IMPERIAL COLLEGE LONDON
Faculty of Natural Sciences

Centre for Environmental Policy

Simulating air pollution exposures in an active population as a function of activity patterns

By

Juan Pablo Orjuela Mendoza

A report submitted in partial fulfilment of the requirements for the MSc and/or DIC.

September 2014

Abstract

The exposure concentrations and contribution to total inhaled dose of four different pollutants is simulated in an active population in Barcelona (Spain) using accelerometry and Global Positioning System (GPS) information from Android smartphones using the application CalFit. The methodology uses a Land Use Regression (LUR) model to estimate ambient concentrations of particulate matter of less than 2.5 microns in diameter ($PM_{2.5}$) and nitrogen dioxide (NO_2) and a dispersion model to estimate particulate matter of less than 10 microns in diameter (PM_{10}), nitrogen oxides (NOx) and NO_2. Temporal corrections are performed to account for specific episodes of high or low concentrations of pollutants using information from a background station in Barcelona. Ratios for different microenvironments (e.g. home, work) and transport modes (e.g. car, bicycle, bus) are applied in a minute-by-minute basis to the concentrations and exposures calculated in time-weighted averages. Using the information from the accelerometer in the smartphone physical activity is estimated and the contribution to the total inhaled dose is computed.

The results show that although the population spends 9% of their time in transit, the contribution of this activity to the total inhaled dose is greater than 20% for all pollutants. This increase is mainly due to high exposure concentrations associated to transport modes and because of increased physical activity when travelling. Recognizing the influence of the correction rations in these results, Monte Carlo simulations are performed in order to estimate the level of uncertainty related to this. From this exercise it is concluded that the contribution to the total inhaled dose can vary between 17% and 45% depending on the parameters used in the model. Future work should be focused on collecting exposure data in order to improve the information on ratios adequate for different transport modes and for the verification of reasonable estimates derived from this methodology.

Acknowledgments

This project was possible thanks to the close collaboration of Antònia XXX, Albert XXX, Marta Cirah, David Donaire and all the rest of the team at CREAL. Thank you all for you help, assistance, and for making me feel at home when in Barcelona.

Many thanks to my supervisors Dr Audrey de Nazelle and Dr Mark Nieuwenhuijsen for their constant supervision and time dedicated to conclude this project successfully. Hopefully it is the beginning of many successful research projects to come.

Very special thanks to Marta Lucía my future wife for being the marvellous reason to keep going, and to my mum for the warm tea.

Table of contents

Introduction 7

Literature Review 10

Methods 13
- Sample 13
- Instruments 13
- Data processing 13
- Assessing exposure concentrations 15
- Ventilation rates 18

Results 20
- Time-activity profiles 20
- Exposure assessment 22
- Monte Carlo simulation for microenvironments ratios 27

Analysis and discussion 30

Conclusions and future work 32
- Future work 32

References 35

Appendix A 37
- Code used in R for parametric analysis 37
- Code used in R for Monte Carlo simulations 45

Appendix B 54
- Tables with parameter values for equations 3 and 4 54

Figure 1. Method for determining exposure concentrations 15
Figure 2. Time allocated to different microenvironments 21
Figure 3. Energy expenditure in different microenvironments 21
Figure 4. Average NO_2 concentrations for different microenvironments 24
Figure 5. Average $PM_{2.5}$ concentration for different microenvironments 24
Figure 6. Average NO_2 concentration for different modes 25
Figure 7. Time weighted average NO_2 concentration 26
Figure 8. Percentage contribution to inhaled dose from different microenvironments 26
Figure 9. Average NO2 concentration after Monte Carlo simulation 28
Figure 10. Average contribution to total inhaled dose after Monte Carlo simulation .. 29

Table 1. NO_2 ratios and limits of the uniform distribution for the sensitivity analysis 18
Table 2. Exposure averages according to different microenvironments 23
Table 3. NO_2 average concentrations with parametric ratios 27
Table 4. Average contribution to the total inhaled dose with parametric ratios 28

Introduction

Air pollution is one of the most relevant issues in today's cities and there are still important challenges to be overcome. According to the World Health Organization (WHO, 2014) outdoor air pollution was estimated to cause 3.7 million premature deaths worldwide in 2012. Cities in Europe frequently exceed both the European maximum levels and WHO recommended limits for local pollutants such as particulates (PM), nitrogen oxides (NOx and NO_2) and ozone (O_3). Barcelona (Spain) is no exception to this problematic and must perform important tasks in order to reduce health hazards to its inhabitants. With this in mind, the Land Use and Sustainability Department of Cataluña has established an air quality action plan seeking to reduce NO_2 and PM_{10} concentration levels in the area that regularly exceed the European legislation limits (Generalitat de Catalunya, 2012). This plan considers a pool of measures to promote the use of public transport and other measures associated to encourage a modal shift towards other modes such as bicycles and walking. Similar actions are being undertaken in many other European cities in order to reduce emissions from mobile sources.

In previous years various modelling exercises have been done in Barcelona, one of the most important being a dispersion Gaussian model done by the Energy and Air Quality Department of Barcelona (Lao & Teixidó, 2011). This model feeds from an emissions inventory that combines a remote sensing campaign and information on fixed sources in order to estimate the dispersion and chemical reactions of the NO_2 in the atmosphere to estimate the concentrations. The model also integrates a photochemical module and street canyon model part of the ADMS-Urban platform with the GIS based information of sources. However, evidence shows that ambient concentrations are not a good proxy to daily personal exposure. Experimental data collected in Barcelona, for example (Minguillón et al., 2012; Valero et al., 2009) show how outdoor concentrations of NO_2 and BC do not have a good correlation with indoor microenvironments. Although there is a clear relationship, estimating personal daily exposures to local pollutants using ambient air quality data or models is not a recommended practice. A similar case happens with the transport mode choice where these differences can be even greater as suggested by the results of several authors (Bigazzi & Figliozzi, 2012; de Nazelle, Seto, et al., 2013; de Nazelle, Rodríguez, & Crawford-Brown, 2009; Dons, Int Panis, Van Poppel, Theunis, & Wets, 2012; Int Panis et al., 2010). Inter alia, these studies point towards the need of more detailed research in daily exposures that not only consider ambient air quality but daily activities and local concentrations to better understand the effects of air quality in today's city dwellers.

The discrepancies mentioned above make the evaluation of the impacts of air pollution measures a difficult but yet very important task for policy makers, environmental researchers and epidemiologists. On one hand, various models have been proposed to assess cities air quality and personal exposure to local pollutants (Gulliver & Briggs, 2011; Pecorari et al., 2013; Physick, Powell, Cope, Boast, & Lee,

2011) but general dispersion models are not always representative of actual exposure levels of city dwellers. On the other hand, a good reasonable amount of experimental data related to exposure to various microenvironments has been collected hinting on some of the most important differences (Both, Westerdahl, Fruin, Haryanto, & Marshall, 2013; de Nazelle et al., 2009; Int Panis et al., 2010; Steinle, Reis, & Sabel, 2013) but these data are expensive to collect and a representative sample is generally unfeasible.

In order to provide enough information for epidemiological studies it is also important to consider the levels of physical activity due to its relation with respiration rates and thus, inhaled doses of air pollutants. Although many authors have used activity diaries to have a broad estimation of this variable (Steinle et al., 2013), the use of accelerometers like the ones fitted in mobile devices today has become increasingly important and even more so with the rise of smartphone users (Donaire et al. 2013).

It is in this context that the present research project has been carries out, providing additional evidence of the applicability of smartphone information to evaluate exposure levels to local pollutants based on existing air quality models and activity diaries. This project builds upon the exploratory research done by de Nazelle et al (2013) and presents the case study of Barcelona to simulate exposures to and inhaled doses of four pollutants, namely NO_2, NOx, PM_{10} and $PM_{2.5}$. The reasoning behind using these pollutants is based on their impacts in public health. As mentioned before, Barcelona has had various episodes of high levels of nitrogen oxides exciding the European regulations and the same can be said for particulates. These pollutants have been associated to respiratory diseases, irritation of the respiratory tracts and other health impacts. Nitrogen oxides are also promoters of tropospheric ozone (O_3), which also has important health impacts, and particulate matter in its smaller fractions has been associated to heart disease and is catalogued by WHO as carcinogen. These pollutants come from various sources, being internal combustion engines one of the most important specially those running on diesel.

The main aim of the project is to <u>simulate air pollution exposure and total inhaled doses for four different pollutants in an active population of 180 individuals as a function of their activity patterns</u>. The governing variables have been studied using Monte Carlo simulations in order to identify the levels of uncertainty associated to the estimation. The project was done in collaboration with the Centre for Research in Environmental Epidemiology (CREAL) in Barcelona that provided experience and the collected data of previous projects. CREAL's researchers gathered the dataset used for this project and thus the data collection process is not described in much detail here.

The specific objectives of the project are:

a.) To use georeferenced information collected from mobile applications and activity diaries from 180 active adults in Barcelona to estimate their activity profile

b.) Combine Barcelona's dispersion and land use regression (LUR) models with the georeferenced information to account for the spatial variability of pollutant ambient concentrations
c.) Use information from air quality network to account for temporal variations in pollutant ambient concentrations
d.) Find adequate correction ratios for different microenvironments and study the impact of their variability in the results
e.) Estimate intake levels by assigning appropriate respiration rates
f.) Assess the contribution of various activities in total inhaled doses

The document is organized in six main chapters, the first one being this introduction. Chapter two presents the most relevant literature associated to the different methodologies in order to give way to the detailed description of the methods employed. The results obtained are presented in chapter four and are then commented in the analysis and discussion. Finally some conclusions are drawn in the last chapter accompanied by some recommendations for future work.

Literature Review

The epidemiological study of air quality in city dwellers has been part of an immense body of literature but there are still some important knowledge gaps mainly due to the spatial and behavioural characteristics of exposure to local pollutants. Although most European cities have a robust air quality network and average concentrations can thus be estimated for the general population, having strong and sufficient data for epidemiological studies is still a big challenge. During a typical day, people move around through various different microenvironments, with a wide range of respiration rates and in various times of the day, all of these contributing to the variability of exposure levels and inhaled dose. In order to resolve for such changes various authors have proposed different methods, some of which will be explained below. The following review is not intended to be an exhaustive summary of all the available methods and studies related to this matters. However, it does touch on some of the most relevant factors to be considered and lessons learned from previous researchers.

As mentioned in the previous chapter, this study focuses on five relevant and widely studied pollutants in urban environments, namely nitrogen dioxides (NOx and NO_2) and particulate matter in both its 10 and 2.5 microns fractions (PM_{10} and $PM_{2.5}$). However, some important observations can also be derived from the study of black carbon (BC), which has been gaining higher importance in later studies. This pollutant is clearly associated to diesel fuels as is the case for NO_2, and it is usually an important fraction of the PM in cities. In this sense, reviewing the literature associated to BC is also relevant. Dons and collaborators (Dons et al., 2012) measured personal exposure to black carbon in various transport microenvironments using portable devices on 62 different individuals. Their results show that although people spent only around 6% of their time in transport, it accounted for 21% of their personal exposure and close to 30% of the inhaled dose. These results highlight the importance of transport in considering air quality impacts in people's daily routines. The times assessed to the various microenvironments where estimated using activity diaries which has been a widely used method to include time and spatial variables in personal exposure methods. The inhaled dose was estimated by defining a gender-specific ventilation rate per activity and per transport mode. A second important observation derived from Dons' results is the fact that transport mode choice is one of the most influential variables for exposure levels and inhaled dose. Although the higher concentrations where found to be related to motorized modes (i.e. cars and buses mainly), the inhaled dose for active modes (i.e. bicycle and walking) was found to be twice as much. These results have important implications in the way exposure to local pollutants is studied, stressing the importance of considering the different transport modes and the ventilation rates.

Studies as the one mentioned above performed by Dons et al., (2012), shed light over the most relevant variables to be considered when assessing exposure to local pollutants and it's public health implications. However, using portable devices to measure BC, PM or NO_2 in a big population may be financially and logistically challenging. It is not always easy to measure actual exposure levels in big populations and this is why many epidemiological studies end up relying on either global concentrations based on air quality network results or dispersion modelling on one hand, or in coarse generalizations form smaller populations. A good exercise to show how reliable these estimates can be was performed by Physick and his associates in Melbourne, Australia (Physick et al., 2011). Their methods consist in a combination of measured and modelled data (what has been sometimes classified as "hybrid" modelling) that consists of passive NO_2 monitors in a limited sample combined with dispersion modelling in order to consider a spatial variability when generalizing to a larger population. In their study Physick et al (2011) use indoor/outdoor ratios for NO_2 in order to estimate microenvironment concentrations from the results of the modelling or the air quality network. The ratios are derived from the collected samples and activity patterns are simplified assuming that everyone in the study is at work between 8:00 and 18:00 and then at home between 18:00 and 8:00. Although this method ignores the effect that moving between activities may have and assumes a timetable which is not necessarily accurate, it does allow expand the studied population by using the different indoor/outdoor ratios.

The use of GIS-tools has also been an interesting a promising new approach in the search of more precise epidemiological studies related to urban air quality exposure. Steinle, Reis and Sabel performed an overview of current practices and proposal of methodology that highlights the shift from static monitoring as the one explained above to more precise spatio-temporally resolved personal exposure assessments (Steinle et al., 2013). This study mentions the use of GPS devices in order to collect specific information on people's exact position while complementing the analysis using time-activity diaries (TADs). It explores the possibility of having personal monitors in combination with GPS and dispersion modelling using GIS tools in order to have good approximations of personal exposures and proposes the use of TADs in order to collect information on the population like gender, age, and other socio-economic and demographic data that may expand the understanding of the context around different exposure assessments. Without going into much detail, this study also mentions the use of smartphones as an important tool to substitute paper TDA's and thus improve the information collected. However, it does not mention the potential of these same devices for route tracking or reporting in real time or their possible use to estimate physical activity through their accelerometer.

Combining the methods mentioned above and making use of the best of each technic may not be an easy task but there are some important examples that seem as a promising route for the accurate estimation of personal exposure and inhaled dose. The present research project followed the methodology proposed by de Nazelle and her colleagues for estimating exposure levels and inhaled dose for 36 subjects in Barcelona (de Nazelle, Seto, et al., 2013). This methodology does not include any actual exposure or concentration measurements but relies on existing data from dispersion models, air quality networks and ratios derived for indoor, outdoor and transport microenvironments. However, it does measure the exact position of individuals by tracking the GPS signal on smartphones and records the level of physical activity through the accelerometer embedded in these devices. It then estimates the concentrations using the position and air quality models, corrects by time and day, uses ratios to account for exposure concentrations according to the microenvironment, and calculates physical activity using the accelerometer data. The detailed methodology will be explained later on but it is important to highlight the potential of this methodology for scale-up. The information is complemented with activity diaries that allow for a more detailed description of the microenvironments and related activities, besides collecting data on gender, age and body mass that will be key in the ventilation rate estimate. Considering that it does not include personal monitoring, this reduces the costs and work force needed to estimate the exposure levels of a selected cohort. In the published study, de Nazelle shows the importance of transport in the total inhaled dose, representing around 24% of the daily NO2 inhaled even though time spent travelling is close to 6%.

In summary, estimating exposure and inhaled dose from city dwellers is not a trivial or easy task and various authors have proposed different methodologies. The reviewed literature suggests that transport and mode choice are key variables in this estimation. The methodology proposed by de Nazelle (2013) may be easily scaled-up to estimate the exposure in a larger population and thus, improving the information for future epidemiological studies.

Methods

As mentioned before, the methodology followed in this study is based on the results published by de Nazelle et al (2013). The detailed process is presented in this chapter explaining each step and the source of the data. The data processing was done using the open source software of free distribution R and the code used is copied in Appendix A for reference.

Sample

The raw data used was collected by CREAL in Barcelona and it consisted of 181 individuals which were given a smartphone with an installed application called CalFit that stores information on the exact position and physical activity of the individual. The individuals were volunteers of an active population living in Barcelona, Spain that were explained the conditions of the study and were given the instructions associated to the use of the phone and the application CalFit. Out of the 181 individuals, seven had no data mainly because battery issues or a non compliance with the instructions. Data was collected all day, for three days in most cases, between beginning of March 2012 and beginning of June 2013. Each individual was also given an activity diary where she/he registered the home address, work address, different modes used and place of modal change in case it applied. Other registered data included demographic information on gender, age and body mass among others.

Instruments

The smartphones fitted with CalFit collected data on position and physical activity and stored it as each individual moved around in its daily activities. CalFit is an open-source software that runs on Android smartphones and was developed by researchers in Computer Science and Environmental Health Sciences from the University of California in Berkeley, USA. It uses the GPS system in the phone to record the position coordinates at 0.1Hz (once every 10s) and in case the satellite signal is not strong enough to generate the exact position, it uses a triangulation method based on the cellular network signal to get an approximate position every 50s. The smartphones with CalFit also record the triaxial accelerometry at 10 Hz that will allow the estimation of the physical activity of each individual while carrying the instrument.

Data processing

The information collected by CalFit needs some processing before using it to adequately predict the position of each individual. Although CalFit collects data form two different sources (GPS and cellular network) sometimes the phone will not receive any signal from either source in an extended period of time. This can be due to a poor reception (e.g. in underground travel) or because of not having enough charge

in the phone battery. In these cases, some assumptions have to be made based on the travel diary information or data of the previous and following points of the blackout. The information provided by CREAL included these corrections.
The data also needs some processing for the indoor places such as home or work locations. Individuals spend a lot of time in the same place indoors and thus, there is a reasonable amount of information that is in theory associated to only one point. However, this is not the case due to some minor shifts in the position recorded by the GPS because of signal reflections and refractions. This results in data clouds that are more accurately represented by one single point. In order to do this, CREAL has a team of experts that has developed an algorithm that identifies such clouds and compares with the address given by the individual in the activity diary. In case the cloud is close enough to the point associated to an address, the information is replaced with this single point. In case it doesn't, the centroid of the cloud is calculated and this point is used instead.

In order to determine the exposure concentrations it is important to distinguish between indoor, outdoor and transport for every point generated by CalFit. However, this information is not easily recovered mainly due to two reasons. On one hand, although individuals are asked to fill the activity diary as accurately as possible, the exact times registered as the time they left home, or got to work cannot always be trusted because people don't always remember to record the exact moment they left a place, or any other reason, so the data on the activity diary is not very useful in this case. On the other hand, for the first minutes after being indoors, it is hard to tell if a point is part of a cloud or the beginning of a trip. In this case, the algorithm developed by CREAL considers the angle between points. When the angle between a certain number of points is consistent with predefined criteria, it is assumed that these points represent the beginning of a route and are assigned as such, while if the angels have an erratic behaviour, it is assumed that these points are part of a cloud of points.

Identifying the exact route of a trip needs also some additional correction. Some points show instantaneous jumps that can be of even more than 500m, which makes no physical sense. Also, some routes seem to be shifted from their real position because when compared with the city map, they appear to be going through buildings or over the rooftops. Although these issues are very easy to spot with a visual inspection, for large data sets as these, some automatic process has to be designed. However, by the time of doing this project, there was no clear solution to this problem. Some clear outliers where identified and aligned, but shift from the roads where not possible to adjust. The shift was assumed to be no more than 50m so this resulted in the need of using a lower resolution of the dispersion models so that these shifts wouldn't be an issue. This implies less accurate results and an induced smoothness in the models that will reduce the maximum concentrations recorded. Nevertheless, manually inspecting and correcting each trip would be an extremely tedious process that would defeat the purpose of having a process able to process information for larger populations.

Finally, the two measurements, the accelerometry and the GPS position must be matched in time in a minute-by-minute base. Due to the high resolution in both the accelerometer and the GPS, this match could be done in 10s intervals. However, this time resolution implies no major change in the analysis considering that it is evaluated for an entire day, but does increase the times of execution and hardware capabilities in an exponential manner. This is why the time resolution was in this case fixed for 1minute steps. When no value was found for an entire minute or more, the position was assumed to be evenly distributed between two existing values. The physical activity and mode were kept constant.

Assessing exposure concentrations

The process of assessing exposure concentrations is illustrated in Figure 1. It consists of three individual processes that evaluate ambient air concentrations, correct for time specific concentrations and then evaluates ratios according to each microenvironment.

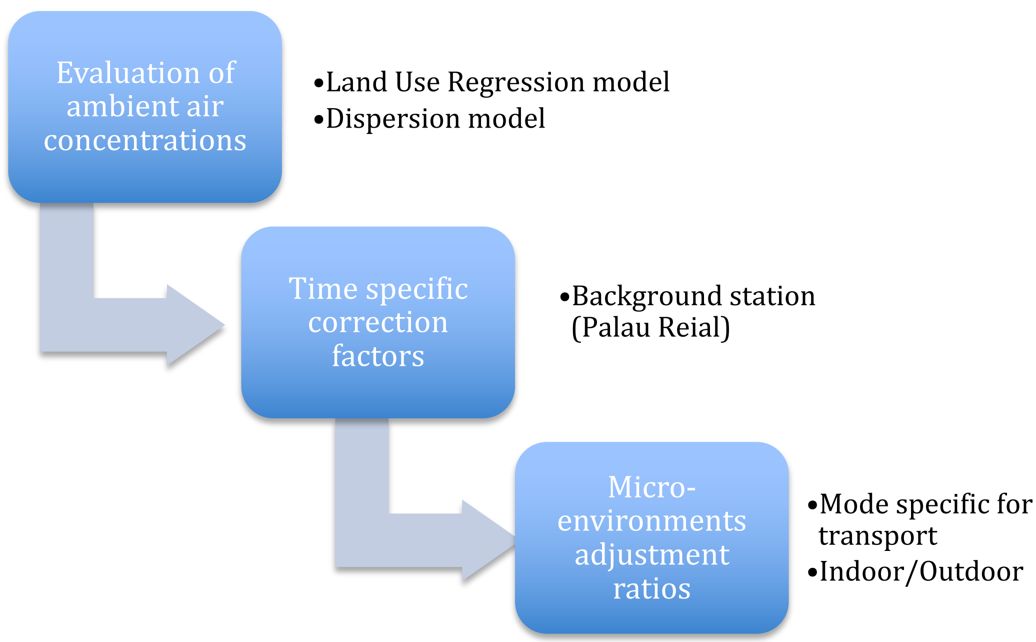

Figure 1. Method for determining exposure concentrations

Evaluation of ambient air concentrations

Ambient air concentrations of the pollutants studied was determined through the use of two different models in order to test the effect that each model may have on the results. The first one is a Land Use Regression (LUR) model based on established regressions between concentrations and different variables such as proximity to busy roads, and the other is a dispersion model that considers meteorological conditions and emissions to estimate the concentrations. Both models where developed previously for Catalonia and Barcelona respectively, and their main characteristics are described below.

The LUR model for Catalonia (de Nazelle, Aguilera, et al., 2013) was developed as part of the European Study of Cohorts for Air Pollution Effects (ESCAPE). It follows the general methodology of a LUR model where monitored pollutant concentrations are associated to land use, traffic intensity, and proximity to major roads among other variables, through the use of regressions. The model was developed using 80 passive NO2 samplers distributed through the area of Catalonia, half of these in Barcelona. The sampling took place in three seasons between February and November in 2009. The predictor variables for this model where defined as the road length within a 1,000m buffer, the traffic intensity within a 25m buffer, and the natural areas within a 5,000m buffer. These variables explained 70% of the variability in the NO2 levels.

The dispersion model developed for Barcelona (Lao & Teixidó, 2011) is an initiative of the city council as part of the Energy, Climate Change and Air Quality Plan for Barcelona 2011-2020 (PECQ) developed for aiding in the reduction of NO2 concentrations in the city. The model starts with an emission module developed through the use of remote sensing devices that characterize the total emissions in the area. These are then dispersed using a Gaussian dispersion model that includes photochemical reactions and street canyon effects that is then integrated into a GIS platform. The model has differentiated grids that become finer closer to busy roads that result in greater resolution and accuracy but at the same time requires more hardware capabilities to run. As a result, more than twelve processors working for thirty days were needed for the 2008 calibration.

The LUR model and the dispersion model operate under very different principles. While the first one is based on experimental regressions explaining most of the variability associated to the concentrations based on three predictor variables, the second one combines emissions inventories, fluid mechanics and a semi-empirical photochemical model to predict the movement of the gases once emitted. However, both models are based on a GIS platform that allows for an easy interface to overlay these concentration results with the locations of the GPS points from CalFit. Once the concentrations grids are defined from the two models, a spatial-specific ambient concentration can be assigned to every point in the georeferenced information from every individual. The resolution defined for both models was 100x100m even though they were available in a 5x5m resolution due to the uncertainty of having the GPS information to a greater precision than 100m. Although this implies less accuracy in the results and the loss of maximum concentrations closer to the sources, the GPS information from the used smartphones did not have such detail.

In order to comply with Spanish regulation that forbids the movement of georeferenced information outside the Spanish borders, this part of the process was developed in Barcelona at CREAL offices. After the concentrations and GPS points are overlaid, the georeferenced information can be eliminated and the result can be exported. It is worth noting that although originally it was thought to simulate some scenarios based on changing the route or the mode choice, this was not possible due to the impossibility of taking this information to London (England) where this analysis was going to take place.

Time specific correction factors

The time variability is accounted for using the historical results of a background station that in the case of Barcelona is the Palau Reial station. Due to a problem in the equipment measuring other pollutants, the station only has quality information for NO2. Considering the relationship between all the pollutants studied and the diesel sources, the correction factors defined for NO2 where considered valid for all other pollutants as well. This may not be the ideal case but there was insufficient data for all other pollutants.

The concentration assigned to each point was therefore corrected in time in accordance with equation 1:

$$C_{p\_t} = C_p \cdot \frac{C_{BG\_t}}{C_{BG\_annual}} \qquad (1)$$

Where C_{p_t} is the concentration in point *p* and time *t*, C_p is the concentration in point *p* given by the dispersion or LUR models, C_{BG_t} is the concentration in the background station in time *t* and C_{BG_annual} is the annual average concentration in the background station.

Microenvironments ratios

By matching the activity diaries with the GPS results it is possible to associate to every point a corresponding microenvironment and transport mode. The activity diaries defined eight different transport modes and five additional microenvironments, as shown on Table 1. For each one of those, an initial ratio was defined in order to adjust the concentrations for each point for the appropriate microenvironment. These values where defined for each pollutant based on the results reported by de Nazelle (2013).

The initial ratio was used as the best estimate of the ratio and was the first run of the model was performed with this value. However, a sensitivity analysis was performed in order to identify the impact of these ratios in the results. This was done through a Monte Carlo analysis with 100 repetitions of the complete set of results and making the ratio a random variable of a uniform distribution between the minimum and maximum values also defined in Table 1.

Table 1. NO_2 ratios and limits of the uniform distribution for the sensitivity analysis

Type	Name	Initial ratio	Minimum value	Maximum value
Transport modes	Walking	1.4	0.8	2.4
	Bicycle	2.3	0.4	12.2
	Car	2.3	1.4	3.9
	Bus/Tram	1.7	0.9	3.3
	Metro/Train	1.7	0.9	3.3
	Other modes	1.0	1.0	1.0
Microenvironments	Home	1.4	1.0	1.8
	Work	2.3	1.9	2.7
	Recreational (outdoors)	2.3	2.3	2.3
	Others (indoors)	1.4	1.4	1.4
	Others (outdoors)	2.3	2.3	2.3

The activity diary allows for the user to include a multimodal trip that consists of up to four different modes. However there was no way of specifying the place or time of the change making it impossible to identify this in the GPS database. For this study multimodal trips where ignored and only the first mode was considered unless no information was provided for it, in which case the second mode was used. Although less than 20% of the trips mentioned a second mode and less that 5% mentioned a third mode, this may be an important source of error and should be avoided in future studies. It is important to develop as part of the data processing algorithm a better estimate of the multimodal trips.

Ventilation rates

Finally, the exposures are also weighted by the ventilation rates defined by the accelerometry from CalFit. This is done by using the vertical displacement from the accelerometer information and calculating the metabolic equivalents (METs) from equation 2 developed by Donaire et al (2013).

$$METs = 1.2907087 + (0.4141791 \cdot VT\ g/min) \qquad (2)$$

Where VT is the vertical component of the CalFit accelerometry.

Once the METs are defined, the oxygen uptake per kg of body mass and the ventilation rate can be computed through the use of the equations defined by Johnson (2002) that are presented below.

$$\frac{VO_2(i,j,k)}{BM} = ECF(k) \cdot METs(i,j,k) \cdot RMR(k)/BM \qquad (3)$$

Where VO_2 is the oxygen uptake of the *i*-th activity of day *j* individual *k*, BM is the individual body mass, ECF is the energy conversion factor defined as the volume of oxygen required to produce one kilocalorie of energy, and RMR is the resting metabolic rate.

And,

$$ln[V_E(i,j,k)/BM(k)] = a + b \cdot ln\left[\frac{VO_2(i,j,k)}{BM}\right] + d(k) + e(i,j,k) \qquad (4)$$

Where V_E is the ventilation rate and *a*, *b*, *d* and *e* are gender and age specific parameters. The tables suggested by Johnson (2002) in order to compute their values as well as the suggested values for ECF and RMR can be found in Appendix B.

Results

Out of the 181 individuals part of the study, there was enough data of exactly 168 from which exactly half were women and the other half were men. The age range was between 18 and 56 years of age and an average of 34 years, having thus a population composed of only young adults. Although some people failed to charge their smartphone of would not open CalFit after turning it on loosing some of the information, no individual had less than six days of complete information and some managed to collect up to eleven days, resulting in an average of the equivalent of 173 hours of useable information per individual. These numbers are based on the raw data provided by CREAL.

Time-activity profiles

On average people spent 57.3% of their time at home followed by 19.8% at work. On average people spent 9.0% of their time in transit but their energy expenditure averaged 3.6 METs while at home and work it averaged 1.7 and 1.9 METs respectively. As shown by Figure 2 and Figure 3, time spent in transit is much less than any other microenvironment and the results are much less dispersed but in terms of the energy expenditure the opposite is observed. This indicates that although not much time is spent in transit and that there is not much difference between individuals, the coarse of their physical activity is performed during their transit between places. Note for example that the category "Others" may include time in the outdoors or such activities related to physical activity. Nevertheless, the average energy expenditure is greater when in transit that in any other microenvironment. The greater dispersion is associated to the broad spectrum of transport modes.

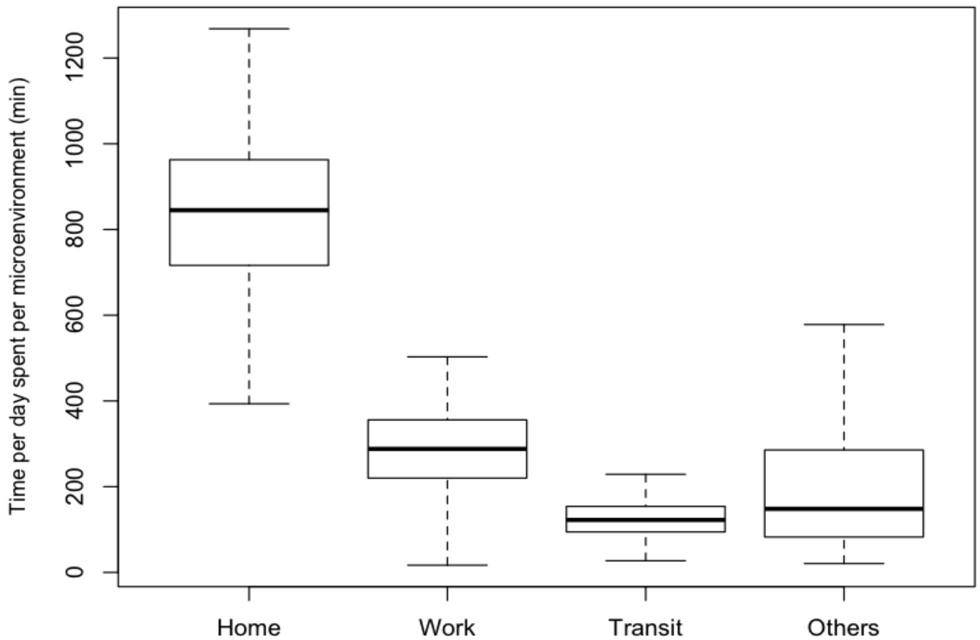

Figure 2. Time allocated to different microenvironments

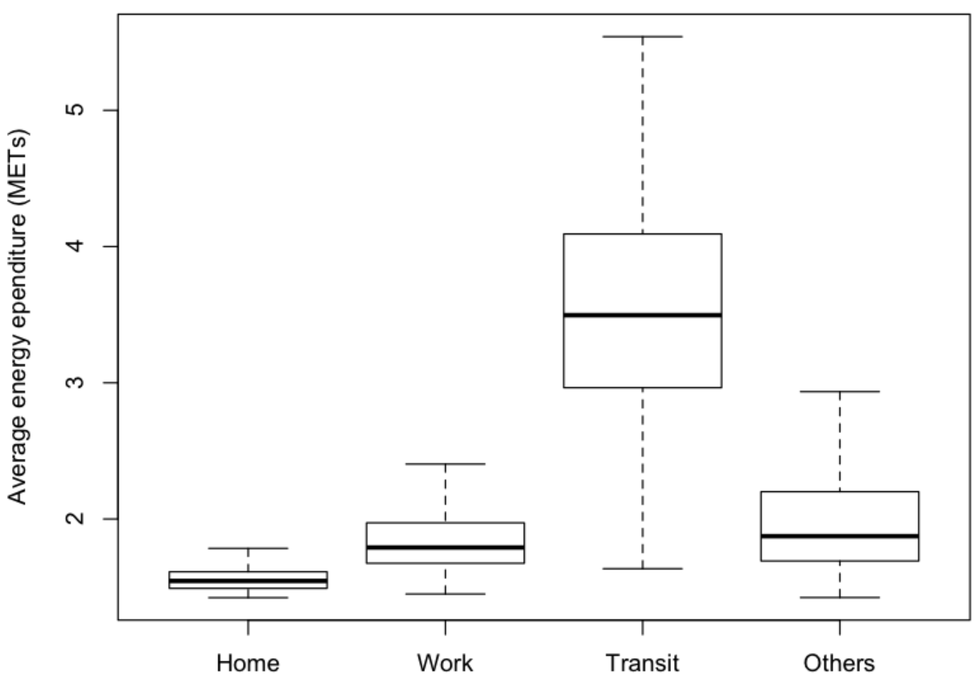

Figure 3. Energy expenditure in different microenvironments

Exposure assessment

The methodology followed allowed to evaluate different models, pollutants and correct for different aspects of the activities carried out by individuals.

Table 2 shows the effect of using the correction factors and the results for the different microenvironments for each pollutant.

In general terms, not including the correction factors for the different microenvironments results in an underestimation of both the exposure concentration and the standard deviation. This difference is especially noticeable for the microenvironments associated to transit, where for example $PM_{2.5}$ concentrations estimated by the LUR model result in a mean of 18.0µg/m^3 (+/- 2.1) but with the adjustments these can rise to 37.0µg/m^3 (+/- 26.5). A similar behaviour is observed for other pollutants like NO_2 that according to de dispersion model concentrations are on average 53.6µg/m^3 (+/- 7.2) but with the correction factors these are calculated as 110.4µg/m^3 (+/- 77.3).

In contrast to the above changes, the changes between the NO2 estimated using the LUR model and the dispersion model are clearly smaller. For the microenvironment Home, for example, the LUR model estimates an average concentration of 54.1µg/m^3 (+/- 16.9) while the dispersion model estimates an average of 50.2µg/m^3 (+/- 10.7). For all microenvironments the LUR model estimates a slightly higher concentration with a larger standard deviation. It is important to note that once the correction factors are applied, these differences are accentuated but are still clearly not the dominating factor. In this sense, the results would suggest that the impact of including the correction factors for time, space and microenvironment have a greater effect on the concentrations than the decision on using the LUR or the dispersion model. Figure 4 shows the average NO_2 concentration using the LUR model after all the adjustment factors have been considered and Figure 5 shows a similar graph for $PM_{2.5}$.

Table 2. Exposure averages according to different microenvironments

Pollutant	Adjustments	Microenvironments							
		Home		Work		Transit		Others	
		Mean (μgm^{-3})	Standard Deviation	Mean (μgm^{-3})	Standard Deviation	Mean (μgm^{-3})	Standard Deviation	Mean (μgm^{-3})	Standard Deviation
$PM_{2.5}$	Annual mean LUR model	17.0	3.1	17.9	3.0	18.0	2.1	16.8	2.6
	With corrections (LUR)	24.5	8.6	40.4	16.9	37.0	26.5	33.9	17.9
PM_{10}	Annual mean dispersion model	37.6	3.7	38.7	3.6	38.6	1.7	37.7	2.7
	With corrections (dispersion)	54.6	18.5	87.2	32.8	79.6	55.5	76.3	38.0
NOx	Annual mean dispersion model	73.0	24.7	81.9	34.7	81.5	17.1	73.3	17.5
	With corrections (dispersion)	104.7	44.4	182.8	93.6	167.9	118.7	147.1	78.9
NO_2	Annual mean LUR model	54.1	16.9	55.0	18.9	56.7	12.5	51.6	15.4
	Annual mean dispersion model	50.2	10.7	54.0	13.4	53.6	7.2	50.2	7.7
	With corrections (LUR)	78.1	34.0	123.3	59.5	116.4	82.9	104.0	61.3
	With corrections (dispersion)	72.5	26.3	120.9	50.0	110.4	77.3	101.1	52.0

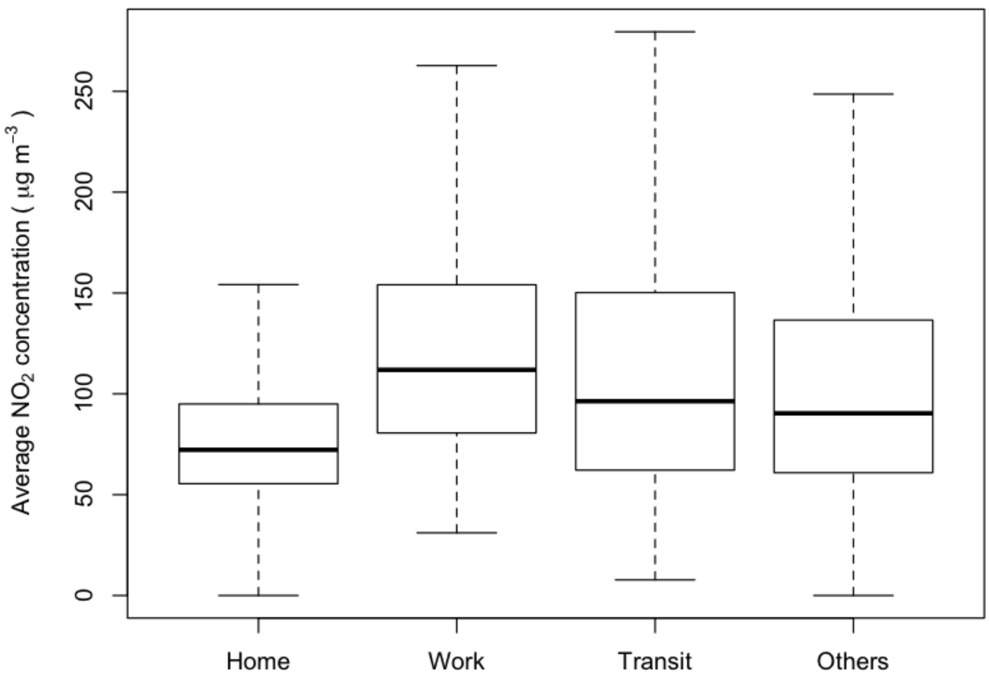

Figure 4. Average NO_2 concentrations for different microenvironments

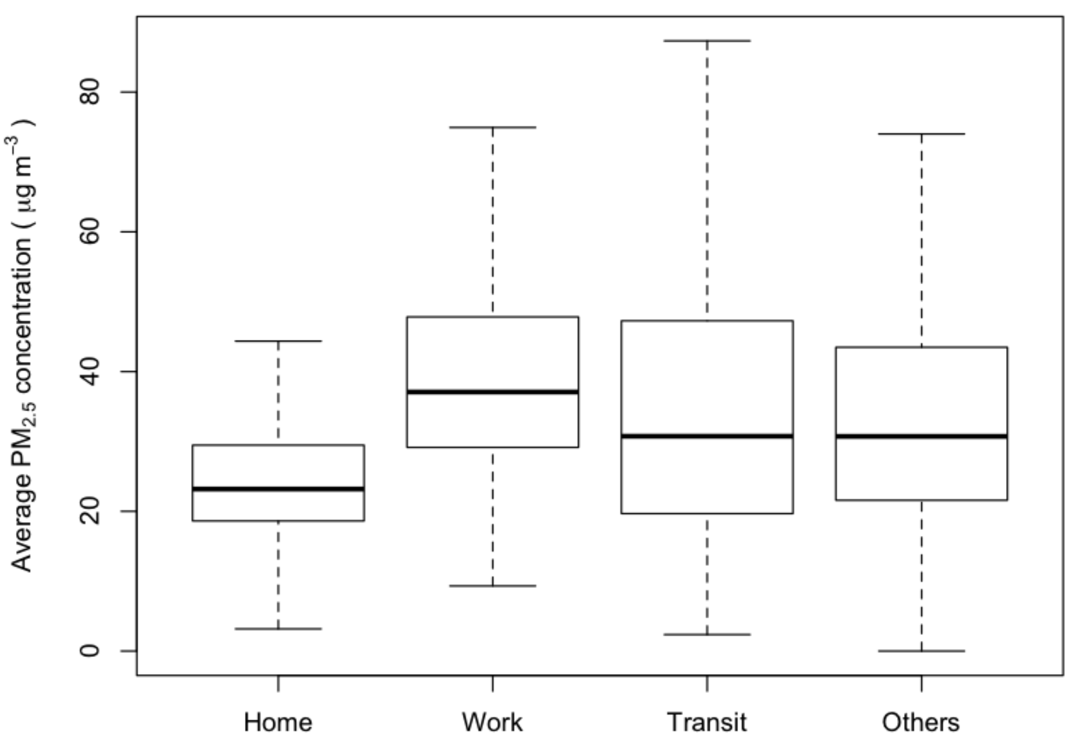

Figure 5. Average $PM_{2.5}$ concentration for different microenvironments

The previous figures show how transit microenvironments have a broader spectrum of exposure concentrations and can reach much higher values than those at home. The distributions of concentrations according to different transport modes are shown in Figure 6. According to these results the bicycle has the highest average of NO_2 concentrations and walking has the lowest ones. The concentrations for car passengers are on average similar to car drivers as could be expected but there is a clear increase in the dispersion of the data. This cannot be explained only by the correction factors used considering that both car driver and passenger where given the same ratio.

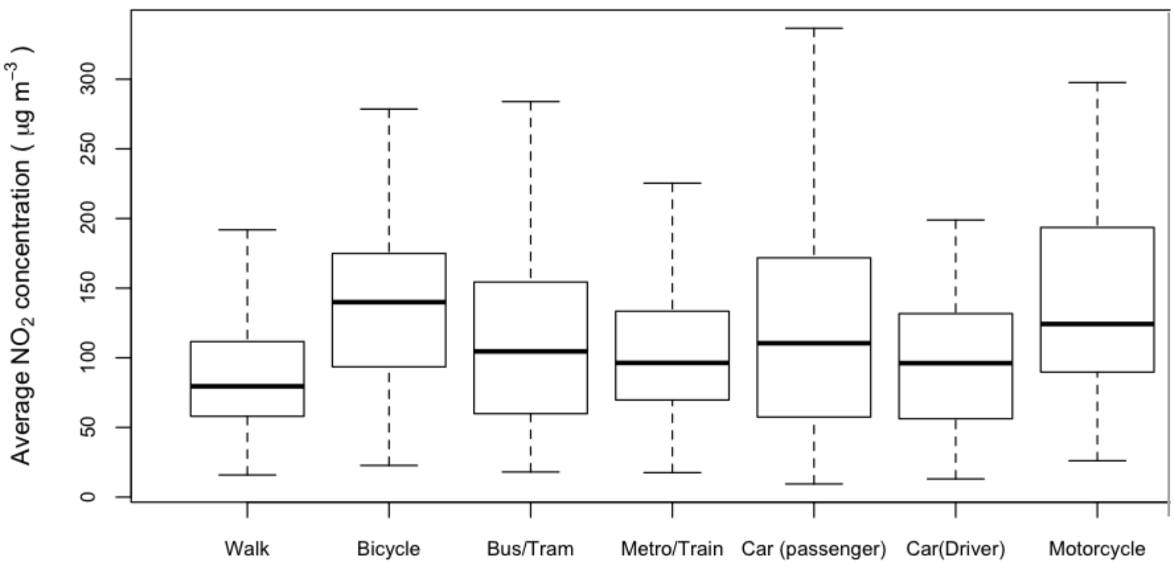

Figure 6. Average NO_2 concentration for different modes

Although on average people spend 9% of their time in transit, the higher concentrations related to these microenvironments result in a higher proportion of transit contribution to time-weighted daily exposures. These results are shown in Figure 7 for NO_2 computed according to the LUR model. The contribution of transit is now raised to an average of 11%.

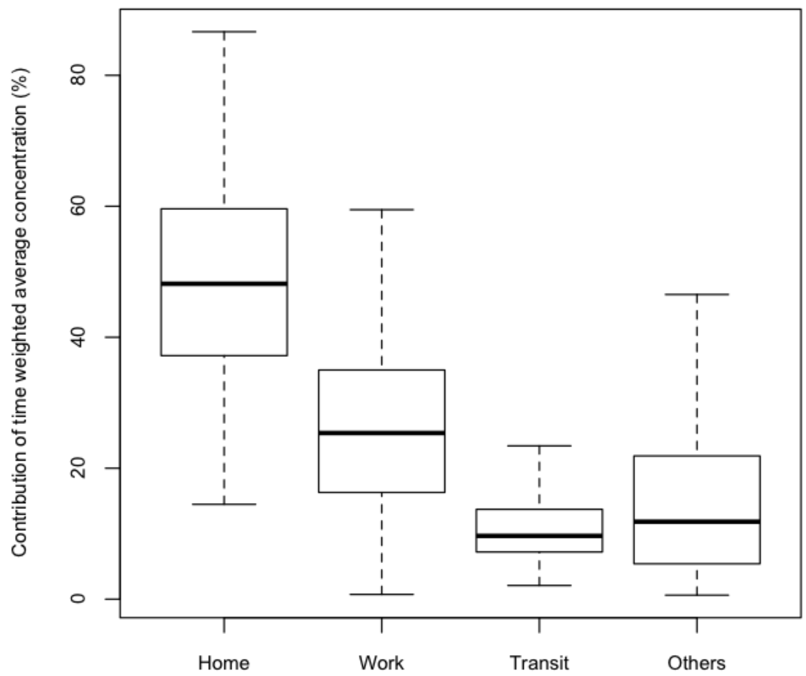

Figure 7. Time weighted average NO_2 concentration

When ventilation rates are considered, transit's share in total inhaled dose is also greater due to higher energy expenditure in transit microenvironments. These results are shown in Figure 8.

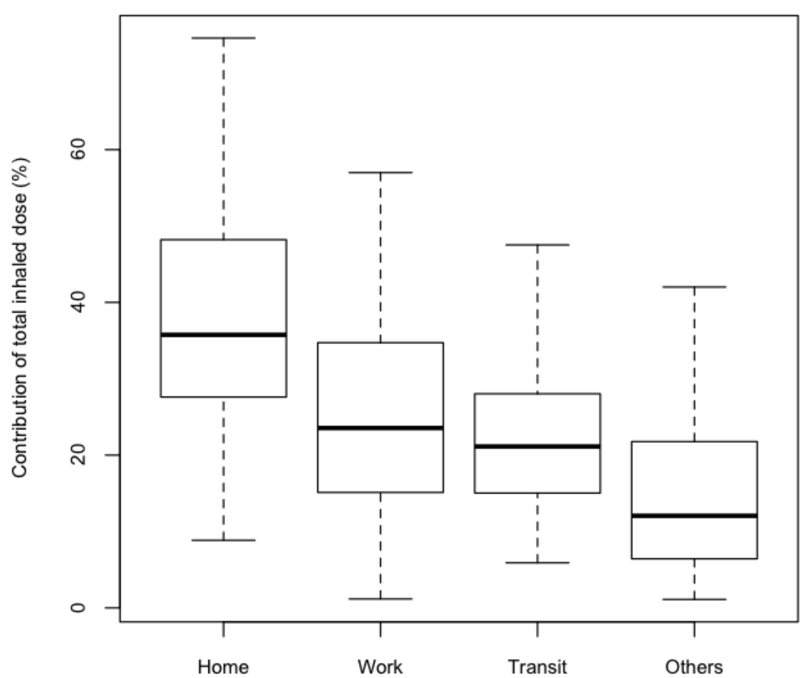

Figure 8. Percentage contribution to inhaled dose from different microenvironments

According to these results, when the energy expenditure is considered, transit's contribution is on average 22.5%. This is consistent with what was found in previous studies done by de Nazelle et al (2013) and Dons et al (2012).

Monte Carlo simulation for microenvironments ratios

From the previous results the influence of the ratios defined for the various microenvironments has been identified as a key parameter. In order to identify the range of probable outcomes, a Monte Carlo analysis has been developed. This consisted of substituting the parametric ratios for random variables from a uniform distribution defined by the maximum and minimum values in Table 1. The simulation was ran 93 times and the results are presented below.

The average of the concentration of NO_2 was computed for the entire population for the different microenvironments. With a fixed parameter this would give the results shown in Table 3. The results after the Monte Carlo simulation are a dispersion of possible results that are shown in Figure 9.

Table 3. NO_2 average concentrations with parametric ratios

Microenvironment	NO_2 Concentration (μgm^{-3})
Home	79.20
Work	119.22
Transit	223.02
Other	100.18

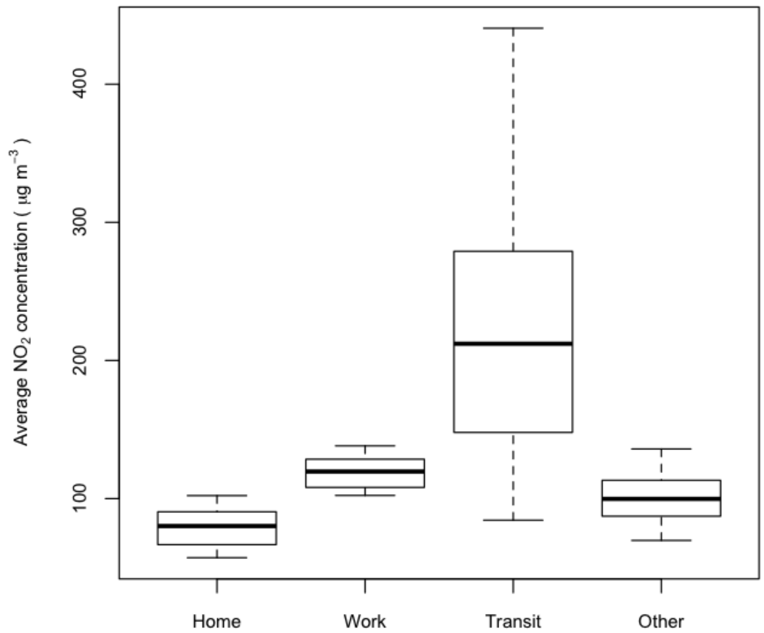

Figure 9. Average NO2 concentration after Monte Carlo simulation

Note that the microenvironment with the most uncertainty is Transit due to the brad range defined for the ratios in the different modes. The maximum average NO2 concentration found for this microenvironment was 462.0µg/m^3 and the minimum was 84.4µg/m^3. Efforts to reduce this uncertainty should be a clear next step in the study of exposure concentrations.

A similar set of results can be obtained for the contribution of the total inhaled dose. Those are shown in Table 4 for a typical result with average parametric values for the ratios and in Figure 10 for the Monte Carlo simulation.

Table 4. Average contribution to the total inhaled dose with parametric ratios

Microenvironment	Contibution to total inhaled dose
Home	32%
Work	22%
Transit	33%
Other	13%

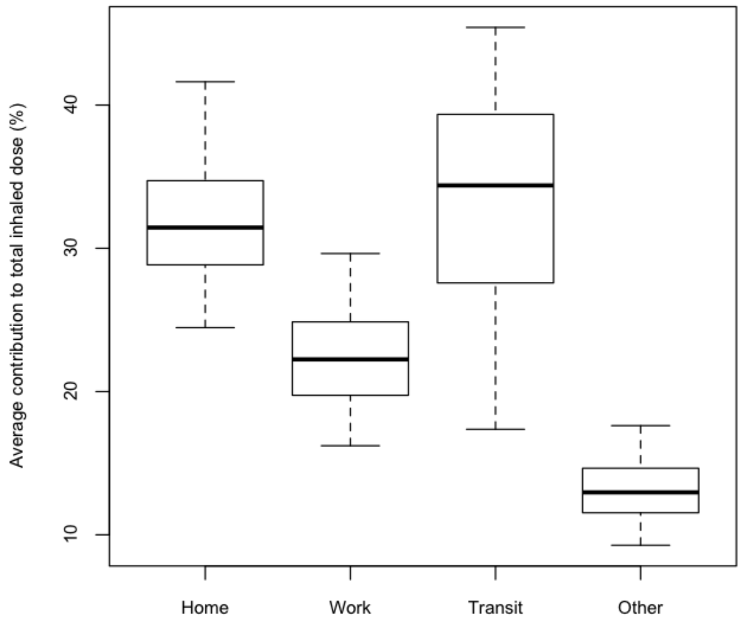

Figure 10. Average contribution to total inhaled dose after Monte Carlo simulation

The previous results indicate that transport's contribution to the total inhaled dose moves between 17% and 45%. The dispersion of these results is associated to the brad distribution used in the ratios for transport modes. Nevertheless the contribution from transportation to the total inhaled dose clearly plays a key role.

Analysis and discussion

The results show how the exposure and inhaled dose of the different individuals may vary and how the different corrections in space, time, microenvironment, and physical activity affect the various estimations. The activity profiles show how the individuals studied spend on average only 9% in transport but it is also one where the most physical activity is recorded, increasing the inhaled dose. This is aggravated by the fact that it is in transit that some of the grater concentrations are estimated, matched only by those estimated for work microenvironments.

When divided into the different modes the results don't show a clear pattern. The range of concentrations perceived in every mode is broad with interquartile ranges (IQR) of more than $50\mu g/m^{-3}$ in all cases. This is probably due to the movement of individuals all around the city. It is important to have in mind that this comparison is not for same routes so it does not imply that one mode or the other is associated to higher pollutant concentrations. Notice for example that the range of results for car passengers (IQR = $150\mu g/m^{-3}$) is greater than that for car drivers (IQR = $75\mu g/m^{-3}$) but if they were following the same route their result would be exactly the same (notice that the ratio used for both is the same).

The results presented show that for a 9% of average time spent in transit, its contribution to the total inhaled dose is of 22.5% for NO_2. This is consistent with other studies that follow similar methodologies like (de Nazelle, Seto, et al., 2013) that suggest a contribution of 24% in a population that spent on average 6% of their time in transit. Dons, Int Panis, Van Poppel, Theunis, & Wets, (2012) estimated the contribution to the total inhaled dose for black carbon (BC) by measuring with portable aethalometers and their results show that for a 6% average time spent in transport, it contributed with as much as 30% of the inhaled dose. Although none of the available models for Barcelona had BC, it was possible to estimate the total inhaled dose for $PM_{2.5}$ (using the LUR model) and PM_{10} (using the dispersion model), somehow associated to BC, and the results where 21.9% and 21.4% respectively. As expected, these figures are smaller than the Dons et al. results since BC is closely related to diesel fuels while PM has a greater variety of possible sources when not in transit.

According to the results, it is key to consider the ventilation rate when addressing issues related to exposure to air pollution. Estimating the total inhaled dose sheds light over the most important variables to be addressed that would not be easy to spot in an analysis based on time-weighted exposure concentrations or ambient air simulations. The ventilation rate results highlight the role of transport in physical activity and the ample spectrum of effort associated to transit from the various individuals. It is also interesting noting the higher average physical activity related to work and other activities than at home. This implies that exposure concentrations associated to home address maybe a good estimate based solely on the amount of time spent at home but misses out on important physical activity and exposure implications of other spaces.

The results highlight the relevance of the correction factors used in the simulation. Using only the annual average maps sub estimates the concentrations in every microenvironment category and presents smaller standard deviations, mainly due to the loss of extreme cases that are accounted for with this methodology. The Monte Carlo analysis performed to these ratios emphasise the importance of the ratios assigned to the various microenvironments and transport modes. Finding the most accurate ratio may be a difficult task and it is still missing more studies and data collection processes. However, the Monte Carlo analysis deals with this issue by changing the parametric ratios by uniform distributions and then giving an idea of the expected results. As seen from this analysis, ratios associated to transport modes seem to have a greater uncertainty and it is there where efforts should be focused in order to reduce the uncertainty in personal exposures to local pollutants. Although in average it is expected a mean concentration of $220 \mu g/m^{-3}$ for transit microenvironments, depending on the ratios chosen this could be as low as $100 \mu g/m^{-3}$ or as high as $400\ \mu g/m^{-3}$. This uncertainty results in a similar range for the contribution of transport in the total inhaled dose from 20% to 40%. Although it seems desirable to collect enough information that may reduce the range of possible ratios for transport, these results do highlight the importance of the exposure during transit as a mean contributor to the total inhaled dose.

Conclusions and future work

This study shows a novel approach to assess the exposure to local pollutants for an active population in Barcelona, Spain. The use of smartphones and the CalFit application allows having high-resolution information on position and physical activity in order to estimate exposure to local pollutants and total inhaled dose. Combining georeferenced data and activity diaries with air quality models and microenvironment ratios, exposures to four different pollutants were simulated for around 180 individuals that carried a smartphone for seven days. The method showed to be an efficient way to evaluate people's exposure and inhaled dose in a cost-effective way and applicable for large populations.

The results show high exposure concentrations in various microenvironments and the correction ratios show these to be much higher than estimated solely by city or regional models for Barcelona. The greater average exposure concentrations where found to be at work followed by transit but the greater share of time-weighted exposures was that from home mainly due to the total time spent there (more than 55% on average). Although people spent on average 9% in transport, this activity is responsible for more than 11% of the time-weighted exposures in all pollutants studied and more than 20% of the total inhaled dose.

The different results show the importance in the analysis of the ratios assigned to the different microenvironments and transport modes. A Monte Carlo simulation was performed in order to identify the critical parameters and the level of the uncertainty associated to the results. The ratios given to different transport modes where identified as the main source of uncertainty. Nevertheless, based on uniform distributions defined for these, it was possible to identify that transport plays a key role in the total inhaled dose that may vary between 20 and 40%. These levels are only mildly matched by the contribution of the home microenvironments.

The most physical activity was recorded during the transit microenvironments and this has important implications in terms of the total inhaled dose. In the search for more healthy livelihoods in European cities, transport should play a key role having the potential of promoting physical activity but with important implications in the inhaled dose of local pollutants. Future epidemiological studies need to consider transport in greater detail and the methodology followed in this study is promising for such evaluations, considering the differences in time, space, activity and levels of physical activity. It is also replicable with relative ease and cost-effectiveness in other cities and larger populations.

Future work

This study serves as evidence of the levels of air pollution to which city dwellers are exposed to on a daily bases in a European city and estimates inhaled doses considering concentrations and various microenvironments and physical activity levels. Nevertheless, in order to use this methodology as an aid in the policy making process there are still some important steps to be taken. Some of them are outlined below:

Data collection to improve the information on transport modes

As suggested by the results, transit plays a key role in daily exposures and inhaled doses in all pollutants studied. However the uncertainty in the ratios used is high and efforts to reduce it should be attempted. The literature in this is scarce and not always available for common and relevant air pollutants like NO_2 or $PM_{2.5}$. There are many variables that may influence these ratios and they have not been properly addressed in the existing literature. The effect of having the windows open or closed, proximity to other sources, age and maintenance of the vehicles are all factors known to influence the exposure levels in transport modes but there is no clear idea of how much can these influence the ratios used. Such issues must me address promptly.

Comparisons with measured exposure levels

This study is based on real-life daily routines and measured levels of physical activity using a smartphone with GPS and accelerometer. However, the level of compliance of these estimates with actual levels of exposure can be in no way verified. In future studies, it would be desirable to include the measurement of air pollutants concentrations using portable devices and be able to compare or eventually calibrate the simulations.

Replication in other cities and with larger populations

This study is in itself based on the results published by de Nazelle et al., (2013) for a smaller population in Barcelona. Doing this new exercise has proven to be useful and has helped in a deeper understanding of the problematic being addressed. In order to identify more strengths and weaknesses of this methodology, it should be applied to larger samples and to different contexts.

Additional analysis and minor corrections

The amount of information available in the resulting databases is still not completely explored. Deeper analysis on the different transport modes used, association with home and work address, impact of the time of year, scenario simulation using different modes and their impact in the total inhaled dose, are all additional analysis that can be done with this database and that would contribute highly to the understanding of the exposure to air pollutants in Barcelona.

Some minor corrections to the computation of results can still be performed. These include the equation for physical activity that has not been calibrated specifically for the phone used and thus the MET calculation may vary slightly; the dispersion model and the LUR model have higher resolutions than 100x100m but such versions could not be used due to the lack of precision in the available GPS information; some people reported using more than one mode but only the first was used due to a lack of information on for how long they took each and where did they change, but this could be solved indirectly using the information on the accelerometer or average speed. All these are suspected not to result in major changes in the information presented here but should be addressed.

References

Bigazzi, A. Y., & Figliozzi, M. A. (2012). Impacts of freeway traffic conditions on in-vehicle exposure to ultrafine particulate matter. *Atmospheric Environment*, *60*, 495–503. doi:10.1016/j.atmosenv.2012.07.020

Both, A. F., Westerdahl, D., Fruin, S., Haryanto, B., & Marshall, J. D. (2013). Exposure to carbon monoxide, fine particle mass, and ultrafine particle number in Jakarta, Indonesia: effect of commute mode. *The Science of the Total Environment*, *443*, 965–72. doi:10.1016/j.scitotenv.2012.10.082

de Nazelle, A., Aguilera, I., Nieuwenhuijsen, M., Beelen, R., Cirach, M., Hoek, G., … Basagaña, X. (2013). Comparison of performance of land use regression models derived for Catalunya, Spain. *Atmospheric Environment*, *77*, 598–606. doi:10.1016/j.atmosenv.2013.05.054

de Nazelle, A., Rodríguez, D. a, & Crawford-Brown, D. (2009). The built environment and health: impacts of pedestrian-friendly designs on air pollution exposure. *The Science of the Total Environment*, *407*(8), 2525–35. doi:10.1016/j.scitotenv.2009.01.006

de Nazelle, A., Seto, E., Donaire-Gonzalez, D., Mendez, M., Matamala, J., Nieuwenhuijsen, M. J., & Jerrett, M. (2013). Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environmental Pollution (Barking, Essex : 1987)*, *176*, 92–9. doi:10.1016/j.envpol.2012.12.032

Donaire-Gonzalez, D., de Nazelle, A., Seto, E., Mendez, M., Nieuwenhuijsen, M., Jerret, M. (2013). Comparison of physical activity measures using mobile phone-based CalFit and Actigraph. *Journal of Medical Internet Research, Res 2013;15(6):e111*. doi: 10.2196/jmir.2470

Dons, E., Int Panis, L., Van Poppel, M., Theunis, J., & Wets, G. (2012). Personal exposure to Black Carbon in transport microenvironments. *Atmospheric Environment*, *55*, 392–398. doi:10.1016/j.atmosenv.2012.03.020

Gulliver, J., & Briggs, D. (2011). STEMS-Air: a simple GIS-based air pollution dispersion model for city-wide exposure assessment. *The Science of the Total Environment*, *409*(12), 2419–29. doi:10.1016/j.scitotenv.2011.03.004

Int Panis, L., de Geus, B., Vandenbulcke, G., Willems, H., Degraeuwe, B., Bleux, N., … Meeusen, R. (2010). Exposure to particulate matter in traffic: A comparison of cyclists and car passengers. *Atmospheric Environment*, *44*(19), 2263–2270. doi:10.1016/j.atmosenv.2010.04.028

Johnson, T. (2002). A guide to selected algorithms, distribution and data bases used in exposure models developed by the Office of Air Quality Planning and Standards. *U.S. Environmental Protection Agency (North Carolina).* RTP

Lao, J., & Teixidó, O. (2011). Air quality model for Barcelona, *147*(2), 25–36. doi:10.2495/AIR110031

Minguillón, M. C., Schembari, A., Triguero-Mas, M., de Nazelle, A., Dadvand, P., Figueras, F., … Querol, X. (2012). Source apportionment of indoor, outdoor and personal PM2.5 exposure of pregnant women in Barcelona, Spain. *Atmospheric Environment*, *59*, 426–436. doi:10.1016/j.atmosenv.2012.04.052

Pecorari, E., Squizzato, S., Masiol, M., Radice, P., Pavoni, B., & Rampazzo, G. (2013). Using a photochemical model to assess the horizontal, vertical and time distribution of PM(2.5) in a complex area: relationships between the regional and local sources and the meteorological conditions. *The Science of the Total Environment*, *443*, 681–91. doi:10.1016/j.scitotenv.2012.11.047

Physick, W., Powell, J., Cope, M., Boast, K., & Lee, S. (2011). Measurements of personal exposure to NO2 and modelling using ambient concentrations and activity data. *Atmospheric Environment*, *45*(12), 2095–2102. doi:10.1016/j.atmosenv.2011.01.063

Steinle, S., Reis, S., & Sabel, C. E. (2013). Quantifying human exposure to air pollution--moving from static monitoring to spatio-temporally resolved personal exposure assessment. *The Science of the Total Environment*, *443*, 184–93. doi:10.1016/j.scitotenv.2012.10.098

Valero, N., Aguilera, I., Llop, S., Esplugues, A., de Nazelle, A., Ballester, F., & Sunyer, J. (2009). Concentrations and determinants of outdoor, indoor and personal nitrogen dioxide in pregnant women from two Spanish birth cohorts. *Environment International*, *35*(8), 1196–201. doi:10.1016/j.envint.2009.08.002

World Health Organization, 2104. Ambient (outdoor) air quality and health. *Media centre, Fact sheet Nº 313 (Updated March 2014)*. Available at: http://www.who.int/mediacentre/factsheets/fs313/en/ [Accessed May 8, 2014].

Appendix A

Code used in R for parametric analysis

```
rm(list=ls(all=TRUE))
### I load all the necessary files and some minor format
 adjustments###
load("/Users/JuanPablo/Desktop/Imperial/Thesis/Info BCN/Archivos
 Albert TD/travel.diary.minutes.modo.final.RData")
load("/Users/JuanPablo/Desktop/Imperial/Thesis/Info BCN/Archivos
 David Donire en R/Acelerometro/bd.mov.orjuela_140627.RData")
load("/Users/JuanPablo/Desktop/Imperial/Thesis/Info BCN/Archivos
 Albert TD/gps.trips.tapasinfo.RData")
Concentracion<-read.csv("~/Desktop/Imperial/Thesis/Info
 BCN/Archivos Antonia/exposures.csv",header=TRUE,sep=",")
Background<-read.csv("~/Desktop/Imperial/Thesis/Info BCN/Archivos
 Marta Cirach/PalauReial/BackgroundFINAL.csv",header=TRUE,sep=";")
Background$TIME <- as.POSIXlt(Background$TIME,"EST")
Background$shortdate<- strftime(Background$TIME,
 format="%Y/%m/%d %H")
ModeRatios<-read.csv("~/Desktop/Imperial/Thesis/R/Pruebas en
 R/ModeRatio.csv",header=TRUE,sep=";")
PlaceRatios<-read.csv("~/Desktop/Imperial/Thesis/R/Pruebas en
 R/PlaceRatio.csv",header=TRUE,sep=";")
options(scipen=999)
###Number of individuals an cycle counters###
i <- 181-8
n <- 1
j <- 1
x <- 1
k <- 1

#Have in mind that there are no results for individuals 9, 28, 60, 67,
 97, 158, 175 and 177#

###Create the principal vectors with the first individual###
ModoId1<-subset(tq.modo, id==1)
ConceId1<-subset(Concentracion, subj==1)
AceId01<-bd.mov[["id001"]]
InfoGPS <- bd.gps
names <- c("id","Gender","Age","new
 epoch","date_time","Place&Mode","MET","rawpm25LUR","rawno2LUR","
 rawpm10dir","rawno2dir","rawnoxdir","background","ResPM25_LUR","V
 entPM25_LUR","ResNO2_LUR","VentNO2_LUR","ResPM10_DIS","VentP
```

```
M10_DIS","ResNO2_DIS","VentNO2_DIS","ResNOX_DIS","VentNOX_DI
S")

###For storing the results###
ListResults <- rep( list(list()), i )

###Start of the loop for all the individuals###
for (n in 1:i) {

 if (n>=9 && n<27) {
   j <- n+1
   } else if (n>=27 && n<58){
   j <- n+2
   }else if (n>=58 && n<64){
   j <- n+3
   }else if (n>=64 && n<93){
   j <- n+4
   } else if (n>=93 && n<153) {
   j <- n+5
   } else if (n>=153 && n<169) {
   j <- n+6
   } else if (n>=169 && n<170) {
   j <- n+7
   } else if (n>=170) {
   j <- n+8
   } else {
      j <- n
   }

 ##Prepare the first mother files of mode, modelling results and
 accelerometer##
 ModoId1<-subset(tq.modo, id==j)
 ConceId1<-subset(Concentracion, subj==j)
 ConceId1$pm25l<- as.numeric(as.character(ConceId1$pm25l))
 ConceId1$no2l<- as.numeric(as.character(ConceId1$no2l))
 ConceId1$pm10br<- as.numeric(as.character(ConceId1$pm10br))
 ConceId1$no2br<- as.numeric(as.character(ConceId1$no2br))
 ConceId1$noxbr<- as.numeric(as.character(ConceId1$noxbr))

 colnames(ModoId1) <- c("id",
   "epochmg","seq.min","where","Mode1","Mode2","Mode3","Mode4")
 if (n>=64 && n< 93){
   l <- n+1
   } else if (n>=93 && n<153){
```

```
  l <- n+2
  } else if (n>=153 && n<169) {
  l <- n+3
  } else if (n>=169 && n<170) {
  l <- n+4
  } else if (n>=170) {
  l <- n+5
  } else{
  l <- n}

AceId01<-bd.mov[[l]]
AceId01["new epoch"]<-AceId01["epoch"]/1000
AceId01["new epoch"]<-AceId01["new epoch"]/60
ConceId1["new epoch"]<-ConceId1["epochmg"]/60
ModoId1["new epoch"] <- ModoId1["epochmg"]/60
ModoId1["new epoch"] <- round(ModoId1["new epoch"])
ConceId1["new epoch"] <- round(ConceId1["new epoch"])
AceId01["new epoch"] <- round(AceId01["new epoch"])
ModoId1$shortdate<- strftime(ModoId1$seq.min, format="%Y/%m/%d
%H")
##asign the correction factors for background, mode and place##
ModoId1$BackgroundNO2 <-
Background$NO2RATIO[match(ModoId1$shortdate,Background$shortd
ate,1)]
ModoId1$PlaceRatio <-
PlaceRatios$Ratio[match(ModoId1$where,PlaceRatios$PlaceNumber,1)]
ModoId1$ModeRatio1 <-
ModeRatios$Ratio[match(ModoId1$Mode1,ModeRatios$ModeNumber,1)
]
ModoId1$ModeRatio2 <-
ModeRatios$Ratio[match(ModoId1$Mode2,ModeRatios$ModeNumber,1)
]
ModoId1$ModeRatio3 <-
ModeRatios$Ratio[match(ModoId1$Mode3,ModeRatios$ModeNumber,1)
]
ModoId1$ModeRatio4 <-
ModeRatios$Ratio[match(ModoId1$Mode4,ModeRatios$ModeNumber,1)
]
ModoId1$ModeRatio <-
apply(data.frame(ModoId1$ModeRatio1,ModoId1$ModeRatio2,ModoId1
$ModeRatio3,ModoId1$ModeRatio4),1,prod,na.rm=TRUE)
##Calculating METs and Ventilation rates##
InfoMets <- aggregate(AceId01$V,list(AceId01$"new
epoch"),FUN=sum,na.rm=TRUE)
```

```
#Donaire et al#
InfoMets$MET <- 1.2907087+(0.4141791*InfoMets$x)
BM <- InfoGPS$peso[match(ModoId1$id,InfoGPS$ID_CAVA,1)][1]
AGE <- InfoGPS$edad[match(ModoId1$id,InfoGPS$ID_CAVA,1)][1]
GENDER <-
InfoGPS$genero[match(ModoId1$id,InfoGPS$ID_CAVA,1)][1]
if (is.na(GENDER)) {GENDER <- "Mujer"}
if (is.na(AGE)) {AGE <- 30}

#Johnsson2002#
if (GENDER =="Mujer" && AGE < 18) {
  a <- 2.898
  b <- 0.056
  sigmae <- 0.47
  } else if (GENDER =="Mujer" && 18<=AGE && AGE<30) {
  a <- 2.036
  b <- 0.062
  sigmae <- 0.50
  }  else if (GENDER=="Mujer" && 30<=AGE && AGE<60) {
  a <- 3.538
  b <- 0.062
  sigmae <- 0.47
  }  else if (GENDER=="Mujer") {
  a <- 2.755
  b <- 0.038
  sigmae <- 0.45
  }  else if (AGE < 18) {
  a <- 2.754
  b <- 0.074
  sigmae <- 0.44
  }  else if (18<=AGE && AGE<30) {
  a <- 2.896
  b <- 0.063
  sigmae <- 0.44
  }  else if (30<=AGE && AGE<60) {
  a <- 3.653
  b <- 0.048
  sigmae <- 0.70
  }  else {
  a <- 2.459
  b <- 0.049
  sigmae <- 0.69}
InfoMets$RMR <-
a+(b*BM)+rnorm(dim(InfoMets)[1],mean=0,sd=sigmae)
```

```
InfoMets$VO2_BM <- 0.198*InfoMets$MET*InfoMets$RMR/BM
if (GENDER =="Mujer" && AGE < 45) {
  a <- 4.357
  b <- 1.276
  sigmad <- 0.1351
  sigmae <- 0.1182
  }  else if (GENDER =="Mujer" && 45<=AGE && AGE<65) {
  a <- 3.454
  b <- 1.021
  sigmad <- 0.1106
  sigmae <- 0.0769
  }  else if (GENDER=="Mujer") {
  a <- 2.956
  b <- 0.908
  sigmad <- 0.0886
  sigmae <- 0.0338
  }  else if (AGE < 45) {
  a <- 3.991
  b <- 1.197
  sigmad <- 0.1228
  sigmae <- 0.1395
  }  else if (45<=AGE && AGE<65) {
  a <- 4.018
  b <- 1.165
  sigmad <- 0.1107
  sigmae <- 0.1112
  } else {
  a <- 3.730
  b <- 1.071
  sigmad <- 0.1082
  sigmae <- 0.0632
  }
InfoMets$LnVe <-
a+b*log(InfoMets$VO2_BM)+rnorm(dim(InfoMets)[1],mean=0,sd=sig
mad)+rnorm(dim(InfoMets)[1],mean=0,sd=sigmae)
InfoMets$Ve <- exp(InfoMets$LnVe)
colnames(InfoMets) <- c("new
epoch","V","MET","RMR","VO2_BM","LnVe","Ve")
ModoId1$Ventilation <- InfoMets$Ve[match(ModoId1$"new
epoch",InfoMets$"new epoch",10)]
ModoId1$MET <- InfoMets$MET[match(ModoId1$"new
epoch",InfoMets$"new epoch",10)]
ModoId1$Gender <-
InfoGPS$genero[match(ModoId1$id,InfoGPS$ID_CAVA,1)][1]
```

```
ModoId1$Age <-
InfoGPS$edad[match(ModoId1$id,InfoGPS$ID_CAVA,1)][1]
##Adding the model concentrations to the mode file##
Concentspm25l <- aggregate(ConceId1$pm25l,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)
Concentsno2l <- aggregate(ConceId1$no2l,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)
Concentspm10br <- aggregate(ConceId1$pm10br,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)
Concentsno2br <- aggregate(ConceId1$no2br,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)
Concentsnoxbr <- aggregate(ConceId1$noxbr,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)
colnames(Concentspm25l) <- c("new epoch","pm25l")
colnames(Concentsno2l) <- c("new epoch","no2l")
colnames(Concentspm10br) <- c("new epoch","pm10br")
colnames(Concentsno2br) <- c("new epoch","no2br")
colnames(Concentsnoxbr) <- c("new epoch","noxbr")
ModoId1$pm25l <- Concentspm25l$pm25l[match(ModoId1$"new
epoch",Concentspm25l$"new epoch",10)]
ModoId1$no2l <- Concentsno2l$no2l[match(ModoId1$"new
epoch",Concentsno2l$"new epoch",10)]
ModoId1$pm10br <- Concentspm10br$pm10br[match(ModoId1$"new
epoch",Concentspm10br$"new epoch",10)]
ModoId1$no2br <- Concentsno2br$no2br[match(ModoId1$"new
epoch",Concentsno2br$"new epoch",10)]
ModoId1$noxbr <- Concentsnoxbr$noxbr[match(ModoId1$"new
epoch",Concentsnoxbr$"new epoch",10)]

##Corrections for concentrations##
x <- 2
k <- dim(ModoId1)[1]
if (ModoId1$pm25l[1]>500 || ModoId1$pm25l[1]=="NaN") {
     ModoId1$pm25l[1] <- 0}

if (ModoId1$no2l[1]>500  || ModoId1$no2l[1]=="NaN") {
     ModoId1$no2l[1] <- 0}

if (ModoId1$pm10br[1]>500 || ModoId1$pm10br[1]=="NaN") {
     ModoId1$pm10br[1] <- 0}

if (ModoId1$no2br[1]>500 || ModoId1$no2br[1]=="NaN") {
     ModoId1$no2br[1] <- 0}
```

```
if (ModoId1$noxbr[1]>500 || ModoId1$no2br[1]=="NaN") {
     ModoId1$noxbr[1] <- 0}

for (x in 2:k) {

  if (ModoId1$pm25l[x]>500 || ModoId1$pm25l[x]=="NaN") {
     ModoId1$pm25l[x] <- ModoId1$pm25l[x-1]}

  if (ModoId1$no2l[x]>500 || ModoId1$no2l[x]=="NaN") {
     ModoId1$no2l[x] <- ModoId1$no2l[x-1]}

  if (ModoId1$pm10br[x]>500 || ModoId1$pm10br[x]=="NaN") {
     ModoId1$pm10br[x] <- ModoId1$pm10br[x-1]}

  if (ModoId1$no2br[x]>500 || ModoId1$no2br[x]=="NaN") {
     ModoId1$no2br[x] <- ModoId1$no2br[x-1]}

  if (ModoId1$noxbr[x]>500 || ModoId1$noxbr[x]=="NaN") {
     ModoId1$noxbr[x] <- ModoId1$noxbr[x-1]}
  }
##Final multiplication and results for each individual##
ModoId1$ResPM25_LUR <-
ModoId1$pm25l*ModoId1$Background*ModoId1$ModeRatio*ModoId1$
PlaceRatio
ModoId1$VentPM25_LUR <-
ModoId1$ResPM25_LUR*ModoId1$Ventilation

ModoId1$ResNO2_LUR <-
ModoId1$no2l*ModoId1$Background*ModoId1$ModeRatio*ModoId1$Pl
aceRatio
ModoId1$VentNO2_LUR <-
ModoId1$ResNO2_LUR*ModoId1$Ventilation

ModoId1$ResPM10_DIS <-
ModoId1$pm10br*ModoId1$Background*ModoId1$ModeRatio*ModoId1
$PlaceRatio
ModoId1$VentPM10_DIS <-
ModoId1$ResPM10_DIS*ModoId1$Ventilation

ModoId1$ResNO2_DIS <-
ModoId1$no2br*ModoId1$Background*ModoId1$ModeRatio*ModoId1$
PlaceRatio
ModoId1$VentNO2_DIS <- ModoId1$ResNO2_DIS*ModoId1$Ventilation
```

```
ModoId1$ResNOX_DIS <-
ModoId1$noxbr*ModoId1$Background*ModoId1$ModeRatio*ModoId1$
PlaceRatio
ModoId1$VentNOX_DIS <-
ModoId1$ResNOX_DIS*ModoId1$Ventilation

##Store a list of results##
ModoId1$Mode1 <- ModoId1$Mode1+10
ModoId1$Mode2 <- ModoId1$Mode2+20
ModoId1$Mode3 <- ModoId1$Mode3+30
ModoId1$Mode4 <- ModoId1$Mode4+40
ModoId1$ModeTOTAL <-
apply(data.frame(ModoId1$where,ModoId1$Mode1,ModoId1$Mode2,Mo
doId1$Mode3,ModoId1$Mode4),1,min,na.rm=TRUE)
IndiResults <-
data.frame(ModoId1$id,ModoId1$Gender,ModoId1$Age,ModoId1$"new
epoch",ModoId1$seq.min,ModoId1$ModeTOTAL,ModoId1$MET,ModoId1
$pm25l,ModoId1$no2l,ModoId1$pm10br,ModoId1$no2br,ModoId1$nox
br,ModoId1$BackgroundNO2,ModoId1$ResPM25_LUR,ModoId1$VentPM
25_LUR,ModoId1$ResNO2_LUR,ModoId1$VentNO2_LUR,ModoId1$ResP
M10_DIS,ModoId1$VentPM10_DIS,ModoId1$ResNO2_DIS,ModoId1$Ve
ntNO2_DIS,ModoId1$ResNOX_DIS,ModoId1$VentNOX_DIS)
colnames(IndiResults) <- names
ListResults[[n]] <- IndiResults
}
save(ListResults, file =
"/Users/JuanPablo/Desktop/Imperial/Thesis/R/versiones
codigo/ListResults.RData")
```

Code used in R for Monte Carlo simulations

```
### I load all the necessary files and some minor format
 adjustments###
###Number of individuals an cycle counters###
i <- 181-8
n <- 1
j <- 1
x <- 1
k <- 1
mc <- 100

#Have in mind that there are no results for individuals 9,
 28, 60, 67, 97, 158, 175 and 177#

###Create the principal vectors with the first individual###
ModoId2<-subset(tq.modo, id==1)
ConceId2<-subset(Concentracion, subj==1)
AceId02<-bd.mov[["id001"]]
InfoGPS <- bd.gps
names <- c("id","new
 epoch","date_time","Place&Mode","ResNO2_LUR","VentNO2_LUR","R
 esNO2_DIS","VentNO2_DIS")

###For storing the results###
ListResultsMC <- rep( list(list()), i )
myListResultsMC <- rep (list(list()),mc)
####For MonteCarlo####

m <- 1
RatioHome <- runif(200,min=1,max=1.8)
RatioWork <- runif(200,min=1.9,max=2.7)
RatioOutdoor <- runif(200,min=1,max=3.6)
RatioWalk <- runif(200,min=0.8,max=2.4)
RatioBike <- runif(200,min=0.4,max=12.2)
RatioCar <- runif(200,min=1.4,max=3.9)
RatioBus <- runif(200,min=0.9,max=3.3)

for (m in 1:mc) {

###Start of the loop for all the individuals###
for (n in 1:i) {
```

```
if (n>=9 && n<27) {
    j <- n+1
    } else if (n>=27 && n<58){
    j <- n+2
    }else if (n>=58 && n<64){
    j <- n+3
    }else if (n>=64 && n<93){
    j <- n+4
    } else if (n>=93 && n<153) {
    j <- n+5
    } else if (n>=153 && n<169) {
    j <- n+6
    } else if (n>=169 && n<170) {
    j <- n+7
    } else if (n>=170) {
    j <- n+8
    } else {
        j <- n
    }

##Prepare the first mother files of mode, modelling results
and accelerometer##
ModoId2<-subset(tq.modo, id==j)
ConceId2<-subset(Concentracion, subj==j)
#ConceId1$pm25l<- as.numeric(as.character(ConceId1$pm25l))#
ConceId2$no2l<- as.numeric(as.character(ConceId2$no2l))
#ConceId1$pm10br<- as.numeric(as.character(ConceId1$pm10br))#
ConceId2$no2br<- as.numeric(as.character(ConceId2$no2br))
#ConceId1$noxbr<- as.numeric(as.character(ConceId1$noxbr))#

colnames(ModoId2) <- c("id",
    "epochmg","seq.min","where","Mode1","Mode2","Mode3","Mode
4")
if (n>=64 && n< 93){
    l <- n+1
    } else if (n>=93 && n<153){
    l <- n+2
    } else if (n>=153 && n<169) {
    l <- n+3
    } else if (n>=169 && n<170) {
    l <- n+4
```

```
    } else if (n>=170) {
    l <- n+5
    } else{
    l <- n}

AceId02<-bd.mov[[l]]
AceId02["new epoch"]<-AceId02["epoch"]/1000
AceId02["new epoch"]<-AceId02["new epoch"]/60
ConceId2["new epoch"]<-ConceId2["epochmg"]/60
ModoId2["new epoch"] <- ModoId2["epochmg"]/60
ModoId2["new epoch"] <- round(ModoId2["new epoch"])
ConceId2["new epoch"] <- round(ConceId2["new epoch"])
AceId02["new epoch"] <- round(AceId02["new epoch"])
ModoId2$shortdate<- strftime(ModoId2$seq.min,
format="%Y/%m/%d %H")

##asign the correction factors for background, mode and
place##
ModoId2$BackgroundNO2 <-
Background$NO2RATIO[match(ModoId2$shortdate,Background$shortd
ate,1)]
##The core of the Monte Carlo##
PlaceRatios$Ratio[2] <- RatioHome[m]
PlaceRatios$Ratio[3] <- RatioWork[m]
PlaceRatios$Ratio[4] <- RatioOutdoor[m]
PlaceRatios$Ratio[5] <- RatioWork[m]
PlaceRatios$Ratio[6] <- RatioOutdoor[m]

ModeRatios$Ratio[2] <- RatioWalk[m]
ModeRatios$Ratio[3] <- RatioBike[m]
ModeRatios$Ratio[4] <- RatioBike[m]
ModeRatios$Ratio[5] <- RatioCar[m]
ModeRatios$Ratio[6] <- RatioCar[m]
ModeRatios$Ratio[7] <- RatioBus[m]
ModeRatios$Ratio[8] <- RatioBus[m]


ModoId2$PlaceRatio <-
PlaceRatios$Ratio[match(ModoId2$where,PlaceRatios$PlaceNumber
,1)]
ModoId2$ModeRatio1 <-
ModeRatios$Ratio[match(ModoId2$Mode1,ModeRatios$ModeNumber,1)
```

```
]
ModoId2$ModeRatio2 <-
ModeRatios$Ratio[match(ModoId2$Mode2,ModeRatios$ModeNumber,1)
]
ModoId2$ModeRatio3 <-
ModeRatios$Ratio[match(ModoId2$Mode3,ModeRatios$ModeNumber,1)
]
ModoId2$ModeRatio4 <-
ModeRatios$Ratio[match(ModoId2$Mode4,ModeRatios$ModeNumber,1)
]
ModoId2$ModeRatio <-
apply(data.frame(ModoId2$ModeRatio1,ModoId2$ModeRatio2,ModoId
2$ModeRatio3,ModoId2$ModeRatio4),1,prod,na.rm=TRUE)
##Calculating METs and Ventilation rates##
InfoMets <- aggregate(AceId02$V,list(AceId02$"new
epoch"),FUN=sum,na.rm=TRUE)
#Donaire et al#
InfoMets$MET <- 1.2907087+(0.4141791*InfoMets$x)
BM <- InfoGPS$peso[match(ModoId2$id,InfoGPS$ID_CAVA,1)][1]
AGE <- InfoGPS$edad[match(ModoId2$id,InfoGPS$ID_CAVA,1)][1]
GENDER <-
InfoGPS$genero[match(ModoId2$id,InfoGPS$ID_CAVA,1)][1]
if (is.na(GENDER)) {GENDER <- "Mujer"}
if (is.na(AGE)) {AGE <- 30}

#Johnsson2002#
if (GENDER =="Mujer" && AGE < 18) {
    a <- 2.898
    b <- 0.056
    sigmae <- 0.47
    } else if (GENDER =="Mujer" && 18<=AGE && AGE<30) {
    a <- 2.036
    b <- 0.062
    sigmae <- 0.50
    }   else if (GENDER=="Mujer" && 30<=AGE && AGE<60) {
    a <- 3.538
    b <- 0.062
    sigmae <- 0.47
    }   else if (GENDER=="Mujer") {
    a <- 2.755
    b <- 0.038
    sigmae <- 0.45
```

```
    }   else if (AGE < 18) {
    a <- 2.754
    b <- 0.074
    sigmae <- 0.44
    }   else if (18<=AGE && AGE<30) {
    a <- 2.896
    b <- 0.063
    sigmae <- 0.44
    }   else if (30<=AGE && AGE<60) {
    a <- 3.653
    b <- 0.048
    sigmae <- 0.70
    }   else {
    a <- 2.459
    b <- 0.049
    sigmae <- 0.69}
InfoMets$RMR <-
a+(b*BM)+rnorm(dim(InfoMets)[1],mean=0,sd=sigmae)
InfoMets$VO2_BM <- 0.198*InfoMets$MET*InfoMets$RMR/BM
if (GENDER =="Mujer" && AGE < 45) {
    a <- 4.357
    b <- 1.276
    sigmad <- 0.1351
    sigmae <- 0.1182
    }   else if (GENDER =="Mujer" && 45<=AGE && AGE<65) {
    a <- 3.454
    b <- 1.021
    sigmad <- 0.1106
    sigmae <- 0.0769
    }   else if (GENDER=="Mujer") {
    a <- 2.956
    b <- 0.908
    sigmad <- 0.0886
    sigmae <- 0.0338
    }   else if (AGE < 45) {
    a <- 3.991
    b <- 1.197
    sigmad <- 0.1228
    sigmae <- 0.1395
    }   else if (45<=AGE && AGE<65) {
    a <- 4.018
    b <- 1.165
```

```
    sigmad <- 0.1107
    sigmae <- 0.1112
    } else {
    a <- 3.730
    b <- 1.071
    sigmad <- 0.1082
    sigmae <- 0.0632
    }
InfoMets$LnVe <-
a+b*log(InfoMets$VO2_BM)+rnorm(dim(InfoMets)[1],mean=0,sd=sig
mad)+rnorm(dim(InfoMets)[1],mean=0,sd=sigmae)
InfoMets$Ve <- exp(InfoMets$LnVe)
colnames(InfoMets) <- c("new
epoch","V","MET","RMR","VO2_BM","LnVe","Ve")
ModoId2$Ventilation <- InfoMets$Ve[match(ModoId2$"new
epoch",InfoMets$"new epoch",10)]
##Adding the model concentrations to the mode file##
#Concentspm25l <- aggregate(ConceId1$pm25l,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)#
Concentsno2l <- aggregate(ConceId2$no2l,list(ConceId2$"new
epoch"),FUN=mean,na.rm=TRUE)
#Concentspm10br <-
aggregate(ConceId1$pm10br,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)#
Concentsno2br <- aggregate(ConceId2$no2br,list(ConceId2$"new
epoch"),FUN=mean,na.rm=TRUE)
#Concentsnoxbr <- aggregate(ConceId1$noxbr,list(ConceId1$"new
epoch"),FUN=mean,na.rm=TRUE)#
#colnames(Concentspm25l) <- c("new epoch","pm25l")#
colnames(Concentsno2l) <- c("new epoch","no2l")
#colnames(Concentspm10br) <- c("new epoch","pm10br")#
colnames(Concentsno2br) <- c("new epoch","no2br")
#colnames(Concentsnoxbr) <- c("new epoch","noxbr")#
#ModoId2$pm25l <- Concentspm25l$pm25l[match(ModoId1$"new
epoch",Concentspm25l$"new epoch",10)]#
ModoId2$no2l <- Concentsno2l$no2l[match(ModoId2$"new
epoch",Concentsno2l$"new epoch",10)]
#ModoId1$pm10br <- Concentspm10br$pm10br[match(ModoId1$"new
epoch",Concentspm10br$"new epoch",10)]#
ModoId2$no2br <- Concentsno2br$no2br[match(ModoId2$"new
epoch",Concentsno2br$"new epoch",10)]
#ModoId1$noxbr <- Concentsnoxbr$noxbr[match(ModoId1$"new
```

```
epoch",Concentsnoxbr$"new epoch",10)]#

##Corrections for concentrations##
x <- 2
k <- dim(ModoId2)[1]
#if (ModoId1$pm25l[1]>500 || ModoId1$pm25l[1]=="NaN") {#
       #ModoId1$pm25l[1] <- 0}#

if (ModoId2$no2l[1]>500  || ModoId2$no2l[1]=="NaN") {
       ModoId2$no2l[1] <- 0}

#if (ModoId1$pm10br[1]>500 || ModoId1$pm10br[1]=="NaN") {#
       #ModoId1$pm10br[1] <- 0}#

if (ModoId2$no2br[1]>500 || ModoId2$no2br[1]=="NaN") {
       ModoId2$no2br[1] <- 0}

#if (ModoId1$noxbr[1]>500 || ModoId1$no2br[1]=="NaN") {#
       #ModoId1$noxbr[1] <- 0}#

for (x in 2:k) {

   #if (ModoId1$pm25l[x]>500 || ModoId1$pm25l[x]=="NaN") {#
       #ModoId1$pm25l[x] <- ModoId1$pm25l[x-1]}#

   if (ModoId2$no2l[x]>500 || ModoId2$no2l[x]=="NaN") {
       ModoId2$no2l[x] <- ModoId2$no2l[x-1]}

   #if (ModoId1$pm10br[x]>500 || ModoId1$pm10br[x]=="NaN")
{#
       #ModoId1$pm10br[x] <- ModoId1$pm10br[x-1]}#

   if (ModoId2$no2br[x]>500 || ModoId2$no2br[x]=="NaN") {
       ModoId2$no2br[x] <- ModoId2$no2br[x-1]}

   #if (ModoId1$noxbr[x]>500 || ModoId1$noxbr[x]=="NaN") {#
       #ModoId1$noxbr[x] <- ModoId1$noxbr[x-1]}#
   }
##Final multiplication and results for each individual##
#ModoId1$ResPM25_LUR <-
ModoId1$pm25l*ModoId1$Background*ModoId1$ModeRatio*ModoId1$Pl
aceRatio#
```

```
#ModoId1$VentPM25_LUR <-
ModoId1$ResPM25_LUR*ModoId1$Ventilation#

ModoId2$ResNO2_LUR <-
ModoId2$no2l*ModoId2$Background*ModoId2$ModeRatio*ModoId2$Pla
ceRatio
ModoId2$VentNO2_LUR <- ModoId2$ResNO2_LUR*ModoId2$Ventilation

#ModoId1$ResPM10_DIS <-
ModoId1$pm10br*ModoId1$Background*ModoId1$ModeRatio*ModoId1$P
laceRatio#
#ModoId1$VentPM10_DIS <-
ModoId1$ResPM10_DIS*ModoId1$Ventilation#

ModoId2$ResNO2_DIS <-
ModoId2$no2br*ModoId2$Background*ModoId2$ModeRatio*ModoId2$Pl
aceRatio
ModoId2$VentNO2_DIS <- ModoId2$ResNO2_DIS*ModoId2$Ventilation

#ModoId1$ResNOX_DIS <-
ModoId1$noxbr*ModoId1$Background*ModoId1$ModeRatio*ModoId1$Pl
aceRatio#
#ModoId1$VentNOX_DIS <-
ModoId1$ResNOX_DIS*ModoId1$Ventilation#

##Store a list of results##
ModoId2$Mode1 <- ModoId2$Mode1+10
ModoId2$Mode2 <- ModoId2$Mode2+20
ModoId2$Mode3 <- ModoId2$Mode3+30
ModoId2$Mode4 <- ModoId2$Mode4+40
ModoId2$ModeTOTAL <-
apply(data.frame(ModoId2$where,ModoId2$Mode1,ModoId2$Mode2,Mo
doId2$Mode3,ModoId2$Mode4),1,min,na.rm=TRUE)
IndiResults <- data.frame(ModoId2$id,ModoId2$"new
epoch",ModoId2$seq.min,ModoId2$ModeTOTAL,ModoId2$ResNO2_LUR,M
odoId2$VentNO2_LUR,ModoId2$ResNO2_DIS,ModoId2$VentNO2_DIS)
colnames(IndiResults) <- names
ListResultsMC[[n]] <- IndiResults
}

myListResultsMC[[m]] <- ListResultsMC
```

```
}
     save(myListResultsMC, file =
"/Users/JuanPablo/Desktop/Imperial/Thesis/R/versiones
codigo/myListResultsMC.RData")
```

Appendix B

Tables with parameter values for equations 3 and 4

Characteristics of the individual		Parameters			
Age group	Gender	Intercept (a)	Slope (b)	Person-level errors (σ_d)	Test-level errors (σ_e)
18-44	Female	4.357	1.276	0.1351	0.1182
	Male	3.991	1,197	0.1228	0.1395
45-64	Female	3.454	1.021	0.1106	0.0769
	Male	4.018	1.165	0.1107	0.1112
65+	Female	2.956	0.908	0.0886	0.0338
	Male	3.730	1.071	0.1082	0.0632

Adapted from Johnson (2002). Table 9-1, pp. 9-3

d(k) is a random variable selected for each person from a normal distribution with mean zero and standard deviation equal to σ_d
e(i,j,k) is a random variable selected for each event from a normal distribution with mean zero and standard deviation equal to σ_e

The recommended value for ECF is established as **0.198** for all activities in this study according to Table 9-10 in Johnson (2002)

According to Johnson, RMR may be assumed to be equal to BMR that is given by the following equation:

$BMR(MJ/day) = a_1 + b_1 BM + e_1$

Where e1 is a random variable from a normal distribution mean zero and standard deviation equal to σ_{e1}

The following table gives the parameters for this regression:

		Regression coefficients		
Gender	Age, years	a_1	b_1	σ_{e1}
Female	18-29.9	2.036	0.062	0.50
Female	30-59.9	3.538	0.034	0.47
Female	>= 60	2.755	0.038	0.45
Male	18-29.9	2.896	0.063	0.64
Male	30-59.9	3.653	0.048	0.70
Male	>= 60	2.459	0.049	0.69

Adapted from Johnson (2002). Table 9-11, pp. 9-18