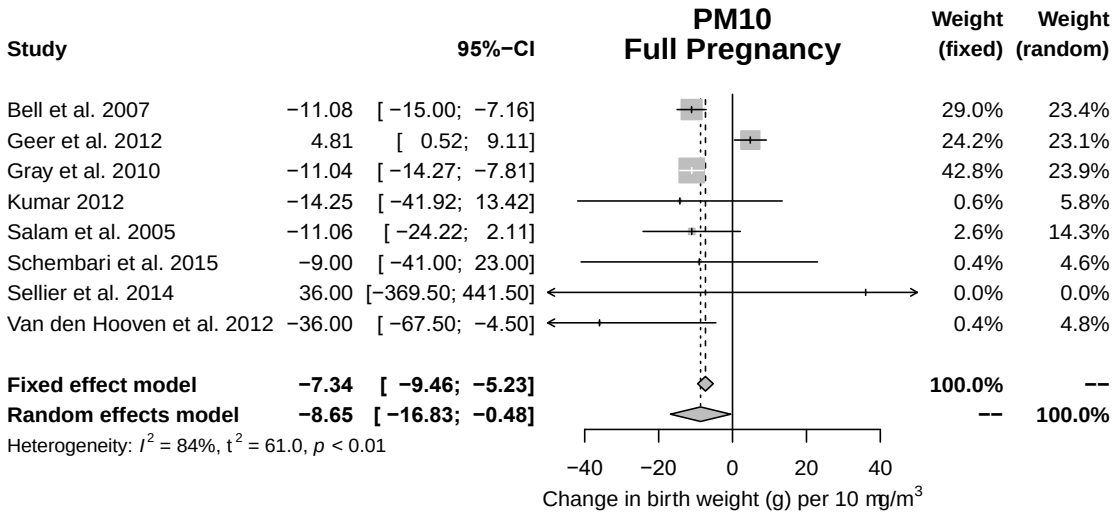
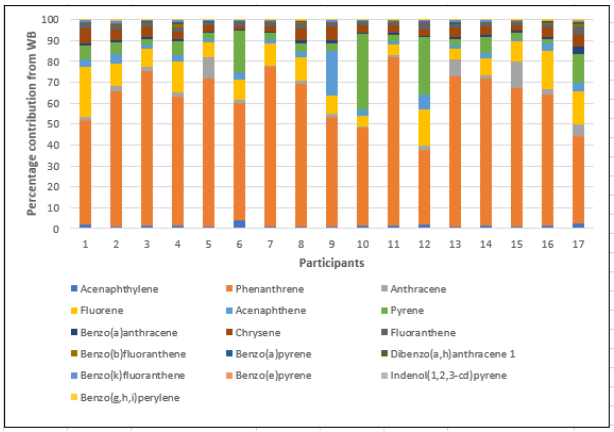
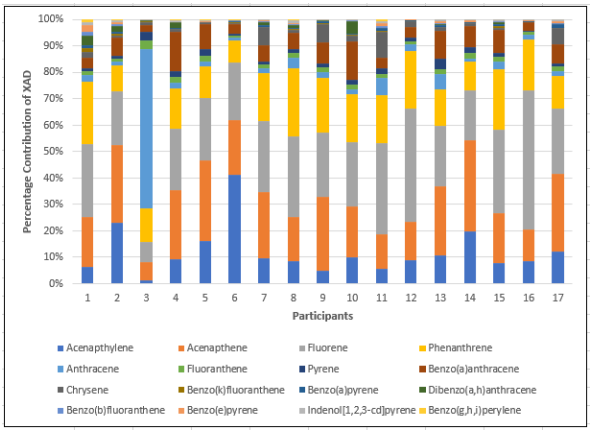


Figure 2.5 PM_{10} (Full Pregnancy) including Studies rated as "Low" or "Probably low" Risk of Bias

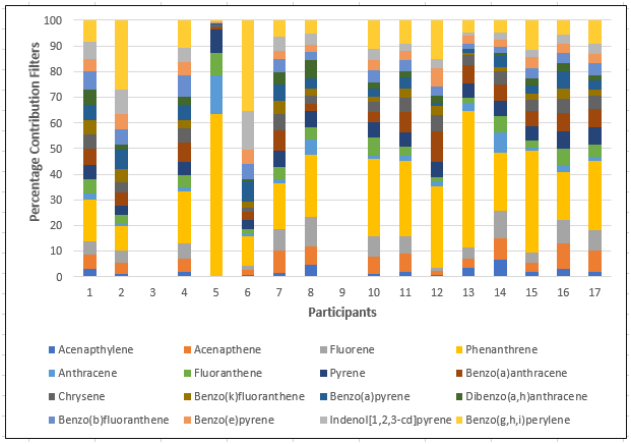
Figure 3.1 Backpack and Personal Air Monitoring Devices: Pump and PDR with Battery pack



Wristband



XAD

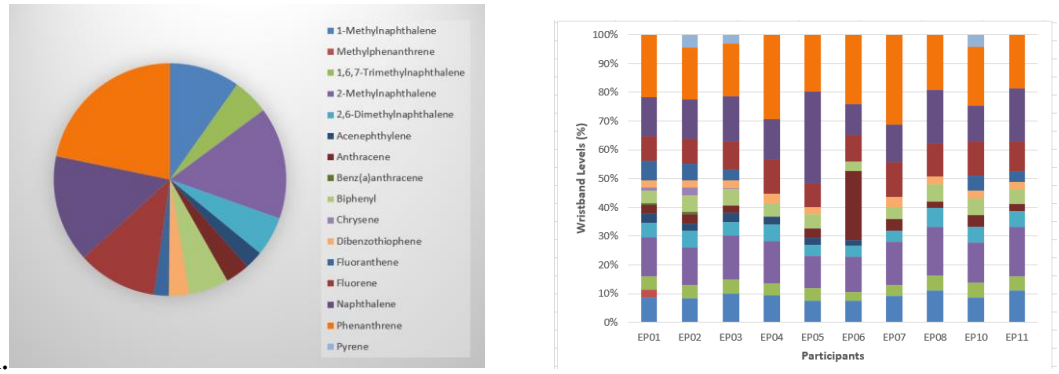


Filter

Figure 3.2 Individual PAH Frequency of Detection and Percentage Contribution for each Participant

Figure 4.1 Map showing the Study Location, Three Border Crossings, and the Continuous Ambient Monitoring Station (CAMS 12) closest to the School.

A.



B.

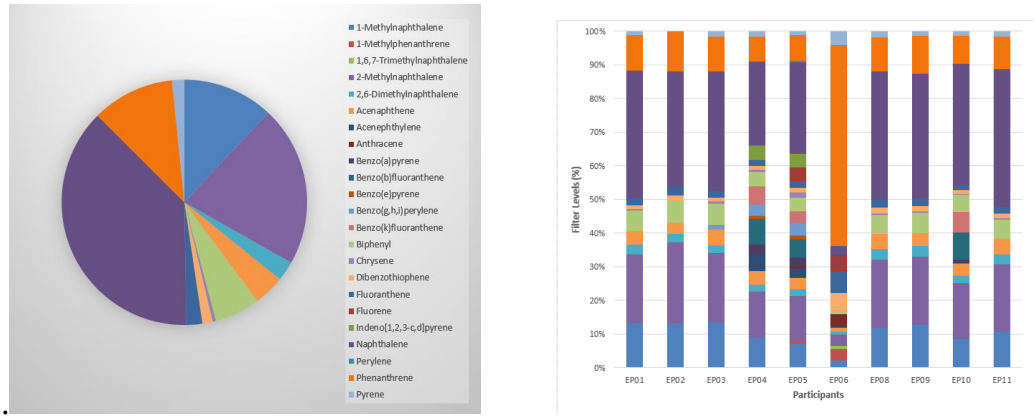
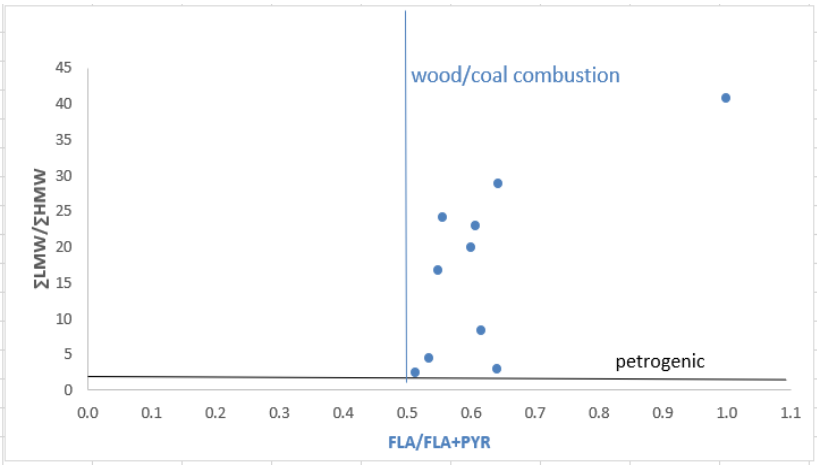


Figure 4.2 Distribution of PAH compounds in Wristbands (A) and Filters (B) for each Participant.

A.



B.

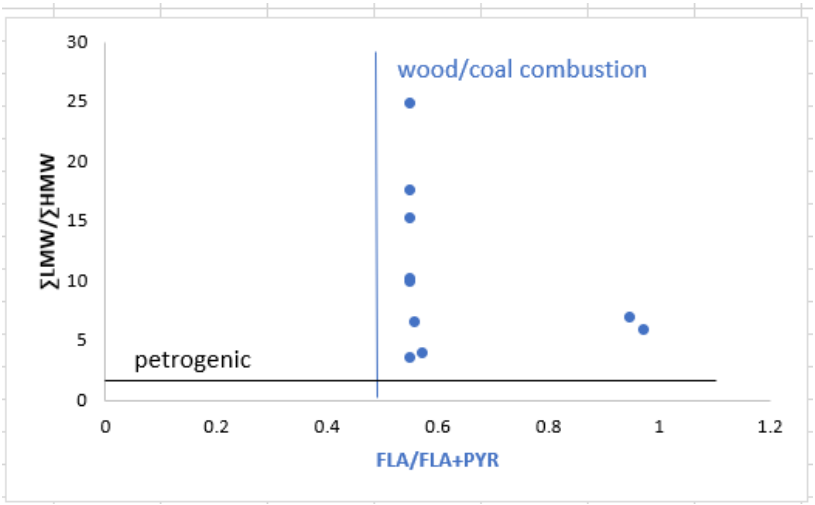


Figure 4.3 Diagnostic Ratio calculated for each Participant from Filter samples (A) and Wristband samples (B).

Table 2.1 Summary of Risk of Bias domains and Criteria for Low Risk of Bias (adapted from Johnson et al. 2014)

Risk of bias domain	Low risk of bias designation
Recruitment strategy	Protocols for recruitment and inclusion/exclusion criteria applied similarly across study groups
Blinding	Knowledge of the exposure ensured when assessing outcome, or judgement that outcome measurement not likely to be influenced by lack of blinding
Exposure assessment	Confidence in the accuracy of the exposure assessment methods that minimizes exposure misclassification, i.e., validity and reliability measures specified for monitoring and modeling
Confounding	All five important potential confounders pre-specified by reviewers are accounted for (i.e., matched, stratified, multivariate analysis or otherwise statistically controlled for)
Incomplete outcome	No missing outcome data, balanced attrition across groups, or for continuous outcome data, plausible effect size among missing outcomes not enough to have a relevant impact on the observed effect size
Selective outcome reporting	All pre-specified outcomes outlined in the protocol, methods, abstract, and/or introduction reported in the pre-specified way
Conflicts of Interest	The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study
Other bias	The study appears to be free of other sources of bias

Table 2.2 Factors for Evaluating the Quality and Strength of the Body of Evidence (adapted from Johnson et al. 2014)

Quality is rated across all studies. Evidence begins as "moderate" and may be downgraded (-1 or -2) or upgraded (+1 or +2) according to factors.		**Strength is rated across all studies.** The final ratings represent the level of certainty of toxicity.	
Downgrade factors	• Risk of bias across studies • Indirectness • Inconsistency • Imprecision • Publication bias	**Considerations**	• Quality of body of evidence • Direction of effect estimates • Confidence in effect estimates • Other compelling attributes of the data that may influence certainty
Upgrade factors	• Large magnitude of effect • Dose response • Confounding minimizes effect		
Quality rating	• High quality • Moderate quality • Low quality	**Strength rating**	• Sufficient evidence • Limited evidence • Inadequate evidence • Evidence of lack of toxicity

Table 2.3 Evidence Table

Reference	Study location	Sample size	Pollutant(s) (exposure assessment method)	Exposure period	Overall ROB rating
(Basu et al. 2014)	California, USA (8 counties)	646,296	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Beland and Oloomi 2019)	southern USA	9,324,839	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy	Probably low
(Bell et al. 2007)	Connecticut and Massachusetts, USA	358,504	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy	Probably low
(Bell et al. 2010)	Connecticut and Massachusetts, USA (4 counties)	76,788	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Bijnens et al. 2016)	Flanders, Belgium	4,760	PM_{10} (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters, last month, last week	Probably high
(Darrow et al. 2011)	Atlanta, USA (5 Counties)	402, 627	$PM_{2.5}$, PM_{10}, $PM_{2.5-10}$ (Ambient monitoring)	Entire pregnancy, 3rd trimester only	Probably high
(Ebisu et al. 2016)	USA (224 Counties)	8,017,865	$PM_{2.5-10}$ (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Erickson et al. 2016)	British Columbia, Canada	231,929	$PM_{2.5}$ (Modeling)	Entire pregnancy	Probably high
(Fong et al. 2019)	Massachusetts, USA	907,766	$PM_{2.5}$ (Modeling)	Entire pregnancy	Probably low
(Geer et al. 2012)	Texas, USA	1,548,904	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy	Probably low
(Giovannini et al. 2018)	Italy	3,614	PM_{10} (Ambient monitoring)	1st, 2nd, and 3rd trimesters	High
(Gouveia et al. 2004)	São Paulo, Brazil	179,460	PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Gray et al. 2010)	North Carolina, USA	350,754	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Gray et al. 2014)	North Carolina, USA	457, 642	$PM_{2.5}$ (Modeling)	Entire pregnancy	Probably low
(Guo et al. 2020)	Guangdong province, China	2,567,457	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy	Probably high
(Han et al. 2018)	Suzhou, China	10,915	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Hannam et al. 2014)	United Kingdom (Northwest England)	203,562	$PM_{2.5}$, PM_{10} (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(He et al. 2018)	Zhengzhou, China	591	PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high

Table 2.3 Continued

Reference	Study location	Sample size	Pollutant(s) (exposure assessment method)	Exposure period	Overall ROB rating
(Huang et al. 2015)	Beijing, China	50,874	PM_{10} (Ambient monitoring)	1st, 2nd, and 3rd trimesters	Probably high
(Hyder et al. 2014)	Connecticut and Massachusetts, USA	834,332	$PM_{2.5}$ (Ambient monitoring and modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Jedrychowski et al. 2009)	Krakow, Poland	481	$PM_{2.5}$ (Personal monitoring)	Entire pregnancy	Low
(Keller et al. 2017)	Georgia, USA	403,881	$PM_{2.5}$ (Modeling)	1st, 2nd, and 3rd trimesters	Probably low
(Kim et al. 2007)	Seoul, Korea	1,514	PM_{10} (Ambient monitoring)	1st, 2nd, and 3rd trimesters	Probably low
(Kirwa et al. 2019)	Puerto Rico	332,129	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy	Probably high
(Kumar 2012)	Chicago, USA	400,000	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Lamichhane et al. 2018)	South Korea	648	PM_{10} (Modeling)	1st, 2nd, and 3rd trimesters	Probably low
(Laurent et al. 2013)	California, USA (2 counties)	105,092	$PM_{2.5}$, PM_{10} (Ambient monitoring and modeling)	Entire pregnancy	High
(Lavigne et al. 2018)	Ontario, Canada	196,171	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Li et al. 2019)	Ningbo, China	170,008	$PM_{2.5}$, PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Mannes et al. 2005)	Sydney, Australia	138,056	$PM_{2.5}$, PM_{10} (Ambient monitoring)	1st, 2nd, and 3rd trimesters	Probably high
(Medeiros and Gouveia 2005)	São Paulo, Brazil	311,735	PM_{10} (Ambient monitoring)	1st, 2nd, and 3rd trimesters	High
(Merklinger-Gruchala and Kapiszewska 2015)	Krakow, Poland	84,842	PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Morello-Frosch et al. 2010)	California, USA	3,545,177	$PM_{2.5}$, PM_{10}, $PM_{2.5\text{-}10}$ (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Parker and Woodruff 2008)	USA (excluding Alaska and Hawaii)	785,965	$PM_{2.5}$, $PM_{2.5\text{-}10}$ (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Parker et al. 2005)	California, USA	18,247	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy	Probably high
(Pedersen et al. 2013)	12 European countries	74,178	$PM_{2.5}$, PM_{10}, $PM_{2.5\text{-}10}$ (Modeling)	Entire pregnancy	High
(Rahmalia et al. 2012)	Poiters and Nancy, France	888	PM_{10} (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high

Table 2.3 Continued

Reference	Study location	Sample size	Pollutant(s) (exposure assessment method)	Exposure period	Overall ROB rating
(Rhee et al. 2019)	Boston, USA	3,366	$PM_{2.5}$ (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Salam et al. 2005)	California, USA	3,901	PM_{10} (Ambient monitoring)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Santos Vde et al. 2014)	São José dos Campos, Brazil	21,591	PM_{10} (Ambient monitoring)	3rd trimester only	High
(Savitz et al. 2014)	New York, USA	252,967	$PM_{2.5}$ (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Schembari et al. 2015)	Bradford, United Kingdom	9,067	$PM_{2.5}$, PM_{10} (Modeling)	Entire pregnancy, 3rd trimester only	Probably low
(Schwarz et al. 2019)	California, USA	2,768,898	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy	Probably low
(Sellier et al. 2014)	Poitiers and Nancy, France	1,026	PM_{10} (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Stieb et al. 2016)	Canada	2,781,940	$PM_{2.5}$ (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(van den Hooven et al. 2012)	Netherlands	7,772	PM_{10} (Modeling)	Entire pregnancy	Probably low
(Vinikoor-Imler et al. 2014)	North Carolina, USA	322,981	$PM_{2.5}$ (Modeling)	1st, 2nd, and 3rd trimesters	Probably low
(Winckelmans et al. 2015)	Flanders, Belgium	525,635	PM_{10} (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Xiao et al. 2018)	Shanghai, China	132,783	$PM_{2.5}$ (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low
(Xue et al. 2018)	USA	18,317,707	$PM_{2.5}$ (Ambient monitoring)	Entire pregnancy	High
(Yang et al. 2003)	Kaohsiung, Taiwan	13,396	PM_{10} (Ambient monitoring)	1st, 2nd, and 3rd trimesters	Probably high
(Ye et al. 2018)	Taizhou, China	24,246	PM_{10} (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably high
(Yuan et al. 2020)	Shanghai, China	3,692	$PM_{2.5}$ (Modeling)	Entire pregnancy, 1st, 2nd, and 3rd trimesters	Probably low

Table 2.4 Summary of Meta-analysis Results and Quality of Evidence rating Conclusion

Exposure Window	No. of studies	Effect estimate g per 10 $\mu g/m^3$ (95% CI)	I^2 (%)	Quality of evidence rating
$PM_{2.5}$				
1st Trimester	11	-6.50 (-15.07, 2.07)	87%	Very low (downgrades for imprecision and inconsistency)
2nd Trimester	12	-5.69 (-10.58, -0.79)	68%	Low (downgrade for imprecision)
3rd Trimester	12	-10.67 (-20.91, -0.43)	84%	Low (downgrade for inconsistency)
Full Pregnancy	15	-27.55 (-48.45, -6.65)	94%	Low (downgrade for inconsistency)
PM_{10}				
1st Trimester	6	3.22 (-3.13, 9.58)	14%	Very low (downgrade for imprecision and inverse effect)
2nd Trimester	6	-3.37 (-8.22, 1.48)	0%	Low (downgrade for imprecision)
3rd Trimester	7	-6.57 (-10.66, -2.48)	0%	Moderate (no changes)
Full Pregnancy	8	-8.65 (-16.83, -0.48)	84%	Moderate (heterogeneity explained by single study with inverse effect)
$PM_{2.5-10}$				
1st Trimester	3	-2.70 (-3.90; -1.49)	0%	Very low (downgrades for risk of bias and imprecision)
2nd Trimester	3	-2.90 (-10.04; 4.23)	70%	Very low (downgrades for risk of bias, imprecision, inconsistency)
3rd Trimester	4	-4.93 (-10.82; 0.96)	76%	Very low (downgrades for risk of bias, imprecision, inconsistency)
Full Pregnancy	5	-8.81 (-10.32; -7.31)	0%	Low (downgrade for risk of bias)

Table 3.1 Participant Characteristics

	N	%
ETHNICITY		
Hispanic	17	100
EDUCATION		
<12 Years	7	41
12 Years	6	35
>12 Years	1	6
Unknown	3	18
SMOKING		
Never	12	71
Before pregnancy	2	12
Unknown	3	18
HEATING SYSTEM		
Central heating	4	24
Single stoves/heaters	3	18
Electric	7	41
Gas	0	0
Unknown	3	18
COOLING SYSTEM		
Central air conditioning	5	29
Window units	6	35
None	1	6
Unknown	5	29
ENERGY SOURCE FOR COOKING		
Electricity	5	29
Gas	9	53
None/no cooking	1	6
Unknown	2	12
EMPLOYMENT STATUS		
Employed	2	12
Unemployed	15	88

Table 3.2 Percentage Contribution for each PAH in WB, Filter and XAD samples

PAH	WB (%)	Filter (%)	XAD (%)
Acenaphthylene	1.38	2.09	12.64
Acenaphthene	3.26	5.99	26.42
Fluorene	10.71	6.12	28.65
Phenanthrene	69.07	26.26	15.13
Anthracene	1.92	1.94	2.37
Fluoranthene	2.17	5.17	1.87
Pyrene	4.46	6.68	1.44
Benzo[a]anthracene	1.10	7.38	8.85
Chrysene	4.69	5.25	0.97
Benzo[k]fluoranthene	0.62	2.94	0.10
Benzo[a]pyrene	0.12	4.24	0.40
Dibenzo[a,h]anthracene	0.14	2.62	0.30
Benzo[b]fluoranthene	0.23	4.72	0.10
Benzo[e]pyrene	0.04	4.31	0.27
Indeno[1,2,3-cd]pyrene	0.04	4.19	0.37
Benzo[g,h,i]perylene	0.05	9.72	0.13

Table 3.3 Median value (Q_{25}-Q_{75}) for Study Participants from Three 24-h Periods

PAH	Molecular Weight	Wristband (ng/band)	Filter (ng/filter)	XAD (ng/xad)
Acenaphthylene	152.2	6.80 (5.52-7.78)	0.08 (0.06-0.15)	3.78 (2.34-6.43)
Acenaphthene	154.2	12.78 (8.95-19.63)	0.31 (0.25-0.42)	7.91 (6.4-12.81)
Fluorene	166.2	44.22 (35.87-60.37)	0.28 (0.25-0.38)	8.58 (6.3-11.76)
Phenanthrene	178.2	385.85 (253.03-439.70)	0.86 (0.55-1.82)	4.53 (3.46-9.43)
Anthracene	178.2	9.01 (1.02-11.61)	0.08 (0.06-0.16)	0.71 (0.38-1.41)
Fluoranthene	202.3	10.17 (8.53-12.51)	0.20 (0.16-0.32)	0.56 (0.41-0.90)
Pyrene	202.3	20.21 (15.79-24.54)	0.29 (0.20-0.40)	0.43 (0.32-0.73)
Benzo(a)anthracene	228.3	4.76 (3.38-6.62)	0.33 (0.25-0.54)	2.65 (2.15-3.62)
Chrysene	228.3	25.84 (2.56-34.99)	0.23 (0.16-0.43)	0.29 (0.16-0.77)
Benzo(k)fluoranthene	252.3	2.71 (1.78-4.66)	0.15 (0.06-0.24)	0.03 (0.01-0.13)
Benzo(a)pyrene	252.3	0.43 (0.25-0.77)	0.22 (0.14-0.35)	0.12 (0.01-0.34)
Dibenzo(a,h)anthracene	278.4	0.40 (0.22-0.86)	0.08 (0.03-0.22)	0.09 (0.02-0.47)
Benzo(b)fluoranthene	252.3	0.52 (0.04-1.06)	0.23 (0.12-0.39)	0.03 (0.01-0.12)
Benzo(e)pyrene	252.3	0.09 (0-0.53)	0.20 (0.11-0.37)	0.08 (0.01-0.25)
Indeno[1,2,3-cd]pyrene	276.3	0.14 (0.09-0.34)	0.17 (0.11-0.37)	0.11 (0.02-0.27)
Benzo(g,h,i)perylene	276.3	0.15 (0.09-0.52)	0.42 (0.20-0.65)	0.04 (0.00-0.11)

Table 3.4 Correlation table for 16 PAHs analyzed WB, Filters and XADs

PAH	WB and Filter		WB and XAD		XAD and Filter	
	r	p-value	r	p-value	r	p-value
Acenaphthylene	0.26	0.1186	0.02	0.9176	0.28	0.0733
Acenaphthene	-0.22	0.1790	0.21	0.1794	0.25	0.1035
Fluorene	-0.14	0.4170	**0.37***	**0.0167**	-0.01	0.9484
Phenanthrene	0.13	0.4273	0.03	0.8733	0.13	0.4191
Anthracene	0.08	0.6516	-0.13	0.4024	**0.48***	**0.0013**
Fluoranthene	0.07	0.6689	0.22	0.1595	**0.37***	**0.0162**
Pyrene	0.07	0.6881	0.04	0.8064	**0.32***	**0.0402**
Benzo[a]anthracene	-0.14	0.4049	-0.07	0.6381	**0.35***	**0.0218**
Chrysene	0.13	0.4394	**-0.53***	**0.0003**	-0.21	0.1759
Benzo[k]fluoranthene	-0.07	0.6866	-0.05	0.7657	-0.01	0.9759
Benzo[a]pyrene	-0.02	0.9059	-0.07	0.6381	0.19	0.2403
Dibenzo[a,h]anthracene	0.16	0.3280	-0.13	0.4154	-0.17	0.2735
Benzo(b)fluoranthene	0.07	0.6694	-0.05	0.7462	-0.01	0.9567
Benzo(e)pyrene	0.05	0.7729	-0.12	0.4983	0.09	0.5716
Indenol[1,2,3-cd]pyrene	-0.01	0.9703	**-0.34***	**0.0278**	-0.04	0.8081
Benzo(g,h,i)perylene	0.20	0.2238	0.09	0.5873	0.11	0.5047

Bold type* indicates statistical significance ($\alpha<0.05$)

Table 3.5 Mean ± Standard Deviation and Median of $PM_{2.5}$ & BC concentrations for each participant from Three Sampling Periods.

Participant	$PM_{2.5}$ (ug/m^3)	Median	BC (ug/m^3)	Median
1	6.21 ± 4.18	5.04	2.47 ± 0.98	2.38
2	31.89 ± 2.60	31.06	2.00 ± 0.51	1.90
3	60.63 ± 16.41	67.56	1.38 ± 0.47	1.25
4	8.22 ± 1.63	7.48	1.24 ± 0.50	1.26
5	23.95 ± 9.75	26.15	1.31 ± 0.48	1.55
6	25.79 ± 7.96	27.72	2.92 ± 2.32	2.44
7	18.63 ± 7.96	20.66	0.98 ± 0.16	0.91
8	46.92 ± 59.20	23.42	1.15 ± 0.31	1.20
9	92.25 ± 50.91	117.07	1.41 ± 0.66	1.40
10	16.04 ± 1.91	16.04	0.95 ± 0.07	0.95
11	34.98 ± 33.64	20.40	1.77 ± 0.44	1.59
12	7.09 ± 5.58	6.09	0.96 ± 0.20	0.97
13	4.69 ± 2.44	5.63	1.07 ± 0.11	1.12
14	22.15 ± 7.97	25.15	1.50 ± 0.71	1.52
15	9.12 ± 4.15	6.83	1.10 ± 0.21	0.99
16	6.01 ± 4.69	4.70	0.81 ± 0.29	0.84
17	23.62 ± 9.51	18.61	1.22 ± 0.13	1.27
Overall	25.78 ± 17.42	20.40	1.43 ± 0.53	1.26

Table 4.1 Participant Demographics

	Number	*Proportion (%)*
Age		
< 35	2	20
> 35	6	60
Unknown	2	20
Gender		
Male	4	40
Female	6	60
Heating		
Gas	1	1
Central	7	70
Fireplace	0	0
Unknown	2	20
Commute		
Personal vehicle	8	80
Public transportation	0	0
Unknown	2	20

Table 4.2 Number of Positive Detections for each of the 25 PAHs measured in each Matrix (ND = Not Detected, *EPA Criteria Pollutants)

PAH Compounds	Wristbands	Filter
1-Methylnaphthalene	10	10
1,6,7-Trimethylnaphthalene	10	1
2-Methylnaphthalene	10	10
2,6-Dimethylnaphthalene	10	10
Naphthalene*	10	10
Biphenyl	10	10
Phenanthrene*	10	10
Fluorene*	10	2
Dibenzothiophene	9	10
Fluoranthene*	5	10
Pyrene*	3	9
Acenaphthene*	ND	10
1-Methylphenanthrene	1	2
Acenaphthylene*	6	3
Anthracene*	9	2
Chrysene*	3	8
Benzo(a)anthracene*	2	ND
Benzo(a)pyrene*	ND	3
Benzo(b)fluoranthene*	ND	3
Benzo(e)pyrene*	ND	2
Benzo[g,h,i]perylene	ND	3
Benzo(k)fluoranthene*	ND	3
Indenol[1,2,3-c,d]pyrene*	ND	2
Dibenzo(a,h)anthracene*	ND	ND
Perylene	ND	1

Table 4.3 Mean (S.D.) of all PAHs measured by Active and Passive Sampling

PAH Compounds	Molecular weight	LOD (Filter)	Active Sampling (ng/filter)	LOD (WB)	Passive sampling (ng/wristband)
1-Methylnaphthalene	142.2	2.09	1.26 (0.65)	0.34	38.81 (13.76)
1,6,7-Trimethylnaphthalene	170.3	6.19	0.13 (0.40)	0.90	18.16 (4.30)
2-Methylnaphthalene	142.2	2.74	2.09 (0.94)	0.22	60.78 (22.64)
2,6-Dimethylnaphthalene	156.2	4.60	0.34 (0.29)	0.75	20.37 (6.49)
1-Methylphenanthrene	192.3	7.40	0.50 (1.54)	1.34	1.01 (3.19)
Acenaphthylene	152.2	8.12	0.17 (0.30)	2.54	7.72 (7.10)
Anthracene	178.2	9.44	0.52 (1.60)	2.94	29.20 (60.00)
Benz(a)anthracene	228.3	16.11	-	4.83	0.79 (1.23)
Biphenyl	154.2	1.61	0.61 (0.26)	0.29	20.18 (4.83)
Chrysene	228.3	7.01	0.05 (0.06)	1.90	2.45 (3.61)
Dibenzothiophene	184.3	4.63	0.87 (2.35)	0.87	9.70 (5.71)
Fluoranthene	202.3	8.28	1.11 (2.88)	1.70	8.61 (10.03)
Fluorene	166.2	4.38	0.71 (2.08)	1.36	44.05 (19.18)
Naphthalene	128.2	2.16	3.40 (0.46)	0.28	65.70 (23.19)
Phenanthrene	178.2	3.53	9.31 (26.52)	0.39	99.27 (56.98)
Pyrene	202.3	8.41	0.70 (1.80)	1.38	3.56 (5.94)
Perylene	252.3	15.47	0.01 (0.02)	6.93	-
Acenaphthene	154.2	3.82	0.51 (0.39)	0.94	-
Benzo(a)pyrene	252.3	17.05	0.08 (0.13)	6.90	-
Benzo(b)fluoranthene	252.3	18.58	0.24 (0.39)	4.30	-
Benzo(e)pyrene	252.3	11.81	0.03 (0.05)	3.62	-
Benzo(g,h,i)perylene	276.3	18.86	0.10 (0.18)	7.91	-
Benzo(k)fluoranthene	252.3	11.55	0.17 (0.28))	4.09	-
Indenol[1,2,3-c,d]pyrene	276.3	24.41	0.10 (0.21)	11.29	-

Table 4.4 Correlation table for 8 PAHs with >/= 50% Detections in both Filter and Wristbands

PAHs	Wristband PAH and Filter PAH	
	r_s	*p*-value
Naphthalene	-0.08	0.8432
1-Methylnaphthalene	0.73	0.0313*
2-Methylnaphthalene	0.40	0.2912
2,6-Dimethylnaphthalene	0.24	0.5250
Biphenyl	0.20	0.6144
Phenanthrene	0.28	0.4663
Dibenzothiophene	-0.33	0.3845
Fluoranthene	-0.70	0.0423*

APPENDIX B

SUPPLEMENTARY MATERIALS

Table B1. Search Terms

1. Exp Particulate Matter/
2. (pm 2* or pm 10 or pm2* or pm10).ti,ab.
3. (particulate adj2 (matter* or pollut*)).ti,ab.
4. Or/1-3
5. Exp "Birth Weight"/
6. (lbw or (low adj1 (birthweight* or birthweight*))).ti,ab.
7. Or/5-6
8. 4 and 7
9. limit 8 to English language

Table B2. HAWC data collection sheet

SOURCE
Reviewer: (identify yourself)
Publication year:
Authors declared conflicts of interest: • *None declared* • *Declared* If declared, provide details:
Study funding source: • *Government grant* • *Industry funded* • *Nonprofit organization grant* • *Other* Study funding source details:
What are the study objectives?
Site(s) of data collection (city, state, country):
METHODS
Study duration:

Study design: • *Cross-sectional* • *Cohort, prospective* • *Cohort, retrospective* • *Case-control* • *Ecological* • *Other (list details below)* Study design details:
Characteristics of study population: • *Cohort (give description, e.g. NHANES 2004-2006)* • *Sample size of total cohort* • *Sample size (each exposure group)* • *Age (each exposure group)* • *Co-morbidities* • *Other relevant details (list below)* Study subject details:
Exposure period: • *Pregnancy* • *Other (details below)* • *Record when exposure occurred or was measured, in relation to outcome measurement*
Source of exposure data: • *Air pollution monitoring (list specific methods)* • Modeling (list specific methods) • *Questionnaire (list specific proxy used to determine of exposure)* • *Other (specify)*
Total number of exposed groups:
Total number of non-exposed groups:
Number of subjects in each group:
If a power calculation was done, was the sample size of the study sufficient? • *Yes* • *No*
Concentrations of PM measured, and units: Frequency of PM measurements if more than once:
Number of replicate measurements taken
Chemical name: • $PM_{2.5}$ • PM_{10} Chemical name details:
Other chemical details:
Outcomes measured:
Method of fetal growth measurement: • *Weight* • *Length* • *Other (details below)*

Method of fetal growth measurement details:
Gestational age at outcome measurements: • *At birth* • *Other (details below)* Birth outcome measurement details:
Unit of measurement (for weight, etc.): • *Grams* • *Millimeters* • *Other (details below)* Unit of measurement (for weight, etc.) details:
Sex (where outcome measured): • *Males only* • *Females only* • *Males and females* • *Other (details below)*
Number subjects analyzed (for exposure and outcome):
Number of missing participants:
RESULTS
Statistical methods: • Statistical tests employed • Statistic (odd ratio, adjusted odds ratio, beta estimate, etc.) • P-values given • Confidence intervals given • Confounding adjustments in statistical tests
Were known confounders accounted for by study design?
Were known confounders accounted for by analysis?
How were data reported (mean, median, raw data, etc.)?:
Growth measurement data for each group (i.e., outcome):
How growth measurement data were reported (table, figures, etc.):
Summary data for each exposure group
Estimate of effect with confidence interval and p-value
How was precision reported (standard error, CI, etc.):? • *Standard error* • *Standard deviation* • *Confidence intervals* • *Other (details below)* • *Not stated* How precision reported details:
Precision estimates: How precision estimates were reported (table, figure, etc.):
Miscellaneous comments by reviewer regarding data analysis:

Table B3. Instructions for Making Risk of Bias Determinations (adapted from Johnson et al 2014)

Please answer LOW RISK, PROBABLY LOW RISK, PROBABLY HIGH RISK, HIGH RISK or NOT APPLICABLE and provide details/justification.

Note: These criteria for judging risk of bias are for human studies only since we are not evaluating animal studies in this case study. These questions have also been modified from previous applications of the Navigation Guide, with edits intended so that answering "No" to each question aligns with a rating of "High risk of bias", "Probably No" → "Probably high risk of bias", "Probably Yes" → "Probably low risk of bias" and "Yes" → "Low risk of bias."

1. Was the strategy for recruiting participants consistent across study groups?

Criteria for a judgment of LOW risk of bias (i.e., answer: "YES"):

Protocols for recruitment and inclusion/exclusion criteria were applied similarly across study groups, and any one of the following:

- Study participants were recruited from the same population at the same time frame; or
- Study participants were not all recruited from the same population, but proportions of participants from each population in each study group are uniform

Criteria for the judgment of PROBABLY LOW risk of bias (i.e., answer: "Probably Yes"):

There is insufficient information about participant selection to permit a judgment of 'YES', but there is indirect evidence that suggests that participant recruitment and inclusion/exclusion criteria was consistent, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of HIGH risk of bias (i.e., answer: "No"):

Any of the following:

- Protocols for recruitment or inclusion/exclusion criteria were applied differently across study groups; or
- Study participants were recruited at different time frames; or
- Study participants were recruited from different populations and proportions of participants from each population in each study group are not uniform
- Differential loss to follow-up between groups
- Reported refusal/non-response is uniform between groups

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e., answer: "Probably No"):

There is insufficient information about participant selection to permit a judgment of 'NO', but there is indirect evidence that suggests that participant recruitment or inclusion/exclusion criteria was inconsistent, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that participant selection is not an element of study design capable of introducing risk of bias in the study.

2. Was knowledge of the exposure adequately prevented during the study?

Criteria for a judgment of LOW risk of bias (i.e., answer: "YES"):

Any of the following:

- No blinding, but the review authors judge that the outcome and the outcome measurement as well as the exposure and exposure measurement are not likely to be influenced by lack of blinding (such as differential outcome assessment where the outcome is assessed using different measurement or estimation metrics across exposure groups, or differential exposure assessment where exposure is assessed using different measurement or estimation metrics across diagnostic or outcome groups); or
- Blinding of key study personnel was ensured, and it is unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but exposure and outcome assessment was blinded and the non-blinding of others is unlikely to introduce bias.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e., answer: "Probably Yes"):
There is insufficient information about blinding to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of low risk of bias. For example, investigators were effectively blinded to the exposure and/or outcome groups, for example if the exposure was measured by a separate entity and the outcome was obtained from a hospital record.

Criteria for the judgment of HIGH risk of bias (i.e., answer: "No"):
Any of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement or exposure and exposure measurement is likely to be influenced by lack of blinding (i.e., differential outcome or exposure assessment); or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken so as to introduce bias; or
- Some key study personnel were not blinded, and the non-blinding of others was likely to introduce bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e., answer: "Probably No"):
There is insufficient information about blinding to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):
There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

3. Were exposure assessment methods robust?
Note: For this risk of bias domain, we will consider exposure assessment metric for PM. Risk of bias will be assessed for each data set. The risk of bias over the body of evidence will be rated by review authors' review of risk of bias across all datasets (not across all studies). Our rationale for breaking up studies into data sets is that: (1) there is empirical evidence that risk of bias varies depending on which air pollution exposure was measured (i.e., chemical component) and how it was measured (i.e.,

exposure metric) (US Environmental Protection Agency 2013)[1]; (2) there is a need to transparently distinguish among these potential biases within a given study; and (3) co-authors did not identify an empirically-based or otherwise scientifically preferable alternative method to address this aspect of heterogeneity in the data.

The following list of considerations represents a collection of factors proposed by experts in various fields that may potentially influence the internal validity of the exposure assessment in a systematic manner (not those that may randomly affect overall study results). ***These should be interpreted only as suggested considerations and should not be viewed as scoring or a checklist.***

List of Considerations:

Exposure assessment metric:

1) *Modeling*
2) *Monitoring*

For each, overall considerations include:

1) *What is the quality of the metric being used?*
2) *Has the metric been validated for the scenario for which it is being used?*
3) *Is the exposure measured in the study a surrogate for air pollution (i.e., distance to freeway)?*
4) *What was the temporal coverage (i.e., whole developmental period, or a shorter duration)?*
5) *Did the analysis account for prediction uncertainty?*
6) *How was missing data accounted for, and any data imputations incorporated?*
7) *Were sensitivity analyses performed?*

In particular, for exposure assessment models:

1) *Were the input data in the study suspected to systematically under- or over-estimate exposure?*
2) *What type of model was used (geostatistical interpolation, land-use regression, dispersion models, personal air sampling models, hybrid models, etc.)?*
3) *Were meteorological variables incorporated in the model and justified by authors in their selection?*
4) *Were data on land use, topography, traffic, monitoring data, emission rates, etc. incorporated and justified by authors in their selection?*
5) *What was the spatial variation (e.g., distance from source) and geographic/spatial accuracy (county, census tract, individual residence)?*
6) *What was the temporal specificity and variation (accuracy to level of the day, pregnancy trimester, year, etc.?)*
7) *What was the address completeness (e.g., only home address at one point in time, or more complete address history throughout pregnancy/postnatal life and other locations such as work)?*
8) *What was the space-time coverage of the model?*
9) *Were time-activity patterns accounted for?*

[1] US Environmental Protection Agency. 2013. America's children and the environment. 3rd ed. Available from: https://www.epa.gov/ace/americas-children-and-environment-third-edition (last accessed 2019-08-06)

10) *Was mixing height considered as a covariate?*

Criteria for a judgment of LOW risk of bias (i.e., answer: "Yes"):
The reviewers judge that there is low risk of exposure misclassification, i.e.:

- There is high confidence in the accuracy of the exposure assessment methods, such as methods that have been tested for validity and reliability in measuring the targeted exposure; or
- Less established or less direct exposure measurements are validated against well-established or direct methods; or:

A) Monitoring: direct and personal monitoring devices that were used that have been validated for the chemical and scenario for which it was used and there is sufficient evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment; or

B) Modeling: the model accounted for the time-activity pattern specific to each research participant, (e.g. includes more than exposure at the residential address) and included air pollution modeling methods that have been validated or shown to have a high degree of spatial accuracy (e.g. point location), and/or methods that are themselves validated with good agreement compared to person-based air data collection; and there is sufficient evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment.

AND if applicable (e.g. for laboratory measurements), appropriate QA/QC for methods are described and are satisfactory, with at least three of the following items reported, or at least two of the following items reported plus evidence of satisfactory performance in a high-quality inter-laboratory comparison:

- Limit of detection or quantification;
- standards recovery;
- measure of repeatability;
- investigation and prevention of blanks contamination.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e., answer: "Probably No"):
There is insufficient information about the exposure assessment methods to permit a judgment of low risk of bias, but there is indirect evidence that suggests that methods were robust, as described by the criteria for a judgment of low risk of bias. Studies only reporting that the QA/QC items above were satisfactory but not reporting all of the actual numbers may receive a judgment of "probably low risk of bias." Additionally:

A) Monitoring: methodologies which directly assess exposure were used, such as personal exposure instruments, but had not been validated for that purpose, or if such instruments were worn for less than 4 hours per day, or there is some evidence that relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment.

B) Modeling: the model used methods that do not meet the criteria of including time-activity patterns AND spatial accuracy, and so may not have the level of validation compared to person-based air measurement, but include measurements that have evidence of quality, such as good-quality data inputs, validation against area-based air measurement, or other establishments of the accuracy of the data inputs and models, or there is some evidence that

relevant factors from the List of Considerations above would imply minimal risk of bias in the exposure assessment.

Criteria for the judgment of HIGH risk of bias (i.e., answer: "No"):
The reviewers judge that there is high risk of exposure misclassification and any one of the following:

- There is low confidence in the accuracy of the exposure assessment methods; or
- Less established or less direct exposure measurements are not validated and are suspected to introduce bias that impacts the outcome assessment (example: participants are asked to report exposure status retrospectively, subject to recall bias); or
- Uncertain how exposure information was obtained; or:

A) Monitoring: Information from databases or otherwise was gathered that indirectly assessed exposure without considering variables noted in the List of Considerations above, such as spatial variability, land use regression, etc., or there is sufficient evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.
B) Modeling: the air pollution model used has been demonstrated not to pertain to area-based or person-based measures or has otherwise been previously demonstrated to be unable to describe air levels of exposure for assigning exposure in a research situation, or there is sufficient evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e., answer: "Probably No"):
There is insufficient information about the exposure assessment methods to permit a judgment of high risk of bias, but there is indirect evidence that suggests that methods were not robust, as described by the criteria for a judgment of high risk of bias. Additionally:

A) Monitoring: measurement of exposures that may not have been validated for use to study air pollution were used, or there is some evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.
B) Modeling: air pollution models were used that have not been compared to person-based or area-based air measurements and have suspicion of problems estimating true exposure because, for example, they do not have spatial accuracy (e.g. county-level measures), do not pertain to the correct time frame, are based on limited data, or differ in methodology between cases and controls in a study, or there is some evidence that relevant factors from the List of Considerations above would imply risk of bias in the exposure assessment.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):
There is evidence that exposure assessment methods are not capable of introducing risk of bias in the study.

4. Was confounding adequately addressed?
Prior to the evaluation of studies, coauthors collectively developed the following list of confounders as well as the rationale for inclusion. The ***top 5 confounders*** *are considered "important" based on strength of association with the outcome as well as exposure variable.*

1. Social class. *This is measured differently from study to study, such as by education and income. Note that variables like marital status and insurance can even reflect aspects of social*

class. Sometimes social class is accounted for by individual-level measurements, and other times by group-level measurements (such as census variables). Rationale: Where people live (neighborhood) is strongly influenced by social class. Additionally, we know that urban residence is associated with higher air pollution (Williams et al. 2007)[2], and the airborne pollutants that someone is exposed to are influenced by neighborhood, so social class is related to neighborhood.

2. Race/ethnicity. *Race and ethnicity are known to influence studies of air pollution and birthweight (Gray et al. 2014).*

3. Tobacco use or Environmental tobacco smoke exposure. *Tobacco use is known to be a form of air pollution and can indirectly influence studies of air pollution and birthweight.*

4. Maternal (and/or paternal) age. *Maternal age is related to social class, because very young mothers tend to be of lower social class and older mothers tend to be of higher social class so maternal age may be correlated with air pollutant levels (i.e., younger women may be of lower social class, and lower social classes may be exposed to higher levels of air pollutants). Note that while paternal age is likely also a risk factor, it is often not adjusted for because this variable has a high degree of missing-ness. Because maternal age and paternal age are highly related, it is often thought that adjusting for maternal age is sufficient, and so may not be necessary to include both.*

5. Season of conception/birth (calendar time of conception/birth). *Air pollutant concentrations also vary by season due to sunlight and other factors. Air pollutants will only vary by season if there is temporal refinement in the air pollutant measure, such as monthly or trimester-long values. A study with annual averages or air pollutant levels, or static levels such as distance to a road, will NOT show a correlation structure between season and air pollutants, and so season will not confound in this type of study.*

6. Comorbidity. Comorbidity is known to influence studies of air pollution and birthweight.

7. Alcohol use. Alcohol use is known to decrease birthweight and be associated with social class. Those who consume alcohol while pregnant are more likely to develop pregnancy complications and have a lower birthweight.

Criteria for a judgment of 'YES' (i.e. low risk of bias):
The study accounted for (i.e., matched, stratified, multivariate analysis or otherwise statistically controlled for) all 5 important potential confounders, or reported that potential confounders were evaluated and omitted because inclusion did not substantially affect the results. The determination of specific confounders may be informed by the data, including the studies included in the review.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

[2] *Williams BL, Pennock-Roman M, Suen HK, Magsumbol MS, Ozdenerol E. 2007. Assessing the impact of the local environment on birth outcomes: A case for hlm. J Expo Sci Environ Epidemiol 17:445-457.*

The study accounted for most at least 3 of the 5 of the important potential confounders AND this lack of accounting is not expected to introduce substantial bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):
The study accounted for two or less of our listed potential confounders.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):
The study accounted for at least two of the important potential confounders but included others potential confounders AND this lack of accounting may have introduced substantial bias.

5. Were incomplete outcome data adequately addressed?
Criteria for a judgment of 'YES' (i.e. low risk of bias):
Participants were followed long enough to obtain outcome measurements;
OR any one of the following:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to introduce bias); or
- Attrition or missing outcome data balanced in numbers across exposure groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a relevant impact on the intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a relevant impact on the observed effect size; or
- Missing data have been imputed using appropriate methods

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):
There is insufficient information about incomplete outcome data to permit a judgment of low risk of bias, but there is indirect evidence that suggests incomplete outcome data was adequately addressed, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):
Participants were not followed long enough to obtain outcome measurements; OR any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across exposure groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- Potentially inappropriate application of imputation.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of high risk of bias, but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):
There is evidence that incomplete outcome data is not capable of introducing risk of bias in the study.

6. Does the study report appear to have been comprehensive in its outcome reporting?
Criteria for a judgment of 'YES' (i.e. low risk of bias):
All the study's pre-specified (primary and secondary) outcomes outlined in the protocol, methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):
There is insufficient information about selective outcome reporting to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):
Any one of the following:

- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, methods, abstract, and/or introduction) have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect); or
- One or more outcomes of interest are reported incompletely

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):
There is insufficient information about selective outcome reporting to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):
There is evidence that selective outcome reporting is not capable of introducing risk of bias in the study.

7. Is the study free of support from any company, study author, or other entity having a financial interest in any of the exposures studied?
Criteria for a judgment of 'YES' (i.e. low risk of bias):
The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;

- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):
There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of 'NO' (i.e. high risk of bias):
The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Chemicals, equipment or testing provided at no cost;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;
- Company limited author access to the data;
- Company was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):
There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):
There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.

8. Did the study appear to be free of other problems that could put it at a risk of bias?
Criteria for a judgment of 'YES' (i.e. low risk of bias):
The study appears to be free of other sources of bias.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):
There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence that suggests the study was free of other threats to validity.

Criteria for the judgment of 'NO' (i.e. high risk of bias):
There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used; or
- Stopped early due to some data-dependent process (including a formal-stopping rule); or
- The conduct of the study is affected by interim results (e.g. recruiting additional participants from a subgroup showing greater or lesser effect); or
- Has been claimed to have been fraudulent; or
- Had some other problem

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):
There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of high risk of bias.

Table B4. Instructions for Grading the Quality and Strength of Evidence (adapted from Johnson et al 2014)

A. Grading Quality

Each of the categories to consider in downgrading or upgrading the evidence is described in detail below. Please record your results on the chart at the end of each category, including a brief explanation for your ratings.

Downgrade Categories

Category 1. Quality of Study Limitations (Risk of Bias)
Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

The evidence from studies can be rated down if most of the relevant evidence comes from studies that suffer from a high risk of bias. Risk of bias is rated by outcome across studies. Study limitations for each outcome for individual studies and across studies are summarized in the heat maps.

GRADE outlines the following principles for moving from risk of bias in individual studies to rating quality of evidence across studies.

1. In deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies is warranted.[3]
2. This judicious consideration requires evaluating the extent to which each study contributes toward the estimate of magnitude of effect. The contribution that each study makes will usually reflect study sample size and number of outcome events. Larger studies with many events will contribute more, much larger studies with many more events will contribute much more.
3. One should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.
4. The risk of bias should be considered in the context of other limitations. If, for instance, reviewers find themselves in a close-call situation with respect to two quality issues (risk

[3] Note: Limitations to GRADE's risk of bias assessments as stated by GRADE: "First, empirical evidence supporting the criteria is limited. Attempts to show systematic difference between studies that meet and do not meet specific criteria have shown inconsistent results. Second, the relative weight one should put on the criteria remains uncertain. The GRADE approach is less comprehensive than many systems, emphasizing simplicity and parsimony over completeness. GRADE's approach does not provide a quantitative rating of risk of bias. Although such a rating has advantages, we share with the Cochrane Collaboration methodologists a reluctance to provide a risk of bias score that, by its nature, must make questionable assumptions about the relative extent of bias associated with individual items and fails to consider the context of the individual items."

of bias and, say, precision), GRADE suggests rating down for at least one of the two.

5. Notwithstanding the first four principles, reviewers will face close-call situations. You should acknowledge that you are in such a situation, make it explicit why you think this is the case, and make the reasons for your ultimate judgment apparent.

Rating for Risk of Bias (Study Limitations) 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 2. Indirectness of Evidence

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

Quality of evidence (your confidence in estimates of effect) may decrease when substantial differences exist between the population, the exposure, or the outcomes measured in research studies under consideration in the review.

Evidence is direct when it directly compares the exposures in which we are interested when applied to the populations in which we are interested and measures outcomes important to the study question (in GRADE the outcomes must be important to patients).

Based on GRADE (GH Guyatt et al. 2011a), evidence can be indirect in one of three ways.[4]

1. The population studied differs from the population of interest (the term applicability is often used for this form of indirectness). GRADE states that in general, one should not rate down for population differences unless one has compelling reason to think that the biology in the

[4] GRADE includes a fourth type of indirectness that occurs when there are no direct (i.e., head-to-head) comparisons between two or more interventions of interest. This criterion is not relevant to our study question; it could be relevant to future case studies.

population of interest is so different than the population tested that the magnitude of effect will differ substantially. According to GRADE, most often, this will not be the case.

2. The intervention (exposure) tested may differ from the exposure of interest, i.e., a difference in the chemical, route and/or dose. Decisions regarding indirectness of populations and exposure depend on an understanding of whether biological or social factors are sufficiently different that one might expect substantial differences in the magnitude of effect. GRADE also states, "As with all other aspects of rating quality of evidence, there is a continuum of similarity of the intervention that will require judgment. It is rare, and usually unnecessary, for the intended populations and interventions to be identical to those in the studies, and we should only rate down if the differences are considered sufficient to make a difference in outcome likely."
3. Outcomes may differ from those of primary interest; for instance, surrogate outcomes that are not themselves important, but measured in the presumption that changes in the surrogate reflect changes in an important outcome. The difference between desired and measured outcomes may relate to time frame. When there is a discrepancy between the time frame of measurement and that of interest, whether to rate down by one or two levels will depend on the magnitude of the discrepancy. Another source of indirectness related to measurement of outcomes is the use of substitute or surrogate endpoints in place of the exposed population's important outcome of interest. In general, the use of a surrogate outcome requires rating down the quality of evidence by one, or even two, levels. Consideration of the biology, mechanism, and natural history of the disease can be helpful in making a decision about indirectness. Surrogates that are closer in the putative causal pathway to the adverse outcomes warrant rating down by only one level for indirectness. GRADE states that rarely, surrogates are sufficiently well established that one should choose not to rate down quality of evidence for indirectness. In general, evidence based on surrogate outcomes should usually trigger rating down, whereas the other types of indirectness will require a more considered judgment.

Rating for Indirectness 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 3. Inconsistency of Evidence

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

According to Cochrane, "when studies yield widely differing estimates of effect (heterogeneity or variability in results) investigators should look for robust explanations for that heterogeneity. …When

heterogeneity exists and affects the interpretation of results, but authors fail to identify a plausible explanation, the quality of the evidence decreases."
Based on GRADE (GH Guyatt et al. 2011b), **a body of evidence is not rated up in quality if studies yield consistent results**, **but may be rated down in quality if inconsistent**. Their stated reason is that a consistent bias will lead to consistent, spurious findings.
GRADE suggests rating down the quality of evidence if large inconsistency (heterogeneity) in study results remains after exploration of a priori hypotheses that might explain heterogeneity. Judgment of the extent of heterogeneity is based on similarity of point estimates, extent of overlap of confidence intervals, and statistical criteria. GRADE's recommendations refer to inconsistencies in effect size, specifically to relative measures (risk ratios and hazard ratios or odds ratios), not absolute measures.

Based on GRADE, reviewers should consider rating down for inconsistency when:
1. Point estimates vary widely across studies;
2. Confidence intervals (CIs) show minimal or no overlap;
3. The statistical test for heterogeneity-which tests the null hypothesis that all studies in a meta-analysis have the same underlying magnitude of effect- shows a low P-value;
4. The I^2 -which quantifies the proportion of the variation in point estimates due to among-study differences-is large. (I.e., the I^2 index quantifies the degree of heterogeneity in a meta-analysis).

GRADE states that inconsistency is important **only when it reduces confidence in results in relation to a particular decision**. Even when inconsistency is large, it may not reduce confidence in results regarding a particular decision. For example, studies that are inconsistent related to the magnitude of a beneficial or harmful effect (but are in the same direction) would not be rated down; in instances when results are inconsistent as to whether there is a benefit or harm of treatment, GRADE would rate down the quality of evidence as a result of variability in results, because the meaning of the inconsistency is so relevant to the decision to treat or not to treat.

Rating for Inconsistency 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 4. Imprecision of Evidence
Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

Cochrane states that when studies have few participants and few events, and thus have wide confidence intervals (CIs), authors can lower their rating of the quality of evidence. These ratings of

precision are made as judgments by review authors. The ratings are made by looking across studies, or, if available, on the results of a meta-analysis.

GRADE defines evidence quality differently for systematic reviews and guidelines. For systematic reviews, quality refers to confidence in the estimates of effect. For guidelines, quality refers to the extent to which confidence in the effect estimate is adequate to support a particular decision (G Guyatt et al. 2011). For the purpose of step 3 of Navigation Guide, we will use the systematic review definition, because the decision phase does not occur until step 4 when recommendations for prevention are made. Thus, when reviewing the data for imprecision, evaluate your confidence in the estimate of the effect.

According to GRADE, to a large extent, CIs inform the impact of random error on evidence quality. Thus, when considering imprecision, the issue is whether the CI around the estimate of exposure effect is sufficiently narrow. If it is not, GRADE rates down the evidence quality by one level (for instance, from high to moderate). If the CI is very wide, GRADE might rate down by two levels.

Rating for Imprecision 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Category 5. Publication Bias
Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

GRADE (GH Guyatt et al. 2011c) and Cochrane (Higgins and Green 2011) assess publication bias in a similar manner. Whereas "selective outcome reporting" is assessed for each study included in the review as part of the risk of bias assessment, "publication bias" is assessed on the body of evidence. GRADE states that "when an entire study remains unreported and the results relate to the size of the effect- publication bias- one can assess the likelihood of publication bias only by looking at a group of studies."

Cochrane's definition of publication bias is "the *publication* or *non-publication* of research findings depending on the nature and direction of the results." Cochrane and GRADE are primarily concerned with *overestimates* of true effects of treatments or pharmaceuticals, especially related to "small studies effects", i.e., the tendency for estimates of an intervention to be more beneficial in smaller

studies. There is empirical evidence in the clinical sciences that publication and other reporting biases result in over estimating the effects of interventions (Higgins and Green 2011).

In contrast, in environmental health, we are primarily concerned with *underestimating* the true effects of a chemical exposure, since in many cases population wide exposure has already occurred. We are also concerned that studies finding no association are less likely to be published because journals are less likely to publish "negative" findings.
Applying this inverted concern to GRADE's assessment for publication bias, leads to these considerations when rating publication bias:

- Early *negative* studies, particularly if small in size, are suspect. (GRADE is concerned with early *positive* studies).
- Authors of systematic reviews should suspect publication bias when studies are uniformly small, particularly when sponsored by the industry. (Same as GRADE)
- Empirical examination of patterns of results (e.g., funnel plots) may suggest publication bias but should be interpreted with caution. (Same as GRADE)
- More compelling than any of these theoretical exercises is authors' success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. (Same as GRADE)
- Comprehensive searches of the literature including unpublished studies, i.e., the grey literature, and a search for research in other languages are important to addressing publication bias. Note that Cochrane also states "comprehensive searching is not sufficient to prevent some substantial potential biases."

Rating for Publication Bias 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels		**Rationale for your judgment**
Human		

Upgrade Categories
GRADE states that the circumstances for upgrading likely occur infrequently and are primarily relevant to observational and other non-randomized studies. Although it is possible to rate up results from randomized controlled trials, GRADE has yet to find a compelling circumstance for doing so (GH Guyatt et al. 2011d).

GRADE specifies 3 categories for increasing the quality of evidence (GH Guyatt et al. 2011d)

Category 6. Large Magnitude of Effect
Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Modeling studies suggests that confounding (from non-random allocation) alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5), and very unlikely to explain associations with an RR greater than 5 (or less than 0.2). Thus, these are the definitions of "large magnitude of effect" used by GRADE to upgrade 1 or 2 levels, respectively. Also, GRADE is more likely to rate up if the effect is rapid and out of keeping with prior trajectory; usually supported by indirect evidence. GRADE presents empirical evidence to support these conclusions, and states that "although further research is warranted, both modeling and empirical work suggest the size of bias from confounding is unpredictable in direction but bounded in size. Hence, the GRADE group has previously suggested guidelines for rating quality of evidence up by one category (typically from low to moderate) for associations greater than 2, and up by two categories for associations greater than 5."

Applying the GRADE definitions of large magnitude of effect i.e., RR greater than 2 or 5 is problematic in environmental health because for dichotomous outcomes RR is a function of the exposure comparator; these definitions also are not applicable to results from continuous variables. At present, we do not have an empirically defined "large magnitude of effect." Therefore, for the purpose of this case study, co-authors should assess whether the results indicate a large magnitude of effect using their expert judgment of "large effects" in environmental health and state their definition for discussion by the group.

Rating for Large Magnitude of Effect 0 no change +1 increase quality 1 level +2 increase quality 2 levels		**Rationale for your judgment**
Human		

Category 7. Dose-response
Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Possible considerations include consistent dose response gradients in one or multiple studies, and/or dose response across studies, depending on the overall relevance to the body of evidence.

Rating for Dose-Response	**Rationale for your judgment**

0 no change +1 increase quality 1 level +2 increase quality 2 levels		
Human		

Category 8. Confounding Minimizes Effect
Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

All plausible residual confounders or biases would reduce a demonstrated effect, or suggest a spurious effect when results show no effect. GRADE provides the following example of grading up evidence when observational studies have failed to demonstrate an association.

Rating for Confounding Minimizes Effect 0 no change +1 increase quality 1 level +2 increase quality 2 levels		**Rationale for your judgment**
Human		

The results of the reviewers' ratings by population will be compiled and discussed leading to a final decision on overall quality of human evidence. The rationale for the decision will be fully documented.

1. Final decision on overall quality of human evidence:
(Example: Moderate quality is upgraded 1 step to high for XYZ reason(s))
---- High
---- Moderate
---- Low
---- Very

B. Rate the Strength of Evidence

The evidence quality ratings will be translated into strength of evidence for each population based on a combination of four criteria: (1) Quality of body of evidence; (2) Direction of effect; (3) Confidence in effect; and (4) Other compelling attributes of the data that may influence certainty. The strength of evidence ratings is summarized in the table below, where their meaning is further defined.

Sufficient evidence of toxicity	A positive relationship is observed between exposure and outcome where chance, bias, and confounding can be ruled out with reasonable confidence. The available evidence includes results from one or more well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies[5].
Limited Evidence of Toxicity	A positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence. Confidence in the relationship is constrained by such factors as: the number, size, or quality of individual studies, or inconsistency of findings across individual studies[6]. As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.
Inadequate Evidence of Toxicity	The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an assessment of effects.
Evidence of Lack of Toxicity	No relationship is observed between exposure and outcome, and chance, bias and confounding can be ruled out with reasonable confidence. The available evidence includes consistent results from more than one well-designed, well-conducted study at the full range of exposure levels that humans are known to encounter, and the conclusion is unlikely to be strongly affected by the results of future studies[7]. The conclusion is limited to the age at exposure and/or other conditions and levels of exposure studied.

[5] The Navigation Guide rates the quality and strength of evidence of human and non-human evidence streams separately as "sufficient", "limited", "inadequate" or "evidence of lack of toxicity" and then these two ratings are combined to produce one of

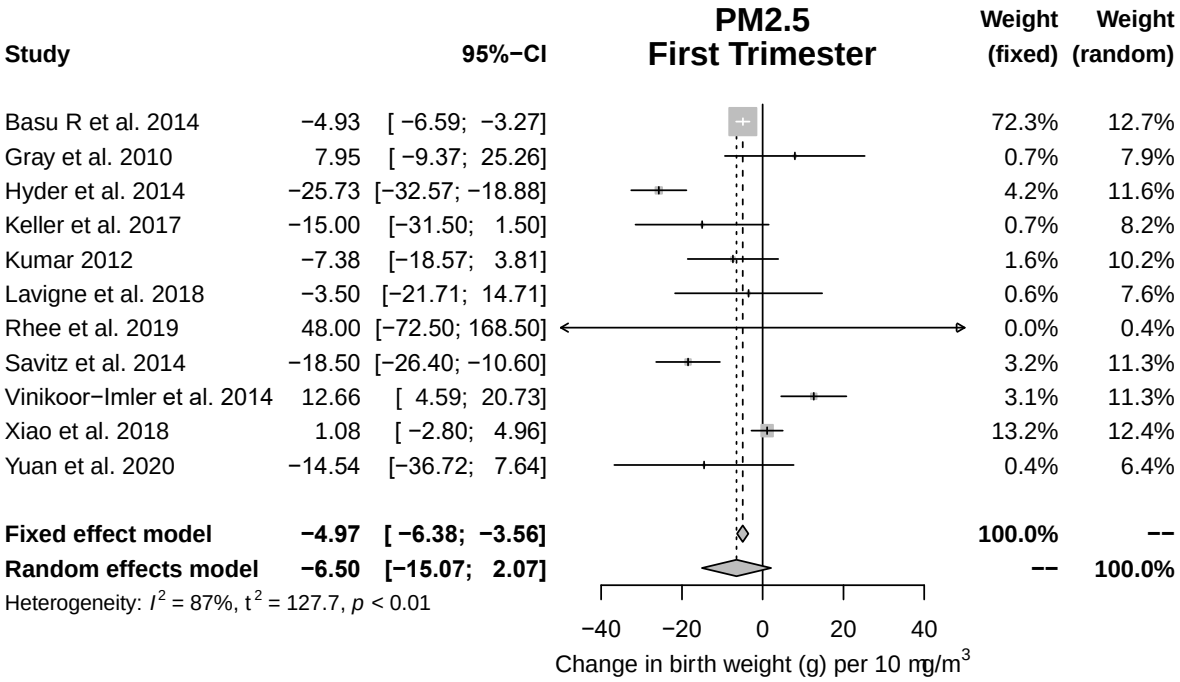
five possible statements about the overall strength of the evidence of a chemical's reproductive/developmental toxicity. The methodology is adapted from the criteria used by the International Agency for Research on Cancer (IARC) to categorize the carcinogenicity of substances International Agency for Research on Cancer. 2006. Iarc monographs on the evaluation of carcinogenic risks to humans: Preamble (amended january 2006). Lyon, France. except as noted.

[6] Language for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit. https://www.uspreventiveservicestaskforce.org/Page/Name/update-on-methods-estimating-certainty-and-magnitude-of-net-benefit (last accessed 2019-08-06)

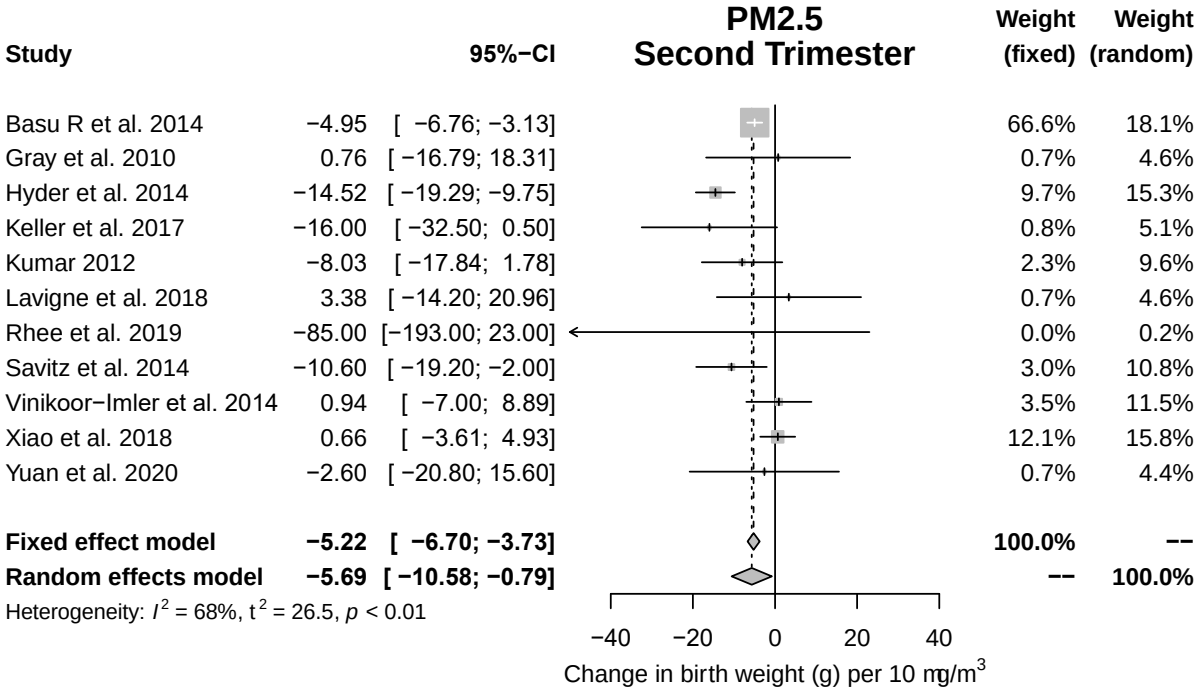
[7] Language for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, Force USPST. 2007. Update on the methods of the U.S. Preventive services task force: Estimating certainty and magnitude of net benefit. Ann Intern Med 147:871-875.

Figure B1. Results of Meta-analysis for studies evaluating $PM_{2.5}$ exposure; overall Quality of Evidence rating.

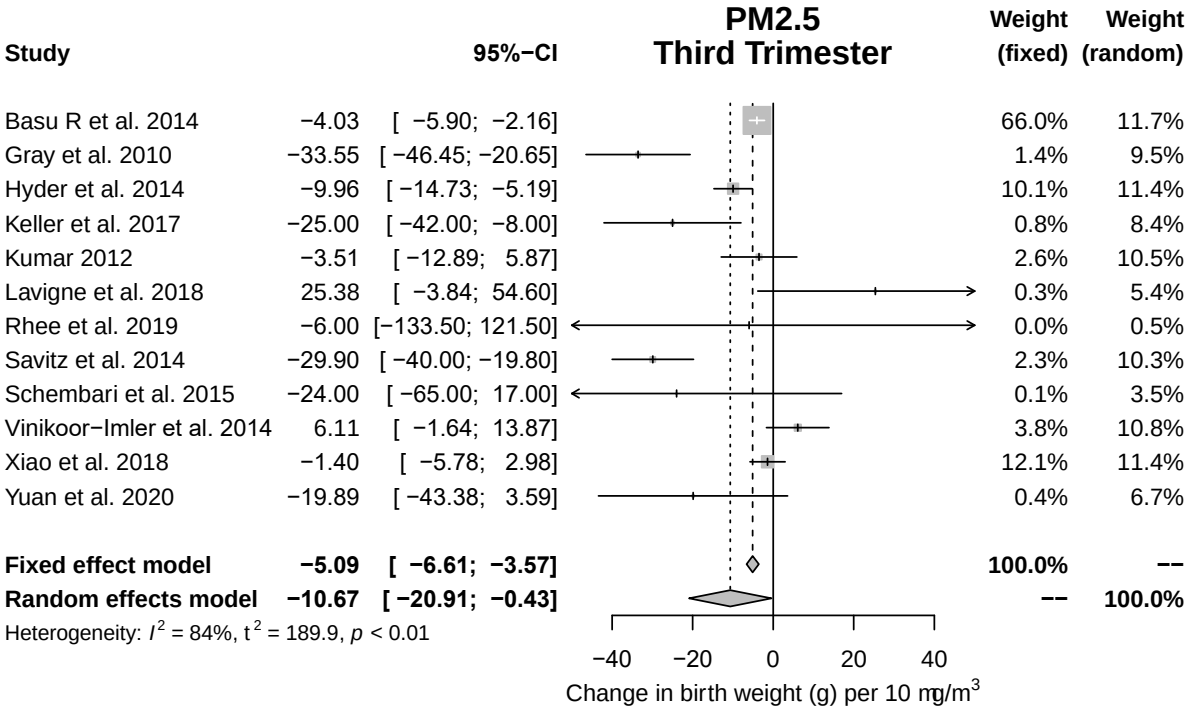
B1.1 $PM_{2.5}$ (First Trimester, n=11 studies rated as "Low" or "Probably low" risk of bias); Very low (downgrades for imprecision and inconsistency)



B1.2 $PM_{2.5}$ (Second Trimester, n=12 studies rated as "Low" or "Probably low" risk of bias); Low (much heterogeneity explained by single study with larger effect, downgrade for imprecision because summary estimate no longer statistically significant if this study is dropped)



B1.3 $PM_{2.5}$ (Third Trimester, n=12 studies rated as "Low" or "Probably low" risk of bias); Low (downgrade for inconsistency)



B1.4 $PM_{2.5}$ (Entire Pregnancy, n=15 studies rated as "Low" or "Probably low" risk of bias); Low (downgrade for inconsistency)

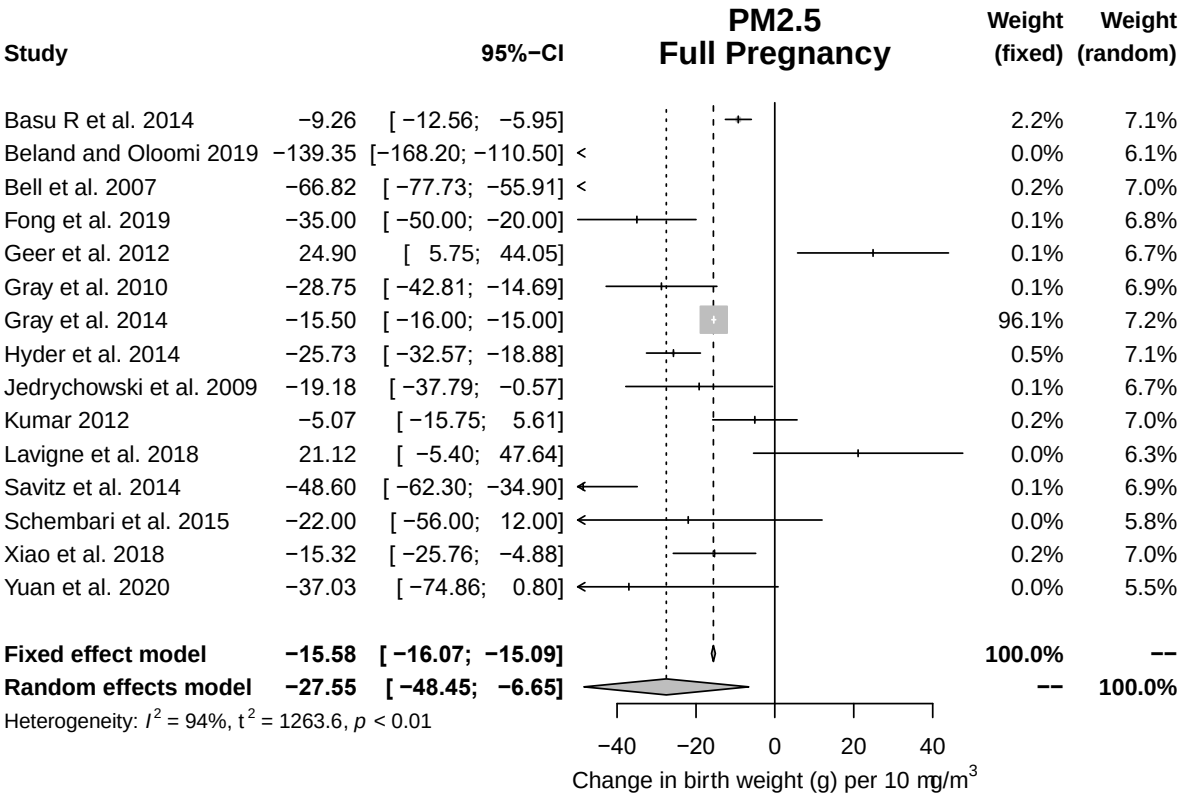
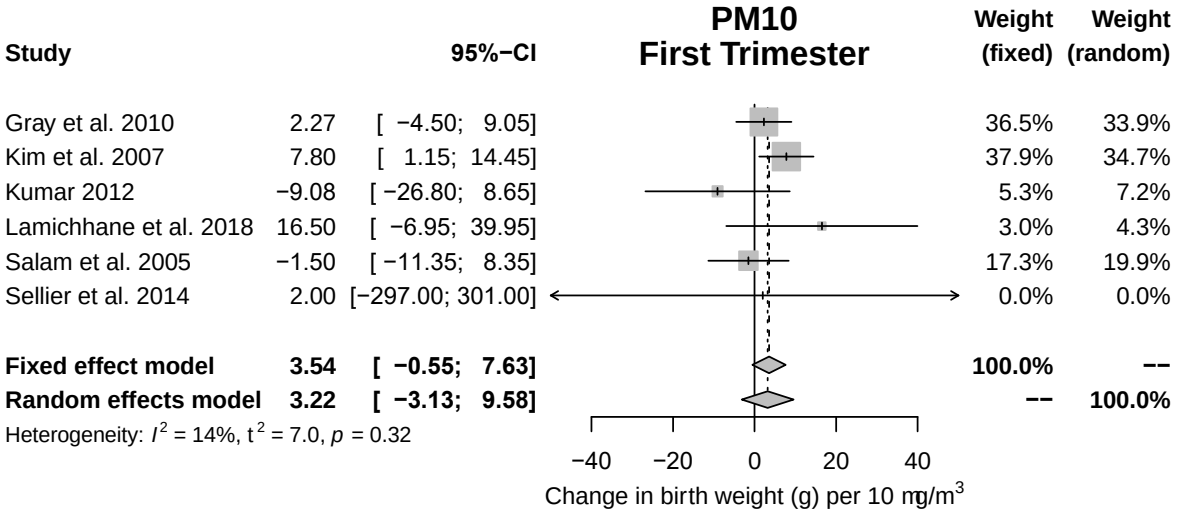
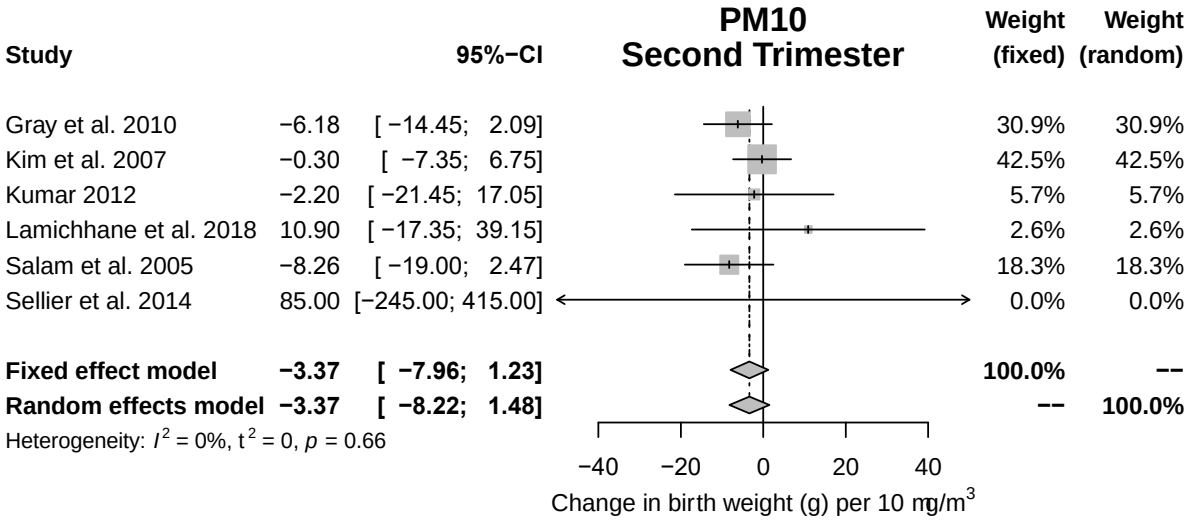


Figure B2. Results of Meta-analysis for studies evaluating PM_{10} exposure; overall Quality of Evidence rating.

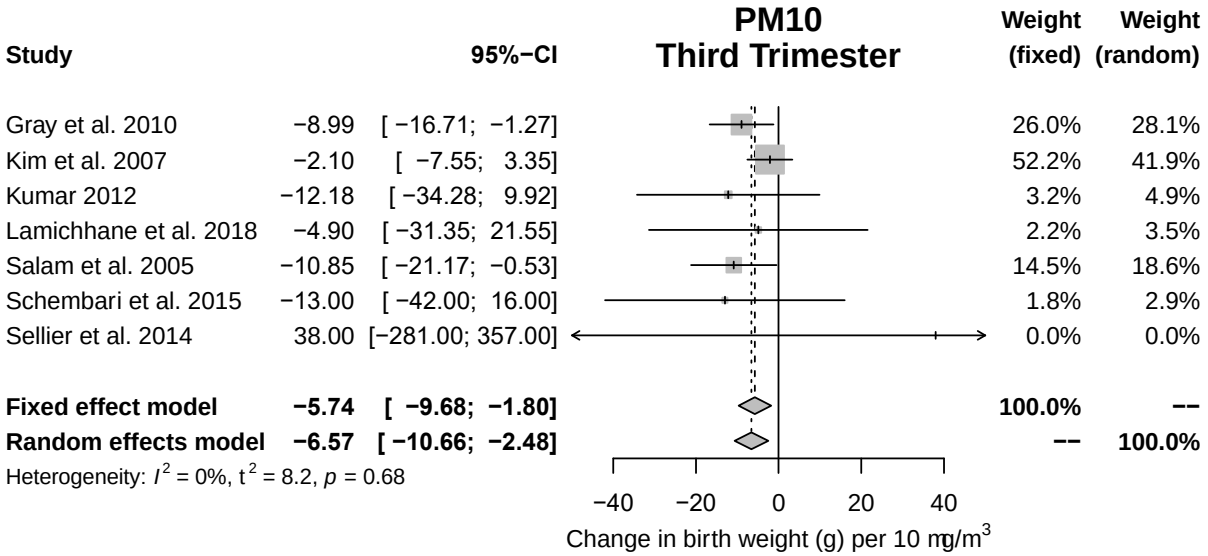
B2.1 PM_{10} (First Trimester, n=6 studies rated as "Low" or "Probably low" risk of bias); Very low (downgrades for imprecision and inverse effect)



B2.2 PM_{10} (Second Trimester, n=6 studies rated as "Low" or "Probably low" risk of bias); Low (downgrade for imprecision)



B2.3 PM_{10} (Third Trimester, n=7 studies rated as "Low" or "Probably low" risk of bias); Moderate (no changes)



B2.4 PM_{10} (Entire Pregnancy, n=8 studies rated as "Low" or "Probably low" risk of bias); Moderate heterogeneity (explained by single study with inverse effect)

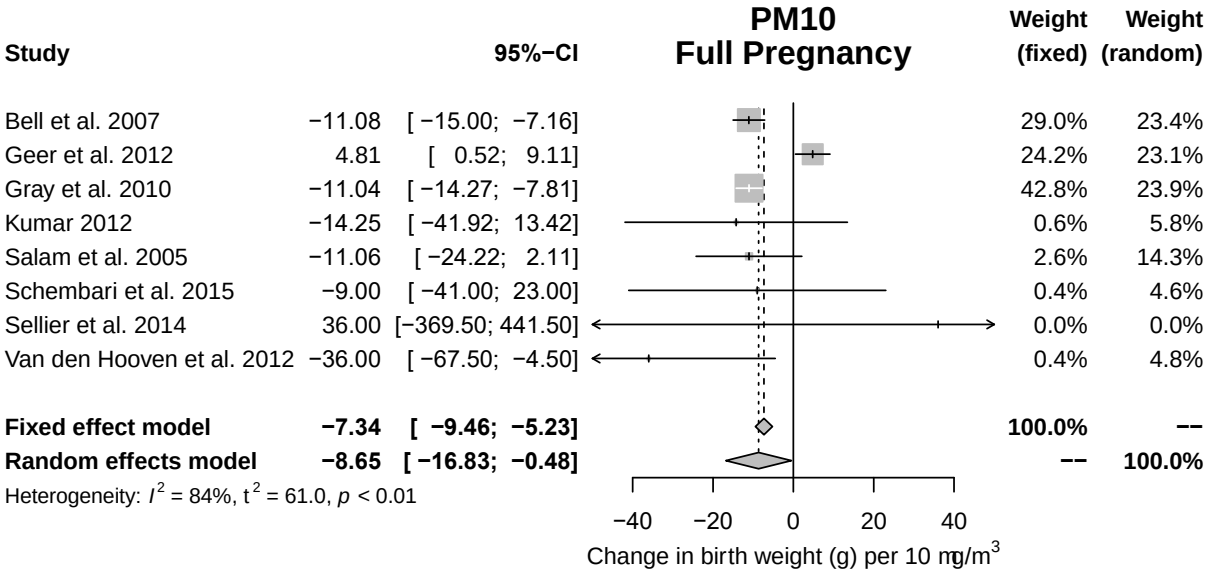
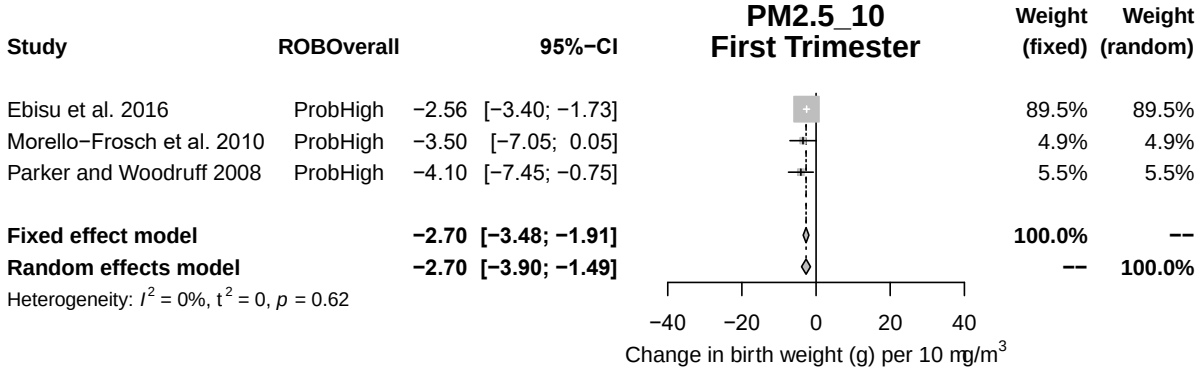
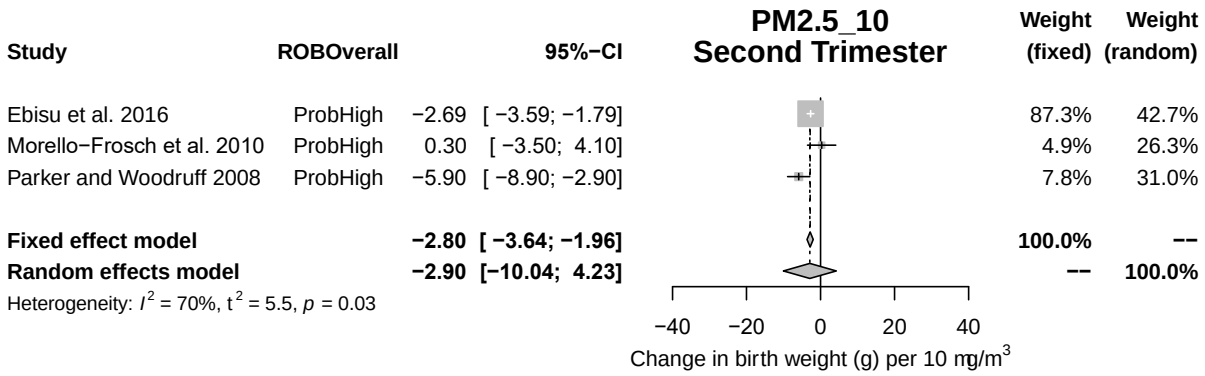


Figure B3. Results of Meta-analysis for studies evaluating $PM_{2.5-10}$ exposure; overall Quality of Evidence rating.

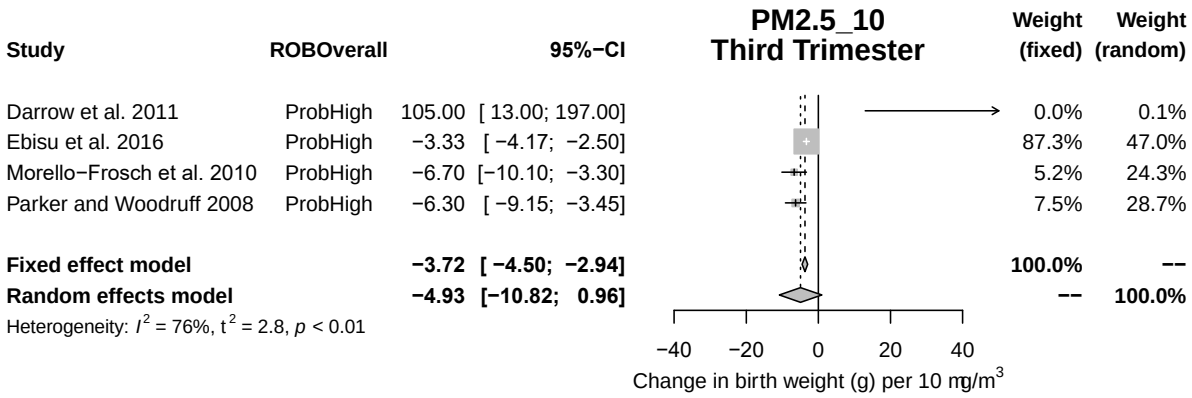
B3.1 $PM_{2.5-10}$ (First Trimester, n=3 studies rated as "Probably high" risk of bias); Very low (downgrades for risk of bias and imprecision)



B3.2 $PM_{2.5-10}$ (Second Trimester, n=3 studies rated as "Probably high" risk of bias); Very low (downgrades for risk of bias, imprecision, inconsistency)



B3.3 $PM_{2.5-10}$ (Third Trimester, n=4 studies rated as "Probably high" risk of bias); Very low (downgrades for risk of bias, imprecision, inconsistency)



B3.4 $PM_{2.5-10}$ (Entire Pregnancy, n=5 studies rated as "High" or "Probably high" risk of bias); Low (downgrade for risk of bias)

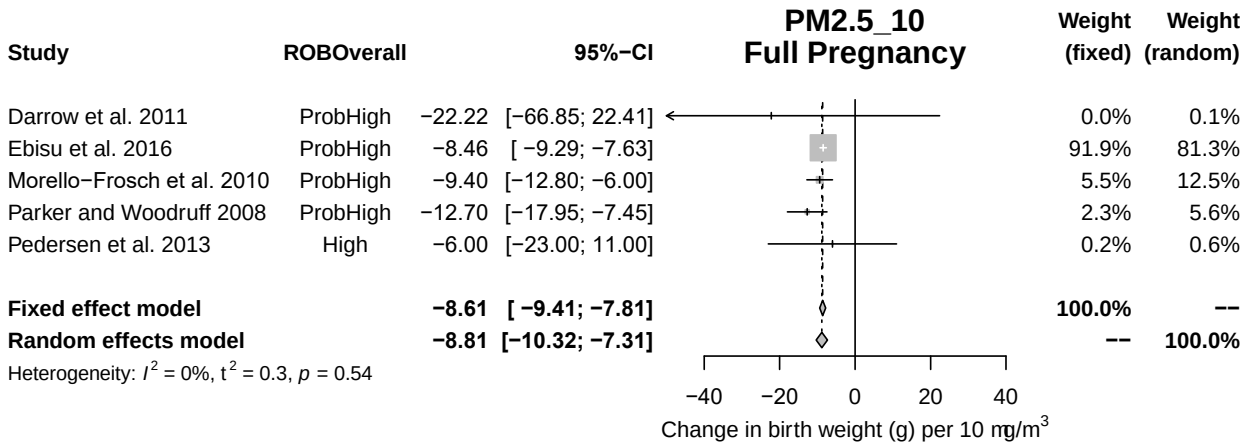
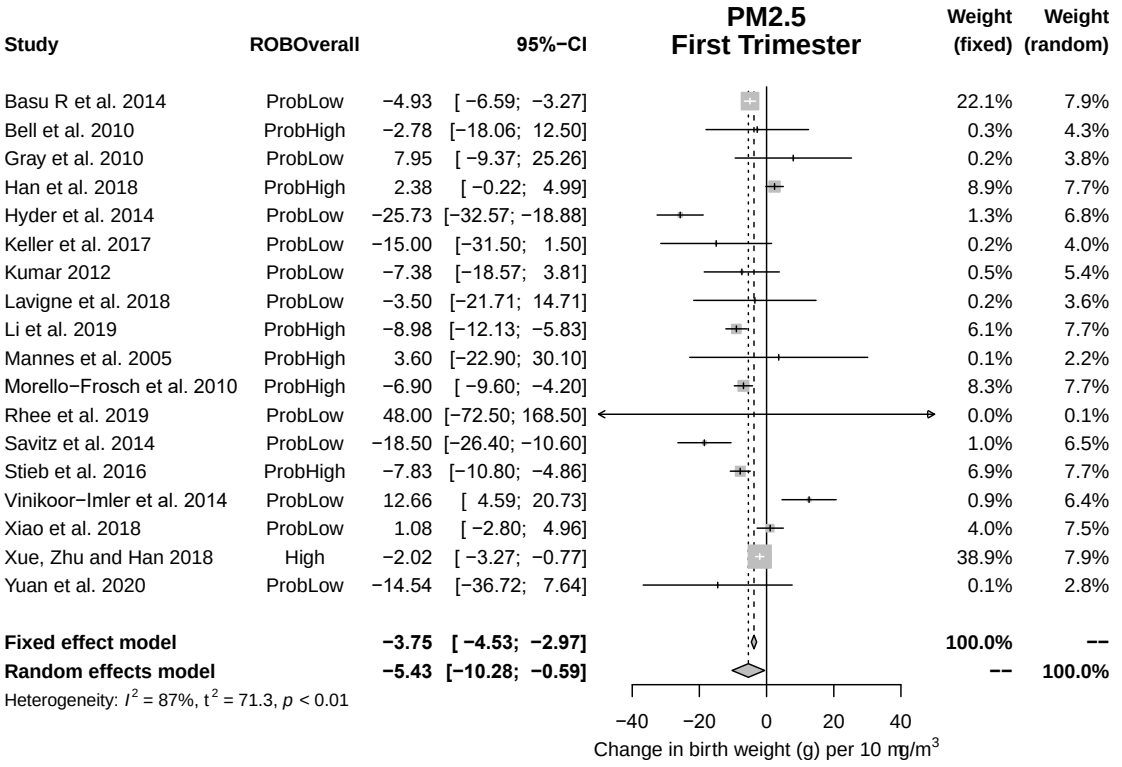
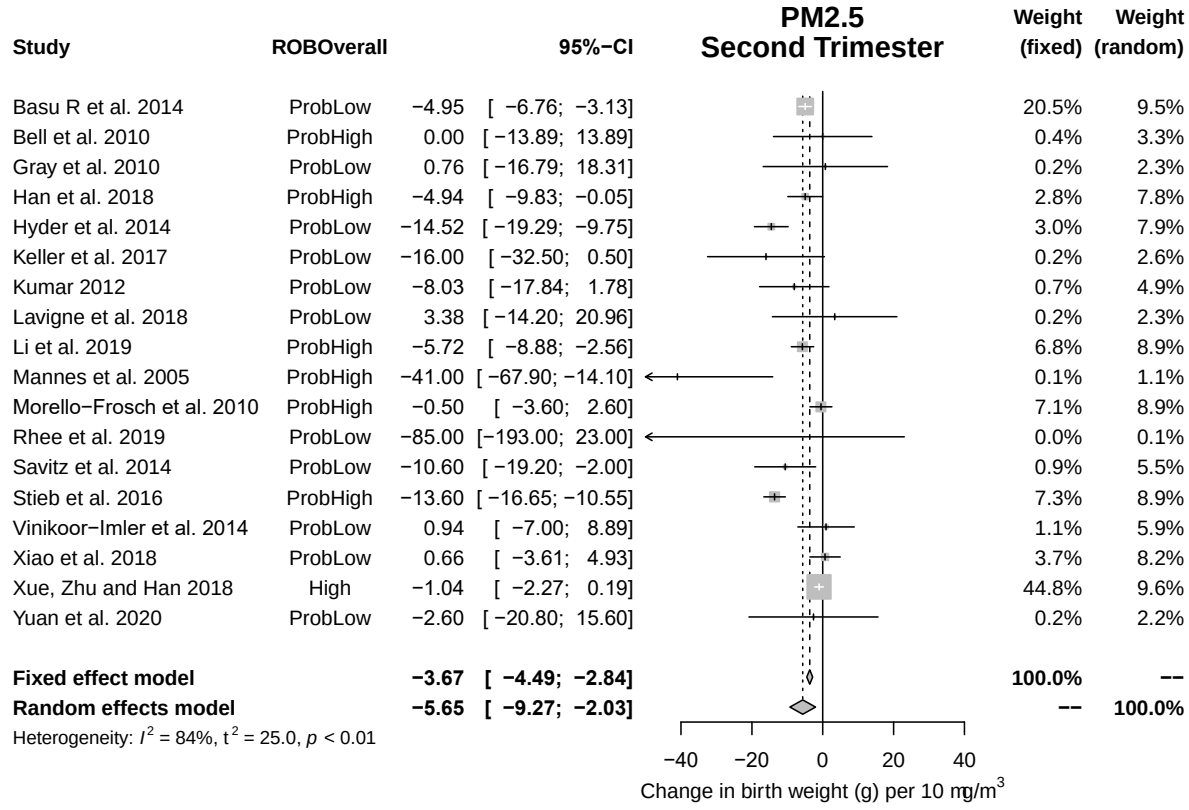


Figure B4. Results of Meta-analysis for studies evaluating $PM_{2.5}$ exposure (all studies included despite risk of bias rating)

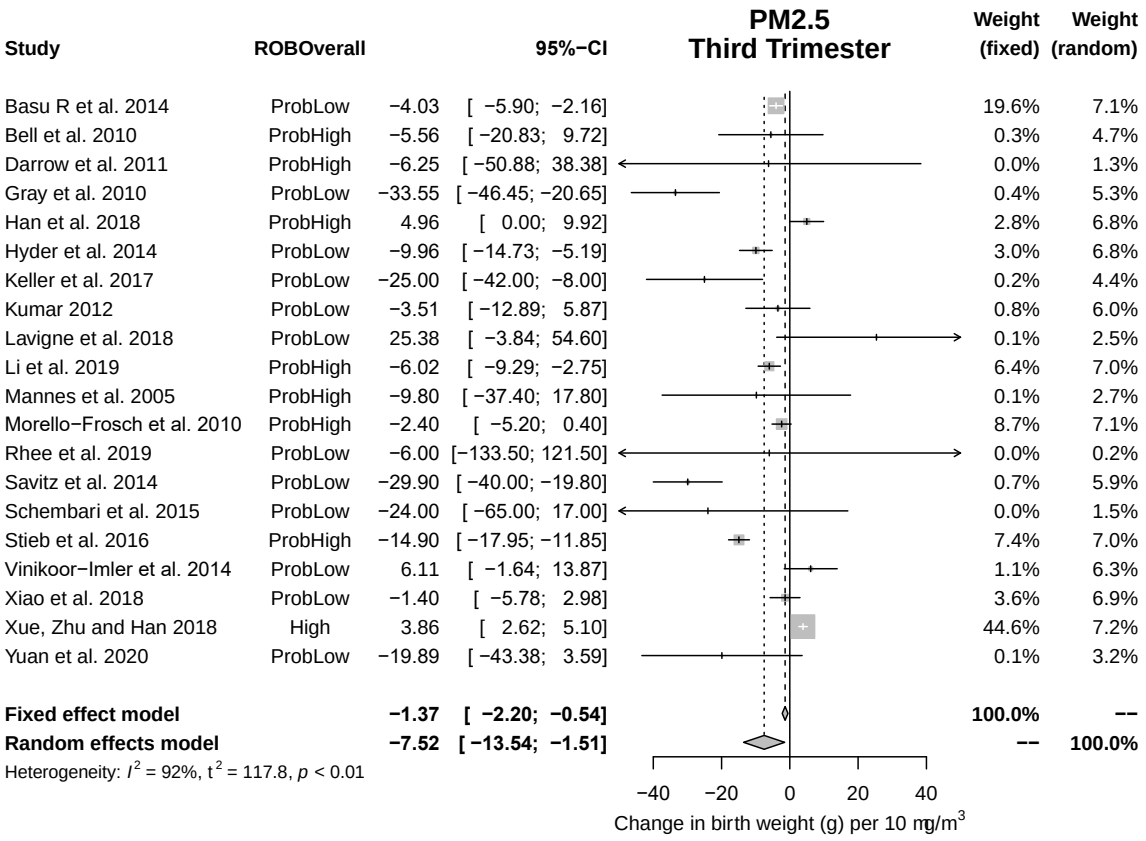
B4.1 $PM_{2.5}$ (First Trimester, n=18 studies)



B4.2 $PM_{2.5}$ (Second Trimester, n=18 studies)



B4.3 $PM_{2.5}$ (Third Trimester, n=20 studies)



B4.4 $PM_{2.5}$ (Entire Pregnancy, n=28 studies)

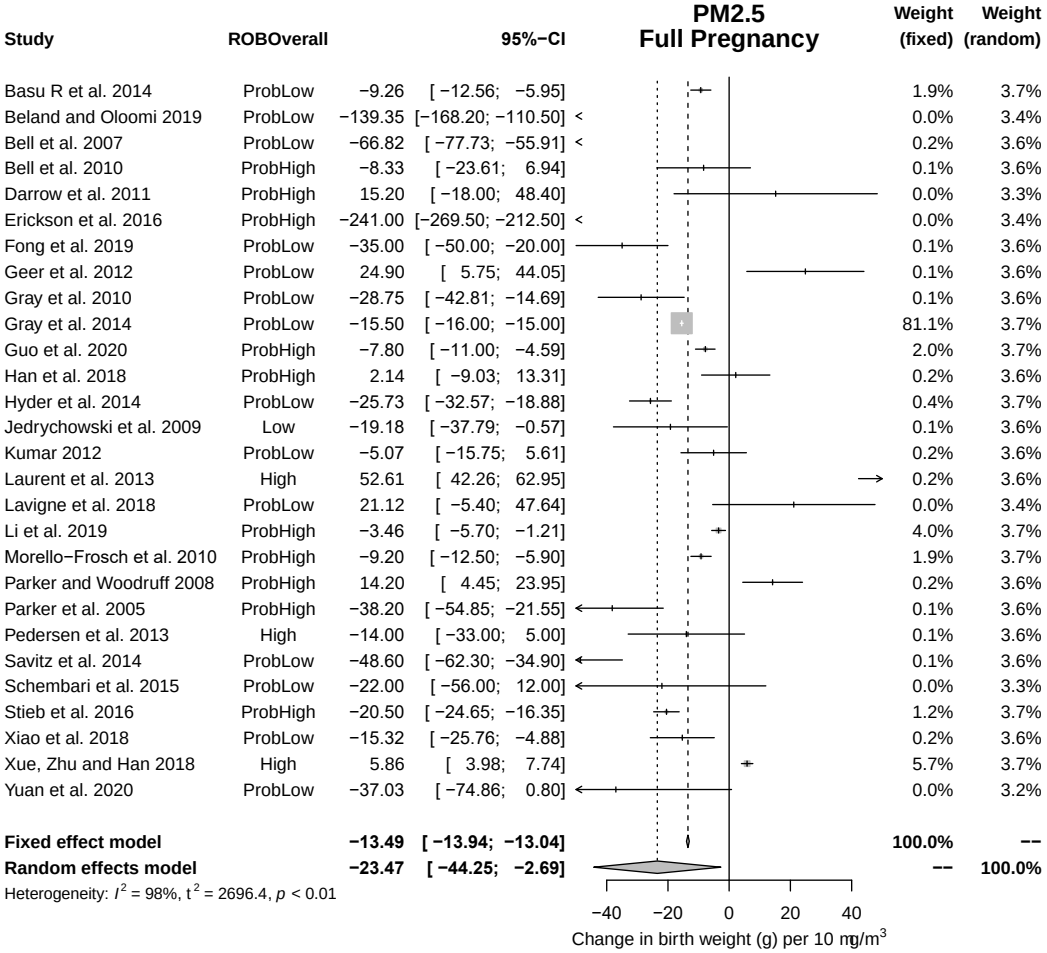
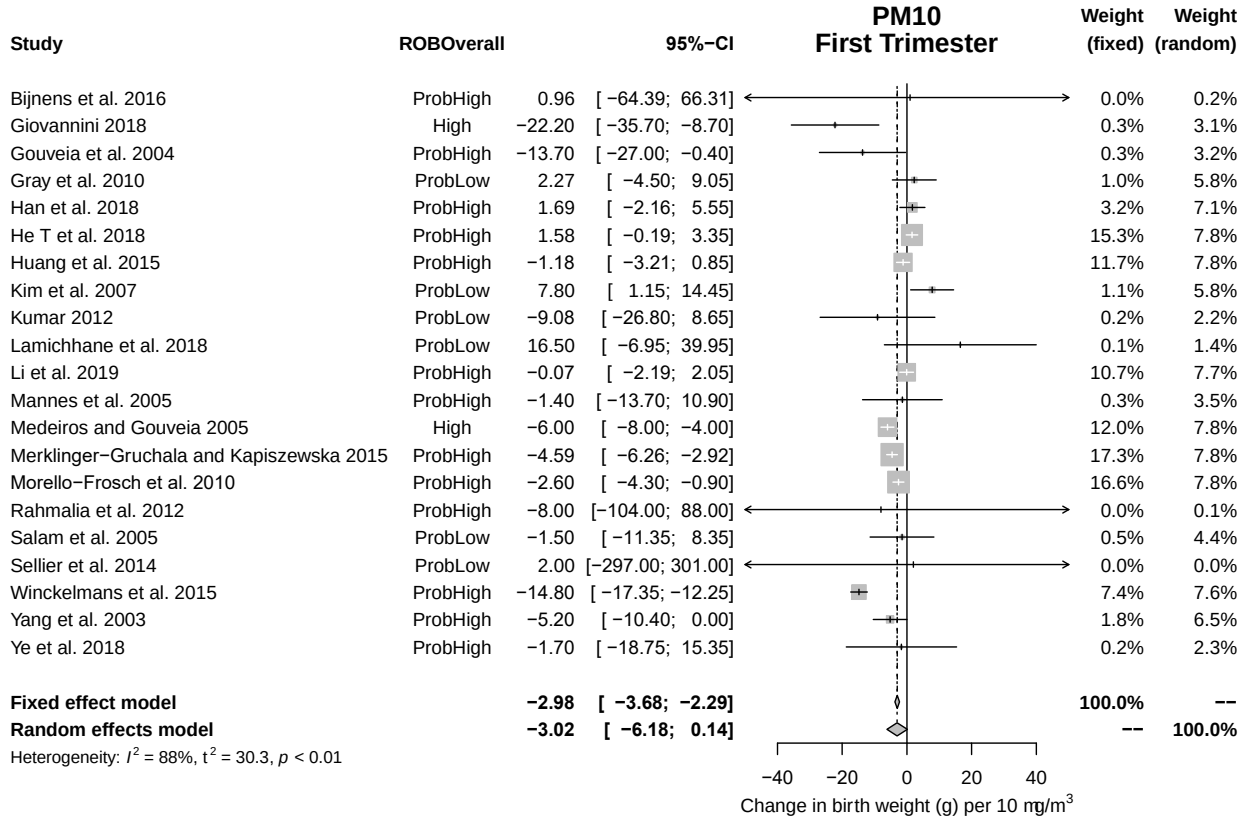
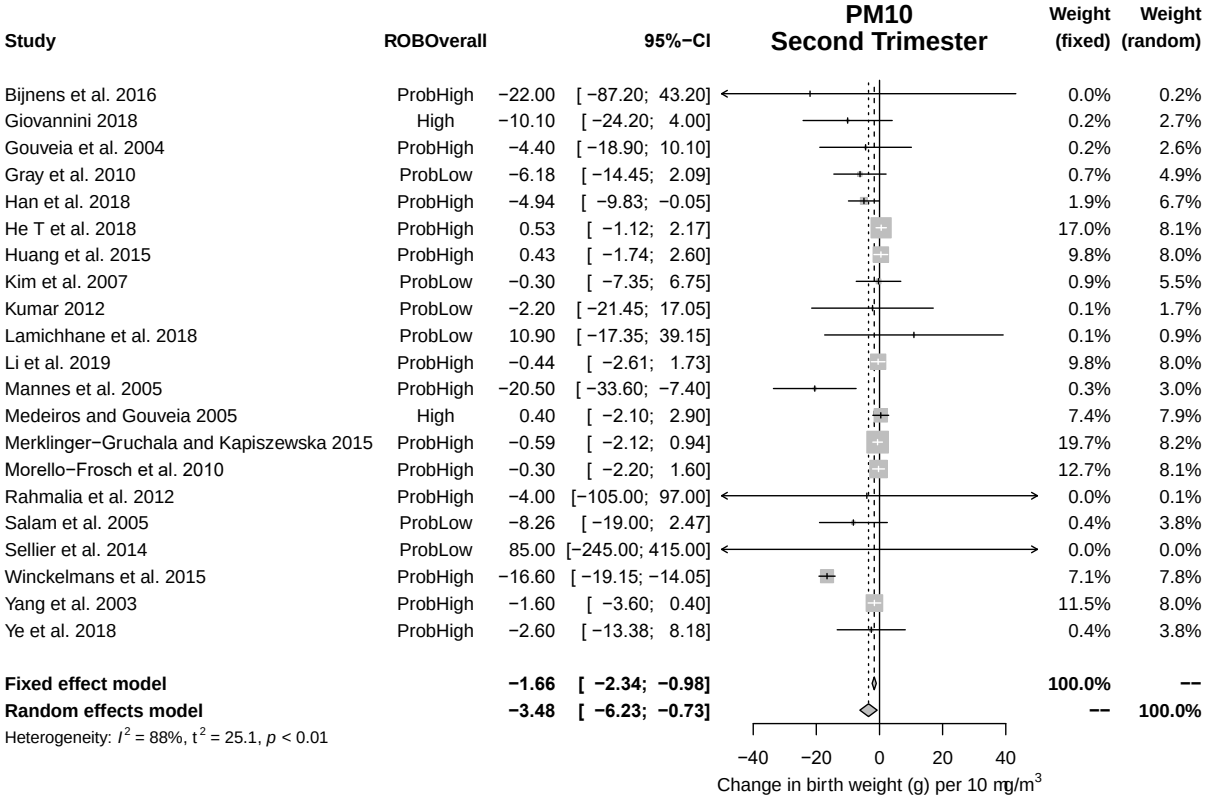


Figure B5. Results of Meta-analysis for studies evaluating PM_{10} exposure (all studies included despite risk of bias rating)

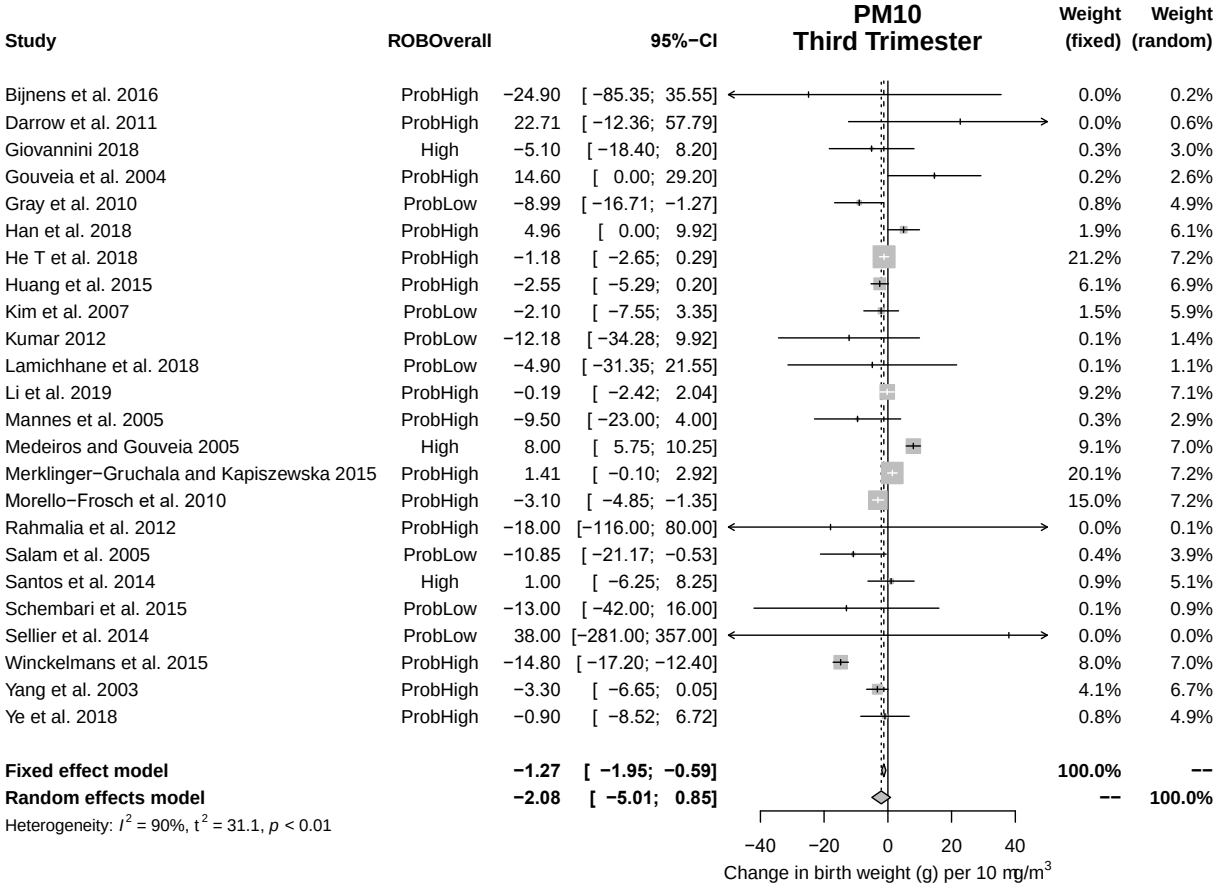
B5.1 PM_{10} (First Trimester, n=21 studies)



B5.2 PM_{10} (Second Trimester, n=21 studies)



B5.3 PM_{10} (Third Trimester, n=24 studies)



B5.4 PM_{10} (Entire Pregnancy, n=21 studies)

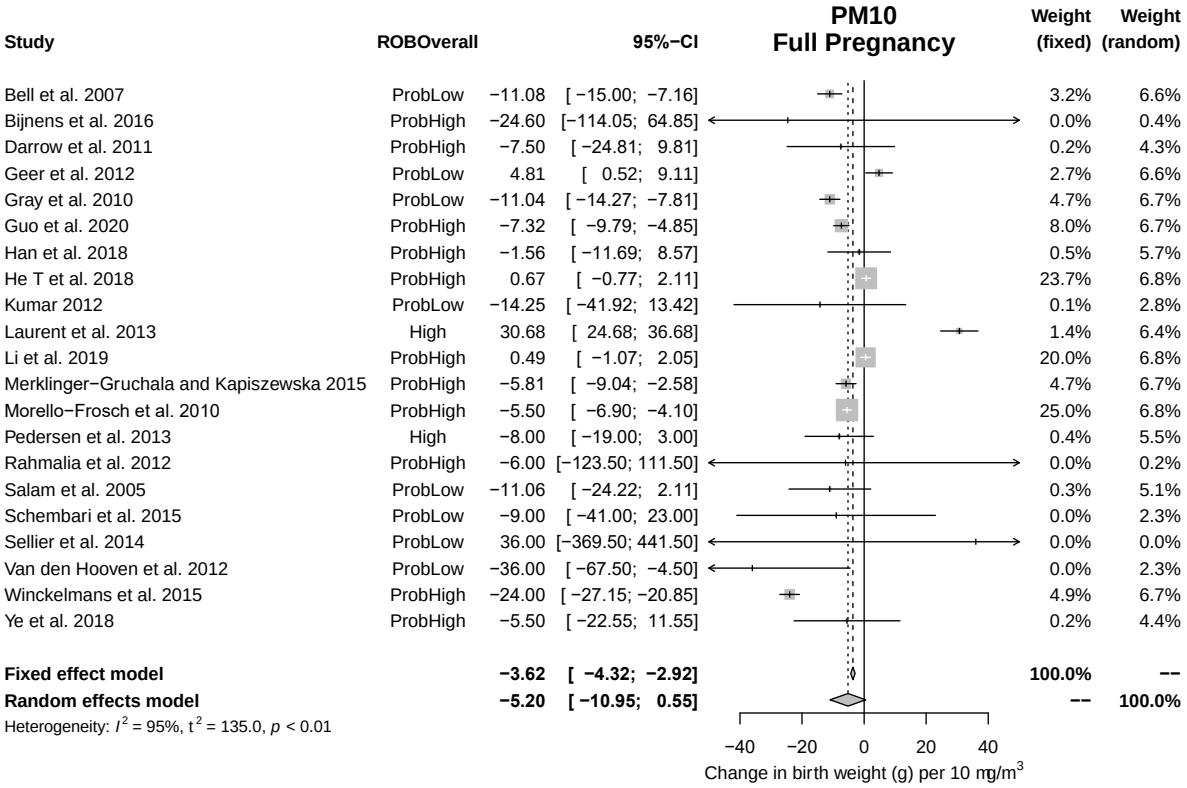
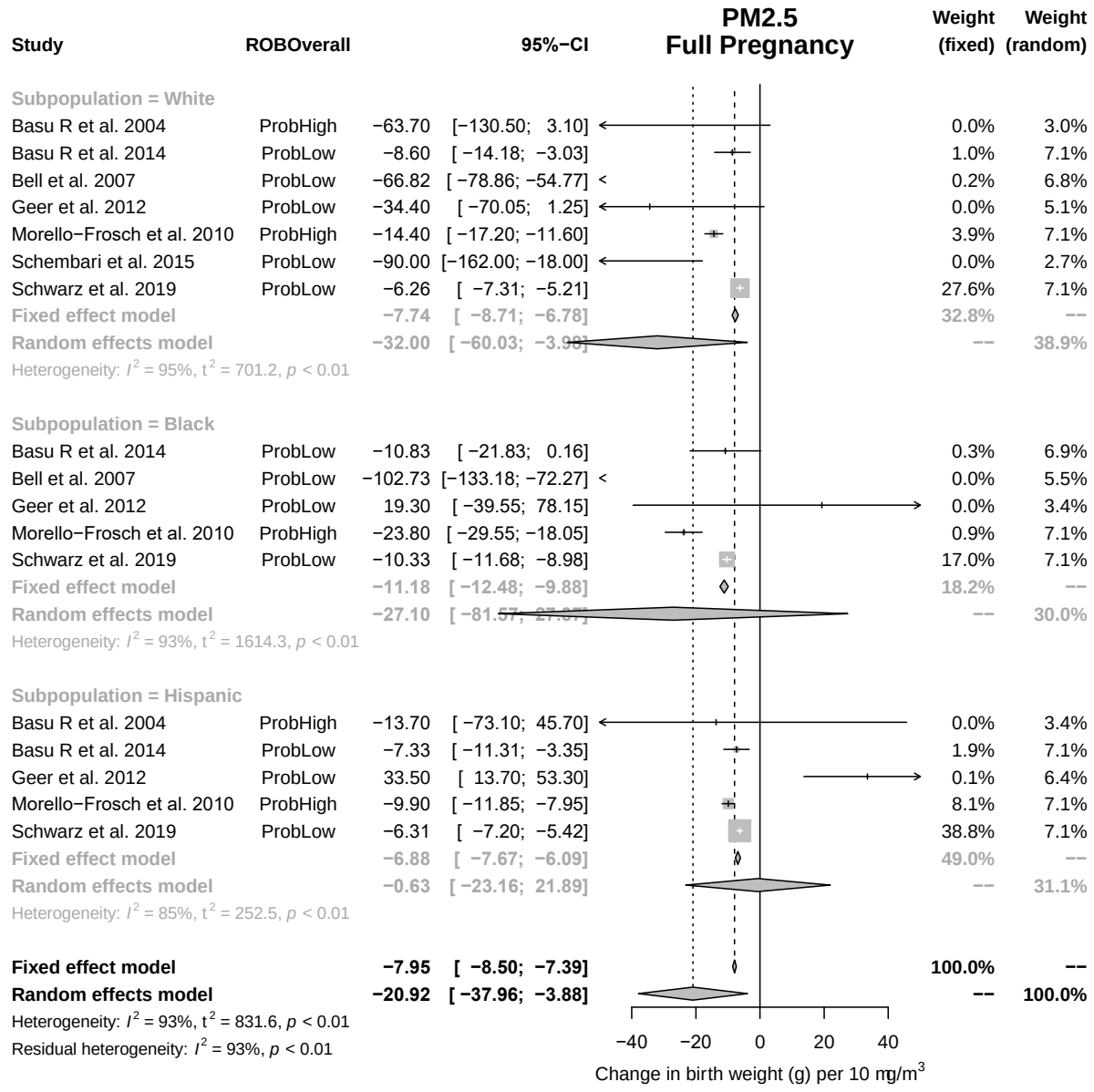


Figure B6. Subgrouping Analysis based on ethnicity (all studies, regardless of risk of bias rating; only analyses with at least 3 studies in at least 2 subgroups included)

B6.1 $PM_{2.5}$ (Entire Pregnancy)



B6.2 PM_{10} (Entire Pregnancy)

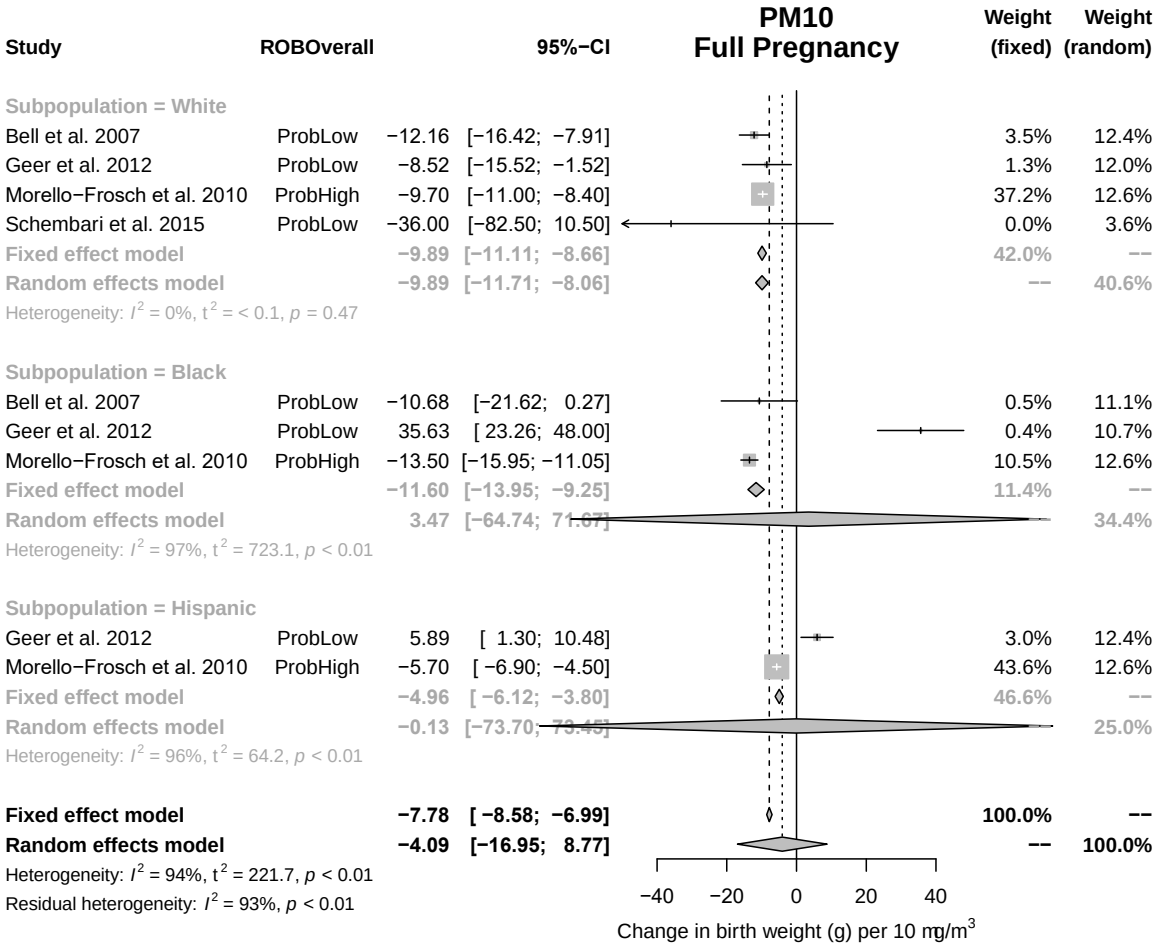
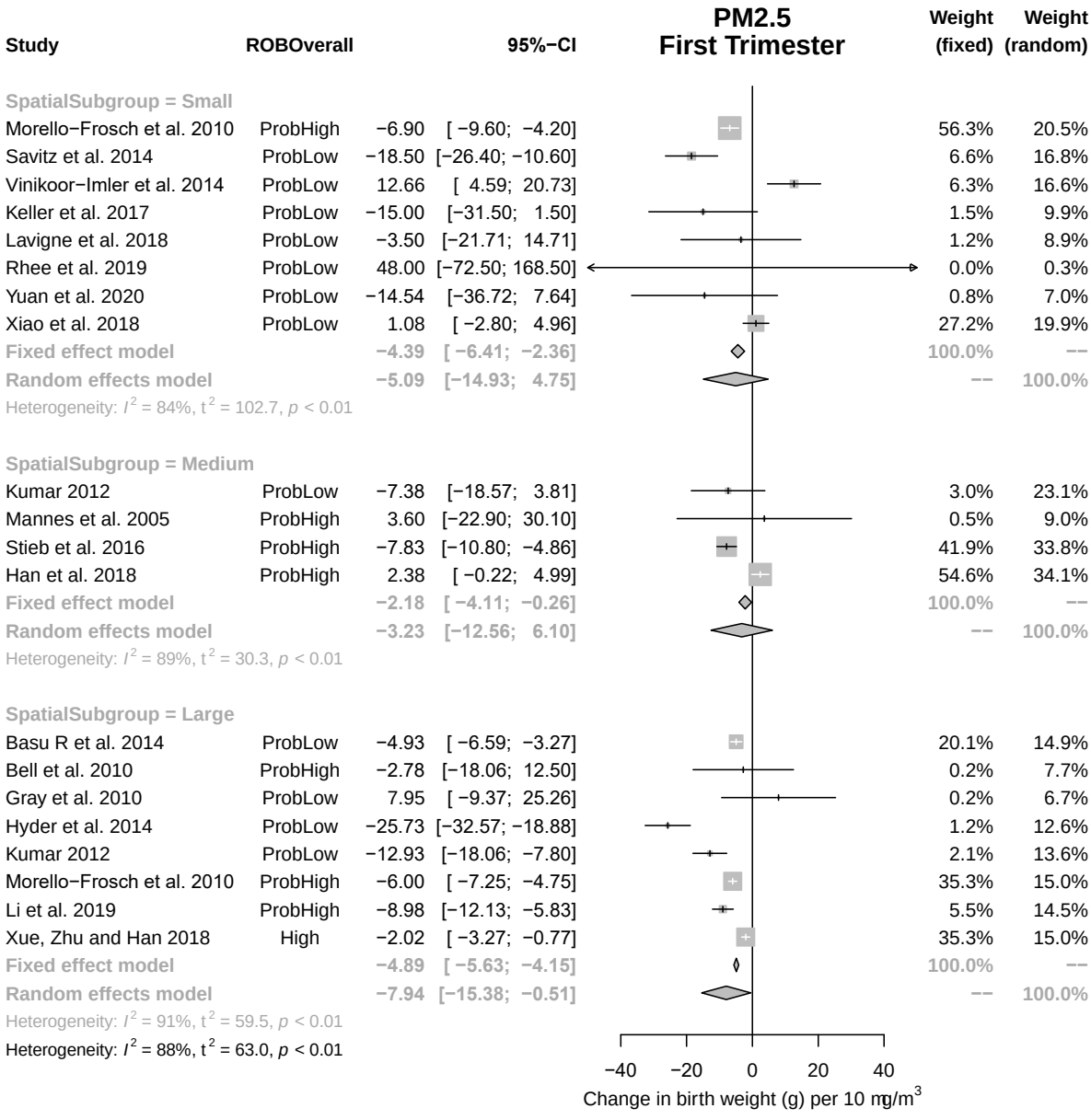
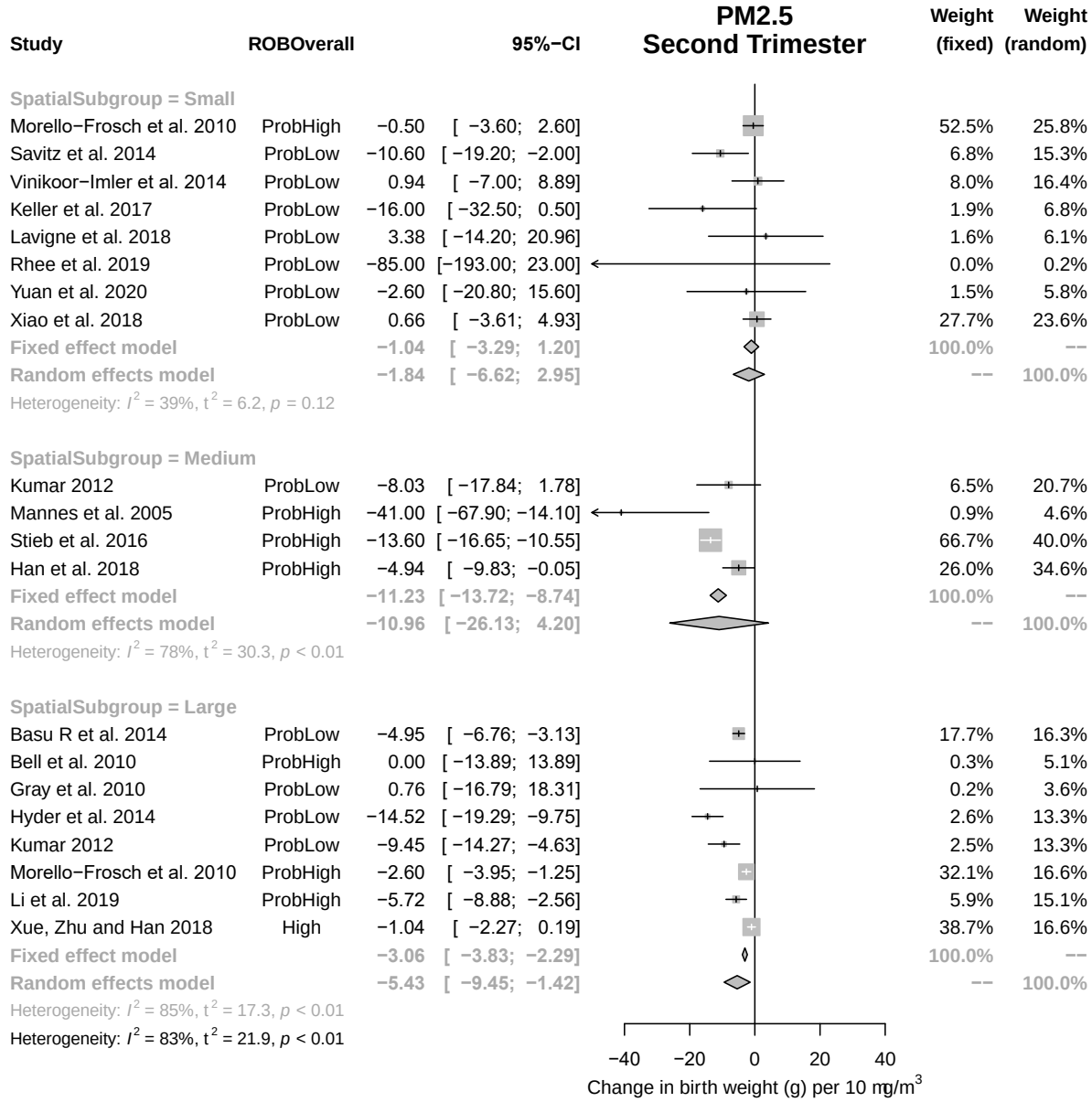


Figure B7. Subgrouping analysis based on spatial scale (all studies, regardless of risk of bias rating; only analyses with at least 3 studies in at least 2 subgroups included)

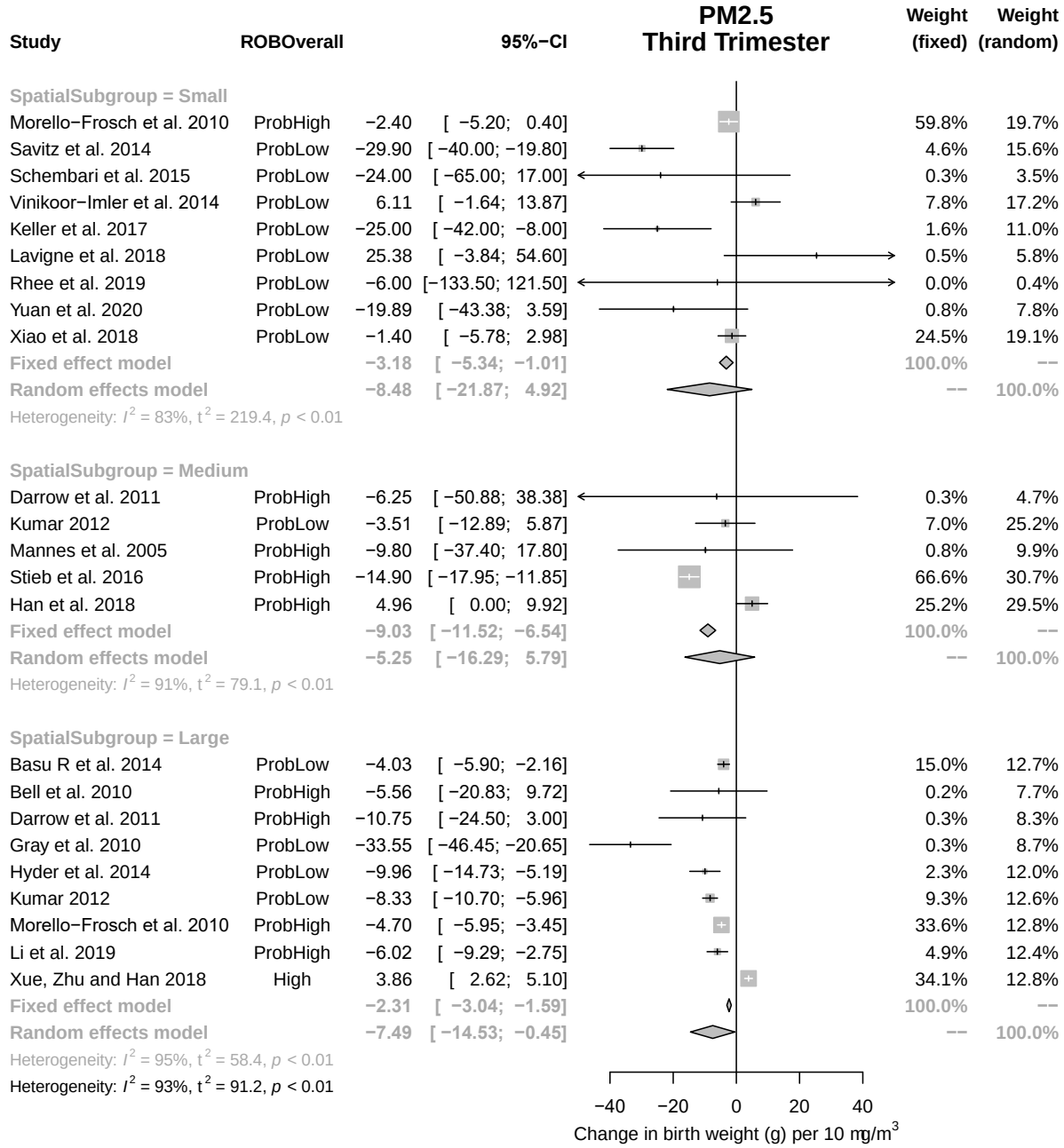
B7.1 $PM_{2.5}$ (first trimester)



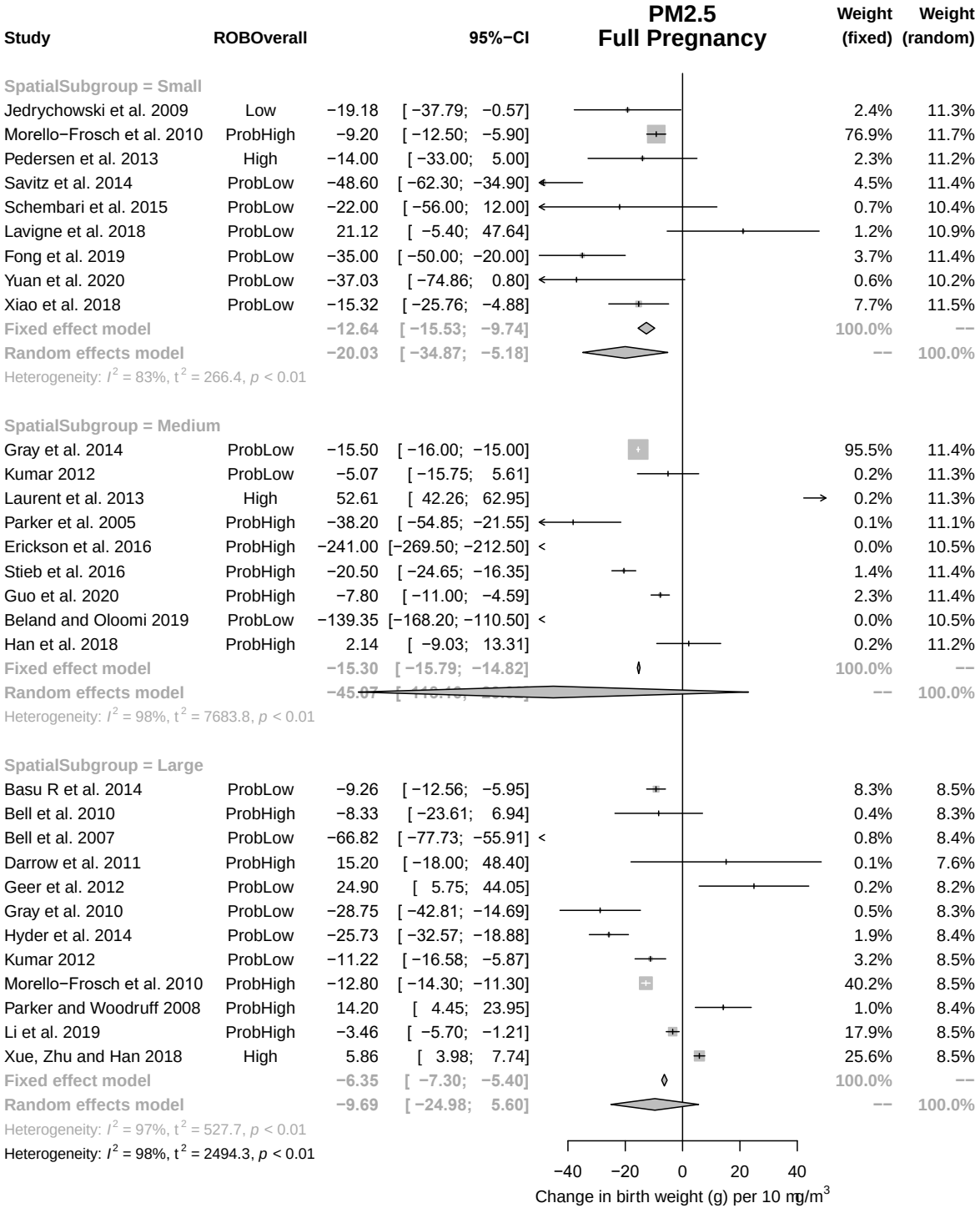
B7.2 $PM_{2.5}$ (Second Trimester)



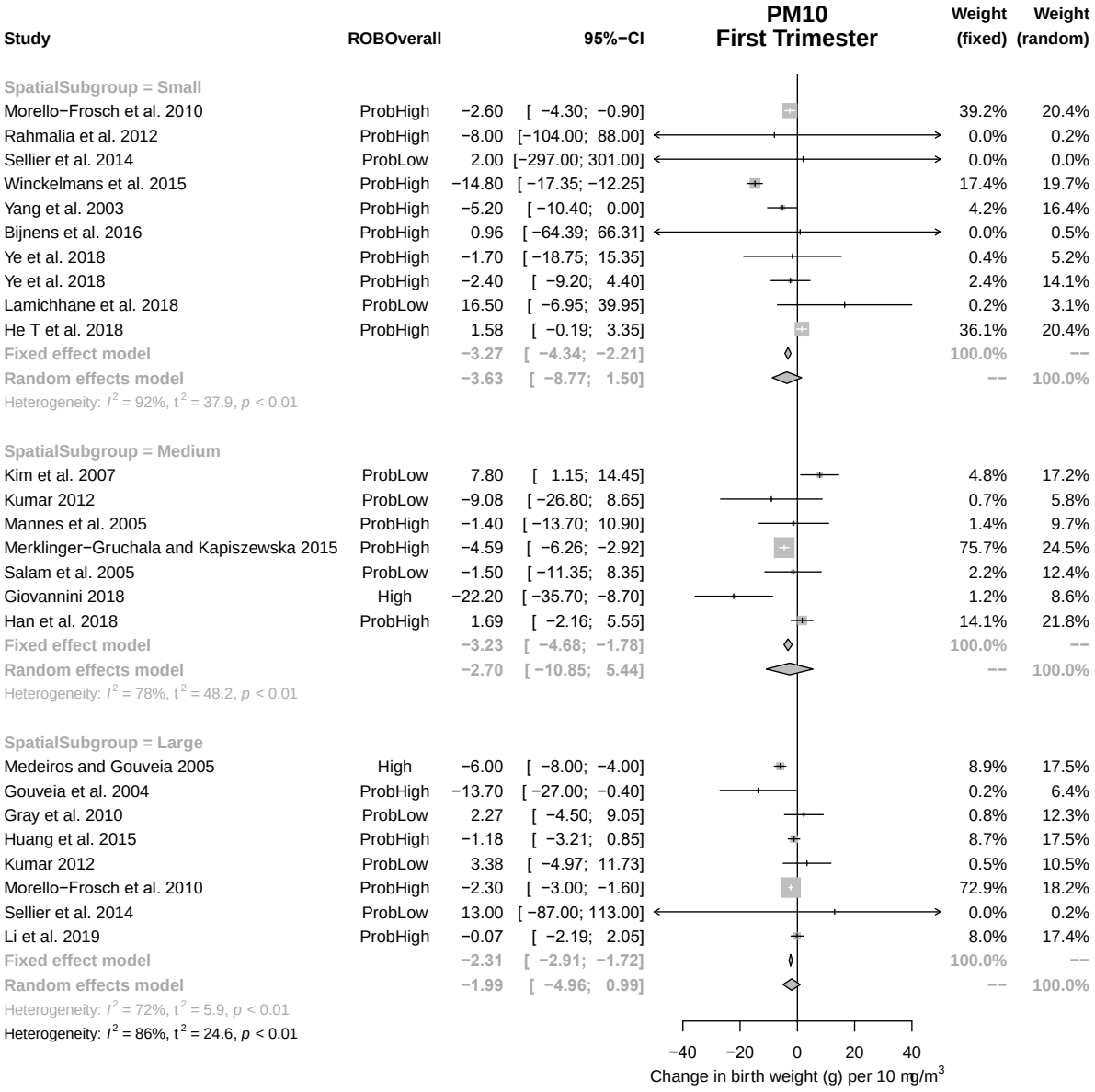
B7.3 $PM_{2.5}$ (Third Trimester)



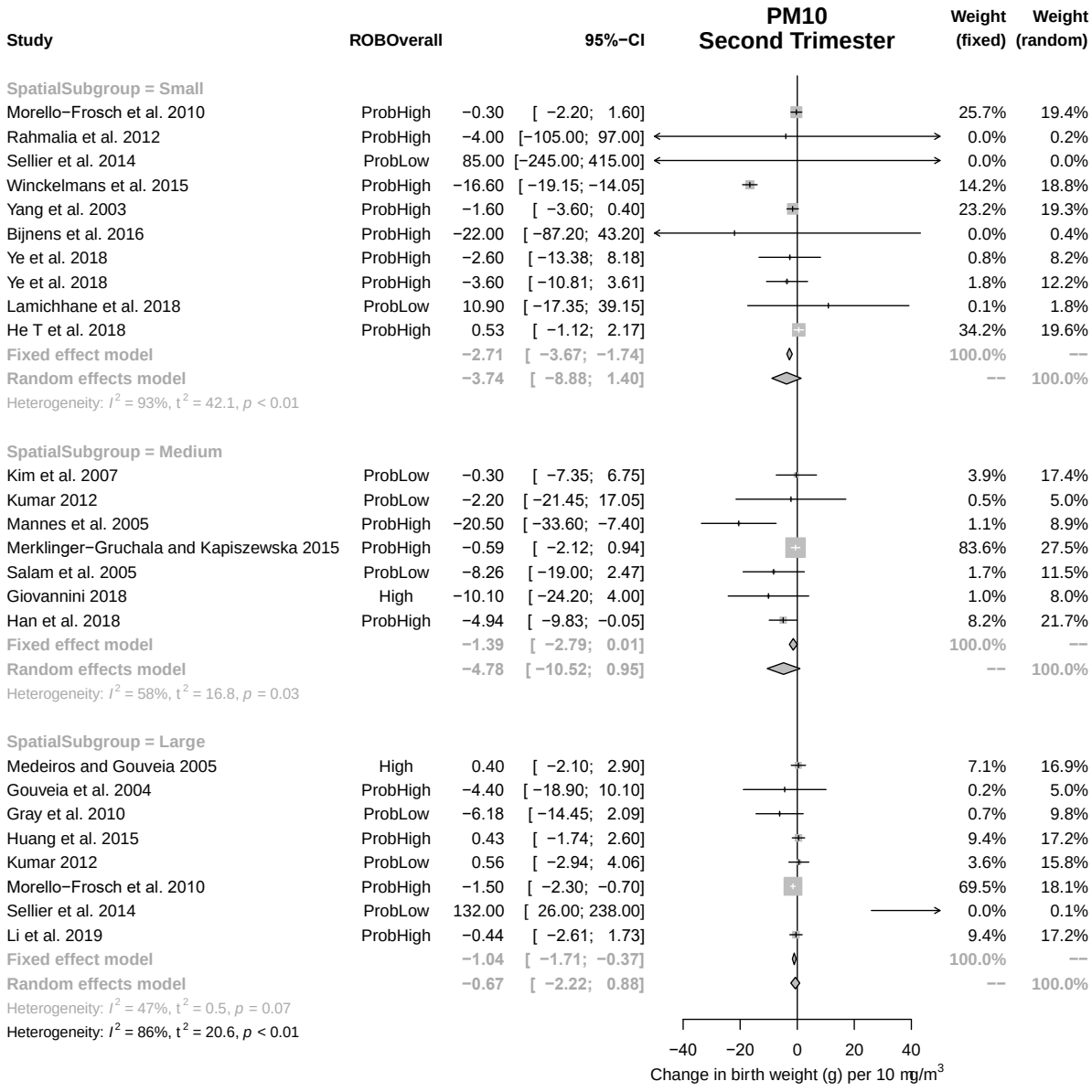
B7.4 $PM_{2.5}$ (Entire Pregnancy)



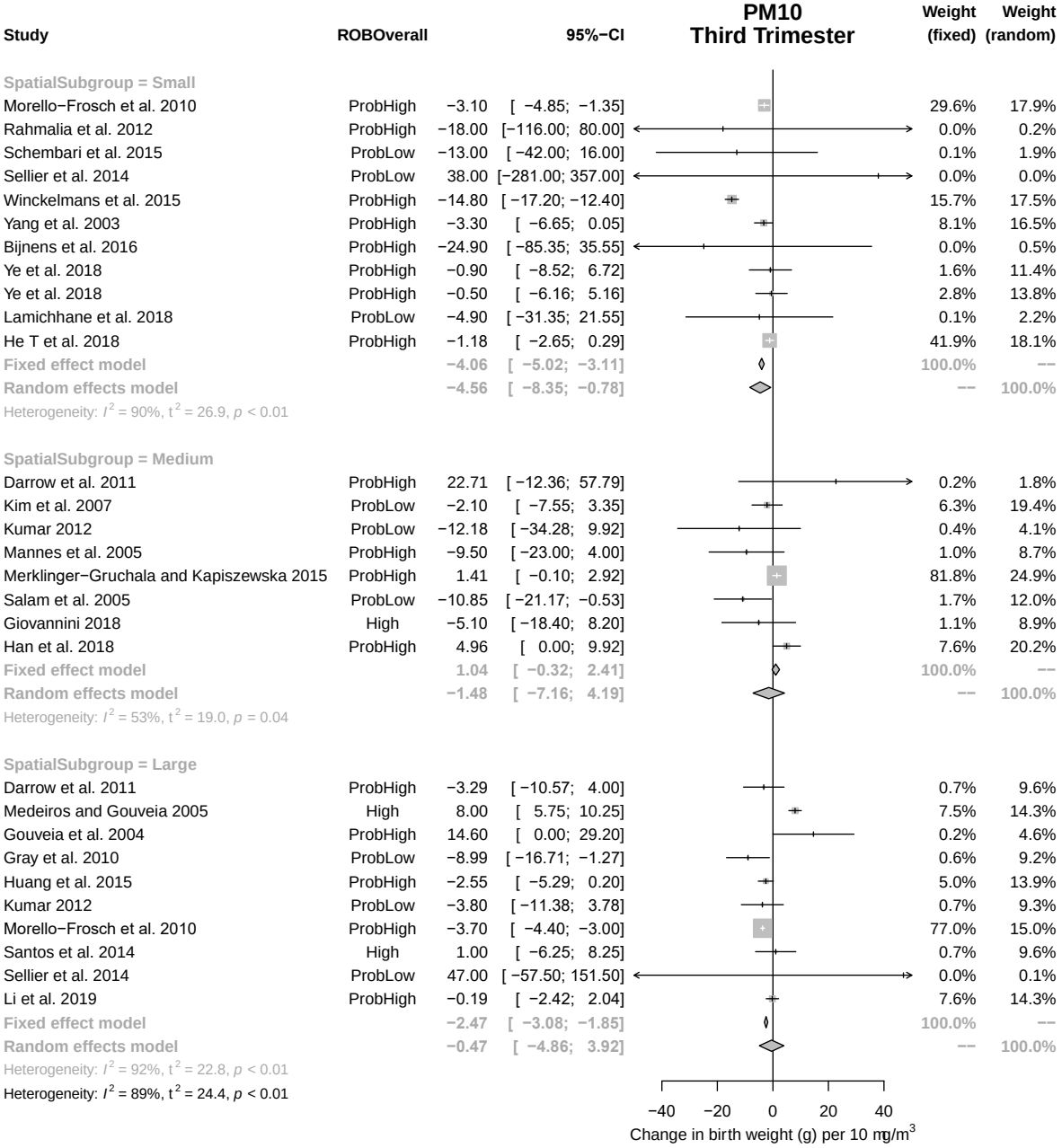
B7.5 PM_{10} (First Trimester)



B7.6 PM_{10} (Second Trimester)



B7.7 PM_{10} (Third Trimester)



B7.8 PM_{10} (Entire Pregnancy)

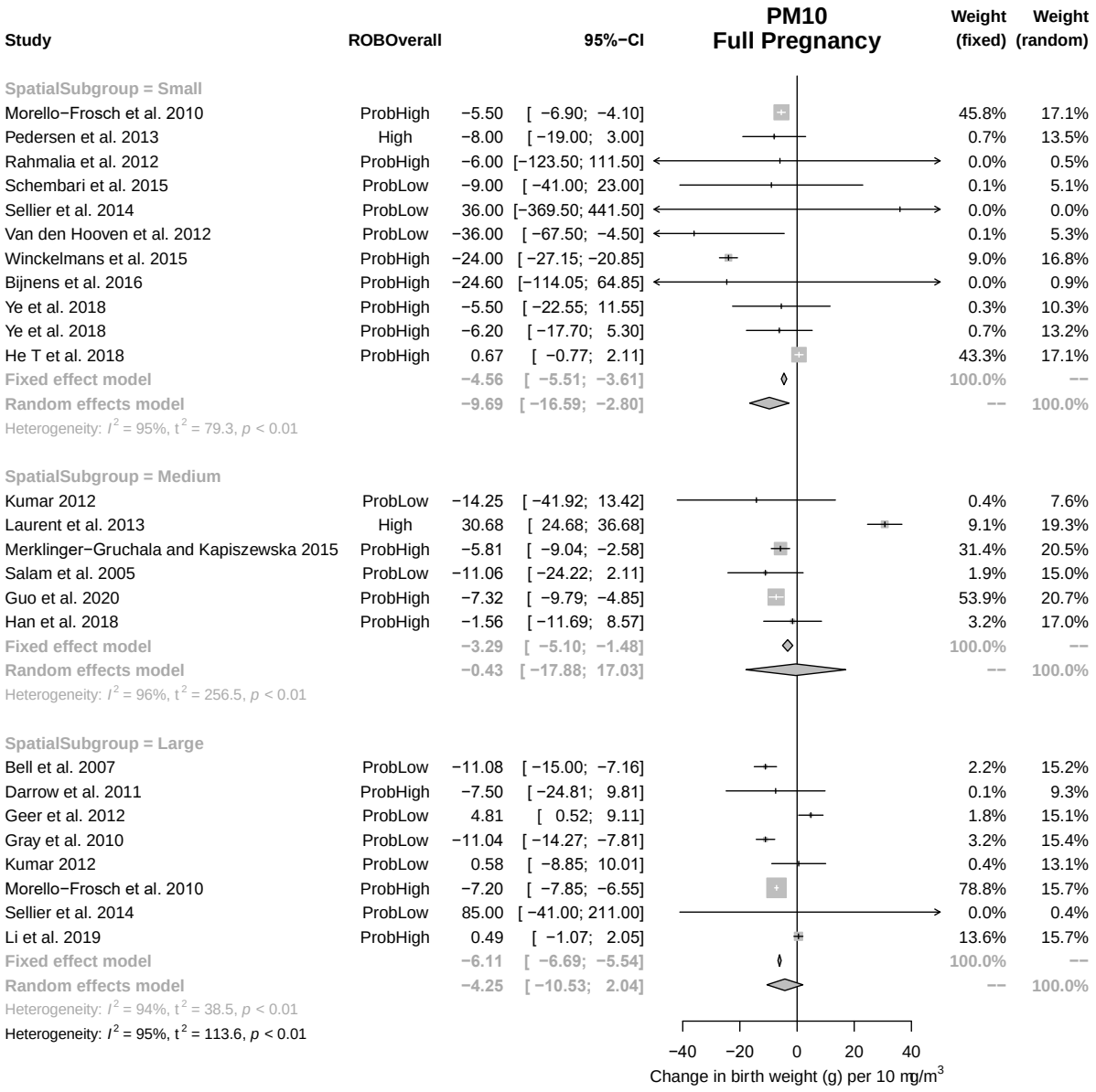
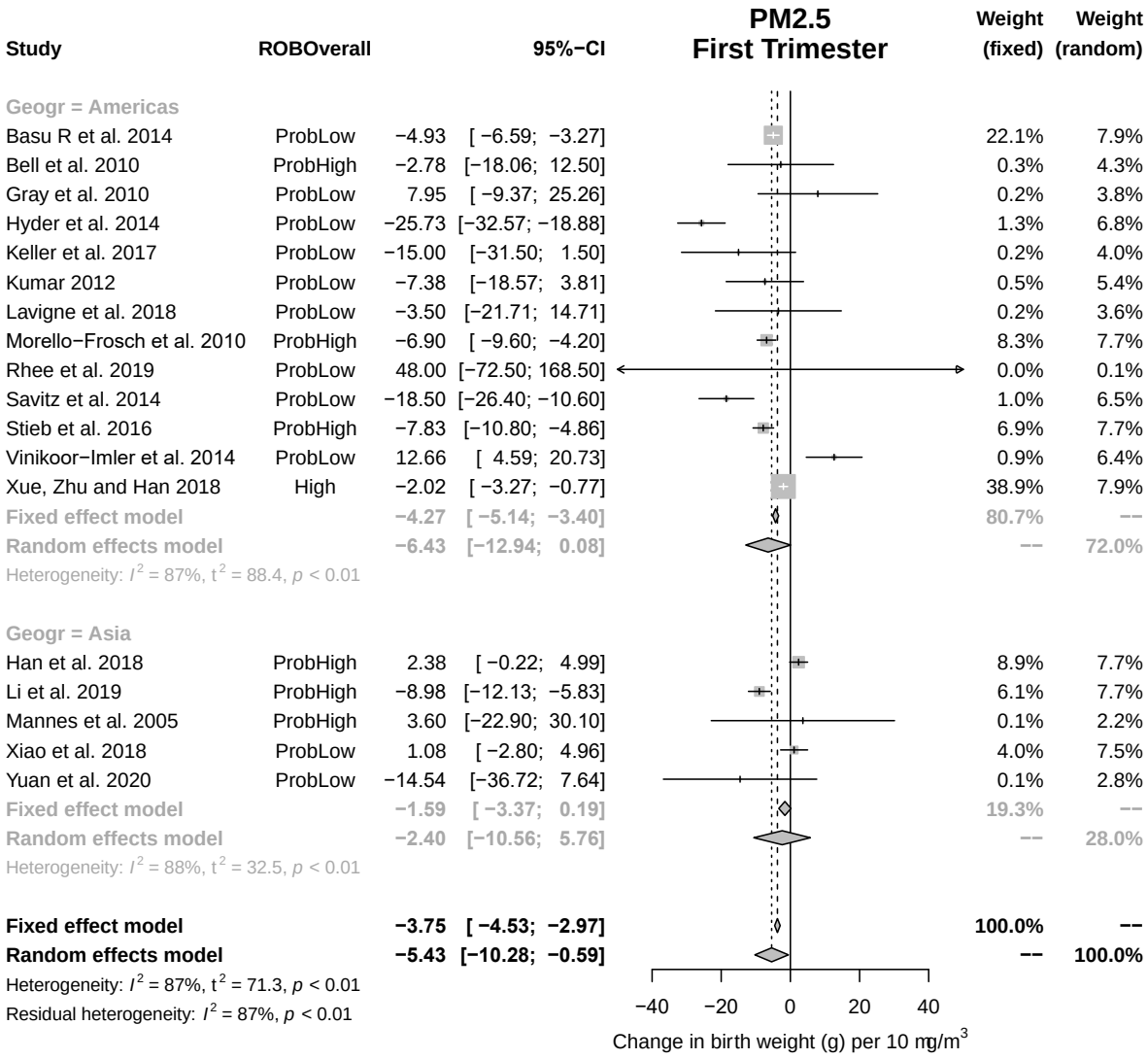
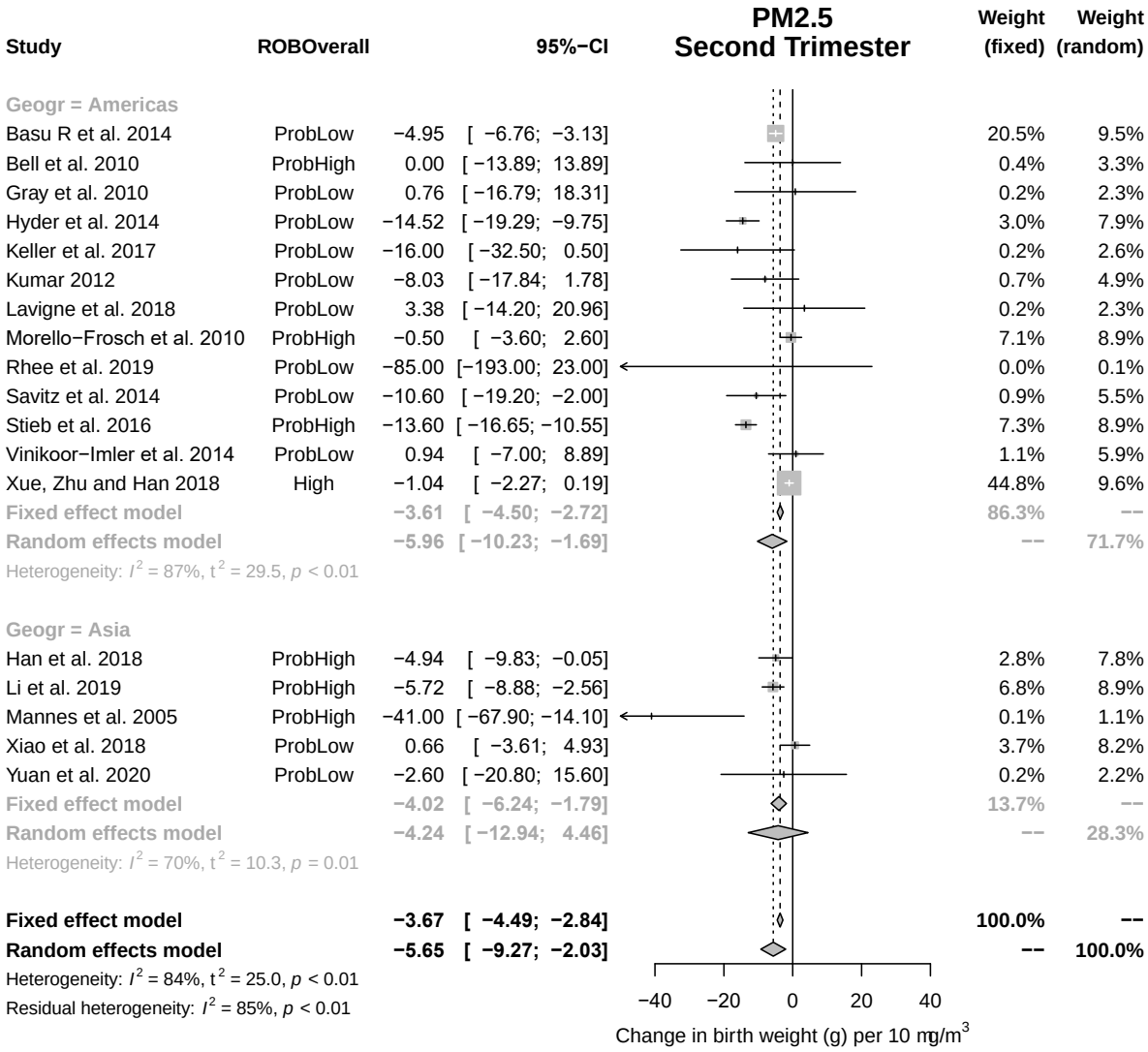


Figure B8. Subgrouping Analysis based on geographic location (all studies, regardless of risk of bias rating; only analyses with at least 3 studies in at least 2 subgroups included)

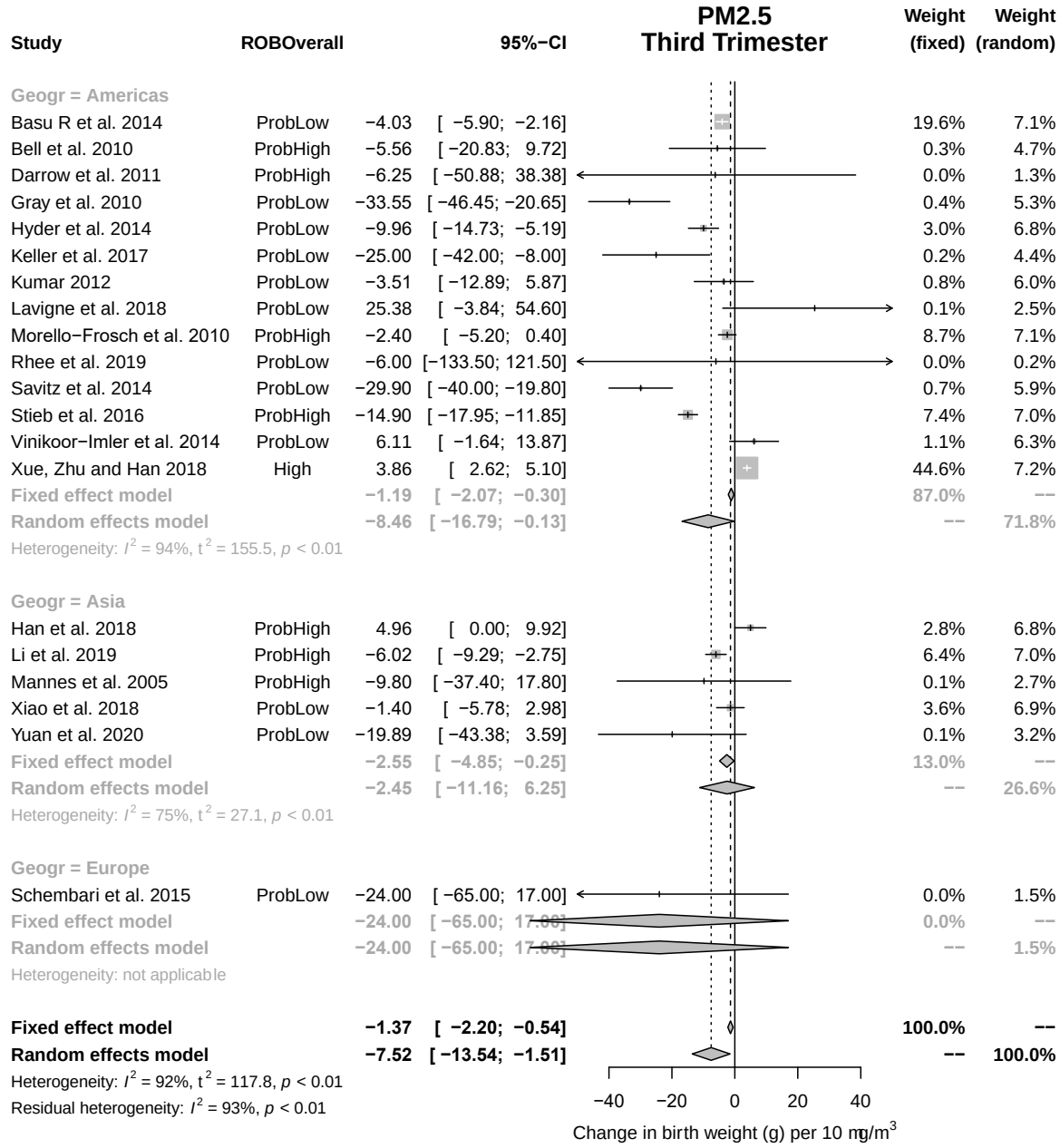
B8.1 $PM_{2.5}$ (first trimester)



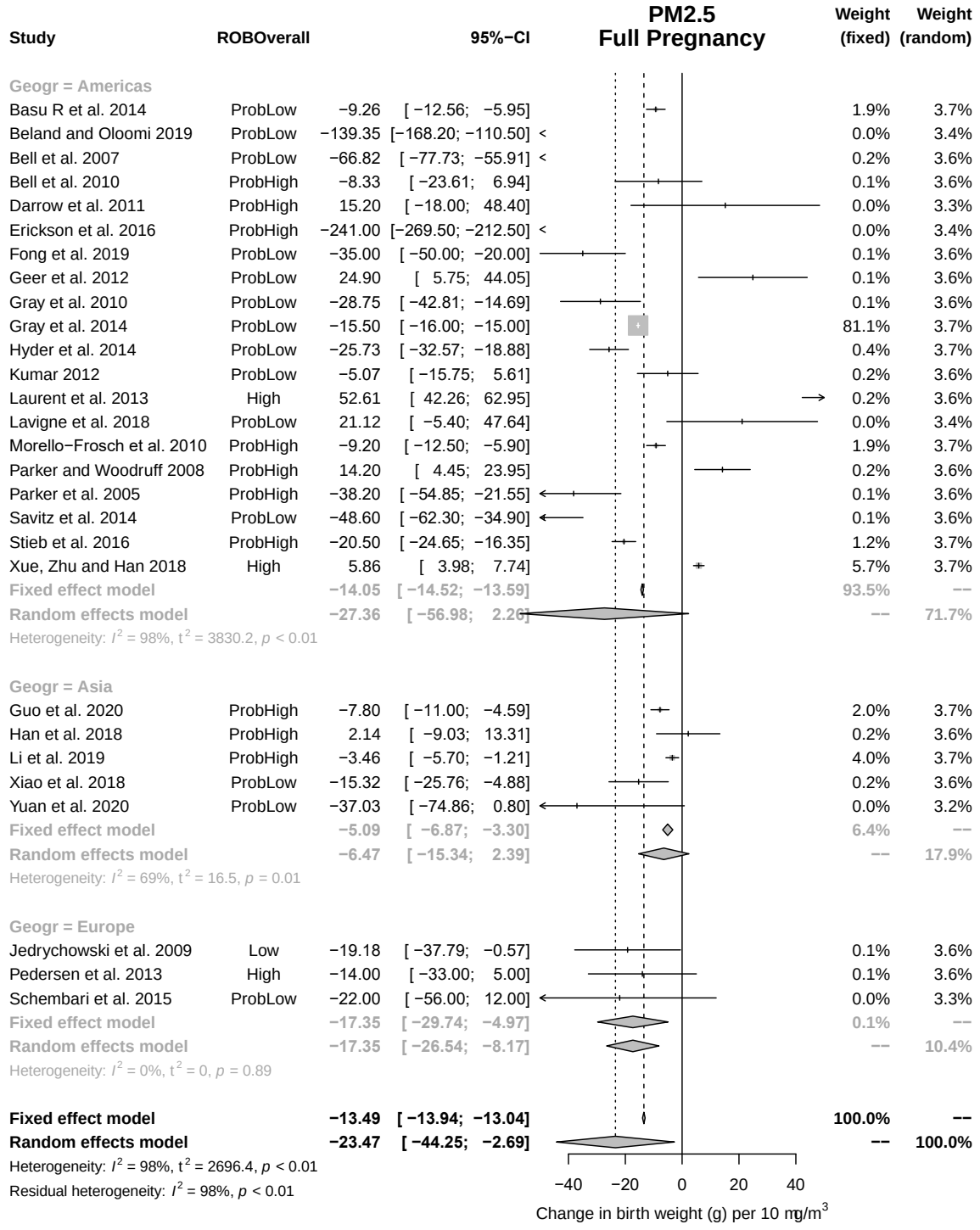
B8.2 $PM_{2.5}$ (second trimester)



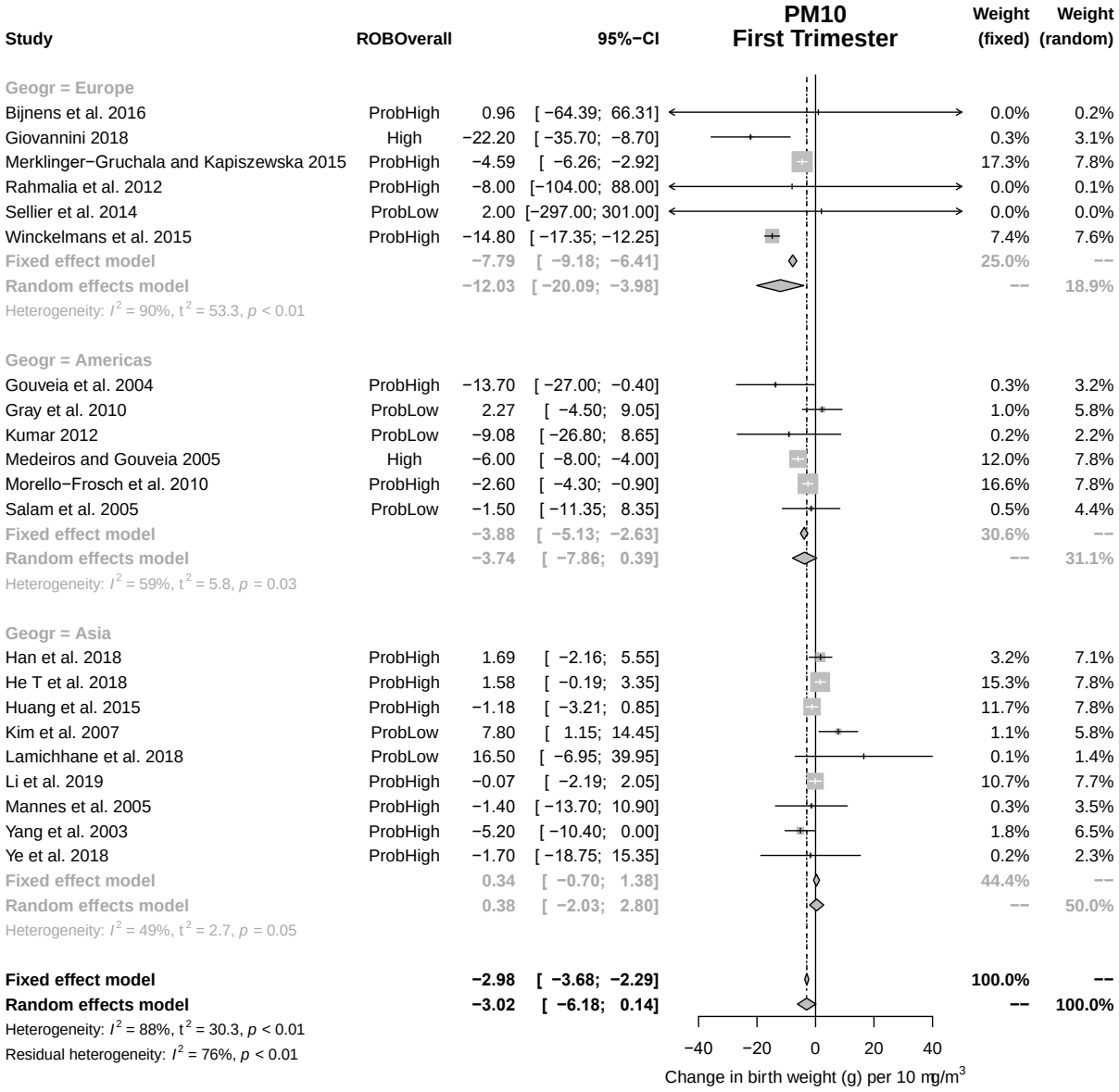
B8.3 $PM_{2.5}$ (third trimester)



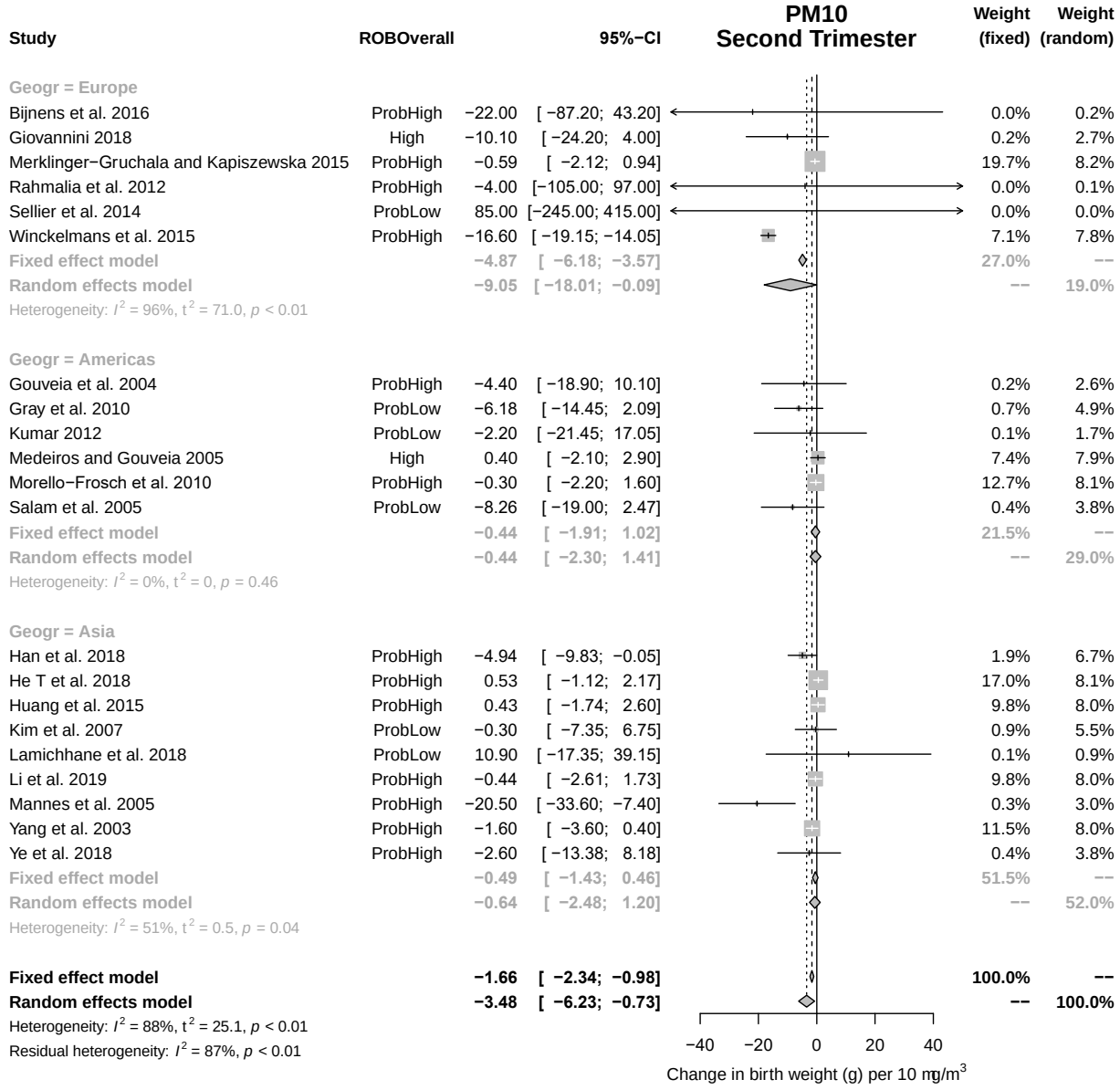
B8.4 $PM_{2.5}$ (entire pregnancy)



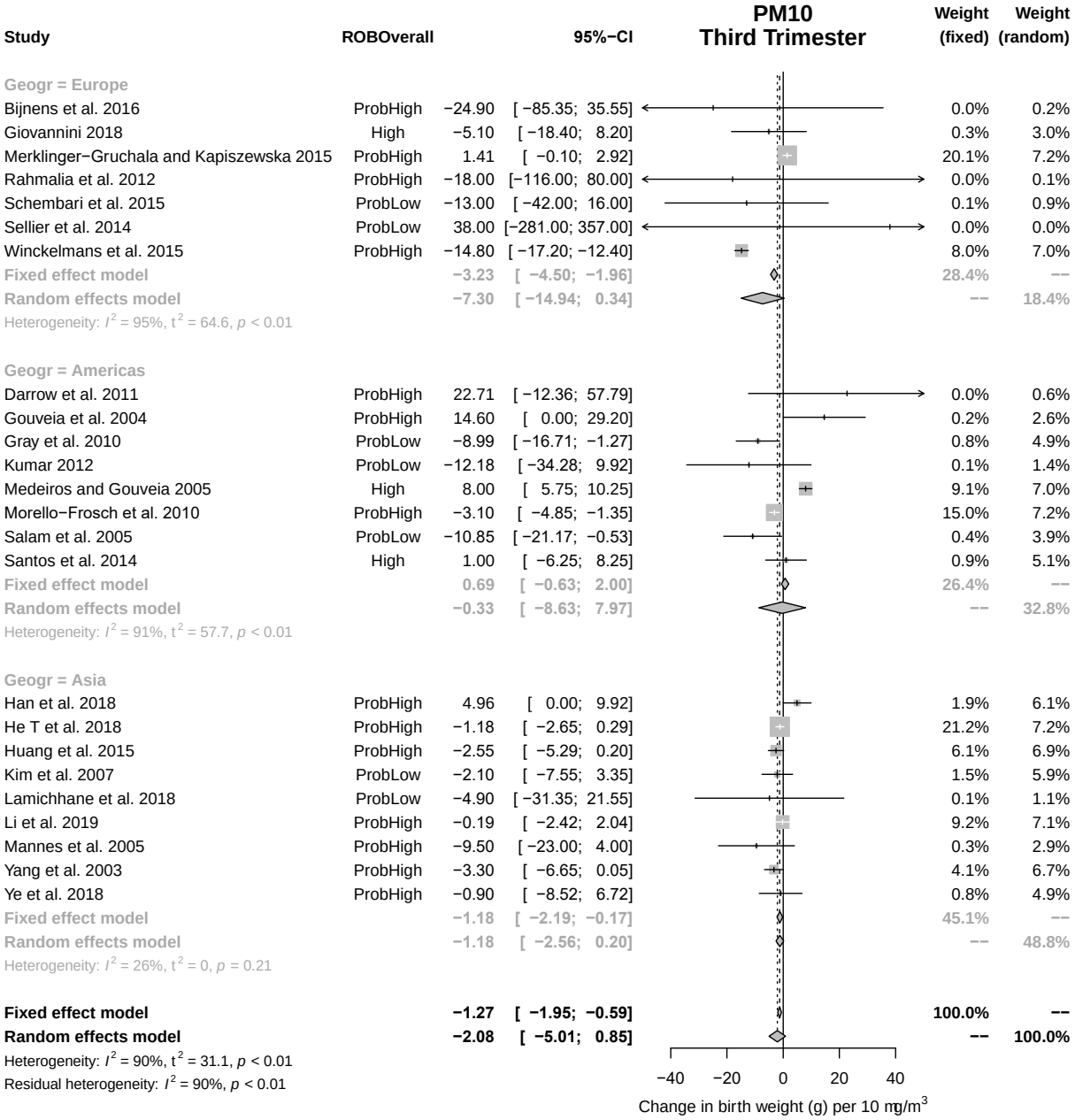
B8.5 PM_{10} (first trimester)



B8.6 PM_{10} (second trimester)



B8.7 PM_{10} (third trimester)



B8.8 PM_{10} (entire pregnancy)

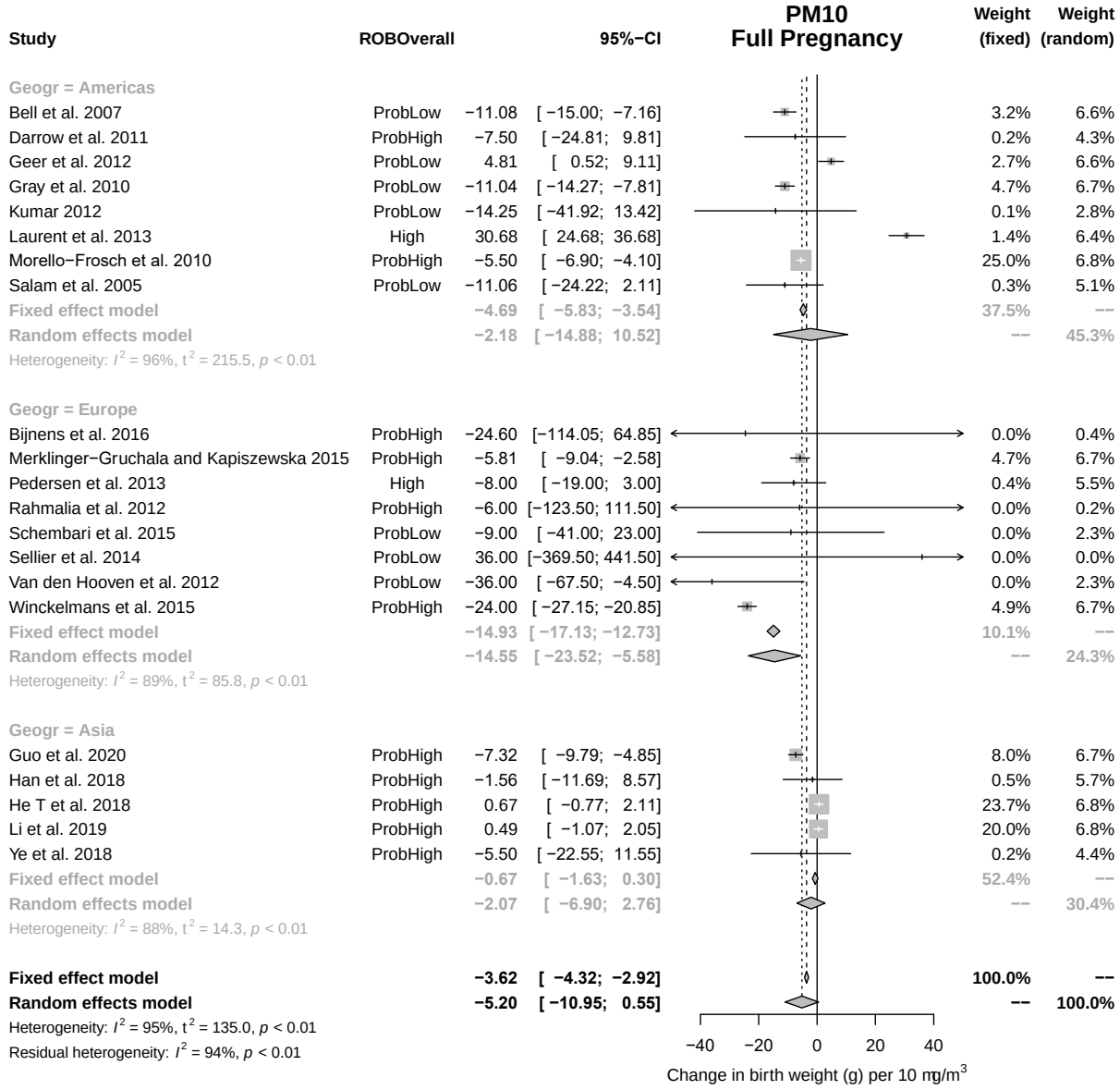
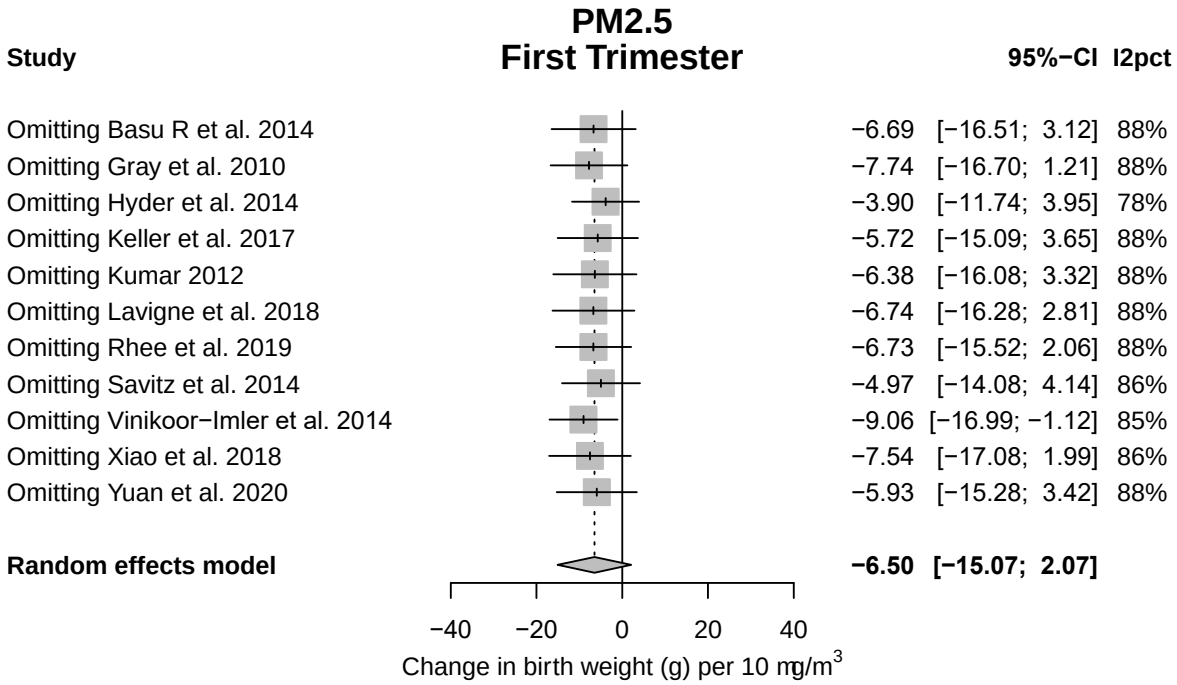
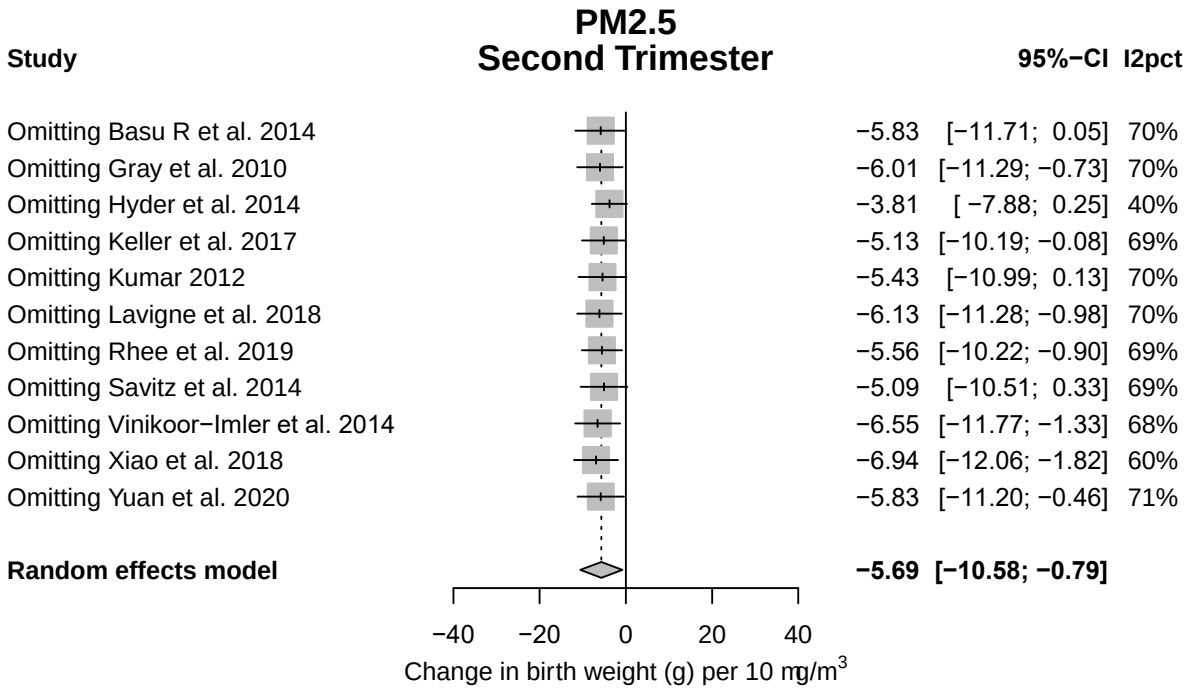


Figure B9. Influence analysis removing one study at a time

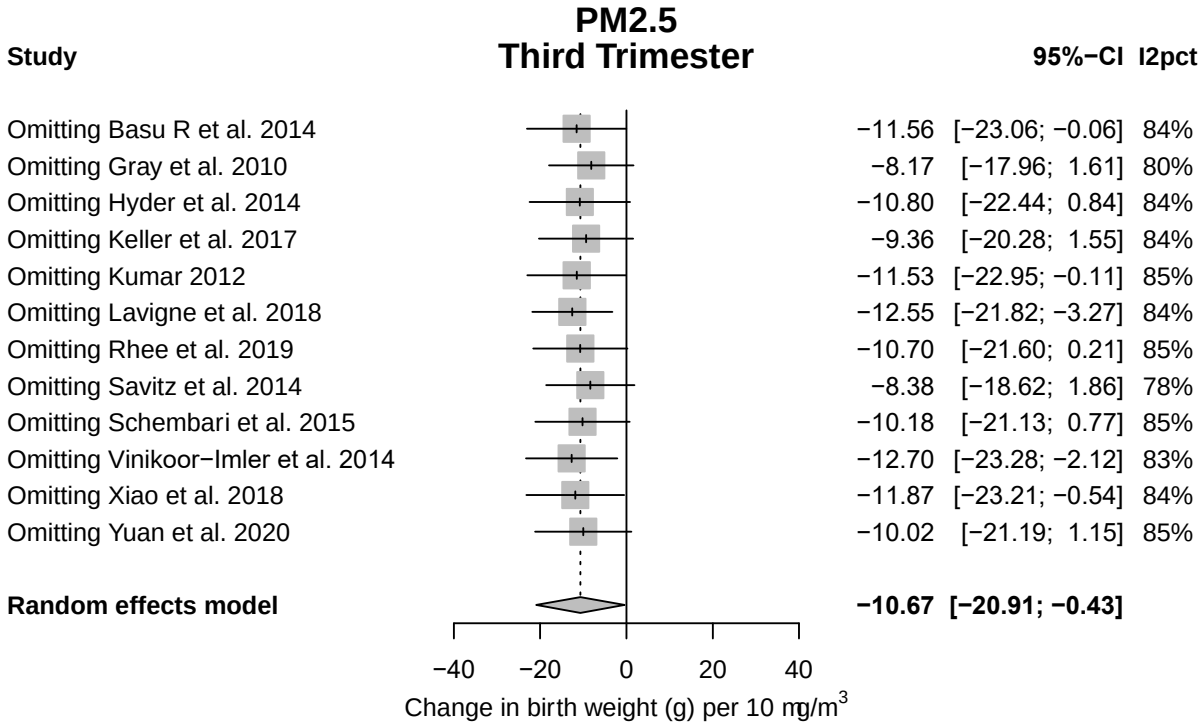
B9.1 $PM_{2.5}$ (first trimester, low or probably low risk of bias studies)



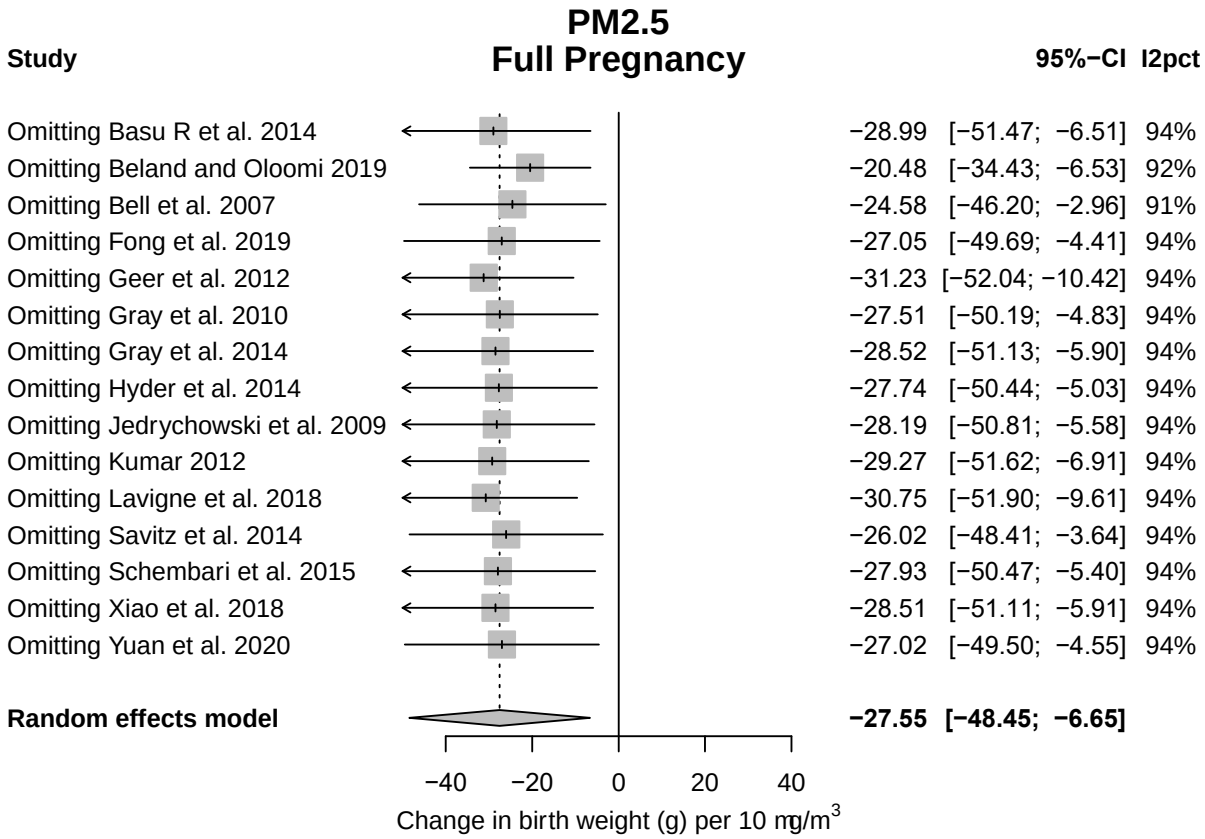
B9.2 $PM_{2.5}$ (second trimester, low or probably low risk of bias studies)



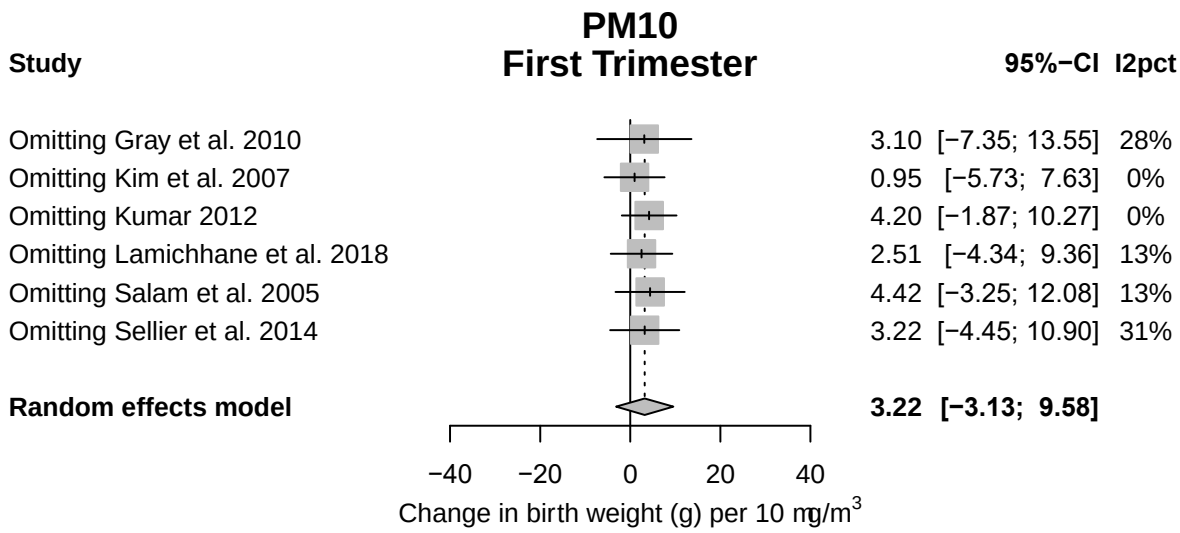
B9.3 $PM_{2.5}$ (third trimester, low or probably low risk of bias studies)



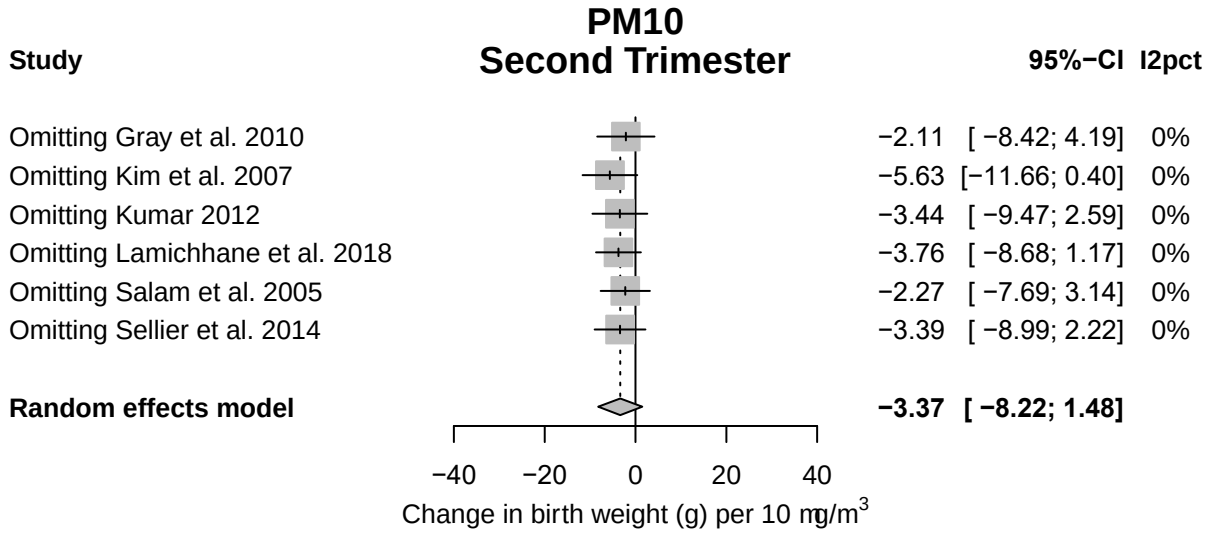
B9.4 $PM_{2.5}$ (entire pregnancy, low or probably low risk of bias studies)



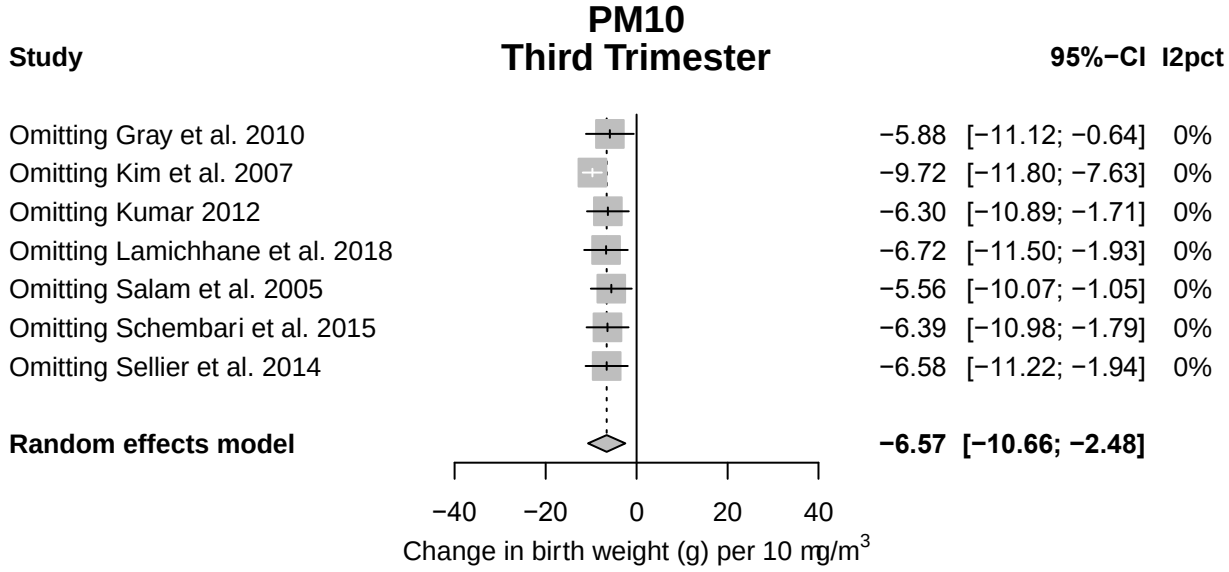
B9.5 PM_{10} (first trimester, low or probably low risk of bias studies)



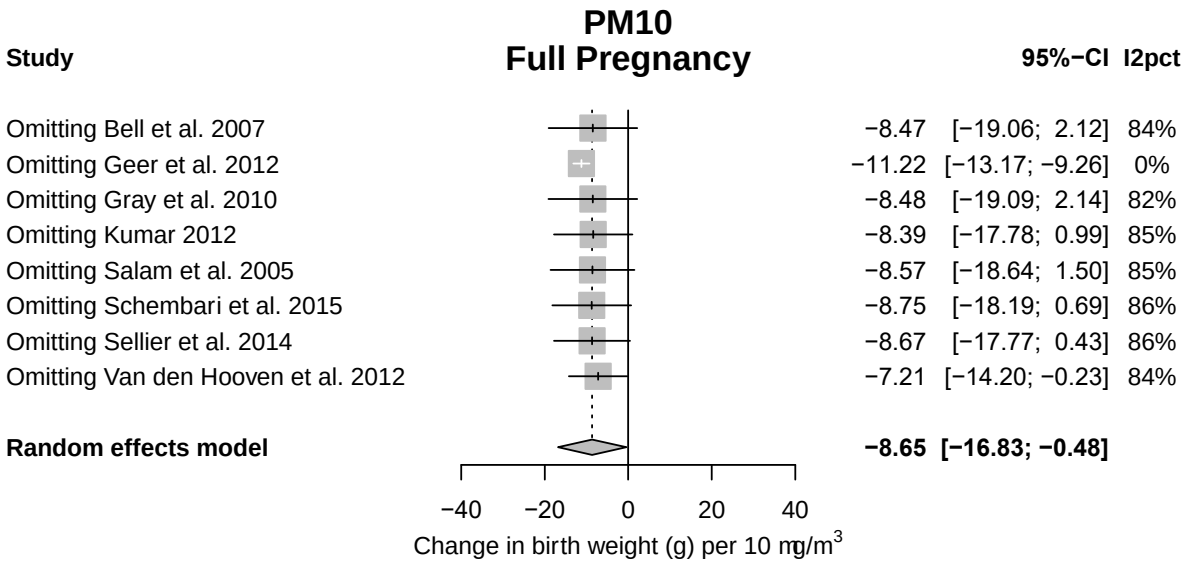
B9.6 PM_{10} (second trimester, low or probably low risk of bias studies)



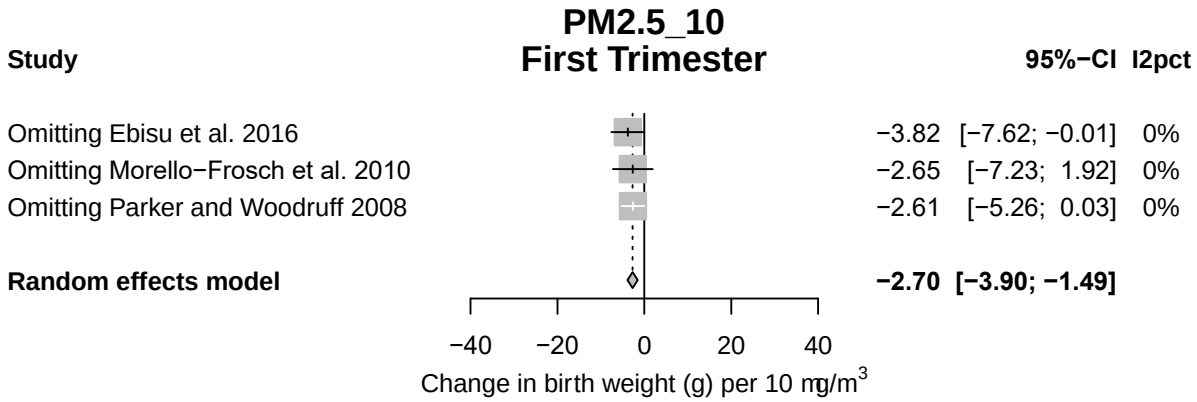
B9.7 PM_{10} (third trimester, low or probably low risk of bias studies)



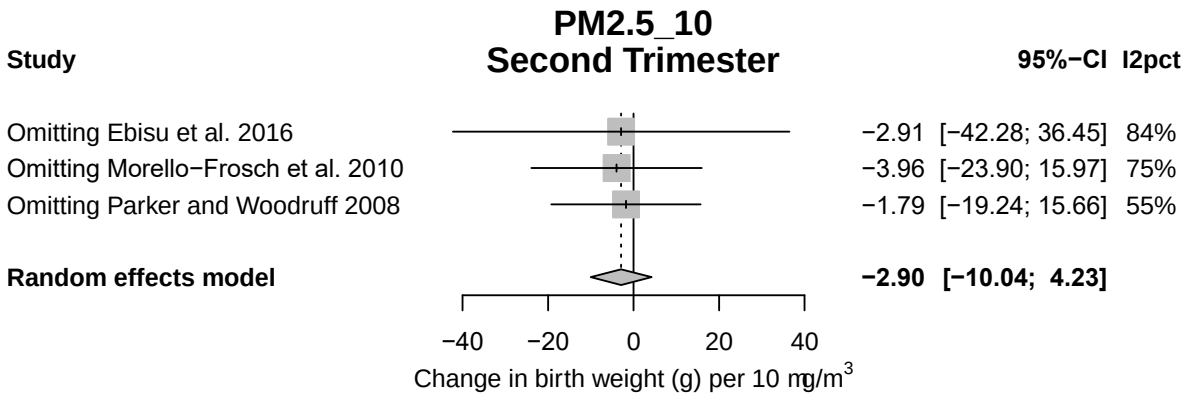
B9.8 PM_{10} (entire pregnancy, low or probably low risk of bias studies)



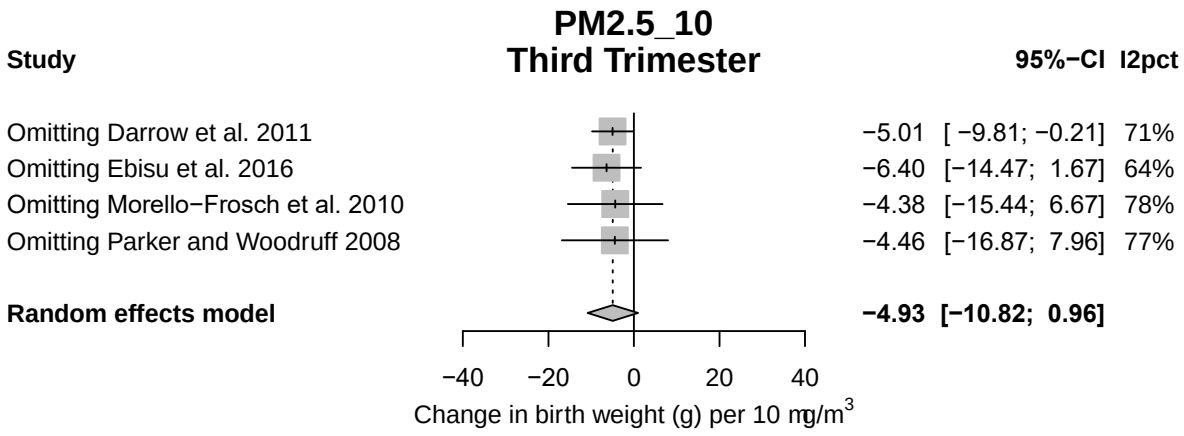
B9.9 $PM_{2.5-10}$ (first trimester)



B9.10 $PM_{2.5\text{-}10}$ (second trimester)



B9.11 $PM_{2.5\text{-}10}$ (third trimester)



B9.12 $PM_{2.5\text{-}10}$ (entire pregnancy)

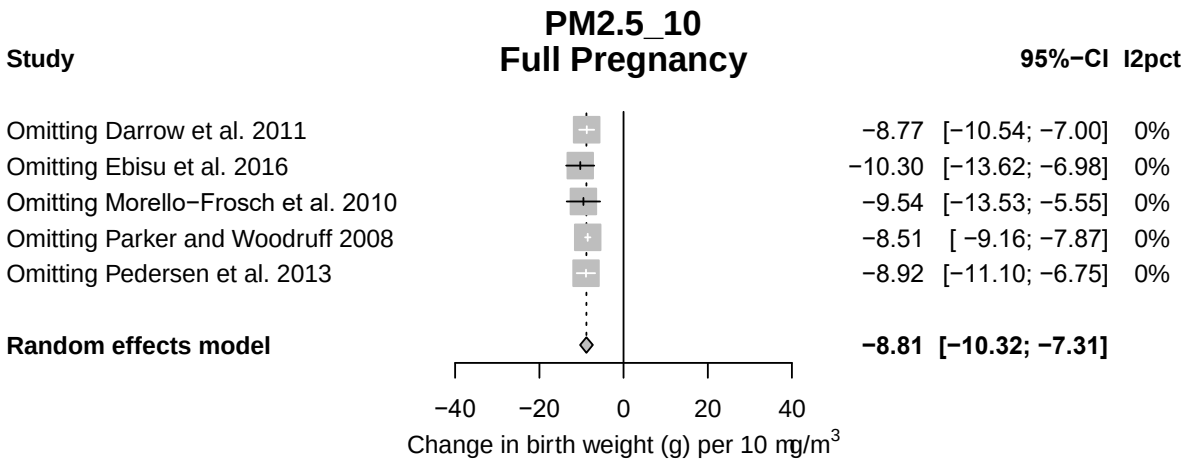


Figure B10. Summary of Risk of Bias judgments across domains for all included studies.