Analytical Plan for Association between organophosphates and hay fever (NHANES 2005–2006): cross-sectional study

DOCUMENT: SAP-2023-033-CM-v01

From: Felipe Figueiredo **To:** (undisclosed MD researcher)

2023-12-08

TABLE OF CONTENTS

Analytical Plan for Association between organophosphates and hay fever (NHANES 2005–2006): cross-sectional study

Document version

Version	Alterations
01	Initial version

1 ABBREVIATIONS

- CI: confidence interval
- NHANES: National Health and Nutrition Examination Survey
- OP: Organophosphate
- OR: odds ratio
- SD: standard deviation

2 CONTEXT

This analysis is based on the 2005–2006 cohort of the NHANES study (Curtin, 2012), which is a nation-wide representative sample for the US population.

2.1 Objectives

To evaluate the association between exposure to organophosphates and hay fever in the US population.

2.2 Hypotheses

1. There is an association between exposure to Dimethylphosphate and hay fever
2. There is an association between exposure to Diethylphosphate and hay fever
3. There is an association between exposure to Dimethylthiophosphate and hay fever
4. There is an association between exposure to Diethylthiophosphate and hay fever
5. There is an association between exposure to Dimethyldithiophosphate and hay fever
6. There is an association between exposure to Diethyldithiophosphate and hay fever

3 DATA

3.1 Raw data

The raw data was spread among several tables, which were joined using the participant ID (sequence number). Starting from the Organophosphate table (OPD_D), the allergy table (AGQ_D) was merged, ensuring that information from new variables are added to the same initial set of individuals available in the OPD_D sub-sample. This procedure was followed by merging other tables in the following order: three Pesticides tables (UPP_D, CARB_D and PP_D) and finally the demographic table (DEMO_D), resulting in a single table with raw data.

After the merging process the resulting table had 140 variables collected on 2756 observations.

Survey weights relevant to the OP sub-sample is recorded in the `WTSC2YR` field, and the survey strata in the `SDMVSTRA` field.

3.2 Analytical dataset

This analysis will focus on the OP insecticides and a small portion of the demographic fiels available. The allergy table encoded the occurrence of hay fever as 1 for positive and 2 for negative, this variable was recoded to represent 0 as negative values. All missing data codes found in the data dictionary (Curtin, 2012) were recoded as `NA`. All unused fields from the merged raw data table were discarded from the analytical table.

After the cleaning process 13 variables were included in the analysis. The total number of observations excluded due to incompleteness and exclusion criteria will be reported in the analysis. Table 1 shows the structure of the analytical dataset.

Table 1 *Analytical dataset structure after variable selection and cleaning.*

id	outcome	WTSC2YR	SDMVSTRA	RIDAGEYR	RIAGENDR	RIDRETH1	URDOP1LC	URDOP2LC	URDOP3LC	URDOP4LC	URDOP5LC	URDOP6LC
1												
2												
3												
…												
N												

All variables in the analytical set were labeled according to the raw data provided and values were labeled according to the data dictionary for the preparation of production-quality results tables and figures.

4 STUDY PARAMETERS

4.1 Study design

Cross-sectional, based on the complex-design NHANES survey 2005–2006 (Curtin, 2012).

The NHANES sample was collected for a stratified random sampling strategy and this analysis accounts for the sampling design.

4.2 Inclusion and exclusion criteria

N/A

4.3 Exposures

The list of OP that will be evaluated is:

1. Dimethylphosphate
2. Diethylphosphate
3. Dimethylthiophosphate
4. Diethylthiophosphate
5. Dimethyldithiophosphate
6. Diethyldithiophosphate

The measurement reflect exposure in the 12 months prior to the survey. For this analysis, we assume the exposure reflects chronic exposure (see Observations).

4.4 Outcomes

Specification of outcome measures (Zarin, 2011):

1. (Domain) Allergies
2. (Specific measurement) Hay fever
3. (Specific metric) End value
4. (Method of aggregation) Proportion of participants that had a diagnosis of hay fever in the past 12 months

Primary outcome

Proportion of participants that had a diagnosis of hay fever in the past 12 months.

4.5 Covariates

1. Age (years)
2. Gender
3. Ethnicity

5 STATISTICAL METHODS

5.1 Statistical analyses

5.1.1 Descriptive analyses

The epidemiological profile of the study participants will be described. Demographic (sex, age and BMI) and clinical variables will be described as mean (SD) or as counts and proportions (%), as appropriate. The distributions of participants' characteristics will be summarized in tables and visualized in exploratory plots.

5.1.2 Inferential analyses

All comparisons between groups will be performed as univariate analyses. Continuous variables will be compared between groups with the independent t test with Welch correction. Differences in distribution of categorical variables will be assessed with the Fisher exact test.

5.1.3 Statistical modeling

A weighted multivariate logistic regression will be fitted to the data to estimate the strength of the association between OP (Section 4.3) and the occurrence of hay fever (section 4.4). These estimates will be adjusted by age, sex and ethnicity (section 4.5).

5.1.4 Missing data

No missing data imputation will be performed. All evaluations will be performed as complete case analyses. Missing data counts and proportions will be reported in tables.

5.2 Significance and Confidence Intervals

All analyses will be performed using the significance level of 5%. All significance hypothesis tests and confidence intervals computed will be two-tailed.

5.3 Study size and Power

N/A

5.4 Statistical packages

This analysis will be performed using statistical software R version 4.3.0.

6 OBSERVATIONS AND LIMITATIONS

Assumptions about the exposures

The data for the exposure to OP originated in laboratory assays, which ensure the validity of those measurements. On the other hand those exposures measure a single event during the two-years period used here.

For the purpose of interpretation of the results of this analysis it is assumed that participants make frequent use of those insecticides, thus having chronic exposure to these compounds. This assumption could imply a risk of information bias if many participants had a single event of exposure coinciding with the data collection by NHANES staff. The assumption can be interpreted as the number of participants in the sample having only such acute exposures being small or negligible.

Recommended reporting guideline

The adoption of the EQUATOR network (http://www.equator-network.org/) reporting guidelines have seen increasing adoption by scientific journals. All observational studies are recommended to be reported following the STROBE guideline (von Elm et al, 2014).

7 REFERENCES

- **SAR-2023-033-CM-v01** – Association between organophosphates and hay fever (NHANES 2005–2006): cross-sectional study
- Curtin LR, Mohadjer L, Dohrmann S, et al. The National Health and Nutrition Examination Survey: Sample design, 1999–2006. National Center for Health Statistics. Vital Health Stat 2(155). 2012.
- Zarin DA, et al. The ClinicalTrials.gov results database – update and key issues. N Engl J Med 2011;364:852-60 (https://doi.org/10.1056/NEJMsa1012065).
- Gamble C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA. 2017;318(23):2337–2343 (https://doi.org/10.1001/jama.2017.18556).
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014 Dec;12(12):1495-9 (https://doi.org/10.1016/j.ijsu.2014.07.013).

8 APPENDIX

This document was elaborated following recommendations on the structure for Statistical Analysis Plans (Gamble, 2017) for better transparency and clarity.

8.1 Availability

All documents from this consultation were included in the consultant's Portfolio.

The portfolio is available at:

https://philsf-biostat.github.io/SAR-2023-033-CM/