Mitigating the Health Effects of Desert Dust Storms Using Exposure-Reduction Approaches (MEDEA) project

Assessment of health outcomes in children with asthma (ACTION C.7) and in adults with atrial fibrillation (ACTION C.8) during DDS events (with vs without intervention measures)

Public Health Intervention Study

Statistical Analysis Plan

(April 2018)

1. Sample Size Determination and Power Calculations

Sample size calculations for the panel study of asthmatic children

Sample size calculations are based on expected changes in asthma symptoms measured by asthma control test (ACT), which is our primary health outcome. Childhood ACT has seven items and provides a score ranging from 0 (poorest asthma control) to 27 (optimal asthma control). A cut-off point of 19 indicates uncontrolled asthma. Previous studies have shown that a minimally meaningful change in the ACT score is ranging from 2 to 3 points (Voorend-van Bergen et al, 2014). To detect a statistically significant difference of 3 points, the minimum sample size needed in each of the 3 groups is 100 participants. This sample size calculation is assuming a level of 0.05 and a power of at least 80% to detect this difference between the comparison groups.

Sample size calculations for the panel study of atrial fibrillation (AF) patients

There are no studies that evaluated the reduction in AF burden attributed to reduction in exposure to ambient air pollution.
We assume that the mean AF burden in the control group will be 15%+/-15%. We conservatively estimated the expected intervention effect to be 20% of relative reduction in AF burden. This is translated into the reduction of AF burden by mean of 43 minutes per 24 hours. Group sample sizes of 108 and 216 achieve 81% power to show a difference in means when there is a difference of -3.0 between the null hypothesis mean difference of 0.0 and the actual mean difference of 3.0 at the 0.05 significance level (alpha) using a two-sided Mann-Whitney-Wilcoxon Test. These results are based on 2000 Monte Carlo samples from the null distributions: WeibullMS(15 15) and WeibullMS(15 15), and the alternative distributions: WeibullMS(15 15) and WeibullMS(12 15).

In our power calculations for the panel studies in asthmatic children and AF patients, the sample size included a dropout rate of 30%, which is an overestimation of the dropout rate observed in previous studies in Cyprus and countries of the region (15-20%). Thus, even if DDS events are 20-30% less we will have ample statistical power to detect a significant effect of each intervention.

2. Data Analyses

Analyses of health outcomes in asthmatic children

Our main outcome for the asthmatic children is the mean monthly change in ACT score. We will have 5 repeated measurements of ACT score per year for each participant and we can assess the effectiveness of each intervention to reduce asthma symptoms during DDS events.
Secondary outcomes will be repeated assessments of asthma medication use, unscheduled visits for asthma, forced expiratory volume in 1 second, peak expiratory flow, and FeNO. For the lung function parameters we will obtain at least three repeated observations for each subject within each group, and this will make possible to assess the health benefits of the intervention and attribute them to the assigned intervention.

Toward this end, we will apply the following regression mixed effects model:

EQUATION1

$$H_{ij} = (\beta_o + v_{oi} + h_{oi}) + (\beta_1 + v_{1,i} + h_{1i}) * (T_{ij} + (1 - T_{ij}) SR_{ij}) * C_{ij} +$$

$$(\beta_2 + v_{2i} + h_{2i}) * D_{ij} + \beta_3 * \text{Temperature}_{ij} + \sum_{m=1}^{M} \gamma_{1m} X_{1mi} + \sum_{n=1}^{N} \gamma_{2n} X_{2nj} + \sum_{p=1}^{P} \gamma_{3p} X_{3pij} + \varepsilon_{ij}$$

$$(v_{oj}\ v_{tj}\ v_{si}\ v_{di}) \sim [(0\ 0\ 0\ 0\ 0), \sigma]$$

$$(h_{oi}, h_{ti}, h_{si}, h_{di}) \sim [(0\ 0\ 0\ 0), \sigma_{\square}]$$

In equation1, Hij is a measured health outcome e.g. ACT score or FEV1 for a subject i on time point j; β0 and v0i are the fixed and subject-specific random intercepts, respectively; β0 and v1i are the fixed and subject-specific random slopes for the exposure term, respectively; β0 and v2i are the fixed and subject-specific random slopes for the physical activity term, Dij respectively; Temperatureij is the ground temperature on time point j corresponding to subject i; β3 is the slope for temperature: xmi is the value of the mth spatial predictor corresponding to subject i; xnij is the value of the nth temporal predictor on time point j corresponding to subject i, and; xpij is the value of the pth spatial-temporal predictor corresponding to subject i on time point j. Finally, h0i, h1i are the subject-specific random intercepts, exposure slopes, and physical activity slopes, respectively, which are specific to control, intervention I1 and intervention I12 groups. The spatial, temporal and spatio-temporal covariates that we will consider for the health effects mixed-effects model are outlined below:

i) Time Activity Data; ii) House Characteristic Data; iii) Meteorological data: We will download weather data from the website of the National Climatic Data Center (NCDC, 2010). We will include weather variables such as temperature, relative humidity, wind speed, visibility, sea level pressure, and planetary boundary layer (PBL); iv) Normalized Difference Vegetation Index (NDVI): We will obtain these data from the U.S., NASA NDVI data from the MODIS sensor. For each subject, we will aggregate NDVI measurements to a 1x1 km grid and a 1-month average. Specifically, we will use the Terra satellite product ID of MOD13A3; v) Land use variables: We will consider nearby emissions of PM2.5, PM10, and NOx from point sources and traffic. For each subject j, we will estimate percent urbanism within a 1x1 km area. We will identify within a 5 km radius geographical predictors such as roads, major buildings, ports, airports, and water bodies using ArcGIS® and ArcMap™ and; vi) Base air pollution emission information: For each home i, we will collect information about the location of small unregistered air pollution sources such as unpaved roads and commercial activities. We will obtain some of this information by analyzing Google Maps.

The advantage of estimating these subject-specific random effects for the control and intervention groups is that we can assess the intervention health benefits for the entire panel. We can do this by subtracting the subject-specific slope*exposure product during the intervention group from that of the control group. In this case, the terms cancel out. As a result, the effectiveness of reducing outdoor exposure, intervention and decreasing home-indoor concentrations in combination with reduction in outdoor exposure, can be assessed individually as follows:

EQUATION2

$$\epsilon_{Di} = D_{i,I1}/D_{i,I0}$$

In equation2, IH11i and IH121i are the health benefits of reducing outdoor exposure and indoor particle concentrations, respectively; h1i0 and h1i1 are the random subject-specific slopes for the control and intervention groups, respectively. Therefore, we can assess which of the intervention types is more effective.

Similarly, we can assess for each subject i the health effects intervention benefit due to the physical activity reduction during the dust storm or anthropogenic pollution episodes, IH2i:

EQUATION3

$$IH_{2i} = h_{2i0} * D_{i0} - h_{2i1} * D_{i1}$$

In equation3, where, h2i0 and h2i1 are the random subject-specific physical activity slopes for the control and intervention groups, respectively. The terms Di0, Di1 are the corresponding activity parameters for subject i during the control and intervention groups.

Furthermore, we can determine the overall panel intervention health benefits by averaging the individual ones:

EQUATION4

$$IH1_1 = \sum_{i=1}^{L} IH1_{1i}/L\,; \quad IH12_1 = \sum_{i=1}^{L} IH12_{1i}/L \text{ and; } \quad IH_2 = \sum_{i=1}^{L} IH1_{2i}/L$$

Analyses of health outcomes in AF patients

For the primary outcome in this panel of patients, which is AF burden, we will have daily recordings of the pacemakers. Therefore we will have daily repeated measurements for participants in all three parallel groups, which can be directly associated with the indoor and outdoor exposure data to be obtained from a random subgroup of participants (20 households) at each project site as mentioned before.

Secondary outcomes-which include blood pressure, heart rate variability and ventricular arrhythmias, will be collected daily. Thus, we will be able using the formulas described above to quantify the benefits of exposure reduction in primary and secondary health outcomes.

Descriptive statistics

Summaries of main variables will be presented in form of means and standard deviations for normal distributed quantitative variables, medians and ranges for non-normal distributed quantitative variables, distribution in percentages for qualitative variables.

Univariate analysis

Univariate analysis will be mostly used for analysis initial datasets consisted of personal data records.

The method of analyses for continuous variables will be parametric using Paired and Unpaired t-test and Repeated Measurements ANOVA.

Non-parametric procedures will be used if parametric assumptions could not be satisfied, even after data transformation attempts. These tests will include Mann-Whitney, Wilcoxon and Spearman Correlation tests.

Categorical variables will be tested using Chi-Square test.

Multivariate analysis

We will apply 2 modelling strategies:

Case- Crossover design

The association between DDS events and episodes of cardiac arrhythmia will be assessed in a case-crossover design [(10)]. This method has been used previously to study triggers of acute cardiovascular events [(11)]. We will define case periods by the time of each confirmed episode of cardiac arrhythmia, rounded to the nearest hour. We will match control periods on weekday and hour of the day within the same calendar month. Conditional logistic regression analysis will be performed to analyze the exposure odds ratio with 95% confidence intervals on risk of cardiac arrhythmia.

Hierarchical model

Generalized Linear Model will be used to fit a Poisson regression of daily AF episodes against temporal variables and DDS data based in hierarchical model. Data model with random effect include repeated episodes of AF in each patient, the intervention groups and the 3 sites of the research. Date model with fixed effect include DDS event during the period of the study.

Analysis will be performed using R Project for Statistical Computing version 3.0 or higher and IBM SPSS software version 23 or higher.

3. References

Voorend-van Bergen S, et al. Monitoring childhood asthma: web-based diaries and the asthma control test. J Allergy Clin Immunol. 2014;133:1599-605.