UNIVERSITY OF MILAN

EPIGET - Epidemiology, Epigenetics and Toxicology Laboratory

Department of Clinical Sciences and Community Health

PhD in Biomedical Statistics

STATISTICAL METHODS TO ASSESS THE SUSCEPTIBILITY TO PARTICULATE MATTER AND HEALTH EFFECTS MEDIATED BY microRNAs CARRIED IN PLASMA EXTRACELLULAR VESICLES

Cicle XXVIII

PhD thesis of: Dott.ssa Laura Angelici

Supervisor: Dott. Federico Ambrogi

Date: 11-12-2015

Academic year 2014-2015
Ogniqualvolta una teoria ti sembra essere l’unica possibile, prendilo come un segno che non hai capito né la teoria né il problema che si intendeva risolvere.

*Karl Popper*

**Ringraziamenti**

Al termine di questi tre anni di dottorato desidero ringraziare tutte le persone che a vario titolo mi hanno accompagnato in questo percorso e senza le quali questo lavoro di tesi non sarebbe stato possibile realizzare.

Innanzitutto ringrazio la Prof.ssa Valentina Bollati per avermi dato l’opportunità di frequentare il corso di dottorato e per il suo contributo teorico e metodologico durante tutte le fasi del mio lavoro di ricerca.

La mia gratitudine va anche al laboratorio EPIGET - Epidemiology, Epigenetics and Toxicology Lab, e in particolare alle mie colleghe biologhe e statistici per il supporto, la pazienza, e l’aiuto.

Ringrazio anche il mio tutor Dott.re Federico Ambrogi per il supporto teorico e metodologico.

Un ringraziamento speciale lo dedico poi al Prof.sore Pier Alberto Bertazzi per l’esempio che è stato per me di intelligenza, correttezza, amore per la ricerca e professionalità che costituirà sempre per me un modello da perseguire nella vita e nel lavoro.

Ringrazio immensamente mio marito Fulvio per l’aiuto in famiglia, i sacrifici sostenuti nell’accompagnarmi in questi tre anni e per non avermi mai fatto mancare il proprio sostegno morale.

Infine ringrazio la mia piccolina Sara per essere stata così buona in assenza della sua mamma, e mi scuso con lei per il tempo sottrattole.
# Table of Contents

## Abstract

1. **INTRODUCTION** .................................................................................................................. 9
   1.1 Air pollution .......................................................................................................................... 9
   1.2 Particulate Matter.................................................................................................................. 11
   1.3 PM and Cardiovascular Health Effects................................................................................. 12
   1.4 Extracellular Vesicles (EVs) and microRNAs in PM-related cardiovascular effects.... 14

2. **AIM AND HYPOTHESES** ................................................................................................. 17

3. **METHODS** ......................................................................................................................... 19
   3.1 Study design ....................................................................................................................... 19
   3.2 Study Population ............................................................................................................... 19
   3.3 Two-stage, split sample study design ................................................................................. 21
   3.4 Laboratory Methods ......................................................................................................... 21
   3.5 Air pollution exposure assessment ..................................................................................... 24

4. **STATISTICAL ANALYSIS** ................................................................................................. 29
   4.1 Ambient exposure attribution ............................................................................................ 29
   4.2 miRNAs expression data analysis ...................................................................................... 33
   4.3 Association between miRNAs expression and PM$_{10}$ exposure .................................... 34
   4.4 Mediation Analysis ............................................................................................................. 40
      4.4.1 The Simple Mediation Model ...................................................................................... 40
      4.4.2 Statistical Inferences ................................................................................................... 43
      4.4.3 The Parallel Multiple Mediator Model ........................................................................ 51
5. RESULTS .................................................................................................................................................. 59

5.1 Study Population description at screening Stage................................................................................. 59

5.2 Comparison between FARM model and monitoring station PM$_{10}$ levels.................. 61

5.3 Comparison between Normalization Strategies............................................................................... 66

5.4 Association between miRNAs expression and PM$_{10}$ exposure .................................................. 69

5.5 Mediation Analysis to investigate the role of miRNAs expression as mediator of the effect of PM$_{10}$ on respiratory, cardiac and inflammatory outcome.............................................. 76

5.5.1 Simple Mediation Models.................................................................................................................. 76

5.5.2 Parallel Multiple Mediation Models.................................................................................................. 104

6. DISCUSSION AND FUTURE DIRECTIONS ......................................................................................... 122

7. APPENDIX 1: RESULTS .......................................................................................................................... 126

8. APPENDIX 2: SAS MACRO PROCESS FOR MEDIATION ANALYSIS ........................................... 158
ABSTRACT

Air pollution exposure is a major problem worldwide and has been linked to many diseases. PM$_{10}$ is one of the components of air pollution and it includes a mixture of compounds. Several studies suggest that PM produces significant effects on respiratory and cardiovascular system, in relation to acute as well as chronic exposure. This process has been extensively studied, but to date it has not yet been fully understood. Ambient particles have been shown to produce a strong inflammatory reaction, and beside pro-inflammatory mediators, cell-derived membrane Extracellular Vesicles (EVs) are also released. EVs (particularly microvesicles) might be the ideal candidate to mediate the effects of air pollution, since potentially they could transfer miRNAs, after internalization within target cells through surface-expressed ligands, enabling intercellular communication in the body. Another gap in our current knowledge regarding PM-related health effects is the identification of susceptible subjects. Recent research findings pointed out obesity as a susceptibility factor to the adverse effects of PM exposure partly due to an increase in particle absorption. According these findings, our hypothesis is that, EVs might be the ideal candidate mechanism to mediate the effects of air pollution, since potentially they could be produced by the respiratory system, reach the systemic circulation and lead to the development of endothelial dysfunction. Moreover, EVs after internalization within target cells through surface-expressed ligands, may transfer miRNAs enabling intercellular communication in the body. Finally, obese individuals might represent one of the best population to investigate the effects of environmental air particles on several molecular mechanisms and, as a final objective, on cardiovascular and respiratory parameters. The main proposal of this research project is to develop the appropriate statistical methodology to address the following specific aims:

- **Aim 1.** Determine whether exposure to air particles and PM-associated metals can modify EVs in plasma in terms of miRNAs content.
- **Aim 2.** Determine whether the changes found in EVs (Aim 1) are associated with respiratory, cardiac and inflammatory outcomes such as: single breath carbon monoxide diffusing capacity DLcoRapp, Forced expiratory volume in the 1st second FEV1, Forced Vital Capacity FVC, Heart Rate, Sistolic Blood Pressure SBP, Diastolic Blood Pressure DBP, C-Reactive Protein CRP, and Fibrinogen.
- **Aim 3.** Investigate the potential role of miRNAs as mediators of the effect of PM$_{10}$ exposure on respiratory, cardiac and inflammatory outcomes listed in Aim2.
We used a cross-sectional study investigating the effects of particulate air pollution on a population of susceptible overweight/obese subjects, recruited in Lombardy Region, Italy. The population study will include 2000 overweight/obese (BMI between 25 and 29.9 is considered overweight and an adult who has a BMI of 30 or higher is considered obese subjects, recruited at the Center for Obesity and Weight Control (Department of Environmental and Occupational Health, University of Milan and IRCCS Fondazione Ca’Granda – Ospedale Maggiore Policlinico). We will follow a two-stage, split sample study design. The first (discovery) stage involves genome-wide miRNA expression profiling, by means of OpenArray technology, among 1000 of the aforementioned 2000 participants (the first 1000 subjects consecutively recruited at the Center for Obesity and Weight Control). The second (replication) stage involves a replication analysis of the top 10 miRNAs that resulted from the first stage. At December 31, 2013 (first stage) we recruited 1303 subjects, 87% of whom living in the province of Milan. At April 2015 we recruited a total of 1786 evaluable subjects. Due to technical problems the replication data were not available for statistical analysis at the time of the layout of the thesis.

Different normalization strategies on miRNAs expression data were evaluated and compared in different set of miRNAs: Endogenous U6, Global Mean and Mean of 4 more stable miRNAs. The performance of the different normalization strategies was assessed by: (1) evaluating their ability to reduce the experimental induced (technical) variation, (2) determining their power to extract true biological variation. We showed for large scale miRNA expression profiling Global Mean normalization strategy outperforms the other normalization strategy in terms of:

- better reduction of technical variation:
  - lower % of miRNAs differentially expressed before and after FDR adjustment
  - lower Fold change range;
- more accurate appreciation of biological changes.
  - higher % of miRNAs differentially expressed before and after FDR adjustment;
  - higher Fold Change range;

PM$_{10}$ exposure assessment is based on daily PM$_{10}$ concentration estimates by the FARM model (the flexible air quality regional model), a three-dimensional Eulerian grid model for dispersion, transformation and deposition of particulates, capable to simulate PM$_{10}$ concentration. By means of ArchGis software the residential address of each subject was georeferenced and the resulting map was superimposed on the map of FARM Model. In this way to each subject was attributed: (a) the estimated daily exposure of the cell containing their residential address; (b) the exposure of the cell containing the address of the Center for Obesity and Work; (c) the daily
average exposure for Milan, calculated as the average of the 22 cells that falls into the city boundaries.

Since in each run of OpenArray were simultaneous analysed up to 4 OpenArray plates, identified by a barcode, for a total of 12 samples (3 per plate) it was possible identify an hierarchical data structure with three levels: sample level (level-1), barcode level (level-2) and run level (level-3). In order to verify the association between miRNAs expression and PM\textsubscript{10} we developed a three-levels hierarchical linear model (HLM) using the MIXED procedure in SAS. The following list of first 10 top miRNAs were identified: miR\textsubscript{106a}_002169, miR\textsubscript{152}_000475, miR\textsubscript{181a}_2\_002317, miR\textsubscript{218}_000521, miR\textsubscript{27b}_000409, miR\textsubscript{30d}_000420, miR\textsubscript{652}_002352, miR\textsubscript{92a}_000431, miR\textsubscript{25}_000403, miR\textsubscript{375}_000564. Simple mediation models were applied in order to investigate the role of miRNAs expression as potential mediator on the effect of PM\textsubscript{10} on respiratory, cardiac and inflammatory outcomes such as: single breath carbon monoxide diffusing capacity DL\textsubscript{co}, Forced expiratory volume in the 1st second FEV\textsubscript{1}, Forced Vital Capacity FVC, Heart Rate, Sistolic Blood Pressure SBP, Diastolic Blood Pressure DBP, C-Reactive Protein CRP, and Fibrinogen. 95% BC bootstrap Confidence intervals for Indirect effect were estimated.

Finally, Multiple Parallel mediation models were applied in order to investigate the role of a set of miRNAs expression identified by means of simple mediation models as potential set of parallel mediator on the effect of PM\textsubscript{10} on respiratory, cardiac and inflammatory outcomes. A significant indirect effect of PM\textsubscript{10} on:

- DL\textsubscript{co}Rapp, was found through the following mediators: mir\textsubscript{106a}_002169, mir\textsubscript{152}_000475, mir\textsubscript{218}_000521 expression;
- FEV\textsubscript{1}Rapp was found through the following mediators: mir\textsubscript{27b}_000409 mir\textsubscript{30d}_000420 mir\textsubscript{92a}_000431 mir\textsubscript{181a}_2\_002317 mir\textsubscript{218}_000521 expression;
- FVC\textsubscript{R} was found through the following mediators: mir\textsubscript{27b}_000409, mir\textsubscript{92a}_000431 and mir\textsubscript{181a}_2\_002317 expression;
- Heart Rate was found through the following mediator: mir\textsubscript{218}_000521 expression;
- Sistolic Blood Pressure was found through the following mediator: mir\textsubscript{92a}_000431 expression;
- CRP was found through the following mediator: mir\textsubscript{106a}_002169 and mir\textsubscript{652}_002352 expression.
- Fibrinogeno was found through the following mediator: mir\textsubscript{375}_000564 expression.

Finally, the total indirect effect of PM\textsubscript{10} exposure:
on DLcoRapp obtained summed the indirect effects across all mediators: mir_106a_002169, mir_152_000475, and mir_218_000521 expression is statistically different from zero;

- on FEV1Rapp obtained summed the indirect effects across all mediators: mir_27b_000409, mir_30d_000420, mir_92a_000431, mir_181a_2_002317, mir_218_000521 expression is statistically different from zero;

- on FVC Rapp obtained summed the indirect effects across all mediators mir_27b_000409, mir_92a_000431 and mir_181a_2_002317 expression is statistically different from zero;

- on CRP obtained summed the indirect effects across all mediators mir_106a_002169 and mir_652_002352 expression is statistically different from zero.
1. INTRODUCTION

1.1 Air pollution

Air pollution exposure is a major problem worldwide and has been linked to many diseases. Air pollution consists of gas and particle contaminants that are present in the atmosphere. Gaseous pollutants include SO\textsubscript{2}, NO\textsubscript{x}, ozone, carbon monoxide (CO), volatile organic compounds (VOCs), certain toxic air pollutants, and some gaseous forms of metals. Particle pollution (PM\textsubscript{2.5} and PM\textsubscript{10}) includes a mixture of compounds. The majority of these compounds can be grouped into five categories: sulfate, nitrate, elemental (black) carbon, organic carbon, and crustal material (Table 1).

Some pollutants are released directly into the atmosphere. These include gases, such as SO\textsubscript{2}, and some particles, such as crustal material and elemental carbon. Other pollutants are formed in the air. Ground-level ozone forms when emissions of NO\textsubscript{x} and VOCs react in the presence of sunlight. Similarly, some particles are formed from other directly emitted pollutants. For example, particle sulfates result from SO\textsubscript{2} and ammonia (NH\textsubscript{3}) gases reacting in the atmosphere. Commonly, emissions come from large stationary fuel combustion sources (such as electric utilities and industrial boilers), industrial and other processes (such as metal smelters, petroleum refineries, cement kilns, manufacturing facilities, and solvent utilization), and mobile sources including highway vehicles and non-road sources (such as recreational and construction equipment, marine vessels, aircraft, and locomotives). Sources emit different combinations of pollutants. For example, electric utilities release SO\textsubscript{2}, NO\textsubscript{x}, and Particles Fossil fuel combustion is the primary source contributing to CO\textsubscript{2} emissions. Major sources of fossil fuel combustion include electricity generation, transportation (including personal and heavy-duty vehicles), industrial processes, residential, and commercial [1].
### Table 1: Health, Environmental, and Climate Effects of Air Pollution Components. *U.S Environmental Protection Agency's Report 2010*

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Health Effects</th>
<th>Environmental and Climate Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ozone (O&lt;sub&gt;3&lt;/sub&gt;)</strong></td>
<td>Decreases lung function and causes respiratory symptoms, such as coughing and shortness of breath; aggravates asthma and other lung diseases leading to increased medication use, hospital admissions, emergency department (ED) visits, and premature mortality.</td>
<td>Damages vegetation by visibly injuring leaves, reducing photosynthesis, impairing reproduction and growth, and decreasing crop yields. Ozone damage to plants may alter ecosystem structure, reduce biodiversity, and decrease plant uptake of CO&lt;sub&gt;2&lt;/sub&gt;. Ozone is also a greenhouse gas that contributes to the warming of the atmosphere.</td>
</tr>
<tr>
<td><strong>Particulate Matter (PM)</strong></td>
<td>Short-term exposures can aggravate heart or lung diseases leading to symptoms, increased medication use, hospital admissions, ED visits, and premature mortality; long-term exposures can lead to the development of heart or lung disease and premature mortality.</td>
<td>Impairs visibility, adversely affects ecosystem processes, and damages and/or soils structures and property. Variable climate impacts depending on particle type. Most particles are reflective and lead to net cooling, while some (especially black carbon) absorb energy and lead to warming. Other impacts include changing the timing and location of traditional rainfall patterns.</td>
</tr>
<tr>
<td><strong>Lead (Pb)</strong></td>
<td>Damages the developing nervous system, resulting in IQ loss and impacts on learning, memory, and behavior in children. Cardiovascular and renal effects in adults and early effects related to anemia.</td>
<td>Harms plants and wildlife, accumulates in soils, and adversely impacts both terrestrial and aquatic systems.</td>
</tr>
<tr>
<td><strong>Oxides of Sulfur (SO&lt;sub&gt;x&lt;/sub&gt;)</strong></td>
<td>Aggravate asthma, leading to wheezing, chest tightness and shortness of breath, increased medication use, hospital admissions, and ED visits; very high levels can cause respiratory symptoms in people without lung disease.</td>
<td>Contributes to the acidification of soil and surface water and mercury methylation in wetland areas. Causes injury to vegetation and local species losses in aquatic and terrestrial systems. Contributes to particle formation with associated environmental effects. Sulfate particles contribute to the cooling of the atmosphere.</td>
</tr>
<tr>
<td><strong>Oxides of Nitrogen (NO&lt;sub&gt;x&lt;/sub&gt;)</strong></td>
<td>Aggravate lung diseases leading to respiratory symptoms, hospital admissions, and ED visits; increase susceptibility to respiratory infection.</td>
<td>Contributes to the acidification and nutrient enrichment (eutrophication, nitrogen saturation) of soil and surface water. Leads to biodiversity losses. Impacts levels of ozone, particles, and methane with associated environmental and climate effects.</td>
</tr>
<tr>
<td><strong>Carbon Monoxide (CO)</strong></td>
<td>Reduces the amount of oxygen reaching the body’s organs and tissues; aggravates heart disease, resulting in chest pain and other symptoms leading to hospital admissions and ED visits.</td>
<td>Contributes to the formation of CO&lt;sub&gt;2&lt;/sub&gt; and ozone, greenhouse gases that warm the atmosphere.</td>
</tr>
<tr>
<td><strong>Ammonia (NH&lt;sub&gt;3&lt;/sub&gt;)</strong></td>
<td>Contributes to particle formation with associated health effects.</td>
<td>Contributes to eutrophication of surface water and nitrate contamination of ground water. Contributes to the formation of nitrate and...</td>
</tr>
</tbody>
</table>
1.2 Particulate Matter

PM is a mixture of suspended particles that vary in chemical composition and size. Particulate matter varies in number, size, shape, surface area, chemical composition and characteristics, and can be of both natural and anthropological origin. The size distribution of urban ambient particle pollution is usually characterized by aerodynamic diameter (AED) which is defined as the diameter of a sphere of unit density (lg/cm$^3$) that has the same inertial properties in the gas as the particle of interest. The size distribution of suspended particles in the atmosphere is tri-modal and includes coarse particles, line particles, and ultrafine particles. Coarse particles (PM2,5-10: often defined as those with an aerodynamic diameter > 2.5μm) are primarily produced by natural mechanisms such as grinding and wind, and derive mainly from soil and other crustal materials. Fine particles (PM$_{2.5}$: diameter < 2.5 μm ) are derived chiefly from combustion processes. In the urban environment these process are associated with transportation, manufacturing, and power generation. Ultrafine particles are often defined as particles less than or equal to 0,1 μm.

Particles can either be directly emitted into the air (primary PM) or be formed in the atmosphere from gaseous precursors such as sulfur dioxide, oxides of nitrogen, ammonia and non-methane volatile organic compounds (secondary particles). Primary PM and the precursor gases can have both man-made (anthropogenic) and natural (non-anthropogenic) sources. Anthropogenic sources include combustion engines (both diesel and petrol), solid-fuel (coal, lignite, heavy oil and biomass) combustion for energy production in households and industry, other industrial activities (building, mining, manufacture of cement, ceramic and bricks, and smelting), and erosion of the pavement by road traffic and abrasion of brakes and tires. Agriculture is the main source of ammonium. Secondary particles are formed in the air through chemical reactions of gaseous pollutants. They are products of atmospheric transformation of nitrogen oxides (mainly emitted by traffic and some industrial processes) and sulfur dioxide resulting from the combustion of sulfur-containing fuels. Secondary particles are mostly found in fine PM.

PM$_{10}$ and PM$_{2.5}$ include inhalable particles that are small enough to penetrate the thoracic region of the respiratory system. The health effects of inhalable PM are well documented. They are due to exposure over both the short term (hours, days) and long term (months, years) and include:

- respiratory and cardiovascular morbidity, such as aggravation of asthma, respiratory symptoms and an increase in hospital admissions;
- mortality from cardiovascular and respiratory diseases and from lung cancer.

There is good evidence of the effects of short-term exposure to PM$_{10}$ on respiratory health, but for mortality, and especially as a consequence of long-term exposure, PM$_{2.5}$ is a stronger risk factor than the coarse part of PM$_{10}$ (particles in the 2.5–10 μm range). All-cause daily mortality
is estimated to increase by 0.2–0.6% per 10 μg/m³ of PM$_{10}$ [2, 3]. Long-term exposure to PM$_{2.5}$ is associated with an increase in the long-term risk of cardiopulmonary mortality by 6–13% per 10 μg/m³ of PM$_{2.5}$ [4-6]. Susceptible groups with pre-existing lung or heart disease, as well as elderly people and children, are particularly vulnerable. For example, exposure to PM affects lung development in children, including reversible deficits in lung function as well as chronically reduced lung growth rate and a deficit in long term lung function [7]. There is no evidence of a safe level of exposure or a threshold below which no adverse health effects occur. The exposure is ubiquitous and involuntary, increasing the significance of this determinant of health. At present, at the population level, there is not enough evidence to identify differences in the effects of particles with different chemical compositions or emanating from various sources. It should be noted, however, that the evidence for the hazardous nature of combustion-related PM (from both mobile and stationary sources) is more consistent than that for PM from other sources [8]. The black carbon part of PM$_{2.5}$, which results from incomplete combustion, has attracted the attention of the air quality community owing to the evidence for its contribution to detrimental effects on health as well as on climate. Many components of PM attached to black carbon are currently seen as responsible for health effects, for instance organics such as PAHs that are known carcinogens and directly toxic to the cells, as well as metals and inorganic salts.

### 1.3 PM and Cardiovascular Health Effects

According to the WHO, diseases of heart and cardiovascular system are the largest single cause of death (accounting for about 3.7 million deaths) in the European Union. Cardiovascular diseases account for the largest number of premature deaths before 75 years of age [9]. Numerous health studies have shown acute [10-15] and chronic [16-18] particulate air pollution exposures to be associated with early death, particularly from cardiovascular and respiratory disease [10, 11, 19]. Metals, which are constituents of particulate air pollution, are also associated with CVD [19-38]. Epidemiological and animal studies have suggested many potential mechanisms by which particles may impact health. Airway or parenchymal inflammatory responses to particulate matter (PM) have been hypothesized to be the inciting events of a cascade of pathophysiologic changes in autonomic cardiac, systemic inflammation, and hemostatic activities. All these processes may ultimately lead to the acute cardiac events associated with PM exposure [39]. However, the relative importance of each potential pathways and the steps along these pathways are not well understood, particularly as to how they relate to specific particle components and sources, for which biological pathways are likely to differ. One of the most important gaps in our current knowledge regarding PM-related health effects is the
identification of susceptible subjects [40]. Recent research findings pointed out obesity as a susceptibility factor to the adverse effects of PM exposure partly due to an increase in particle absorption [40, 41]. A positive correlation between exhaled nitric oxide, a marker of pulmonary inflammation, and Body Mass Index (BMI) have been shown in healthy adults [42]. BMI was associated with a graded increase in the estimated total lung dose of deposited fine particles in an inhalation study of healthy children (6–13 years of age) [43]. In a panel study of 44 senior citizens, vascular inflammatory response (measured by C-reactive protein) to ambient levels of PM$_{2.5}$ (particulate matter with aerodynamic diameter $\leq$2.5μm) averaged over 1–7 days was greater in obese (BMI $\geq$ 30 kg/m$^2$) than in non-obese subjects [44]. Moreover, a differential autonomic cardiac response (measured as heart rate variability) to metal particulates have been observed between obese and non-obese individuals [40]. These findings suggest that obese individuals might represent a candidate population to investigate PM mediated cardiovascular effects and related pathogenetic mechanisms. The mechanisms linking PM exposure and cardiovascular disease have not yet been fully elucidated. It has been proposed that inhaled fine particulate matter translocates directly into systemic circulation through the pulmonary capillary bed, promoting atherothrombosis by breaching endothelial integrity and inciting a local inflammatory reaction [45].

![Proposed mechanism for air pollution effects on exosomes release and cell-to-cell communication.](image)

Figure 1: Proposed mechanism for air pollution effects on exosomes release and cell-to-cell communication.
However, just a very small fraction of these fine and ultrafine particles accumulate in extra pulmonary organs such as the liver and the spleen, [46] and currently there is no final evidence that fine particles physically enter and deposit in blood vessels. An alternative hypothesis is that ambient particles produce a strong inflammatory reaction in the lungs followed by the release of proinflammatory mediators that are able to reach the systemic circulation (Figure 1) [47, 48]. In spite of more than two decades of mechanistic research, at front of high degree of consistency of the epidemiology findings showing increased cardiovascular risk, the evidence on intermediate mechanisms remains moderate or weak [49]. In addition, inflammatory and oxidative responses have little specificity and can be activated by a multitude of trigger, thus limiting our capability to correlate them to air pollution exposure as well as to provide the groundwork for the development of targeted biomarkers and prevention strategies. Beside the release of proinflammatory mediators, cell-derived membrane vesicles are also released, representing another new mode of intercellular communication that has recently become the subject of increasing interest [48].

1.4 Extracellular Vescicles (EVs) and microRNAs in PM-related cardiovascular effects

Intercellular communication is an essential hallmark of multicellular organisms and can be mediated through direct cell–cell contact or transfer of secreted molecules. In the last two decades, a third mechanism for intercellular communication has emerged that involves intercellular transfer of extracellular vesicles (EVs). EVs might be the ideal candidate to mediate the effects of air pollution, since potentially they could be produced by the respiratory system, [50, 51] reach the systemic circulation [52] and lead to the development of endothelial disfunction [53]. EVs are spherical structures limited by a lipid bilayer that can be generated by cells and secreted into the extracellular space and are likely composed of both exosomes (EXs) and microvesicles (MVs). There are various types of secreted membrane microvesicles that have distinct structural and biochemical properties depending on their intracellular site of origin, and these features probably affect their function. Microparticles originating from platelets, endothelial cells and monocytes have been most extensively studied [54]. Platelet microparticles were originally studied because of their procoagulant activity [54] and recent studies have investigated their involvement in the pathophysiology of vascular disorders [54]. They could also participate in a defensive shedding of complement attack complexes [55] or in deployment of immunomodulating activities [56]. The term exosome is used to identify a particular subgroup of vesicles, ranging from 40 to 100 nanometers (nm), released as a consequence of multivesicular endosome (a membrane-bound intracellular vesicle, containing EXs) fusion with the plasma
membrane [57], whereas the term microvesicle is used for those EVs, larger than 100 nm, that are shed from the plasma membrane. EVs are released from cell membranes by triggers such as endotoxin encounter, hypoxia or oxidative stress conditions, cytokines release, thrombin production [58] and could be one of the means used by tissues to adapt to these stimuli [59]. In particular, DNA damaging conditions have been related to activation of the p53 that leads to increased exosomes secretion [60]. EV membranes are enriched in molecules characteristic of their parent cell and express adhesion molecules on their surface (i.e., ICAM1), which could favor their capture by recipient cells. The fate of exosomes after binding the surface of recipient cells is not known but recent evidence suggests that they might fuse with recipient cell membranes and deliver their content directly into the cytoplasm of the recipient cells. It has been suggested that EVs, after internalization in the target cells through surface-expressed ligands, may transfer microRNAs (miRNAs) [61, 62] allowing intercellular and inter-organ communication in the body [62]. Moreover, miRNA expression in circulating exosomes has been detected also in plasma of normal subjects and a predictive role of peripheral blood miRNA signatures in human disease has been also hypothesized. MiRNAs are small, endogenous, single stranded noncoding RNAs of 20-22 nucleotides [63] that post-transcriptionally regulate gene expression by either triggering mRNA cleavage or repressing translation [64]. One single miRNA can regulate hundreds of mRNAs in interrelated gene pathways and a single mRNA can be targeted by several different miRNAs [65]. Moreover, miRNA expression in circulating EVs has been detected also in plasma of normal subjects and a predictive role of peripheral blood miRNA signatures in human diseases has been also hypothesized [62]. Changes in the expression of several miRNAs have been implicated in disease mechanisms that may be related to PM exposure such as oxidative stress [66] and regulation of inflammation [67]. Recently, our group showed in foundry workers of an electric-steel plant facility that air particles, particularly those rich in lead and cadmium, are able to modify miRNAs expression in blood [68].
2. **AIM AND HYPOTHESES**

According these findings, our hypothesis is that obese individuals might represent one of the best population to investigate the effects of environmental air particles on several molecular mechanisms and, as a final objective, on cardiovascular and respiratory parameters. EVs might be the ideal candidate mechanism to mediate the effects of air pollution, since potentially they could be produced by the respiratory system, reach the systemic circulation and lead to the development of endothelial dysfunction. Moreover, EVs after internalization within target cells through surface-expressed ligands, may transfer miRNAs enabling intercellular communication in the body. The main proposal of this research project is to develop the appropriate statistical methodology to address the following specific aims:

- **Aim 1.** Determine whether exposure to air particles and PM-associated metals can modify EVs in plasma in terms of miRNAs content.

- **Aim 2.** Determine whether the changes found in ECVs (Aim 1) are associated with respiratory, cardiac and inflammatory outcomes such as: single breath carbon monoxide diffusing capacity DLco, Forced expiratory volume in the 1st second FEV1, Forced Vital Capacity FVC, Heart Rate, Sistolic Blood Pressure SBP, Diastolic Blood Pressure DBP, C-Reactive Protein CRP, and Fibrinogen.

- **Aim 3.** Investigate the potential role of miRNAs as mediators of the effect of PM$_{10}$ exposure on respiratory, cardiac and inflammatory outcomes listed in Aim2.
3. METHODS

3.1 Study design

The SPHERE study is a cross-sectional study investigating the effects of particulate air pollution on a population of susceptible overweight/obese subjects, recruited in Lombardy Region, Italy. Lombardy is situated in the Northern part of Italy and is divided physically into three parts from north to south: Alpine and pre-Alpine mountains, foothills and a zone of alluvial plains sloping to the Po river. The region covers an area of 23,864 km² with a population of about 10 million people [69] and consists of 12 provinces, among which Milan is the regional capital. The Milan metropolitan area is home to 7 million inhabitants with 1.3 million residing in the core municipality [70].

3.2 Study Population

The population study will include 2000 overweight/obese (BMI between 25 and 29.9 is considered overweight and an adult who has a BMI of 30 or higher is considered obese subjects, recruited at the Center for Obesity and Weight Control (Department of Environmental and Occupational Health, University of Milan and IRCCS Fondazione Ca’Granda – Ospedale Maggiore Policlinico). This study population has been chosen because some evidence shows that obesity may bring greater susceptibility [71, 72] to the adverse cardiovascular effects of PM exposure. In addition, it has been shown that obese people are more susceptible to the effect of inflammation in the short-term exposure to PM, presenting higher levels of C-reactive protein and interleukin-6, compared with normal-weight subjects, at the same exposure of PM level [44]. The recruitment period started in September 2010 and will continue until the end of 2015. The eligibility criteria for participants are as follows: 1) older than 18 years at enrollment; 2) obese/overweight according to the following definition: overweight is defined as a BMI between 25 and 30 kg/cm², obesity is defined as a BMI of 30 kg/cm² or more; 4) domiciled in Lombardy at the time of the recruitment; 5) agreement to sign an informed consent and donate a blood sample. Exclusion criteria include: experienced previously diagnosed cancer, heart disease or stroke in the last year or other chronic diseases such as multiple sclerosis, Alzheimer’s disease, Parkinson’s disease, depression, bipolar disorder, schizophrenia and epilepsy.

Epidemiological and Clinical Data collection

At recruitment, each study subject is asked to:

- fill in a lifestyle and a diet questionnaire,
- donate a 15 ml blood sample (for molecular tests),
- provide a 50 ml urine sample (for metal internal dose assessment),
- provide a lock of hair cut next to the root in the occipital area of the head (for metal internal dose assessment)

As part of the routine protocol, for each subject presenting at the Center, extensive physical examination, spirometry, ECG are performed and biochemical tests are also collected, including Emocrome, Fibrinogen, C-reactive protein, Total cholesterol, HDL, LDL, Triglyceride, Serum creatinine, AST, ALT, Gamma-Glutamyltransferase, Glucose, Homocysteine, TSH, Glycated haemoglobin, Postprandial glycaemia, Insulin level, 2-hour post glucose insulin level, Urinary pH, Uric acid.

**Lifestyle questionnaire**

The lifestyle questionnaire collects information on socio-demographic data, residential area (complete address, characteristics of the house, and traffic), education, past and present health status of the subjects and their first-degree relatives, medications in the last year, employment history and for employed subjects address of the plant of their current work, smoking history, passive smoking at home and at workplace, physical activity levels and sedentary behavior commuting time and transport mode.

**Diet questionnaire**

The questionnaire on eating habits included questions on the number of servings from each food in a usual week or month. Several different types of food were investigated, including: legumes, vegetables, fruits, nuts, red and white meat, fish, eggs, dairy products, cereals, snacks, oil and butter, alcoholic beverage, thea and coffee. Number of servings from each food were translated into usual daily micronutrients intake weighting for serving size, age class and gender. Both questionnaires were checked for completeness at the time of data collection in order to ensure high quality data.

**Lung and cardiac function**

Pulmonary functions are measured, at the same day of blood drawing, with an electronic flow volume spirometer V-max 22 with Autobox (SensorMedics), according to European Respiratory Society/American Thoracic Society guidelines (ERS/ATS 2005) [73]. Tests are performed on patients in the sitting position, and are repeated until at least three reproducible forced expiratory curves have been obtained. Lung function parameters are: forced expiratory volume in one second (FEV1); forced vital capacity (FVC); best peak expiratory flow (PEF); forced expiratory flows at 25%, 50%, and 75% of FVC (FEF25, FEF50, FEF75); mid-expiratory flow (FEF25–
75) derived from the best maneuver (defined as the one with the highest sum of FEV1 + FVC). The single breath carbon monoxide diffusing capacity (DLco) is also measured [74]. All parameters are expressed as a percentage of the predicted normal values, [75] and adjusted for sex, age, height. A resting ECG and rhythm strip is also recorded and blood pressure is measured with the participant supine, after 5 minutes of rest.

### 3.3 Two-stage, split sample study design

We will follow a two-stage, split sample study design. We foresee to collect at least 600 subjects/year. The first (discovery) stage involves genome-wide miRNA expression profiling among 1000 of the aforementioned 2000 participants (the first 1000 subjects consecutively recruited at the Center for Obesity and Weight Control). The second (replication) stage involves a replication analysis of the top 10 miRNAs that resulted from the first stage.

**First stage**

At December 31, 2013 we recruited 1303 subjects, 87% of whom living in the province of Milan. Evaluable patients were 1250. Mean BMI of our study population is 33.5 Kg/cm² (±5.5): nearly 27.8% are overweight, 38.6% are obese, and 33.6% severe obese. The percentage of obese individuals in Lombardy is about 10.3% of the total adult population [76]. The participation rate in the years 2010-2013 was 90%. The study population is composed by 74% of female, mean aged 52 years.

**Second Stage**

At April 2015 we recruited a total of 1786 evaluable subjects. Due to technical problems the replication data were not available at the time for statistical analysis at the time of the layout of the thesis. Statistical analyses results were performed only on data coming from First Stage.

### 3.4 Laboratory Methods

Stage 1 will involve using the OpenArray technology (Applied Bioscience) that allows to run 2,688 TaqMan® qPCR reactions in parallel. In Stage 2, we will replicate the results obtained on the discovery set by standard real time PCR on an Applied Biosystems 7900HT Real-Time PCR System. Laboratory methods can be summarized as follows:

- **Isolations ECVs:** Plasma was thawed on ice and subjected to 3 consecutive centrifuged, this passage has been useful to eliminate dead cells and debris. The supernatant was
subjected to ultracentrifuge and then it was discarded and the pellet remaining were stored at -20 ° C.

- **miRNAs extraction:** miRNAs were extracted using the Qiagen Kit miRNReasy Mini Kit and RNeasyMinEluteClean Up Kit.

- **High throughput miRNAs:** high throughput analysis is the set of scientific analysis able to perform tests on a large number of data in a limited time due to machinery and automated instrumentation. For a complete analysis of all miRNAs we performed a reverse transcription, followed by a reaction of preamplification and then we evaluated the expression of miRNAs through a Real-Time PCR.

- **Reverse transcription:** this technique of reverse transcription allows to produce cDNA starting from RNA and for this reason is a technique that is exploited to study gene expression.

- **Reaction of preamplification:** knowing that the amount of miRNAs contained in the microvesicles extracted from 1.5 mL of plasma is very low, before the Real-time PCR must perform a reaction of preamplification.

- **Real Time PCR:** The product of the reaction of preamplification was transferred from the plate 96 to 384, performing an ottuplicato. Finally, using the Robot AccuFillTM System (Life Technologies) was set up the Open Array Plate formed by 48 subarray with 64 holes, for a total of 3072 holes of 300 μm in diameter and 300 μm of depth (*Figure 2*). Each hole has an outer hydrophobic, while inside it has a hydrophilic coating (the reagents are kept inside by means of surface tension).

![Image](image.png)

*Figure 2: OpenArray Plate.*

In this system, each real time PCR reaction has a final volume of 33 nanoliters. After loading the sample, the plate was covered with adhesive Lid and filled with fluid immersion.
The plate thus prepared was loaded in QuantStudioTM 12K Flex OpenArray (Life technologies) which allows the simultaneous analysis of up to 4 OpenArray plates, for a total of 12 samples (3 per plate). The Real-Time PCR, also called quantitative PCR or real-time quantitative PCR (RTQ-PCR), is a method of amplification (polymerase chain reaction or PCR) and simultaneous quantification of samples for DNA and cDNA. The fluorescence is generated during PCR due to several possible chemical reactions. (Figure 3). The main chemicals are based on the binding of fluorescent dyes that intercalate in the double helix of DNA, such as SYBR Green, or the hybridisation of specific probes, such as TaqMan probes. In this experiment it were used TaqMan probes. The use of probes revealing fluorescence is one of the most reliable and accurate method because the TaqMan probes are designed so as to pair up to specific target sequences. A typical cycle of OpenArrayflex normal PCR and includes 3 phases:
1. Heat denaturation of template DNA (94-99 °C);
2. Annealing to the sequence of the oligonucleotides (30-65 °C);
3. Extension by the DNA polymerase from the heat-resistant primers (65-72 °C).

Figure 3: QuantStudioTM 12K OpenArray and OpenArray flex Real Time PCR cycle.

The real time PCR is quantitative, as the data is collected during the exponential growth phase, when the quantity of the product of the reaction is directly proportional to the initial amount of nucleic acid. The quantification of the miRNA is based on the fluorescence detected at each reaction cycle: every cycle, the cDNA molecules doubled, reaching a plateau when the reactants...
are exhausted and the operation of the enzyme decreases drastically. After finishing the reaction, the instrument software provides results in terms of: Related Cycle Threshold (CRT): the number of PCR cycles at which the efficiency of the reaction Real-Time PCR miRNA target is maximal. The higher the CRT, the lower the expression of miRNA. In order to evaluate the goodness of the reaction a parameter called Ampscore was used. A good amplified has a Ampscore exceeding 1.24. This means that the curve representing its amplification, as a result of the Real-Time PCR, has the characteristic of an exponential growth curve. With this information, determine the Ampscore can also help to discriminate false positives and false negatives.

3.5 Air pollution exposure assessment

Exposure is defined using a multifaceted approach, which integrates information coming from PM$_{10}$ assessment with personal dose of metals in urine and hair. In particular, PM$_{10}$ is assigned to each subject following two approaches: I. use of daily PM$_{10}$ concentration series from air quality monitors; II. use of daily PM$_{10}$ concentration estimates by the FARM model (the flexible air quality regional model), a three-dimensional Eulerian grid model for dispersion, transformation and deposition of particulates, capable to simulate PM$_{10}$ concentration using a 4 km–dispersion grid [15].

1. Air Monitoring Stations data from ARPA Lombardy: we collected recordings of daily PM$_{10}$ data by 81 fixed monitoring stations selected by the Regional Environmental Protection Agency (ARPA Lombardy) from the approximately 150 monitors of the Regional Air Monitoring Network throughout Lombardy region, (Figure 4) on the basis of their reliability determined by standardized quality control procedures and correlation with in-situ measurements, of continuity of recording and of the ability to represent local background air pollution. Three stations are located in the city of Milan ("Verziere", "Pascal-Città Studi" and “via Senato”). We used daily concentration measured by single monitors in the study area to characterize PM$_{10}$ exposure, at the date of recruitment and until 365 days before, for each subject. Thus, we are able to estimate both short- and long-term exposure to the pollutant investigated. As air monitoring station recordings for the area of interest are available from 2001, older exposures might also be estimated. Moreover, as patients were not recruited all at once, but during the first visit to the Center for Obesity and Work, PM$_{10}$ exposure was not tied to a specific date, but randomly assigned during the work week, ensuring a great variability of exposure distributions.
Data on meteorological variables were obtained from the monitoring stations of the Regional Environmental Protection Agency measuring respectively temperature (233 monitors) and relative humidity (163 monitors). (Figure 5) The apparent temperature were calculated using the following formula:

\[ T_{app} = -2.653 + 0.994 \times TEMP + 0.0153 \times DEW^2 \]

where the dew point DEW is defined as:

\[ DEW = \left( \frac{u}{100} \right) \times \left( 112 + (0.9 \times TEMP) \right) + 0.1 \times TEMP - 112 \]
2. **FARM model** (flexible air quality regional model) by ARPA Lombardy: Estimated daily average concentrations of PM$_{10}$ for the years 2007-2012 based on the FARM model were obtained from ARPA Lombardy. The FARM model is a chemical transport model able to treat the main processes of chemical and physical nature of formation and removal of pollutants, in addition to the transport and dispersion by the action of wind and atmospheric mixing. The model provides an estimate of PM$_{10}$ concentration in the atmosphere, from a considerable number of data that can be grouped as follows: information needed to characterize in space and time the emissions in the area of interest, information on meteorological variables involved in the processes of transport and dispersion of pollutants in the air, concentrations of conditions present at the boundary of this area and the beginning of the simulation period. The input data are built from meteorological observations of weather and hydrological network of ARPA and from the processing of the results of the global meteorological model of the European Centre ECMWF (European Centre for Medium-Range Weather Forecasts). The initial and boundary conditions are obtained starting from the data of the ARPA network of air quality and the results of the model CHIMERE (http://www.lmd.polytechniques.fr/chimere/chimere.php) applied at continental scale within the France forecasting system Prev’air. Finally, data on measured and simulated concentrations are harmonized through interpolation techniques [65, 77]. Data elaborated by modeling systems integrates the ones of the monitoring network and allow to know air quality state on an extensive way on the territory.

![Figure 6: Grid 4x4Km from FARM model by ARPA Lombardy applied on the map of Lombardy Region.](image-url)
FARM model allows to assess the population's risk of exposure to air pollution in regions where there are no direct observations. By this model the Lombardy region is divided into a grid of 1678 cells (4x4 km), each associated with daily PM10 concentration estimates. (Figure 6). Thus, the advantage of FARM dispersion modeling, on monitors, is that it allows to assess the population's risk of exposure to air pollution in regions where there are no direct observations. Data estimated from the models are available until 2012, since the data validation imply a lag time of nearly 6 months and will soon be available for further data analysis. Data estimated from the models are available until 2012, since the data validation imply a lag time of nearly 6 months and will soon be available for further data analysis. Given the time window of available data, will be possible to estimate both short-term (days) and long-term (months, years) exposure to the pollutants investigated at or before the index date (date of recruitment for each subject).
4. STATISTICAL ANALYSIS

Statistical analysis follows different strategies in the two different stages. We proceeded according to the steps explained in the following paragraphs.

4.1 Ambient exposure attribution

Verification and possible correction of address of domicile / residence of subjects was performed as first step. Allocation of geographic coordinates to addresses of domicile / residence was performed using the GPSvisualizer software. ArchGIs was used for the attribution of individual ambient exposure to each subjects of the pilot study. In detail for the previous points I and II:

1. **Air Monitoring Stations data from ARPA Lombardy:** We geocoded the monitoring stations and the addresses of study subjects in order to assign them the daily PM$_{10}$ concentration from: (1) the monitor at the lower distance to home address, defined “subject’s residence”; (2) the nearest monitor to the Center for Obesity and Work (“Verziere”), defined “Policlinico”; (3) daily exposure for Milan created averaging the three available city monitors, defined as “average Milan” (*Figure 7*).

![Figure 7: Attribution of PM10 from exposure Monitoring Stations ARPA to subjects.](image)

*Handling missing data in Air Monitoring Stations data series*

In case of incomplete series, the missing data is estimated based on the values measured on the same day in other stations and the ratio between the annual average recorded in the station that
has the missing data and annual averages of the other stations \((63,64)\). The other stations were chosen among those closest and strongest correlation. The final series has been created averaging over monitors and imputing missing values under proportionality assumptions. Each station and pollutants had missing daily averages data. To fill the gaps for each pollutant \(A\), first the \(q\) stations with yearly data coverage higher than 90\% were selected \((1 \leq q < n)\). Then a daily variability profile \(p_d\) was calculated as follows, based on the observation that relationships between observed concentrations of \(\text{PM}_{10}\) (expressed as Pearson correlation coefficient) at the different selected sites were always higher than 0.7:

\[
p_d = \frac{\sum_{k=1}^{q}[A]_{d,k}}{\sum_{K=1}^{q} \beta_{d,k} \times [A]_{y,k}}
\]

Where:

\[
\beta_{d,k} = 0 \text{ if } [A]_{d,k} = \text{missing}
\]

\[
\beta_{d,k} = 1 \text{ if } [A]_{d,k} = \text{not missing}
\]

\([A]_{d,k}\) = daily average concentration of the pollutant \(A\) measured at the station \(k\).

\([A]_{y,k}\) = yearly average concentration of the pollutant \(A\) measured at the station \(k\).

Then the data series for the \(q\) stations and the missing data days \(j\) were completed as follows:

\([A]_{j,k} = p_j \times [A]_{y,k}\)

To complete the time series of the \(n-q\) stations with yearly data coverage less than 90\%, first an overall pollutants daily average was calculated from the \(q\) stations:

\([A]_{d} = \frac{1}{q} \times \sum_{k=1}^{q}[A]_{d,k}\)

Thus a correlation line was built for the \(z\) \((n-q)\) station with data coverage < 90\%, starting from the \(i\) daily average available:

\([A]_{i,z} = a \times [A]_{i,d} + b\)

The calculated slope \((a)\) and intercept \((b)\) was finally used to estimate the lost values \(j\) at each \(z\) station:

\([A]_{j,z} = \hat{a} \times [A]_{j} + \hat{b}\)
With such an approach it was possible to improve the data coverage to almost 100% for all the monitoring stations selected; only a few days were still missing for PM$_{10}$ (no measurement available at all, 2 days out of 1460 days of observations).

2. **FARM model by ARPA Lombardy:** The residential address of each subject was georeferenced and the resulting map was superimposed on the map of FARM Model (Figure 8). In this way each subject was attributed: (a) the estimated daily exposure of the cell containing their residential address; (b) the exposure of the cell containing the address of the Center for Obesity and Work; (c) the daily average exposure for Milan, calculated as the average of the 22 cells that falls into the city boundaries.

![Figure 8: Attribution of PM$_{10}$ exposure from FARM Model to subjects.](image)

A map of predicted PM$_{10}$ concentration for the study period 2010-2012 is shown in Figure 9. We assume that the day of the visit subjects are exposed to levels of PM$_{10}$ measured by the “Policlinico” station, as the time elapsed in the Hospital to perform all the examinations (approximately five hours) is supposed to be sufficient to experience outcomes related to very short-term effect. About 57% of SPHERE subjects live in the city and an additional 28% work in Milan (even if they live outside the city), overall a 67% of subjects spent many hours a day in the city or travelling from workplace to residence. Thus, longer exposure effects are evaluated by both residential and Milan monitors using appropriate lag time from recruitment date, as sensitivity analysis.
However, very high correlation was observed among the three sources of exposition (RPoliclinico vs Average Milan=0.99; RSubjects’ Residence vs Average Milan=0.94; RPoliclinico vs Subjects’ Residence=0.99). With the intent of giving a map representation of PM$_{10}$ from monitoring stations, we applied the Empirical Bayesian kriging (EBK) to expand the monitors PM$_{10}$ point observation to the whole Lombardy territory. Empirical Bayesian kriging (EBK) is a geostatistical interpolation method that automates the most difficult aspects of building a valid kriging model. Other kriging methods in Geostatistical Analyst require you to manually adjust parameters in order to receive accurate results, but EBK automatically calculates these parameters through a process of subsetting and simulations. Empirical Bayesian kriging also differs from other kriging methods by accounting for the error introduced by estimating the underlying semivariogram [78, 79].

A map of observed PM$_{10}$ mean concentration for the whole study period 2010-2013 is shown in (Figure 10).
We assume that the day of the visit subjects are exposed to levels of PM$_{10}$ measured by the “Policlinico” station, as the time elapsed in the Hospital to perform all the examinations (approximately five hours) is supposed to be sufficient to experience outcomes related to very short-term effect. The use of both monitoring stations and Eulerian model is forced for very recent data, when data estimated from FARM model are not available, while monitors data are.

4.2 miRNAs expression data analysis

Stage 1
The new technology used for real-time PCRs allowed to get a quantitative measurement in terms of Related Cycle Threshold (Crt) of the expression of the entire miRNome. Crt is the number of PCR cycles at which the efficiency of the reaction Real-Time PCR miRNA target is maximal. Thus the starting point was a dataset containing the Crt values of 733 human miRNAs for each subject of the first stage analysis (N=1250). First of all it was applied an automatic miRNAs selection, setting missing values and values with Crt >27 or Amp score >1.24 equal to the detection limit of CRT=28. In this way, miRNAs assumed not to be detected because having a Crt above the threshold of 28 was set at 28. The choice of Crt=28 (lower than the usual value of detection limit used in Real-time PCR) is due to the need of preamplification reaction, the large
number of miRNAs target and low volume of Real Time reaction. Afterwards, in order to try to remove the background noise, three different strategies of miRNAs selection were performed.

- **SET1**: Removing miRNAs completely not expressed (Crt=28) in all samples: 527 miRNAs remaining;
- **SET2**: Removing miRNAs do not expressed (Crt=28) in at least 30% of samples: 152 miRNAs remaining (keeping miRNAs expressed (Crt≠28) in at least 70% of samples);
- **SET3**: Removing miRNAs do not expressed (Crt=28) in at least 10% of samples: 105 miRNAs remaining (keeping miRNAs expressed (Crt≠28) in at least 90% of samples);

On each of these miRNAs sets three different normalization methods were applied. The normalization aimed at removing experiment-specific effects in order to maximize the true biological information contained in expression measures. No consensus exists about which procedure performs best [80, 81], the following normalization methods were applied:

- **Endogenous U6**:
  \[ \text{DELTAC}\text{rt} = \text{Crt} - (\text{Arithmetic Mean Ct of U6 across samples}) \]
- **Global Mean**:
  \[ \text{DELTAC}\text{rt} = \text{Crt} - (\text{Arithmetic Mean Crt of each miRNAs across samples}) \]
- **Mean of 4 more stable miRNAs**:
  \[ \text{DELTAC}\text{RT} = \text{CRT} - (\text{Arithmetic Mean CRT of 4 more stable miRNAs across samples}). \]

Finally, after calculating DELTACRT, relative quantification defined as: \[ \text{RQ} = 2^{(-\text{DELTACRT})} \] were calculated for each normalization methods applied. The performance of the different normalization strategies is assessed by [80, 81]: (1) evaluating their ability to reduce the experimental induced (technical) variation, (2) determining their power to extract true biological variation.

### 4.3 Association between miRNAs expression and PM_{10} exposure

Standard descriptive statistics were used to summarize data. In order to verify the association between miRNAs expression and PM_{10} we first fitted multiple linear regression models. We checked the assumptions of models by means of both graphical inspections and formal test which were particularly useful given the huge amount of miRNA to be tested.

We will validate the "iid" assumption of linear regression by examining the residuals of our final model and testing the heteroscedasticity by means of the White test, moreover the Durbin-Watson statistic was used to test for first order correlation of error terms. We tested the normality assumption of errors using the normal probability plot and the Shapiro-Wilks statistic. Finally
plots of residuals versus predicted values and Lack of Fit tests were used to explore potential nonlinearity. Logarithmic transformation on miRNA expression was performed in order to improve the normality and linearity assumptions. The following adjusting variables were selected a priori, based on previous work investigating associations between miRNAs and particles in foundry workers [68]: age, body mass index, cigarette smoking (never, former, or current), and pack-years. We adjusted for percent of granulocytes (to control for possible shifts in leukocyte differential count), date, Seasonality (using sine and cosine) and apparent temperature. It was assumed that:

- **Exposure:** Is the daily PM$_{10}$ exposure estimate (µg/m$^3$) from Eulerian model for the 4x4 km cell containing the address of the Center for Obesity and Weight Control. The exposure lag period chosen for the analysis is of zero days (daily exposure of blood collection day). The hypothesis underlying this choice is that, the mechanism starting with PM$_{10}$ inhalation, with subsequent inflammatory reaction in lungs, release of ECVs by the respiratory system in systemic circulation, and finally the potential transfer of miRNAs by ECVs, is a short term mechanism. Thus, ideally, the subject arrived at the Center for Obesity and Weight Controls inhaled the exposure estimated by Eulerian model in the 4x4 km cell containing the address of the Centre which triggers the above mechanisms.

- **Outcome:** Is the Relative quantification RQ of each miRNAs. log2 transformation was applied in order to satisfy the normality assumption of linear regression model.

Since in each run of OpenArray were simultaneous analysed up to 4 OpenArray plates, identified by a barcode, for a total of 12 samples (3 per plate) it was possible identify an hierarchical data structure with three levels: sample level (level-1), barcode level (level-2) and run level (level-3) (Figure 11).

![Figure 11: Three levels hierarchical data structure: run level (level-1), barcode level (level-2) and sample level (level-3). In each run of OpenArray were simultaneous analysed up to 4 OpenArray plates, identified by a barcode, for a total of 12 samples (3 per plate).](image)
We developed a three-levels hierarchical linear model (HLM) [83, 84] using the MIXED procedure in SAS. The equations necessary for estimating three-levels mixed effect linear model are:

- **Level-1 i samples**: The level-1 model for three-level HLM, for sample i, barcode j, and run k, can be given as follows:
  \[ Y_{ijk} = \pi_{0jk} + \pi_{1jk} A_{ijk} + e_{ijk} \]  
  \( (Eq.1) \)
  In (Eq.1), \( Y_{ijk} \) is represented as the miRNAs expression of sample i associated with barcode j, and run k. Predictor is \( A_{ijk} \) (say, A). The coefficient \( \pi_{0jk} \) is the intercept, \( \pi_{1jk} \) is the slope for A. Further, \( A_{ijk} \) is a continuous predictor, and is grand-mean centered.
  Grand mean centering was defined previously, under two-level HLM. The term \( e_{ijk} \) represents the random effect for sample i, barcode j, and run k, which is normally distributed with mean zero and variance \( \sigma^2 \).

- **Level-2 j barcode**: The level-2 model is formulated by using level-1 intercept \( \pi_{0jk} \) and slopes \( \pi_{1jk} \) as outcomes. The level-2 equation, where \( X_{ijk} \) is a continuous predictor and grand-mean centered, can be given as follows:
  \[ \pi_{0jk} = \beta_{00k} + \beta_{01k} X_{ijk} + r_{0jk} \]  
  \[ \pi_{1jk} = \beta_{10k} + \beta_{11k} X_{ijk} + r_{1jk} \]  
  \( (Eq.2) \)
  The parameters \( \beta_{00k}, \beta_{10k} \) are level-2 intercepts. Further, the coefficients \( \beta_{01k}, \beta_{11k} \), are level-2 slopes. The terms \( r_{0jk}, r_{1jk} \), and are random effects for barcode j, and run k.

- **Level-3 k barcode**: For run k, the level-3 model can be formulated as follows:
  \[ \beta_{00k} = \gamma_{000} + \gamma_{001} W_{ik} + u_{00k} \]  
  \[ \beta_{01k} = \gamma_{010} + \gamma_{011} W_{ik} + u_{01k} \]  
  \[ \beta_{10k} = \gamma_{100} + \gamma_{101} W_{ik} + u_{10k} \]  
  \[ \beta_{11k} = \gamma_{110} + \gamma_{111} W_{ik} + u_{11k} \]  
  \( (Eq.3) \)
  In the level-three model, the level-2 intercepts and slopes are used as outcomes. The terms \( u_{00k}, u_{02k}, u_{10k}, u_{11k} \) in (Eq.3) are random effects associated with run k. The single-equation can be formulated as follows, by substituting (Eq.3) in (Eq.2), and then substituting the newly produced (Eq.2) in (Eq.1).
  \[ Y_{ijk} = \pi_{0jk} + \pi_{1jk} A_{ijk} + e_{ijk} \]  
  \[ Y_{ijk} = \beta_{00k} + \beta_{01k} X_{ijk} + r_{0jk} + \beta_{10k} A_{ijk} + \beta_{11k} X_{ijk} A_{ijk} + r_{1jk} A_{ijk} + e_{ijk} \]  
  \[ Y_{ijk} = \gamma_{000} W_{ik} + \gamma_{010} X_{ijk} + \gamma_{011} W_{ik} X_{ijk} + \gamma_{100} A_{ijk} + \gamma_{101} W_{ik} A_{ijk} + \gamma_{110} X_{ijk} A_{ijk} + \gamma_{111} W_{ik} X_{ijk} A_{ijk} + u_{00k} + u_{01k} X_{ijk} + u_{10k} A_{ijk} + u_{11k} X_{ijk} A_{ijk} + r_{0jk} + r_{1jk} A_{ijk} + e_{ijk} \]  
  \( (Eq.4) \)
In (Eq.4), the parameter $\gamma_{000}$ is interpreted as the predicted miRNAs expression for a reference sample associated with a reference barcode in a reference run (in this case, we assume all $X$s, $W$s, and $A$s equal to zero). The terms $\gamma_{001}$, $\gamma_{010}$, $\gamma_{100}$, are the simple effects of individual sample, barcode, and run level predictors. The parameters $\gamma_{011}$, $\gamma_{101}$, $\gamma_{110}$, represent simple two-way interaction effects of the sample, barcode, and run level predictor. The terms $\gamma_{111}$, represent the three-way interaction effect on miRNAs expression due to sample, barcode, and run level predictors. Equation (Eq.4) also consists of residual terms associated with all three levels. The level-1 residual, $e_{ijk}$, is the unique effect of sample $i$ on miRNA expression, associated with barcode $j$, and run $k$. Similarly, $r_{0jk}$ is the unique effect of barcode $j$ from run $k$, and $u_{00k}$ is the unique run effect for $k$th run for a reference barcode and reference sample. The terms r’s and u’s are level-2 and level-3 residual terms respectively associated with slopes. The interaction terms of random effect and individual predictor or cross-level predictors are also present in the above model. For example, the term $\{ r_{ijk} A_{ijk} \}$ is the interaction of the unique effect associated with sample A-slope, and $\{ u_{11k} X_{ijk} A_{ijk} \}$ is the three-way interaction between $X$, $A$, and the residual term associated with the slope of $X$.

Assumptions

The following assumptions can be made for three-level HLM.

a) The error terms of each level-1 unit should have a mean of zero, and the error terms should be multivariate normally distributed. If, for example, we consider level-1 and level-2 units as sample and barcode, respectively, then the mean of the error within each barcode should be zero, and these error terms should be multivariate normally distributed.

b) It is assumed that the relationship between predictors and outcome variables, at all three levels, is linear.

c) Another assumption is the homogeneity of variance. That is, all barcodes should have equal variances in the sample.

d) Level-1 predictors are independent of the level-1 error term. In other words, the covariance between the level-1 predictors and the error term should equal zero.

e) Level-2 and level-3 error terms have a mean of zero and follow a multivariate normal distribution.

f) Level-2 predictors are independent of all level-2 error terms and level-3 predictors are independent of all level-3 error terms.

g) The level-1 error terms are independent of (uncorrelated to) level-2, and level-3 error terms in the model. That is, the correlation is zero between the level-1 error term and the
level-2 error term in the model for the level-1 intercept, or the error term in any of the equations used to estimate the slopes of level-1 variables.

The use of three-levels hierarchical linear models allowed to investigate other variability sources linked to the outcome. In particular we inspect the following research questions:

1) how much of the variability in miRNAs expression is attributable to barcodes and runs?
2) does the association between the level-1 predictor PM$_{10}$ vary among barcode or run?

To answer to these questions we proceeded according the following model selection strategy:

- **Model 1 - Unconditional model:** For the three-level HLM, the unconditional model is formulated by using no predictors in the model and just random effect for the intercept to estimate the amount of variance in miRNA expression attributable to barcodes and runs. The resulting level-1, level-2 and level-3 unconditional models are:

  **Level –1 Unconditional model:**
  \[ Y_{ijk} = \pi_{0jk} + e_{ijk} \]  
  \( (Eq.5) \)

  **Level-2 Unconditional**
  \[ \pi_{0jk} = \beta_{00k} + r_{0jk} \]  
  \( (Eq.6) \)
  \[ \pi_{ijk} = \beta_{10k} + r_{ijk} \]

  **Level-3 Unconditional model:**
  \[ \beta_{00k} = \gamma_{000} + u_{00k} \]  
  \( (Eq.7) \)
  \[ \beta_{10k} = \gamma_{100} + u_{10k} \]

**SAS code**

```
proc mixed data=dataset method=ml;
   class barcode run;
   model mirna=/solution;
   random intercept/sub=run type=vc;
   random intercept/sub=barcode(run) type=vc;run;
```

This model allows to answer to the first research question, we used the three variance estimates to calculate the intra-class correlation coefficients ICCs for barcode and run:

\[ ICC_{\text{barcode}} = \frac{\sigma^2_{\text{barcode}}}{\sigma^2_{\text{barcode}} + \sigma^2_{\text{run}} + \sigma^2_{\text{error}}} \]  
\( (Eq.8) \)

\[ ICC_{\text{run}} = \frac{\sigma^2_{\text{run}}}{\sigma^2_{\text{barcode}} + \sigma^2_{\text{run}} + \sigma^2_{\text{error}}} \]  
\( (Eq.9) \)

\( ICC_{\text{barcode}} \) expresses the similarity (correlation between) of samples in the same barcode (within the same run); or alternatively informs us of how much of the total variation in
miRNAs expression exists between barcode. ICC\textsubscript{run} expresses the similarity of barcode within the same run, ignoring within-class variation.

- **Model 2: Model 1 + level - 1 fixed effect.** We included the sample-level predictor PM\textsubscript{10}

  \`\texttt{proc mixed data= dataset covtest method=ml;}
  \texttt{class barcode run;}
  \texttt{model mirna=PM10_policlinico /solution;}
  \texttt{random intercept /sub=run type=vc;}
  \texttt{random intercept /sub=barcode(run) type=vc;run;}

  Results indicate the relationship between level-1 predictor PM\textsubscript{10} and the outcome miRNAs expression.

- **Model 3: Model2 + random slopes for level-1 predictor PM\textsubscript{10.** We expanded Model 2 specifying the PM\textsubscript{10} predictor as random slope at both barcode and run level.

  \`\texttt{proc mixed data=dataset covtest method=ml;}
  \texttt{class barcode run;}
  \texttt{model mirna=PM10_policlinico /solution;}
  \texttt{random intercept PM10_policlinico /sub=run type=vc;}
  \texttt{random intercept PM10_policlinico /sub=barcode(run) type=vc;run;}

  This model allows to answer to the second research question: fixed effects results provide the same information as Model2, random slope results reveal if the relationships between the level-1 predictor PM\textsubscript{10} and the outcome miRNAs expression vary between barcode and run. This analysis was conducted for exploratory purposes initially on the first top 10 miRNAs identified by a simple multivariable regression model and then applied to the three sets of miRNAs identified. The final model was adjusted for age, body mass index, cigarette smoking (never, former, or current), and pack-years. We adjusted for percent of granulocytes (to control for possible shifts in leukocyte differential count), date, Seasonality (using sine and cosine) and apparent temperature. The results were reported as Variation(\%)=[2^*(\beta*10) -1]\times100 expresses the percentage variation in miRNAs expression associated with an increase of 10 (µg/m^3) in PM\textsubscript{10}.

  Finally, to control the expected proportion of incorrectly rejected null hypothesis in multiple comparisons, false discovery rate control (FDR) [85, 86] was applied. On the basis of FDR p-value significance (threshold of 0.20 or 0.10) and a set of top miRNAs were identified.
4.4 Mediation Analysis

4.4.1 The Simple Mediation Model
Mediation analysis is a statistical method used to help answer the question as to how some causal agent X transmits its effect on Y. What is the mechanism, be it emotional, cognitive, biological, or otherwise, by which X influences Y? The most basic of mediation models—the simple mediation model—is represented in conceptual diagram form in Figure 12. As can be seen, this model contains two consequent variables (M) and (Y) and two antecedent variables (X) and (M), with X causally influencing Y and M, and M causally influencing Y. A simple mediation model is any causal system in which at least one causal antecedent X variable is proposed as influencing an outcome Y through a single intervening variable M. In such a model, there are two distinct pathways by which a specific X variable is proposed as influencing Y. These pathways are found by tracing every way one can get from X to Y while never tracing in a direction opposite to the direction an arrow points. One pathway leads from X to Y without passing through M and is called the direct effect of X on Y. The second pathway from X to Y is the indirect effect of X on Y through M. It first passes from antecedent X to consequent M and then from antecedent M to consequent Y. The indirect effect represents how Y is influenced by X through a causal sequence in which X influences M, which in turn influences Y. In a mediation model, M is typically called a mediator variable. Once X exerts its effect on M, then M’s causal influence on Y produces variation in Y.

When thinking about whether a phenomenon or theory you are studying could be conceptualized as a mediation process, it is important to keep in mind that mediation is ultimately a causal explanation. It is assumed that the relationships in the system are causal, and, importantly, that M is causally located between X and Y. It must be assumed, if not also empirically substantiated, that X causes M, which in turn causes Y. M cannot possibly carry X’s effect on Y if M is not located causally between X and Y.

Estimation of the Direct, Indirect, and Total Effects of X

When empirically testing a causal process that involves a mediation component, of primary interest is the estimation and interpretation of the direct and indirect effects along with inferential tests thereof. To derive these effects, one must also estimate the constituent components of the indirect effect, meaning the effect of X on M as well as the effect of M on Y, although the constituent components of the indirect effect are not of primary interest in modern mediation analysis. Many researchers often estimate the total effect of X on Y as well, although doing so is not required for the purpose of interpretation.
The simple mediation model represented in the form of a statistical diagram can be found in Figure 12 (Panel B). Notice that in comparing Panel A and B, there is little difference between the conceptual and statistical diagrams representing a simple mediation model. As there are two consequent variables in this diagram, two linear models are required, one for each consequent. This statistical diagram represents two equations:

\[ M = i_1 + aX + e_M \]  
\[ Y = i_2 + c'X + bM + e_Y \]  

where \( i_1 \) and \( i_2 \) are regression intercepts, \( e_M \) and \( e_Y \) are errors in the estimation of \( M \) and \( Y \), respectively, and \( a \), \( b \), and \( c' \) are the regression coefficients given to the antecedent variables in the model in the estimation of the consequents. The coefficients of the model are treated as estimates of the putative causal influences of each variable in the system on others, and the analytical goal is to estimate these coefficients, piece them together, and interpret. These coefficients can be estimated by conducting two OLS regression analyses.

**The Direct Effect of \( X \) on \( Y \)**

In (Eq.11), \( c' \) estimates the direct effect of \( X \) on \( Y \). A generic interpretation of the direct effect is that two cases that differ by one unit on \( X \) but are equal on \( M \) are estimated to differ by \( c' \) units on \( Y \). More formally,

\[ c' = \left[ \hat{Y} \mid (X = x, M = m) \right] - \left[ \hat{Y} \mid (X = x - 1, M = m) \right] \]  

where \( m \) is any value of \( M \), \( \mid \) means conditioned on or given, and the hat over \( Y \) means estimated or expected from the model. In other words, for two cases with \( M = m \) but that differ by one unit on \( X \), \( c' \) is the estimated value of \( Y \) for the case with \( X = x \) minus the estimated value of \( Y \) for the case with \( X = x - 1 \). As can be determined looking at equation 4.3, the sign of \( c' \) tells whether the case one unit higher on \( X \) is estimated to be higher (\( c' = + \)) or lower (\( c' = - \)) on \( Y \). So a positive direct effect means that the case higher on \( X \) is estimated to be higher on \( Y \), whereas a negative direct effect means that the case higher on \( X \) is estimated to be lower on \( Y \). In the special case
where $X$ is dichotomous, with the two values of $X$ differing by a single unit (e.g., $X = 1$ and $X = 0$), $\hat{Y}$ can be interpreted as a group mean, so $c' = [\hat{Y} | (X = x, M = m)] - [\hat{Y} | (X = x - 1, M = m)]$, meaning $c'$ estimates the difference between the two group means holding $M$ constant. This is equivalent to what in analysis of covariance terms is called an adjusted mean difference.

**The Indirect Effect of $X$ on $Y$**

Before defining the indirect effect, it is first necessary to discuss what $a$ and $b$ estimate. In this model, $a$ quantifies how much two cases that differ by one unit on $X$ are estimated to differ on $M$, with the sign determining whether the case higher on $X$ is estimated to be higher (+) or lower (−) on $M$. That is, $a = [\hat{M} | (X = x)] - [\hat{M} | (X = x - 1)]$, when $X$ is a dichotomous variable coded by a unit difference, $a$ in (Eq.10) represents the difference between the two group means on $M$:

$$a = [M | (X = x)] - [M | (X = x - 1)]$$

The $b$ coefficient from equation (Eq.11) has an interpretation analogous to $c'$, except with $M$ as the antecedent. Two cases that differ by one unit on $M$ but that are equal on $X$ are estimated to differ by $b$ units on $Y$. As with $a$ and $c'$, the sign of $b$ determines whether the case higher on $M$ is estimated as higher (+) or lower (−) on $Y$:

$$b = [\hat{Y} | (M = m, X = x)] - [\hat{Y} | (M = m - 1, X = x)]$$

The indirect effect of $X$ on $Y$ through $M$ is the product of $a$ and $b$. The indirect effect tells us that two cases that differ by one unit on $X$ are estimated to differ by $ab$ units on $Y$ as a result of the effect of $X$ on $M$ which, in turn, affects $Y$. The indirect effect will be positive (meaning the case higher on $X$ is estimated to be higher on $Y$) if $a$ and $b$ are both positive or both negative, whereas it will be negative (meaning the case higher on $X$ is estimated to be lower on $Y$) if either $a$ or $b$, but not both, is negative. Although one can interpret the indirect effect without considering the signs of $a$ and $b$, doing so can be dangerous, because the sign of $ab$ is determined by two different configurations of the signs of $a$ and $b$. A certain theory you are testing might predict $ab$ to be positive because, according to the process the theory explains, $a$ and $b$ should both be positive. But what if, after estimation, $a$ and $b$ turned out to be negative? This would yield a positive indirect effect as predicted, yet this pattern of results for $a$ and $b$ is exactly opposite to what the theory predicts, and this should cast some doubt on whether the theory is adequately describing the process generating your data.
The Total Effect of X on Y

The direct and indirect effects perfectly partition how differences in X map on to differences in Y, the so-called total effect of X, denoted here as c. The total effect c quantifies how much two cases that differ by one unit on X are estimated to differ on Y. That is,

\[c = [\hat{Y}(X = x)] - [\hat{Y}(X = x - 1)]\]

In a simple mediation model, c can be derived by estimating Y from X alone:

\[Y = i_3 + cX + eY\]

(Eq.13)

When X is a dichotomous variable coded by a single unit difference, c is the difference between the group means on Y: \(c = [Y \mid (X = x)] - [Y \mid (X = x-1)]\). Regardless of whether X is dichotomous, the total effect of X on Y is equal to the sum of the direct and indirect effects of X: \(c = c' + ab\). This relationship can be rewritten as \(ab = c-c'\), which provides another definition of the indirect effect. The indirect effect is the difference between the total effect of X on Y and the effect of X on Y controlling for M, the direct effect.

That the total effect of X is the sum of the direct and indirect effects can be illustrated by substituting (Eq.11) into equation (Eq.12), thereby expressing Y as a function of only X:

\[Y = i_2 + b(i_1 + aX + e_M) + c'X + e_Y\]

which can be equivalently written as:

\[Y = (i_2 + bi_1) + (ab + c')X + (e_Y + be_M)\]

(Eq.14)

Although it may not look obvious, (Eq.14) is a simple linear function of X, just as (Eq.13). In fact, equations (Eq.13) and (Eq.14) are identical if you make the following substitutions: \(c = ab + c'\), \(i_3 = i_2 + bi_1\), and \(e_Y\) from equation (Eq.13) = \((e_Y + be_M)\) from equation (Eq.14). So \(ab + c'\) has the same interpretation as c. The sum of the direct and indirect effects quantifies how much two cases that differ by a unit on X are estimated to differ on Y.

4.4.2 Statistical Inferences

Inference about the Direct Effect of X on Y

The direct effect quantifies the estimated difference in Y between two cases that differ by one unit on X independent of M’s influence on Y. Inference for the direct effect of X on Y in a mediation analysis is typically undertaken using the standard method used for inference for any regression coefficient in a regression model. This involves testing a null hypothesis about \(c'\) against an alternative hypothesis or the construction of a confidence interval for \(c'\). Except in unusual circumstances, researchers focus on ascertaining whether a claim that \(c'\) is different from zero is justified based on the data available. If so, this supports the argument that X is
related to \( Y \) independent of the mechanism represented by \( M \). If not, one can claim that there is no evidence of association between \( X \) and \( Y \) when the mechanism through \( M \) is accounted for. In other words, \( X \) does not affect \( Y \) independent of \( M \)'s effect on \( Y \). In terms of a null hypothesis, this means testing \( H_0 : \tau c' = 0 \) against the alternative \( H_a : \tau c' \neq 0 \). Framed in terms of a confidence interval, this involves determining whether an interval estimate for \( Tc' \) includes zero.

**Inference about the Indirect Effect of \( X \) on \( Y \) through \( M \)**

The indirect effect quantifies how much two cases that differ by a unit on \( X \) are estimated to differ on \( Y \) as a result of \( X \)'s influence on \( M \), which in turn influences \( Y \). The indirect effect is relevant as to whether \( X \)'s effect on \( Y \) can be said to be transmitted through the mechanism represented by the \( X \rightarrow M \rightarrow Y \) causal chain of events. As with the direct effect, investigators typically want to know whether the data allow for the claim that this estimated difference in \( Y \) attributable to this mechanism can be said to be different from zero. If so, one can claim \( M \) serves as a mediator of the effect of \( X \) on \( Y \). As with inference about the direct effect, this inference can be formulated in terms of a null hypothesis test about \( \tau ab \) or by constructing an interval estimate. There are more than a dozen of available approaches [87-90] to statistical inference for the indirect effect, the most popular are:

- **The Normal Theory Approach (Sobel Test).** Also called the *product of coefficients* approach to inference, the *delta method*, or the *Sobel test*, the normal theory approach is based on the same theory of inference used for inference about the direct effect, as well as other inferential tests widely used in the social sciences and described in elementary statistics books. The indirect effect \( ab \) is a sample-specific instantiation of \( \tau ab \), which is subject to sampling variance. With an estimate of the standard error of \( ab \) and assuming the sampling distribution of \( ab \) is normal, a \( p \)-value for \( ab \) can be derived given a specific null hypothesized value of \( \tau ab \), or an interval estimate can be generated. Before the normal theory approach can be implemented, an estimate of the standard error of \( ab \) is needed. There are a few such estimators circulating in the literature that have been used in mediation analysis [91-94]. The simplest is a function of \( a \) and \( b \) and their standard errors:

\[
se_{ab} = \sqrt{a^2 se_b^2 + b^2 se_a^2}
\]  
(Eq.15)

where \( se_a^2 \) and \( se_b^2 \) are the squared standard errors of \( a \) and \( b \), respectively. A slightly more complex estimator includes an additional term:

\[
se_{ab} = \sqrt{a^2 se_b^2 + b^2 se_a^2 + se_a^2 se_b^2}
\]  
(Eq.16)
In practice, it typically makes little difference which estimator is used [92, 95]. (Eq. 15) is sometimes called the “first-order” delta estimator of the standard error and (Eq. 16) the “second-order” estimator.

With an estimate of the standard error of the indirect effect, the null hypothesis that \( \tau ab = 0 \) can be tested against the alternative that \( \tau ab \neq 0 \) by taking the ratio of \( ab \) to its standard error:

\[
Z = \frac{ab}{se_{ab}}
\]

and deriving the proportion of the standard normal distribution more extreme than ±Z.

Using confidence intervals over null hypothesis testing, the standard error of \( ab \) can be used to generate an interval estimate for \( \tau ab \) by assuming normality of the sampling distribution of \( ab \) and applying the following equation:

\[
ab - Z_{ci}\% se_{ab} \leq \tau ab \leq ab + Z_{ci}\% se_{ab}
\]  

(Eq. 17)

where \( ci \) is the confidence desired (e.g., 95) and \( Z_{ci}\% \) is the value of the standard normal distribution above which \((100 - ci)/2\%\) percent of the distribution resides.

The normal theory approach is simple enough to conduct, and it can be conducted by hand fairly easily if one is careful using the output from any statistical software that estimates \( a, b \), and their standard errors. Moreover the normal theory approach can be conducted even if one does not have the data used to estimate \( a, b \), and their standard errors. Although most researchers would have the original data from their own studies, there could be some circumstances in which it is not available (time has passed; the data were destroyed, lost, or stored on an obsolete storage medium; etc.). In addition, one could apply this approach using the regression coefficients and standard errors provided in the tables or text of published studies conducted by someone else that include a mediation analysis but not a formal test of the indirect effect.

These benefits aside (ease of computation, not requiring the data), the normal theory approach suffers from two flaws that make it difficult to recommend. First, whether inference is based on a hypothesis test or the construction of a confidence interval, this method assumes that the sampling distribution of \( ab \) is normal. But it has been shown analytically and through simulation that the distribution is quite irregular in sample sizes that characterize most empirical studies [96, 97]. Because it is never possible to know for certain whether the sampling distribution is close enough to normal given the characteristics of one’s problem to safely apply a method that assumes normality, it is desirable to use a test that does not require this assumption, if one is available. There are several inferential tests available that do not require this assumption and that better
respect the irregularity of the sampling distribution of \( ab \) than does the normal theory approach. Second, simulation research that has compared this approach to various competing inferential methods has shown that it is one of the lowest in power and generates confidence intervals that tend to be less accurate than some other methods described next [98]. If \( X \) does influence \( Y \) indirectly through \( M \), the normal theory approach is relatively less likely to detect it than competing alternatives. So its relatively low power combined with the unrealistic normality assumption suggest to avoid the Sobel test when possible. A possible alternative is the calculation of Bootstrap Confidence Interval.

- **Bootstrap Confidence Intervals.** The downfall of the normal theory approach is the assumption it makes about the shape of the sampling distribution of the indirect effect over repeated sampling from the population. As a member of a class of procedures known as resampling methods, bootstrapping [99-103] is a versatile method that can be applied to many inferential problems a researcher might confront. It is especially useful when the behaviour of a statistic over repeated sampling is either not known, too complicated to derive, or highly context dependent. Regardless of the inferential problem, the essence of bootstrapping remains constant across applications. The original sample of size \( n \) is treated as a miniature representation of the population originally sampled. Observations in this sample are then “resampled” with replacement, and some statistic of interest is calculated in the new sample of size \( n \) constructed through this resampling process. Repeated over and over—thousands of times ideally—a representation of the sampling distribution of the statistic is constructed empirically, and this empirical representation is used for the inferential task at hand. In mediation analysis, bootstrapping is used to generate an empirically derived representation of the sampling distribution of the indirect effect, and this empirical representation is used for the construction of a confidence interval for \( \gamma_{a\beta} \). Unlike the normal theory approach, no assumption is made about the shape of the sampling distribution of \( ab \). Bootstrap confidence intervals better respect the irregularity of the sampling distribution of \( ab \) and, as a result, yield inferences that are more likely to be accurate than when the normal theory approach is used. When used to test a hypothesis, the result is a test with higher power. There are six steps involved in the construction of a bootstrap confidence interval for \( \gamma_{a\beta} \):

1. Take a random sample of \( n \) cases from the original sample, sampling those cases with replacement, where \( n \) is the size of the original sample. This is called a bootstrap sample.
2. Estimate the indirect effect $ab^*$ in the bootstrap sample, where $ab^*$ is the product of $a$ and $b$ from (Eq.10) and (Eq.11).

3. Repeat (1) and (2) above a total of $k$ times, where $k$ is some large number, saving the value of $ab^*$ each time. Generally, $k$ of at least a few thousand is preferred. More than 10,000 typically is not necessary, but in principle, the more the better.

4. Sort the $k$ indirect effects $ab$ estimated from steps (1), (2), and (3) from low to high.

5. For a $ci\%$ confidence interval, find the value of $ab$ in this distribution of $k$ estimates that defines the $0.5(100 − ci)$th percentile of the distribution. This is the lower bound of a $ci\%$ confidence interval. It will be the value of $ab$ in ordinal position $0.005k(100 − ci)$ of the sorted distribution.

6. Find the value of $ab^*$ in this distribution of $k$ estimates that defines the $[100 − 0.5(100 − ci)]$th percentile of the distribution. This is the upper bound of a $ci\%$ confidence interval. It will be the value of $ab$ in ordinal position $k[1 − 0.005(100 − ci)] + 1$ of the sorted distribution.

To illustrate steps (1), (2), and (3) of this bootstrap sampling and estimation process, Table 2 provides a small-scale example. Suppose to have a sample of $n = 10$ cases in a study measured on variables $X$, $M$, and $Y$, and suppose to generate a bootstrap sampling distribution of the indirect effect of $X$ on $Y$ through $M$. Using the original data in the leftmost columns of the table, the obtained indirect effect is $ab = 0.770$. This is a point estimate of $TaTb$. A bootstrap confidence interval for $TaTb$ is constructed by repeatedly taking a random sample of size $n$ from the original sample, with replacement, and estimating the indirect effect in each resample.

Table 2: Bootstrap Estimates of $a$, $b$, and the Indirect Effect $ab$ When Taking Two Bootstrap Samples from an Original Sample of Size $n = 10$

<table>
<thead>
<tr>
<th>Original Sample</th>
<th>Bootstrap sample1</th>
<th>Bootstrap sample2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>X</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td>4.1</td>
<td>3.4</td>
</tr>
<tr>
<td>2</td>
<td>4.2</td>
<td>5.4</td>
</tr>
<tr>
<td>3</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td>4</td>
<td>4.3</td>
<td>8.6</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>5.3</td>
</tr>
<tr>
<td>6</td>
<td>4.5</td>
<td>2.5</td>
</tr>
<tr>
<td>7</td>
<td>3.8</td>
<td>6.3</td>
</tr>
<tr>
<td>8</td>
<td>4.9</td>
<td>2.9</td>
</tr>
<tr>
<td>9</td>
<td>4.5</td>
<td>3.3</td>
</tr>
<tr>
<td>10</td>
<td>4.8</td>
<td>2.1</td>
</tr>
</tbody>
</table>

| $a$ | 1.16 | $a$ | 2.14 | $a$ | 1.16 |
| $b$ | 3.24 | $b$ | 3.15 | $b$ | 3.18 |
| $ab$ | 3.76 | $ab^*$ | 6.74 | $ab^*$ | 3.69 |
The middle columns of Table 2 contain one such bootstrap sample, which yields an indirect effect of \( ab = 0.526 \). The rightmost columns contain a second bootstrap sample with an indirect effect of \( ab = 1.039 \). As this process is repeated over and over, a distribution of \( ab \) is built which functions as an empirical proxy for the unknown sampling distribution of \( ab \) when taking a random sample of size \( n \) from the original population. Steps (5) and (6) are generic ways of describing how the endpoints of a confidence interval are constructed given \( k \) bootstrap estimates of the indirect effect. A specific example will help. If a \( CI = 95\% \) confidence interval is desired, the lower and upper bounds of the interval are defined as the bootstrap values of \( ab \) that define the 2.5th and 97.5th percentiles in the distribution of \( k \) values of \( ab \). Suppose \( k = 10,000 \). In that case, after sorting the 10,000 values of \( ab \) obtained from repeated bootstrap sampling from low to high, the 2.5th and 97.5th percentiles of the distribution will be in ordinal positions \( 0.005(10,000)(100-95) = 250 \) and \( (10,000)[1-0.005(100-95)] + 1 = 9,751 \) in the sorted list, respectively. These are the lower and upper bounds of the 95% confidence interval for \( TaTb \). Obviously, this is a computationally intensive process that requires a computer.

The use of bootstrap confidence as inferential approach it is not without its pitfalls and criticisms:

- first, in order to have much confidence in bootstrap-based inference, it is clearly important that one is able to muster some faith in the quality of one’s sample as a reasonable representation of the population with respect to the distribution of the measured variables. Bootstrapping is founded on the notion that resampling with replacement from one’s sample mimics the original sampling process. But if the sample does not adequately represent the population from which the sample was derived, then bootstrapping will produce results that are hard to trust. It is not required that the original sample be obtained randomly from the population, but merely that the distribution of the measured variables roughly mirrors the population distributions. Random sampling facilitates this representativeness, of course, but it isn’t required;

- second, bootstrapping is particularly useful relative to the normal theory approach in smaller samples, because it is in smaller samples that the non-normality of the sampling distribution of \( ab \) is likely to be most severe, the large sample asymptotics of the normal theory approach are harder to trust, and the power advantages of bootstrapping are more pronounced. But if the original sample is
very small, in principle, there is a strong potential for one or two cases to distort a bootstrap analysis even more than they do a more traditional inferential procedure. If the original sample is very small, an unusual case or two are highly likely to appear in a bootstrap sample multiple times, and this can distort a bootstrap analysis;

third, because bootstrap confidence intervals are based on random resampling of the data, the endpoints of the confidence interval are not fixed quantities. Rather, each time a bootstrap confidence interval is produced from the same data, a slightly different confidence interval will result, however the sampling variation from analysis to analysis can be made arbitrarily small simply by setting the number of bootstrap samples to an arbitrarily large number. This raises the question as to how many bootstrap samples is enough. It can be shown that the variation in the estimation of the limits of a confidence interval shrinks remarkably quickly as the number of bootstrap samples increases. Generally speaking, 5,000 to 10,000 bootstrap samples is sufficient in most applications. There is relatively little added value to increasing it above 10,000, as the gain in precision is fairly marginal beyond that.

A bootstrap confidence interval calculated using the approach just described is called a percentile bootstrap confidence interval, because it is based entirely on values of \( ab^* \) that demarcate the upper and lower \( (100 - ci)/2 \% \) of the distribution of \( k \) bootstrap estimates of the indirect effect. It is also possible calculate bias-corrected bootstrap confidence intervals (BC bootstrap confidence intervals) are like percentile confidence intervals but the endpoints are adjusted as a function of the proportion of \( k \) values of \( ab^* \) that are less than \( ab \), the point estimate of the indirect effect calculated in the original data. The endpoints will be adjusted upward or downward to varying degrees depending on that proportion. A variation on this variation, known as the bias-corrected and accelerated bootstrap confidence interval, makes an additional adjustment based on the skew of the distribution of \( k \) bootstrap estimates. The following steps [87, 104-106]; allow to generate a bias-corrected bootstrap confidence interval for the indirect effect:

1. Follow steps (1) through (4) to generate \( k \) bootstrap estimates of the indirect effect, \( ab^* \) as described before.

2. Calculate \( Z(\bar{p}) \), the Z-score that cuts off the lower 100 \( \bar{p} \)% of the standard normal distribution from the rest of the distribution, and \( \bar{p} \) is the proportion of the \( k \) values of \( ab^* \) that are less than \( ab \) calculated using the original data.
3. Calculate $Z_{low} = Z_{ci} + 2Z(p)$ and $Z_{high} = -Z_{ci} + 2Z(p)$, where $Z_{ci}$ is the $Z$-score that cuts off the lower $(100 - ci\%) / 2$ percent of the standard normal distribution from the rest of the distribution. For instance, for a 95% confidence interval, $Z_{95} = -1.96$.

4. Calculate $plow$ and $phigh$, the proportion of the standard normal distribution the left of $Z_{low}$ and $Z_{high}$, respectively.

5. Find the value of $ab^*$ in the distribution of $k$ estimates that defines the 100$plow$ percentile of the distribution. This is the lower bound of a $ci\%$ bias-corrected bootstrap confidence interval, and will be the value of $ab^*$ in ordinal position $(plow)k$ of the sorted distribution. If $(plow)k$ is not an integer, round it down to the lowest integer.

6. Find the value of $ab^*$ in the distribution of $k$ estimates that defines the 100$phigh$ percentile of the distribution. This is the upper bound of a $ci\%$ bias-corrected bootstrap confidence interval, and will be the value of $ab^*$ in ordinal position $(phigh)k$ of the sorted distribution. If $(phigh)k$ is not an integer, round it up to the next highest integer.

Observe that the upper and lower bounds of the 95% bootstrap confidence intervals calculated earlier are not equidistant from the point estimate. This is not due to the random resampling process but instead reflects the actual asymmetry of the sampling distribution of $ab$. Confidence intervals based on the normal theory approach to inference, by contrast, impose a symmetry constraint on this distance. The endpoints of a 95% confidence interval using equation 4.9 are necessarily 1.96 standard errors from the point estimate. The endpoints are symmetrical around the point estimate. Thus, percentile-based and BC bootstrap confidence intervals are called “asymmetric,” whereas normal theory confidence intervals are “symmetric.” Asymmetric approaches to interval estimation are preferred when the sampling distribution of the estimator is asymmetric and non-normal, as is the case for the sampling distribution of $ab$.

Bootstrapping is not the only approach to the construction of asymmetric confidence intervals. Although bootstrapping is recommended, it does have a few weaknesses, among them that it requires the original data (not usually a real problem typically), the endpoints of the confidence interval will vary from run to run (but not if you seed the random number generator yourself), and it isn’t implemented in all software one might choose to use. An alternative to get around these problems: Monte Carlo confidence intervals. Monte Carlo confidence intervals [87, 98, 107] are simulation-based. This approach relies on the fact that though the distribution of $ab$ is not normal, the sampling distributions of $a$ and $b$ tend to be nearly so. Furthermore, in simple mediation analysis using OLS
regression, a and b are independent across repeated sampling (i.e., their covariance is zero). Thus, an empirical approximation of the sampling distribution of ab can be generated by randomly sampling values of a and b from normally distributed populations with \( \mu = a, \sigma = se_a \) and \( \mu = b, \sigma = se_b \), respectively, where a, b, sea, and seb are the OLS regression coefficients and standard errors from the mediation analysis. The sampled values of a and b are then multiplied together to produce \( ab^* \), and this process is repeated \( k \) times. Over the \( k \) replications, the upper and lower bounds of the confidence interval for \( ab \) can be generated using the procedure described in steps (4) through (6) for the construction of bootstrap confidence intervals. Monte Carlo method is almost as good as bootstrapping and better than the normal theory approach.

**Inference about the Total Effect of X on Y**

In a simple mediation model, the total effect of \( X \) on \( Y \) is the sum of the direct effect of \( X \) on \( Y \) and indirect effect of \( X \) on \( Y \) through \( M \). Whereas there are many choices available for inferences about the indirect effect, inference for the total effect is simple and straightforward. Although the total effect is the sum of two pathways of influence, it can be estimated simply by regressing \( Y \) on \( X \). The regression coefficient for \( X \) in that model, \( c \) in equation 4.4, is the total effect of \( X \). Inference can be framed in terms of a null hypothesis test (\( H_0 : \gamma_c = 0 \) versus the alternative \( H_a : \gamma_c \neq 0 \)) or whether an interval estimate for \( \gamma_c \) includes zero.

**4.4.3 The Parallel Multiple Mediator Model**

In a parallel multiple mediator model, antecedent variable \( X \) is modelled as influencing consequent \( Y \) directly as well as indirectly through two or more mediators, with the condition that no mediator causally influences another. *Figure 13* depicts a parallel multiple mediator model with \( k \) mediators in conceptual form (Panel A) and in statistical diagram (Panel B). Observe that the parallel multiple mediator model looks much like a simple mediation model except that it includes more than one mediator. A defining feature of the parallel multiple mediator model that distinguishes it from an alternative multiple mediator model, for example the serial multiple mediator model, is the constraint that no mediator is modelled as influencing another mediator in the model. This constraint is apparent in *Figure 13* by the absence of any unidirectional arrows linking any mediator to any other mediator. This is not to say that the mediators are assumed to be independent. In fact, in most circumstances, the mediators are likely to be correlated. Even if they are not, there still may be some advantage to estimating a parallel multiple mediator model with \( k \) mediators rather than \( k \) simple mediation models.
Figure 13 Conceptual (Panel A) and Statistical (Panel B) representing a parallel multiple mediator model with k mediators.

Doing so could result in a power boost for tests of indirect effects if each mediator is correlated with Y, and doing so affords the ability to compare the sizes of the indirect effects through different mediators. In principle, the number of mediators one can include in a parallel multiple mediator model is limited only by the number of cases in one’s data file and the number of variables one has the foresight to measure as possible mediators. In practice, models with two mediators are most commonly estimated. But parallel multiple mediator models can be found with three, six, and even as many as seven mediators in a model simultaneously. As can be seen in Figure 13, a parallel multiple mediator model with k mediators has k + 1 consequent variables (one for each of the k mediators M and one for the outcome variable Y) and so requires k + 1 equations to estimate all the effects of X on Y. These equations are:

\[ M_i = i_{M_i} + a_i X + e_{M_i} \text{ for all } i = 1 \text{ to } k \] (Eq.17)

\[ Y = i_Y + c'X + \sum_{i=1}^{k} b_i M_i + e_Y \] (Eq.18)

In this set of equations, \( a_i \) estimates the effect of \( X \) on \( M_i \), \( b_i \) estimates the effect of \( M_i \) on \( Y \) controlling for \( X \) and the other \( k - 1 \) \( M \) variables, and \( c' \) estimates the effect of \( X \) on \( Y \) holding all \( k \) \( M \) variables constant. Consider a parallel multiple mediator with three proposed mediators. With \( k = 3 \) mediators, four equations are needed:

\[ M_1 = i_{M_1} + a_1 X + e_{M_1} \] (Eq.19)

\[ M_2 = i_{M_2} + a_2 X + e_{M_2} \] (Eq.20)

\[ M_3 = i_{M_3} + a_3 X + e_{M_3} \] (Eq.21)
\[ Y = i_Y + c'M + b_1M_1 + b_2M_2 + b_3M_3 + e_m3 \]  
(Eq.22)

In equations (Eq.19), (Eq.20), and (Eq.21), \( a_1 \), \( a_2 \), and \( a_3 \) quantify the amount by which two cases that differ by one unit on \( X \) are estimated to differ on \( M_1 \), \( M_2 \), and \( M_3 \), respectively. In equation (Eq.22), \( b_1 \) estimates the amount by which two cases that differ by a unit on \( M_1 \) differ on \( Y \) holding \( M_2 \), \( M_3 \), and \( X \) constant. Similarly, \( b_2 \) estimates the amount by which two cases that differ by a unit on \( M_2 \) differ on \( Y \) holding \( M_1 \), \( M_3 \), and \( X \) constant, and \( b_3 \) estimates the amount by which two cases that differ by a unit on \( M_3 \) differ on \( Y \) holding \( M_1 \), \( M_2 \), and \( X \) constant. Finally, \( c' \) estimates the amount by which two cases that differ by one unit on \( X \) differ on \( Y \) holding \( M_1 \), \( M_2 \), and \( M_3 \) constant. The interpretations of \( a_i \) and \( c' \) are not dependent on the scale of measurement of \( X \). Whether \( X \) is a dichotomous variable or a continuum, the interpretation is the same. However, when \( X \) is a dichotomous variable with the two groups coded by a one unit difference, these can be interpreted as estimated mean differences. For instance, suppose the two groups are coded with \( X = 0 \) or \( X = 1 \). In that case:

\[
\begin{align*}
    a_i &= [\bar{M}_i|(X = 1)] - [\bar{M}_i|(X = 0)] \\
    a_i' &= [\bar{Y}^*|(X = 1)] - [\bar{Y}^*|(X = 0)]
\end{align*}
\]

where \( \bar{Y}^* \) is an adjusted mean, with all mediators set to their sample means:

\[
\bar{Y}^* = i_Y + c'X + \sum_{i=1}^{k} b_i \bar{M}_i
\]

**Direct and Indirect Effects in a Parallel Multiple Mediator Model**

In a parallel multiple mediator as in Figure 13, \( X \) is modeled to exert its effect on \( Y \) through \( k + 1 \) pathways. One pathway is direct, from \( X \) to \( Y \) without passing through any of the proposed mediators, and the other \( k \) pathways are indirect, each through a single mediator. In a multiple mediator model, the indirect effects are referred to as *specific indirect effects*. Thus, a model with \( k \) mediators has \( k \) specific indirect effects, one through \( M_1 \) (\( X \to M_1 \to Y \)), one through \( M_2 \) (\( X \to M_2 \to Y \)), and so forth, up through \( M_k \) (\( X \to M_k \to Y \)). As in a simple mediation model, the indirect effect of \( X \) on \( Y \) through a given mediator \( M_i \) is quantified as the product of paths linking \( X \) to \( Y \) through \( M_i \). In a parallel multiple mediator model, only two paths link \( X \) to \( Y \) through \( M_i \). The first of these paths is the effect of \( X \) to \( M_i \), and the second is the path from \( M_i \) to \( Y \). The regression coefficients corresponding to these paths, when multiplied together, yield the specific indirect effect of \( X \) on \( Y \) through \( M_i \). So consider the three-mediator parallel multiple mediator model estimated with equations (Eq.19) through (Eq.20). In this model, the specific indirect of \( X \) on \( Y \) through \( M_1 \) is \( a_1b_1 \), the specific indirect effect through \( M_2 \) is \( a_2b_2 \), and the specific indirect effect of \( X \) through \( M_3 \) is \( a_3b_3 \). Most generally, regardless of the number of mediators, the specific indirect effect of \( X \) on \( Y \) through \( M_i \) is estimated as
\( a_i b_i \) from equations (Eq.17) and (Eq.18). A specific indirect effect is interpreted just as in the simple mediation model, except with the addition of controlling for all other mediators in the model. Thus, the specific indirect effect of \( X \) on \( Y \) through \( M_i \) is the estimated amount by which two cases that differ by a unit on \( X \) are estimated to differ on \( Y \) as a result of the effect of \( X \) on \( M_i \), which in turn affects \( Y \), holding all other mediators constant. When added together, the specific indirect effects yield the total indirect effect of \( X \) on \( Y \) through all mediators in the model. In a model with \( k \) mediators:

\[
\text{Total indirect effect of } X \text{ on } Y = \sum_{i=1}^{k} a_i b_i
\]

For example, in a parallel multiple mediator model with three mediators represented by equations (Eq.19) through (Eq.22), the total indirect effect of \( X \) on \( Y \) is \( a_1 b_1 + a_2 b_2 + a_3 b_3 \). The direct effect of \( X \) quantifies how much two cases that differ by a unit on \( X \) are estimated to differ on \( Y \) independent of all mediators. As discussed earlier, this is \( c' \) in the model of \( Y \) from \( X \) and all mediators (e.g., equation (Eq.22) for the three-mediator model, or equation (Eq.18) more generally). As in the simple mediation model, the sum of the direct and indirect effects is the total effect of \( X \). In a model with \( k \) mediators, from the coefficients in equations (Eq.17) and (Eq.18).

\[
c = c' + \sum_{i=1}^{k} a_i b_i
\]  

(Eq.23)

where \( c \) is the total effect of \( X \). The total effect can also be estimated by regressing \( Y \) on \( X \) alone. For instance, in the three mediator model, \( c = c' + a_1 b_1 + a_2 b_2 + a_3 b_3 \). Isolation of the total indirect effect in equation (Eq.23) shows that the total indirect effect is equal to the difference between the total and the direct effects of \( X \):

\[
c - c' = \sum_{i=1}^{k} a_i b_i
\]

Inference about the Direct Effect

As in the simple mediation model, inference about the direct effect of \( X \) on \( Y \) is straightforward. A test of the null hypothesis \( H_0 : \gamma c' = 0 \) versus the alternative \( Ha : \gamma c' \neq 0 \) is available in the output from any statistical package that can estimate (Eq.18) using OLS regression. Alternatively, a confidence interval can be constructed using equation \( c' - t_{c%} S_e c' \leq Tc' \leq c' - t_{c%} S_e c' \).
**Inference about Specific Indirect Effects**

The normal theory approach for the indirect effect in a simple mediation model can be used for statistical inference about specific indirect effects in a parallel multiple mediator model without modification. For the specific indirect effect of X on Y through $M_i$, the first-order standard error estimator is:

$$Se_{a_ib_i} = \sqrt{a_i^2 Se_{b_i}^2 + b_i^2 Se_{a_i}^2}$$

where $Se_{b_i}^2$ and $Se_{a_i}^2$ are the squared standard errors of $a_i$ and $b_i$. The second order estimator is:

$$Se_{a_ib_i} = \sqrt{a_i^2 Se_{b_i}^2 + b_i^2 Se_{a_i}^2 + Se_{b_i}^2 Se_{a_i}^2}$$

A test of the null hypothesis that $\gamma a_i b_i = 0$ is constructed by dividing $a_i b_i$ by the estimated standard error and deriving a $p$-value from the standard normal distribution. Alternatively, a $c_i$% confidence interval can be constructed as $a_i b_i - Zc_i % Se_{a_ib_i} \leq \gamma a_i b_i \leq a_i b_i + Zc_i % Se_{a_ib_i}$ where $c_i$ is the confidence desired (e.g., 95) and $Zc_i$% is the value under the normal distribution that cuts off the upper $(100 - c_i)/2$% of the distribution from the rest.

However this approach is hard to trust. It makes the unrealistic assumption of normality of the sampling distribution of the specific indirect effect, and it is one of the more conservative tests available. Also in this case bootstrap confidence intervals are the better approach to inference when the original data are available for analysis. No assumptions about the shape of the sampling distribution of $a_i b_i$ are made, and bootstrap confidence intervals tend to be more powerful than competing methods such as the normal theory approach [108]. Using the same procedure described before, a bootstrap confidence interval for a specific indirect effect is constructed by taking a random sample with replacement of size $n$ from the sample, estimating each specific indirect effect $a_i b_i$ in the resulting data, and repeating this resampling and estimation many times. With several thousand bootstrap estimates of each specific indirect effect, endpoints of the confidence interval are calculated using either the percentile or bias corrected method. If zero is outside of a $c_i$% confidence interval, then $\gamma a_i b_i$ is declared different from zero with $c_i$% confidence, whereas if the confidence interval straddles zero, the conclusion is that there is insufficient evidence that X affects Y through $M_i$.

**Pairwise Comparisons between Specific Indirect Effects**

In a multiple mediator model, it is sometimes of interest to test whether one indirect effect is statistically different from another. If the indirect effect of X through mediator $i$ (i.e., $a_i b_i$) is pertinent to the mechanism postulated by one theory and the indirect effect of X through mediator $j$ (i.e., $a_j b_j$) quantifies the mechanism relevant to a second theory, an inference about whether $\gamma a_i b_i = \gamma a_j b_j$ affords a claim as to whether one mechanism accounts for more of the effect of X on Y than the other mechanism [109, 110],
with an important caveat described below. Although it might seem that such a comparison between specific indirect effects would be impossible if the mediators are measured on different metrics, it turns out this is not a problem at all. Remember that the specific indirect effect is interpreted as the amount by which two cases differing by a unit on X are estimated to differ on Y through the intervening variable independent of the other intervening variables. Notice that this interpretation does not include the metric of the intervening variable. Specific indirect effects are scaled entirely in terms of the metrics of X and Y [68, 89, 111], so two specific indirect effects of the same antecedent on the same consequent can be meaningfully compared even if the mediator variables are measured on entirely different scales. Thus, standardization is not necessary to conduct an inferential test of the equality of specific indirect effects from X to Y in a multiple mediator model.

Two inferential approaches have been most widely discussed and disseminated in the literature. A normal theory approach is described by [68, 111] based on dividing $a_i b_i - a_j b_j$ by an estimate of its standard error. One estimator of the standard error of the difference is:

$$Se_{a_i b_i - a_j b_j} = \sqrt{b_i^2 Se_{a_i}^2 - 2 b_i b_j COV_{a_i, a_j} + b_j^2 Se_{a_j}^2 + a_j^2 Se_{b_j}^2 - 2 a_i a_j COV_{b_i b_j} + a_i^2 Se_{b_i}^2}$$

where $COV_{a_i a_j}$ is the covariance between $a_i$ and $a_j$, and $COV_{b_i b_j}$ is the covariance between $b_i$ and $b_j$.

[111] Offers a different standard error estimator that does not require the covariance between $a_i$ and $a_j$ by assuming it is zero, which is equivalent to constraining the correlation between the residuals in the models of $\hat{M}_i$ and $\hat{M}_j$ to be zero:

$$Se_{a_i b_i - a_j b_j} = \sqrt{b_i^2 Se_{a_i}^2 + b_j^2 Se_{a_j}^2 + a_j^2 Se_{b_j}^2 - 2 a_i a_j COV_{b_i b_j} + a_i^2 Se_{b_i}^2}$$

The ratio of the difference to its standard error is then calculated and a $p$-value for a test of the null hypothesis that $\tau a_i \tau b_i = \tau a_j \tau b_j$ can be derived using the standard normal distribution. Alternatively, a 95% confidence interval for the difference can be computed as:

$$(a_i b_i - a_j b_j) \pm 1.96 Se_{a_i b_i - a_j b_j} \quad (Eq.23)$$

In this expression, 1.96 can be replaced with an appropriate critical $Z$ from a table of normal probabilities for different confidence levels (e.g., 1.645 for 90% or 2.57 for 99% confidence).

Like all normal theory approaches discussed thus far, this method requires the assumption that the sampling distribution of the difference between specific indirect effects is normal. It turns out that this is a fairly reasonable assumption, but since an assumption can never be proven true, bootstrapping offers an alternative test without requiring this assumption. A bootstrap confidence interval is derived by estimating the difference between specific indirect effects over repeated bootstrap sampling and model estimation. Using the resulting empirical approximation of the sampling distribution of the difference between specific indirect effects, a confidence interval for the difference can be constructed using either the percentile method described earlier or through bias correction.

In a model with $k$ mediators, will be conducted $k(k - 1)/2$ pairwise comparisons, one for each possible difference between specific indirect effects. A confidence interval that does not contain zero provides
evidence that the two indirect effects are statistically different from each other, whereas a confidence interval that straddles zero supports the claim of no difference between the specific indirect effects. It is tempting to treat this as a test of the difference in strength of the mechanisms at work linking X to Y, or that one indirect effect is larger than another in an absolute sense. However, such an interpretation is justified only if the point estimates for the two specific indirect effects being compared are of the same sign. Consider, for instance, the case where $ab_i = -0.30$ and $ab_j = 0.30$. A test of the difference between these specific indirect effects may lead to the claim that their difference is not zero, but this does not imply the mechanisms are of different strength or that one indirect effect is bigger. The point estimates suggest one mechanism results in a positive difference in Y, whereas the other yields a negative difference of equal magnitude. In an absolute sense, they are equal in size by the point estimates, yet statistically different by an inferential test which considers their sign. But one indirect effect is not stronger than the other. Nor can we say that X exerts a larger effect on Y through one of the mediators relative to the other.

**Inference about the Total Indirect Effect**

A multiple mediator model also contains a total indirect effect, defined as the sum of all specific indirect effects. It is possible to conduct an inferential test of the total indirect effect using either the normal theory approach or a bootstrap confidence interval. The normal theory approach requires an estimate of the standard error of the total indirect effect, but the formula for constructing it is quite complicated even in multiple mediator models with only two mediators. Given such complicated expressions is recommended to use a bootstrap confidence interval obtained as described before.
5. RESULTS

5.1 Study Population description at screening Stage.

Table 1 and Table 2 summarize continuous and categorical variables collected in the study at stage 1. A total of 1250 subjects, mainly women (74%), were enrolled (participation rate in the years 2010-2013 was 90%) with a mean age of 51.9±13.6 years and a mean BMI of 33.5±5.5 kg/m^2: 7.8% are overweight, 38.6% obese, and 33.6% severe obese (BMI ≥ 35 Kg/cm^2). About 34% of subjects are former smokers and about 15% are current smokers with a median of 14.5 Pack/years. About 57% of SPHERE subjects live in the city and an additional 28% work in Milan (even if they live outside the city), overall a 67% of subjects spent many hours a day in the city or travelling from workplace to residence. Thus, longer exposure effects are evaluated by both residential and Milan monitors using appropriate lag time from recruitment date, as sensitivity analysis.

Table 3: Characteristics of study participants and data collected from the self-reported questionnaire at December 31, 2013.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Categories</th>
<th>n=1250 (Stage 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>330 (26.4%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>920 (73.6%)</td>
</tr>
<tr>
<td>Age</td>
<td>Years (mean±SD)</td>
<td>51.9±13.6</td>
</tr>
<tr>
<td>Education</td>
<td>Primary school or less</td>
<td>105 (8.4%)</td>
</tr>
<tr>
<td></td>
<td>Secondary school</td>
<td>325 (26.0%)</td>
</tr>
<tr>
<td></td>
<td>High school</td>
<td>493 (39.4%)</td>
</tr>
<tr>
<td></td>
<td>University</td>
<td>188 (15.0%)</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>87 (7.0%)</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>52 (4.2%)</td>
</tr>
<tr>
<td>Occupation</td>
<td>Employee</td>
<td>714 (57.1%)</td>
</tr>
<tr>
<td></td>
<td>Unemployed</td>
<td>102 (8.2%)</td>
</tr>
<tr>
<td></td>
<td>Pensioner</td>
<td>304 (24.3%)</td>
</tr>
<tr>
<td></td>
<td>Housewife</td>
<td>93 (7.4%)</td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td>37 (3.0%)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>White</td>
<td>1198(95.8%)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>11 (0.9%)</td>
</tr>
<tr>
<td></td>
<td>Asian</td>
<td>3 (0.3%)</td>
</tr>
<tr>
<td></td>
<td>South America</td>
<td>38 (3.0%)</td>
</tr>
<tr>
<td>Year of enrollment</td>
<td>2010</td>
<td>129 (10.3%)</td>
</tr>
<tr>
<td></td>
<td>2011</td>
<td>419 (33.5%)</td>
</tr>
<tr>
<td></td>
<td>2012</td>
<td>385 (30.8%)</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>317 (25.4%)</td>
</tr>
<tr>
<td>Season of enrollment</td>
<td>Winter</td>
<td>320 (25.6%)</td>
</tr>
<tr>
<td></td>
<td>Spring</td>
<td>313 (25.0%)</td>
</tr>
<tr>
<td></td>
<td>Summer</td>
<td>190 (15.2%)</td>
</tr>
<tr>
<td></td>
<td>Autumn</td>
<td>427 (34.2%)</td>
</tr>
<tr>
<td>Smoking</td>
<td>Never</td>
<td>599 (47.9%)</td>
</tr>
<tr>
<td></td>
<td>Former</td>
<td>431 (34.5%)</td>
</tr>
<tr>
<td></td>
<td>Current</td>
<td>190 (15.2%)</td>
</tr>
</tbody>
</table>
Table 4: Characteristics of the study subjects and mean levels of the clinical measure investigated at December 31, 2013.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N</th>
<th>Mean±SD or N(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI, Kg/cm²</td>
<td>1247</td>
<td>33.5±5.5</td>
</tr>
<tr>
<td>BMI categorical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;30 Kg/cm²</td>
<td>347</td>
<td>27.8%</td>
</tr>
<tr>
<td>30-35 Kg/cm²</td>
<td>483</td>
<td>38.6%</td>
</tr>
<tr>
<td>≥35 Kg/cm²</td>
<td>420</td>
<td>33.6%</td>
</tr>
<tr>
<td>Waist circumference, cm</td>
<td>1237</td>
<td>101.3±13.1</td>
</tr>
<tr>
<td>Blood pressure, mmHg</td>
<td>1247</td>
<td></td>
</tr>
<tr>
<td>Sistolic</td>
<td>1254</td>
<td>15.8</td>
</tr>
<tr>
<td>Diastolic</td>
<td>78.5</td>
<td>9.5</td>
</tr>
<tr>
<td>Above 140/90 mmHg</td>
<td>60</td>
<td>4.8%</td>
</tr>
<tr>
<td>Below 140/90 mmHg</td>
<td>1190</td>
<td>95.2%</td>
</tr>
<tr>
<td>Heart rate, bPM</td>
<td>1243</td>
<td>67.6±10.4</td>
</tr>
<tr>
<td>Uric acid</td>
<td>1163</td>
<td>5.2±1.4</td>
</tr>
<tr>
<td>Fibrinogen, mg/dl</td>
<td>1129</td>
<td>335±59</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>1160</td>
<td>0.3 [0.1-0.5]</td>
</tr>
<tr>
<td>Total cholesterol, mg/dl</td>
<td>1165</td>
<td>215.1±41</td>
</tr>
<tr>
<td>HDL</td>
<td></td>
<td>59.2±15.5</td>
</tr>
<tr>
<td>Test</td>
<td>Reference</td>
<td>Value</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>LDL</td>
<td>1164</td>
<td>134.7±36.3</td>
</tr>
<tr>
<td>Triglyceride</td>
<td>1164</td>
<td>107 [77-145.5]</td>
</tr>
<tr>
<td>Serum creatinine, mg/dL</td>
<td>1165</td>
<td>0.8±0.3</td>
</tr>
<tr>
<td>AST, U/I</td>
<td>1159</td>
<td>19 [16-23]</td>
</tr>
<tr>
<td>ALT, U/I</td>
<td>1160</td>
<td>21 [16-30.5]</td>
</tr>
<tr>
<td>Gamma-Glutamyltransferase, U/L</td>
<td>1162</td>
<td>19 [13-30]</td>
</tr>
<tr>
<td>Glucose</td>
<td>1155</td>
<td>92 [86-101]</td>
</tr>
<tr>
<td>Homocysteine</td>
<td>1151</td>
<td>10.4 [8.6-12.7]</td>
</tr>
<tr>
<td>TSH</td>
<td>1163</td>
<td>1.7 [1.2-2.5]</td>
</tr>
<tr>
<td>Glycated hemoglobin, mmol/mol</td>
<td>1159</td>
<td>39 [36.6-43]</td>
</tr>
<tr>
<td>Postprandial glycaemia, mg/dl</td>
<td>1162</td>
<td>99 [90-112]</td>
</tr>
<tr>
<td>Insulin level</td>
<td>1158</td>
<td>12.3 [8.8-18]</td>
</tr>
<tr>
<td>2-hour post glucose insulin level</td>
<td>1155</td>
<td>46.4 [27.6-73]</td>
</tr>
<tr>
<td>Urinary pH</td>
<td>1144</td>
<td>5.6±0.7</td>
</tr>
<tr>
<td>Emocrome</td>
<td>1156</td>
<td></td>
</tr>
<tr>
<td>White blood cells</td>
<td></td>
<td>6.8±1.7</td>
</tr>
<tr>
<td>Red blood cells</td>
<td></td>
<td>4.8±0.4</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td></td>
<td>13.8±1.4</td>
</tr>
<tr>
<td>Hematocrit</td>
<td></td>
<td>40.7±3.4</td>
</tr>
<tr>
<td>Mean Corpuscolar Volume</td>
<td></td>
<td>85.1±6.4</td>
</tr>
<tr>
<td>Platelets</td>
<td></td>
<td>249.7±59</td>
</tr>
</tbody>
</table>

5.2 Comparison between FARM model and monitoring station PM$_{10}$ levels.

Air quality modeling and ambient measurements are two different ways to estimate actual ambient concentrations of pollutants in the atmosphere. Both modeling and measurements have some degree of uncertainty associated with their estimates. The uncertainty of the FARM model results both from that of the model (due to the inability to perfectly describe the physical phenomena) and from that associated with the input data. Figure 14 shows the box plots comparing PM$_{10}$ concentrations of the two methods of exposure assessment by year: estimated PM$_{10}$ concentrations are slightly lower than observed across investigated years. Table 3 reports a description of PM$_{10}$ levels between 2010 and 2012 and by season, and selected weather variable distribution.
Figure 14: Mean PM10 level observed by monitors and estimated by FARM model by place and across years.

PM10 air concentrations depends, besides emissions, on weather conditions during the days, in particular by rainfall, atmospheric stability and wind. PM10 levels follows a seasonal trend, with critic periods concentrated in autumn and winter seasons, which are characterized by atmospheric stability, calm wind and absence of precipitations. PM10 air concentrations depends, besides emissions, on weather conditions during the days, in particular by rainfall, atmospheric stability and wind. Figure 15 shows the distributions of daily mean PM10 concentrations of all monitors and of all grid cells with a monitor falling into their boundary.

Table 5: PM10 profile and weather variables at the time of blood sampling for the entire period of the SPHERE Study and by seasons.

<table>
<thead>
<tr>
<th>PM10 (µg/m³)</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monitory station (n=1250)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policlinico</td>
<td>47.0</td>
<td>30.5</td>
<td>7.0</td>
<td>26.0</td>
<td>38.0</td>
<td>59.0</td>
<td>174.0</td>
</tr>
<tr>
<td>Subjects' residence</td>
<td>44.2</td>
<td>28.3</td>
<td>3.0</td>
<td>24.0</td>
<td>36.0</td>
<td>56.0</td>
<td>171.0</td>
</tr>
<tr>
<td>Average Milan</td>
<td>46.7</td>
<td>29.5</td>
<td>7.7</td>
<td>25.7</td>
<td>37.7</td>
<td>60.0</td>
<td>170.7</td>
</tr>
<tr>
<td><strong>FARM Model Estimate (n=931)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policlinico</td>
<td>34.5</td>
<td>18.0</td>
<td>6.2</td>
<td>21.6</td>
<td>29.8</td>
<td>41.8</td>
<td>104.7</td>
</tr>
<tr>
<td>Subjects' residence</td>
<td>33.8</td>
<td>18.3</td>
<td>4.0</td>
<td>20.4</td>
<td>29.7</td>
<td>42.1</td>
<td>113.1</td>
</tr>
<tr>
<td>Average Milan</td>
<td>35.6</td>
<td>18.7</td>
<td>6.4</td>
<td>22.8</td>
<td>30.4</td>
<td>43.1</td>
<td>113.0</td>
</tr>
</tbody>
</table>
Autumn/Winter

Monitory station (n=747)

<table>
<thead>
<tr>
<th></th>
<th>Policlinico</th>
<th>Subjects’ residence</th>
<th>Average Milan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58.8</td>
<td>33.5</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>33.0</td>
<td>52.0</td>
<td>76.0</td>
</tr>
<tr>
<td></td>
<td>55.0</td>
<td>30.8</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>48.0</td>
<td>72.0</td>
<td>171.0</td>
</tr>
<tr>
<td></td>
<td>58.6</td>
<td>31.8</td>
<td>11.3</td>
</tr>
<tr>
<td></td>
<td>34.3</td>
<td>51.3</td>
<td>76.7</td>
</tr>
<tr>
<td></td>
<td>52.0</td>
<td>76.0</td>
<td>170.7</td>
</tr>
</tbody>
</table>

FARM Model Estimate (n=569)

<table>
<thead>
<tr>
<th></th>
<th>Policlinico</th>
<th>Subjects’ residence</th>
<th>Average Milan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40.1</td>
<td>19.4</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>26.6</td>
<td>35.9</td>
<td>52.8</td>
</tr>
<tr>
<td></td>
<td>39.5</td>
<td>19.6</td>
<td>4.9</td>
</tr>
<tr>
<td></td>
<td>25.8</td>
<td>35.6</td>
<td>51.4</td>
</tr>
<tr>
<td></td>
<td>41.7</td>
<td>20.0</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>27.2</td>
<td>37.0</td>
<td>55.3</td>
</tr>
</tbody>
</table>

Spring/Summer

Monitory station (n=503)

<table>
<thead>
<tr>
<th></th>
<th>Policlinico</th>
<th>Subjects’ residence</th>
<th>Average Milan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>29.4</td>
<td>11.7</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>22.0</td>
<td>27.0</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>28.2</td>
<td>12.2</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>20.0</td>
<td>26.0</td>
<td>35.0</td>
</tr>
<tr>
<td></td>
<td>28.9</td>
<td>11.8</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>21.0</td>
<td>26.7</td>
<td>36.3</td>
</tr>
</tbody>
</table>

FARM Model Estimate (n=362)

<table>
<thead>
<tr>
<th></th>
<th>Policlinico</th>
<th>Subjects’ residence</th>
<th>Average Milan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.6</td>
<td>10.7</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>18.6</td>
<td>23.9</td>
<td>30.6</td>
</tr>
<tr>
<td></td>
<td>24.9</td>
<td>11.2</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>17.6</td>
<td>22.3</td>
<td>30.9</td>
</tr>
<tr>
<td></td>
<td>26.0</td>
<td>10.8</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>18.8</td>
<td>24.2</td>
<td>30.8</td>
</tr>
</tbody>
</table>

Weather variables (all seasons) (n=1250)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative humidity</td>
<td>69.3</td>
<td>18.4</td>
<td>9.0</td>
</tr>
<tr>
<td>Temperature</td>
<td>12.6</td>
<td>7.8</td>
<td>-8.0</td>
</tr>
<tr>
<td>Apparent temperature</td>
<td>11.2</td>
<td>8.8</td>
<td>-9.0</td>
</tr>
</tbody>
</table>

SD: standard deviation, PM: particulate matter, Min: minimum, Q1: first quartile, Q3: third quartile, Max=maximum.

Monitoring stations available from 2010-2013.

FARM Model Estimate available from 2010-2012.

PM$_{10}$ air concentrations depends, besides emissions, on weather conditions during the days, in particular by rainfall, atmospheric stability and wind. Figure 15 shows the distributions of daily mean PM$_{10}$ concentrations of all monitors and of all grid cells with a monitor falling into their boundary. Darker area highlights winter months, characterized by major differences between the two methods of exposure assessment.

The winter months (October to February) are those where there is a more obvious difference between the two methods of exposure assessment; for the rest of the year the distributions are similar. In order to examine model performance of FARM model we used a standard performance metric: the mean fractional bias (MFB), which is a measure of the tendency of the model to over or under predict the observation from monitors. The MFB is defined as the normalized average difference between all model-observed pairs:
\[ MFB = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{C_m - C_o}{C_o} \right) \]

and can vary between 200% and 200%. The model performance goal is met when the MFB is between ±30% [112]. (76,77). We obtain a MFB=-16% for the entire period of comparison and of MFB2010=-3%, MFB2011=-29%, MFB2012=-28%, by year.

Figure 15: Estimated and observed daily mean PM\(_{10}\) concentrations (2010-2012).

These results suggest the fulfillment of the objective of performance, although with a tendency to underestimate the observed concentrations (negative values for the statistical index MFB). The differences between the two methods are consistent with what reported by ARPA Lombardy in the Annual Assessment Of Air Quality Modeling for years 2009-2011. (78).

Figure 16 shows mean concentration (2010-2012) measured by monitors and estimated by the FARM model in the corresponding cell. Each point is the PM\(_{10}\) mean concentration of a single monitor. Very high correlation was observed among the three sources of exposition (RPoliclinico vs Average Milan=0.99; RSubjects’ Residence vs Average Milan=0.94; RPoliclinico vs Subjects’ Residence=0.99). The cone dotted lines represents the range of quality data established by law [113] for the mean of the pollutant over the period, that is equal to ±50% [112, 114]. All the average concentrations fall into the quality range, except for three monitors, quite isolated and far from subjects’ residences. The analysis was performed both by year and over the period 2010-2012 and comparable results were found. When comparing modeling results to observations, the measurements should not be considered the absolute truth. The differences
between the two methods of exposure assessment may be partly explained by the issue related to comparing a point measurement to a volumetric grid cell averaged modeled concentration.

![Graph showing PM\textsubscript{10} concentration comparison](image)

**Figure 16:** PM\textsubscript{10} mean concentration (2010-2012) measured by monitors and estimated by the FARM model in the corresponding cell.

PM modeling use a 4 km on a side in the horizontal and between 10 to 10000 meters in the vertical. Since the modeling results represent an average concentration over the entire grid cell and the observations likely do not represent the average concentration over the same volume of air, it is not possible to exactly match modeled results to measurements\[112]\.

We can summarize as follows the pros e cons of using PM\textsubscript{10} exposure from FARM model air and from monitoring stations:

**FARM model:**
- provides an estimate of PM\textsubscript{10} for a grid of 1678 cells 4x4Km;
- no missing values in PM\textsubscript{10} series;
- is not easy to obtain (formal request to ARPA Lombardy);
- validation requires long times (7-8 months);
- high complexity of the model and the variables that feed it;

**Monitoring Stations:**
- easy availability from the site ARPA air quality;
- validation times from 3 to 6 months, but not validated series are also available;
✓ provides a punctual $\text{PM}_{10}$ measure by 81 air monitoring stations;
✓ lengthy procedure of estimation of missing data that is not always possible when the number of missing is high.

In conclusion, individual air pollution exposure assessment is determined using two sources of information: actual monitor measurements for each and every day starting from January 1st, 1990 and a regional well validated modelling systems applied starting from 2007. This allows us to estimate both short-term (days) and long-term (months, years) exposure to the pollutants. Personal exposure of each study subject will be attributed based on their residential and work addresses and on questionnaire information on their time patterns (time at home, time at work, communing time, number of days at work, etc.). We decide to use FARM model data exposure for the analysis of the association between $\text{PM}_{10}$ exposure and miRNAs expression.

5.3 Comparison between Normalization Strategies.

The performance of the different normalization strategies was assessed by [81]: (1) evaluating their ability to reduce the experimental induced (technical) variation, (2) determining their power to extract true biological variation. The point (1) was assessed calculating the standard deviation (SD) for each individual miRNAs across all samples upon applying different normalization procedures and plotting the cumulative distribution of the deltaCrt SD values (Figure 17).

![Cumulative distribution of deltaCrt SD values upon applying different normalization procedures applied on miRNAs SET3.](image)

---

Lower standard deviations denote better removal of experimentally induced noise: technical variation +
The comparison of the mean standard deviation between the three normalization strategies for the 50% least and 50% more variable miRNAs was performed by means of paired t test (Figure 18), the results can be summarized as follows:

✓ For the 50% least variable miRNAs all three normalization strategies result in a significant decrease of the mean SD value respect to Not normalized data;
✓ For the 50% most variable miRNAs only d and c normalization strategy result in a significant decrease of the mean SD value respect to Not normalized data;
✓ For both 50% least and most variable miRNAs Global mean result in a significant decrease in of the mean SD value respect to Mean 4 more miRNAs data (p < 0.001);
✓ As true differentially expressed miRNAs predominantly reside in the most variable half of the dataset, only Global mean normalization is capable of reducing the number of false negatives. Reduction of false positives is possible with both normalization strategies (d and c) but to different extents as Global mean normalization results in a stronger decrease of technical variation for the 50% least variable miRNAs.

1) 50% least variable miRNAs

2) 50% most variable miRNAs

Figure 18: Comparison of Mean SD between the three normalization strategies for the 50% least and 50% more variable miRNAs was performed by means of paired t test.

The point (2) was assessed by a new experiment aimed to evaluate how different normalization strategies affect biological changes. We consider a single OpenArray plate with three sample and proceeded in the whole-genome profiling of 733 human miRNAs for the three samples, with two replicates performed by the same operator in the same day. Repeated measures analysis by means of proc mixed in SAS was performed in order to:
SAS code

```sas
PROC MIXED DATA=miRNA ;
class time sample ;
model log2_RQ_1 = sample time / ddfm=kenwardroger;
repeated time ;
random samplename;
lsmeans Replicates "replicate 2 vs replicate 1" 1 1 /;
lsmeans Sample "sample 1 vs sample 2" 1 0 -1;
lsmeans Sample "sample 1 vs sample 3" 1 0 -1;
lsmeans Sample "sample 2 vs sample 3" 0 1 -1;
run;
```

1) **Evaluate the normalization strategy performance in reducing technical variation on SET3 (136miRNAs):**

Comparison between the two replicates in terms of miRNAs expression obtained for the three samples. The results are summarized in Table 4 and showed that the lowest percentage (8.82%) of miRNAs differentially expressed is obtained with the Global Mean normalization strategy, as well as the lowest Fold Change range (0.2-1.1).

Table 6: Results of evaluation of normalization strategy performance in reducing technical variation on SET3 (136 miRNAs).

<table>
<thead>
<tr>
<th>Normalization strategy</th>
<th>% of miRNAs differentially expressed replicate 2 vs replicate 1</th>
<th>% of miRNAs differentially expressed replicate 2 vs replicate 1 after FDR adjustment</th>
<th>Fold change range replicate 2 vs replicate 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endogenous</td>
<td>13.50%</td>
<td>0%</td>
<td>0.4-1.4</td>
</tr>
<tr>
<td>Global Mean</td>
<td>8.82%</td>
<td>0%</td>
<td>0.2-1.1</td>
</tr>
<tr>
<td>Mean of 4 more stable</td>
<td>12.20%</td>
<td>0%</td>
<td>0.4-1.4</td>
</tr>
</tbody>
</table>

2) **Evaluate the normalization strategy performance in extracting true biological variation on SET3 (136 miRNAs):**

Comparison between the three samples in terms of miRNAs expression obtained at the two replicates. Results were summarized in Table5. The comparison between the three samples in terms of miRNAs expression obtained at the two replicates showed that the higher percentage (1vs2:47.8%, 1vs3:78.7%, 2vs3:50.0%) of miRNAs differentially expressed is obtained with the Global Mean normalization strategy, as well as the higher Fold Change range (1vs2:0.37-5.7, 1vs3:0.7-9.14, 2vs3:0.6-6.3).
Table 7: Results of evaluation of normalization strategy performance in extracting true biological variation on SET3 (136 miRNAs).

<table>
<thead>
<tr>
<th>Normalization strategy</th>
<th>% of miRNAs differentially expressed</th>
<th>% of miRNAs differentially expressed after FDR adjustment</th>
<th>Fold Change range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1vs2</td>
<td>1vs3</td>
<td>2vs3</td>
</tr>
<tr>
<td>Endogenous</td>
<td>41.9%</td>
<td>47.8%</td>
<td>61.8%</td>
</tr>
<tr>
<td>Global Mean</td>
<td>47.6%</td>
<td>78.7%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Mean of 4 more stable</td>
<td>49.2%</td>
<td>31.6%</td>
<td>43.4%</td>
</tr>
</tbody>
</table>

In conclusion we can state that for large scale miRNA expression profiling Global Mean normalization strategy outperforms the other normalization strategy in terms of:

- better reduction of technical variation:
  - lower % of miRNAs differentially expressed before and after FDR adjustment
  - lower Fold change range;
- more accurate appreciation of biological changes.
  - higher % of miRNAs differentially expressed before and after FDR adjustment;
  - higher Fold Change range;

5.4 Association between miRNAs expression and PM$_{10}$ exposure

In order to verify the association between miRNAs expression and PM$_{10}$ we first fitted multiple linear regression models. As exposure was considered the daily PM$_{10}$ exposure estimate ($\mu$g/m$^3$) from Eulerian model for the 4x4 km cell containing the address of the Center for Obesity and Weight Control. The exposure lag period chosen for the analysis is of zero days (daily exposure of blood collection day). As outcome was considered the Relative quantification RQ of each miRNAs. log2 transformation was applied in order to satisfy the normality and linearity assumptions of linear regression model. We tested the normality assumption of errors using the normal probability plot and the Shapiro-Wilks statistic. Finally, for each miRNA, plots of residuals versus predicted values and Lack of Fit tests were used to explore potential nonlinearity of PM$_{10}$ and apparent temperature. Moreover, test we tested for nonlinearity of PM$_{10}$ and apparent temperature using penalized splines in generalized linear models for all miRNAs included in the analysis by mean of graphical inspection and by formal test of linearity. We excluded a nonlinear relation between PM$_{10}$ or apparent temperature and miRNAs expression. In particular Figure19 reports the results of non-linearity testing for the first nine top miRNAs selected.
The following adjusting variables were selected: age, body mass index, cigarette smoking (never, former, or current), and pack-years. We adjusted for percent of granulocytes (to control for possible shifts in leukocyte differential count), date, Seasonality (using sine and cosine) and apparent temperature. Since in each run of OpenArray were simultaneous analysed up to 4 OpenArray plates, identified by a barcode, for a total of 12 samples (3 per plate) it was possible to identify a hierarchical data structure with three levels: sample level (level-1), barcode level (level-2) and run level (level-3). We developed three-levels HLM using the MIXED procedure in SAS.

Figure 19: Penalized Splines: testing Non linearity on the association of PM$_{10}$ of and first 9 top miRNAs selected.

The use of three-levels hierarchical linear models allowed to investigate other variability sources linked to the outcome. In particular we inspect the following research questions:

1) how much of the variability in miRNAs expression is attributable to barcodes and runs?
2) does the association between of the level-1 predictor PM$_{10}$ vary among barcode or run?

To answer to this questions we proceeded according the following model selection strategy:

- **Model1- Unconditional model:** We fit the following model to answer to the first question. As miRNA variable we use miR$_{106a}$002169 which is the first top miRNA identified in the multivariable simple regression analysis

  SAS code
proc mixed data=dataset method=ml;  
class barcode run;  
model mirna=solution;  
random intercept/sub=run type=vc;  
random intercept/sub=barcode(run) type=vc;  
run;  

We used the three variance estimates to calculate the intraclass correlation coefficients ICCs for barcode and for run:

\[
 ICC_{\text{barcode}} = \frac{\sigma_{\text{barcode}}^2}{\sigma_{\text{barcode}}^2 + \sigma_{\text{run}}^2 + \sigma_{\text{error}}^2} = \frac{0.1407}{0.1407 + 0.7912 + 0.6875} = 0.0868
\]

\[
 ICC_{\text{run}} = \frac{\sigma_{\text{run}}^2}{\sigma_{\text{barcode}}^2 + \sigma_{\text{run}}^2 + \sigma_{\text{error}}^2} = \frac{0.7912}{0.1407 + 0.7912 + 0.6875} = 0.4885
\]

The 8.6% of the variation in miRNAs expression exists between barcode and 48.8% exist between run, leaving 42.6% of the variance in miRNAs expression within samples. Thus a practically meaningful proportion of the variance in miRNA expression exists at the barcode and above all at run levels, providing support for the use of a three-level analytical model.

- **Model 2: Model 1 + level-1 fixed effect.** We included the sample-level predictor PM_{10}  

SAS code

```
proc mixed data=dataset covtest method=ml; 
class barcode run; 
model mirna=PM10_policlinico /solution; 
random intercept /sub=run type=vc; 
random intercept /sub=barcode(run) type=vc;run; 
run; 
```

Results from fixed effects indicate the relationship between level-1 predictor PM_{10} and the outcome miRNAs expression

- **Model 3: Model 2 + random slopes for level-1 predictor PM$_{10}$.** We expanded Model 2 specifying the PM$_{10}$ predictor as random slope at both barcode and run level.

SAS code

```
proc mixed data=dataset covtest method=ml; 
class barcode run; 
model mirna=PM10_policlinico /solution; 
random intercept PM10_policlinico /sub=run type=vc; 
random intercept PM10_policlinico /sub=barcode(run) type=vc;run; 
```
This model allows to answer to the second research question: fixed effects results provide the same information as Model2, random slope results reveal if the relationships between level-1 predictor PM$_{10}$ and the outcome miRNAs expression vary between barcode and run.

Results from the three models can be summarize in Table 8:

In both Model2 and Model3 sample level predictor PM$_{10}$ is significant, suggesting that an increase of 1 (µg/m$^3$) in PM$_{10}$ is associated with a decrease (respectively $\beta$=-0.020 Model2 and $\beta$=-0.012) in miR$_{106a}$ expression. Examining the random effects section, we see that significant variability in barcode and run intercept, as well as the coefficients associated with PM$_{10}$ existed even after controlling for the sample level fixed effect.

Table 8. Estimates from three-level hierarchical linear model predicting miRNA expression (miR$_{106a}$).

<table>
<thead>
<tr>
<th></th>
<th>Model1</th>
<th>Model2</th>
<th>Model3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>11.84</td>
<td>12.13*</td>
<td>12.05</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.10)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>$PM_{10}$</td>
<td>-0.020*</td>
<td>-0.012*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td><strong>Error variance</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level-1</td>
<td>0.6875*</td>
<td>0.6786*</td>
<td>0.6576*</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Intercept(barcode)</td>
<td>0.1407*</td>
<td>0.1434*</td>
<td>0.1334*</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Intercept (run)</td>
<td>0.7912</td>
<td>0.7529</td>
<td>0.8327</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Slope ($PM_{10}$)B</td>
<td></td>
<td></td>
<td>0.03*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Slope ($PM_{10}$)R</td>
<td></td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td><strong>Model Fit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>2418.7</td>
<td>2410.2</td>
<td>2254.8</td>
</tr>
<tr>
<td>BIC</td>
<td>2529.1</td>
<td>2521.9</td>
<td>2389.7</td>
</tr>
</tbody>
</table>

Note: *Statistically significant $p<0.05$; Barcode ICC=, Run ICC= B=random effect at barcode level, R=random effect at run level. Value based on SAS PROC MIXED. Entries show parameter estimates with standard error in parentheses. Estimation Method=ML; Satterthwaite degrees of freedom.

Thus the association between PM$_{10}$ and miRNAs expression varies significantly among barcode and run (as denoting by sub-B and sub R respectively). Thus coefficient associated with this variable may be stronger/weaker from barcode to barcode within a run. Examining the AIC and BIC values at the bottom of Table 6, we can see that each progressive model exhibited better fit to the data. The results obtained from the final model, Model3 adjusted for age, body mass index, cigarette smoking, percent of granulocytes, date, Seasonality (using sine and cosine) and
apparent temperature, were reported in Table 7, Table 8 and Table 9, one for each set of miRNAs used (Set1, Set2 and Set3). Each table reports the significant associations found between miRNAs expression and exposure according to the FDR p-value threshold 0.10 for the first 10 top miRNAs. We can observe that for each set of miRNAs we obtain almost the same first 10 top miRNAs even if with a different order given by the raw and FDR p-value. Looking at the results obtained on the three sets of miRNAs the following list of first 10 top miRNAs were identified: miR_106a_002169, miR_152_000475, miR_181c_002333, miR_218_000521, miR_27b_000409, miR_30d_000420, miR_652_002352, miR_92a_000431, miR_25_000403, miR_375_000564. These miRNAs are now being confirmed by Real Time PCR on the next 1000 enrolled patients enrolled.

**RESULTS: miRNAs SET 1**

Table 9: Association between FARM model estimate of PM$_{10}$ daily exposure and miRNAs expression (SET1: 527 miRNAs)

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>Estimate</th>
<th>95% CI</th>
<th>VARIATION (%)</th>
<th>RAW pvalue</th>
<th>FDR pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR_106a_002169</td>
<td>-0.012</td>
<td>-0.017</td>
<td>-0.007</td>
<td>-8.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_152_000475</td>
<td>-0.013</td>
<td>-0.019</td>
<td>-0.007</td>
<td>-8.69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_181c_002333</td>
<td>0.01</td>
<td>0.006</td>
<td>0.014</td>
<td>6.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_218_000521</td>
<td>-0.015</td>
<td>-0.022</td>
<td>-0.008</td>
<td>-9.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_27b_000409</td>
<td>-0.016</td>
<td>-0.024</td>
<td>-0.009</td>
<td>-10.74</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_30d_000420</td>
<td>-0.016</td>
<td>-0.022</td>
<td>-0.009</td>
<td>-10.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_652_002352</td>
<td>-0.023</td>
<td>-0.033</td>
<td>-0.013</td>
<td>-14.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_674_002021</td>
<td>0.01</td>
<td>0.005</td>
<td>0.014</td>
<td>6.97</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_92a_000431</td>
<td>-0.011</td>
<td>-0.017</td>
<td>-0.006</td>
<td>-7.53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_181a_2_002317</td>
<td>-0.015</td>
<td>-0.022</td>
<td>-0.007</td>
<td>-9.79</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Three-level HLM model adjusted for age, body mass index, cigarette smoking, percent of granulocytes, date, Seasonality (using sine and cosine) and apparent temperature. Only significant FDR-pvalues according to the threshold of 0.10 were shown. Estimate=$2^{(\beta\times10)}$. Variation(%)=$(2^{(\beta\times10)} - 1)\times100$ expresses the percentage variation in miRNAs expression associated with an increase of 10 (µg/m³) in PM$_{10}$.

**RESULTS: miRNAs SET 2**

Table 10: Association between FARM model estimate of PM$_{10}$ daily exposure and miRNAs expression (SET2: 152 miRNAs)

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>Estimate</th>
<th>95% CI</th>
<th>VARIATION (%)</th>
<th>RAW pvalue</th>
<th>FDR pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR_106a_002169</td>
<td>-0.012</td>
<td>-0.017</td>
<td>-0.007</td>
<td>-8.11</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
###RESULTS: miRNAs SET 3

Figure 19: Association between FARM model estimate of PM$_{10}$ daily exposure and miRNAs expression (SET2: 105 miRNAs)

<table>
<thead>
<tr>
<th>miRNAs</th>
<th>Estimate</th>
<th>95% CI</th>
<th>VARIATION (%)</th>
<th>RAW pvalue</th>
<th>FDR pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR_106a_002169</td>
<td>-0.012</td>
<td>-0.017</td>
<td>-0.007</td>
<td>-8.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_152_000475</td>
<td>-0.013</td>
<td>-0.019</td>
<td>-0.007</td>
<td>-8.69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_218_000521</td>
<td>-0.015</td>
<td>-0.022</td>
<td>-0.008</td>
<td>-9.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_27b_000409</td>
<td>-0.016</td>
<td>-0.024</td>
<td>-0.009</td>
<td>-10.74</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_30d_000420</td>
<td>-0.016</td>
<td>-0.022</td>
<td>-0.009</td>
<td>-10.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_652_002352</td>
<td>-0.023</td>
<td>-0.033</td>
<td>-0.013</td>
<td>-14.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_92a_000431</td>
<td>-0.011</td>
<td>-0.017</td>
<td>-0.006</td>
<td>-7.53</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_25_000403</td>
<td>-0.012</td>
<td>-0.018</td>
<td>-0.006</td>
<td>-8.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>miR_720_002895</td>
<td>-0.011</td>
<td>-0.017</td>
<td>-0.006</td>
<td>-7.32</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>let_7c_000379</td>
<td>-0.014</td>
<td>-0.022</td>
<td>-0.007</td>
<td>-9.45</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Three-level HLM model adjusted for age, body mass index, cigarette smoking, percent of granulocytes, date, Seasonality (using sine and cosine) and apparent temperature. Only significant FDR-pvalues according to the threshold of 0.10 were shown. Estimate=$2^\beta\times10$, Variation(%)=$2^\beta\times10-1\times100$ expresses the percentage variation in miRNAs expression associated with an increase of 10 (µg/m$^3$) in PM$_{10}$. 

Three-level HLM model adjusted for age, body mass index, cigarette smoking, percent of granulocytes, date, Seasonality (using sine and cosine) and apparent temperature. Only significant FDR-pvalues according to the threshold of 0.10 were shown. Estimate=$2^\beta\times10$, Variation(%)=$2^\beta\times10-1\times100$ expresses the percentage variation in miRNAs expression associated with an increase of 10 (µg/m$^3$) in PM$_{10}$. 

---
5.5 Mediation Analysis to investigate the role of miRNAs expression as mediator of the effect of PM$_{10}$ on respiratory, cardiac and inflammatory outcome.

5.5.1 Simple Mediation Models

Simple mediation models were applied in order to investigate the role of miRNAs expression as potential mediator on the effect of PM$_{10}$ on respiratory, cardiac and inflammatory outcomes such as: single breath carbon monoxide diffusing capacity DLco, Forced expiratory volume in the 1st second FEV1, Forced Vital Capacity FVC, Heart Rate, Sistolic Blood Pressure SBP, Diastolic Blood Pressure DBP, C-Reactive Protein CRP, and Fibrinogen. Following tables reports the results of simple mediation model involving miRNAs showed a significant indirect effect of PM$_{10}$ on respiratory, cardiac and inflammatory outcomes. The results obtained for the remaining miRNAs were reported in Appendix1.

**DLco**

![Diagram](image)

Figure 21: Statistical Diagram for the Simple mediation model with PM$_{10}$ as independent variable, M as miRNAs expression and DLcoRapp= (Measured DLco / Theoretical DLco)*100 as dependent variable.

_mir_106a_002169_: Results of Simple Mediation Analysis for mediator mir_106a_002169 were reported in Table11:

- the regression analysis representing path c is significant. The F-statistic was 10.428 and the p-value was <0.001. PM$_{10}$ is positively associated with DLcoRapp and the regression coefficient for PM$_{10}$ ($\beta=0.063$ 95% CI: -0.0001-0.127) is statistically significantly
different from zero (p=0.051), this means that two patients who differ by 1 µg/m³ in PM₁₀ exposure level are estimated to differ by 0.063 in DLcoRapp level, the positive sign suggests that patients with higher PM₁₀ exposure show higher DLcoRapp level. This results is quite controversial suggesting the intriguing possibility that short term PM₁₀ exposure differentially alter cytokine pathways;

Table 11: Simple Mediation Analysis results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)*100$ for an increase 1 µg/m³ in PM10.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

DLcoRapp = (Measured DLco / Theoretical DLco)*100

*Sobel test.

- the regression analysis representing path a is significant. The F-statistic was 5.351 and the p-value was <0.001. The regression coefficient for PM₁₀ is statistically significantly different from zero (p<0.001) and PM₁₀ is negatively correlated with miR_106a_002169 expression, in particular, back-transforming results due to the log2 transformation of miRNAs expression data appears that miR_106a_002169 expression decreases by 0.773% for an increase of 1 µg/m³ in PM₁₀;

- the regression analysis representing path b is also significant. The F-statistic was 9.602 and the p-value was <0.001. The regression coefficient for miR_106a_002169 is statistically significantly different from zero (p=0.037) and miR_106a_002169 expression is positively correlated with DLcoRapp, in particular, back-transforming results due to the log2 transformation of miRNAs expression data appears that 1% change in miR_106a_002169 expression is associated with an increase of 0.014 in DLcoRapp;
the regression analysis representing path c' estimated a direct effect c'=0.074 (95% CI:0.010-0.138 p-value=0.023). This represents the estimated difference in DLcoRapp between two patients with the same miR_106a_002169 expression level but who differ by 1 µg/m³ in their PM₁₀ exposure level. The coefficient is positive, meaning that patients with higher PM₁₀ exposure but with the same miR_106a_002169 expression level is estimated to be 0.074 units higher in DLcoRapp. The indirect effect -0.011 means that two patients who differ by 1 µg/m³ in their PM₁₀ exposure level are estimated to by differ -0.011 in DLcoRapp level as a result of the tendency of those with higher PM₁₀ exposure level to have lower miR_106a_002169 expression level (because a is negative), which in turn translates into higher DLcoRapp level (because b is positive). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely below zero (-0.024; -0.003). In this case the normal theory-based Sobel test (Z= -1.852, p=0.064) does not agree with the inference made using a biased correct bootstrap confidence interval. mir_152_000475: Results of Simple Mediation Analysis for mediator mir_152_000475 were reported in Table11.

Table 12: Simple Mediation Analysis results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : (2^β -1)*100 for an increase 1 µg/m³ in PM10. Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp : β log2(101/100) associated with 1%change in miRNAs expression. DLcoRapp= (Measured DLco / Theoretical DLco)*100 *Sobel test.

the regression analysis representing path c is significant. The F-statistic was 10.428 and the p-value was <0.001. PM₁₀ is positively associated with DLcoRapp and the regression coefficient for PM₁₀ (β=0.063 95% CI: -0.0001-0.127) is statistically and significantly
different from zero (p=0.051), this means that two patients who differ by 1 µg/m^3 in PM_{10} exposure level are estimated to differ by 0.063 in DLcoRapp level, the positive sign suggests that patients with higher PM_{10} exposure show higher DLcoRapp level;

- the regression analysis representing path a is significant. The F-statistic was 2.623 and the p-value was 0.016. The regression coefficient for PM_{10} is significantly different from zero (p=0.0319) and PM_{10} is negatively correlated with mir_152_000475 expression, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that mir_152_000475 expression decreases by 0.491% for an increase of 1 µg/m^3 in PM_{10};

- the regression analysis representing path b is also significant. The F-statistic was 10.032 and the p-value was p<0.001. The regression coefficient for mir_152_000475 is statistically significantly different from zero (p=0.008) and mir_152_000475 expression is positively correlated with DLcoRapp, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir_152_000475 expression is associated with an increase of 0.014 in DLcoRapp;

- the regression analysis representing path c’ estimated a direct effect c’=0.070 (95% CI:0.007-0.133 pvalue=0.300). This represents the estimated difference in DLcoRapp between two patients with the same mir_152_000475 expression level but who differ by 1 µg/m^3 in their PM_{10} exposure level. The coefficient is positive, meaning that patients with higher PM_{10} exposure but with the same mir_152_000475 expression level is estimated to be 0.070 units higher in DLcoRapp.

The indirect effect -0.007 means that two patients who differ by 1 µg/m^3 in their PM_{10} exposure level are estimated to by differ -0.007 in DLcoRapp level as a result of the tendency of those with higher PM_{10} exposure level to have lower mir_152_000475 expression level (because a is negative), which in turn translates into higher DLcoRapp level (because b is positive). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely below zero (-0.016; -0.002). In this case the normal theory-based Sobel test (Z= -1.608, p=0.108) does not agree with the inference made using a biased correct bootstrap confidence interval.

mir_218_000521: Results of Simple Mediation Analysis for mediator mir_218_000521 were reported in Table13:

- the regression analysis representing path c is significant. The F-statistic was 10.428 and the p-value was <0.001. PM_{10} is positively associated with DLcoRapp and the regression coefficient for PM_{10} (β=0.063 95% CI: -0.0001-0.127) is statistically significantly
different from zero (p=0.051), this means that two patients who differ by 1 µg/m^3 in PM_{10} exposure level are estimated to differ by 0.063 in DLcoRapp level, the positive sign suggests that patients with higher PM_{10} exposure show higher DLcoRapp level;

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM_{10})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i_j</td>
<td>50.645</td>
<td>9.366</td>
<td>32.257</td>
</tr>
<tr>
<td>M(miR_{218_000521})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c'</td>
<td>0.050</td>
<td>0.032</td>
<td>-0.013</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>b</td>
<td>-1.103</td>
<td>0.320</td>
<td>-1.730</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM_{10} on DLcoRapp</td>
<td>c 0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>Direct Effect of PM_{10} on DLcoRapp</td>
<td>c' 0.050</td>
<td>0.032</td>
<td>-0.013</td>
</tr>
<tr>
<td>Indirect Effect of PM_{10} on DLcoRapp</td>
<td>0.013</td>
<td>0.006</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Boot SE</th>
<th>Boot 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect Effect of PM_{10} on DLcoRapp</td>
<td>0.013</td>
<td>0.006</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 13: Simple Mediation Analysis results for mediator miR_{218_000521} on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM_{10} as dependent variable it was obtained the %change in miRNAs expression: (2^β - 1)*100 for an increase 1 µg/m^3 in PM_{10}.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: β log2(101/100) associated with 1% change in miRNAs expression.

DLcoRapp = (Measured DLco / Theoretical DLco)*100.

*Sobel test.

- the regression analysis representing path a is significant. The F-statistic was 2.773 and the p-value was p=0.011. The regression coefficient for PM_{10} is statistically significantly different from zero (p=0.001) and PM_{10} is negatively correlated with miR_{218_000521} expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that miR_{218_000521} expression decreases by 0.821% for an increase of 1 µg/m^3 in PM_{10};

- the regression analysis representing path b is also significant. The F-statistic was 10.770 and the p-value was <0.001. The regression coefficient for miR_{218_000521} is statistically significantly different from zero (p=0.037) and miR_{218_000521} expression is negatively correlated with DLcoRapp, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in miR_{218_000521} expression is associated with a decrease of 0.016 in DLcoRapp.

- the regression analysis representing path c’ estimated a direct effect c’=0.050 (95% CI:-0.013-0.113 pvalue=0.121). This represents the estimated difference in DLcoRapp between two patients with the same miR_{218_000521} expression level but who differ by 1 µg/m^3 in their PM_{10} exposure level. The coefficient is positive, meaning that patients
with higher PM\textsubscript{10} exposure but with the same miR\textsubscript{218}_000521 expression level is estimated to be 0.050 units higher in DLcoRapp.

The indirect effect 0.013 means that two patients who differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to by differ 0.013 in DLcoRapp level as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have lower miR\textsubscript{218}_000521 expression level (because a is negative), which in turn translates into lower DLcoRapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely below zero (-0.004; -0.027). In this case the normal theory-based Sobel test (Z= 2.313, p=0.027) agrees with the inference made using a biased correct bootstrap confidence interval.

\textit{FEV\textsubscript{1}Rapp}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{diagram.png}
\caption{Statistical Diagram for the Simple mediation model with PM\textsubscript{10} as independent variable, M as miRNAs expression and FEV\textsubscript{1}Rapp= (Measured FEV\textsubscript{1} / Theoretical FEV\textsubscript{1})*100 as dependent variable.}
\end{figure}

\textit{mir\textunderscore 27b\textunderscore 000409}: Results of Simple Mediation Analysis for mediator mir\textunderscore 27b\textunderscore 000409 were reported in Table14:

- the regression analysis representing path c is significant. The F-statistic was 7.217 and the p-value was <0.001. PM\textsubscript{10} is positively associated with FEV1Rapp and the regression coefficient for PM\textsubscript{10} (β=0.025 95% CI: -0.044-0.094) this means that two patients who differ by 1 µg/m\textsuperscript{3} in PM\textsubscript{10} exposure level are estimated to differ by 0.025 in FEV1Rapp level, the positive sign suggests that patients with higher PM\textsubscript{10} exposure show higher FEV1Rapp level. However this relationship is not statistically significantly different from zero (p=0.473);
Table 14: Simple Mediation Analysis results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed. 

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : $2^\beta - 1)*100$ for an increase 1 µg/m^3 in PM10.

Back transforming in model with FEV1 as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV1Rapp : $\beta \log2(101/100)$ associated with 1% change in miRNAs expression.

FEV1Rapp = (Measured FEV1 / Theoretical FEV1) *100

*Sobel test

- the regression analysis representing path a is significant. The F-statistic was 6.203 and the p-value was p<0.001. The regression coefficient for PM10 is statistically significantly different from zero (p=0.001) and PM10 is negatively correlated with mir_27b_000409 expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir_27b_000409 expression decreases by 0.931% for an increase of 1 µg/m^3 in PM10;

- the regression analysis representing path b is also significant. The F-statistic was 6.930 and the p-value was <0.001. The regression coefficient for mir_27b_000409 is statistically significantly different from zero (p=0.031) and mir_27b_000409 expression is negatively correlated with FEV1Rapp, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir_27b_000409 expression decreases by 0.931% for an increase of 1 µg/m^3 in PM10;

- the regression analysis representing path c’ estimated a direct effect $c’=0.016$ (95% CI:-0.053-0.085 pvalue=0.644). This represents the estimated difference in FEV1Rapp between two patients with the same mir_27b_000409 expression level but who differ by 1 µg/m^3 in their PM10 exposure level. The coefficient is positive, meaning that patients with higher PM10 exposure but with the same mir_27b_000409 expression level is estimated to be 0.016 units higher in FEV1Rapp. The indirect effect 0.009 means that two patients who differ by 1 µg/m^3 in their PM10 exposure level are estimated to by
differ 0.009 in FEV1Rapp level as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower mir$_{27b\_000409}$ expression level (because a is negative), which in turn translates into lower FEV1Rapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.002; 0.020). In this case the normal theory-based Sobel test (Z= 1.748, p=0.080) does not agree with the inference made using a biased correct bootstrap confidence interval.

**mir$_{30d\_000420}$:** Results of Simple Mediation Analysis for mediator mir$_{30d\_000420}$ were reported in Table 15:

<table>
<thead>
<tr>
<th>$X(PM_{10})$</th>
<th>$c'$</th>
<th>$M(miR_{30d_000420})$</th>
<th>$Y(FEV_{1_Rapp})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.025</td>
<td>0.035</td>
<td>-0.0436</td>
</tr>
<tr>
<td>$a$</td>
<td>-0.015</td>
<td>0.004</td>
<td>-0.023</td>
</tr>
<tr>
<td>$b$</td>
<td>-0.663</td>
<td>0.336</td>
<td>-1.322</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X(PM_{10})$</th>
<th>$c'$</th>
<th>$M(miR_{30d_000420})$</th>
<th>$Y(FEV_{1_Rapp})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.015</td>
<td>0.035</td>
<td>-0.054</td>
</tr>
<tr>
<td>$a$</td>
<td>-0.015</td>
<td>0.004</td>
<td>-0.023</td>
</tr>
<tr>
<td>$b$</td>
<td>-0.663</td>
<td>0.336</td>
<td>-1.322</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X(PM_{10})$</th>
<th>$c'$</th>
<th>$M(miR_{30d_000420})$</th>
<th>$Y(FEV_{1_Rapp})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.010</td>
<td>0.006</td>
<td>-</td>
</tr>
</tbody>
</table>

The regression analysis representing path c is significant. The F-statistic was 7.217 and the p-value was <0.001. PM$_{10}$ is positively associated with FEV1Rapp and the regression coefficient for PM$_{10}$ ($\beta$=0.025 95% CI: -0.044-0.094) this means that two patients who differ by 1 µg/m$^3$ in PM$_{10}$ exposure level are estimated to differ by 0.025 in FEV1Rapp level, the positive sign suggests that patients with higher PM$_{10}$ exposure show higher FEV1Rapp level. However this relationship is not statistically significantly different from zero (p=0.473);

- the regression analysis representing path a is significant. The F-statistic was 4.073 and the p-value was p<0.001. The regression coefficient for PM$_{10}$ is statistically significantly
different from zero (p<0.001) and PM$_{10}$ is negatively correlated with mir$_{30d\_000420}$ expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir$_{30d\_000420}$ expression decreases by 1.062% for an increase of 1 µg/m$^3$ in PM$_{10}$;

- the regression analysis representing path b is also significant. The F-statistic was 6.827 and the p-value was <0.001. The regression coefficient for mir$_{30d\_000420}$ is statistically significantly different from zero (p=0.048) and mir$_{30d\_000420}$ expression is negatively correlated with FEV$_1$Rapp, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir$_{30d\_000420}$ expression is associated with a decrease of 0.010 in FEV$_1$Rapp.

- the regression analysis representing path c’ estimated a direct effect c’=0.015 (95% CI:-0.054-0.084 pvalue=0.673). This represents the estimated difference in FEV$_1$Rapp between two patients with the same mir$_{30d\_000420}$ expression level but who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level. The coefficient is positive, meaning that patients with higher PM$_{10}$ exposure but with the same mir$_{30d\_000420}$ expression level is estimated to be 0.015 units higher in FEV$_1$Rapp.

The indirect effect 0.010 means that two patients who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to by differ 0.010 in FEV$_1$Rapp level as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower mir$_{30d\_000420}$ expression level (because a is negative), which in turn translates into lower FEV$_1$Rapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.001; 0.023). In this case the normal theory-based Sobel test (Z= 1.742, p=0.082) does not agree with the inference made using a biased correct bootstrap confidence interval.

**mir$_{92a\_000431}$:** Results of Simple Mediation Analysis for mediator mir$_{92a\_000431}$ were reported in Table16:

- the regression analysis representing path c is significant. The F-statistic was 7.217 and the p-value was <0.001. PM$_{10}$ is positively associated with FEV$_1$Rapp and the regression coefficient for PM$_{10}$ ($\beta=0.025$ 95% CI: -0.044-0.094) this means that two patients who differ by 1 µg/m$^3$ in PM$_{10}$ exposure level are estimated to differ by 0.025 in FEV$_1$Rapp level, the positive sign suggests that patients with higher PM$_{10}$ exposure show higher FEV$_1$Rapp level. However this relationship is not statistically significantly different from zero (p=0.473);
Table 16: Simple Mediation Analysis results for mediator mir_92a_000431 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : (2^β -1)*100 for an increase 1 µg/m^3 in PM10.

Back transforming in model with FEV1 as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV1Rapp: β log2(101/100) associated with 1%change in miRNAs expression.

FEV1Rapp = (Measured FEV1 / Theoretical FEV1)*100

*Sobel test

- the regression analysis representing path a is significant. The F-statistic was 4.177 and the p-value was p<0.001. The regression coefficient for PM10 is statistically significantly different from zero (p<0.001) and PM10 is negatively correlated with mir_92a_000431 expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir_92a_000431 expression decreases by 0.808% for an increase of 1 µg/m^3 in PM10;

- the regression analysis representing path b is also significant. The F-statistic was 7.041 and the p-value was <0.001. The regression coefficient for mir_92a_000431 is statistically significantly different from zero (p=0.019) and mir_92a_000431 expression is negatively correlated with FEV1Rapp, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir_92a_000431 expression is associated with a decrease of 0.014 in FEV1Rapp.

- the regression analysis representing path c’ estimated a direct effect c’=0.014 (95% CI:-0.055-0.083 pvalue=0.694). This represents the estimated difference in FEV1Rapp between two patients with the same mir_92a_000431 expression level but who differ by 1 µg/m^3 in their PM10 exposure level. The coefficient is positive, meaning that patients with higher PM10 exposure but with the same mir_92a_000431 expression level is estimated to be 0.014 units higher in FEV1Rapp.
The indirect effect 0.011 means that two patients who differ by 1 µg/m^3 in their PM_{10} exposure level are estimated to by differ 0.011 in FEV_{1}Rapp level as a result of the tendency of those with higher PM_{10} exposure level to have lower mir_92a_000431 expression level (because a is negative), which in turn translates into lower FEV_{1}Rapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.003; 0.027). In this case the normal theory-based Sobel test (Z= 1.953, p=0.080) does not agree with the inference made using a biased correct bootstrap confidence interval.

**mir_181a_2_002317**: Results of Simple Mediation Analysis for mediator mir_181a_2_002317 were reported in Table 17:

- the regression analysis representing path c is significant. The F-statistic was 7.217 and the p-value was <0.001. PM_{10} is positively associated with FEV_{1}Rapp and the regression coefficient for PM_{10} (β=0.025 95% CI: -0.0436-0.094) this means that two patients who differ by 1 µg/m^3 in PM_{10} exposure level are estimated to differ by 0.025 in FEV_{1}Rapp level, the positive sign suggests that patients with higher PM_{10} exposure show higher FEV_{1}Rapp level. However this relationship is not statistically significantly different from zero (p=0.473);

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM_{10})</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>constant</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>c</td>
<td>0.025</td>
<td>0.035</td>
<td>-0.0436</td>
<td>0.094</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consequent</th>
<th>M(mir_181a_2_002317)</th>
<th>Y(FEV_{1}Rapp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antecedent</td>
<td>Coeff</td>
<td>SE</td>
</tr>
<tr>
<td>X(PM_{10})</td>
<td>-0.017</td>
<td>0.004</td>
</tr>
<tr>
<td>constant</td>
<td>1.956</td>
<td>1.153</td>
</tr>
</tbody>
</table>

R^2 = 0.061
F(7,780)=7.217, p<0.001

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM_{10} on FEV_{1}Rapp</td>
<td>c</td>
<td>0.025</td>
<td>0.035</td>
</tr>
<tr>
<td>Direct Effect of PM_{10} on FEV_{1}Rapp</td>
<td>c'</td>
<td>0.011</td>
<td>0.035</td>
</tr>
<tr>
<td>Indirect Effect of PM_{10} on FEV_{1}Rapp</td>
<td>0.014</td>
<td>0.006</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Boot SE</th>
<th>Boot 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect Effect of PM_{10} on FEV_{1}Rapp</td>
<td>0.014</td>
<td>0.005</td>
<td>-0.005</td>
</tr>
</tbody>
</table>

Table 17: Simple Mediation Analysis results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : (2^β -1)*100 for an increase 1 µg/m^3 in PM10.

Back transforming in model with FEV1 as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV1Rapp: β log2(101/100) associated with 1%change in miRNAs expression.

FEV1Rapp = (Measured FEV1 / Theoretical FEV1)*100

*Sobel test.

- the regression analysis representing path a is significant. The F-statistic was 2.870 and the p-value was p<0.001. The regression coefficient for PM_{10} is statistically significantly
different from zero (p<0.001) and PM$_{10}$ is negatively correlated with mir$_{181a_2_002317}$ expression, in particular, back-trasforming results due to the log$_2$ trasformation of miRNAs expression data appears that mir$_{181a_2_002317}$ expression decreases by 1.137% for an increase of 1 µg/m$^3$ in PM$_{10}$;

- the regression analysis representing path b is also significant. The F-statistic was 7.395 and the p-value was <0.001. The regression coefficient for mir$_{181a_2_002317}$ is statistically significantly different from zero (p=0.004) and mir$_{181a_2_002317}$ expression is negatively correlated with FEV$_1$Rapp, in particular, back-trasforming results due to the log$_2$ trasformation of miRNAs expression data appears that 1% change in mir$_{181a_2_002317}$ expression is associated with a decrease of 0.012 in FEV$_1$Rapp.

- the regression analysis representing path c’ estimated a direct effect c’=0.011 (95% CI:-0.058-0.080 pvalue=0.754). This represents the estimated difference in FEV$_1$Rapp between two patients with the same mir$_{181a_2_002317}$ expression level but who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level. The coefficient is positive, meaning that patients with higher PM$_{10}$ exposure but with the same mir$_{181a_2_002317}$ expression level is estimated to be 0.011 units higher in FEV$_1$Rapp.

The indirect effect 0.014 means that two patients who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to by differ 0.014 in FEV$_1$Rapp level as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower mir$_{181a_2_002317}$ expression level (because a is negative), which in turn translates into lower FEV$_1$Rapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.005; 0.026). In this case the normal theory-based Sobel test (Z= 2.269, p=0.023) agrees with the inference made using a biased correct bootstrap confidence interval.

**mir$_{218_000521}$**: Results of Simple Mediation Analysis for mediator mir$_{218_000521}$ were reported in Table18:

- the regression analysis representing path c is significant. The F-statistic was 7.217 and the p-value was <0.001. PM$_{10}$ is positively associated with FEV$_1$Rapp and the regression coefficient for PM$_{10}$ (β=0.025 95% CI: -0.044-0.094) this means that two patients who differ by 1 µg/m$^3$ in PM$_{10}$ exposure level are estimated to differ by 0.025 in FEV$_1$Rapp level, the positive sign suggests that patients with higher PM$_{10}$ exposure show higher FEV$_1$Rapp level. However this relationship is not statistically significantly different from zero (p=0.473);
### Table 18: Simple Mediation Analysis results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^{\beta -1}) \times 100\) for an increase 1 µg/m\(^3\) in PM10.

Back transforming in model with FEV1 as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV1Rapp: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

FEV1Rapp = (Measured FEV1 / Theoretical FEV1)*100

\*Sobel test.

- the regression analysis representing path a is significant. The F-statistic was 3.773 and the p-value was p<0.001. The regression coefficient for PM\(_{10}\) is statistically significantly different from zero (p<0.001) and PM\(_{10}\) is negatively correlated with mir_218_000521 expression, in particular, back-traasforming results due to the log2 trasformation of miRNAs expression data appears that mir_218_000521 expression decreases by 1.110% for an increase of 1 µg/m\(^3\) in PM\(_{10}\);

- the regression analysis representing path b is also significant. The F-statistic was 6.841 and the p-value was p<0.001. The regression coefficient for mir_218_000521 is statistically significantly different from zero (p=0.046) and mir_218_000521 expression is negatively correlated with FEV1Rapp, in particular, back-traasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir_218_000521 expression is associated with a decrease of 0.009 in FEV1Rapp.

- the regression analysis representing path c’ estimated a direct effect c’=0.015 (95% CI:-0.055-0.084 pvalue=0.682). This represents the estimated difference in FEV1Rapp between two patients with the same mir_218_000521 expression level but who differ by 1 µg/m\(^3\) in their PM\(_{10}\) exposure level. The coefficient is positive, meaning that patients with higher PM\(_{10}\) exposure but with the same mir_218_000521 expression level is estimated to be 0.015 units higher in FEV1Rapp.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Consequent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM(_{10}))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M(miR_218_000521)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c</td>
<td>0.025</td>
<td>0.035</td>
<td>-0.0436</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c’</td>
<td>0.015</td>
<td>0.035</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>b</td>
<td>-0.016</td>
<td>0.004</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c’</td>
<td>0.011</td>
<td>0.006</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c’</td>
<td>-0.658</td>
<td>0.329</td>
<td>-1.303</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>i(_2)</td>
<td>1.036</td>
<td>0.971</td>
<td>2.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>i(_2)</td>
<td>1.036</td>
<td>0.971</td>
<td>2.007</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.061 \]
\[ F(7,780)=7.217, p<0.001 \]

\[ R^2 = 0.033 \]
\[ F(8,779)=6.841, p<0.001 \]

\[ R^2 = 0.066 \]
\[ F(7,780)=3.773, p<0.001 \]
The indirect effect 0.011 means that two patients who differ by 1 µg/m^3 in their PM_{10} exposure level are estimated to by differ 0.011 in FEV_1Rapp level as a result of the tendency of those with higher PM_{10} exposure level to have lower mir_218_000521 expression level (because a is negative), which in turn translates into lower FEV_1Rapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.001; 0.023). In this case the normal theory-based Sobel test (Z= 1.771, p=0.077) does not agree with the inference made using a biased correct bootstrap confidence interval.

**FVC Rapp**

![Diagram](image)

Figure 23: Statistical Diagram for the Simple mediation model with PM_{10} as independent variable, M as miRNAs expression and FVC Rapp=(Measured FVC / Theoretical FVC)*100 as dependent variable.

**mir_27b_000409**: Results of Simple Mediation Analysis for mediator mir_27b_000409 were reported in Table19:

- the regression analysis representing path c is significant. The F-statistic was 15.999 and the p-value was <0.001. PM_{10} is positively associated with FVC Rapp and the regression coefficient for PM_{10} (β=0.051 95% CI: -0.016-0.117) this means that two patients who differ by 1 µg/m^3 in PM_{10} exposure level are estimated to differ by 0.051 in FVC Rapp level, the positive sign suggests that patients with higher PM_{10} exposure show higher FVC Rapp level. However this relationship is not statistically significantly different from zero (p=0.134);
Table 19: Simple Mediation Analysis results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

\[FVCRapp = (\text{Measured FVC} / \text{Theoretical FVC}) \times 100\]

* Sobel test.

- the regression analysis representing path a is significant. The F-statistic was 6.203 and the p-value was \(p<0.001\). The regression coefficient for PM10 is statistically significantly different from zero (\(p=0.001\)) and PM10 is negatively correlated with mir_27b_000409 expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir_27b_000409 expression decreases by 0.931% for an increase of 1 µg/m^3 in PM10;

- the regression analysis representing path b is also significant. The F-statistic was 14.739 and the p-value was \(p<0.001\). The regression coefficient for mir_27b_000409 is statistically significantly different from zero (\(p=0.022\)) and mir_27b_000409 expression is negatively correlated with FVCRapp, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir_27b_000409 expression is associated with a decrease of 0.010 in FVCRapp.

- the regression analysis representing path c’ estimated a direct effect \(c’=0.042\) (95% CI:-0.025-0.108 pvalue=0.222). This represents the estimated difference in FVCRapp between two patients with the same mir_27b_000409 expression level but who differ by 1 µg/m^3 in their PM10 exposure level. The coefficient is positive, meaning that patients with higher PM10 exposure but with the same mir_27b_000409 expression level is estimated to be 0.042 units higher in FVCRapp.
The indirect effect 0.009 means that two patients who differ by 1 µg/m^3 in their PM\textsubscript{10} exposure level are estimated to by differ 0.009 in FVCRapp level as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have lower mir\textsubscript{27}b_000409 expression level (because a is negative), which in turn translates into lower FVCRapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.001; 0.021). In this case the normal theory-based Sobel test (Z= 1.826, p=0.068) does not agree with the inference made using a biased correct bootstrap confidence interval.

**mir\textsubscript{92a}_000431**: Results of Simple Mediation Analysis for mediator mir\textsubscript{92a}_000431 were reported in Table20:

- the regression analysis representing path c is significant. The F-statistic was 15.999 and the p-value was <0.001. PM\textsubscript{10} is positively associated with FVCRapp and the regression coefficient for PM\textsubscript{10} (β=0.051 95% CI: -0.016-0.117) this means that two patients who differ by 1 µg/m^3 in PM\textsubscript{10} exposure level are estimated to differ by 0.051 in FVCRapp level, the positive sign suggests that patients with higher PM\textsubscript{10} exposure show higher FVCRapp level. However this relationship is not statistically significantly different from zero (p=0.134);

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM\textsubscript{10})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.051</td>
<td>0.034</td>
<td>-0.0155</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i\textsubscript{3}</td>
<td>97.508</td>
<td>9.289</td>
<td>79.274</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.126 \]
\[ F(7,780)=15.999, p<0.001 \]

\[
\text{mir}_92a_000431: \text{ Simple Mediation Analysis results for mediator mir}_92a_000431 \text{ on log}_2 \text{ scale. Bootstrap standard error and bootstrap 95\% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log}_2 \text{ transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the } \%	ext{change in miRNAs expression} : (2^{\beta} -1)*100 \text{ for an increase } 1 \mu g/m^3 \text{ in PM10. Back transforming in model with FVCapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCapp: } \beta \log 2(101/100) \text{ associated with } 1\% \text{change in miRNAs expression. FVCapp} = (\text{Measured FVC} / \text{Theoretical FVC})*100 \\
*\text{Sobel test.}
\]

91
• the regression analysis representing path a is significant. The F-statistic was 4.177 and the p-value was p<0.001. The regression coefficient for PM\textsubscript{10} is statistically significantly different from zero (p<0.001) and PM\textsubscript{10} is negatively correlated with mir\_92a\_000431 expression, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that mir\_92a\_000431 expression decreases by 0.808\% for an increase of 1 µg/m\textsuperscript{3} in PM\textsubscript{10};

• the regression analysis representing path b is also significant. The F-statistic was 14.531 and the p-value was p<0.001. The regression coefficient for mir\_92a\_000431 is statistically significantly different from zero (p=0.050) and mir\_92a\_000431 expression is negatively correlated with FVCRapp, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir\_92a\_000431 expression is associated with a decrease of 0.011 in FVCRapp.

• the regression analysis representing path c’ estimated a direct effect c’=0.042 (95% CI:-0.025-0.108 p value=0.222). This represents the estimated difference in FVCRapp between two patients with the same mir\_92a\_000431 expression level but who differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level. The coefficient is positive, meaning that patients with higher PM\textsubscript{10} exposure but with the same mir\_92a\_000431 expression level is estimated to be 0.042 units higher in FVCRapp.

The indirect effect 0.009 means that two patients who differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to by differ 0.009 in FVCRapp level as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have lower mir\_92a\_000431 expression level (because a is negative), which in turn translates into lower FVCRapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.001; 0.023). In this case the normal theory-based Sobel test (Z= 1.699, p=0.089) does not agree with the inference made using a biased correct bootstrap confidence interval.

**mir\_181a\_2\_002317**: Results of Simple Mediation Analysis for mediator mir\_181a\_2\_002317 were reported in Table 21:

• the regression analysis representing path c is significant. The F-statistic was 15.999 and the p-value was <0.001. PM\textsubscript{10} is positively associated with FVCRapp and the regression coefficient for PM\textsubscript{10} (β=0.051 95% CI: -0.016-0.117) this means that two patients who differ by 1 µg/m\textsuperscript{3} in PM\textsubscript{10} exposure level are estimated to differ by 0.051 in FVCRapp level, the positive sign suggests that patients with higher PM\textsubscript{10} exposure show higher
FVCRapp level. However this relationship is not statistically significantly different from zero (p=0.134);

- the regression analysis representing path a is significant. The F-statistic was 2.870 and the p-value was p=0.006. The regression coefficient for PM$_{10}$ is statistically significantly different from zero (p<0.001) and PM$_{10}$ is positively correlated with mir_181a_2_002317 expression, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that mir_181a_2_002317 expression increases by 0.292% for an increase of 1 µg/m$^3$ in PM$_{10}$;

- the regression analysis representing path b is also significant. The F-statistic was 15.029 and the p-value was p<0.001. The regression coefficient for mir_181a_2_002317 is statistically significantly different from zero (p=0.007) and mir_181a_2_002317 expression is negatively correlated with FVCRapp, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir_181a_2_002317 expression is associated with a decrease of 0.011 in FVCRapp.

- the regression analysis representing path c’ estimated a direct effect c’=0.038 (95% CI:-0.029-0.104 pvalue=0.266). This represents the estimated difference in FVCRapp between two patients with the same mir_181a_2_002317 expression level but who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level. The coefficient is positive, meaning that

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(mir_181a_2_002317)</th>
<th>Y(FVCRapp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>c 0.051 0.034 -0.0155 0.117 0.134</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>j 97.508 9.289 79.274 115.742 &lt;0.001</td>
<td></td>
</tr>
<tr>
<td>c’ 0.038 0.034 -0.029 0.104 0.266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b -0.778 0.287 -1.342 -0.214 0.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i 2 99.029 9.269 80.835 117.224 &lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 21: Simple Mediation Analysis results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : (2$^β$ -1)*100 for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: $β \log2(101/100)$ associated with 1%change in miRNAs expression. FVCRapp = (Measured FVC / Theoretical FVC)*100

*Sobel test.
patients with higher PM\textsubscript{10} exposure but with the same \textit{mir\textunderscore181a\textunderscore2\textunderscore002317} expression level is estimated to be 0.038 units higher in FVCRapp. The indirect effect 0.013 means that two patients who differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to by differ 0.013 in FVCRapp level as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have higher \textit{mir\textunderscore181a\textunderscore2\textunderscore002317} expression level (because a is positive), which in turn translates into lower FVCRapp level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.003; 0.026). In this case the normal theory-based Sobel test (\textit{Z}=2.186, \textit{p}=0.029) agrees with the inference made using a biased correct bootstrap confidence interval.

\textit{Heart Rate}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig24.png}
\caption{Statistical Diagram for the Simple mediation model with PM\textsubscript{10} as independent variable, M as miRNAs expression and Heart Rate as dependent variable.}
\end{figure}

\textit{mir\textunderscore218\textunderscore000521}: Results of Simple Mediation Analysis for mediator \textit{mir\textunderscore218\textunderscore000521} were reported in Table 22:

- the regression analysis representing path c is significant. The F-statistic was 10.428 and the p-value was <0.001. The regression coefficient for PM\textsubscript{10} is significantly different from zero (p=0.026) and PM\textsubscript{10} is positively associated with Heart Rate and the regression coefficient for PM\textsubscript{10} (β=0.044 95% CI: -0.005-0.082) this means that two patients who differ by 1 µg/m\textsuperscript{3} in PM\textsubscript{10} exposure level are estimated to differ by 0.044 bpm in Heart Rate, the positive sign suggests that patients with higher PM\textsubscript{10} exposure show higher Heart Rate;
- the regression analysis representing path a is significant. The F-statistic was 3.411 and the p-value was p=0.020. The regression coefficient for PM\textsubscript{10} is statistically significantly
different from zero (p<0.001) and PM$_{10}$ is negatively correlated with mir$_{218\_000521}$ expression, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that mir$_{218\_000521}$ expression decreases by 1.151% for an increase of 1 µg/m$^3$ in PM$_{10}$;

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(mir$_{218_000521}$)</th>
<th>Y(Heart Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-0.012 0.003 -0.018 -0.006 &lt;0.001</td>
<td>c' 0.039 0.020 0.001 0.078 0.048</td>
</tr>
<tr>
<td>costant</td>
<td>4.110 0.532 3.066 5.154 &lt;0.001</td>
<td>i$_3$ 51.717 3.265 45.308 58.126 &lt;0.001</td>
</tr>
</tbody>
</table>

Table 22: Simple Mediation Analysis results for mediator mir$_{218\_000521}$ on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1)*100$ for an increase 1 µg/m$^3$ in PM10.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate : $\beta \log2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test.

- the regression analysis representing path b is also significant. The F-statistic was 9.686 and the p-value was p<0.001. The regression coefficient for mir$_{218\_000521}$ is not statistically significantly different from zero (p=0.068) and mir$_{218\_000521}$ expression is negatively correlated with FVCRapp, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir$_{218\_000521}$ expression is associated with a decrease of 0.006 bpm in Heart Rate.

- the regression analysis representing path c' estimated a direct effect c'=0.039 (95% CI:-0.001-0.078, pvalue=0.048). This represents the estimated difference in Heart Rate between two patients with the same mir$_{218\_000521}$ expression level but who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level. The coefficient is positive, meaning that patients with higher PM$_{10}$ exposure but with the same mir$_{218\_000521}$ expression level is estimated to be 0.039 bpm higher in Heart Rate.

The indirect effect 0.004 means that two patients who differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to by differ 0.004 bpm in Heart Rate as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower mir$_{218\_000521}$ expression level (because a is
negative), which in turn translates into lower Heart Rate (because \( b \) is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.001; 0.010). In this case the normal theory-based Sobel test (\( Z=1.762, p=0.114 \)) does not agree with the inference made using a biased correct bootstrap confidence interval.

**SBP: Sistolic Blood Pressure**

Figure 25: Statistical Diagram for the Simple mediation model with \( \text{PM}_{10} \) as independent variable, \( M \) as miRNAs expression Systolic Blood Pressure as dependent variable.

**mir_92a_000431:** Results of Simple Mediation Analysis for mediator mir_92a_000431 were reported in Table 23:

- the regression analysis representing path \( c \) is significant. The F-statistic was 27.761 and the p-value was <0.001. \( \text{PM}_{10} \) is negatively associated with SBP and the regression coefficient for \( \text{PM}_{10} \) (\( \beta=-0.012 \) 95% CI: -0.072; -0.047) this means that two patients who differ by 1 \( \mu g/m^3 \) in \( \text{PM}_{10} \) exposure level are estimated to differ by 0.012 mmHg in SBP level, the negative sign suggests that patients with higher \( \text{PM}_{10} \) exposure show lower SBP level. However this relationship is not statistically significantly different from zero (\( p=0.682 \));
- the regression analysis representing path \( a \) is significant. The F-statistic was 5.454 and the p-value was \( p=0.037 \). The regression coefficient for \( \text{PM}_{10} \) is statistically significantly different from zero (\( p<0.001 \)) and \( \text{PM}_{10} \) is negatively correlated with mir_92a_000431 expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir_92a_000431 expression decreases by 0.821% for an increase of 1 \( \mu g/m^3 \) in \( \text{PM}_{10} \);
Table 23: Simple Mediation Analysis results for mediator mir_92a_000431 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. 
miRNAs expression was log2 transformed. 
Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : (2^β -1)*100 for an increase 1 µg/m^3 in PM10.
Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: β log2(101/100) associated with 1%change in miRNAs expression. 
*Sobel test.

- the regression analysis representing path b is also significant. The F-statistic was 24.172 and the p-value was p<0.001. The regression coefficient for mir_92a_000431 is not statistically significantly different from zero (p=0.124) and mir_92a_000431 expression is negatively correlated with SBP, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in mir_92a_000431 expression is associated with a decrease of 0.011 mmHg in SBP level.

- the regression analysis representing path c’ estimated a direct effect c’=-0.019 (95% CI:-0.079-0.041 pvalue=0.533). This represents the estimated difference in SBP between two patients with the same mir_92a_000431 expression level but who differ by 1 µg/m^3 in their PM10 exposure level. The coefficient is negative, meaning that patients with higher PM10 exposure but with the same mir_92a_000431 expression level is estimated to be 0.019 mmHg lower in SBP level.

The indirect effect 0.007 means that two patients who differ by 1 µg/m^3 in their PM10 exposure level are estimated to by differ 0.007mmHg in SBP level as a result of the tendency of those with higher PM10 exposure level to have lower mir_92a_000431 expression level (because a is negative), which in turn translates into lower SBP level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.001; 0.017). In this case the normal theory-
based Sobel test ($Z=1.409$, $p=0.159$) does not agree with the inference made using a biased correct bootstrap confidence interval.

**DBP: Diastolic Blood Pressure**

No miRNAs showed a significant indirect effect of PM$_{10}$ on Diastolic Blood Pressure.

![Diagram for Simple mediation model with PM$_{10}$ as independent variable, M as miRNAs expression and Diastolic Blood Pressure as dependent variable.](image)

**CRP: C-Reactive Protein**

![Diagram for Simple mediation model with PM$_{10}$ as independent variable, M as miRNAs expression and C-reactive protein as dependent variable.](image)

**mir_106a_002169**: Results of Simple Mediation Analysis for mediator mir_106a_002169 were reported in Table 24:

- the regression analysis representing path $c$ is significant. The F-statistic was 25.951 and the $p$-value was <0.001. and PM$_{10}$ is positively associated with Heart Rate and the regression coefficient for PM$_{10}$ ($\beta=0.002$ 95% CI: -0.002-0.006) this means that two
patients who differ by 1 µg/m^3 in PM_{10} exposure level are estimated to differ by 0.002 mg/l in CRP, the positive sign suggests that patients with higher PM_{10} exposure show higher CRP. However the regression coefficient for PM_{10} is not significantly different from zero (p=0.351);

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM_{10})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>constant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M(miR_{106a_002169})</td>
<td>a</td>
<td>-0.012</td>
<td>0.003</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td>i_1</td>
<td>13.102</td>
<td>0.394</td>
<td>12.330</td>
</tr>
<tr>
<td>c</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.006</td>
</tr>
<tr>
<td>i_3</td>
<td>-4.345</td>
<td>0.317</td>
<td>-4.967</td>
<td>-3.722</td>
</tr>
</tbody>
</table>

R^2 = 0.156

F(6,840)=25.951, p<0.001

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of Pm_{10} on CRP</td>
<td>c</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>Direct Effect of Pm_{10} on CRP</td>
<td>c'</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>Indirect Effect of Pm_{10} on CRP</td>
<td>0.0008</td>
<td>0.0004</td>
<td>-</td>
</tr>
</tbody>
</table>

R^2 = 0.045

F(6,840)=6.532, p=0.001

\[ 2^\beta - 1 \] for an increase 1 µg/m^3 in PM_{10}.

The %change in miRNAs expression: \( 100(1.01^\beta - 1) \) associated with 1% change in miRNAs expression.

*Sobel test.

- the regression analysis representing path a is significant. The F-statistic was 6.532 and the p-value was p<0.001. The regression coefficient for PM_{10} is statistically significantly different from zero (p<0.001) and PM_{10} is negatively correlated with mir_{106a_002169} expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir_{106a_002169} expression decreases by 0.849% for an increase of 1 µg/m^3 in PM_{10};

- the regression analysis representing path b is also significant. The F-statistic was 23.049 and the p-value was p<0.001. The regression coefficient for mir_{106a_002169} is statistically significantly different from zero (p=0.027) and mir_{106a_002169} expression is negatively correlated with CRP, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression and CRP level appears that 1% change in mir_{106a_002169} expression is associated with a decrease of 0.03 mg/l in CRP level.

- the regression analysis representing path c' estimated a direct effect c' = 0.001 (95% CI: 0.003-0.005 pvalue=0.581). This represents the estimated difference in CRP between two

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect Effect of Pm_{10} on CRP</td>
<td>0.0008</td>
<td>0.0004</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Table 24: Simple Mediation Analysis results for mediator mir_{106a_002169} on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature, miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \( 2^\beta - 1 \) for an increase 1 µg/m^3 in PM10.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: \( 100(1.01^\beta - 1) \) associated with 1% change in miRNAs expression.
patients with the same mir_92a_000431 expression level but who differ by 1 µg/m^3 in their PM_{10} exposure level. The coefficient is positive, meaning that patients with higher PM_{10} exposure but with the same mir_106a_002169 expression level is estimated to be 0.001 mg/l lower in CRP level.

The indirect effect 0.0008 means that two patients who differ by 1 µg/m^3 in their PM_{10} exposure level are estimated to by differ 0.0008 mg/l in CRP level as a result of the tendency of those with higher PM_{10} exposure level to have lower mir_106a_002169 expression level (because a is negative), which in turn translates into lower CRP level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.0002; 0.0017). In this case the normal theory-based Sobel test (Z=1.985, p=0.047) agrees with the inference made using a biased correct bootstrap confidence interval.

**mir_652_002352:** Results of Simple Mediation Analysis for mediator mir_652_002352 were reported in Table 24:

- the regression analysis representing path c is significant. The F-statistic was 25.951 and the p-value was <0.001. and PM_{10} is positively associated with Heart Rate and the regression coefficient for PM_{10} (β=0.002 95% CI: -0.002-0.006) this means that two patients who differ by 1 µg/m^3 in PM_{10} exposure level are estimated to differ by 0.002 mg/l in CRP, the positive sign suggests that patients with higher PM_{10} exposure show higher CRP. However the regression coefficient for PM_{10} is not significantly different from zero (p=0.351);

- the regression analysis representing path a is significant. The F-statistic was 5.596 and the p-value was p<0.001. The regression coefficient for PM_{10} is statistically significantly different from zero (p<0.001) and PM_{10} is negatively correlated with mir_652_002352 expression, in particular, back-transforming results due to the log2 trasformation of miRNAs expression data appears that mir_652_002352 expression decreases by 1.513% for an increase of 1 µg/m^3 in PM_{10};

- the regression analysis representing path b is also significant. The F-statistic was 22.949 and the p-value was p<0.001. The regression coefficient for mir_652_002352 is statistically significantly different from zero (p=0.038) and mir_652_002352 expression is negatively correlated with CRP, in particular, back-transfoming results due to the log2 trasformation of miRNAs expression appears that 1% change in mir_652_002352 expression is associated with a decrease of 0.004 mg/l in CRP level.
Table 25: Simple Mediation Analysis results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1)\)*100 for an increase 1 µg/m^3 in PM10.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: \(100(1.01^\beta - 1)\) associated with 1% change in miRNAs expression.

*Sobel test.

- the regression analysis representing path \(c'\) estimated a direct effect \(c' = 0.001\) (95% CI: -0.003-0.005 pvalue=0.534). This represents the estimated difference in CRP between two patients with the same mir_652_002352 expression level but who differ by 1 µg/m^3 in their PM10 exposure level. The coefficient is positive, meaning that patients with higher PM10 exposure but with the same mir_652_002352 expression level is estimated to be 0.001 mg/l higher in CRP level.

The indirect effect 0.0006 means that two patients who differ by 1 µg/m^3 in their PM10 exposure level are estimated to by differ 0.0006 mg/l in CRP level as a result of the tendency of those with higher PM10 exposure level to have lower mir_652_002352 expression level (because \(a\) is negative), which in turn translates into lower SBP level (because \(b\) is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.0001; 0.0014). In this case the normal theory-based Sobel test (Z=1.828, p=0.068) does not agree with the inference made using a biased correct bootstrap confidence interval.
**Fibrinogen**

Figure 28: Statistical Diagram for the Simple mediation model with PM$_{10}$ as independent variable, M as miRNAs expression and Fibrinogen as dependent variable.

**mir\_375\_000564**: Results of Simple Mediation Analysis for mediator mir\_375\_000564 were reported in Table 26:

- the regression analysis representing path $c$ is significant. The F-statistic was 18.709 and the p-value was $<0.001$. PM$_{10}$ is positively associated with Fibrinogen and the regression coefficient for PM$_{10}$ ($\beta=0.217$ 95% CI: -0.016-0.450) this means that two patients who differ by 1 µg/m$^3$ in PM$_{10}$ exposure level are estimated to differ by 0.217 g/l in Fibrinogen, the positive sign suggests that patients with higher PM$_{10}$ exposure show higher Fibrinogen level. Hower the regression coefficient for PM$_{10}$ is not significantly different from zero ($p=0.068$)

- the regression analysis representing path $a$ is significant. The F-statistic was 5.116 and the p-value was $<0.001$. The regression coefficient for PM$_{10}$ is statistically significantly different from zero ($p<0.001$) and PM$_{10}$ is negatively correlated with mir\_375\_000564 expression, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that mir\_375\_000564 expression decreases by 1.185% for an increase of 1 µg/m$^3$ in PM$_{10}$;

- the regression analysis representing path $b$ is also significant. The F-statistic was 16.865 and the p-value was $<0.001$. The regression coefficient for mir\_375\_000564 is statistically significantly different from zero ($p=0.023$) and mir\_375\_000564 expression is negatively correlated with Fibrinogen, in particular, back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in
mir_375_000564 expression is associated with a decrease of 0.030 g/l in Fibrinogen level.

Table 26: Simple Mediation Analysis results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: 
\[(2^\beta - 1)\times 100\] for an increase 1 µg/m^3 in PM10.

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: 
\[\beta \log_2(101/100)\] associated with 1%change in miRNAs expression.
* Sobel test.

- the regression analysis representing path \(c\)' estimated a direct effect \(c' = 0.181\) (95% CI: -0.054-0.415 pvalue=0.130). This represents the estimated difference in Fibrinogen between two patients with the same mir_375_000564 expression level but who differ by 1 µg/m^3 in their PM10 exposure level. The coefficient is positive, meaning that patients with higher PM10 exposure but with the same mir_375_000564 expression level is estimated to be 0.181 g/l higher in Fibrinogen level.

The indirect effect 0.036 means that two patients who differ by 1 µg/m^3 in their PM10 exposure level are estimated to by differ 0.036 g/l in Fibrinogen level as a result of the tendency of those with higher PM10 exposure level to have lower mir_375_000564 expression level (because a is negative), which in turn translates into lower Fibrinogen level (because b is negative). Finally, the indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence interval that is interely above zero (0.002; 0.099). In this case the normal theory-based Sobel test (Z=1.914, p=0.056) agrees with the inference made using a biased correct bootstrap confidence interval.
5.5.2 Parallel Multiple Mediation Models

**DLco**

![Diagram of Parallel Multiple Mediation Model]

Figure 29: Statistical Diagram for the Parallel Multiple Mediation model with PM\(_{10}\) as independent variable, M\(_1\) (miR\(_{106a\_00216}\)), M\(_2\) (miR\(_{152\_000475}\)) and M\(_3\) (miR\(_{218\_000521}\)) as miRNAs expression and DLcoRapp as dependent variable.

Table 27 shows the results of the four best fitting OLS regression models that define the parallel multiple mediator model represented in Figure 29. Regression analysis representing paths \(a_1, a_2\) and \(a_3\) are significant as show the p-values for the F statistics. Moreover, back-trasforming results due to the log2 trasformation of miRNAs expression data, appears that:

- mir\(_{106a\_00216}\) expression decreases by 0.773\% for an increase of 1 \(\mu g/m^3\) in PM\(_{10}\);
- mir\(_{152\_000475}\) expression decreases by 0.491\% for an increase of 1 \(\mu g/m^3\) in PM\(_{10}\);
- mir\(_{218\_000521}\) expression decreases by 0.821\% for an increase of 1 \(\mu g/m^3\) in PM\(_{10}\);

Regression analysis representing paths \(b_1, b_2\) and \(b_3\) is also significant (F statistic 12.973, \(p<0.001\)) and back-trasforming results due to the log2 trasformation of miRNAs expression data appears that 1\% change in:

- mir\(_{375\_000564}\) expression is associated with a decrease of 0.016 in DLcoRapp;
- mir\(_{152\_000475}\) expression is associated with an increase of 0.048 in DLcoRapp;
- mir\(_{218\_000521}\) expression is associated with an of 0.037 in DLcoRapp;

The total effect from estimating DLcoRapp from PM\(_{10}\) alone is statistically significantly different from zero \(\beta=0.064\) (95\%CI: 0.0004;0.127; p-value=0.049), this means that two patients who differ by 1 \(\mu g/m^3\) in PM\(_{10}\) exposure level are estimated to differ by 0.064 in DLcoRapp, the positive sign suggests that patients with higher PM\(_{10}\) exposure show higher DLcoRapp level. Very little of the variance in mir\(_{106a\_00216}\) or mir\(_{152\_000475}\) or mir\(_{218\_000521}\)
expression is explained by PM$_{10}$ (respectively $R^2 = 0.041$, $R^2 = 0.021$, $R^2 = 0.022$), and about a sixth of the variance in DLcoRapp is accounted for by proposed mediators and PM$_{10}$, $R^2 = 0.136$.

The most relevant information pertinent to the process being modeled is the direct and indirect effects of PM$_{10}$ on DLcoRapp. Starting first with the indirect effect through mir$_{106a_00216}$ expression, this indirect effect is estimated as 0.012, meaning that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by 0.012 units in their DLcoRapp through mir$_{106a_00216}$ expression, having higher DLcoRapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM$_{10}$ exposure level to have higher mir$_{106a_00216}$ expression level (because $a_1$ is positive), which in turn translates into higher DLcoRapp level (because $b_1$ is positive).

The second indirect effect of PM$_{10}$ exposure on DLcoRapp modeled through mir$_{152_000475}$ expression, is estimated as -0.024, meaning that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by -0.024 in their DLcoRapp through mir$_{152_000475}$ expression, having lower DLcoRapp (because the indirect effect is negative). This indirect effect is negative as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower mir$_{152_000475}$ expression level (because $a_2$ is negative), which in turn translates into higher DLcoRapp level (because $b_2$ is positive).

The third indirect effect of PM$_{10}$ exposure on DLcoRapp modeled through mir$_{2018_000521}$ expression, is estimated as 0.031, meaning that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by 0.031 in their DLcoRapp through mir$_{2018_000521}$ expression, having higher DLcoRapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower mir$_{2018_000521}$ expression level (because $a_3$ is negative), which in turn translates into lower DLcoRapp level (because $b_3$ is negative).

The total indirect effect of PM$_{10}$ exposure on DLcoRapp obtained summed the indirect effects across all mediators is 0.020, this is positive, meaning that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by 0.020 in their DLcoRapp as a result of the effect of PM$_{10}$ exposure on the mediators, which in turn influence DLcoRapp. The total indirect effect often is not of much interest in a multiple mediator model, and sometimes it will be small even when the specific indirect effects are relatively large, which seems paradoxical. The direct effect, $c' = 0.044$, quantifies the effect of PM$_{10}$ exposure on DLcoRapp independent of the effect of the proposed mediators on DLcoRapp. Irrespective of the effects of PM$_{10}$ exposure on mediators (mir$_{106a_00216}$, mir$_{152_000475}$ and mir$_{2018_000521}$...
expression level) and how those mediators relate to DLcoRapp, an increase of 1 μg/m^3 in PM_{10} exposure is associated with an increase of 0.044 in DLcoRapp (because c’ is positive). The total effect of PM_{10} exposure on DLcoRapp is not determined at all by the mediators proposed as intervening between X and Y. As it was in the simple mediation model, c = 0.064. As promised, this total effect partitions cleanly into the direct effect plus the sum of the specific indirect effects:

\[ c = c’ + a_1b_1 + a_2b_2 + a_3b_3 = 0.044 + (-0.011*1.105) + (-0.007*3.332) + (-0.012*2.599) = 0.064 \]

meaning that the total indirect effect of PM_{10} exposure (i.e., the sum of the specific indirect effects) is difference between the total and direct effects of PM_{10} exposure:

\[ c - c’ = a_1b_1 + a_2b_2 + a_3b_3 = 0.064 - 0.044 = 0.020 \]

The total indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals (0.001;0.039), this supports the claim that mir_106a_00216, mir_152_000475 and mir_218_000521 expression mediate the effect of PM_{10} exposure on DLcoRapp. Also the indirect effects for the mediators mir_152_000475 and mir_2018_000521 expression are statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals. Respectively for mir_152_000475 expression it is interely below zero (-0.005; -0.004) and for mir_2018_000521 expression it is interely above zero (0.016;0.051). These results agree with the inferences made using the normal theory-based Sobel test. The estimate for the specific indirect effects (C1) through miR_106A_002169 minus the specific indirect effect through miR_152_000475 (a_1b_1 . a_2b_2) is 0.012-(-0.024)=0.036 and the 95% BC bootstrap confidence interval is entirely above zero (0.009;0.082) meaning that these specific indirect effect are statistically different from each other.

The estimate for the specific indirect effects (C2) through miR_106A_002169 minus the specific indirect effect through miR_218_000521 (a_1b_1 - a_3b_3) is 0.012-(-0.031)=0.019 and the 95% BC bootstrap confidence interval straddles zero (-0.040;0.000) meaning that these specific indirect effect are not statistically different from each other. Finally, the estimate for the specific indirect effects (C3) through mir_152_000475 minus the specific indirect effect through miR_218_000521 (a_2b_2 - a_3b_3) is -0.024 - 0.031=-0.055 and the 95% BC bootstrap confidence interval is entirely below zero (-0.099;-0.024) meaning that these specific indirect effect are statistically different from each other. It is tempting to treat this as a test of the difference in strength of the mechanisms at work linking X to Y, or that one indirect effect is larger than another in an absolute sense.

However, such an interpretation is justified only if the point estimates for the two specific indirect effects being compared are of the same sign. Consider, for instance, the case where aibi = −0.30 and ajbj = 0.30. A test of the difference between these specific indirect effects may lead
to the claim that their difference is not zero, but this does not imply the mechanisms are of different strength or that one indirect effect is bigger. The point estimates suggest one mechanism results in a positive difference in Y, whereas the other yields a negative difference of equal magnitude. In an absolute sense, they are equal in size by the point estimates, yet statistically different by an inferential test which considers their sign. But one indirect effect is not stronger than the other. Nor can we say that X exerts a larger effect on Y through one of the mediators relative to the other.
Table 27: Parallel Multiple Mediation Analysis results for mediators M1 (miR_106a_002169), M2 (miR_152_000475) and M3 (miR_218_000521) on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. 

miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)*100$ for an increase 1 µg/m³ in PM10.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: $\beta \log_{2}(101/100)$ associated with 1%change in miRNAs expression.

Contrasts for pairwise comparisons between specific indirect effects: (C1) miR_106A_002169 minus miR_152_000475, (C2) miR_106A_002169 minus miR_218_000521, (C3) miR_152_000475 minus miR_218_000521. *Sobel test.
Figure 30: Statistical Diagram for the Parallel Multiple Mediation model with PM$_{10}$ as independent variable, M$_1$ (miR$_{27b\_000409}$), M$_2$ (miR$_{30d\_000420}$), M$_3$ (miR$_{92a\_000431}$), M$_4$ (miR$_{181a\_2\_002317}$), M$_5$ (miR$_{218\_000521}$) as miRNAs expression and FEV$_1$ as dependent variable.

Table 28 shows the results of the four best fitting OLS regression models that define the parallel multiple mediator model represented in Figure 30. Regression analysis representing paths $a_1$, $a_2$, $a_3$, $a_4$, $a_5$ are significant as show the p-values for the F statistics. Moreover, back-transforming results due to the log2 trasformation of miRNAs expression data, appears that:

- miR$_{27b\_000409}$ expression decreases by 0.931 for an increase of 1 µg/m$^3$ in PM$_{10}$;
- miR$_{30d\_000420}$ expression decreases by 1.062 for an increase of 1 µg/m$^3$ in PM$_{10}$;
- miR$_{92a\_000431}$ expression decreases by 0.808 for an increase of 1 µg/m$^3$ in PM$_{10}$;
- miR$_{181a\_2\_002317}$ expression decreases by 1.137 for an increase of 1 µg/m$^3$ in PM$_{10}$;
- miR$_{218\_000521}$ expression decreases by 1.110 for an increase of 1 µg/m$^3$ in PM$_{10}$;
Regression analysis representing paths b1, b2, b3, b4 and b5 is also significant (F statistic 5.078, p<0.001) and back-transforming results due to the log2 trasformation of miRNAs expression data appears that 1% change in:

- miR_27b_000409 expression is associated with a decrease of 0.017 in FEV1Rapp;
- miR_30d_000420 expression is associated with an increase of 0.603 in FEV1Rapp;
- miR_92a_000431 expression is associated with a decrease of 0.730 in FEV1Rapp;
- miR_181a_2_002317 expression is associated with a decrease of 0.877 in FEV1Rapp;
- miR_218_000521 expression is associated with a decrease of 0.157 of in FEV1Rapp;

The total effect from estimating FEV1Rapp from PM10 alone is 0.025 (95%CI: -0.044;0.094; p-value=0.473), this means that two patients who differ by 1 µg/m^3 in PM10 exposure level are estimated to differ by 0.025 in FEV1Rapp, the positive sign suggests that patients with higher PM10 exposure show higher FEV1Rapp level. However this total effect is not statistically significant.

The most relevant information pertinent to the process being modeled is the direct and indirect effects of PM10 on FEV1Rapp. Starting first with the indirect effect through miR_27b_000409 expression, this indirect effect is estimated as 0.0002, meaning that two patients that differ by 1 µg/m^3 in their PM10 exposure level are estimated to differ by 0.0002 in their FEV1Rapp level through miR_27b_000409 expression, having higher FEV1Rapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM10 exposure level to have lower miR_27b_000409 expression level (because a1 is negative), which in turn translates into lower FEV1Rapp level (because b1 is negative).

The second indirect effect of PM10 exposure on FEV1Rapp modeled through miR_30d_000420 expression, is estimated as -0.009, meaning that two patients that differ by 1 µg/m^3 in their PM10 exposure level are estimated to differ by -0.009 in their FEV1Rapp through miR_30d_000420 expression, having lower FEV1Rapp (because the indirect effect is negative). This indirect effect is negative as a result of the tendency of those with higher PM10 exposure level to have lower miR_30d_000420 expression level (because a2 is negative), which in turn translates into higher FEV1Rapp level (because b2 is positive). The third indirect effect of PM10 exposure on FEV1Rapp modeled through miR_92a_000431 expression, is estimated as 0.009, meaning that two patients that differ by 1 µg/m^3 in their PM10 exposure level are estimated to differ by 0.009 in their FEV1Rapp through miR_92a_000431 expression, having higher FEV1Rapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM10 exposure level to have lower miR_92a_000431 expression level (because a3 is negative), which in turn translates into lower FEV1Rapp level (because b3 is negative). The fourth indirect effect of PM10 exposure on FEV1Rapp modeled through miR_92a_000432 expression, is estimated as 0.009, meaning that two patients that differ by 1 µg/m^3 in their PM10 exposure level are estimated to differ by 0.009 in their FEV1Rapp through miR_92a_000432 expression, having higher FEV1Rapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM10 exposure level to have lower miR_92a_000432 expression level (because a4 is negative), which in turn translates into lower FEV1Rapp level (because b4 is negative). The fifth indirect effect of PM10 exposure on FEV1Rapp modeled through miR_92a_000433 expression, is estimated as 0.009, meaning that two patients that differ by 1 µg/m^3 in their PM10 exposure level are estimated to differ by 0.009 in their FEV1Rapp through miR_92a_000433 expression, having higher FEV1Rapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM10 exposure level to have lower miR_92a_000433 expression level (because a5 is negative), which in turn translates into lower FEV1Rapp level (because b5 is negative).
expression level (because a3 is negative), which in turn translates into lower FEV1Rapp level (because b3 is negative).

The fourth indirect effect of PM$_{10}$ exposure on FEV1Rapp modeled through miR$_{181a\_2\_002317}$ expression, is estimated as 0.015, meaning that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by 0.015 in their FEV1Rapp through miR$_{181a\_2\_002317}$ expression, having higher FEV1Rapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower miR$_{181a\_2\_002317}$ expression level (because a4 is negative), which in turn translates into lower FEV1Rapp level (because b4 is negative).

The fifth indirect effect of PM$_{10}$ exposure on FEV1Rapp modeled through miR$_{218\_000521}$ expression, is estimated as 0.003, meaning that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by 0.003 in their FEV1Rapp through miR$_{218\_000521}$ expression, having higher FEV1Rapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM$_{10}$ exposure level to have lower miR$_{218\_000521}$ expression level (because a5 is negative), which in turn translates into lower FEV1Rapp level (because b5 is negative).

The total indirect effect of PM$_{10}$ exposure on FEV1Rapp obtained summed the indirect effects across all mediators is 0.017, this is positive, meaning that that two patients that differ by 1 µg/m$^3$ in their PM$_{10}$ exposure level are estimated to differ by 0.017 in their FEV1Rapp as a result of the effect of PM$_{10}$ exposure on the mediators, which in turn influence FEV1Rapp. The direct effect, c$'$ = 0.009, quantifies the effect of PM$_{10}$ exposure on FEV1Rapp independent of the effect of the proposed mediators on FEV1Rapp. Irrespective of the effects of PM$_{10}$ exposure on mediators (miR$_{27b\_000409}$, miR$_{30d\_000420}$, miR$_{92a\_000431}$, miR$_{181a\_2\_002317}$ and miR$_{218\_000521}$ expression level) and how those mediators relate to FEV1Rapp, an increase of 1 µg/m$^3$ in PM$_{10}$ exposure is associated with an increase of 0.009 in FEV1Rapp (because c$'$ is positive). The total effect of PM$_{10}$ exposure on DLcoRapp is not determined at all by the mediators proposed as intervening between X and Y. As it was in the simple mediation model, c = 0.025. As promised, this total effect partitions cleanly into the direct effect plus the sum of the specific indirect effects:

$$c = c' + a_1b_1 + a_2b_2 + a_3b_3 = 0.044 + (-0.011\times1.105) + (-0.007\times3.332) + (-0.012\times2.599) = 0.064$$

meaning that the total indirect effect of PM$_{10}$ exposure (i.e., the sum of the specific indirect effects) is difference between the total and direct effects of PM$_{10}$ exposure:

$$c - c' = a_1b_1 + a_2b_2 + a_3b_3 = 0.064 - 0.044 = 0.012 - 0.024 + 0.031 = 0.020$$

The total indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals (0.001;0.039), this supports the claim that mir$_{106a\_00216}$,
mir_152_000475 and mir_218_000521 expression mediate the effect of PM$_{10}$ exposure on DLcoRapp. Also the indirect effects for the mediators mir_152_000475 and mir_2018_000521 expression are statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals. Respectively for mir_152_000475 expression it is interely below zero (-0.005; -0.004) and for mir_2018_000521 expression it is interely above zero (0.016;0.051). These results agree with the inferences made using the normal theory-based Sobel test. The estimate for the specific indirect effects (C1) through miR_106A_002169 minus the specific indirect effect through miR_152_000475 ($a_1b_1 - a_2b_2$) is 0.012-(-0.024)=0.036 and the 95% BC bootstrap confidence interval is entirely above zero (0.009;0.082) meaning that these specific indirect effect are statistically different from each other.

The estimate for the specific indirect effects (C2) through miR_106A_002169 minus the specific indirect effect through miR_218_000521 ($a_1b_1 - a_3b_3$) is 0.012-(-0.031)=0.019 and the 95% BC bootstrap confidence interval straddles zero (-0.040;0.000) meaning that these specific indirect effect are not statistically different from each other. Finally, the estimate for the specific indirect effects (C3) through mir_152_000475 minus the specific indirect effect through miR_218_000521 ($a_2b_2 - a_3b_3$) is (-0.024 - 0.031)=-0.055 and the 95% BC bootstrap confidence interval is entirely below zero (-0.099;-0.024) meaning that these specific indirect effect are statistically different from each other. It is tempting to treat this as a test of the difference in strength of the mechanisms at work linking X to Y, or that one indirect effect is larger than another in an absolute sense.
### Table 28: Parallel Multiple Mediation Analysis results for mediators M1 (miR_106a_002169), M2 (miR_152_000475), M3 (miR_218_000521), M4 (miR_181a_2_002317), M5 (miR_218_000521) on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1)\times 100\) for an increase 1 µg/m³ in PM10.

Back transforming in model with FEV1Rapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV1Rapp: \(\beta \log2(101/100)\) associated with 1%change in miRNAs expression.

**Contrasts for pairwise comparisons between specific indirect effects:**
- \((C1)\) miR_27B_000409 minus miR_30D_000420
- \((C2)\) miR_27B_000409 minus miR_92A_000431
- \((C3)\) miR_27B_000409 minus miR_181A_2_002317
- \((C4)\) miR_27B_000409 minus miR_218_000521
- \((C5)\) miR_30D_000420 minus miR_92A_000431
- \((C6)\) miR_30D_000420 minus miR_181A_2_002317
- \((C7)\) miR_30D_000420 minus miR_218_000521
- \((C8)\) miR_92A_000431 minus miR_181A_2_002317
- \((C9)\) miR_92A_000431 minus miR_218_000521
- \((C10)\) miR_181A_2_002317 minus miR_218_000521

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M1 (miR_27B_000409)</th>
<th>M2 (miR_30D_000420)</th>
<th>M3 (miR_82a_000431)</th>
<th>M4 (miR_181a_2_002317)</th>
<th>M5 (miR_218_000521)</th>
<th>F(7,780) = 6.203, p&lt;0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.002</td>
<td>0.007</td>
<td>-</td>
<td>0.973*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>-0.009</td>
<td>0.009</td>
<td>-</td>
<td>0.326*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.009</td>
<td>0.009</td>
<td>-</td>
<td>0.258*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.015</td>
<td>0.010</td>
<td>-</td>
<td>0.140*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>0.071</td>
<td>0.071</td>
<td>-</td>
<td>0.731*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.007</td>
<td>0.001</td>
<td>0.322</td>
</tr>
<tr>
<td>M2</td>
<td>0.007</td>
<td>0.014</td>
<td>0.132</td>
</tr>
<tr>
<td>M3</td>
<td>0.011</td>
<td>0.013</td>
<td>0.099</td>
</tr>
<tr>
<td>M4</td>
<td>0.007</td>
<td>0.002</td>
<td>0.264</td>
</tr>
<tr>
<td>M5</td>
<td>0.011</td>
<td>0.013</td>
<td>0.099</td>
</tr>
</tbody>
</table>

**Pairwise comparisons between specific indirect effects:**

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>0.008</td>
<td>0.013</td>
<td>0.100</td>
</tr>
<tr>
<td>M2</td>
<td>0.008</td>
<td>0.018</td>
<td>0.086</td>
</tr>
<tr>
<td>M3</td>
<td>0.012</td>
<td>0.014</td>
<td>0.067</td>
</tr>
<tr>
<td>M4</td>
<td>0.006</td>
<td>0.012</td>
<td>0.029</td>
</tr>
<tr>
<td>M5</td>
<td>0.012</td>
<td>0.016</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Sobel test.
Figure 31: Statistical Diagram for the Parallel Multiple Mediation model with PM\textsubscript{10} as independent variable, M\textsubscript{1} (miR\textsubscript{27b}_000409), M\textsubscript{2} (miR\textsubscript{92a}_000431), M\textsubscript{3} (miR\textsubscript{181a}_2_002317) as miRNAs expression and FVCRapp as dependent variable.

Table 29 shows the results of the the four best fitting OLS regression models that define the parallel multiple mediator model represented in Figure 31. Regression analysis representing paths a\textsubscript{1},a\textsubscript{2} and a\textsubscript{3} are significant as show the p-values for the F statistics. Moreover, back-transforming results due to the log2 trasformation of miRNAs expression data, appears that:
- mir\textsubscript{27b}_000409 expression decreases by 0.931\% for an increase of 1 µg/m\textsuperscript{3} in PM\textsubscript{10};
- mir\textsubscript{92a}_000431 expression decreases by 0.808\% for an increase of 1 µg/m\textsuperscript{3} in PM\textsubscript{10};
- mir\textsubscript{181a}_2_002317 expression decreases by 1.137\% for an increase of 1 µg/m\textsuperscript{3} in PM\textsubscript{10};

Regression analysis representing paths b\textsubscript{1},b\textsubscript{2} and b\textsubscript{3} is also significant (F statistic 12.050, p<0.001) and back-transforming results due to the log2 trasformation of miRNAs expression data appears that 1\% change in:
- mir\textsubscript{27b}_000409 expression is associated with a decrease of 0.003 g/l in FVCRapp;
- mir\textsubscript{92a}_000431 expression is associated with a decrease of 0.003 g/l in FVCRapp;
- mir\textsubscript{181a}_2_002317 expression is associated with an increase of 0.006 g/l in FVCRapp;

The estimate for the total effect obtained estimating FVCRapp from PM\textsubscript{10} alone is $\beta=0.051$ (95\%CI: -0.016;0.117; p-value=0.134), this means that two patients who differ by 1 µg/m\textsuperscript{3} in PM\textsubscript{10} exposure level are estimated to differ by 0.051 in FVCRapp, the positive sign suggests that patients with higher PM\textsubscript{10} exposure show higher FVCRapp level. However the total effect is not
statistically significantly different from zero. Very little of the variance in mir_106a_00216 or mir_152_000475 or mir_2018_000521 expression is explained by PM\textsubscript{10} (respectively R\textsuperscript{2} = 0.053, R\textsuperscript{2} = 0.036, R\textsuperscript{2} = 0.025), and about a sixth of the variance in FVCRapp is accounted for by proposed mediators and PM\textsubscript{10}, R\textsuperscript{2} = 0.134.

The most relevant information pertinent to the process being modeled is the direct and indirect effects of PM\textsubscript{10} on FVCRapp. Starting first with the indirect effect through mir\textsubscript{27b}_000409 expression, this indirect effect is estimated as 0.002, meaning that two patients that differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to differ by 0.002 in their FVCRapp through mir\textsubscript{27b}_000409 expression, having higher FVCRapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have lower mir\textsubscript{27b}_000409 expression level (because a\textsubscript{1} is negative), which in turn translates into lower FVCRapp level (because b\textsubscript{1} is negative).

The second indirect effect of PM\textsubscript{10} exposure on FVCRapp modeled through mir\textsubscript{92a}_000431 expression, is estimated as 0.002, meaning that two patients that differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to differ by 0.002 in their FVCRapp through mir\textsubscript{92a}_000431 expression, having lower FVCRapp (because the indirect effect is negative). This indirect effect is negative as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have lower mir\textsubscript{92a}_000431 expression level (because a\textsubscript{2} is negative), which in turn translates into lower FVCRapp level (because b\textsubscript{2} is negative).

The third indirect effect of PM\textsubscript{10} exposure on FVCRapp modeled through mir\textsubscript{181a}_2_002317 expression, is estimated as 0.010, meaning that two patients that differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to differ by 0.010 in their FVCRapp through mir\textsubscript{181a}_2_002317 expression, having higher FVCRapp (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM\textsubscript{10} exposure level to have lower mir\textsubscript{181a}_2_002317 expression level (because a\textsubscript{3} is negative), which in turn translates into lower FVCRapp level (because b\textsubscript{3} is negative).

The total indirect effect of PM\textsubscript{10} exposure on FVCRapp obtained summed the indirect effects across all mediators is 0.014, this is positive, meaning that two patients that differ by 1 µg/m\textsuperscript{3} in their PM\textsubscript{10} exposure level are estimated to differ by 0.014 in their FVCRapp as a result of the effect of PM\textsubscript{10} exposure on the mediators, which in turn influence FVCRapp. The direct effect, c\textsuperscript{′} = 0.036, quantifies the effect of PM\textsubscript{10} exposure on FVCRapp independent of the effect of the proposed mediators on FVCRapp. Irrespective of the effects of PM\textsubscript{10} exposure on mediators (mir\textsubscript{27b}_000409, mir\textsubscript{92a}_000431 and mir\textsubscript{181a}_2_002317 expression level) and how those mediators relate to FVCRapp, an increase of 1 µg/m\textsuperscript{3} in PM\textsubscript{10} exposure is associated
with an increase of 0.036 in FVCRapp (because $c'$ is positive). The total effect of PM$_{10}$ exposure on FVCRapp is not determined at all by the mediators proposed as intervening between X and Y. As it was in the simple mediation model, $c = 0.051$. As promised, this total effect partitions cleanly into the direct effect plus the sum of the specific indirect effects:

$$c = c' + a_1b_1 + a_2b_2 + a_3b_3 = 0.036 + (-0.014* -0.175) + (-0.012* -0.182) + (-0.017* -0.588) = 0.051$$
meaning that the total indirect effect of PM$_{10}$ exposure (i.e., the sum of the specific indirect effects) is difference between the total and direct effects of PM$_{10}$ exposure:

$$c - c' = a_1b_1 + a_2b_2 + a_3b_3 = 0.051 - 0.036 = 0.002 + 0.002 + 0.010 = 0.051$$

The total indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals (0.002; 0.029), this supports the claim that mir$_{27b}$ $_{000409}$, mir$_{92a}$$_{000431}$ and mir$_{181a}$$_{2}_002317$ expression mediate the effect of PM$_{10}$ exposure on FVCRapp. However, the indirect effects for all the mediators are not statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals which straddle zero. These results agree with the inferences made using the normal theory-based Sobel test.

The estimate for the specific indirect effects (C1) through miR$_{27b}$$_{000409}$ minus the specific indirect effect through miR$_{92a}$$_{000431}$ ($a_1b_1 - a_2b_2$) is 0.0024 - 0.0021 = 0.0003.

The estimate for the specific indirect effects (C2) through miR$_{27b}$$_{000409}$ minus the specific indirect effect through miR$_{181a}$$_{2}_002317$ ($a_1b_1 - a_3b_3$) is 0.0024 - 0.0097 = -0.007.

The estimate for the specific indirect effects (C3) through miR$_{92a}$$_{000431}$ minus the specific indirect effect through miR$_{181a}$$_{2}_002317$ ($a_2b_2 - a_3b_3$) is 0.0021 - 0.0097 = -0.008.

These specific indirect effect are not statistically different from zero as shown by the 95% BC bootstrap confidence intervals.
Table 29: Parallel Multiple Mediation Analysis results for mediators M1 (miR_27b_000409), M2 (miR_92a_000431) and M3 (miR_181a_2_002317) on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycaated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1)\times 100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: \(\beta \log2(10^{1/100})\) associated with 1% change in miRNAs expression.

Contrasts for pairwise comparisons between specific indirect effects: (C1) miR_27b_000409 minus miR_92A_000431, (C2) miR_27b_000409 minus miR_181a_2__002317, (C3) miR_92a_000431 minus miR_181a_2__002317.

*Sobel test.
Table 30 shows the results of the three best fitting OLS regression models that define the parallel multiple mediator model represented in Figure 32. Regression analysis representing paths $a_1$, $a_2$ are significant as show the p-values for the F statistics. Moreover, back-transforming results due to the log2 transformation of miRNAs expression data, appears that:

- mir_27b_000409 expression decreases by -0.979% for an increase of 1 µg/m$^3$ in PM$_{10}$;
- mir_652_002352 expression decreases by -1.459% for an increase of 1 µg/m$^3$ in PM$_{10}$;

Regression analysis representing paths $b_1$, and $b_2$ is also significant (F statistic 17.781, p<0.001) and back-transforming results due to the log2 trasformation of miRNAs expression data and CRP levels appears that 1% change in:

- mir_27b_000409 expression is associated with a decrease of 0.011 g/l in CRP;
- mir_652_002352 expression is associated with a decrease of 0.024 g/l in CRP;

The estimate for the total effect obtained estimating CRP from PM$_{10}$ alone is $\beta=0.002$ (95%CI: -0.002; 0.006; p-value=0.3946), this means that two patients who differ by 1 µg/m$^3$ in PM$_{10}$ exposure level are estimated to differ by 0.002 mg/l in CRP level, the positive sign suggests that patients with higher PM$_{10}$ exposure show higher CRP level. However the total effect is not statistically significantly different from zero. Very little of the variance in mir_27b_000409 or mir_92a_000431 expression is explained by PM$_{10}$ (respectively $R^2 = 0.053$, $R^2 = 0.037$), and about a sixth of the variance in FVCRapp is accounted for by proposed mediators and PM$_{10}$, $R^2 = 0.162$. 

Figure 32: Statistical Diagram for the Parallel Multiple Mediation model with PM$_{10}$ as independent variable, $M_1$ (miR_27b_000409), $M_2$ (miR_652_002352) as miRNAs expression and CRP as dependent variable.
The most relevant information pertinent to the process being modeled is the direct and indirect effects of PM\(_{10}\) on CRP. Starting first with the indirect effect through mir\(_{27b\_000409}\) expression, this indirect effect is estimated as 0.0002, meaning that two patients that differ by 1 \(\mu\text{g/m}^3\) in their PM\(_{10}\) exposure level are estimated to differ by 0.0002 mg/l in their CRP through mir\(_{27b\_000409}\) expression, having higher CRP (because the indirect effect is positive). This indirect effect is positive as a result of the tendency of those with higher PM\(_{10}\) exposure level to have lower mir\(_{27b\_000409}\) expression level (because \(b_1\) is negative), which in turn translates into lower CRP level (because \(b_2\) is negative).

The second indirect effect of PM\(_{10}\) exposure on CRP modeled through mir\(_{652\_002352}\) expression, is estimated as 0.0005, meaning that two patients that differ by 1 \(\mu\text{g/m}^3\) in their PM\(_{10}\) exposure level are estimated to differ by 0.0005 in their CRP through mir\(_{652\_002352}\) expression, having higher CRP (because the indirect effect is positive). This indirect effect is negative as a result of the tendency of those with higher PM\(_{10}\) exposure level to have lower mir\(_{652\_002352}\) expression level (because \(b_2\) is negative), which in turn translates into lower CRP level (because \(b_2\) is negative).

The total indirect effect of PM\(_{10}\) exposure on CRP obtained summed the indirect effects across all mediators is 0.0007, this is positive, meaning that that two patients that differ by 1 \(\mu\text{g/m}^3\) in their PM\(_{10}\) exposure level are estimated to differ by 0.0007 in their CRP as a result of the effect of PM\(_{10}\) exposure on the mediators, which in turn influence CRP. The direct effect, \(c' = 0.001\), quantifies the effect of PM\(_{10}\) exposure on CRP independent of the effect of the proposed mediators on CRP. Irrespective of the effects of PM\(_{10}\) exposure on mediators (mir\(_{27b\_000409}\), mir\(_{652\_002352}\) expression level) and how those mediators relate to CRP, an increase of 1 \(\mu\text{g/m}^3\) in PM\(_{10}\) exposure is associated with an increase of 0.001 in CRP (because \(c'\) is positive).

The total effect of PM\(_{10}\) exposure on CRP is not determined at all by the mediators proposed as intervening between X and Y. As it was in the simple mediation model, \(c = 0.002\). As promised, this total effect partitions cleanly into the direct effect plus the sum of the specific indirect effects:

\[
c = c' + a_1b_1 + a_2b_2 = 0.001 + (-0.014\times -0.011) + (-0.021\times -0.024) = 0.002
\]

meaning that the total indirect effect of PM\(_{10}\) exposure (i.e., the sum of the specific indirect effects) is difference between the total and direct effects of PM\(_{10}\) exposure:

\[
c - c' = a_1b_1 + a_2b_2 = 0.002 - 0.001 = 0.0002 +0.0005= 0.0007
\]

The total indirect effect is statistically different from zero, as revealed by the 95% BC bootstrap confidence intervals (0.0001;0.0014), this supports the claim that mir\(_{27b\_000409}\), mir\(_{652\_002352}\) expression mediate the effect of PM\(_{10}\) exposure on CRP. However, the indirect effects for all the mediators are not statistically different from zero, as revealed by the
95% BC bootstrap confidence intervals which straddle zero. These results agree with the inferences made using the normal theory-based Sobel test.

The estimate for the specific indirect effects (C1) through miR_27b_000409 minus the specific indirect effect through miR_652_002352 \((a_1b_1 - a_2b_2)\) is \(0.0002 - 0.0005 = -0.0004\), and the 95% BC bootstrap confidence interval straddles zero meaning that it is not statistically different from zero.
Table 30: Parallel Multiple Mediation Analysis results for mediators M1 (miR_27b_000409) and M2 (miR_652_002352) on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, temperature.

miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression : $(2^β - 1)\times 100$ for an increase 1 µg/m$^3$ in PM10.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100(1.01^β - 1)$ associated with 1% change in miRNAs expression.

Contrasts for pairwise comparisons between specific indirect effects: (C1) miR_27b_000409 minus miR_652_002352

*Sobel test.
The proposed research will help to shed light on the chain of events that from air pollution exposure leads to CVD trying to explore a new mechanism which involves alteration of extracellular vesicles production and content. Our findings, if confirmed, could lead to the identification of potentially-reversible alterations that might be also considered as potential target for new diagnostic and therapeutic interventions. Results on the first 1000 enrolled subjects have highlighted a statistical significant association at level 0.10 after FDR adjustment of miR_106a_002169, miR_152_000475, miR_181a_2__002317, miR_218_000521, miR_27b_000409, miR_30d_000420, miR_652_002352, miR_92a_000431, miR_25_000403, miR_375_000564 expression with PM$_{10}$ exposure estimated by Eulerian model of the cell containing the address of the Center for Obesity. The use of the daily FARM PM$_{10}$ exposure estimate of the 4*4 km cell of subject residence for the same day of blood collection produced similar results. The candidate miRNAs were subjected to pathway exploration using the Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA) . Using this software, top-ranked pathways were also determined. Moreover are provided the biological functions associated with these miRNAs, such as the diseases in which they are involved (Figure 33).

![Figure 33: Biological functions and diseases associated with candidate miRNAs.](image-url)
Inflammatory diseases, inflammatory response and respiratory disease resulted the first three top-ranked pathways.

A new experiment was set up in order to evaluate the performance of different normalization strategies in reducing technical variation and extracting true biological variation. Results showed that for large scale miRNA expression profiling Global Mean normalization strategy outperforms the other normalization strategy in terms of:

- better reduction of technical variation:
  - lower % of miRNAs differentially expressed before and after FDR adjustment
  - lower Fold change range;
- more accurate appreciation of biological changes.
  - higher % of miRNAs differentially expressed before and after FDR adjustment;
  - higher Fold Change range;

Individual air pollution exposure assessment is determined using two sources of information: actual monitor measurements for each and every day starting from January 1st, 1990 and The FARM model, a regional well validated modelling systems applied starting from 2007. The winter months (October to February) are those where there is a more obvious difference between the two methods of exposure assessment; for the rest of the year the distributions are similar. The examination of model performance of FARM model suggests the fulfillment of the objective of performance, although with a tendency to underestimate the observed concentrations. However the differences between the two methods are consistent with what reported by ARPA Lombardy in the Annual Assessment Of Air Quality Modeling for years 2009-2011, and the huge amount of available data offer the opportunity to implement new modelling techniques as long as they might come out.

Potential non linearity, was investigated by means of generalized additive models using penalized splines. We tested for nonlinearity of PM$_{10}$ and apparent temperature using penalized splines in generalized linear models. We excluded a nonlinear relation between PM$_{10}$ or apparent temperature and miRNAs expression. Mixed effect models was developed in order to take into account of other variability sources linked to the outcome due to the hierarchical data structure with three levels: sample level (level-1), barcode level (level-2) and run level (level-3). Results showed that the association between PM$_{10}$ and miRNAs expression varies significantly among barcode and run, providing inputs for the biologists in order to improve the repeatability of the experiment in future analysis.

Simple Mediation Models and Parallel Multiple Mediation Models have proven useful and appropriate tools to investigate the role of miRNAs expression as mediator of the effect of PM$_{10}$
on respiratory, cardiac and inflammatory outcomes. We estimated the Indirect Effect of PM$_{10}$ on outcome through the mediator as the product $a*b$, and applied Sobel test for testing the null hypothesis that the “true” indirect effect is zero, with the p-value derived from the standard normal distribution (Normal Theory Approach). However the sampling distribution of the indirect effect $a*b$ tends to be asymmetric, with nonzero skewness and kurtosis. We applied bootstrapping as alternative more powerful than the Normal Theory Approach. In particular we calculated percentiles bootstrap confidence intervals, because they are based entirely on values of $ab^*$ that demarcate the upper and lower $(100 - ci)/2\%$ of the distribution of k bootstrap estimates of the indirect effect, and bias-corrected bootstrap confidence intervals (BC bootstrap confidence intervals). They are like percentile confidence intervals but the endpoints are adjusted as a function of the proportion of k values of $ab^*$ that are less than ab, the point estimate of the indirect effect calculated in the original data. The endpoints will be adjusted upward or downward to varying degrees depending on that proportion. Although bootstrapping is recommended, it does have a few weaknesses, among them that it requires the original data (not usually a real problem typically), the endpoints of the confidence interval will vary from run to run (but not if you seed the random number generator yourself), and it isn’t implemented in all software one might choose to use. An alternative to get around these problems: Monte Carlo confidence intervals. A significant indirect effect of PM$_{10}$ on:

- DLcoRapp, was found through the following mediators: mir_106a_002169, mir_152_000475, mir_218_000521 expression;
- FEV1Rapp was found through the following mediators: mir_27b_000409 mir_30d_000420 mir_92a_000431 mir_181a_2_002317 mir_218_000521 expression;
- FVC_Rapp was found through the following mediators: mir_27b_000409, mir_92a_000431 and mir_181a_2_002317 expression;
- Heart Rate was found through the following mediator: mir_218_000521 expression;
- Sistolic Blood Pressure was found through the following mediator: mir_92a_000431 expression;
- CRP was found through the following mediator: mir_106a_002169 and mir_652_002352 expression.
- Fibrinogeno was found through the following mediator: mir_375_000564 expression.

Finally, a the total indirect effect of PM$_{10}$ exposure:

- on DLcoRapp obtained summed the indirect effects across all mediators: mir_106a_002169, mir_152_000475, and mir_218_000521 expression was statistically different from zero;
- on FEV1Rapp obtained summed the indirect effects across all mediators: mir_27b_000409 mir_30d_000420 mir_92a_000431 mir_181a_2_002317 mir_218_000521 expression was statistically different from zero;

- on FVCRapp obtained summed the indirect effects across all mediators mir_27b_000409, mir_92a_000431 and mir_181a_2_002317 expression was statistically different from zero;

- on CRP obtained summed the indirect effects across all mediators mir_106a_002169 and mir_652_002352 expression was statistically different from zero;

From a literature review [115,118], has emerged an important role of these miRNAs, for example mir_106a_002169, mir_92a_000431 and mir_652_002352 effectively regulate macrophage inflammatory responses through modulating leukocyte SIRPa synthesis. As an important signaling molecule, SIRPa modulates many aspects of leukocyte inflammatory responses, including activation, chemotaxis, and phagocytosis. These miRNAs are involved in macrophage infiltration through the direct suppression of expression of signal-regulatory protein α (SIRP α). A downregulation of mir_106a_002169 and mir_92a_000431 seems to act in favor of the differentiation of monocytes into macrophages. Macrophage infiltration occurs in many tissue types, such as adipose tissue, the vascular system, liver and muscle and it is involved in the processes of formation of atherosclerotic plaques. The resident macrophages in tissues and organs play critical roles in controlling physiological functions and systemic homeostasis in tissues. miRNAs are a set of potent regulators of macrophage differentiation, infiltration, activation and cell-cell interactions.

Future directions in research activities will be:

- Apply mediation analysis to the three-levels HLM models developed.
- Deepen and apply a the modern approach to mediation analysis based on counterfactual framework;
- Replicate the analysis on data coming from standard real time-PCR validation on the set of 10 top miRNAs identified in stage 1 on 1000 subjects enrolled in stage2;
- Using exposure measured by personal passive samplers on a subgroup of 200 selected subjects;
- Using exposure to metals as measured in urine and hair to estimate short/medium term exposure;
APPENDIX 1: RESULTS

DLcoRapp: Single breath carbon monoxide diffusing capacity

### mir_25_000403

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>50.645</td>
<td>9.366</td>
<td>32.257</td>
</tr>
</tbody>
</table>

$R^2 = 0.076$

$F(7,743) = 10.428, p < 0.001$

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>a</td>
<td>-0.007</td>
<td>0.003</td>
<td>-0.014</td>
<td>-0.001</td>
<td>0.0244</td>
<td>c'</td>
<td>0.068</td>
</tr>
<tr>
<td>M(mir_25_000403)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>b</td>
<td>0.585</td>
<td>0.366</td>
<td>-0.133</td>
</tr>
<tr>
<td>costant</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>8.959</td>
<td>0.938</td>
<td>7.118</td>
<td>10.799</td>
<td>&lt;0.001</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>45.405</td>
</tr>
</tbody>
</table>

$R^2 = 0.083$

$F(7,743) = 9.322, p < 0.001$

#### Table 1: Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin. miRNAs expression was log<sub>2</sub> transformed.

Back transforming in model with miRNAs expression as independent variable and PM<sub>10</sub> as dependent variable it was obtained the %change in miRNAs expression: $(2^{\beta} - 1) \times 100$ for an increase 1 µg/m<sup>3</sup> in PM<sub>10</sub>.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: $\beta \log_{2}(101/100)$ associated with 1% change in miRNAs expression.

DLcoRapp = (Measured DLcoRapp / Theoretical DLcoRapp) \times 100

*Sobel test

### mir_27b_000409

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>50.645</td>
<td>9.366</td>
<td>32.257</td>
</tr>
</tbody>
</table>

$R^2 = 0.078$

$F(7,743) = 10.428, p < 0.001$

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>a</td>
<td>-0.005</td>
<td>0.004</td>
<td>-0.013</td>
<td>0.003</td>
<td>0.2064</td>
<td>c'</td>
<td>0.069</td>
</tr>
<tr>
<td>M(mir_27b_000409)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>b</td>
<td>1.053</td>
<td>0.297</td>
<td>0.470</td>
</tr>
<tr>
<td>costant</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>6.436</td>
<td>1.148</td>
<td>4.182</td>
<td>8.690</td>
<td>&lt;0.001</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>43.869</td>
</tr>
</tbody>
</table>

$R^2 = 0.083$

$F(7,743) = 9.602, p < 0.001$

#### Table 2: Simple Mediation Analysis Results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin. miRNAs expression was log<sub>2</sub> transformed.

Back transforming in model with miRNAs expression as independent variable and PM<sub>10</sub> as dependent variable it was obtained the %change in miRNAs expression: $(2^{\beta} - 1) \times 100$ for an increase 1 µg/m<sup>3</sup> in PM<sub>10</sub>.

Back transforming in model with DLcoRapp as dependent variable and miRNAs expression as independent variable it was obtained the %change in DLcoRapp: $\beta \log_{2}(101/100)$ associated with 1% change in miRNAs expression.

DLcoRapp = (Measured DLcoRapp / Theoretical DLcoRapp) \times 100

*Sobel test
### mir_30d_000420

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>a</td>
<td>0.004</td>
<td>-0.009</td>
<td>-0.016</td>
<td>-0.002</td>
<td>0.013</td>
<td>0.063</td>
<td>0.032</td>
</tr>
<tr>
<td>M(mir_30d_000420)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>50.645</td>
<td>9.366</td>
<td>32.57</td>
</tr>
<tr>
<td>costant</td>
<td>8.459</td>
<td>1.041</td>
<td>6.415</td>
<td>10.502</td>
<td>&lt;0.001</td>
<td>i&lt;sub&gt;2&lt;/sub&gt;</td>
<td>51.363</td>
<td>9.779</td>
</tr>
</tbody>
</table>

R<sup>2</sup> = 0.078
F(7,743) = 8.936, p<0.001

### mir_92a_000431

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>M(mir_92a_000431)</td>
<td>a</td>
<td>-0.007</td>
<td>0.003</td>
<td>-0.012</td>
<td>-0.001</td>
<td>0.0259</td>
<td>0.068</td>
<td>0.032</td>
</tr>
<tr>
<td>costant</td>
<td>12.122</td>
<td>0.853</td>
<td>10.448</td>
<td>13.796</td>
<td>&lt;0.001</td>
<td>i&lt;sub&gt;1&lt;/sub&gt;</td>
<td>41.983</td>
<td>10.547</td>
</tr>
</tbody>
</table>

R<sup>2</sup> = 0.078
F(7,743) = 8.936, p<0.001

### Table 3: Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated hemoglobin. miRNAs expression was log<sub>2</sub> transformed.

Back transforming in model with miRNAs expression as independent variable and PM<sub>10</sub> as dependent variable it was obtained the %change in miRNAs expression : \((2^{\beta} - 1)\times 100\) for an increase 1 µg/m<sup>3</sup> in PM<sub>10</sub>.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp : \(\beta \log_{2}(101/100)\) associated with 1%change in miRNAs expression.

DLcoRapp= (Measured DLcoRapp / Theoretical DLcoRapp)*100

*Sobel test.

### mir_92a_000431

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.0001</td>
</tr>
<tr>
<td>M(mir_92a_000431)</td>
<td>a</td>
<td>0.005</td>
<td>0.004</td>
<td>-0.016</td>
<td>-0.001</td>
<td>0.004</td>
<td>0.068</td>
<td>0.032</td>
</tr>
<tr>
<td>costant</td>
<td>0.005</td>
<td>0.004</td>
<td>0.000</td>
<td>0.016</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

R<sup>2</sup> = 0.082
F(6,744) = 1.484, p<0.001

### Table 4: Simple Mediation Analysis Results for mediator mir_92a_000431 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated hemoglobin. miRNAs expression was log<sub>2</sub> transformed.

Back transforming in model with miRNAs expression as independent variable and PM<sub>10</sub> as dependent variable it was obtained the %change in miRNAs expression : \((2^{\beta} - 1)\times 100\) for an increase 1 µg/m<sup>3</sup> in PM<sub>10</sub>.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp : \(\beta \log_{2}(101/100)\) associated with 1%change in miRNAs expression.

DLcoRapp= (Measured DLcoRapp / Theoretical DLcoRapp)*100

*Sobel test.
Table 5: Simple Mediation Analysis Results for mediator mir_181a_2_0002317 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated hemoglobin.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

\[DLcoRapp = \frac{\text{Measured DLcoRapp}}{\text{Theoretical DLcoRapp}} \times 100\]

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM 10)</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.001</td>
<td>0.127</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>i_3</td>
<td>50.645</td>
<td>9.366</td>
<td>32.257</td>
<td>69.033</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R^2 = 0.078

F(7,743) = 10.428, p < 0.001

| Total effect of PM 10 on DLcoRapp | c | 0.063 | 0.032 | -0.001  | 0.127 | 0.051 |
| Direct Effect of PM 10 on DLcoRapp | c' | 0.062 | 0.033 | -0.002  | 0.126 | 0.059 |
| Indirect Effect of PM 10 on DLcoRapp | 0.002 | 0.004 | - | - | 0.007 | 0.012 | 0.709* |

R^2 = 0.022

F(6,744) = 2.823, p = 0.010

Table 6: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated hemoglobin.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

\[DLcoRapp = \frac{\text{Measured DLcoRapp}}{\text{Theoretical DLcoRapp}} \times 100\]

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM 10)</td>
<td>c</td>
<td>0.063</td>
<td>0.032</td>
<td>-0.001</td>
<td>0.127</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>i_3</td>
<td>50.645</td>
<td>9.366</td>
<td>32.257</td>
<td>69.033</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R^2 = 0.078

F(7,743) = 10.428, p < 0.001

| Total effect of PM 10 on DLcoRapp | c | 0.063 | 0.032 | -0.001  | 0.127 | 0.051 |
| Direct Effect of PM 10 on DLcoRapp | c' | 0.062 | 0.033 | -0.002  | 0.126 | 0.058 |
| Indirect Effect of PM 10 on DLcoRapp | 0.001 | 0.005 | - | - | 0.009 | 0.012 | 0.790* |

R^2 = 0.040

F(6,744) = 4.736, p < 0.001

Table 6: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated hemoglobin.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

\[DLcoRapp = \frac{\text{Measured DLcoRapp}}{\text{Theoretical DLcoRapp}} \times 100\]

*Sobel test.
Table 7: Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated hemoglobin.

miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the % change in miRNAs expression: \((2^\beta - 1)\times 100\) for an increase 1 µg/m3 in PM10.

Back transforming in model with DLcoRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in DLcoRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression. DLcoRapp = \((\text{Measured DLcoRapp} / \text{Theoretical DLcoRapp})\times 100\).

*Sobel test.

Table 8: Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the % change in miRNAs expression: \((2^\beta - 1)\times 100\) for an increase 1 µg/m3 in PM10.

Back transforming in model with FEV1 Rapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV1 Rapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression. FEV1 Rapp = \((\text{Measured FEV1 Rapp} / \text{Theoretical FEV1 Rapp})\times 100\).

*Sobel test.
Table 9: Simple Mediation Analysis Results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $2^{\beta-1} \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.
Back transforming in model with FEV$_1$ as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV$_1$ Rapp: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.
FEV$_1$ Rapp = (Measured FEV$_1$/ Theoretical FEV$_1$) * 100
*Sobel test

Table 10: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $2^{\beta-1} \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.
Back transforming in model with FEV$_1$ as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV$_1$ Rapp: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.
FEV$_1$ Rapp = (Measured FEV$_1$/ Theoretical FEV$_1$) * 100
*Sobel test
Table 11: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM_{10} as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m\(^3\) in PM_{10}.

Back transforming in model with FEV\(_1\) as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV\(_1\) Rapp: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

\[
\text{FEV}_1\text{Rapp} = \left(\frac{\text{Measured FEV}_1}{\text{Theoretical FEV}_1}\right) \times 100
\]

* Sobel test

Table 12: Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM\(_{10}\) as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m\(^3\) in PM\(_{10}\).

Back transforming in model with FEV\(_1\) as independent variable and miRNAs expression as dependent variable it was obtained the change in FEV\(_1\) Rapp: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

\[
\text{FEV}_1\text{Rapp} = \left(\frac{\text{Measured FEV}_1}{\text{Theoretical FEV}_1}\right) \times 100
\]

* Sobel test
**FVCRapp:**

**mir_27b_000409**

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM_{10})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M(mir_{25_000403})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>i_1</td>
<td>9.740</td>
<td>0.926</td>
<td>7.924</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.126 \]

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM10 on FVCRapp</td>
<td>c</td>
<td>0.051</td>
<td>0.034</td>
</tr>
<tr>
<td>Direct Effect of PM10 on FVCRapp</td>
<td>c'</td>
<td>0.048</td>
<td>0.034</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Boot SE</th>
<th>Boot 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM10 on FVCRapp</td>
<td>0.003</td>
<td>0.005</td>
<td>-0.008</td>
</tr>
</tbody>
</table>

---

**mir_27b_000409**

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM_{10})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M(mir_{30d_000420})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>i_1</td>
<td>9.233</td>
<td>1.029</td>
<td>7.213</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.129 \]

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM10 on FVCRapp</td>
<td>c</td>
<td>0.051</td>
<td>0.034</td>
</tr>
<tr>
<td>Direct Effect of PM10 on FVCRapp</td>
<td>c'</td>
<td>0.042</td>
<td>0.034</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Boot SE</th>
<th>Boot 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM10 on FVCRapp</td>
<td>0.009</td>
<td>0.006</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

---

Table 13: Simple Mediation Analysis Results for mediator mir_{27b_000409} on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^{\beta-1})\times100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

FVCRapp = \((\text{Measured FVC} / \text{Theoretical FVC})\times100\)

*Sobel test

---

Table 14: Simple Mediation Analysis Results for mediator mir_{27b_000409} on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^{\beta-1})\times100\) for an increase 1 µg/m^3 in PM10.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

FVCRapp = \((\text{Measured FVC} / \text{Theoretical FVC})\times100\)

*Sobel test

---

132
Table 15: Simple Mediation Analysis Results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.
miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1)*100\) for an increase 1 µg/m^3 in PM10.
Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

\[
FVCRapp = \frac{\text{Measured FVC}}{\text{Theoretical FVC}} \times 100
\]

* Sobel test.

### Table 16: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.
miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1)*100\) for an increase 1 µg/m^3 in PM10.
Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

\[
FVCRapp = \frac{\text{Measured FVC}}{\text{Theoretical FVC}} \times 100
\]

* Sobel test.
Table 17: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

FVCRapp = (Measured FVC / Theoretical FVC) x 100

Sobel test.

Table 18: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

FVCRapp = (Measured FVC / Theoretical FVC) x 100

Sobel test.
### mir_652_002352

**Table 19:** Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.051</td>
<td>0.034</td>
<td>0.0155</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i$_3$</td>
<td>97.508</td>
<td>9.289</td>
<td>79.274</td>
</tr>
</tbody>
</table>

$R^2$ = 0.126

$F(7,780)$ = 15.999, $p$ < 0.001

### Heart Rate:

**Table 20:** Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>a</td>
<td>-0.021</td>
<td>0.005</td>
<td>-0.032</td>
<td>-0.011</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M(mir_652_002352)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c'</td>
<td>0.049</td>
<td>0.034</td>
<td>0.018</td>
</tr>
<tr>
<td>costant</td>
<td>i$_1$</td>
<td>8.632</td>
<td>1.480</td>
<td>5.726</td>
<td>11.538</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ = 0.037

$F(7,780)$ = 4.290, $p$ < 0.001

**Table 19:** Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with FVCRapp as independent variable and miRNAs expression as dependent variable it was obtained the change in FVCRapp: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*FVCapp = (Measured FVC / Theoretical FVC) * 100

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>a</td>
<td>-0.012</td>
<td>0.003</td>
<td>-0.019</td>
<td>-0.006</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M(mir_25_000403)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c'</td>
<td>0.035</td>
<td>0.022</td>
<td>0.007</td>
</tr>
<tr>
<td>costant</td>
<td>i$_2$</td>
<td>9.823</td>
<td>2.074</td>
<td>5.779</td>
<td>11.768</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ = 0.034

$F(8,779)$ = 7.920, $p$ < 0.001

### Heart Rate:

**Table 20:** Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>a</td>
<td>-0.035</td>
<td>0.022</td>
<td>-0.007</td>
<td>0.077</td>
<td>0.105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M(mir_25_000403)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c'</td>
<td>0.035</td>
<td>0.022</td>
<td>0.007</td>
</tr>
<tr>
<td>costant</td>
<td>i$_2$</td>
<td>51.705</td>
<td>4.038</td>
<td>43.780</td>
<td>59.630</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ = 0.034

$F(8,747)$ = 7.920, $p$ < 0.001

**Table 19:** Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

**Table 20:** Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test
Table 21: Simple Mediation Analysis Results for mediator mir_27_000409 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: \(((2^{\beta} - 1)\times 100)\) for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>X(Pm$_{10}$)</td>
<td>a</td>
<td>-0.016</td>
<td>0.004</td>
<td>-0.024</td>
</tr>
<tr>
<td>M(mir_27b_000409)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>$i_1$</td>
<td>8.550</td>
<td>0.616</td>
<td>7.340</td>
</tr>
<tr>
<td>$i_3$</td>
<td>51.197</td>
<td>3.339</td>
<td>44.645</td>
<td>57.750</td>
</tr>
</tbody>
</table>

$R^2$ = 0.061
$F(6,848) = 9.242, p<0.001$

Table 22: Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: \(((2^{\beta} - 1)\times 100)\) for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>X(Pm$_{10}$)</td>
<td>a</td>
<td>-0.017</td>
<td>0.004</td>
<td>-0.024</td>
</tr>
<tr>
<td>M(mir_30d_000420)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>$i_1$</td>
<td>9.084</td>
<td>0.562</td>
<td>7.981</td>
</tr>
<tr>
<td>$i_2$</td>
<td>52.051</td>
<td>3.820</td>
<td>44.553</td>
<td>59.549</td>
</tr>
</tbody>
</table>

$R^2$ = 0.062
$F(7,847) = 7.944, p<0.001$

<table>
<thead>
<tr>
<th>Effect</th>
<th>Boot SE</th>
<th>Boot 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of Pm$_{10}$ on Heart Rate</td>
<td>$c$</td>
<td>0.036</td>
<td>0.021</td>
</tr>
<tr>
<td>Direct Effect of Pm$_{10}$ on Heart Rate</td>
<td>$c'$</td>
<td>0.034</td>
<td>0.022</td>
</tr>
<tr>
<td>Indirect Effect of Pm$_{10}$ on Heart Rate</td>
<td>-</td>
<td>0.002</td>
<td>0.004</td>
</tr>
</tbody>
</table>

$R^2$ = 0.054
$F(6,848) = 7.989, p<0.001$

$R^2$ = 0.062
$F(7,847) = 7.934, p<0.001$
### mir_92a_000431

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm\textsubscript{10})</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.036</td>
<td>0.021</td>
<td>-0.0062</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i\textsubscript{3}</td>
<td>51.197</td>
<td>3.339</td>
<td>44.645</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.061 \]

\[ F(6,848) = 9.242, p < 0.001 \]

**Table 23:** Simple Mediation Analysis Results for mediator mir\textsubscript{92a}_000431 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM\textsubscript{10} as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m\(^3\) in PM\textsubscript{10}.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test

### mir\textsubscript{106a}_002169

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm\textsubscript{10})</td>
<td>a</td>
<td>-0.012</td>
<td>0.003</td>
<td>-0.018</td>
<td>-0.006</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M(mir\textsubscript{106a}_002169)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c'</td>
<td>0.035</td>
<td>0.022</td>
<td>-0.007</td>
</tr>
<tr>
<td>costant</td>
<td>i\textsubscript{1}</td>
<td>13.438</td>
<td>0.451</td>
<td>12.552</td>
<td>14.323</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ R^2 = 0.038 \]

\[ F(6,848) = 5.608, p = 0.0003 \]

**Table 24:** Simple Mediation Analysis Results for mediator mir\textsubscript{106a}_002169 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM\textsubscript{10} as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta - 1) \times 100\) for an increase 1 µg/m\(^3\) in PM\textsubscript{10}.

Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test
**mir_152_000475**

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(miR_152_000475)</th>
<th>Y(Heart Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>SE</td>
</tr>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.061$

$F(6,848) = 9.242, p < 0.001$

| M(miR_152_000475) | X(Pm$_{10}$) | a    | 0.013 | 0.003 | -0.019 | -0.006 | <0.001 | c'  | 0.033 | 0.022 | -0.009 | 0.075 | 0.124 |
| costant         | i$_1$    | 7.031| 0.512 | 6.027 | 8.035  | <0.001 | i$_2$  | 52.565 | 3.692 | 45.318 | 59.812 | <0.001 |

$R^2 = 0.036$

$F(6,848) = 5.239, p < 0.001$

**mir_181a_2_002317**

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(miR_181a_2_002317)</th>
<th>Y(Heart Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>SE</td>
</tr>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.061$

$F(6,848) = 9.242, p < 0.001$

| M(miR_181a_2_002317) | X(Pm$_{10}$) | a    | -0.017 | 0.004 | 1.206 | 3.606 | <0.001 | c'  | 0.034 | 0.022 | -0.009 | 0.076 | 0.117 |
| costant         | i$_1$    | 2.406| 0.611 | 1.206 | 3.606  | <0.001 | i$_2$  | 51.440 | 3.370 | 44.825 | 58.055 | <0.001 |

$R^2 = 0.028$

$F(6,848) = 3.586, p < 0.001$

$F(6,848) = 7.956 p < 0.001$

**Table 25:** Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)$*100 for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test

**Table 26:** Simple Mediation Analysis Results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)$*100 for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate: $\beta \log2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test
Table 27: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM\textsubscript{10} as dependent variable it was obtained the %change in miRNAs expression : \((2^\beta -1)*100\) for an increase 1 µg/m\(^3\) in PM\textsubscript{10}.
Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate : \(\beta \log_2(101/100)\) associated with 1%change in miRNAs expression.
*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm\textsubscript{10})</td>
<td>0.036</td>
<td>0.021</td>
<td>-0.006</td>
<td>0.077</td>
</tr>
<tr>
<td>costant</td>
<td>51.197</td>
<td>3.339</td>
<td>44.645</td>
<td>57.750</td>
</tr>
</tbody>
</table>

\(R^2 = 0.061\)
\(F(6,848) = 9.242, \ p<0.001\)

Table 28: Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% CI were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM\textsubscript{10} as dependent variable it was obtained the %change in miRNAs expression : \((2^\beta -1)*100\) for an increase 1 µg/m\(^3\) in PM\textsubscript{10}.
Back transforming in model with Heart Rate as independent variable and miRNAs expression as dependent variable it was obtained the change in Heart Rate : \(\beta \log_2(101/100)\) associated with 1%change in miRNAs expression.
*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm\textsubscript{10})</td>
<td>-0.018</td>
<td>0.004</td>
<td>-0.027</td>
<td>-0.009</td>
</tr>
<tr>
<td>M(mir_652_002352)</td>
<td>0.033</td>
<td>0.022</td>
<td>-0.009</td>
<td>0.075</td>
</tr>
<tr>
<td>costant</td>
<td>3.7337</td>
<td>0.686</td>
<td>2.387</td>
<td>5.080</td>
</tr>
</tbody>
</table>

\(R^2 = 0.0356\)
\(F(6,848) = 4.5185, \ p<0.001\)

\(R^2 = 0.0040\)
\(F(6,848) = 8.1345, \ p<0.001\)

\(R^2 = 0.00630\)
\(F(7,847) = 8.1345, \ p<0.001\)
### Table 29: Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM10. Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm10)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>M(mir_25_000403)</td>
<td>9.796</td>
<td>0.494</td>
<td>8.826, 10.765</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>costant</td>
<td>72.370</td>
<td>3.724</td>
<td>65.061, 79.680</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.059 \]

\[ F(6,851) = 8.960, p < 0.001 \]

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm10)</td>
<td>a</td>
<td>-0.012</td>
<td>0.003, -0.018</td>
</tr>
<tr>
<td>M(mir_25_000403)</td>
<td>b</td>
<td>0.034</td>
<td>0.214, -0.386</td>
</tr>
<tr>
<td>costant</td>
<td>c</td>
<td>72.699</td>
<td>3.078, 66.658</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.059 \]

\[ F(6,851) = 8.960, p < 0.001 \]

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm10)</td>
<td>a</td>
<td>-0.016</td>
<td>0.004, -0.023</td>
</tr>
<tr>
<td>M(mir_27b_000409)</td>
<td>b</td>
<td>-0.192</td>
<td>0.172, -0.530</td>
</tr>
<tr>
<td>costant</td>
<td>c</td>
<td>74.342</td>
<td>3.411, 67.647</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.059 \]

\[ F(6,851) = 8.960, p < 0.001 \]

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm10)</td>
<td>a</td>
<td>-0.036</td>
<td>0.020, -0.075</td>
</tr>
<tr>
<td>M(mir_27b_000409)</td>
<td>b</td>
<td>-0.192</td>
<td>0.172, -0.530</td>
</tr>
<tr>
<td>costant</td>
<td>c</td>
<td>74.342</td>
<td>3.411, 67.647</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.061 \]

\[ F(7,850) = 7.860, p < 0.001 \]

Table 30: Simple Mediation Analysis Results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM10. Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression.

*Sobel test
**Table 31:** Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta -1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_{2}(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(mir_30d_000420)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Y(DBP)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>c</td>
<td>-0.033</td>
<td>0.020</td>
<td>-0.071</td>
<td>0.006</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>i$_2$</td>
<td>72.699</td>
<td>3.078</td>
<td>66.658</td>
<td>78.741</td>
</tr>
<tr>
<td>X(Pm$_{10}$)</td>
<td>a</td>
<td>-0.017</td>
<td>0.004</td>
<td>-0.024</td>
<td>-0.010</td>
<td>c’</td>
<td>-0.031</td>
<td>0.020</td>
<td>-0.070</td>
<td>0.009</td>
</tr>
<tr>
<td>M(mir_30d_000420)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>i$_1$</td>
<td>9.072</td>
<td>0.560</td>
<td>7.974</td>
<td>10.170</td>
</tr>
<tr>
<td>costant</td>
<td>i$_2$</td>
<td>71.674</td>
<td>3.553</td>
<td>64.760</td>
<td>78.589</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2 = 0.059$

$F(6,851) = 8.960, p < 0.001$

**Table 32:** Simple Mediation Analysis Results for mediator mir_92a_000431 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta -1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_{2}(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(mir_92a_000431)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Y(DBP)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-0.012</td>
<td>0.003</td>
<td>-0.018</td>
<td>-0.006</td>
<td>c</td>
<td>-0.032</td>
<td>0.020</td>
<td>-0.071</td>
<td>0.008</td>
</tr>
<tr>
<td>M(mir_92a_000431)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>i$_1$</td>
<td>13.433</td>
<td>0.449</td>
<td>12.552</td>
<td>14.315</td>
</tr>
<tr>
<td>costant</td>
<td>i$_2$</td>
<td>71.486</td>
<td>4.410</td>
<td>62.829</td>
<td>80.143</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R^2 = 0.037$

$F(6,851) = 5.454, p < 0.001$

$R^2 = 0.060$

$F(7,850) = 7.725, p < 0.001$
Table 33: Simple Mediation Analysis Results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

Back transforming in model with miRNAs expression as independent variable and PM\textsubscript{10} as dependent variable it was obtained the %change in miRNAs expression : (2^{\beta - 1})*100 for an increase 1 µg/m\textsuperscript{3} in PM\textsubscript{10}.

Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: \beta log2(101/100) associated with 1%change in miRNAs expression.

*Sobel test

Table 34: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

Back transforming in model with miRNAs expression as independent variable and PM\textsubscript{10} as dependent variable it was obtained the %change in miRNAs expression : (2^{\beta - 1})*100 for an increase 1 µg/m\textsuperscript{3} in PM\textsubscript{10}.

Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: \beta log2(101/100) associated with 1%change in miRNAs expression.

*Sobel test
Table 35: Simple Mediation Analysis Results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.
Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.
*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>-0.033</td>
<td>0.020</td>
<td>-0.0711</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i$_3$</td>
<td>72.699</td>
<td>3.078</td>
<td>66.658</td>
</tr>
</tbody>
</table>

$R^2 = 0.059$
$F(6,851) = 8.960, p<0.001$

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>a</td>
<td>-0.017</td>
<td>0.004</td>
<td>-0.024</td>
<td>-0.009</td>
<td>c’</td>
<td>-0.035</td>
<td>0.020</td>
</tr>
<tr>
<td>M(mir_181a_2_002317)</td>
<td>2.360</td>
<td>0.609</td>
<td>1.166</td>
<td>3.555</td>
<td>&lt;0.001</td>
<td>b</td>
<td>-0.150</td>
<td>0.173</td>
</tr>
<tr>
<td>costant</td>
<td>i$_3$</td>
<td>73.053</td>
<td>3.106</td>
<td>66.958</td>
<td>79.149</td>
<td>i$_2$</td>
<td>73.053</td>
<td>3.106</td>
</tr>
</tbody>
</table>

$R^2 = 0.027$
$F(7,850) = 7.785, p<0.001$

Table 36: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.
Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.
*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>-0.033</td>
<td>0.020</td>
<td>-0.0711</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i$_3$</td>
<td>72.699</td>
<td>3.078</td>
<td>66.658</td>
</tr>
</tbody>
</table>

$R^2 = 0.059$
$F(6,851) = 8.960, p<0.001$

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>a</td>
<td>-0.016</td>
<td>0.004</td>
<td>-0.023</td>
<td>-0.009</td>
<td>c’</td>
<td>-0.029</td>
<td>0.020</td>
</tr>
<tr>
<td>M(mir_218_000521)</td>
<td>4.643</td>
<td>0.556</td>
<td>3.551</td>
<td>5.734</td>
<td>&lt;0.001</td>
<td>b</td>
<td>0.201</td>
<td>0.190</td>
</tr>
<tr>
<td>costant</td>
<td>i$_3$</td>
<td>71.768</td>
<td>3.201</td>
<td>65.484</td>
<td>78.051</td>
<td>i$_2$</td>
<td>71.768</td>
<td>3.201</td>
</tr>
</tbody>
</table>

$R^2 = 0.060$
$F(7,850) = 7.785, p<0.001$

Table 36: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.
Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.
*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>-0.033</td>
<td>0.019</td>
<td>-0.0711</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i$_3$</td>
<td>72.699</td>
<td>3.078</td>
<td>66.658</td>
</tr>
</tbody>
</table>

$R^2 = 0.032$
$F(6,851) = 4.718, p<0.001$

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(Pm$_{10}$)</td>
<td>a</td>
<td>-0.016</td>
<td>0.004</td>
<td>-0.023</td>
<td>-0.009</td>
<td>c’</td>
<td>-0.029</td>
<td>0.019</td>
</tr>
<tr>
<td>M(mir_218_000521)</td>
<td>4.643</td>
<td>0.556</td>
<td>3.551</td>
<td>5.734</td>
<td>&lt;0.001</td>
<td>b</td>
<td>0.201</td>
<td>0.190</td>
</tr>
<tr>
<td>costant</td>
<td>i$_3$</td>
<td>71.768</td>
<td>3.201</td>
<td>65.484</td>
<td>78.051</td>
<td>i$_2$</td>
<td>71.768</td>
<td>3.201</td>
</tr>
</tbody>
</table>

$R^2 = 0.061$
$F(7,850) = 7.841, p<0.001$

Table 36: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.
Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.
*Sobel test
Table 37: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^{β} - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $β \log2(101/100)$ associated with 1%change in miRNAs expression.

* Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$(PM$_{10}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M(mir_375_000564)</th>
<th>Y(DBP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coeff.</td>
<td>SE</td>
</tr>
<tr>
<td>$c$</td>
<td>-0.033</td>
</tr>
<tr>
<td>$i_3$</td>
<td>72.699</td>
</tr>
</tbody>
</table>

$R^2$ = 0.059

F(6,851) = 8.960, p < 0.001

Table 38: Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^{β} - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with DBP as independent variable and miRNAs expression as dependent variable it was obtained the change in DBP: $β \log2(101/100)$ associated with 1%change in miRNAs expression.

* Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$(PM$_{10}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M(mir_652_002352)</th>
<th>Y(DBP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coeff.</td>
<td>SE</td>
</tr>
<tr>
<td>$c$</td>
<td>-0.035</td>
</tr>
<tr>
<td>$i_2$</td>
<td>73.647</td>
</tr>
</tbody>
</table>

$R^2$ = 0.059

F(6,851) = 8.960, p < 0.001

$R^2$ = 0.060

F(7,850) = 7.789, p < 0.001
SBP: Systolic Blood Pressure

### Table 39: Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)*100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Y(SBP)</th>
<th>Total Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-0.012</td>
<td>0.003</td>
<td>-0.018</td>
<td>-0.006</td>
<td>&lt;0.001</td>
<td>c' -0.015</td>
<td>0.030</td>
<td>-0.075</td>
<td>0.045</td>
</tr>
<tr>
<td>costant</td>
<td>9.796</td>
<td>0.494</td>
<td>8.826</td>
<td>10.765</td>
<td>&lt;0.001</td>
<td>i</td>
<td>99.357</td>
<td>4.720</td>
<td>90.093</td>
</tr>
</tbody>
</table>

$R^2$ = 0.164

F(6,851) = 27.761, p<0.001

### Table 40: Simple Mediation Analysis Results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)*100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Y(SBP)</th>
<th>Total Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-0.031</td>
<td>0.004</td>
<td>-0.032</td>
<td>-0.008</td>
<td>&lt;0.001</td>
<td>c' -0.015</td>
<td>0.031</td>
<td>-0.074</td>
<td>0.045</td>
</tr>
<tr>
<td>costant</td>
<td>8.546</td>
<td>0.613</td>
<td>7.343</td>
<td>9.748</td>
<td>&lt;0.001</td>
<td>i</td>
<td>100.500</td>
<td>5.234</td>
<td>90.227</td>
</tr>
</tbody>
</table>

$R^2$ = 0.164

F(6,851) = 7.969, p<0.001

F(7,850) = 23.811, p<0.001

---

145
**Table 41:** Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. MiRNAs expression was log2 transformed.

Back transforming in model with MiRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in MiRNAs expression: \( (2^{\beta} - 1)*100 \) for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and MiRNAs expression as dependent variable it was obtained the change in SBP: \( \beta \log(101/100) \) associated with 1% change in MiRNAs expression.

*Sobel test

---

**Table 42:** Simple Mediation Analysis Results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. MiRNAs expression was log2 transformed.

Back transforming in model with MiRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in MiRNAs expression: \( (2^{\beta} - 1)*100 \) for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and MiRNAs expression as dependent variable it was obtained the change in SBP: \( \beta \log(101/100) \) associated with 1% change in MiRNAs expression.

*Sobel test

---

Table 41: Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. MiRNAs expression was log2 transformed.

Back transforming in model with MiRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in MiRNAs expression: \( (2^{\beta} - 1)*100 \) for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and MiRNAs expression as dependent variable it was obtained the change in SBP: \( \beta \log(101/100) \) associated with 1% change in MiRNAs expression.

*Sobel test

---

Table 42: Simple Mediation Analysis Results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. MiRNAs expression was log2 transformed.

Back transforming in model with MiRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in MiRNAs expression: \( (2^{\beta} - 1)*100 \) for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and MiRNAs expression as dependent variable it was obtained the change in SBP: \( \beta \log(101/100) \) associated with 1% change in MiRNAs expression.

*Sobel test
### Table 43: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the % change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(PM_{10})$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.164$

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM$_{10}$ on SBP</td>
<td>-0.012</td>
<td>0.030</td>
<td>0.072</td>
</tr>
<tr>
<td>Direct Effect of PM$_{10}$ on SBP</td>
<td>-0.013</td>
<td>0.030</td>
<td>0.072</td>
</tr>
<tr>
<td>Indirect Effect of PM$_{10}$ on SBP</td>
<td>0.001</td>
<td>0.003</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

Table 44: Simple Mediation Analysis Results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the % change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(PM_{10})$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.164$

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM$_{10}$ on SBP</td>
<td>-0.017</td>
<td>0.004</td>
<td>-0.024</td>
</tr>
<tr>
<td>Direct Effect of PM$_{10}$ on SBP</td>
<td>-0.019</td>
<td>0.031</td>
<td>-0.079</td>
</tr>
<tr>
<td>Indirect Effect of PM$_{10}$ on SBP</td>
<td>0.007</td>
<td>0.004</td>
<td>-0.170*</td>
</tr>
</tbody>
</table>

$R^2 = 0.166$

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM$_{10}$ on SBP</td>
<td>-0.012</td>
<td>0.030</td>
<td>-0.072</td>
</tr>
<tr>
<td>Direct Effect of PM$_{10}$ on SBP</td>
<td>-0.019</td>
<td>0.031</td>
<td>-0.079</td>
</tr>
<tr>
<td>Indirect Effect of PM$_{10}$ on SBP</td>
<td>0.007</td>
<td>0.004</td>
<td>-0.170*</td>
</tr>
</tbody>
</table>

Table 43: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the % change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X(PM_{10})$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.164$

<table>
<thead>
<tr>
<th>Effect</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total effect of PM$_{10}$ on SBP</td>
<td>-0.012</td>
<td>0.030</td>
<td>-0.072</td>
</tr>
<tr>
<td>Direct Effect of PM$_{10}$ on SBP</td>
<td>-0.019</td>
<td>0.031</td>
<td>-0.079</td>
</tr>
<tr>
<td>Indirect Effect of PM$_{10}$ on SBP</td>
<td>0.007</td>
<td>0.004</td>
<td>-0.170*</td>
</tr>
</tbody>
</table>

$R^2 = 0.166$
### Table 45: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale.

Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM 

-10 as dependent variable it was obtained the %change in miRNAs expression : \((2^{\beta -1})\times 100\) for an increase 1 µg/m^3 in PM 

10 .

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
</table>
| X(PM 

| 10 ) | - | - | - | - | c -0.012 0.030 -0.0716 0.047 0.682 |
| costant | - | - | - | - | i3 99.357 4.720 90.093 108.621 <0.001 |

\(R^2=0.164\)

\(F(6,851)=27.761, p<0.001\)

### Table 46: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale.

Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM 

-10 as dependent variable it was obtained the %change in miRNAs expression : \((2^{\beta -1})\times 100\) for an increase 1 µg/m^3 in PM 

10 .

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
</table>
| X(PM 

| 10 ) | - | - | - | - | c -0.016 0.004 -0.023 -0.009 <0.001 |
| M(miR_218_000521) | a | -0.016 0.004 0.023 0.009 >0.001 |
| costant | i3 4.643 0.556 3.551 5.734 <0.001 |

\(R^2=0.032\)

\(F(6,851)=4.718, p<0.001\)

### Table 47: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale.

Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM 

-10 as dependent variable it was obtained the %change in miRNAs expression : \((2^{\beta -1})\times 100\) for an increase 1 µg/m^3 in PM 

10 .

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: \(\beta \log2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
</table>
| X(PM 

| 10 ) | - | - | - | - | c -0.018 0.004 0.026 0.009 >0.001 |
| M(miR_375_000564) | a | -0.018 0.004 0.026 0.009 >0.001 |
| costant | i3 3.789 0.685 2.445 5.134 <0.001 |

\(R^2=0.036\)

\(F(6,851)=5.322, p<0.001\)

\(R^2=0.164\)

\(F(7,850)=23.773, p<0.001\)

\(F(6,851)=27.761, p<0.001\)

\(R^2=0.165\)

\(F(7,850)=23.922, p<0.001\)
**mir_652_002352**

Table 47: Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.164$

$F(6,851) = 27.761, p < 0.001$

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with SBP as independent variable and miRNAs expression as dependent variable it was obtained the change in SBP: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test.

**CRP:**

**mir_25_000403**

Table 48: Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$R^2 = 0.164$

$F(6,851) = 23.776, p < 0.001$

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100(1.01^\beta - 1)$ associated with 1% change in miRNAs expression.

*Sobel test.
**mir_27b_000409**

Table 49: Simple Mediation Analysis Results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta -1)*100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100(1.01^\beta -1)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff. SE 95% CI</th>
<th>p</th>
<th>Coeff. SE 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.006</td>
</tr>
<tr>
<td>i$_3$</td>
<td>-4.345</td>
<td>0.317</td>
<td>-4.967</td>
<td>-3.722</td>
</tr>
</tbody>
</table>

$R^2=0.156$

F(6,840)=25.951, p<0.001

---

**mir_30d_000420**

Table 50: Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta -1)*100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100(1.01^\beta -1)$ associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff. SE 95% CI</th>
<th>p</th>
<th>Coeff. SE 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.006</td>
</tr>
<tr>
<td>i$_3$</td>
<td>-4.345</td>
<td>0.317</td>
<td>-4.967</td>
<td>-3.722</td>
</tr>
</tbody>
</table>

$R^2=0.156$

F(6,840)=25.951, p<0.001
Table 51: Simple Mediation Analysis Results for mediator mir_92a_000431 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression and CRP was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM<sub>10</sub> as dependent variable it was obtained the %change in miRNAs expression : \( (2^{\beta-1}) \times 100 \) for an increase 1 µg/m<sup>3</sup> in PM<sub>10</sub>.
Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: \( 100(1.01^{\beta-1}) \) associated with 1% change in miRNAs expression.

*Sobel test

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM&lt;sub&gt;10&lt;/sub&gt;)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Consequent**

<table>
<thead>
<tr>
<th>M(mir_92a_000431)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.0021</td>
<td>0.006</td>
</tr>
<tr>
<td>i&lt;sub&gt;3&lt;/sub&gt;</td>
<td>-4.345</td>
<td>0.317</td>
<td>-4.967</td>
<td>-3.722</td>
</tr>
</tbody>
</table>

\( R^2 = 0.156 \)

\( F(6,840)=25.951, p<0.001 \)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Boot SE</th>
<th>Boot 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Effect of PM&lt;sub&gt;10&lt;/sub&gt; on CRP</td>
<td>0.002</td>
<td>0.002</td>
<td>-0.0021</td>
</tr>
<tr>
<td>Direct Effect of PM&lt;sub&gt;10&lt;/sub&gt; on CRP</td>
<td>0.0003</td>
<td>0.0002</td>
<td>-0.0002</td>
</tr>
<tr>
<td>Indirect Effect of PM&lt;sub&gt;10&lt;/sub&gt; on CRP</td>
<td>-0.013</td>
<td>-0.019</td>
<td>-0.016</td>
</tr>
</tbody>
</table>

\( R^2 = 0.036 \)

\( F(6,840)=5.194, p<0.001 \)

Table 52: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression and CRP was log2 transformed.
Back transforming in model with miRNAs expression as independent variable and PM<sub>10</sub> as dependent variable it was obtained the %change in miRNAs expression : \( (2^{\beta-1}) \times 100 \) for an increase 1 µg/m<sup>3</sup> in PM<sub>10</sub>.
Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: \( 100(1.01^{\beta-1}) \) associated with 1% change in miRNAs expression.

*Sobel test
Table 53: Simple Mediation Analysis Results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100(1.01^\beta - 1)$ associated with 1% change in miRNAs expression.

*Sobel test

Table 54: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1)\times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100(1.01^\beta - 1)$ associated with 1% change in miRNAs expression.

*Sobel test
Table 55: Simple Mediation Analysis Results for mediator mir_375_000564 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature, miRNAs expression and CRP was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with CRP as independent variable and miRNAs expression as dependent variable it was obtained the percentage change in CRP: $100 \times (1.01^\beta - 1)$ associated with 1% change in miRNAs expression.

*Sobel test

Table 56: Simple Mediation Analysis Results for mediator mir_25_000403 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression: $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: $\beta \log_2(101/100)$ associated with 1% change in miRNAs expression.

*Sobel test
Table 57: Simple Mediation Analysis Results for mediator mir_27b_000409 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression. 
*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>0.217</td>
<td>0.119</td>
<td>-0.0159</td>
<td>0.450</td>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i$_3$</td>
<td>149.649</td>
<td>18.752</td>
<td>112.841</td>
</tr>
</tbody>
</table>

R$^2$ = 0.121
F(6,818)=18.709, p<0.001

Table 58: Simple Mediation Analysis Results for mediator mir_30d_000420 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed. Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : $(2^\beta - 1) \times 100$ for an increase 1 µg/m$^3$ in PM$_{10}$. Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: $\beta \log_2(101/100)$ associated with 1%change in miRNAs expression. 
*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>0.204</td>
<td>0.120</td>
<td>-0.031</td>
<td>0.439</td>
<td>0.088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

R$^2$ = 0.121
F(7,817)=16.115, p<0.001

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>0.202</td>
<td>0.120</td>
<td>-0.034</td>
<td>0.438</td>
<td>0.094</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

R$^2$ = 0.121
F(7,817)=16.122, p<0.001
### mir_92a_000431

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Consequent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM(10))</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.217</td>
<td>0.119</td>
<td>-0.0159</td>
<td>0.450</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i(_j)</td>
<td>149.649</td>
<td>18.752</td>
<td>112.841</td>
<td>186.456</td>
</tr>
</tbody>
</table>

\(R^2 = 0.121\)

### mir_106a_002169

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Consequent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM(10))</td>
<td>-0.011</td>
<td>0.003</td>
<td>-0.017</td>
<td>-0.005</td>
<td>&lt;0.001</td>
<td>c'</td>
<td>0.213</td>
<td>0.120</td>
<td>-0.022</td>
</tr>
<tr>
<td>M(mir_92a_000431)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i(_j)</td>
<td>13.512</td>
<td>0.466</td>
<td>12.598</td>
<td>14.426</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i(_j)</td>
<td>154.228</td>
<td>26.727</td>
<td>101.766</td>
<td>206.690</td>
</tr>
</tbody>
</table>

\(R^2 = 0.037\)

\(F(6,818)=5.153, p<0.001\)

\(R^2 = 0.121\)

\(F(7,817)=16.026, p<0.001\)

\(R^2 = 0.121\)

\(F(6,818)=18.709, p<0.001\)

Table 59: Simple Mediation Analysis Results for mediator mir_92a_000431 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM\(10\) as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta -1)*100\) for an increase 1 µg/m\(^3\) in PM\(10\).

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: \(\beta \log_2(101/100)\) associated with 1%change in miRNAs expression.

*Sobel test

### mir_106a_002169

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Consequent</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM(10))</td>
<td>-0.011</td>
<td>0.003</td>
<td>-0.017</td>
<td>-0.007</td>
<td>&lt;0.001</td>
<td>c'</td>
<td>0.197</td>
<td>0.120</td>
<td>-0.039</td>
</tr>
<tr>
<td>M(mir_106a_002169)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i(_j)</td>
<td>13.146</td>
<td>0.403</td>
<td>12.356</td>
<td>13.937</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i(_j)</td>
<td>171.042</td>
<td>28.454</td>
<td>115.191</td>
<td>226.894</td>
</tr>
</tbody>
</table>

\(R^2 = 0.045\)

\(F(6,818)=6.392, p<0.001\)

\(R^2 = 0.122\)

\(F(7,817)=16.179, p<0.001\)

\(R^2 = 0.121\)

\(F(6,818)=18.709, p<0.001\)

Table 60: Simple Mediation Analysis Results for mediator mir_106a_002169 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM\(10\) as dependent variable it was obtained the %change in miRNAs expression: \((2^\beta -1)*100\) for an increase 1 µg/m\(^3\) in PM\(10\).

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: \(\beta \log_2(101/100)\) associated with 1%change in miRNAs expression.

*Sobel test
Table 61: Simple Mediation Analysis Results for mediator mir_152_000475 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : \(2^\beta - 1\) * 100 for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(mir_152_000475)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Y(Fibrinogen)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>c</td>
<td>0.217</td>
<td>0.119</td>
<td>-0.0159</td>
<td>0.450</td>
</tr>
<tr>
<td>costant</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i$_1$</td>
<td>149.649</td>
<td>18.752</td>
<td>112.841</td>
<td>186.456</td>
</tr>
</tbody>
</table>

\(R^2 = 0.121\)

\(F(6,818)=18.709, p<0.001\)

\(R^2 = 0.0370\)

\(F(6,818)=5.239, p<0.001\)

Table 62: Simple Mediation Analysis Results for mediator mir_181a_2_002317 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM$_{10}$ as dependent variable it was obtained the %change in miRNAs expression : \(2^\beta - 1\) * 100 for an increase 1 µg/m$^3$ in PM$_{10}$.

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: \(\beta \log_2(101/100)\) associated with 1% change in miRNAs expression.

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>M(mir_181a_2_002317)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
<th>Y(Fibrinogen)</th>
<th>Coeff.</th>
<th>SE</th>
<th>95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM$_{10}$)</td>
<td>a</td>
<td>-0.012</td>
<td>0.003</td>
<td>-0.019</td>
<td>-0.006</td>
<td>&lt;0.001</td>
<td>c'</td>
<td>0.194</td>
<td>0.120</td>
<td>-0.040</td>
</tr>
<tr>
<td>M(mir_181a_2_002317)</td>
<td>costant</td>
<td>i$_1$</td>
<td>7.121</td>
<td>0.528</td>
<td>6.085</td>
<td>8.157</td>
<td>&lt;0.001</td>
<td>i$_2$</td>
<td>163.145</td>
<td>20.717</td>
</tr>
</tbody>
</table>

\(R^2 = 0.123\)

\(F(7,817)=16.395, p<0.001\)

\(R^2 = 0.0263\)

\(F(6,818)=5.239, p<0.001\)

\(R^2 = 0.126\)

\(F(7,817)=16.746, p<0.001\)

\(R^2 = 0.121\)

\(F(6,818)=18.709, p<0.001\)

\(R^2 = 0.0263\)

\(F(6,818)=3.689, p<0.001\)

\(R^2 = 0.0263\)

\(F(7,817)=16.746, p<0.001\)
### mir_218_000521

Table 63: Simple Mediation Analysis Results for mediator mir_218_000521 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature. miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^{\beta} - 1) \times 100\) for an increase 1 µg/m³ in PM10.

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: \(\beta \log_2(101/100)\) associated with 1%change in miRNAs expression.

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff. SE 95% CI</th>
<th>p</th>
<th>Indirect Effect of PM10 on Fibrinogen Coeff. SE 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM10)</td>
<td>- - - - - -</td>
<td>-</td>
<td>0.0360 0.027 1.007 -0.007 0.130 -</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>4.637 0.578 3.502 5.772</td>
<td>&lt;0.001</td>
<td>i₁ 159.732 19.443 121.568 197.896</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### mir_652_002352

Table 64: Simple Mediation Analysis Results for mediator mir_652_002352 on log2 scale. Bootstrap standard error and bootstrap 95% were obtained for Indirect effect. Models adjusted for age, sex, bmi, smoking status, glycated haemoglobin, temperature.

miRNAs expression was log2 transformed.

Back transforming in model with miRNAs expression as independent variable and PM10 as dependent variable it was obtained the %change in miRNAs expression: \((2^{\beta} - 1) \times 100\) for an increase 1 µg/m³ in PM10.

Back transforming in model with Fibrinogen as independent variable and miRNAs expression as dependent variable it was obtained the change in Fibrinogen: \(\beta \log_2(101/100)\) associated with 1%change in miRNAs expression.

*Sobel test.

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Coeff. SE 95% CI</th>
<th>p</th>
<th>Indirect Effect of PM10 on Fibrinogen Coeff. SE 95% CI</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>X(PM10)</td>
<td>- - - - - -</td>
<td>-</td>
<td>0.0180 0.024 -0.020 -0.075 -</td>
<td>-</td>
</tr>
<tr>
<td>costant</td>
<td>8.055 0.818 6.449 9.661</td>
<td>&lt;0.001</td>
<td>i₁ 156.314 19.831 117.389 195.240</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

| Total effect of PM10 on Fibrinogen | 0.217 0.119 -0.016 0.450 | 0.068 |
| Direct Effect of PM10 on Fibrinogen | 0.199 0.120 -0.037 0.434 | 0.098 |
| Indirect Effect of PM10 on Fibrinogen | 0.018 0.024 -0.020 0.075 | -    |

| Total effect of PM10 on Fibrinogen R² | 0.121 |
| Direct Effect of PM10 on Fibrinogen R² | 0.040 |
| Indirect Effect of PM10 on Fibrinogen R² | 0.040 |

**Effect SE Boot SE Boot 95% CI p**

**Effect SE Boot 95% CI p**
8. APPENDIX 2: SAS MACRO PROCESS FOR MEDIATION ANALYSIS

* PROCESS for SAS v2.13 */.
* Copyright 2012-2014 */.
* by Andrew F. Hayes */.
* www.afhayes.com */.

* Documentation available in Appendix A of http://www.guilford.com/p/hayes3;

* Permission is hereby granted, free of charge, to any person obtaining a
  copy of this software;
* and associated documentation files (the "Software"), to use the software in
  this form. Distribution;
* after modification is prohibited, as is its use for any commercial purpose
  without authorization;
* This software should not be posted or stored on any webpage, server, or
directory accessible to;
* the public whether free or for a charge unless written permission has been
  granted by the copyright;
* holder. The copyright holder requests that this software be distributed by
  directing users to;
* proccessmacro.org where the latest release of the software and related
documentation is archived and;
* can be downloaded;

* THIS SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, ;
* EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF ;
* MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ;
* IN NO EVENT SHALL THE COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, ;
* DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT ;
* OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE ;
* SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE ;

* The above text should be included in any distribution of the software;

%macro bcboot (databcbt=,estmte=9999);
  temp=&databcbt;
  temp[rank(temp)]=&databcbt;
  badlo=0;
  badhi=0;
  if (&estmte ^= 9999) then;
    do;pv=(temp < &estmte);pv=pv[+] / boot;
    ppv=ppv;
    if (pv > .5) then;do;ppv=1-pv;end;
    y5=sqrt(-2*log(ppv));
    xp=y5+((((y5*p4+p3)*y5+p2)*y5+p1)*y5+p0)/((((y5*q4+q3)*y5+q2)*y5+q1)*y5+q0);
    if (pv <= .5) then;do;xp=-xp;end;
    cilow=round(boot*(probnorm(2*xp+xp2)));
    cihigh=int(boot*(probnorm(2*xp+(-xp2))))+1;
    if (cilow < 1) then;do;cilow=1;booterr=1;badlo=1;end;
    if (cihigh > boot) then;do;cihigh=boot;booterr=1;badhi=1;end;
    llcit=temp[cilow,1];
    ulcit=temp[cihigh,1];
    end;
  if (&estmte=9999) then;do;llcit=temp[cilow,1];ulcit=temp[cihigh,1];end;
%mend;

%macro bcci (data=,var=,point=9999,conf=95);
proc iml;
use &data;
read all var{&var} into perdata;
x = (perdata = .); xx = xx[,] + 1;
j = 1; do i = 1 to nrow(perdata); if xx[i,1] = 0 then; perdata[j,] = perdata[i,]; j = j + 1; end; end;
perdata = perdata[1:j - 1, ];
berror = 0;
p0 = -.322232431088;
p1 = -1;
p2 = -.342242088547;
p3 = -.0204231210245;
p4 = -.0000453642210148;
q0 = .0993484626060;
q1 = .588581570495;
q2 = .531103462366;
q3 = .103537752850;
q4 = .0038560700634;
conf = &conf;
if ((floor(conf) >= 100) | (floor(conf) <= 50)) then; do; conf = 95; end;
alpha2 = (1 - (conf/100)) / 2;
y5 = sqrt(-2 * log(alpha2));
xp2 = 
(y5 + (((y5*p4 + p3)*y5 + p2)*y5 + p1)*y5 + p0) / (((y5*q4 + q3)*y5 + q2)*y5 + q1)*y5 + q0));
boot = nrow(perdata);
temp = perdata;
temp[rank(temp)] = perdata;
badlo = 0; badhi = 0;
if (&point ^= 9999) then;
do; pv = (temp < &point); pv = pv[+] / boot;
if ((pv = 0) | (pv = 1)) then; do; berror = 2; end;
if (berror = 0) then; do;
ppv = pv;
if (pv > 0.5) then; do; ppv = 1 - pv; end;
y5 = sqrt(-2 * log(ppv));
xp = y5 + (((y5*p4 + p3)*y5 + p2)*y5 + p1)*y5 + p0) / (((y5*q4 + q3)*y5 + q2)*y5 + q1)*y5 + q0));
if (pv <= .5) then; do; xp = -xp; end;
cilow = round(boot * probnorm(2*xp + xp2));
cihih = int(boot * probnorm(2*xp + (-xp2)) + 1;
if (cilow < 1) then; do; berror = 1; end;
if (cihih > boot) then; do; berror = 1; end;
if (berror = 0) then; do;
llcit = temp[cilow,1];
ulcit = temp[cihih,1];
nametmp = "Conf%" // "LLCI" // "ULCI";
outp = conf // llcit // ulcit;
print outp [label = "Bias corrected confidence interval:" rowname = nametmp format = 10.4];
end;
end;
end;
end;
end;
end;
if (&point = 9999) then; do;
cilowp = round(boot * probnorm(xp2));
cihihp = int(boot * probnorm(-xp2)) + 1;
if (cilowp < 1) then; do; berror = 1; end;
if (cihihp > boot) then; do; berror = 1; end;
if (berror = 0) then; do;
llcitp = temp[cilowp,1];
ulcitp = temp[cihihp,1];
nametmp = "Conf%" // "LLCI" // "ULCI";
outp = conf // llcitp // ulcitp;
print outp [label = "Percentile confidence interval:" rowname = nametmp format = 10.4];
end;
if (berror=1) then;
do;
print "Error: Decrease your confidence or increase the number of bootstrap estimates.";
end;
if (berror=2) then;
do;
print "Error: Impossible point estimate provided";
end;
quit;
%mend

%macro process (data=, vars=, model=77, y=, m=, x=, w=, z=, v=, q=, conf=95, hc3=0, cluster=, wmodval=999, zmodval=999, vmodval=999, qmodval=999, mmodval=999, xmodval=999, boot=1000, center=0, quantile=0, effsize=0, normal=0, varorder=2, total=0, plot=0, detail=1, iterate=10000, converge=0.00000001, percent=0, jn=0, coeffci=1, covmy=0, contrast=0, seed=0, save=xxx, mc=0, decimals=10.4, covcoeff=0, olsdichy=0, olsdichm=0, ws=0);
options pagesize=32767;
proc iml;
use &data;
read all var{"vars} into dat;
vnames={&vars};
yname={&y};xname={&x};mnames={&m};
if ("&w" = "") then wname="xxx";else wname = "&w";
if ("&z" = "") then zname="xxx";else zname = "&z";
if ("&v" = "") then vname="xxx";else vname = "&v";
if ("&q" = "") then qname="xxx";else qname = "&q";
if ("&save" = "") then saveboot="xxx";else saveboot="&save";
if ("&cluster" = "") then clname="xxx";else clname = "&cluster";

wname=upcase(wname);zname=upcase(zname);vname=upcase(vname);qname=upcase(qname);
clname=upcase(clname);
ninit=nrow(dat);
xx=(dat = .);xx=xx[,+];
j=1;do i = 1 to nrow(dat);if xx[i,1]=0 then;do;dat[j,]=dat[i,];j=j+1;end;end;
dat=dat[1:j-1,];
n=nrow(dat);
p0=-.32232431088;
p1 = -1;
p2 = -.342242088547;
p3 = -.0204231210245;
p4 = -.0000453642210148;
p0 = .0993484626060;
p1 = .588581570495;
p2 = .531103462366;
p3 = .10357752850;
p4 = .0038560700634;
badend=0;priorlo=-9999999;priorhi=9999999;
criterr=0;cluster=0;clsdmy=0;jndich=0;booterr=0;
wvdich=0;mod74dic=0;
effsize=(&effsize=1);
covcoeff=(&covcoeff=1);
notes=j(10,1,0);notes=1;iterr=0;clsmtch=0;
quantile=(&quantile=1);jn=(&jn=1);contrast=(&contrast=1);
center=(&center=1);detail=(&detail=1);coeffci=(&coeffci=1);
conf=(&conf);bconoff=(&percent ^= 1);
covmy=floor(&covmy);if (covmy < 0 | covmy > 2) then;do;covmy=0;end;
if ((floor(conf) >= 100) | (floor(conf) <= 50)) then;do;
conf=95;note[notes,1]=1;notes=notes+1;
if (n < ninit) then; do; nmiss = ninit - n; note[notes, 1] = 11; notes = notes + 1; end;
errs = 0; quantd = j(1, 6, 0); quantc = j(1, 6, 0); mcheck = 0; tttt = 0;
plot = (&plot ^= 0);
runerrs = j(50, 1, 0);
model = floor(&model);
normal = (&normal);
ws = (&ws);
olsdichm = (&olsdichm = 1); olsdichy = (&olsdichy = 1);
if (ws = 1) then; do;
  if (effsize = 1) then; do;
    note[notes, 1] = 19; notes = notes + 1; effsize = 0;
  end;
  if (normal = 1) then; do;
    note[notes, 1] = 16; notes = notes + 1; normal = 0;
  end;
end;
if (model ^= 4) then; do; ws = 0; end;
if (((jn = 1) & (model ^= 1) & (model ^= 3)) then; do;
  note[notes, 1] = 7; notes = notes + 1;
end;
if ((model > 76) | (model < 1)) then; do;
  model = 77; criterr = 1; errs = errs + 1; runerrs[errs, 1] = 19;
end;
toteff = 0; toteff = ((&total = 1) * (model = 4) | (model = 6));
varorder = (&varorder);
hc3 = (&hc3 ^= 0);
centvar = 'xxx';
modeln =
  [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1,
  0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3,
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 4,
  0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 5,
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6,
  1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 7,
  1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 8,
  1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 9,
  1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 10,
  1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 11,
  1 1 1 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 12,
  1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 13,
  0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 14,
  0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 15,
  0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 16,
  0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 17,
  0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 18,
  0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 19,
  0 0 0 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 20,
  1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 21,
  1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 22,
  1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 23,
  1 1 0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 24,
  1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 25,
  1 1 1 1 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 26,
  1 1 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 27,
  1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 28,
  1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 29,
  1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 30,
  1 1 0 1 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 31,
  1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 32,
  1 1 1 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 33,
  1 1 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 34,
if (ncol(yname) ^= 1) then; do; errs=errs+1; runerrs[errs,1]=21; criterr=1; end;
if ((&mmodval=0) & (&wmodval = 0) & (model=3 | model=2) & (contrast=1))
then;
do;
   errs=errs+1; runerrs[errs,1]=31; criterr=1;
end;
xlist = ((wm=1) | (zm=1) | (wzm=1) | (wy=1) | (zy=1) | (vzy=1) | (vxy=1) | (vqxy=1) | (wvxy=1)) | (wxmy=1));
mlist = ((vy=1) | (qy=1) | (vqy=1) | (vy=1) | (vqy=1) | (wvy=1) | (vvy=1) | (wvy=1) | (wzm=1) | (wxmy=1));
bad=0;
if (criterr=0) then; do;
werr=0; verr=0; qerr=0; zerr=0; yerr=1; xerr=1;
   wlist=((wm=1) | (wzm=1) | (wy=1) | (vzy=1) | (wvy=1) | (vvy=1) | (wvmy=1) | (wvy=1) | (wzm=1) | (wxmy=1));
   if ((wlist=1) & (wname = "XXX"))
      then; do; werr=1; wlist=0; errs=errs+1; runerrs[errs,1]=4;
      end;
   if ((wlist=1) & ((wname=qname) | (wname=vname) | (wname=zname) | (wname=xname) | (wname=yname))))
      then; do;
         werr=4; errs=errs+1; runerrs[errs,1]=12;
      end;
   if ((wlist=1) & (zname = "XXX"))
      then; do; zerr=1; zlist=0; errs=errs+1; runerrs[errs,1]=5;
      end;
   if ((zlist=1) & (zname=qname) | (zname=vname) | (zname=wname) | (zname=xname) | (zname=yname)))
      then; do;
         zerr=4; errs=errs+1; runerrs[errs,1]=13;
      end;
   if ((zlist=1) & (qname = "XXX"))
      then; do; qerr=1; qlist=0; errs=errs+1; runerrs[errs,1]=6;
      end;
   if ((qlist=1) & (qname=zname) | (qname=vname) | (qname=wname) | (qname=xname) | (qname=yname)))
      then; do;
         qerr=4; errs=errs+1; runerrs[errs,1]=14;
      end;
   if ((qlist=1) & (vname = "XXX"))
      then; do; vlist=((vy=1) | (vqy=1) | (vqy=1) | (vqxy=1)) | (wvy=1) | (vvy=1) | (wvy=1) | (vvy=1));
      end;
   if ((vlist=0) & (vname = "XXX"))
      then; do; verr=2; vlist=0; runerrs[errs,1]=8;
      end;
   if ((vlist=0) & (vname = "XXX"))
      then; do; vlist=0; runerrs[errs,1]=9;
      end;
   if ((vlist=0) & (vname = "XXX"))
      then; do; vlist=0; runerrs[errs,1]=10;
      end;
   if ((vlist=0) & (vname = "XXX"))
      then; do; vlist=0; runerrs[errs,1]=11;
      end;
   if (hc3=1) then; do; note[notes,1]=3; notes=notes+1;
end;
end;
alpha2=(1-(conf/100))/2;
y5=sqrt(-2*log(alpha2));
xp2=(
(y5*(y5+p4+p3)*y5+p2)*y5+p1)*y5+p0)/((y5*(y5+q4+q3)*y5+q2)*y5+q1)*y5+q0));
cons=j(n,1,1);
temp=(n*t(dat)*dat)-(t(dat[+,])*dat[+]);
temp=temp/(n*(n-1));temp=sqrt(vecdiag(temp));temp=(temp = 0);temp=temp[+];
temp2=1;
do i = 1 to ncol(vnames);
  if (vnames[i]=xname) then;
doi=dat[i];
end;
if (ws=1) then;
temp2=(n*t(tmp2)*tmp2)-(t(tmp2[+])*tmp2[+]);temp2=temp2/(n*(n-1));
end;
if ((temp > 0) & ws=0) | ((temp > 0) & (ws=1) & (temp2 ^=0)) then;
criterr=1;errs=errs+1;runerrs[errs,1]=27;
end;
nmeds=ncol(mnames);
mcmats=I(nmeds*2);
mccoef=j((nmeds*2),1,0);
sobel=j(nmeds,4,-999);
if ((model = 6) & (nmeds > 4)) then;
dois=errs+1;runerrs[errs,1]=2;
end;
if ((model < 4) & (nmeds > 1)) then;
dois=errs+1;runerrs[errs,1]=3;
end;
nmods=(model=74);
bad=0;intcnt=1;modvals=0;modvalsd=0;
yintemp=INT_1,INT_2,INT_3,INT_4,INT_5,INT_6,INT_7,INT_8,INT_9,INT_10,INT_11,INT_12;
yintemp=yintemp||INT_13,INT_14,INT_15,INT_16,INT_17,INT_18,INT_19,INT_20,INT_21,INT_22;
yintemp=yintemp||INT_33,INT_34,INT_35,INT_36,INT_37,INT_38,INT_39,INT_40,INT_41,INT_42;
yintemp=yintemp||INT_43,INT_44,INT_45,INT_46,INT_47,INT_48,INT_49,INT_50,INT_51,INT_52;
yintemp=yintemp||INT_53,INT_54,INT_55,INT_56,INT_57,INT_58,INT_59,INT_60,INT_61,INT_62;
cntname=(C1)/(C2)/(C3)/(C4)/(C5)/(C6)/(C7)/(C8)/(C9)/(C10);
cntname=cntname/(C11)/(C12)/(C13)/(C14)/(C15)/(C16)/(C17)/(C18)/(C19)/(C20);
cntname=cntname/(C21)/(C22)/(C23)/(C24)/(C25)/(C26)/(C27)/(C28)/(C29)/(C30);
cntname=cntname/(C31)/(C32)/(C33)/(C34)/(C35)/(C36)/(C37)/(C38)/(C39)/(C40);
cntname=cntname/(C41)/(C42)/(C43)/(C44)/(C45)/(C46)/(C47)/(C48)/(C49)/(C50);
cntname=cntname/(C51)/(C52)/(C53)/(C54)/(C55)/(C56)/(C57)/(C58)/(C59)/(C60);
cntname=cntname/(C61)/(C62)/(C63)/(C64)/(C65)/(C66)/(C67)/(C68)/(C69)/(C70);
cntname="cntname""/(C71)""/(C72)""/(C73)""/(C74)""/(C75)""/(C76)""/(C77)""/(C78)""/(C79)""/(C80)"; 

cntname="cntname""/(C81)""/(C82)""/(C83)""/(C84)""/(C85)""/(C86)""/(C87)""/(C88)""/(C89)""/(C90)"; 

cntname="cntname""/(C91)""/(C92)""/(C93)""/(C94)""/(C95)""/(C96)""/(C97)""/(C98)""/(C99)""/(C100)"; 

cntname="cntname""/(C101)""/(C102)""/(C103)""/(C104)""/(C105)"; 

apathnam="a path""/(a1 path)""/(a2 path)""/(a3 path)""/(a4 path)""/(a5 path)""/(a6 path)""/(a7 path)""/(a8 path)""/(a9 path)""/(a10 path)"; 

apathnam="b path""/(b1 path)""/(b2 path)""/(b3 path)""/(b4 path)""/(b5 path)""/(b6 path)""/(b7 path)""/(b8 path)""/(b9 path)""/(b10 path)"; 

modvnm = {"xxxxxxxxxxxxxxxx"}||{"xxx"}||{"xxx"}||{"xxx"}||{"xxx"}; 

modvnm2 = modvnm; 

mlab="M1    =""M2    =""M3    =""M4    =""M5    =""M6    =""M7    =""M8    =""M9    =""M10   ="; 

m=j(n,nmeds,1); ymat=j(16,nmeds,0); deco=j(10,1,0); modmat=j(5,999); 

modmatv=j(1,5,1); modmatp=j(1,5,0); modprod=modmatv; 

iterate=abs(floor(&iterate)); converge=abs(&converge); 

boot=abs(floor(&boot)); adjust=0; 

if ((ws=1) & (&mc > 0)) then; do; note[notes,1]=17; notes=notes+1; end; 

mc=(abs(floor(&mc))*((1-ws))); 

if ((mc > 0) & (model > 5)) then; do; 

if (boot = 0) then; do; boot=mc; end; 

mc=0; 

note[notes,1]=12; notes=notes+1; 
end; 

if (boot ^= 0) then; do; 

cilow=floor(boot*\(1-\(100/100\))\(1/2\)); 

cihigh=floor((boot*\(100\) + (boot*\((1-(conf/100))/2\)) +1; 

do until ((cilow > 0) & (cihigh <= boot)); 

cilow=floor(boot*\(1-(conf/100))\(1/2\)); 

cihigh=floor((boot*\(100\) + (boot*\((1-conf/100))/2\)) +1; 

if ((cilow < 1) | (cihigh > boot)) then; do; 

boot=floor((boot*1000)/1000)*1000; adjust=1; 
end; 

if (adjust = 1) then; do; 

note[notes,1]=6; notes=notes+1; end; 

if ((mc > 0) & (model > 3) & (model < 6)) then; do; boot=0; bconoff=0; end; 

if ((boot > 0) & (mc > 0)) then; do; mc=0; end; 

savboot=0; 

if ((savboot ^= "xxx") & (boot > 0) & (model > 3)) then; do; savboot=1; end; 

if ((savboot=1) & (ws=1)) then; do; 

note[notes,1]=20; notes=notes+1; savboot=0; 
end; 

if ((boot = 0) | (mc = 0)) then; do; 

bootsz=boot; 

if (mc > 0) then; do; bootsz=mc; end; 

cilow=round(bootsz*\(1-\(100/100\))\(1/2\)); 

cihigh=int((bootsz*\(100\) + (bootsz*\((1-(conf/100))/2\)))+1; 

do until ((cilow > 0) & (cihigh <= bootsz)); 

cilow=floor(bootsz*\(1-(conf/100))\(1/2\)); 

cihigh=floor((bootsz\(100\) + (bootsz*\((1-conf/100))/2\)) +1; 

if ((cilow < 1) | (cihigh > bootsz)) then; do; 

bootsz=floor((bootsz*1000)/1000)*1000; adjust=1; 
end; 

end; 

end;
boot=bootsz;
if (mc > 0) then;do;mc=bootsz;end;
if ((boot > 0) & (mc > 0) & (model > 3) & (model < 6))
then;do;boot=0;end;
if ((adjust = 1) & (boot > 0)) then;do;note[notes,1]=6;notes=notes+1;end;
if ((adjust = 1) & (mc > 0)) then;do;note[notes,1]=13;notes=notes+1;end;
end;
if ((model = 6) & (nmeds > 1)) then;do;
indboot=j((boot+1),999);
if (nmeds = 2) then;do;indboot=j((boot+1),3,999);
end;
if (nmeds = 3) then;do;indboot=j((boot+1),7,999);
end;
if (nmeds = 4) then;do;indboot=j((boot+1),15,999);
end;
end;
if ((model = 6) & (nmeds > 1)) then;do;
mmaths=j((nmeds+2),(nmeds+2),0);
if (nmeds = 2) then;do;indboot=j((boot+1),3,999);
end;
if (nmeds = 3) then;do;indboot=j((boot+1),7,999);
end;
if (nmeds = 4) then;do;indboot=j((boot+1),15,999);
end;
end;
if (model < 4) then;do;
boot=0;cmat=j(10,1,0);zmat=j(10,1,0);
end;
nvarch=j(1,ncol(dat),0);xmatch=0;zmatch=0;vmatch=0;gmatch=0;
minprobe=0;maxprobe=0;
do i = 1 to ncol(vnames); *[b];
if (vnames[i]=yname) then;do;
y=dat[i];nvarch[1,i]=1;yerr=0;
if ((yname=xname)|(yname=wname)|(yname=zname)|(yname=vname)|(yname=qname))
then;do;
errs=errs+1;runerrs[errs,1]=17;
end;
end;
if (vnames[i]=xname) then;do;
x=dat[i];nvarch[1,i]=1;xdich=1;xerr=0;
do jj = 1 to n;
if ((x[jj,1]^=max(x)) & (x[jj,1]^=min(x))) then;do;xdich=0;goto leave;
end;
leave:
end;
end;
xmean=x[+,]/n;
if ((center = 1) & ((model < 4) | (xlist > 0))) then;do;
meancen=1;meanvec=j(n,1,xmean);x=x-meanvec;centvar=centvar||xname;
end;
xmean=x[+,]/n;
tmp=x-(cons*xmean);
xsd=sqrt((1/(n-1))*(t(tmp)*tmp));
if (xdich=0) then;do;
quantc[1,6]=1;matx=(xmean-xsd)//xmean//((xmean+xsd);
if ((xmodval = 999) & (quantile = 0)) then;do;
if ((matx[1,1]<min(x)) & (model=74))
then;do;matx[1,1]=min(x);minprobe=1;end;
if ((matx[3,1]>max(x)) & (model=74))
then;do;matx[3,1]=max(x);maxprobe=1;end;
end;
if (quantile = 1) then;do;
quantd[1,6]=1;quantc[1,6]=0;tmp=x;tmp[rank(tmp)]=x;
matx=tmp[floor(n*0.10),1]//tmp[floor(n*0.25),1]//tmp[floor(n*0.5),1]//tmp[floor(n*0.75),1]//tmp[floor(n*0.9),1];
end;
end;
if (xdich=1) then;do;matx=min(x)//max(x);
   if (model=74) then;do;matx=min(x);mod74dic=1;end;
end;
if (&xmodval ^= 999) then;do;matx=&xmodval;quantd[1,6]=0;quantc[1,6]=0;end;
end;
if ((werr=0) & (wlist=1)) then;do;
   if (vnames[,i]=wname) then;do;
      werr=0;wmatch=1;w=dat[,i];
      if (center = 1) then;do;
         wmean=w[+]/n;meanvec=j(n,1,wmean);w=w-meanvec;
      end;
   end;
   leave2:
      if (model = 3) then;do;jndich=wdich;jnmin=min(w);jnmax=max(w);
      end;
      if (wdich=0) then;
         matw=(wmean-wsd)//wmean//(wmean+wsd);
         if ((&wmodval = 999) & (quantile=0)) then;do;
            if (matw[1,1] < min(w)) then;do;minprobe=1;end;
            if (matw[3,1] > max(w)) then;do;maxprobe=1;end;
         end;
      end;
      end;
   end;
leave2:
   if (wdich=1) then;do;
      matw=min(w)//max(w);wvdich=1;cmaxw=max(w);cminw=min(w);
   end;
   if (&wmodval = 999) then;do;
      matw=&wmodval;quantd[1,1]=0;quantc[1,1]=0;
   end;
   modmatv[1,1]=nrow(matw);
   modmat[1:nrow(matw),1]=matw;
   modvnm[1,1]=wname;
   modmatp[1,1]=1;
end;
end;
if ((zerr=0) & (zlist=1)) then;do;
   if (vnames[,i]=zname) then;do;
      zerr=0;zmatch=1;z=dat[,i];
      if (center = 1) then;do;
         zmean=z[+]/n;meanvec=j(n,1,zmean);z=z-meanvec;
      end;
end;
nvarch[1,i]=1; nmods=nmods+1; zmean=z[+,]/n; tmp=z-(cons*zmean);
zsd=sqrt((1/(n-1))*t(tmp)*tmp));
zdic=1;
do jj=1 to n;
if ((z[jj,1] ^= max(z)) & (z[jj,1] ^= min(z))) then; do; zdich=0; goto leave3;
end;
end;
leave3:
if (zdich = 0) then; do;
matz=(zmean-zsd)//zmean//(zmean+zsd);
if ((&zmodval=999) & (quantile=0)) then; do;
    if (matz[1,1] < min(z)) then; do; matz[1,1]=min(z); minprobe=1; end;
    if (matz[3,1] > max(z)) then; do; matz[3,1]=max(z); maxprobe=1; end;
end;
if (quantile = 1) then; do;
    quantd[1,2]=1; quantc[1,2]=0;
    tmp=z; tmp[rank(tmp)]=z;
    matz=tmp[floor(n*0.10),1]//tmp[floor(n*0.25),1]//tmp[floor(n*0.5),1]//tmp[floor(n*0.75),1]//tmp[floor(n*0.9),1];
end;
end;
if (zdich=1) then; do; matz=min(z)//max(z);
end;
if (&zmodval ^= 999) then; do; matz=&zmodval; quantd[1,2]=0; quantc[1,2]=0;
end;
modmatv[1,2]=nrow(matz); modmat[[1:nrow(matz)],2]=matz;
modvnm[1,2]=zname;
modmatp[1,2]=1;
end;
if ((verr=0) & (vlist=1)) then; do;
if (vnames[,i]=vname) then; do;
    verr=0; vmatch=1; v=dat[,i];
    if (center = 1) then; do;
        vmean=v[+,]/n; meanvec=j(n,1,vmean); v=v-meanvec; centvar=centvar||vname;
    end;
end;
nvarch[1,i]=1; nmods=nmods+1; vmean=v[+,]/n; tmp=v-(cons*vmean);
vsd=sqrt((1/(n-1))*t(tmp)*tmp));
vdic=1;
do jj=1 to n;
if ((v[jj,1] ^= max(v)) & (v[jj,1] ^= min(v))) then; do; vdic=0; goto leave4;
end;
end;
leave4:
if (vdich = 0) then; do;
    matv=(vmean-vs)/vmean//vmean+vmean); quantc[1,3]=1;
    if (&vmodval=999) then; do;
        if (matv[1,1] < min(v)) then; do; matv[1,1]=min(v); minprobe=1; end;
        if (matv[3,1] > max(v)) then; do; matv[3,1]=max(v); maxprobe=1; end;
    end;
    if (quantile = 1) then; do;
        quantd[1,3]=1; quantc[1,3]=0;
        tmp=v; tmp[rank(tmp)]=v;
matv = tmp[floor(n*0.10),1]//tmp[floor(n*0.25),1]//tmp[floor(n*0.5),1]//tmp[floor(n*0.75),1]//tmp[floor(n*0.9),1];
end;
end;
if (vdich=1)
then;do;matv=min(v)//max(v);wvdich=1;cmaxv=max(v);cminv=min(v);
end;
end;
if (&vmodval ^= 999)
then;do;matv=&vmodval;quantd[1,3]=0;quantc[1,3]=0;
end;
modmatv[1,3]=nrow(matv);
modmat([1:nrow(matv)),3]=matv;
modvnm[1,3]=vname;
modmatp[1,3]=1;
end;
end;
if ((qerr=0) & (qlist=1)) then;do;
if (vnames[,i]=qname) then;do;
qerr=0;qmatch=1;q=dat[,i];
if (center = 1) then;do;
qmean=q[+,]/n;meanvec=j(n,1,qmean);q=q-meanvec;centvar=centvar||qname;
end;
end;
nvarch[1,i]=1;nmods=nmods+1;qmean=q[+,]/n;temp=q-(cons*qmean);
qsd=sqrt((1/(n-1))*(t(tmp)*tmp));
qdich=1;
do jj=1 to n;
if ((q[jj,1] ^= max(q)) & (q[jj,1] ^= min(q))) then;do;qdich=0;goto leave5;
end;
leave5:
if (qdich = 0) then;do;
matq=(qmean-qsd)//qmean//((qmean+qsd);quantc[1,4]=1;
if ((&qmodval=999) & (quantile = 0)) then;do;
if (matq[1,1] < min(q))
then;do;matq[1,1]=min(q);minprobe=1;end;
if (matq[3,1] > max(q))
then;do;matq[3,1]=max(q);maxprobe=1;end;
end;
if (quantile = 1) then;do;
quantd[1,4]=1;quantc[1,4]=0;temp=q;tmp[rank(tmp)]=q;
end;
end;
matq = tmp[floor(n*0.10),1]//tmp[floor(n*0.25),1]//tmp[floor(n*0.5),1]//tmp[floor(n*0.75),1]//tmp[floor(n*0.9),1];
end;
end;
if (qdich=1) then;do;matq=min(q)//max(q);
end;
end;
if (&qmodval ^= 999)
then;do;matq=&qmodval;quantd[1,4]=0;quantc[1,4]=0;
end;
modmatv[1,4]=nrow(matq);
modmat([1:nrow(matq)),4]=matq;
modvnm[1,4]=qname;
modmatp[1,4]=1;
end;
end;
if (vnames[,i]=clname) then;do;
cld=dat[,i];cvname=vnames[,i];nvarch[1,i]=1;clsmtch=1;
end;
do j = 1 to ncol(mnames);
if (vnames[,i]=mnames[1,j]) then; do;
  mmatch=mmatch+1;m[,j]=dat[,i];
  if (center=1) & (nvarch[1,i]=0) & (mlist > 0) then; do;
    tmp=m[,j];
    meanvec=j(n,1,tmp+/n);m[,j]=m[,j]-meanvec;mmmm=m[,j];
    centvar=centvar||mnames[1,j];
  end;
  nvarch[1,i]=1;
  dichm=1;
  do jj=1 to n;
    if ((m[jj,j] ^= max(m[,j])) & (m[jj,j] ^= min(m[,j])))
      then; do; dichm=0; goto leave6; end;
  end;
  leave6:
  if ((dichm=1) & (olsdichm=1))
    then; do; note[notes,1]=18; notes=notes+1; end;
  if ((dichm=1) & (model > 3) & (mcheck = 0) & (olsdichm=0))
    then; do; errs=errs+1; runerrs[errs,1]=1; mcheck=1; end;
  if ((model <= 3) & (ncol(mnames) = 1)) then; do;
    tmp=m[,j];
    nmods=nmods+1; mmean=tmp+; tmp=m[,j]-(cons*mmean);
    msd=sqrt((1/(n-1))*t(tmp)*tmp));
    mdich=1;
    do jj=1 to n;
      if ((m[jj,j] ^= max(m[,j])) & (m[jj,j] ^= min(m[,j])))
        then; do; mdich=0; goto leave7; end;
    end;
  leave7:
  if (model = 1)
    then; do; jndich=mdich; jnmin=min(m[,j]); jnmax=max(m[,j]); end;
  if (mdich=0) then; do;
    matm=(mmean-msd)//mmean//mmean+msd); quantc[1,5]=1;
    if ((&modval=999) & (quantile=0)) then; do;
      if (matm[1,1] < min(m[,j]))
        then; do; matm[1,1]=min(m[,j]); minprobe=1; end;
      if (matm[3,1] > max(m[,j]))
        then; do; matm[3,1]=max(m[,j]); maxprobe=1; end;
    end;
  if (quantile = 1) then; do;
    quantd[1,5]=1; quantc[1,5]=0; tmp=m[,j]; tmp[rank(tmp)]=m[,j];
  end;
end;
if (minprobe=1) then; do; note[notes,1]=14; notes=notes+1; end;
if (maxprobe=1) then; do; note[notes,1]=15; notes=notes+1; end;
if ((classname ^= "XXX") & (clsmatch = 0)) then; do; note[notes,1]=15; notes=notes+1; end;
if (classname ^= "XXX") then; do;
if ((classname=zname) | (classname=vname) | (classname=wname) | (classname=xname) | (classname=yname) | (classname=qname)) then; do;
   errs=errs+1; runerrs[errs,1]=23;
end;
if ((wlist=1) & (werr=0) & (wmatch=0)) then; do; werr=3; errs=errs+1; runerrs[errs,1]=4;
end;
if ((qlist=1) & (qerr=0) & (qmatch=0)) then; do; qerr=3; errs=errs+1; runerrs[errs,1]=5;
end;
if ((vlist=1) & (verr=0) & (vmatch=0)) then; do; verr=3; errs=errs+1; runerrs[errs,1]=7;
end;
if (yerr = 1) then; do; errs=errs+1; runerrs[errs,1]=16;
end;
if (xerr = 1) then; do; errs=errs+1; runerrs[errs,1]=32;
end;
if ((model=6) & (nmeds < 2)) then; do; errs=errs+1; runerrs[errs,1]=18;
end;
if (mmatch < ncol(mnames)) then; do; errs=errs+1; runerrs[errs,1]=25;
end;
end;
end;
if (classname ^= "XXX") then; do;
cld=design(cld); cluster=ncol(cld); cld=cld[2:ncol(cld)]; clsdmy=ncol(cld);
if (clsdmy > 19) then; do; errs=errs+1; runerrs[errs,1]=26;
end;
dichy=1;
do jj=1 to n;
if ((y[jj,1] ^= max(y)) & (y[jj,1] ^= min(y))) then; do; dichy=0; goto leave8;
end;
leave8:
end;
if ((dichy=1) & (olds dichy = 1)) then; do; note[notes,1]=18; notes=notes+1; end;
if (dichy = 1) then; do; jncrit=xp2*xp2;
   if (ws=1) then; do; critterr=1; errs=errs+1; runerrs[errs,1]=28; end;
end;
covs = ncol(dat)-nvarch[,+];
if ((effsize=1) & (covmy ^= 0) & (ncovs > 0) & (model > 3) & (model < 7)) then; do;
   note[notes,1]=22; notes=notes+1; effsize=0;
end;
if (ws=1 & ncovs < nmeds) then; do; critterr=1; errs=errs+1; runerrs[errs,1]=29; end;
if (ws=1 & ncovs > nmeds) then; do; critterr=1; errs=errs+1; runerrs[errs,1]=30; end;
if (errs = 0) then; do; *[cccc];
   tmpl=quantd[,+]; tmp2=quantc[,+];
   if (tmpl > 0) then; do; note[notes,1]=4; notes=notes+1; end;
   if (tmp2 > 0) then; do; note[notes,1]=5; notes=notes+1;
end;
if (ncovs > 0) then;
c=j(n,ncovs,0);cnames=\{"x"\};j=1;
do i = 1 to ncol(vnames);
   if (nvarch[1,i]=0) then;
      c[,]=dat[,i];nvarch[1,i]=1;j=j+1;cnames=cnames||vnames[,i];
   end;
end;
cnames=cnames[1,2:ncol(cnames)];
if (ws=1) then;
covmean=c[+,:]/n;
do i = 1 to ncovs;
   meanvec=j(n,1,covmean[1,i]);
   c[,i]=c[,i]-meanvec;
end;
centvar=centvar||cnames;
x=c[,1:ncovs];
if (ncovs=nmeds) then;do;ncovs=0;end;
end;
end;
names=yname||xname||mnames||wname||zname||vname||qname;
if (ncovs > 0) then;
names=names||cnames;
end;
if ((dichy=1) & (effsize=1)) then;
   note[notes,1]=2;notes=notes+1;
end;
if ((model > 3) & (model < 6)) then;
   indeff=j(nmeds,1,0);indboot=j((boot+1),nmeds,999);
   if (mc > 0) then;do;indboot=j((mc+1),nmeds,999);end;
   if ((effsize=1) & (dichy=0)) then;
      rmeff=j((boot+1),(nmeds+1),999);
      abpseff=j((boot+1),(nmeds+1),999);
      abcseff=j((boot+1),(nmeds+1),999);
      pmeff=j((boot+1),(nmeds+1),999);
      r245=j((boot+1),1,999);
      kappa2=j((boot+1),1,999);
   end;
end;
if ((model = 6) & (effsize=1) & (dichy=0)) then;
   rmeff=j((boot+1),ncol(indboot),999);
   abpseff=j((boot+1),ncol(indboot),999);
   abcseff=j((boot+1),ncol(indboot),999);
   pmeff=j((boot+1),ncol(indboot),999);
end;
if (nmods > 0) then;
do i = 1 to 5;
   if (modmatp[1,i]=1) then;
      modmat[,tmp]=modmat[,i];modvnm[1,tmp]=modvnm[1,i];modmatv[1,tmp]=modmatv[1,i];
      tmp=tmp+1;
end;
end;
modmat=modmat[,1:nmods];modvnm=modvnm[,1:nmods];modmatv=modmatv[,1:nmods];
do i = 1 to ncol(modmatv)-1;
tmp=1;
do j = (i+1) to ncol(modmatv);
   tmp=tmp*modmatv[1,j];
end;

modprod[1,i]=tmp;
end;
modvals=j((modmatv[1,1]*modprod[1,1]),nmods,0);
do i = 1 to nmods;
    strt=1;fnsh=0;
do j = 1 to modmatv[1,i];
    tmp=j(modprod[1,i],1,modmatv[j,i]);
    fnsh=fnsh+nrow(tmp);
    modvals[(strt:fnsh),i]=tmp;
    strt=fnsh+1;
end;
end;
end;
if (model = 74) then;do;modvals=matx;modvnm=xname;
end;
vmat=j(8,nrow(modvals),0);
vmat[1,1:nrow(modvals)]=j(1,nrow(modvals),1);
vmat[5,1:nrow(modvals)]=j(1,nrow(modvals),1);
indeff=j(nrow(modvals),1,0);
if (model ^= 5) then;do;
    indboot=j(((boot+1)*nmeds),nrow(modvals),-99999999);
    indbootp=j((boot+1),nmeds,-99999999);
end;
end;
if (nmods > 0) then;do;
do i = 1 to ncol(modvals);
    if (modvnm[1,i]=wname) then;do;wcol=i;
    end;
    if (modvnm[1,i]=zname) then;do;zcol=i;
    end;
    if (modvnm[1,i]=qname) then;do;qcol=i;
    end;
    if (modvnm[1,i]=vname) then;do;vcol=i;
    end;
end;
end;
if (dichy=1) then;do;
    omx=max(y);omn=min(y);y=(y=omx);rcd=omn||0;rcd1=omx||1;rcd=rcd//rcd1;
end;
data=cons||y||m||x;
datanm="CONSTANT"//yname//t(mnames)//xname;
datanmm="CONSTANT"//yname//t(mnames)//xname;
datanm="CONSTANT"//yname//t(mnames)//xname;
yintkey=" "||" "||" "||" "||" "||" ";
if ((model < 4) & (errs = 0)) then;do;
    yintkey=yintemp[1,intcnt]||xname||" X"||mnames||" "||" ";
yintkey=yintkey//yintkey;
datamed=data;datayed=data;
datanm=yname//t(mnames)//xname;
datanm=yname//t(mnames)//xname;
yintkey=" "||" "||" "||" "||" "||" ";
if ((model < 4) & (errs = 0)) then;do;
    yintkey=yintemp[1,intcnt]||xname||" X"||mnames||" "||" ";
yintkey=yintkey//yintkey;
datamed=data;datayed=data;
datanm=yname//t(mnames)//xname;
datanm=yname//t(mnames)//xname;
yintkey=" "||" "||" "||" "||" "||" ";
if ((model = 2) | (model = 3)) then;do;
    int1=x#w;
datamed=data;datayed=data;
yintkey=yintemp[1,intcnt]||xname||" X"||wname||" "||" ";
yintkey=yintkey//yintkey;
datanm=wname//yintemp[1,intcnt];
intcnt=intcnt+1;
do\ i=1\ to\ nrow(modvals);
\ vmat[2,i]=modvals[i,2];
\ vmat[3,i]=modvals[i,1];
\ vmat[4,i]=modvals[i,1]\#modvals[i,2];
end;
end;
if\ (model=3)\ then;do;
\ yintkey=yintkey||mnames||" X"||wname||" ";
\ datanmy=datanmy||yintkey;
\ intcnt=intcnt+1;
\ int1=w\#m;\ int2=x\#w\#m;
\ datayed=datayed||int1||int2;
\ yintkey=yintkey||xname||" X"||mnames||" X"||wname;
\ datanmy=datanmy||yintkey;
\ intcnt=intcnt+1;
end;
if\ ((model=4)\ \|\ (model=5))\ then;do;
\ vmat=j(8,1,1);
end;
yintkey2=yintkey;
if\ (wm=1)\ then;do;
\ int1=x\#w;
\ datamed=datamed||w||int1;
\ yintkey=yintkey||xname||" X"||wname||" ";
\ datanmm=datanmm||wname||yintkey;
\ intcnt=intcnt+1;
do\ i=1\ to\ nrow(modvals);
\ vmat[2,i]=modvals[i,wcol];
end;
if\ (zm=1)\ then;do;
\ int1=x\#z;
\ datamed=datamed||z||int1;
\ yintkey=yintkey||xname||" X"||zname||" ";
\ datanmm=datanmm||zname||yintkey;
\ intcnt=intcnt+1;
do\ i=1\ to\ nrow(modvals);
\ vmat[3,i]=modvals[i,zcol];
end;
end;
if\ (wzm=1)\ then;do;
\ yintkey=yintkey||wname||" X"||zname;
\ datanmm=datanmm||yintkey;
\ intcnt=intcnt+1;
do\ i=1\ to\ nrow(modvals);
\ vmat[4,i]=modvals[i,wcol]\#(modvals[i,zcol]);
end;
end;
datamed=ncol(datamed);
mintkey=yintkey;
yintkey=""||""||""||""||""||""||""||"";
medints=intcnt-1;
if ((vy=1) | (xmy=1)) then;do;
  mp=1;
  do i = 1 to nrow(modvals);
    vmat[6,i]=modvals[i,1];
  end;
if (vy=1) then;do;
  datayed=datayed||v;
  datanmy=datanmy//vname;
  mm=1;
  do i = 1 to nrow(modvals);
    vmat[6,i]=modvals[i,vcol];
  end;
if (qy=1) then;do;
  mp=2;datayed=datayed||q;
  datanmy=datanmy//qname;
  mm=2;
  do i = 1 to nrow(modvals);
    vmat[7,i]=modvals[i,qcol];
  end;
end;
if (vqy=1) then;do;
  mp=3;datayed=datayed||(v#q);
  datanmy=datanmy//mnames[1,1]
  do i = 1 to nrow(modvals);
    vmat[8,i]=modvals[i,vcol]#modvals[i,qcol];
  end;
end;
end;
mints=j(n,(nm=(nm=1)),0);
do i = 0 to (nm=1);
  if (i = 0) & (vqy=1) then;do;
    yintkey=yintemp[1,intcnt]|vname|" X"||qname|" "||"
  "|yintkey=yintkey//yintkey;
  datanmy=datanmy//yintemp[1,intcnt];
  intcnt=intcnt+1;
end;
if (vy=1) then;do;
  mints[,(i=1)+1]=m[,(i=1)]|v;
  yintkey=yintemp[1,intcnt]|mnames[1,1]|" X"||qname|" "||"
  "|yintkey=yintkey//yintkey;
  datanmy=datanmy//yintemp[1,intcnt];
  intcnt=intcnt+1;
end;
if (xmy=1) then;do;
  mints[,(i=1)+1]=m[,(i=1)]|x;
  yintkey=yintemp[1,intcnt]|mnames[1,1]|" X"||xname|" "||"
  "|yintkey=yintkey//yintkey;
  datanmy=datanmy//yintemp[1,intcnt];
  intcnt=intcnt+1;
end;
if (qy=1) then;do;
  mints[,(i=1)+2]=m[,(i=1)]|q;
  yintkey=yintemp[1,intcnt]|mnames[1,1]|" X"||qname|" "||"
  "|yintkey=yintkey//yintkey;
  datanmy=datanmy//yintemp[1,intcnt];
  intcnt=intcnt+1;
if (vqy=1) then;do;
  mints[,(i=1)+3]=m[,(i=1)]|v#q;
  yintkey=yintemp[1,intcnt]|mnames[1,1]|" X"||qname|" "||"
  "|yintkey=yintkey//yintkey;
  datanmy=datanmy//yintemp[1,intcnt];
\begin{verbatim}
  intcnt=intcnt+1;
  end;
  end;
  datayed=datedayd||mints;
  end;
  mp=1;
  if (wvmy=1) then;
    do i = 1 to nrow(modvals);
      vmat[8,i]=modvals[i,wcol]#modvals[i,vcol];
    end;
    end;
  mints2=j(n,(nmeds*mp),0);
  if (wmy=1) then;
    do i = 1 to nrow(modvals);
      vmat[7,i]=modvals[i,wcol];
    end;
    if (wvmy=1) then;
      do i = 0 to (nmeds-1);
        if ((i = 0) & (wvy = 1)) then;
          datayed=datayed||(w#v);
          yintkeyt=yintemp[1,intcnt]||wname||" X"||vname||" X"||vname;
          yintkey=yintkey//yintkeyt;
          datanmy=datanmy//yintemp[1,intcnt];
          intcnt=intcnt+1;
        end;
        mints2[,(i*mp)+1]=m[,,(i+1)]#w;
        yintkeyt=yintemp[1,intcnt]||mnames[1,(i+1)]||" X"||wname||" X"||vname;
        yintkey=yintkey//yintkeyt;
        datanmy=datanmy//yintemp[1,intcnt];
        intcnt=intcnt+1;
      end;
      if (wmy=1) then;
        do i = 1 to nrow(modvals);
          vmat[6,i]=modvals[i,zcol];
        end;
        if (wzmy=1) then;
          do i = 1 to nrow(modvals);
            vmat[8,i]=modvals[i,zcol]#modvals[i,wcol];
          end;
          end;
      end;
    end;
  mints3=j(n,(nmeds*mp),0);
  if (zy=0) then;
    datayed=datayed||z;
    datanmy=datanmy//zname;
  end;
\end{verbatim}
do i = 0 to (nmeds-1);
   if (((i=0) & (wzmy = 1) & (wzy=0)) then; do;
      datayed=datayed||(w#z);
      yintkey=yintemp[1,intcnt]||wname||" X"||zname||" "||" "
   );
   yintkey=yintkey//yintkeyt;
   datanmy=datanmy//yintemp[1,intcnt];
   intcnt=intcnt+1;
   end;
   mints3[,((i*mp)+1)]=m[,((i+1)]#z;
   yintkey=yintemp[1,intcnt]||mnames[1,(i+1)]||" X"||zname||" "||wname||" "||" "
   );
   yintkey=yintkey//yintkeyt;
   datanmy=datanmy//yintemp[1,intcnt];
   intcnt=intcnt+1;
   if (wzmy = 1) then; do;
      datanmy=datanmy//yintemp[1,intcnt];
      intcnt=intcnt+1;
      mints3[,((i*mp)+2)]=m[,((i+1)]#w#z;
      yintkey=yintemp[1,intcnt]||mnames[1,(i+1)]||" X"||wname||" "||zname||" "
      );
      yintkey=yintkey//yintkeyt;
      datanmy=datanmy//yintemp[1,intcnt];
      intcnt=intcnt+1;
      end;
   end;
   datayed=datayed||mints3;
end;

decoc=1;
modmat=j(5,5,999);modmatv=j(1,5,1);modmatp=j(1,5,0);modprod=modmatv;
if ((wy = 1) & (model > 3)) then; do;
   datayed=datayed||w|w|x#w;
   decoc=decoc+1;
   deco[decoc,1]=ncol(datayed)-1;
   modmatv[1,1]=nrow(matw);
   modmat[1:nrow(matw),1]=matw;
   modvnm2[1,1]=wname;
   modmatp[1,1]=1;
   yintkey=yintemp[1,intcnt]|xname||" X"||wname||" "||" "
   );
   yintkey=yintkey//yintkeyt;
   datanmy=datanmy//wname//yintemp[1,intcnt];
   intcnt=intcnt+1;
   end;
   if (zy = 1) then; do;
      datanmy=datanmy//yintemp[1,intcnt];
      intcnt=intcnt+1;
      yintkey=yintemp[1,intcnt]|xname||" X"||zname||" "||" "
   );
   yintkey=yintkey//yintkeyt;
   datanmy=datanmy//zname//yintemp[1,intcnt];
   intcnt=intcnt+1;
   end;
   if (wzy = 1) then; do;
      datanmy=datanmy//yintemp[1,intcnt];
      intcnt=intcnt+1;
      yintkey=yintemp[1,intcnt]|xname||" X"||wname||" "||" "
   );
   yintkey=yintkey//yintkeyt;
   datanmy=datanmy//wname//yintemp[1,intcnt];
   intcnt=intcnt+1;
end;
end;
if (vxy = 1) then;
datayed=datayed||x#v;
decoc=decoc+1;
deoc[decoc,1]=ncol(datayed)-1;
modmatv[1,3]=nrow(matv);
modmat[(1:nrow(matv)),3]=matv;
modvnm2[1,3]=vname;
modmatp[1,3]=1;
yintkey=yintemp[1,intcnt]||xname||" X"||vname||" ";
yintkey=yintkey//yintkeyt;
datanmy=datanmy//yintemp[1,intcnt];
intcnt=intcnt+1;
if (qxy = 1) then;
datayed=datayed||x#q;
decoc=decoc+1;
deoc[decoc,1]=ncol(datayed)-1;
modmatv[1,4]=nrow(matq);
modmat[(1:nrow(matq)),4]=matq;
modvnm2[1,4]=qname;
modmatp[1,4]=1;
yintkey=yintemp[1,intcnt]||xname||" X"||qname||" ";
yintkey=yintkey//yintkeyt;
datanmy=datanmy//yintemp[1,intcnt];
intcnt=intcnt+1;
if (vqxy = 1) then;
datayed=datayed||x#v#q;
decoc=decoc+1;
deoc[decoc,1]=ncol(datayed)-1;
yintkey=yintemp[1,intcnt]||xname||" X"||vname||" X"||qname;
yintkey=yintkey//yintkeyt;
datanmy=datanmy//yintemp[1,intcnt];
intcnt=intcnt+1;
end;
end;
if (wvxy = 1) then;
datayed=datayed||x#w#v;
decoc=decoc+1;
deoc[decoc,1]=ncol(datayed)-1;
yintkey=yintemp[1,intcnt]||xname||" X"||wname||" X"||vname;
yintkey=yintkey//yintkeyt;
datanmy=datanmy//yintemp[1,intcnt];
intcnt=intcnt+1;
end;
modvalsd=0;ttt=modmatp[,+];ssss=modmatp[,+];
if (ssss > 0) then;
tmp=1;
do i = 1 to 5;
  if (modmatp[1,i]=1) then;
    modmat[,tmp]=modmat[,i];
    modvnm2[1,tmp]=modvnm2[1,i];
    modmatv[1,tmp]=modmatv[1,i];
    tmp=tmp+1;
  end;
end;
modmat=modmat[,1:ttt];
modvnm2=modvnm2[,1:ttt];
modmatv=modmatv[,1:ttt];
do i = 1 to (ncol(modmatv)-1); 
tmp=1;
do j = (i+1) to ncol(modmatv);
  tmp=tmp*modmatv[1,j];
```plaintext
end;
modprod[i,i]=tmp;
end;
modvalsd=j((modmatv[1,1]*modprod[1,1]),ttt,0);
do i = 1 to ttt;
   strt=1;fnsh=0;
do while (fnsh < nrow(modvalsd));
do j = 1 to modmatv[1,i];
   tmp=j(modprod[1,i],1,modmat[j,i]);
   fnsh=fnsh+nrow(tmp);
   modvalsd[(strt:fnsh),i]=tmp;
   strt=fnsh+1;
end;
end;
end;
if (ttt > 0) then;
do i = 1 to ncol(modvalsd);
   if (modvnm2[1,i] = wname) then;
      wcol=i;
   end;
   if (modvnm2[1,i] = zname) then;
      zcol=i;
   end;
   if (modvnm2[1,i] = vname) then;
      vcol=i;
   end;
   if (modvnm2[1,i] = qname) then;
      qcol=i;
   end;
end;
directv=j(nrow(modvalsd),1,1);
if (wy=1) then;
directv=directv||modvalsd[,wcol];
end;
if (zy=1) then;
directv=directv||modvalsd[,zcol];
end;
if (wzy=1) then;
directv=directv||(modvalsd[,wcol])#(modvalsd[,zcol]);
end;
if (vxy=1) then;
directv=directv||modvalsd[,vcol];
end;
if (qxy=1) then;
directv=directv||modvalsd[,qcol];
end;
if (vqxy=1) then;
directv=directv||(modvalsd[,vcol])#(modvalsd[,qcol]);
end;
if (wvxy=1) then;
directv=directv||(modvalsd[,vcol])#(modvalsd[,wcol]);
end;
ydatacol=ncol(datayed);
if (ncovs > 0) then;
do i = 1 to ncol(datamed);
   if (covmy ^= 2) then;
      datamed=datamed||c;
   end;
   if (covmy ^= 1) then;
      datayed=datayed||c;
   end;
covmeans=c[+,-]/n;
end;
if (cluster > 0) then;
do i = 1 to ncol(datamed);
   datamed=datamed||cl;
   datayed=datayed||cl;
   clmeans=cl[+,-]/n;
end;
mst=3;
mnd=mst+nmeds-1;
ydatacol=ncol(datayed);
ddatacol=ncol(datamed);
if (ncovs > 0) then;
do;
end;
```
datanmy=datanmy/(cnames');
if (model > 3) then;do;
   datanmm=datanmm/(cnames');
end;
end;
datanmy="constant"//datanmy[3:nrow(datanmy),1];
if (model > 3) then;do;
   datanmm="constant"//datanmm[3:nrow(datanmm),1];
end;
end;
datanmy={"constant"}//datanmy[3:nrow(datanmy),1];
if (model > 3) then;do;
   datanmm={"constant"}//datanmm[3:nrow(datanmm),1];
end;
end;
amm=j(2,1,0);abmm=j(2,1,0);mnv=datayed[,2]/n;mnv=mnv[+,:];mnv=j(n,1,mnv);
ssty=(datayed[,2]-mnv)##2;ssty=ssty[+,:];
sigma=(n*t(datayed)*datayed)-t(datayed[+,:]*datayed[+,:]);
sigma=sigma/(n*(n-1));
if (ws=1) then;do;
   stddevm=vecdiag(sigma);stddevm=sqrt(stddevm[3:(nmeds+2),1]);
end;
stddevy=sqrt(sigma[2,2]);
stddevx=sqrt(sigma[3,3]);
r2xy=(sigma[2,3]/(stddevy*stddevx))##2;
r2my=(sigma[2,3]/(stddevy*sqrt(sigma[3,3])))##2;
ctot=sigma[2,3]/sigma[3,3];
if ((model = 4) & (nmeds = 1) & (cluster = 0) & (ncovs = 0)) then;do;
kappaa=sigma[2,3]*sigma[2,4];
kappab=sqrt((sigma[3,3]*sigma[2,2])-(sigma[2,3]*sigma[2,3]));
kappac=sqrt((sigma[4,4]*sigma[2,2])-(sigma[2,4]*sigma[2,4]));
kappad=sqrt((sigma[4,4]*sigma[3,3])-(sigma[3,4]*sigma[3,4]));
amm[1,1]=(kappaa+kappab)/kappad;
amm[2,1]=(kappaa-kappab)/kappad;
if (sigma[3,4] < 0) then;do;
   amma=min(amm);
end;
if (sigma[3,4] > 0) then;do;
   amma=max(amm);
end;
abmm[1,1]=-amma*(kappac/kappae);
abmm[2,1]=amma*(kappac/kappae);
end;
datatm=datamed;
dataty=datayed;
mdlnms2=compress(char(model))//yname//xname;
mdlnms="Model ="//"Y = "//"X = ";
if (ws=1) then;do;
   mdlnms2=compress(char(model))//yname;
   mdlnms="Model ="//"Y = ";
end;
do i = 1 to ncol(mnames);
   mdlnms2=mdlnms2//mnames[1,i];
if ((i=1) & (ncol(mnames)=1)) then;
   mdlnms=mdlnms//"M = ";
   else mdlnms=mdlnms//mlab[i,1];
end;
if (wname ^= "XXX") then;do;
   mdlnms2=mdlnms2//wname;
   mdlnms=mdlnms//"W = ";
end;
if (zname ^= "XXX") then;do;
   mdlnms2=mdlnms2//zname;
   mdlnms=mdlnms//"Z = ";
end;
if (vname ^= "XXX") then;do;
   mdlnms2=mdlnms2//vname;
   mdlnms=mdlnms//"V = ";
end;
if (qname ^= "XXX") then; do;
    mdlnms2 = mdlnms2 // qname;
    mdlnms=mdlnms//{"Q     = "};
end;
if ((jn = 1) & (model = 1) & (jndich = 1)) then; do;
    note[notes,1]=8;
    notes=notes+1;
end;
if ((jn = 1) & (model = 3) & (jndich = 1)) then; do;
    note[notes,1]=8;
    notes=notes+1;
end;
yes=((nrow(modvals)=1) & (contrast=1) & (model = 3 | model = 2));
yes2=((nrow(modvals)=1) & (contrast=1) & (model = 2));

print "************************* PROCESS Procedure for SAS Release 2.13
*************************"
print "Written by Andrew F. Hayes, Ph.D. http://www.afhayes.com";
print "****************************************************************************
**************
print mdlnms2 [rowname = mdlnms label = "Model and Variables"];
if (ncovs > 0) then; do;
    print cnames [label = "Statistical controls:"];
end;
print n [label = "Sample size:" format = 7.0];
if (cluster > 0) then; do;
    print cluster [label = "Clustering variable and number of clusters:"];
    rowname = cvname];
end;
if ((model > 3) & (&seed ^= 0)) then; do;
    seedt=&seed;
    print seedt [label = "Custom seed:"];
end;
do bt = 1 to (boot+1); *[b];
if ((bt = 2) & (savboot = 1)) then; do;
    bootstrap=j(boot,ncol(bootcoef),-999);
end;
bootcoef=0;
if (bt > 1) then; do;
    rk=1;
do while (rk = 1);
    v=floor((ranuni(j(n,1,&seed))*n)+1;
    datayed=dataty[v,];
    detcheck=det(t(datayed)*datayed);
    rk=(detcheck=0);
    if (model > 3) then; do;
        datamed=datatm[v,];
        detcheck=det(t(datamed)*datamed);
        if (rk=1) then; do;rk=(detcheck=0);end;
    end;
    sigma=(n*t(datayed)*datayed)-(t(datayed[+])*datayed[+]);
    sigma=sigma/(n*(n-1));
    temp=vecdiag(sigma);
    * rk=temp[2:nrow(temp),1];*rk=(rk = 0);*rk=rk[+];
    bad=bad+rk;
    false=1;
end;
stddevy=sqrt(sigma[2,2]);
stddevx=sqrt(sigma[(3+nmeds),(3+nmeds)]);
ctot = \sigma[2, (3+nmeds)] / \sigma[(3+nmeds), (3+nmeds)];
if ((model = 4) & (nmeds = 1) & (cluster = 0) & (ncovs = 0)) then; do;
    r2xy = (\sigma[2, 4] / \text{stddevy*stddevx})^2;
    r2my = (\sigma[2, 3] / \text{stddevy*sqrt(\sigma[3, 3])})^2;
ss tot = \sigma[2, 2] * (n-1);
    kappaa = \sigma[2, 3] * \sigma[2, 4];
kappab = sqrt((\sigma[3, 3] * \sigma[2, 2]) - (\sigma[2, 3] * \sigma[2, 3]));
kappac = sqrt((\sigma[4, 4] * \sigma[2, 2]) - (\sigma[2, 4] * \sigma[2, 4]));
kappad = \sigma[4, 4] * \sigma[2, 2];
kappae = sqrt((\sigma[4, 4] * \sigma[3, 3]) - (\sigma[3, 4] * \sigma[3, 4]));
    am m[1, 1] = (kappaa + (kappab * kappac)) / kappad;
    am m[2, 1] = (kappaa - (kappab * kappac)) / kappad;
if (\sigma[3, 4] < 0) then; do;
    am m = min(amm);
end;
if (\sigma[3, 4] > 0) then; do;
    am m = max(amm);
end;
ab mm[1, 1] = -amma * (kappac / kappae);
ab mm[2, 1] = amma * (kappac / kappae);
end;
* mediator model;
if (model > 3) then; do;
do im = 1 to nmeds;
xm = cons | datamed[, (mnd+1):mdatacol];
xmm = "Constant" / datanmm[(2+nmeds):nrow(datanmm), 1];
invxtx = inv(xm`*xm);
    coeff = invxtx * xm`*datamed[, (2+im)];
if (model = 6) then; do;
    if (im = 1) then; do;
        xm = cons | datamed[, (mnd+1):mdatacol];
        invxtx = inv(xm`*xm);
        coeff = invxtx * xm`*datamed[, (2+im)];
    end;
    if (im > 1) then; do;
        xm = cons | datamed[, (3:(im+1))] || datamed[, ((mnd+1):mdatacol)];
xmm = "Constant" / datanmm[(2+im), 1] / datanmm[(mnd:nrow(datanmm)), 1];
        invxtx = inv(xm`*xm);
        coeff = invxtx * xm`*datamed[, (2+im)];
    mmpaths[(im+1), (2:im)] = (coeff[(2:im), 1])`;
end;
end;
if (ws = 1) then; do;
    coeff[1, 1] = datamed[+, (2+im)] / n;
end;
bootcoef = bootcoef || (coeff[1: (nrow(coeff) - clsdmy), 1])`;
if (bt = 1) then; do;
    resid = datamed[, (2+im)] - xm*coeff;
    mse = ssq(resid);
    ssse = ssq(resid);
    mnnv = data[, (2+im)] / mnnv[+,] / n; mnnv = j(n, 1, mnnv);
    sstm = (data[, (2+im)] - mnnv)^2; sstm = sstm[+];
    k3 = nrow(coeff);
    if (hc3 = 1) then; do;
        h = xm[, 1];
        do i3 = 1 to n;
            h[i3, 1] = xmi3, 1]*invxtx*(xm[i3, 1])`;
        end;
        do i3 = 1 to k3;
            xmi3 = (resid[, ncol(resid)] / (1-h))#(xm[, i3]);
        end;
end;
end;
if (hc3 ^= 1) then;do;
    do i3 = 1 to k3;
        xm[,i3]=sqrt(mse)#xm[,i3];
    end;
end;
lmat=I(nrow(coeff));
lmat=lmat[,2:ncol(lmat)];
hccov=invxtx*xm`*xm*invxtx;
mcmats[im,im]=hccov[2,2];
dfnum=nrow(coeff)-1;
dfden=n-dfnum-1;
fratio=((lmat`*coeff)`*inv(lmat`*hccov*lmat)*(lmat`*coeff))/dfnum;
coeff=coeff[1:(nrow(coeff)-clsdmy),1];
mccoeff[im,1]=coeff[2,1];
standerr=sqrt(vecdiag(invxtx*xm`*xm*invxtx));
standerr=standerr[1:(nrow(standerr)-clsdmy),1];
if (ws=1) then;do;
    standerr[1,1]=stddevm[im,1]/sqrt(n);
end;
tratio=coeff/standerr;
p=2*(1-probt(abs(tratio),(n-ncol(xm))));
temp=(n-ncol(xm));
if (ws=1) then;do;
    p=2*(1-probt(abs(tratio),(n-1)));
    temp=n-1;
end;
xd=abs(xp2);
temp=(temp*(exp((temp-(5/6))*(xd/(temp-(2/3)+(.11/temp)))+(xd/temp-(2/3)+(.11/temp))))-1));
temp1=coeff-sqrt(abs(temp))*standerr;
temp2=coeff+sqrt(abs(temp))*standerr;
if (ws=1) then;do;
op=coeff||standerr||tratio||p||temp1||temp2;
    temp=mnames[1,im];
    r2full=1-(sse/sstm);
pfr=1-probf(fratio,dfnum,dfden);
    summ=sqrt(r2full)||r2full||mse||fratio||dfnum||dfden||pfr;
    print "****************************************************************************
*************";
    print temp [rowname = "Outcome:" label = " " ];
    clnm ={"R" "R-sq" "MSE" "F" "df1" "df2" "p"};
    if (ws ^= 1) then;do;
    print summ [label = "Model Summary" colname = clnm format=&decimals];
end;
if (coefficient=0) then;do;
op=op[1:(ncol(op)-2)];
end;
clnm ={"coeff" "se" "t" "p" "LLCI" "ULCI"};
print op [label = "Model" colname = clnm rowname = xmmn format=&decimals];
if (coeffci=1) then;do;
hccovtmp=hccov[1:nrow(op),1:nrow(op)];
cnamextp=xmmn;
print hccovtmp [label = "Covariance matrix of regression parameter estimates" rowname=xmnm colname=cnamestp format=&decimals];
end;
if ((nmods > 0) & (nrow(mintkey) > 1)) then;do;
print mintkey [label = "Interactions:"];
end;
if (ws = 1) then;do;
wsdf=n-1;
tmpnamb="df = ";
print wsdf [label = " " rowname=tmpnamb];
end;
end;
end;
ymat[1,im]=coeff[2,1];
if (ws=1) then;do;
 ymat[1,im]=coeff[1,1];
end;
if (wm = 1) then;do;
 ymat[2,im]=coeff[4,1];
if (zm = 1) then;do;
 ymat[3,im]=coeff[6,1];
if (wzm = 1) then;do;
 ymat[4,im]=coeff[8,1];
end;
end;
end;
end;
end;
end;
end;
if (model = 6) then;do;
mmpaths[(im+1),1]=coeff[(im+1),1];
end;
end;
end;

* estimate model of outcome;
do totlp = 1 to (1+(toteff*(bt=1))); *[d];
xy=cons||datayed[,3+nmeds]:ydatacol];
if ((toteff=1) & (totlp = 2)) then;do;
xy=cons||datayed[,3]:ydatacol];
end;
if (dichy = 1) then;do;
 meany=datayed[,2];meany=meany[+]/n;
 pt2=j(nrow(datayed[,2]),1,meany);
 LL3 = datayed[,2]#log(pt2)+(1-datayed[,2])#log(1-pt2);
 LL3 = -2*LL3[+,
 pt1=j(n,1,0.5));
 bt1=j(ncol(xy),1,0);
 LL1=0;
 LL2=LL3;
 xy22=xy;
do jjj= 1 to ncol(xy);
 xy22[,ijk]=xy[,ijk]#pt1#(1-pt1);
end;
coeff=bt1+inv(xy22\*xy)*xy\**(datayed[,2]-pt1);
 pt1 = 1/(1+exp(-(xy*coeff))));
templ=((pt1 < 0.00000000000001) | (pt1 > 0.9999999999999));
itprob=templ[+,
 if (itprob = 0) then;do;
 LL=datayed[,2]#log(pt1)+(1-datayed[,2])#log(1-pt1);
 LL2=2*LL[+,
 end;
 bt1=coeff;
end;
if ((jjj >= iterate) & (iterr = 0)) then; 
    errs=errs+1; runerrs[err,1]=22; iter=1;
end;
do ijk = 1 to ncol(xy);
    xy22[ijk]=xy[ijk]*pt1*(1-pt1);
end;
covmat=inv(xy22`*xy);
    if (totlp ^= 2) then; 
        bootcoef=bootcoef||(coeff[1:(nrow(coeff)-clsdmy),1])`;
    end;
end;
if (dichy = 0) then; 
    invxtx=inv(xy`*xy);
    coeff=invxtx*xy`*datayed[,2];
if (effsize=1 & model > 3 & model < 7 & (ncovs > 0 | clsdmy > 0)) then; 
    xysdy=cons||datayed[,4:nmeds]:ydatacol];
    coeffsd=inv(xysdy`*xysdy)*xysdy`*datayed[,2];
    resid=datayed[,2]-(xysdy*coeffsd);
    sdy cov=sqrt(ssq(resid)/(ncol(xysdy)));
    xvaron=datayed[,3:nmeds];
    coeffx2=inv(xysdy`*xysdy)*xysdy`*xvaron;
    residx2=xvaron-(xysdy*coeffx2);
    sdxcov=sqrt(ssq(residx2)/(ncol(xysdy))); 
end;
if (totlp ^= 2) then; 
    bootcoef=bootcoef||(coeff[1:(nrow(coeff)-clsdmy),1])`;
end;
if ((nmeds = 1) & (ncovs = 0) & (cluster = 0) & (model = 4) & (bt > 1)) then; 
    resid=datayed[,2]-xy*coeff; sse=ssq(resid); r2full=1-(sse/sstot);
end;
if (bt = 1) then; 
    resid=data[,2]-xy*coeff;
    k3=nrow(coeff);
    sse=ssq(resid); mse=sse/(ncol(xy));
    if (hc3 = 1) then; 
        h=xy[,1];
        do i3 = 1 to n;
            h[i3,1]=xy[i3,]*invxtx*xy[i3,]`;
        end;
        do i3 = 1 to k3;
            xy[,i3]=resid[,ncol(resid)]/(1-h)#xy[,i3];
        end;
    end;
    if (hc3 ^= 1) then; 
        do i3 = 1 to k3;
            xy[,i3]=sqrt(mse)#xy[,i3];
        end;
    end;
covmat=invxtx*xy`*xy*invxtx;
end;
if (bt = 1) then; 
    *[f];
if (model=2) then; 
    xy2=cons||datayed[,3:ydatacol];
    temp=ncol(xy2);
    if (temp > 6) then; 
        xy3=xy2[,7:temp];
    end;
    xy2=xy2[,1:3]||xy2[,5];
if (temp > 6) then; 
    xy2=xy2||xy3;
invxtx = inv(xy2 * xy2);
coeff2 = invxtx * xy2 * datayed[,2];
ssem2 = ssq(datayed[,2] - xy2 * coeff2);
end;
standerr = sqrt(vecdiag(covmat));
if (totlp = 1) then; do;
mcmats[(nmeds+1):ncol(mcmats), (nmeds+1):ncol(mcmats)] = covmat[2:(1+nmeds), 2:(1 + nmeds)];
end;
standerr = standerr[1:(nrow(standerr) - clsdmy), 1];
if (ws = 1 & totlp = 2) then; do;
    standerr[1,1] = stddevy / sqrt(n);
end;
coeffplt = coeff;
lmat = I(nrow(coeff));
lmat = lmat[, 2:ncol(lmat)];
dfnum = nrow(coeff) - 1;
dfden = n - dfnum - 1;
fratio = (((lmat`) * coeff` * inv(lmat` * covmat * lmat) * (lmat` * coeff)) / dfnum;
if (totlp = 1) then; do;
mcoeff[(nmeds+1):nrow(mcoeff)] = coeff[2:(1+nmeds), 1];
end;
bbbb = coeff[2,1];
if (totlp = 1) then; do;
deco[1,1] = 2 + nmeds;
deco = deco[1:deco, 1];
covdirt = j((nrow(covmat) - clsdmy), (ncol(covmat) - clsdmy), 0);
covdirt = covmat[deco,];
covdir = j(nrow(covdirt), nrow(covdirt), 0);
covdir = covdirt[, deco`];
deco = coeff[deco, 1];
if (ttt > 0) then; do;
sedir = sqrt(vecdiag(directv*covdir*directv`));
directv = directv * deco;
end;
end;
if (varorder ^= 2) then; do;
end;
end;
if (varorder = 2) then; do;
end;
if (dichy = 0) then; do;
tratio = coeff / standerr;
p = 2 * (1 - probt(abs(tratio), (n - ncol(xy))));
if ((ws = 1) & (totlp = 2)) then; do;
p = 2 * (1 - probt(abs(tratio), (n - 1)));
dfden = n - 1;
end;
end;
cnms = "coeff" || "se" || "t" || "p" || "LLCI" || "ULCI";
op = coeff || standerr || tratio || p;
if(dichy=1) then;do;
  tratio=coeff/standerr;
p=2*\(1\-probnorm(\text{abs(tratio)})\);
  wald=tratio\#tratio;
  cnms="coeff"||"se"||"Z"||"p"||"LLCI"||"ULCI";
  temp=coeff-abs(xp2)*standerr;
  op=coeff|standerr|tratio||p||temp;
  temp=coeff+abs(xp2)*standerr;
  op=op||temp;
end;
if (detail = 1) then;do;
  if (totlp=2) then;do;
    print "******************************* TOTAL EFFECT MODEL
*******************************************";
    print yname [rowname = "Outcome:" label = " "];
  end;
  if (totlp ^= 2) then;do;
    print "****************************************************************************
*************";
    print yname [rowname = "Outcome:" label = " "];
  end;
end;
if ((dichy=1) & (bt=1) & (totlp=1)) then;do;
nmsd=yname||"Analysis";
print rcd [label = "Coding of binary DV for analysis" colname =
  nmsd format = 9.2];
end;
if (dichy = 0) then;do;
r2full=1-(sse/ssty);
pfr=1-probf(fratio,dfnum,dfden);
jndf=dfden;
if (ws=1) then;do;
  wddf=dfden;
end;
x=abs(xp2);
jncrit=(dfden*(exp((dfden-(5/6))\*((xd/(dfden-(2/3))+(.11/dfden))))\-(1));
summ=sqrt(r2full)||r2full||mse||fratio||dfnum||dfden||pfr;
temp=coeff-sqrt(jncrit)*standerr;
temp2=coeff+sqrt(jncrit)*standerr;
   op=coeff|standerr|tratio||p||temp1||temp2;
if (detail = 1) then;do;
  if (ws ^= 1) then;do;
    clnm = {"R" "R-sq" "MSE" "F" "df1" "df2" "p"};
    print summ [label = "Model Summary" colname = clnm format =
      &decimals];
  end;
end;
if (dichy = 1) then;do;
  LLdiff=LL3-LL2;
  mcf=LLdiff/LL3;
  cox=1-exp(-LLdiff/n);
  nagel=cox/(1-exp(-\(\text{LL3})/n));
  pf=LL2||LLdiff||mcf||cox||nagel||n;
  if (detail = 1) then;do;
    clnm = {"-2LL" "Model LL" "McFadden" "CoxSnell" "Nagelkrk" "n"};
    print pf [label = "Logistic Regression Summary" colname = clnm
      format = &decimals];
end;
end;
if (totlp=2) then;do;
datanmy="constant"//datanmy[(nmeds+2):nrow(datanmy),1];
end;
if (detail = 1) then;do;
if (coeffci = 0) then;do;
op=op[1:(ncol(op)-2)];
end;
if (ws = 1) then;do;
op2=op[1+nmecs,];
datanmy2="c path"//bpathnam[1+(nmecs>1):(1+(nmecs>1)+(nmecs-1)),1];
if (totlp=2) then;do;
datanmy2="c path";op2=op[1,];wsdf=n-1;
end;
print op2 [label = "Model" rowname=datanmy2 colname=cnms format = &decimals];
tmpnamb="df = ";
print wsdf [label = " " rowname=tmpnamb];
end;
if (ws ^= 1) then;do;
print op [label = "Model" rowname = datanmy colname=cnms format = &decimals];
if (covcoeff=1) then;do;
covmattp=covmat[1:nrow(op),1:nrow(op)];
cnamestp=datanmy`;
print covmattp [label="Covariance matrix of regression parameter estimates" rowname=datanmy colname=cnamestp format=&decimals];
end;
end;
end;
if ((ttt = 0) & (totlp = 1)) then;do;
deco=op[(nmeds+2),];
if (ws=1) then;do;
deco=op[1,];
end;
end;
if ((ttt = 0) & (totlp = 2)) then;do;
decotot = op[2,];
if (ws=1) then;do;
decotot=op[1,];
end;
end;
if ((nmods > 0) & (model > 4) & (detail = 1) & (nrow(yintkey) > 1)) then;do;
print yintkey [label = "Interactions:"];
end;
if ((nmods > 0) & (model < 4) & (detail = 1)) then;do;
print yintkey2 [label = "Interactions:"];
if (((model = 1) | (model = 2)) & (dichy = 0) & (hc3 = 0)) then;do;
temp = (((op[4,3]#op[4,3])*(1-r2full))/dfden)||((op[4,3]#op[4,3]))||1||dfden||op[4,4];
runms=yintkey2[2,1];
if (model = 2) then;do;
temp2=(((op[6,3]#op[6,3])*(1-r2full))/dfden)||((op[6,3]#op[6,3]))||1||dfden||op[6,4];
temp = temp//temp2;
frat2=(dfden*(r2full-(1-(ssem2/ssty)))/(2*(1-r2full)));
temp2=(r2full-(1-(ssem2/ssty))||frat2||2||dfden||(1-
probf(frat2,2,dfden));
temp = temp//temp2;
rnms=rnms//yintkey2[3,1]//"Both";
end;
clnm = {"R2-chng", "F", "df1", "df2", "p"};

print temp [label = "R-square increase due to interaction(s):", rowname=rnms, colname=clnm, format=&decimals];
end;
if ((model = 3) & (dichy = 0) & (hc3 = 0)) then; do;
temp = (((op[8,3] # op[8,3]) * (1 - r2full))/dfden) || (op[8,3] # op[8,3]) || dfden || op[8,4];
rnms=yintkey2[5,1];
clnm = {"R2-chng", "F(1,df2)", "df2", "p"};
print temp [label = "R-square increase due to three-way interaction:", rowname=rnms, colname=clnm, format=&decimals];
end;
end; *[f];
if ((model = 6) & (totlp = 1)) then; do;
mmpaths[nrow(mmpaths),1]=coeff[nrow(mmpaths),1];
mmpaths[nrow(mmpaths),2:(nmeds+1)]=coeff[2:(nmeds+1),1]`
end;
if (totlp = 1) then; do; *
end;
di m = 1 to nmeds; *
end;
if (model < 4) then; do;
yn[1,im]=coeff[3,1]*1-yes);
ymat[2,im]=coeff[4,1]; cmat[1,im]=covmat[3,3]; cmat[2,im]=covmat[4,4];

jnsb1b3=covmat[3,4];
if ((model = 2) | (model = 3)) then; do;

yn[3,im]=coeff[6,1]; cmat[3,im]=covmat[6,6]; cmat[6,im]=covmat[3,6]; cmat[8,im]=covmat[4,6];
end;
if (model = 3) then; do;

yn[4,im]=coeff[8,1]; cmat[4,im]=covmat[8,8]; cmat[7,im]=covmat[3,8]; cmat[9,im]=covmat[4,8];

jnsb1b3=covmat[8,8];

end;
end;
if (model > 3) then; do;
yn[5,im]=coeff[(1+im),1];
end;
if (xmy = 1) then; do;
yn[6,im]=coeff[(2+nmeds+im),1];
end;
if (vy = 1) then; do;
yn[6,im]=coeff[(3+nmeds+im),1];
end;
if (((qy = 1) & (vy = 1)) then; do;
yn[6,im]=coeff[(5+nmeds+((im-1)*2)),1];
yn[7,im]=coeff[(6+nmeds+((im-1)*2)),1];
end;
if (wmy = 1) then; do;
yn[7,im]=coeff[(3+nmeds+im-wy),1];
end;
if ((wmy = 1) & (vy = 1)) then; do;
  ymat[7, im] = coeff[(4 + (nmeds * 2) + im - wy), 1];
end;
if ((wmy = 1) & (vy = 1) & (wvmy = 1)) then; do;
  ymat[7, im] = coeff[(6 + (nmeds * 2) + ((im - 1) * 2) - wy), 1];
end;
if ((wmy = 1) & (vy = 1) & (zmy = 1)) then; do;
  ymat[6, im] = coeff[((6 - (wy * 3) + (wzm - 1) + (nmeds * 2) + ((im - 1) * 2) * wzm) + (((im - 1) * (1 - wzm)) - ((zy - wy) * 2)), 1];
  ymat[7, im] = coeff[(3 - zy + im + nmeds), 1];
  if (wzmy = 1) then; do;
    ymat[8, im] = coeff[((7 - (wy * 3)) + (nmeds * 2) + ((im - 1) * 2)), 1];
  end;
end;
if ((nmods > 0) & (model ^= 5)) then; do;
  do indlp = 1 to nrow(modvals);
    temp1 = ymat[1:4, im] # vmat[1:4, indlp];
    temp1 = temp1 + ,;
    indeff[indlp, 1] = temp1;
  end;
  if (model > 6) then; do;
    temp2 = ymat[5:8, im] # vmat[5:8, indlp];
    temp2 = temp2 + ,;
    indeff[indlp, 1] = temp1 * temp2;
  end;
end;
indboot[(bt + (im - 1) * (boot + 1)), ] = indeff';
if ((model = 8) | (model = 7)) then; do;
  indbootp[bt, im] = ymat[2, im] * ymat[5, im];
  if (which = 1) then; do;
    indbootp[bt, im] = indbootp[bt, im] * (cmaxw - cminw);
  end;
end;
if ((model = 14) | (model = 15) | (model = 74)) then; do;
  indboot[bt, im] = ymat[1, im] * ymat[6, im];
  if (wvdich = 1) then; do;
    indboot[bt, im] = indboot[bt, im] * (cmaxv - cminv);
  end;
end;
if (model = 12) then; do;
  indboot[bt, im] = ymat[4, im] * ymat[5, im];
end;
if ((model = 58) | (model = 59) & (wvdich = 1)) then; do;
end;
if (ctot = 0) then; do; ctot = .00000000000001; end;
  sumind = indboot[bt, ];
  if (im = nmeds) then; do;
    pmeff[bt, 2: (im + 1)] = indboot[bt, ] / (sumind + coeff[2 + nmeds], 1)];
  end;
  rmef[bt, (im + 1)] = indboot[bt, im] / stddevy;
  abcseff[bt, (im + 1)] = abpseff[bt, (im + 1)] * stddevx;
end;
if (ncovs > 0 & clsdmy > 0) then; do;
end;

pmeff[btt] = indboot[btt] / (indboot[btt] + mmpaths[nrow(mmpaths), bt]);
rmeff[btt] = indboot[btt] / mmpaths[nrow(mmpaths), bt];
abpseff[btt] = indboot[btt] / stddevy;
abcseff[btt] = (stddevx * indboot[btt]) / stddevy;

if (ctot = 0) then; do; ctot = .0000000000000001; end;

if (nmeds = 6) then; do;
if (model = 6) then; do;
end;
end;

if (nmeds = 4) then; do;
end;

end;

if ((effsize = 1) & (dichy = 0)) then; do;
if (ctot = 0) then; do; ctot = .0000000000000001; end;

pmeff[btt] = indboot[btt] / (indboot[btt] + mmpaths[nrow(mmpaths), bt]);
rmeff[btt] = indboot[btt] / mmpaths[nrow(mmpaths), bt];
abpseff[btt] = indboot[btt] / stddevy;
abcseff[btt] = (stddevx * indboot[btt]) / stddevy;
if (ncovs > 0 & clsdmy > 0) then; do;
end;
abps\text{eff}[bt,]=\text{indboot}[bt,]/\text{sdycov};

abc\text{seff}[bt,]=(\text{sdxcov}\times\text{indboot}[bt,]/\text{sdycov});

end;

if ((\text{nmeds} = 1) \& (\text{ncovs} = 0) \& (\text{cluster} = 0) \& (\text{model} = 4)) then;do;

r245[bt,]=r2\text{my}-(r2\text{full}-r2\text{xy});

end;

end;

end; end;

end; *[h];

end; *[g];

end; *[d];

if ((\text{savboot} = 1) \& (\text{bt} > 1)) then;do;\text{bootstrp}[(\text{bt}-1),]=\text{bootcoef};

end;

end; *[b];

if (\text{savboot}=1) then;do;\text{bootstrp}=\text{bootstrp}[,2:\text{ncol(bootstrp)}];

create \&\text{save from bootstrp};

append from bootstrp;

end;

if (\text{mc} > 0) then;do;

x1=\sqrt(-2\times\text{log(}\text{ranuni}(j(\text{mc},\text{nrow(mcmats)},&\text{seed})))\times(2\times3.14159265358979)\times(\text{ranuni}(j(\text{mc},\text{nrow(mcmats)},&\text{seed}))));

x1=x1\times\text{root(mcmats)};

do i = 1 to \text{nrow(x1)};

x1[i,]=x1[i,]+mc\text{coeff};

end;

do i = 1 to \text{nmeds};

\text{indboot}[2:\text{nrow(indboot)},i]=x1[,i]\times x1[,,(i+\text{nmeds})];

end;

end;

if ((\text{ttt} = 0) \& (\text{model} > 3)) then;do;

if (\text{toteff} = 0) then;do;

print "********************************************************************** DIRECT AND INDIRECT EFFECTS

**********************************************************************";

end;

if (\text{toteff} ^= 0) then;do;

print "********************************************************************** TOTAL, DIRECT AND INDIRECT EFFECTS

**********************************************************************";

end;

if (\text{model} ^= 74) then;do;

if (\text{dichy} = 0) then;do;

\text{clnm} = \{\text{"Effect"} "\text{SE}" "\text{t}" "\text{p}" "\text{LLCI}" "\text{ULCI}"\};

if (\text{toteff} = 1) then;do;

print \text{decotot} [\text{label} = "\text{Total effect of X on Y}" \text{colname} = \text{clnm} \text{format} = \&\text{decimals}];

end;

print \text{deco} [\text{label} = "\text{Direct effect of X on Y}" \text{colname} = \text{clnm} \text{format} = \&\text{decimals}];

end;

if (\text{dichy} = 1) then;do;

\text{clnm} = \{\text{"Effect"} "\text{SE}" "\text{Z}" "\text{p}" "\text{LLCI}" "\text{ULCI}"\};

if (\text{toteff} = 1) then;do;

print \text{decotot} [\text{label} = "\text{Total effect of X on Y}" \text{colname} = \text{clnm} \text{format} = \&\text{decimals}];

end;

print \text{deco} [\text{label} = "\text{Direct effect of X on Y}" \text{colname} = \text{clnm} \text{format} = \&\text{decimals}];

end;

end;

end;
end;
if (ttt > 0) then;do;
  print "***************************************************************************** DIRECT AND INDIRECT EFFECTS*****************************************************************************";
  clbs=modvnm2||"Effect"||"SE"||"t"||"p"||"LLCI"||"ULCI";
  tratio=directv/sedir;
  p=2*(1-probt(abs(tratio),(n-ncol(xy))));
  outp=modvalsd||directv||sedir||tratio||p;
if (dichy = 0) then;do;
  temp1=directv-sqrt(jncrit)*sedir;
  temp2=directv+sqrt(jncrit)*sedir;
  outp=outp||temp1||temp2;
end;
if (dichy = 1) then;do;
  p=2*(1-probnorm(abs(tratio)));
  temp=directv-abs(xp2)*sedir;
  outp=outp||temp;
  temp=directv+abs(xp2)*sedir;
  outp=outp||temp;
  clbs=modvnm2||"Effect"||"SE"||"Z"||"p"||"LLCI"||"ULCI";
end;
if (coeffci=0) then;do;
  outp=outp[,1:(ncol(outp)-2)];
end;
print outp [label = "Conditional direct effect(s) of X on Y at values of the moderator(s)"
format = &decimals colname = clbs rowname = " "];
end;
if ((nmods > 0) & (model ^= 5)) then;do; *[xx];
  if (model < 4) then;do;
  zmat[1,1]=1*(1-yes);
  cfse=j(nrow(modvals),1,0);
  do mmm = 1 to nrow(modvals);
    if (model = 1) then;do;
      zmat[2,1]=modvals[mmm,1]**2;
      zmat[5,1]=2*modvals[mmm,1];
    end;
    if ((model = 2) | (model = 3)) then;do;
      zmat[2,1]=modvals[mmm,2]**2;
      zmat[3,1]=modvals[mmm,1]**2;
      zmat[4,1]=(modvals[mmm,1]**2)*(modvals[mmm,2]**2)*(1-yes2);
      zmat[5,1]=2*modvals[mmm,2]*(1-yes);
      zmat[6,1]=2*modvals[mmm,1]*(1-yes);
      zmat[7,1]=2*modvals[mmm,1]*modvals[mmm,2]*(1-yes);
      zmat[8,1]=2*modvals[mmm,1]*modvals[mmm,2];
      zmat[9,1]=2*modvals[mmm,1]*(modvals[mmm,2]**2)*(1-yes2);
      zmat[10,1]=2*(modvals[mmm,1]**2)*modvals[mmm,2]*(1-yes2);
    end;
  temp=zmat#cmat;cfse[mmm,1]=sqrt(temp[+,]);
end;
end;
if (nmods > 0) then;do; *[bbb];
  clbs=modvnm||"Effect";
  do im = 1 to nmeds;
    obs=indboot[(1+(im-1)*(boot+1))];
    outp=modvals|||obs;
    if (model < 4) then;do;
      tstat=obs/cfse;
      if (dichy=0) then;do;
        pval=2*(1-probt(abs(tstat),(n-ncol(xy))));
      end;
temp = obs - sqrt(jncrit) * cfse;
outp = outp || "t" || pval || temp;

if (dichy = 1) then;
pval = 2 * (1 - probnorm(abs(tstat)));

end;

if (boot > 0) then;
ones = j(boot, 1, 1);
estmte = indboot[(1+(im-1)*(boot+1)),] ;
indboot2 = indboot[(2+(im-1)*(boot+1))*boot,];
mind = indboot2[+]/boot;

stdind = (sqrt((tmp - ((indboot2[+]/boot))/boot-1))) ;
llci = j(1, ncol(indboot2), -999);
ulci = j(1, ncol(indboot2), -999);
do eee = 1 to ncol(indboot2);
inpt = indboot2[*, eee];
inpt2 = (estmte[1, eee]*bconoff)+(9999*(1-bconoff));
%
bcboot (databcbt = inpt, estmte = inpt2);
llci[1, eee] = llcit;
ulci[1, eee] = ulcit;
if ((badlo = 1) & (llcit ^= priorlo)) then;
badend = badend || llcit;
priorlo = llcit;
end;
if ((badhi = 1) & (ulcit ^= priorhi)) then;
badend = badend || ulcit;
priorhi = ulcit;
end;
end;
outp = outp || stdind || llci` || ulci` ;
clbs = modvnm || "Boot SE" || "BootLLCI" || "BootULCI" ;
end;
mtemp = mnames[1, im];
rlbs = j(nrow(modvals), 1, mnames[1, im]);
if (model < 4) then;
if (coeffci = 0) then;
outp = outp[1:(ncol(outp)-2)];
end;
if (yes = 0) then;
print outp [label = "Conditional effect of X on Y at values of the moderator(s)"
colname = clbs
format=&decimal
rowname = " "];
end;
if (yes = 1) then;
outp = outp[3:ncol(outp)];
clbs = clbs[3:ncol(clbs)];
clbs[1, 1] = "Contrast" ;
print outp [label = "Contrast of conditional effects of X on Y"
colname = clbs
format=&decimal
rowname = " "];
end;
end;
if ((model > 5) & (mod74dic ^= 1)) then;do;
  if (im = 1) then;do;
    print "Conditional indirect effect(s) of X on Y at values of
    the moderator(s)";
  end;
  print outp [label = " " rowname = rlbs colname = clbs format =
&decimals];
end;
if ((model = 74) & (mod74dic = 1)) then;do;
  if (im = 1) then;do;
    print "Indirect effect(s) of X on Y:";
    clbs3=clbs[1,2:ncol(clbs)];
  end;
  outp3=outp[1,2:ncol(outp)];
  print outp3 [label = " " rowname = rlbs colname=clbs3 format =
&decimals];
end;
do i = notes to 1 by -1;
  if ((note[i,1]=4) & (yes=0)) then;do;
    print "Values for quantitative moderators are the 10th, 25th,
50th, 75th, and 90th percentiles.";
    print "Values for dichotomous moderators are the two values of
the moderator.";
  end;
  if ((note[i,1]=5) & (yes=0)) then;do;
    print "Values for quantitative moderators are the mean and
plus/minus one SD from mean.";
    print "Values for dichotomous moderators are the two values of
the moderator.";
  end;
  if ((note[i,1]=14) & (yes=0)) then;do;
    print "NOTE: For at least one moderator in the conditional effect
table above, one SD";
    print "     below the mean was replaced with the minimum
because one SD below the mean";
    print "     is outside of the range of the data.";
  end;
  if (note[i,1]=15) then;do;
    print "NOTE: For at least one moderator in the conditional effect
table above, one SD";
    print "     below the mean was replaced with the maximum
because one SD above the mean";
    print "     is outside of the range of the data.";
  end;
if ((model = 3) & (yes=0)) then;do;
jnvals=j(nrow(matw),7,0);
jnvals[,1]=matw;
jnvals[,2]=jnbl+jnbs3*jnvals[,1];
jnvals[,3]=sqrt(jnslb+2*jnvals[,1]*jnbs3+(jnvals[,1]**2)*jnsb3);
jnvals[,4]=jnvals[,2]/jnvals[,3];
  if (dichy=0) then;do;
    jnvals[,5]=2*(1-probt(abs(jnvals[,4]),jndf));
  end;
  if (dichy=1) then;do;
    jnvals[,5]=2*(1-probnorm(abs(jnvals[,4])));end;
jnvals[,6]=jnvals[,2]-sqrt(jncrit)#jnvals[,3];
jnvals[,7]=jnvals[,2]+sqrt(jncrit)#jnvals[,3];
clbs=clbs[1]|clbs[3:ncol(clbs)];
  if (coeffci = 0) then;do;
    jnvals=jnvals[,1:(ncol(jnvals)-2)];
end;
print jnvals [label = "Conditional effect of X*M interaction at values of W" colname = clbs rowname = " " format = &decimals];
end;
if (((jn = 1) & (model = 1) | (model = 3)) & (jndich=0) & (yes=0)) then;
ajn=(jncrit*jnsb3)-(jnbs*jnbs);bjn=2*((jncrit*jnsb1b3)-(jnbs1*jnbs3));
cjn=(jncrit*jnsb1)-(jnbs1*jnbs1);radarg=(bjn*bjn)-(4*ajn*cjn);
den=2*ajn;nrts=0;
print "****************************************************************************** *
JOHNSON-NEYMAN TECHNIQUE
*******************************************************************************
 if ((radarg >= 0) & (den ^= 0)) then;
        x21=(-bjn+sqrt(radarg))/den;
        x22=(-bjn-sqrt(radarg))/den;
        roots = 0;
        if ((x21 >= jnmin) & (x21 <= jnmax)) then;do;
          nrts=1;roots=roots//x21;
        end;
        if ((x22 >= jnmin) & (x22 < jnmax)) then;do;
          nrts=nrts+1;roots=roots//x22;
        end;
        roots=roots||j(nrow(roots),2,0);
        modtemp=m;
        if (model=3) then;do;modtemp=w;end;
        if (nrts > 0) then;do;
          roots = roots[2:nrow(roots),1:3];
          rootsum=(modtemp < roots[1,1]);roots[1,2]=(rootsum+[+]/n)*100;
          rootsum=(modtemp > roots[1,1]);roots[1,3]=(rootsum+[+]/n)*100;
          if (nrow(roots)=2) then;do;
            rootsum=(modtemp < roots[2,1]);roots[2,2]=(rootsum+[+]/n)*100;
            rootsum=(modtemp > roots[2,1]);roots[2,3]=(rootsum+[+]/n)*100;
          end;
          lohilbs="Value"||"% below"||"% above";
          print roots [label = "Moderator values(s) defining Johnson-NEYMAN significance region(s)" colname=lohilbs format = &decimals];
jnvals=j((21+nrts),7,0);
doi = 0 to 20;
        jnvals[(i+1),1]=jnmin+(i*((jnmax-jnmin)/20));
        end;
do i = 1 to nrts;
do j = 2 to nrow(jnvals);
        if ((roots[i,1] > jnvals[(j-1),1]) & (roots[i,1] < jnvals[j,1])) then;do;
          jnvals[(j+1):21+i,1]=jnvals[j:(20+i),1];
          jnvals[j,1]=roots[i,1];
        end;
end;
do i = 1 to nrow(jnvals);
        jnvals[i,2]=jnbs1+jnbs3*jnvals[i,1];
jnvals[i,3]=sqrt(jnbs1+2*jnvals[i,1]*jnbs1b3+jnvals[i,1]*jnbs3);jnvals[i,4]=jnvals[i,2]/jnvals[i,3];
        if (dichy = 0) then;do;
          jnvals[i,5]=2*(1-probt(abs(jnvals[i,4]),jndf));
        end;
        if (dichy = 1) then;do;
          jnvals[i,5]=2*(1-probnorm(abs(jnvals[i,4])));
        end;
        jnvals[i,6]=jnvals[i,2]-sqrt(jncrit)*jnvals[i,3];
jnvals[i,7]=jnvals[i,2]+sqrt(jncrit)*jnvals[i,3];
end;
if (model = 1) then; do;
    print jnvals [label = "Conditional effect of X on Y at values of the moderator (M)"]
    colname=jnclbs rowname = " " format = &decimals;
end;
    if (model = 3) then; do;
        jnclbs=jnclbs[,1]|jnclbs[,3:ncol(jnclbs)];
        print jnvals [label = "Conditional effect of X*M on Y at values of the moderator (W)"]
        colname=jnclbs rowname = " " format=&decimals;
end;
end;
end;
end;
if (nrts = 0) then; do;
    print "There are no statistical significance transition points within the observed range of the moderator";
end;
end;
end; *[bbb];
if ((model < 4) & (plot = 1)) then; do;
    dataplot=j((nrow(modvals)*nrow(matx)),(ncol(modvals)+1),0);
tmp=1;
do i = 1 to nrow(modvals);
do j = 1 to nrow(matx);
    dataplot[tmp,]=matx[j,1]|modvals[i,];
tmp=tmp+1;
end;
end;
dataplot=dataplot||j(nrow(dataplot),(1+dichy),0);
dataplot2=j(nrow(dataplot),1,1);
if (model = 1) then; do;
    dataplot2=dataplot2||dataplot[,2]||dataplot[,1]||(dataplot[,1]#dataplot[,2]);
end;
if ((model = 2) | (model = 3)) then; do;
    dataplot2=dataplot2||dataplot[,3]||dataplot[,1]||(dataplot[,1]#dataplot[,3])||dataplot[,2]||(dataplot[,1]#dataplot[,2]);
if (model = 3) then; do;
    dataplot2=dataplot2||(dataplot[,2]#dataplot[,3])||(dataplot[,1]#dataplot[,2]#dataplot[,3]);
end;
end;
do i = 1 to nrow(dataplot);
tmp=dataplot2[i,];
    if (ncovs > 0) then; do;
        tmp=tmp||covmeans;
    end;
    if (cluster > 0) then; do;
        tmp=tmp||cldmeans;
    end;
    dataplot[i,(ncol(dataplot)-(dichy))]=tmp*coeffplt;
if (dichy=1) then; do;
    dataplot[i,(ncol(dataplot))]=exp(tmp*coeffplt)/(1+exp(tmp*coeffplt));
end;
end;
clbs=xname||modvnm||"yhat";
if (dichy = 1) then; do;
    clbs=xname||modvnm||"ln(odds)||"prob";
end;
print "*********************************************
*******************************
*************";

print dataplot [label = "Data for visualizing conditional effect of X on Y" colname = clbs rowname = " " format = &decimals];
if (ncovs > 0) then;do;
    print "Estimates in this table are based on setting covariates to their sample means";
    end;
end; *[xx];
if (((model = 8) | (model = 12) | (model = 7) | (model=14) | (model=15) | (model=74) & (mod74dic=0))) | (((model = 58) | (model = 59)) & (wvdich = 1)) then;do;
    obsprod=indbootp[1,];
    if (boot > 0) then;do;
        ones=j(boot,1,1);
        estmte=indbootp[1,];
        indbootp=indbootp[2:(boot+1),];
        mnnindp=indbootp[+,]/boot;mnnindp=mnnindp`
        tmp=indbootp##2;tmp=tmp[+,];
        stdindp=(sqrt((tmp-((indbootp[+,]##2)/boot))/(boot-1)))`
        llcip=j(1,ncol(indbootp),-999); ulcip=j(1,ncol(indbootp),-999);
        do eee = 1 to ncol(indbootp);
            inpt=indbootp[,eee];inpt2=(estmte[1,eee]*bconoff)+((9999*(1-bconoff)));
            %bcboot (databcbt=inpt,estmte=inpt2);
            llcip[1,eee]=llcit;
            ulcip[1,eee]=ulcit;
            if ((badlo=1) & (llcit ^= priorlo)) then;do;
                badend=badend||llcit;
                priorlo=llcit;
            end;
            if ((badhi=1) & (ulcit ^= priorhi)) then;do;
                badend=badend||ulcit;
                priorhi=ulcit;
            end;
        end;
        outp=obsprod||stdindp||llcip`||ulcip`;
        clbs="Effect"||"Boot SE"||"BootLLCI"||"BootULCI";
        if ((model=8) | (model=12)) then;do;
            print outp [label = "Indirect effect of highest order interaction" colname = clbs rowname = mnames format = &decimals];
        end;
    end;
end; [*
end;

if (boot = 0) then;do;
    if ((model = 8) | (model = 12)) then;do;
        clnm43="Effect";
        print obsprod [label = "Indirect effect of highest order interaction" colname = clnm43 rowname = mnames format = &decimals];
    end;
if (model ^= 12) then;do;
print "**********************************INDEX OF MODERATED MEDIATION
**********************************";
clnm43="Index"
print obsprod [label = " " colname = clnm43 rowname = mnames format = &decimals]
end;
end;
end;
end;
conmake=0;concols=0;
if (((model > 3) & (model < 7)) & (contrast=1) & (nmods = 0) & (nmeds > 1)) then;do;
concols=(ncol(indboot)* (ncol(indboot)-1))/2;
indcon=j(nrow(indboot),concols,-999);
conkey=" ";
temp=1;
conmake=1;
do i = 1 to (ncol(indboot)-1);
do j = (i+1) to (ncol(indboot));
indcon[,temp]=indboot[,i]-indboot[,j];
if (model ^= 6) then;do;
conkey=mnames[1,i]||"minus"||mnames[1,j];conkey=conkey||/\nkey;
end;
if (model = 6) then;do;
conkey=indlbl2[i,1]||"minus"||indlbl2[j,1];conkey=conkey||/\nkey;
end;
end;
end;
end;
if ((model = 4) | (model = 5)) then;do; *[ddd];
clbs="Effect";rlbs="TOTAL"//mnames`
ob=indboot[1,];obs=obs[+],//obs;
if (conmake=1) then;do;
ob=obs//indcon[1,];
rlbs=rlbs//cntname[1:ncol(indcon),1];
end;
outp=obs;
outp2=outp;
if ( (effsize=1) & (dichy=0) ) then;do;
tmp=pmeff[,2:ncol(pmeff)];pmeff[,1]=tmp[,+];
tmp=rmeff[,2:ncol(rmeff)];rmeff[,1]=tmp[,+];
tmp=abpseff[,2:ncol(abpseff)];abpseff[,1]=tmp[,+];
tmp=abcseff[,2:ncol(abcseff)];abcseff[,1]=tmp[,+];
if (nmeds=1) & (ncovs = 0) & (cluster = 0) & (model = 4)) then;do;
short=eff||/\nseff||/\nseff||/\nseff;
psobs=abpseff[1,1:(nmeds+1)]
rmobs=rmeff[1,1:(nmeds+1)]
pseobs=abpseff[1,1:(nmeds+1)]
csobs=abcseff[1,1:(nmeds+1)]
if (contrast = 0) then;do;
outp2=obs||psobs||csobs||pmobs||rmobs;
end;
if (contrast = 1) then;do;
obs=obs[1:nrow(psobs),];
outp=obs2||psobs||csobs||pmobs||rmobs;
clbs="ab"||"ab_ps"||"ab_cs"||"ab/c"||"ab/c'";
if ((nmeds = 1) & (ncovs = 0) & (cluster = 0) & (model = 4)) then;
  outp2=outp2||r245obs||(obs/abmmr);
  clbs=clbs||"R-sq_med"||"kappa2";
end;

if ((boot = 0) & (mc = 0)) then;
  if (nmeds = 1) then;
    outp2=outp2[2,];
    rlbs=rlbs[2,1];
  end;
  if (model=5) then;
    outp2=outp2[1:3];
  end;
  print outp2 [label = "Indirect effect(s) of X on Y" rowname = rlbs
colname = clbs format = &decimals];
  if ((contrast=1) & (effsize=1) & (nmeds > 1)) then;
    outp2=indcon[1,`;rlbs2=cntname[1:ncol(indcon),1];
    print outp2 [label = "Contrast(s) between indirect effects" rowname = rlbs2
colname = clbs format = &decimals];
  end;
end;

if ((boot > 0) | (mc > 0)) then;
  temp=indboot[,]|indboot=temp||indboot;
  bootsz=boot;
  if (mc > 0) then;do;bootsz=mc;end;
  ones=j(bootsz,1,1);
  if (conmake=1) then;
    indboot=indboot||indcon;
  end;
estmte=indboot[1,];
  indboot=indboot[2:(bootsz+1),];
  mnind=(indboot[+,]/bootsz)`;
tmp=indboot#2;tmp=tmp[+,];
  llci=[j(1,ncol(indboot),-999);ulci=[j(1,ncol(indboot),-999);do eee = 1 to ncol(indboot);
    inpt=indboot[,]|eee];inpt2=(estmte[1,eee]*bconoff)+(9999*(1-bconoff));
%bcboot (databcbt=inpt,estmte=inpt2);
llci[1,eee]=llcit;
ulci[1,eee]=ulcit;
  if ((badlo=1) & (llcit ^= priorlo)) then;do;
    badend=badend||llcit;
priorlo=llcit;
  end;
  if ((badhi=1) & (ulcit ^= priorhi)) then;do;
    badend=badend||ulcit;
priorhi=ulcit;
  end;
  if ((effsize=1) & (dichy=0)) then;
estmte=eff[1,];
  eff=eff[2:nrow(eff),];
tmp=eff#2;tmp=tmp[+,];
  stdindf=(sqrt((tmp-(eff[+,]##2)/boot))/(boot-1)))`;
llcif=j(1,ncol(eff),-999);ulcif=j(1,ncol(eff),-999);do eee = 1 to ncol(eff);
  inpt=eff[,]|eee];inpt2=(estmte[1,eee]*bconoff)+(9999*(1-bconoff));
%bcboot (databc=inp1,estmte=inp2);
llcif[1,eee]=llcit;
ulcif[1,eee]=ulcit;
if ((badlo=1) & (llcit ^= priorlo)) then;do;
badend=badend||llcit;
priorlo=llcit;
end;
if ((badhi=1) & (ulcit ^= priorhi)) then;do;
badend=badend||ulcit;
priorhi=ulcit;
end;
end;
end;
end;
if ((boot > 0) | (mc > 0)) then;do;
outp=obs||stdind||llci`||ulci`;
if (nmeds = 1) then;do;
outp=outp[2,];rlbs=rlbs[2,1];
end;
clbs="Effect"||"Boot SE"||"BootLLCI"||"BootULCI";
if (mc > 0) then;do;
clbs="Effect"||"MC SE"||"MC LLCI"||"MC ULCI";
end;
print outp [label = "Indirect effect of X on Y" rowname = rlbs colname = clbs format = &decimals];
if ((dichy=0) & (effsize = 1)) then;do;
outp=psobs||stdindf[(2*(nmeds+1)+1):(3*(nmeds+1))+1]|llcif[1,((2*(nmeds+1))+1):((3*(nmeds+1))+1)]`||ulcif[1,((2*(nmeds+1))+1):((3*(nmeds+1))+1)]`; if (nmeds = 1) then;do;
outp=outp[2,];
end;
print outp [label = "Partially standardized indirect effect of X on Y" rowname = rlbs colname = clbs format = &decimals];
outp=csobs||stdindf[(3*(nmeds+1)+1):(4*(nmeds+1))+1]|llcif[1,((3*(nmeds+1))+1):((4*(nmeds+1))+1)]`||ulcif[1,((3*(nmeds+1))+1):((4*(nmeds+1))+1)]`; if (nmeds = 1) then;do;
outp=outp[2,];
end;
print outp [label = "Completely standardized indirect effect of X on Y" rowname = rlbs colname = clbs format = &decimals];
outp=pmoobs||stdindf[1:(nmeds+1),1]|llcif[1,1:(nmeds+1)]`||ulcif[1,1:(nmeds+1)]`; if (model = 4) then;do;
if (nmeds = 1) then;do;
outp=outp[2,];
end;
print outp [label = "Ratio of indirect to total effect of X on Y" rowname = rlbs colname = clbs format = &decimals];
outp=rmoobs||stdindf[((nmeds+1)+1):(2*(nmeds+1))+1]|llcif[1,((nmeds+1)+1):((2*(nmeds+1))+1)]`||ulcif[1,((nmeds+1)+1):((2*(nmeds+1))+1)]`; if (nmeds = 1) then;do;
outp=outp[2,];
end;
print outp [label = "Ratio of indirect to direct effect of X on Y" rowname = rlbs colname = clbs format = &decimals];
if ((nmeds = 1) & (cluster = 0) & (ncovs = 0) & (model = 4)) then;do;
r245obs=r245obs[1,1];
outp=r245obs||stdindf[(4*(nmeds+1)+1):(4*(nmeds+1)+1),1]||(llcif[1,(4*(nmeds+1)+1):(4*(nmeds+1)+1))`||ulcif[1,(4*(nmeds+1)+1):(4*(nmeds+1)+1))`];
print outp [label = "R-squared mediation effect size" rowname = rlbs colname = clbs format = &decimals];
outp=kappa2ob||stdindf[nrow(stdindf),1]||llcif[1,ncol(llcif)]||ulcif[1,ncol(u lcif)];
print outp [label = "Preacher and Kelley (2011) Kappa-squared" rowname = rlbs colname = clbs format = &decimals];
end;
end;
end;
end;
if (normal = 1) then;do;
clbs2="Effect"||"se"||"Z"||"p";
if (nmeds = 1) then;do;
print sobel [label = "Normal theory test for indirect effect" colname = clbs2 format = &decimals];
end;
if (nmeds > 1) then;do;
rlbs2=rlbs[2:nrow(rlbs),1];
print sobel [label = "Normal theory tests for specific indirect effects" rowname = rlbs2 colname = clbs2 format = &decimals];
end;
if (conmake = 1) then;do;
conkey=conkey[2:nrow(conkey),];
conlbs=cntname[1:ncol(indcon),1];
print conkey [label = "Specific indirect effect contrast definitions" rowname = conlbs];
end;
end; *[dddd];
if (model = 6) then;do;
clbs="Effect";
rlbs="TOTAL"//mnames`;
obs=indboot[1,]``;
obsfsum=obs[+1,];
obs=obsfsum//obs;
indlbl=indlbl[1:nrow(obs),1];
if (conmake = 1) then;do;
obs=obs//indcon[1,];
indlbl=indlbl//cntname[1:ncol(indcon),1];
end;
obs2=obs;
if (boot = 0) then;do;
if ((dichy=0) & (effsize = 1)) then;do;
clbs="eff"||"eff_ps"||"eff_cs"||"eff/c"||"eff/c'";
if (ncovs=0) then;do;
obs=obs||obs/stddevy||obs*stddevx/stddevy||obs/ctot||obs/mmpaths[nrow(mmpaths),1];
end;
if (ncovs > 0) | (clsdmy > 0)) then;do;
obstable=obs||obs/ctddevy||obs*stddexy/ctddevy||obs/ctot||obs/mmpaths[nrow(mmpaths),1];
end;
obs2=obs;
if (contrast=1) then;do;
obs2=obs[1:(nrow(boot)-concols),];
end;
end;
print obs2 [label = "Indirect effect(s) of X on Y" rowname = indlbl colname = clbs format = &decimals];
    if ((contrast=1) & (effsize=1)) then;do;
        outp2=indcon[1,];
        rlbs2=cnname[1:ncol(indcon),1];
        print outp2 [label = "Contrast(s) between indirect effects" rowname=rlbs2 colname=clbs format=&decimals];
    end;
end;
if (boot > 0) then;do;
    ones=j(boot,1,1);
    indboot=indboot[,]||indboot;
    if (conmake=1) then;do;
        indboot=indboot||indcon;
    end;
estmte=indboot[1,];
    /* here it is */;
    indboot=indboot[2:(boot+1),];
    mnind=indboot[+,]/boot;mnind=mnind`;
    stdind=(sqrt((tmp-((indboot[+,]##2)/boot))/(boot-1))))`;
    temp=nrow(indboot);
    llci=j(1,ncol(indboot),-999);
    ulci=j(1,ncol(indboot),-999);
    do eee = 1 to ncol(indboot);
        inpt=indboot[,eee];inpt2=(estmte[1,eee]*bconoff)+(9999*(1-bconoff));
        %bcboot (databcbt=inpt,estmte=inpt2);
        llci[1,eee]=llcit;
        ulci[1,eee]=ulcit;
        if ((badlo=1) & (llcit ^= priorlo)) then;do;
            badend=badend||llcit;
            priorlo=llcit;
        end;
        if ((badhi=1) & (ulcit ^= priorhi)) then;do;
            badend=badend||ulcit;
            priorhi=ulcit;
        end;
    end;
obs=obs||stdind||llci`||ulci`;
clbs="Effect"||"Boot SE"||"BootLLCI"||"BootULCI";
print obs [label = "Indirect effect(s) of X on Y" rowname = indlbl colname = clbs format = &decimals];
    if ((effsize=1) & (dichy=0)) then;do;
        indboot=indboot[,1:(ncol(indboot)-concols)];
        eff=abpseff[,]||abpseff||abcseff[,]||abceff[,]||pmeff[,]||pmeff||rmeff[,]||rmeff;
        effobs=eff[1,];
        /* here it is */;
        eff=eff[2:nrow(eff),];
        tmp=eff##2;tmp=tmp[+];
        stdind=(sqrt((tmp-((eff[+,]##2)/boot))/(boot-1))))`;
        llci=j(1,ncol(eff),-999);
        ulci=j(1,ncol(eff),-999);
        do eee = 1 to ncol(eff);
            inpt=eff[,eee];inpt2=(effobs[1,eee]*bconoff)+(9999*(1-bconoff));
            %bcboot (databcbt=inpt,estmte=inpt2);
            llci[1,eee]=llcit;
            ulci[1,eee]=ulcit;
            if ((badlo=1) & (llcit ^= priorlo)) then;do;
                badend=badend||llcit;
                priorlo=llcit;
            end;
end;
if ((badhi=1) & (ulcit ^= priorhi)) then;do;
    badend=badend||ulcit;
    priorhi=ulcit;
end;
end;
temp2=stdindf[1:ncol(indboot),1];
temp3=effobs[1,1:ncol(indboot)];
templow=llcif[1,1:ncol(indboot)];
temphi=ulcif[1,1:ncol(indboot)];
outp=temp3'||temp2'||templow'||temphi';
print outp [label = "Partially standardized indirect effect of X on Y" colname = clbs rowname = indlbl format = &decimals];
temp2=stdindf[(ncol(indboot)+1):(2*ncol(indboot)),1];
temp3=effobs[1,(ncol(indboot)+1):(2*ncol(indboot))];
templow=llcif[1,(ncol(indboot)+1):(2*ncol(indboot))];
temphi=ulcif[1,(ncol(indboot)+1):(2*ncol(indboot))];
outp=temp3'||temp2'||templow'||temphi';
print outp [label = "Completely standardized indirect effect of X on Y" colname = clbs rowname = indlbl format = &decimals];
temp2=stdindf[(3*ncol(indboot)+1):(4*ncol(indboot)),1];
temp3=effobs[1,(ncol(indboot)+1):(4*ncol(indboot))];
templow=llcif[1,(ncol(indboot)+1):(4*ncol(indboot))];
temphi=ulcif[1,(ncol(indboot)+1):(4*ncol(indboot))];
outp=temp3'||temp2'||templow'||temphi';
print outp [label = "Ratio of indirect to total effect of X on Y" colname = clbs rowname = indlbl format = &decimals];
end;
end;
if (nmeds = 2) then;do;
    effkey=xname||"-"||mnames[1,1]|"-"||yname||" "||" ";
tempkey=xname||"-"||mnames[1,1]|"-"||mnames[1,2]|"-"||yname;
effkey=effkey//tempkey;
effkey=indlbl[2:4,1]|effkey;
end;
if (nmeds = 3) then;do;
effkey=xname||"-"||mnames[1,1]|"-"||mnames[1,2]|"-"||mnames[1,3]|"-"||yname;
tempkey=xname||"-"||mnames[1,1]|"-"||mnames[1,2]|"-"||mnames[1,3]|"-"||yname;
effkey=effkey//tempkey;
effkey=indlbl[2:8,1]|effkey;
end;
if (nmeds = 4) then;do;
effkey=xname||"->"||mnames[1,1]||"->"||yname||" "||" "||" "||" "||" "||" "||"
"||"

tempkey=xname||"->"||mnames[1,1]||"->"||mnames[1,2]||"->"||yname||" ";

memory: ;

tempkey=xname||"->"||mnames[1,1]||"->"||mnames[1,3]||"->"||yname;

tempkey=xname||"->"||mnames[1,1]||"->"||mnames[1,4]||"->"||yname;

tempkey=xname||"->"||mnames[1,1]||"->"||mnames[1,2]||"->"||mnames[1,3]||"->"||yname;

tempkey=xname||"->"||mnames[1,1]||"->"||mnames[1,4]||"->"||mnames[1,2]||"->"||yname;

tempkey=xname||"->"||mnames[1,2]||"->"||mnames[1,3]||"->"||yname;

tempkey=xname||"->"||mnames[1,2]||"->"||mnames[1,4]||"->"||yname;

tempkey=xname||"->"||mnames[1,2]||"->"||mnames[1,3]||"->"||yname;

tempkey=xname||"->"||mnames[1,3]||"->"||mnames[1,4]||"->"||yname;

tempkey=xname||"->"||mnames[1,4]||"->"||yname;

tempkey=xname||"->"||mnames[1,4]||"->"||yname;

tempkey=xname||"->"||mnames[1,4]||"->"||yname;

effkey=effkey//tempkey;

tempkey=xname||"->"||mnames[1,1]||"->"||mnames[1,2]||"->"||mnames[1,3]||"->"||mnames[1,4]||"->"||yname;

effkey=effkey//tempkey;
effkey=indlbl[2:16,1]|effkey;
end;

print effkey [label = "Indirect effect key"]; if (conmake = 1) then;do;
conkey=conkey[2:nrow(conkey),];
conlbs=cntname[1:ncol(indcon),1];
print conkey [label = "Specific indirect effect contrast definitions"
rowname = conlbs];
end;
end;
end; *[cccc]
if (bad > 0) then;do;note[notes,1]=9;notes=notes+1;
end;
print "****************************** ANALYSIS NOTES AND WARNINGS
******************************";
do i = 1 to errs;
if (runerrs[i,1]=1) then;do;
print "ERROR: One of your declared mediators is dichotomous. This
procedure cannot be used.";
end;
if (runerrs[i,1]=2) then;do;
print "ERROR: For model 6, this procedure limits the number of mediators
to four.";
end;
if (runerrs[i,1]=3) then;do;
print "ERROR: For models 1, 2, and 3, only a single variable can be
listed in the M list.";
end;
if (runerrs[i,1]=4) then;do;
print "ERROR: You requested a model involving W but did not provide a
valid W variable name.";
end;
if (runerrs[i,1]=5) then;do;
print "ERROR: You requested a model involving Z but did not provide a valid Z variable name.";
end;
if (runerrs[i,1]=6) then;
print "ERROR: You requested a model involving Q but did not provide a valid Q variable name.";
end;
if (runerrs[i,1]=7) then;
print "ERROR: You requested a model involving V but did not provide a valid V variable name.";
end;
if (runerrs[i,1]=8) then;
print "ERROR: You specified a W variable for a model that does not need it.";
end;
if (runerrs[i,1]=9) then;
print "ERROR: You specified a Z variable for a model that does not need it.";
end;
if (runerrs[i,1]=10) then;
print "ERROR: You specified a Q variable for a model that does not need it.";
end;
if (runerrs[i,1]=11) then;
print "ERROR: You specified a V variable for a model that does not need it.";
end;
if (runerrs[i,1]=12) then;
print "ERROR: The variable specified for W has already been assigned.";
end;
if (runerrs[i,1]=13) then;
print "ERROR: The variable specified for Z has already been assigned.";
end;
if (runerrs[i,1]=14) then;
print "ERROR: The variable specified for Q has already been assigned.";
end;
if (runerrs[i,1]=15) then;
print "ERROR: The variable specified for V has already been assigned.";
end;
if (runerrs[i,1]=16) then;
print "ERROR: You did not provide a valid Y variable name.";
end;
if (runerrs[i,1]=17) then;
print "ERROR: The variable specified for Y has already been assigned.";
end;
if (runerrs[i,1]=18) then;
print "ERROR: Model 6 requires more than one mediator.";
end;
if (runerrs[i,1]=19) then;
print "ERROR: You have not specified a valid model number.";
end;
if (runerrs[i,1]=20) then;
print "ERROR: At least one and only one variable must be listed for X.";
end;
if (runerrs[i,1]=21) then;
print "ERROR: At least one and only one variable must be listed for Y.";
end;
if (runerrs[i,1]=22) then;
print "ERROR: Iteration didn't converge to a solution. Interpret results with caution.";
end;
if (runerrs[i,1]=23) then;
print "ERROR: You specified a clustering variable that does not exist in your variable list.";
end;
if (runerrs[i,1]=24) then;do;
  print "ERROR: You specified a clustering variable that has already been assigned.";
end;
if (runerrs[i,1]=25) then;do;
  print "ERROR: One of more of your M variables is not listed in the variables list.";
end;
if (runerrs[i,1]=26) then;do;
  print "ERROR: A maximum of 20 cluster units is allowed. Use multilevel modeling instead.";
end;
if (runerrs[i,1]=27) then;do;
  print "ERROR: One of the variables in your model is a constant.";
end;
if (runerrs[i,1]=28) then;do;
  print "ERROR: Dichotomous Y is not permitted with WS option.";
end;
if (runerrs[i,1]=29) then;do;
  print "ERROR: Insufficient number of variables in vars= list when using WS option.";
end;
if (runerrs[i,1]=30) then;do;
  print "ERROR: Too many number of variables in vars= list when using WS option. Covariates not allowed.";
end;
if (runerrs[i,1]=31) then;do;
  print "ERROR: mmodval and wmodval cannot both be set to zero with contrast option.";
end;
if (runerrs[i,1]=32) then;do;
  print "ERROR: You did not provide a valid X variable name.";
end;
end;
if (errs = 0) then;do;
  if ((boot > 1) | (mc > 0)) then;do;
    if ((bconoff = 1) & (boot > 0)) then;do;
      print boot [label = "Number of bootstrap samples for bias corrected bootstrap confidence intervals:" format = 10.0];
    end;
    if ((bconoff = 0) & (boot > 0)) then;do;
      print boot [label = "Number of bootstrap samples for percentile bootstrap confidence intervals:" format = 10.0];
    end;
    if (mc > 1) then;do;
      print mc [label = "Number of samples for Monte Carlo confidence intervals:" format = 10.0];
    end;
    if (booterr = 1) then;do;
      badend=badend[1,2:(ncol(badend))];
      badend=badend';
      print badend [label = "WARNING: Bootstrap CI endpoints below not trustworthy. Decrease confidence or increase bootstraps" format = &decimals];
    end;
  end;
  print conf [label = "Level of confidence for all confidence intervals in output:" format = 10.4];
  if (((center = 1) | (ws=1)) & (ncol(centvar) > 0)) then;do;
    centvar=centvar[1,2:ncol(centvar)];
  end;
end;
print centvar [label = "NOTE: The following variables were mean centered prior to analysis:"]; end;
warnrep=0;
do i = 1 to notes;
  if (note[i,1]=1) then;
    print "NOTE: Confidence level restricted to between 50 and 99.9999%. 95% confidence is provided.";
  end;
  if (note[i,1]=2) then;
    print "NOTE: Effect size measures not available for models with dichotomous outcomes.";
  end;
  if (note[i,1]=3) then;
    print "NOTE: All standard errors for continuous outcome models are based on the HC3 estimator.";
  end;
  if (note[i,1]=6) then;
    print "NOTE: The number of bootstrap samples was adjusted upward given your desired confidence.";
  end;
  if (note[i,1]=7) then;
    print "NOTE: The Johnson-Neyman method is available only for models 1 and 3.";
  end;
  if (note[i,1]=8) then;
    print "NOTE: The Johnson-Neyman method cannot be used with a dichotomous moderator.";
  end;
  if (note[i,1]=9) then;
    print bad [label = "NOTE: Some bootstrap samples had to be replaced. The number of such replacements was:" format = 10.0];
  end;
  if (note[i,1]=11) then;
    print nmiss [label = "NOTE: Some cases were deleted due to missing data. The number of such cases was:" format = 10.0];
  end;
  if (note[i,1]=12) then;
    print "NOTE: Monte Carlo method available only for models 4 and 5. Bootstrapping was used instead.";
  end;
  if (note[i,1]=13) then;
    print "NOTE: The number of Monte Carlo samples was adjusted upward given your desired confidence.";
  end;
  if (note[i,1]=19) then;
    print "NOTE: Effect sizes not available for within-subject analyses.";
  end;
  if (note[i,1]=16) then;
    print "NOTE: Normal theory tests not available for within-subject analyses.";
  end;
  if (note[i,1]=17) then;
    print "NOTE: Monte Carlo confidence intervals not available for within-subject analyses.";
  end;
  if (note[i,1]=18 & warnrep=0) then;
    print "WARNING: You have requested OLS estimation with a dichotomous criterion.";
    print "Interpret model coefficients and inferential statistics with caution.";
    warnrep=1;
  end;
end;
if (note[i,1]=20) then; do;
    print "NOTE: Saving of bootstrap estimates not available for within-
subject analyses.";
    end;
    if (note[i,1]=22) then; do;
    print "NOTE: Effect size option with covariates requires covariates in
models of M and Y.";
    end;
    end;
quit;
options pagesize=54;
%mend;
REFERENCES


69. Regional Statistical Yearbook [http://www.asr-lombardia.it/]
70. Comune di Milano - Settore Statistica [http://allegati.comune.milano.it/Statistica/Popolazione/Popolazione%202013/clreta_zone_eta_2013.pdf].
77. Regional Environmetal Protection Agency (ARPA Lombardy) [http://www2.arpalombardia.it/qarifiles/varie/Report_VMQA2011.pdf].


100. Lunneborg, C.E., *Data analysis by resampling* 2000, Duxburg: Pacific Grove, CA.


