PROTOCOL FOR STATISTICAL ANALYSES IN THE STUDY ENTITLED "PESTICIDE EXPOSURE, ASTHMA AND DIABETES IN UGANDA (PEXADU)"

Version 1.0

Last edited on Wednesday, September 4, 2019

Martin Rune Hassan Hansen[1,2*], Erik Jørs[3], Annelli Sandbæk[1,4], Daniel Sekabojja[5], John Ssempebwa[6], Ruth Mubeezi[6], Philipp Staudacher[7,8], Samuel Fuhrimann[9], Wajd Abbas Hassan Hansen[1], Vivi Schlünssen[1,2]

1. Aarhus University, Aarhus C, Denmark
2. National Research Center for the Working Environment, Copenhagen, Denmark
3. Odense University Hospital, Odense, Denmark
4. Steno Diabetes Center Aarhus, Aarhus, Denmark
5. Uganda National Association of Community and Occupational Health, Kampala, Uganda
6. Makerere University, Kampala, Uganda
7. Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
8. Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zürich, Switzerland
9. Utrecht University, Utrecht, Netherlands

* Corresponding author:

Martin Rune Hassan Hansen

Section for Environment, Work and Health, Danish Ramazzini Centre, Department of Public Health, Aarhus University

Bartholins Allé 2, Building 1260

DK-8000 Aarhus C

Denmark

martinrunehassanhansen@ph.au.dk

Contents

Preface

This document describes the statistical analyses planned as part of the scientific project entitled "Pesticide Exposure, Asthma and Diabetes in Uganda" (PEXADU). The document is based partly on the project protocol that was approved by the institutional review board at Makerere University School of Public Health (Kampala, Uganda) on July 2, 2018. The document also contains information from a presentation of preliminary results[1] made at the 2019 EPICOH (Epidemiology in Occupational Health) conference. The abstract for the latter is freely available from https://oem.bmj.com/content/76/Suppl_1/A3.1

1 Abstract

BACKGROUND

Insecticides are important both to control agriculture and for the fight against vector-borne diseases such as malaria. Previous studies show that relatively low exposure to organochlorine insecticides is associated with diabetes mellitus, but less is known for other more widely used classes of insecticides such as organophosphates and other insecticides. Other studies suggest an association between insecticide exposure and lung function impairment, but the publications are few and hampered by weak study designs and insufficient confounder control.

AIMS

1) Investigate the possible association between insecticide exposure, diabetes and lung function impairment.
2) Assess whether the effects are acute or chronic, reversible or irreversible.
3) Establish exposure-response relationships between specific insecticides and health outcomes for risk assessment and management.

METHODS AND MATERIALS

Repeated cross-sectional study among two groups of small-scale farmers from the Wakiso District of Uganda: conventional farmers and semi-organic farmers. The main outcome of interest are temporal changes in blood glucose levels (determined fasting plasma glucose and glycosylated hemoglobin A) and objective lung function parameters. Insecticide exposure is determined by passive sampling using silicone bracelets, measurement of blood acetylcholine esterase and subjective exposure information.

PERSPECTIVES

If exposure to insecticides increase the risk of diabetes or lung function impairment, it will have to be taken into account when planning public health interventions using insecticides against vector-borne diseases and when using insecticides in agriculture.

2 Acronyms and abbreviations

- ACh = acetylcholine
- AChE = acetylcholine esterase
- ATS = American Thoracic Society
- AU = Aarhus University, Aarhus, Denmark
- DM = diabetes mellitus
- DM1 = diabetes mellitus type 1
- DM2 = diabetes mellitus type 2
- FEV_1 = forced expiratory volume in 1 second
- FEV_6 = forced expiratory volume in 6 seconds
- FEV_1/FVC = FEV_1 divided by FVC
- FPG = fasting plasma glucose
- FVC = forced vital capacity
- HbA_{1c} = glycosylated hemoglobin A
- MU = Makerere University, Kampala
- NRCWE = National Research Center for the Working Environment, Copenhagen, Denmark
- ODK = Open Data Kit (https://opendatakit.org/)
- OP = organophosphates (refers to organophosphate insecticides only)
- OUH = Odense University Hospital, Odense, Denmark
- PEXADU = “Pesticide Exposure, Asthma and Diabetes in Uganda” (project title)
- RCM = Random Coefficient Model

- UNACOH = Uganda National Association of Community and Occupational Health

3 Definitions

- ***Acetylcholine***

  Neurotransmitter substance. The main neurotransmitter of the parasympathetic nervous system.

- ***Acetylcholine esterase***

  Enzyme that catalyzes the breakdown of acetylcholine. Important for the normal function of nerves that use acetylcholine as a transmitter substance.

- ***Asthma***

  Chronic inflammatory airways disease characterized by hyper-reactive airways and intermittent bronchoconstriction, mucus hypersecretion and edema of the airway mucosa.

- ***Bronchoconstriction***

  Narrowing of the airways.

- ***Bronchodilator***

  Medication that cause airway smooth muscle cells to relax so that the airways open up more.

- ***Chronic bronchitis***

  Chronic inflammation in the bronchi (lower airways). Clinically defined as coughing and bringing up phlegm for most days for at least three months two years in a row.

- ***Chronic obstructive lung disease***

  Chronic pulmonary disease with varying components of emphysema and chronic bronchitis. Often known as “smoker's lung”, but other influences than smoking can also lead to the disease.

- ***Diabetes***

  Common term for diabetes mellitus. Technically, the word “diabetes” can also mean “diabetes insipidus”, but that is an entirely different disease that will not be discussed further in this protocol.

- ***Diabetes mellitus***

  A heterogeneous group of diseases characterized by varying degrees of hyperglycemia, insulin resistance and decreased insulin production. The main types are diabetes mellitus type 1 (characterized by autoimmune destruction of the insulin-producing beta-cells of the pancreas) and diabetes mellitus type 2 (characterized primarily by insulin resistance).

- ***Dyspnea***

  Shortness of breath.

- ***Emphysema***

  Destruction of septa between the alveoli, leading to

  - Decreased gas exchange in the lungs
  - Decreased elasticity of the lung tissue (and thus, dynamic airway collapse during exhalation)

- ***Forced expiratory volume in 1 second***

  The amount of air that is exhaled in the first second of a forced exhalation following a maximal inhalation. Measured by spirometry.

- ***Forced vital capacity***

  The total amount of air that can be exhaled following a maximal inhalation. Measured by spirometry.

- ***FEV1/FVC ratio***

  Forced expiratory volume in 1 second divided by forced vital capacity. A measure of airway obstruction.

- ***Glycosylated hemoglobin A***

  Hemoglobin A (the main oxygen-transporting molecule in the blood) that has non-enzymatically reacted with glucose. The speed of this reaction depends on the glucose concentration in the blood, so glycosylated hemoglobin A can be used as a measure of mean blood glucose level during the last 6-8 weeks.

- ***Herbicide***

  Pesticide targeted against plants.

- ***Hyperglycemia***

  Increased blood glucose level.

- ***Insecticide***

  A pesticide targeted against insects.

- ***Mucus hypersecretion***

  Abnormally high production of mucus in the airways.

- ***Pesticide***

  A chemical compound used to kill organisms considered unwanted by humans. In the context of this protocol, the term only considers synthetic compounds and not natural compounds such as plant extracts, or insect-killing bacteria.

- ***Spirometry***

  Clinical examination of a person's ability to exhale air (amount and velocity). Commonly referred to as "lung function testing", but does not test the ability of the lungs to perform gas exchange.

5 Background

Use of pesticides is important for both modern agriculture and control of vector-borne diseases such as malaria. However, scientific studies indicate that chronic exposure to pesticides can lead to health damage – even at low levels without any acute symptoms. Among other things, associations with both diabetes mellitus, respiratory symptoms, and decreased lung function are suspected.

A systematic review has indicated an association between exposure to insecticides and the risk of diabetes mellitus,[2] a serious disease affecting an estimated 2.8 percent of the population of Uganda.[3] The majority of the studies included in the systematic review only considered organochlorine insecticides such as DDT (dichloro-diphenyl-trichloroethane)[2]. Less is known about more widely used groups of insecticides such as organophosphates. However, there are indications that also organophosphates can cause diabetes mellitus. A study from a rural area in Iran found significant higher fasting blood glucose (FBG) levels among organophosphate-exposed farmers than among matched non-exposed controls (mean FBG 84.90 vs. 78.31 mg/dL, respectively).[4] Furthermore, a recent publication indicated that organophosphate exposure among Indian farmers changed the gut microbiome, causing increased production of short-chain fatty acids in the intestines and thus, increased risk of diabetes mellitus.[5]

Chronic obstructive lung disease is a common disease affecting 16.2% of men and women in rural Uganda.[6] In recent years, it has been suggested that low-dose pesticide exposure is linked to respiratory symptoms and decreased lung function. A 2014 review concluded that exposure to pesticides may be associated with airway obstruction, and that the evidence was stronger for asthma than for chronic obstructive pulmonary disease.[7] However, the authors found that many of the studies had weak designs and did not adequately deal with confounders.[7] A recently published study from Ethiopia demonstrated that the FEV_1 (forced expiratory volume in 1 second) was significantly lower among farmers directly exposed to pesticides, and they found an exposure-response relationship with higher-exposed persons performing worse.[8] An association between organophosphate and carbamate insecticides and respiratory symptoms, asthma and chronic obstructive pulmonary disease is biologically plausible. The primary effect of organophosphates is inhibition of the enzyme acetylcholine esterase, the enzyme responsible for degrading acetylcholine.[9] Acetylcholine is the neurotransmitter of the parasympathetic nervous system (PNS). Increased activity in the PNS leads to bronchoconstriction and increased mucus production - both well-known signs of acute organophosphate intoxication.[10]

6 Hypotheses

1. Exposure to organophosphates and other insecticides increases the blood sugar level.
2. Exposure to organophosphates and other insecticides increases the risk of diabetes.
3. Exposure to organophosphates and other insecticides leads to bronchoconstriction.
4. The effects described in hypothesis 1-3 follow a positive exposure-response relationship.

7 Objectives

7.1 General objective

To examine the temporal relationship between exposure to insecticides and changes in glycemic status and lung function among a group of occupationally exposed farmers.

7.2 Specific objectives

To statistically describe changes in the following primary parameters in relation to a two-month spraying season:

- Fasting plasma glucose,
- Glycosylated hemoglobin (HbA_{1c}),
- Lung function measurements (FEV_1, FVC and FEV_1/FVC),

and to model these changes as a function of each participant's exposure level.

8 Data collection methodology

8.1 Setting

Data was collected from a group of 364 small-scale farmers from the Wakiso District in central Uganda. The farmers were recruited from two farmer's organizations – one organization for conventional farmers (with assumed high level of exposure to pesticides) and one organization of farmers working towards organic certification (with an assumed lower level of pesticide exposure). Participants were recruited in July 2018 by visiting individual farmer's groups associated with each of the two larger farmer's organizations. Participants were interviewed and examined at an examination center that was set up for this project.

8.2 Study design

We carried out three repeated cross-sectional studies in the same study population:

- Phase 1: Early September – early October 2018
- Phase 2: Mid-November – early December 2018
- Phase 3: Early January – early February 2019

The timing of the study was chosen because we expected farmers to use the most pesticides in October-November, and smaller amounts before and after. This expectation was based on knowledge about the timing of the agricultural seasons in the Wakiso District (personal communication, Aggrey Atuhaire, UNACOH).

8.2.1 Criteria for inclusion and exclusion

8.2.1.1 High-exposed persons

- Inclusion criteria: Members of the selected farmers' groups from the conventional farmers' organization. Both females and males above the age of 18 years will be recruited.
- Exclusion criteria:
    - Persons who refuse to sign the informed consent form.
    - Women who report pregnancy.
    - Lung function testing will *not* be performed for participants who report any of the following:
        - Myocardial infarction in the last 3 months.
        - Suffering from angina pectoris.
        - Suffering from hemoptysis
        - Any kind of surgery in the last 3 months
        - Aortic aneurism
        - History of pulmonary embolism

8.2.1.2 Low-exposed persons

- Inclusion criteria: Members of the selected farmers' groups from the semi-organic farmers' group. Both females and males above the age of 18 years were recruited.
- Exclusion criteria: Same criteria as for exposed persons (see above).

8.3 Data collection

8.3.1 Questionnaire-based structured interview

Insecticide exposure was determined primarily by interviewer-administrated questionnaires. We collected data on the duration of working with agricultural pesticides, the intensity of insecticide exposure during work, the personal protective equipment and the types of insecticides used. Information on known risk factors for diabetes and pulmonary disease were collected using an adapted version of standardized STEPS instrument[11] (developed by the WHO), supplemented with questions from other standardized questionnaires. The questionnaire included questions on diet, physical activity level and other possible confounders.

8.3.2 Biological samples

At each examination, one or two finger-prick capillary blood samples and one 4ml venous blood sample were taken for point-of-care biochemical analysis. In addition, a random sample of participants gave spot urine samples that were stored for later analysis:

Sample type	Parameter measured	Purpose	Eligible participants	Measurement device
Venous full blood (anticoagulant: K_2-EDTA)	HbA1c	Primary measure of glycemic regulation	Everyone	HemoCue Hba1c 501
Capillary full blood	Fasting plasma glucose	Secondary measure of glycemic regulation	Those who came fasting in the morning	HemoCue Glucose 201 RT
Capillary full blood	AChE Hemoglobin concentration AChE normalized by hemoglobin concentration	Biomarker of exposure to organophosphate and carbamate insecticides	Everyone	Test-mate ChE Cholinesterase Test System (Model 400)
Spot urine	Pesticide metabolites	Objective measure of recent pesticide exposure	Random sub-sample of participants (selected using pseudo-random number generator). 50% of conventional farmers, 20% semi-organic farmers.	Not yet measured. Planned to be analyzed using LC-MS.

Glycosylated hemoglobin A (HbA_{1c}) is a measure of the average blood glucose levels for the last 6-8 weeks[12].

8.3.3 Silicone bracelets

All participants were asked to wear a silicone bracelet from phase 1 to phase 2. A random subsample of approximately 100 persons were also asked to wear a silicone bracelet from phase 2 to phase 3. The latter group was randomly selected among participants eligible to give urine samples – see "8.3.2 Biological samples". The silicone bracelets passively absorb insecticides at a rate relative to the exposure levels.[13 14]

In phase 2 and 3, we collected the bracelets, packed them in diffusion-proof bags and stored them at -20 degrees Celsius. The bracelets will be shipped to Denmark for analysis of insecticide residues using liquid chromatography – mass spectrometry (LC-MS).

8.3.4 Lung function testing

At each examination round, eligible participants performed spirometry without reversibility testing using a handheld spirometer and according to guidelines from the American Thoracic Society.[15] Reversibility testing was not carried out, as we did not have ethical clearance for the administration of bronchodilator medicine. During each testing session, each participant blew a minimum of five and a maximum of nine times.

Spirometry was carried out using a diagnostic-quality spirometer (MicroMedical MicroDL). In addition, participants eligible for spirometry in phase 1 were also tested using a mini-spirometer (Vitalograph copd-6). The order of testing (MicroDL or copd-6 first) was determined using a pseudo-random number generator. Participants always blew into the copd-6 three times. The purpose of including the copd-6 device in the project was to evaluate is precision and accuracy, and hence to judge whether this cheaper device can reliably be used for lung function testing in future studies on environmental and occupational health.

8.3.5 Anthropometry

Height and weight were measured to determine BMI (body mass index). We also measured participants' hip and waist circumferences for calculation of the hip-to waist ratio, and measured participants' blood pressure and pulse rate.

8.3.6 Data management

In the field, the majority of the collected data was collected in digital format using ODK (Open Data Kit) software. Only minor amounts of data (participant registration sheets, information from filled-out informed consent forms and registration sheets with biochemical results) were collected in paper form, and later digitized in duplicate using ODK.

Data will be managed and analyzed using the Stata 15 statistical software package (StataCorp LLC, College Station, Texas).

9 General considerations for statistical analyses

9.1 Overview of analyses

The planned analyses are presented below. Analyses are placed in four groups, corresponding to the subjects of the planned papers from the PEXADU project:

- 10 Analysis plan for organophosphate exposure

- Analysis plan for glycemic regulation
- 12 Analysis plan for lung function tests
- 13 Analysis plan for validation of Vitalograph copd-6

9.2 Level of significance

p-values ≤ 0.05 will be considered significant.

A relatively high number of statistical tests will be carried out because of the many independent variables we want to examine - e.g. HbA_{1c} (continuous), FPG (continuous), diabetes (yes/no), FEV_1 z-score (continuous), FVC z-score (continuous) and FEV_1/FVC z-score (continuous). The number of tests means that there is a risk of mass significance, i.e. finding statistically significant results where no true differences exist. By definition, this will happen in 5% (= the level of significance) of all tests. While we will not try to account formally for this (e.g. by Bonferroni correction[16]), it will be kept in mind when interpreting results.

9.3 Interdependence of data

Because of the way participants were recruited for the PEXADU project, some participants were relatives. Before any more advanced analyses are carried out, we will tabulate the number of participants who stated that they were related to another participant. Depending on the result, we may decide that the interdependence between observations is negligible, or we may decide that it is sufficiently large to warrant consideration in the statistical analyses (e.g., by including family as a random effect in mixed effect models).

10 Analysis plan for organophosphate exposure

10.1 Purpose and introduction

Most of the available information on pesticide exposure in the PEXADU project is questionnaire-based. Measurements of red blood cell acetylcholine esterase (AChE) are available for all persons in all phases, but express exposure to only organophosphate and carbamate insecticides in the last approximately three months (see details in section 1.1.1). In order to assess health effects of any other pesticides, as well as effects of long-term exposure to organophosphate and carbamate insecticides, we will create and validate a pesticide-exposure score with information on specific pesticide compounds. Our primary exposure score will be based on a deterministic model, but we will also develop a statistical model in order to get an empirically based exposure score.

At the time of writing, the only biomarker of pesticide exposure available from the PEXADU project is AChE. As preliminary analyses of the data has shown that very few persons in the study had used carbamates, and none had done so within the last week before the interview (data not shown), we will only be able to validate the exposure scores for organophosphate insecticides. We are still working on getting the necessary permits to export the urine samples and silicone bracelets from Uganda to Denmark for analysis so that we can also validate scores for other classes of pesticides.

10.2 Statistical procedures

The derivation and validation of the pesticide exposure score will be based on self-reported exposure to organophosphates before each interview (phases 1/2/3) and validated against AChE measurements. To derive the best possible exposure score, we will create and compare the performance of two statistical models: A deterministic model and an empirical model.

10.3 Descriptive statistics

Before the analyses are carried out, we will draw a table of demographics, stratified by organization membership (outlined in

Table 1). For the persons who reported ever having mixed or applied pesticides in the baseline interview, we will draw a table specifying the frequency of use of all listed types of PPE, and hygienic practices (outlined in Table 2).

TABLE 1: DEMOGRAPHICS

Characteristic	All participants	Conventional farmer's group	Semi-organic farmer's group
Total n			
Sex			
Male, n (%)			
Female, n (%)			
Age in years: Median, IQR			
Educational level (years of full-time schooling): Median, IQR)			
Alcohol consumption in the last week (unit): Median, IQR			
Farming as main occupation: n (%)			
Ever mixed or applied pesticides, n (%)			
Mixed or applied pesticides in week prior to interview			
Phase 1, n (%)			
Phase 2, n (%)			
Phase 3, n (%)			
Number of hours of farm work (excluding work with pesticides) in the last week			
Phase 1: Median, IQR			
Phase 2: Median, IQR			
Phase 3: Median, IQR			
Number of hours of farm work (excluding work with pesticides) in the last week, limited to farms where pesticides are used			
Phase 1: Median, IQR			
Phase 2: Median, IQR			
Phase 3: Median, IQR			

TABLE 2: DESCRIPTIVE STATISTICS FOR PPE USE AND HYGIENIC PRACTICES

		All participants	Conventional farmer's group	Semi-organic farmer's group
Total n				
PPE type	Frequency of use, n (%)			
Dust mask	All the time (100%)			
	Most of the time (75%)			
	Often (50%)			
	Rarely (25%)			
	Never (0%)			
(…)				
Cap	All the time (100%)			
	Most of the time (75%)			
	Often (50%)			
	Rarely (25%)			
	Never (0%)			
Time of showering	Immediately or within one hour, n (%)			
	A few hours later, n (%)			
	Many hours later, n (%)			
	Next day, n (%)			
Time of changing clothes	Immediately or within one hour, n (%)			
	A few hours later, n (%)			
	Many hours later, n (%)			
	Next day, n (%)			

10.4 Description of the deterministic exposure model

The deterministic exposure model will take into account the amount of organophosphate insecticides used, the timing of exposure, the potency of individual compounds, PPE use and hygienic measures. These factors are described in details below.

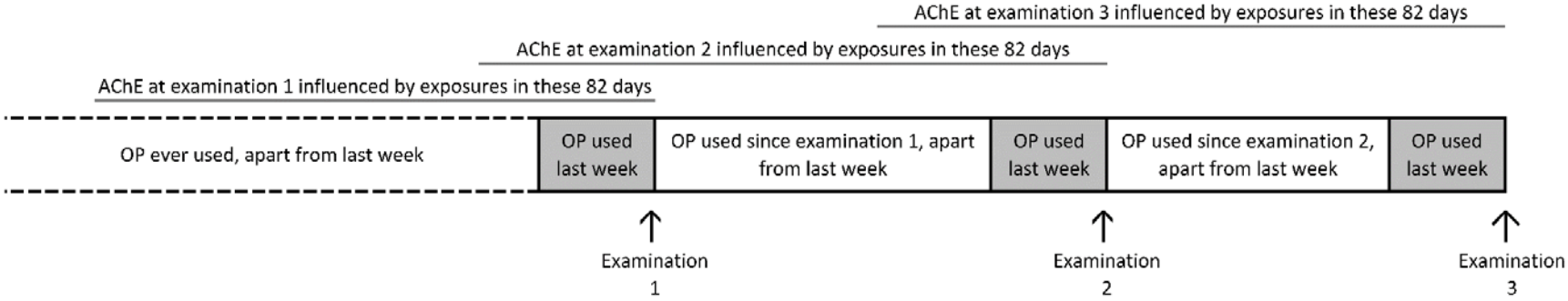
10.4.1 Weighting factor for time since exposure

While carbamate insecticides reversibly inhibit AChE,[17] the inhibition caused by organophosphate insecticides can be irreversible.[17] This means that even though organophosphates are not biologically persistent in the human body, the inhibition caused by organophosphate insecticide exposure can persist for a considerable amount of time after the exposure happened.[18]

A study among workers occupationally exposed to the organophosphate insecticide dichlorvos showed that RBC AChE did not revert to baseline until after approximately 82 days after the exposure ended, and the increase in RBC AChE was a linear (as opposed to exponential) function of time.[18] In our analyses of subjective exposure information vs. AChE, we have to take into account this duration of AChE inhibition after exposure.

The interval between examinations of each participant in our study was approximately 2 months. At each examination, the participants were asked about the pesticides that they had used most in the last week before the examination. In phases 2 and 3, they were additionally asked about the pesticides they had used most since the last examination, but not in the last week. In phases 1, we instead asked them about the pesticides they had used most since they started using pesticides, but not in the last week. I.e., in phase 1 we do not know the timing of the exposure that happened more than one week before the examination. An overview is provided in Figure 1.

FIGURE 1: OVERVIEW OF TIMING OF EXAMINATIONS AND TIMING OF EXPOSURE INFORMATION

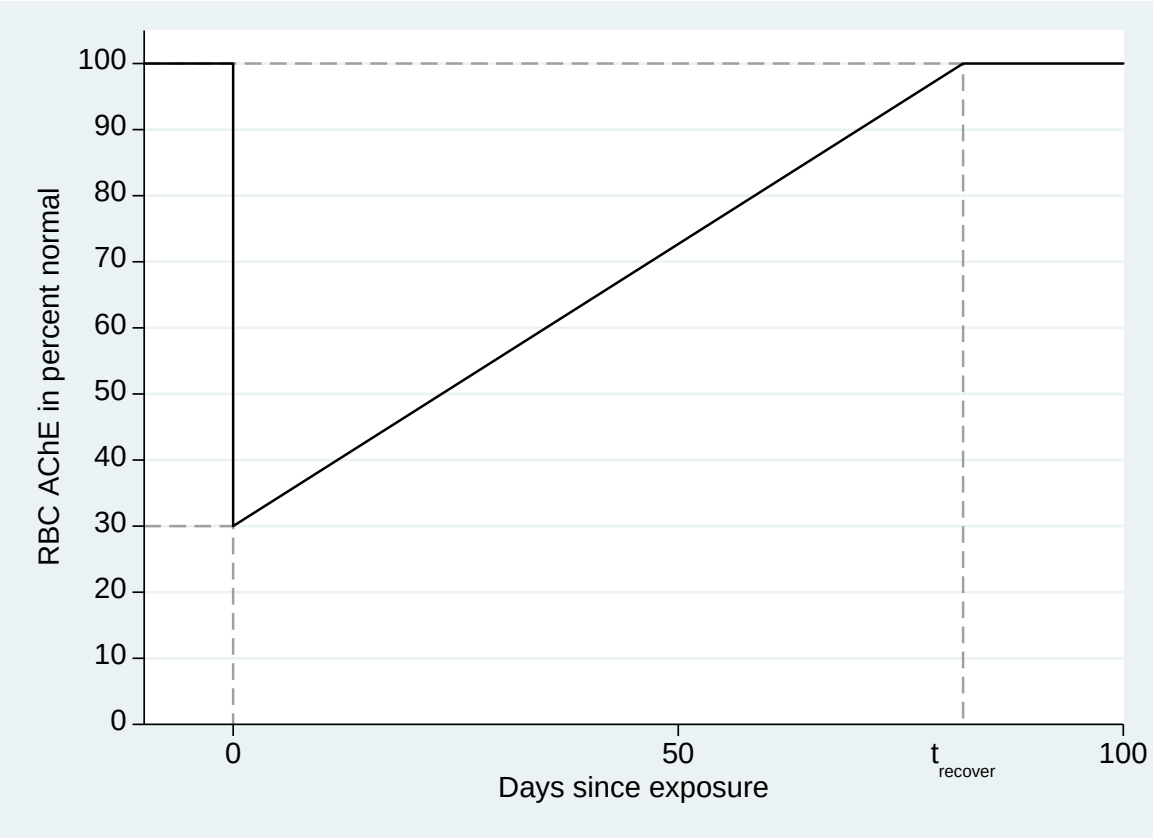


OP = Organophosphate insecticides. The durations of the time windows are not to scale.

As outlined in Figure 1, we expect AChE measurements in both of the phases 1 and 2 to be influenced by exposure in the time preceding the phase 1 examination by more than one week (i.e., in a time interval where we do not know the timing of exposure). This means that it will be difficult to estimate how much participants have been exposed within the time window where the exposure could influence AChE in phases 1 and 2. Only in phase 3 do we have a good idea of the timing of all organophosphate exposures that could have influenced AChE. Hence, we will base our validation on the subjective organophosphate exposure within the last 82 days before the phase 3 examination, and on the AChE measurement in phase 3. We have to weight each exposure in this time window according to the time that has elapsed since the exposure, and the weighting strategy is described in details below.

We expect the RBC inhibition to decrease (or in other words, the enzyme activity to increase towards normal) as a linear function of time since exposure.[18] A schematic representation of this model is shown in Figure 2, where a single exposure happens at time 0, and $t_{recover}$ is the time from exposure termination to normalization of RBC AChE.

FIGURE 2: MODEL OF RBC ACHE AS FUNCTION OF TIME AFTER A SINGLE EXPOSURE TO A SINGLE ORGANOPHOSPHATE INSECTICIDE



In the following, D will denote the inhibition/decrease of RBC AChE relative to the normal, in percent:

EQUATION 10-1

$$D = AChE_{normal} - AChE$$

At time 0, we will assume that the degree of inhibition is the product of the absorbed dose of organophosphate insecticide and its absolute potency ($potency_{abs}$ = its ability to inhibit the enzyme at a given dose). In section 10.4.3, we describe how the absorbed dose is a function of the external exposure to the compound and the use of personal protective equipment, and in section 10.4.2 we further explain the concept of potency.

EQUATION 10-2

$$D(0) = potency_{abs} \times dose$$

Inspecting Figure 2, we can derive the following three equations for D as a function of time after exposure (t):

EQUATION 10-3

$$D(t) = \begin{cases} 0 \; for \; t < 0 \\ potency_{abs} \times dose - \dfrac{potency_{abs} \times dose}{t_{recover}} \times t \; for \; t \in [0, t_{recover}] \\ 0 \; for \; t > t_{recover} \end{cases}$$

We can rewrite this as

EQUATION 10-4

$$D(t) = potency_{abs} \times dose \times \begin{cases} 0 \; for \; t < 0 \\ 1 - \dfrac{t}{t_{recover}} \; for \; t \in [0, t_{recover}] \\ 0 \; for \; t > t_{recover} \end{cases}$$

Therefore, we define the time weight *w* as the last factor in Equation 10-4:

EQUATION 10-5

$$w = \begin{cases} 0 \; for \; t < 0 \\ 1 - \dfrac{t}{t_{recover}} \; for \; t \in [0, t_{recover}] \\ 0 \; for \; t > t_{recover} \end{cases}$$

Where *t* is the number of days since exposure and $t_{recover}$ is the number of days it takes for AChE to revert to normal after a single organophosphate exposure. Human data for workers exposed to dichlorvos suggest the best estimate is $t_{recover} = 82$, but with a 95% CI ranging from 72 to 98.[18] We will use the best estimate in the primary analysis, but perform sensitivity analyses with both the upper and lower limit of the 95% CI.

10.4.2 Weighting factor for potency of each organophosphate insecticide

Before we try to correlate organophosphate insecticide exposures with AChE measurements, the exposures must be weighted not only by the time since exposure, but also by the potency of the compound (i.e., its ability

to induce acetylcholine esterase inhibition), as the potencies of different organophosphate insecticides can vary by up to four orders of magnitude (see below). The absolute potency is defined in Equation 10-2.

Based on published data from animal experiments, the US EPA has published lists of the relative oral potencies (i.e., the potency compared to that of methamidophos) of a number of organophosphate insecticides, based on both brain[19] and RBC[20] (red blood cell) enzyme isoforms. The potencies are listed in Table 3. In the primary model, we will weight exposures by their relative RBC potency, while the brain potency will be used in secondary analyses. For some compounds, only the brain potency was available from the US EPA. In these cases, we have predicted the RBC potency as a function of the brain potency and the lipofilicity of the compound, as detailed in Appendix A. Potential parameters for the prediction model were chosen based on literature.[21]

TABLE 3: RELATIVE POTENCIES OF ORGANOPHOSPHATE INSECTICIDES REGARDING ACHE INHIBITION

	Relative potency	
	Brain	**Red blood cells**
Acephate	0.08	0.0211
Azinphos-methyl	0.1	0.3504
Bensulide	0.003	0.0113
Chlorethoxyfos	0.13	0.7357 *
Chlorpyrifos	0.06	0.1002
Chlorpyrifos-methyl	0.005	0.0255 *
Diazinon	0.01	0.2205
Dichlorvos	0.03	0.1453
Dicrotophos	1.91	2.1779 *
Dimethoate	0.32	0.4187
Disulfoton	1.26	4.5647
Ethoprop	0.06	0.2397 *
Fenamiphos	0.04	0.6504
Fenthion	0.33	1.5692 *
Fosthiazate	0.07	0.3964
Malathion	0.0003	0.0041
Methamidophos	1	1.0969
Methidathion	0.32	0.2658
Methyl-parathion	0.12	0.2690
Mevinphos	0.76	0.5391
Naled	0.08	0.0326
Omethoate	0.93	0.7751 *
Oxydemeton-methyl	0.86	0.5741
Phorate	0.39	3.5171
Phosalone	0.01	0.0722
Phosmet	0.02	0.1296
Phostebupirim	0.22	1.0832 *
Pirimiphos-methyl	0.04	0.0326
Profenofos	0.004	0.0234 *

Terbufos	0.85	2.9736
Tetrachlorvinphos	0.001	0.0028
Tribufos	0.02	0.2493
Trichlorfon	0.003	0.0047

* = RBC potency estimated as detailed in Appendix A.

10.4.3 Weighting factor for use of personal protective equipment and hygienic measures

In phase 1 of the PEXADU project, all participants in the PEXADU project who reported that they had ever sprayed or mixed pesticides were asked how often they used each of the following categories of personal protective equipment (PPE), among others, when handling pesticides:

- Dust mask
- Mask with carbon filter
- Goggles
- Gloves
- Long-sleeved shirt
- Rubber apron
- Rain poncho
- Overalls
- Long pants
- Gaiters
- Water proof pants
- Rubber boots

For each of each of these kinds of PPE, participants were asked how often they used them. The following categories were available as answers:

- Always (100%)
- Often (75%)
- Sometimes (50%)
- Rarely (25%)
- Never (0%)

Participants could also state “I don’t know” or refuse to answer the question. If participants stated in phase 1 that they had never used pesticides, yet stated in phase 2 or 3 that they had used pesticides since the last examination, we do not have any information on their use of PPE. In the primary analysis, the missing values will be replaced with values imputed based on a linear regression model of PPE use vs. participant sex, age and farmer’s organization (conventional farmer’s group vs. semi-organic farmer’s group) among the participants where the information is available. In sensitivity analyses, participants with any missing information will be excluded.

We will assume that the protection offered by the protective equipment can be expressed as a weighting factor *PPE*. Furthermore, we assume that the *PPE* protection factor is the sum of the protection offered to six different body parts:

$$PPE = PPE^{UPPERBODY} + PPE^{EYE} + PPE^{MOUTH} + PPE^{HAND} + PPE^{LEG} + PPE^{FEET}$$

This formula is based on an existing pesticide exposure score developed for use among small-holder farmers in Costa Rica.[22] Each type of PPE is assumed to protect a specific part of the body only, and lowers exposure to that body part by a specific fraction, estimated from literature. Deterministic exposure scores for different types of PPE are seen in Table 4. If a participant has reported use of two kinds of PPE that protect the same body part, we will base our calculations on the PPE that protects the most.

Participants were also asked how soon after pesticide exposure they showered and changed their clothes. The protection factor offered by these hygienic measures are seen in Table 5.

10.4.4 Creating the deterministic model:

As suggested in previous work,[22] for each organophosphate insecticide reported by participants, we will assume that the intensity of external pesticide exposure per session is a function of both preparing/mixing the compound (denoted MIX below) and applying it (APPLY):

$$INTENSITY_{EXTERNAL} = MIX + APPLY$$

The magnitude of the internal intensity per exposure is the intensity of external exposure, multiplied by factors accounting for the use of personal protective equipment, changing clothes (CHANGE) and/or showering (SHOWER) after handling pesticides:[22]

$$INTENSITY_{INTERNAL} = INTENSITY_{EXTERNAL} \times PPE \times CHANGE \times SHOWER$$

As described above, the degree of acetylcholine inhibition induced by the exposure is the product of the potency of the compound, a weighting factor *w* taking into account the amount of time since the exposure, and the internal intensity of the exposure:

$$D(t) = potency_{abs} \times INTENSITY_{INTERNAL} \times w$$

In the study population, participants were reported exposed to multiple different organophosphate insecticides, and exposure was repeated (rather than at a single point in time). We will assume that the inhibition caused by these exposures is additive:

EQUATION 10-6

$$D(t)_{total} = \sum_{i,j} D(t)_{i,j}$$

Where $D(t)_{i,j}$ us the inhibition caused by the i^{th} exposure to the j^{th} organophosphate insecticide compound.

Since we do not know the absolute potency of all included organophosphates, but rather the relative potency, we will use the latter instead. The final deterministic exposure score SCORE that we will validate against the measured AChE activity is therefore given by Equation 10-7:

EQUATION 10-7

$$SCORE = PPE \times CHANGE \times SHOWER \times (MIX + APPLY) \times \sum_{i,j} potency_{rel,j} \times w_i$$

10.4.4.1 Information that will not be included in the deterministic model

The deterministic model will not include time spent farming in general (excluding the time working with pesticides), and whether other people on the farm use pesticides. While we do have this information from participant interviews, we cannot use it to obtain data on specific pesticidal compounds or classes of compounds. We recognize that pesticide exposure during work in pesticide-treated fields or stables can be significant, but we will focus on compound-specific information regarding participant's own use of pesticides, as we deem that more helpful in the assessment of health effects of specific pesticides.

We will also not include information on the use of insecticide-treated mosquito nets, even though this information is also available from the interviews. In phase 1, 313 out of 364 participants (86%) reported that they had used an insecticide-treated net within the last year, but only 19 persons (5%) reported that more pesticide was ever applied to the nets. Unfortunately, we did not ask participants with which insecticides their mosquito net was treated (if so). The insecticides recommended by the WHO for use on mosquito nets are pyrethroid insecticides,[23 24] but any participants retreating their own nets at home may have used different insecticides. Because of the lack of compound-specific data and the low number of persons reporting re-treatment of the nets, it would be hard to validate an exposure score including mosquito nets against biomarkers of insecticide exposure.

TABLE 4: DETERMINISTIC EXPOSURE SCORES FOR DIFFERENT KINDS OF PPE

Exposure pathway	Inhalation		Dermal exposure							Whole body
Body part	Mouth		Eyes	Hands	Upper body		Legs		Feet	
Relative contribution in case no PPE is used	0.1		0.1	0.4	0.2		0.1		0.1	1
Type of PPE	Dust mask	Mask with carbon filter	Goggles	Gloves	Overall or long-sleeved shirt	Rubber apron or rain poncho	Overall or long pants	Gaiters or water-proof pants	Rubber boots	
PPE quality	W	NWP	NWP	W or NWP	W	NWP	W	NWP	NWP	
Fractional exposure when using PPE										
All the time (100%)	0.30	0.10	0.10	0.20	0.30	0.10	0.30	0.10	0.10	
Most of the time (75%)	0.48	0.33	0.33	0.40	0.48	0.33	0.48	0.33	0.33	
Often (50%)	0.65	0.55	0.55	0.60	0.65	0.55	0.65	0.55	0.55	
Rarely (25%)	0.83	0.78	0.78	0.80	0.83	0.78	0.83	0.78	0.78	
Never (0%)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
Actual exposure score: Example for best possible protection	0.03	0.01	0.01	0.08	0.06	0.02	0.03	0.01	0.01	0.14

W = woven. NWP = non-woven permeable
Protection factors from [22]

TABLE 5: DETERMINISTIC EXPOSURE SCORES FOR BEHAVIORAL FACTORS

Variable	Deterministic exposure score	Reference
MIX	5 if mixing pesticides, otherwise 0	Thomas et al 2010[25]
APPLY	8 if applying pesticides, otherwise 0	Thomas et al 2010[25]
CHANGE	Next day: 1	Dosemeci et al 2002[26]
	Many hours later: 0.9	
	A few hours later: 0.8	
	Immediately or within one hour: 0.7	
SHOWER	Next day: 1	Dosemeci et al 2002[26]
	Many hours later: 0.9	
	A few hours later: 0.8	
	Immediately or within one hour: 0.7	

10.5 Description of the empirical model

As described above, the deterministic model will make many assumptions regarding the ways that exposure to organophosphates is related to AChE (time, potency, PPE, hygienic factors). We will therefore also develop a simple empirical model containing fewer variables. The size of the influence of these variables on AChE will be determined in a linear mixed effect model.

As we have too little data to create compound-specific models for all reported organophosphate insecticides, we will calculate the sum of the potency-adjusted number of times individual organophosphate insecticides were used between phases 1 and 3.

EQUATION 10-8

$$TIMES_{total} = \sum_{j} TIMES_j \times potency_{rel,j}$$

Where $TIMES_{total}$ is the potency-adjusted number of times of use, $TIMES_j$ is the number of times that the j^{th} organophosphate insecticide has been used, and $potency_{rel,j}$ is its relative potency. The relative potencies of each organophosphate insecticide can be seen above in section 10.4.2. We will take into account all organophosphate exposure between phases 1 and 3, but not consider the time from exposure to the phase 3 examination.

We will assume that using a specific type of PPE will reduce exposure by a fraction proportional to the frequency of usage of that PPE:

$$INTENSITY_{INTERNAL} = TIMES_{total} \times \left(1 - \propto_{PPE\_0} \times f_{PPE\_0}\right)$$

Where $\propto_{PPE,0}$ is the protection offered by PPE_0 if it was used 100% of the time when handling pesticides, and f_{PPE_0} is the fraction of time that PPE_0 is actually used. This can be rewritten as:

$$INTENSITY_{INTERNAL} = TIMES_{total} - \propto_{PPE\_0} \times f_{PPE\_0} \times TIMES_{total}$$

Based on analyses of Directed Acyclic Graphs (see Appendix C, section 17.1.1) we will also include the three following predictors of AChE:

- Age (continuous variable). Can influence AChE independently of organophosphate exposure,[27] and may determine actual exposure.
- Sex. Can influence AChE independently of organophosphate exposure,[27] and may determine actual exposure.
- Years of schooling (continuous variable, proxy for socioeconomic status).

10.6 Running the analyses in Stata

10.6.1 Primary analysis

10.6.1.1 Deterministic model

To validate the deterministic model, we will first calculate the deterministic score SCORE as detailed in Equation 10-7. We will then run a linear mixed effect model with AChE as the dependent variable and SCORE as the independent variable. Age, sex and years of schooling (proxy for socioeconomic status) will be included as covariates, as analysis of a Directed Acyclic Graph suggests that they are can influence the relationship between exposure and AChE (see Appendix C, section 17.1.1).

SCORE, age and years of schooling will be modelled using restricted cubic splines with four knots (location determined by the distributions of the variables), as implemented in the Stata command 'mkspline':

```
mkspline age_spline = age, cubic nknots(4)
mkspline years_school_spline = years_school, cubic nknots(4)
mkspline score_spline = score, cubic nknots(4)
```

The linear mixed effects model (with fixed effects only) can then be run using the following command:

```
mixed ache c.score_spline* c.age_spline* c.years_school_spline* i.male ||, reml
```

The performance of model will primarily be evaluated using the residual variance, compared to the total AChE variance (lower residual variance is better). But the evaluation will also include QQ-plot and histogram of residuals, as well as Bland-Altmann plots.

10.6.1.2 Empirical model

We only have sufficient data to include one kind of PPE in our empirical model, and we will not make any a priori assumptions about which kind of PPE best predicts exposure. E.g., using a mask with a carbon filter likely protects better than a simple dust mask, but if almost no one uses masks with carbon filters, we cannot predict AChE using this variable. To create the best possible empirical model, we will fit a number of linear fixed effect models, where each model includes a specific type of PPE. Each model will be run using a command of the following structure:

```
gen ppe_adjusted_times = times_total * frequency_ppe
mixed ache c.times_total c.ppe_adjusted_times c.age_spline*
  c.years_school_spline* i.male ||, reml
```

The types of PPE considered are those mentioned in section 10.4.3. We will only build models for types of PPE that have been used by at least 15 people who also reported having used organophosphate insecticides between phases 1 and 3.

Note that because of the limited amount of data available, we will assume that **`times_total`** and **`ppe_adjusted_times`** affect AChE in a linear manner. Age and years of schooling will be modelled using restricted cubic splines.

The performance of each model will be evaluated in the same manner as for the deterministic model. The final deterministic model is the one with the numerically lowest residual variance (we will not consider statistical significance when choosing the best model).

10.6.1.3 Comparison

To determine whether the deterministic or the empirical model has the highest explanatory power, we will compare the residual variance of the two models. The best model is the one with the numerically lowest residual variance. We will not consider whether any difference between the models is statistically significant.

10.6.2 Secondary analysis

As seen above, the primary analyses are multivariate. The models will be supplemented with the following simpler models:

10.6.2.1 Deterministic model with exposure metric only

Stata code

```
mixed ache c.score_spline* ||, reml
```

10.6.2.2 Empirical model with exposure metric only

Stata code

```
mixed ache c.times_total c.ppe_adjusted_times ||, reml
```

10.6.2.3 Model only including non-insecticide determinants of AChE

Stata code

```
mixed ache c.age_spline* c.years_school_spline* i.male ||, reml
```

10.6.3 Sensitivity analyses

10.6.3.1 Sensitivity analysis # 1

In this sensitivity analysis, we will re-analyze the primary deterministic model (described in section 10.6.1.1). But instead of weighting each organophosphate compound by its relative potency for AChE inhibition in RBC, we will use its potency in brain (see Table 3).

10.6.3.2 Sensitivity analysis # 2

In the primary deterministic model, the time from organophosphate exposure to total recovery of AChE is assumed to be 82 days. When we do not know the exact date of an exposure with size E, we will model it as n repeated exposures of size E/n, where n is the number of possible dates where it could have happened.

To test the robustness of the model, it will be re-analyzed under the following four scenarios:

1. Recovery time = 72 days. Otherwise similar to primary deterministic model.
2. Recovery time = 98 days. Otherwise similar to primary deterministic model.
3. Recovery time = 82 days. If an exposure E has happened in an interval of n days, we will model it as a point exposure on the first of these days.
4. Recovery time = 82 days. If an exposure E has happened in an interval of n days, we will model it as a point exposure on the last of these days.

10.6.3.3 Sensitivity analysis # 3

In this sensitivity analysis, we will rerun the primary analysis, assuming linearity between organophosphate insecticide exposure, age and AChE. I.e., we will not use splines. In Stata, the models will be run like this:

```
mixed ache c.score c.age c.years_school i.male ||, reml
mixed ache c.times_total c.ppe_adjusted_times c.age c.years_school i.male ||,
   reml
```

The performance of the models will be evaluated in the same way as in the primary analysis.

10.6.3.4 Sensitivity analysis # 4

In this sensitivity analysis, we will not impute missing information on hygienic practices or use of PPE. Instead, we leave out persons with missing information in any of the variables in the model.

10.6.3.5 Sensitivity analysis # 5

During data collection, AChE measurements were repeated for a few people. This was done when the primary investigator suspected that measurements were erroneous, e.g. in case of very high or very high values for either AChE or hemoglobin, as this could indicate that a mistake might have happened during the sample analysis. While the decision to repeat the analysis was not wittingly based on participant's pesticide exposure levels, it is possible that the decision could have been subconsciously biased. Therefore, in the main analyses we will use the first AChE value for each person in a given phase, no matter if we think that value is correct or not. Any errors in these data are expected to be non-differential and therefore unable to introduce bias into or results. As a sensitivity analysis, we will repeat the primary analysis (recreating both the deterministic and the empirical models), but this time using the AChE value we deem most likely to be correct. This judgment will be based on the consistency between all AChE and hemoglobin values from each phase.

10.6.3.6 Sensitivity analysis # 6

Preliminary descriptive analyses have shown that participants used a considerable amount of other pesticides, apart from organophosphate insecticides (date not shown). To investigate whether participants were able to accurately report compound-specific pesticide usage data, we will repeat the primary analysis, but using all other classes of pesticides than organophosphate insecticides as the relevant exposure. A priory, we do not expect these other pesticides to be able to inhibit AChE. Exposure to different classes of chemical is expected to covariate. However, if participants were able to reliably report compound-specific data, we expect the residual variance in these models to be higher than in the primary analyses.

As we have no a priori expectations that other classes of pesticides than organophosphates (or carbamates) can inhibit AChE, we will not weight the amounts of these other pesticides by any potency factor.

10.6.3.7 Sensitivity analysis # 7

In this sensitivity analysis, we will not only take into account the number of times that participants have handled organophosphate insecticides, but also the amount of each compound used each time.

We will thus define the deterministic exposure score as

$$SCORE = PPE \times CHANGE \times SHOWER \times (MIX + APPLY) \times \sum_{i,j} potency_{rel,j} \times AMOUNT_i \times w_i$$

Where $AMOUNT_i$ is the amount used of organophosphate *j* during exposure *i*.

In the empirical model, we will replace the potency-adjusted number of times of exposure ($TIMES_{total}$) with the potency-adjusted amount of organophosphate insecticide used ($AMOUNT_{total}$):

$$AMOUNT_{total} = \sum_{j} AMOUNT_j \times potency_{rel,j}$$

Where $AMOUNT_j$ is the amount used of the j^{th} insecticide.

10.6.3.8 Sensitivity analysis # 8

As previously described, our outcome of interest is red blood cell acetylcholine esterase activity, expressed in units of enzyme activity per gram hemoglobin (U/g). In phase 1 of the PEXADU project, we noticed an apparent discrepancy between hemoglobin values reported by the Test-Mate AChE system and hemoglobin values measured with a clinical hemoglobinometer in a subsample of participants (data not shown). Our results indicated that the Test-Mate might be underestimating the hemoglobin levels. As the difference was systematic, it was possible to derive an adjustment equation.

To investigate whether inaccuracy in the Test-Mate hemoglobin values could have any considerable effects on our models, we will repeat our primary analyses after normalizing the acetylcholine esterase activity with the adjusted hemoglobin values (calculated using before-mentioned equation).

11 Analysis plan for glycemic regulation

The exposure models used in analyses of pesticide exposure vs. glycemic regulation will depend on results from the analyses of organophosphate exposure vs. AChE activity. The analysis plan for glycemic regulation will therefore be written once these results are ready.

We will present both raw and adjusted results. In our adjusted analyses, we will adjust for age, socioeconomic status and sex. These confounders have been selected a priori based on Directed Acyclic Graphs,[28] as described in Appendix C, section 17.2.

12 Analysis plan for lung function tests

The exposure models used in analyses of pesticide exposure vs. lung function tests will depend on results from the analyses of organophosphate exposure vs. AChE activity. The analysis plan for lung function tests will therefore be written once these results are ready.

We will present both raw and adjusted results. In our adjusted analyses, we will adjust for age, socioeconomic status and sex. These confounders have been selected a priori based on Directed Acyclic Graphs,[28] as described in Appendix C, section 17.3.

13 Analysis plan for validation of Vitalograph copd-6

13.1 Purpose

To evaluate the accuracy and precision on the Vitalograph copd-6 mini-spirometer in a Ugandan population. Spirometry is an important examination in clinical practice, as well as in studies of occupational and environmental determinants of poor lung function. However, diagnostic-quality spirometers can be prohibitively expensive for use in developing countries such as Uganda. The mini-spirometer Vitalograph copd-6[29] is marketed as screening device for COPD[30] and is much cheaper than diagnostic-quality spirometers. We wanted to examine whether the copd-6 is sufficiently accurate and precise to allow future studies on pulmonary health in Uganda to rely only on the copd-6.

13.2 Statistical procedures

After a test, the copd-6 reports FEV_1 and FEV_6. Analyses will be based on FEV_1 and FEV_6 as continuous variables. Due to a low number of study participants with airway obstruction, we have insufficient statistical strength to examine the sensitivity and specificity of the copd-6 for the diagnosis of obstruction.

13.2.1 Descriptive statistics

In phase 1, participants eligible for spirometry performed spirometry both with the diagnostic-quality device MicroMedical MicroDL and with the copd-6.

We will start by creating a table with the following descriptive metrics:

	MicroDL	**copd-6**		
		Before MicroDL	**After MicroDL**	**Total**
Number of persons	n	n	n	n
Number of blows per person	Median (95% PI)	Median (95% PI)	Median (95% PI)	Median (95% PI)
FEV_1 (l)	Median (95% PI)	Median (95% PI)	Median (95% PI)	Median (95% PI)
FEV_6 (l)	[Not reported by device]	Median (95% PI)	Median (95% PI)	Median (95% PI)
FVC (l)	Median (95% PI)	[Not reported by device]	[Not reported by device]	[Not reported by device]

It should be noted that the MicroDL reports FEV_1 and FVC, while the copd-6 reports FEV_1 and FEV_6.

"Before MicroDL" and "After MicroDL" means that a participant was tested with the copd-6 before or after being tested with the MicroDL, respectively. Data are presented separately for the two conditions, as the order of testing may influence results (learning effects, fatigue, etc.).

Continuous data will be presented as median rather than mean/average, as data may be non-normally distributed. Non-normally distributed data will be transformed to obtain normality, and summary metrics back-transformed to the original scale.

13.2.2 Primary analysis

For each participant, we will calculate the difference in FEV_1 measured by the copd-6 and the MicroDL:

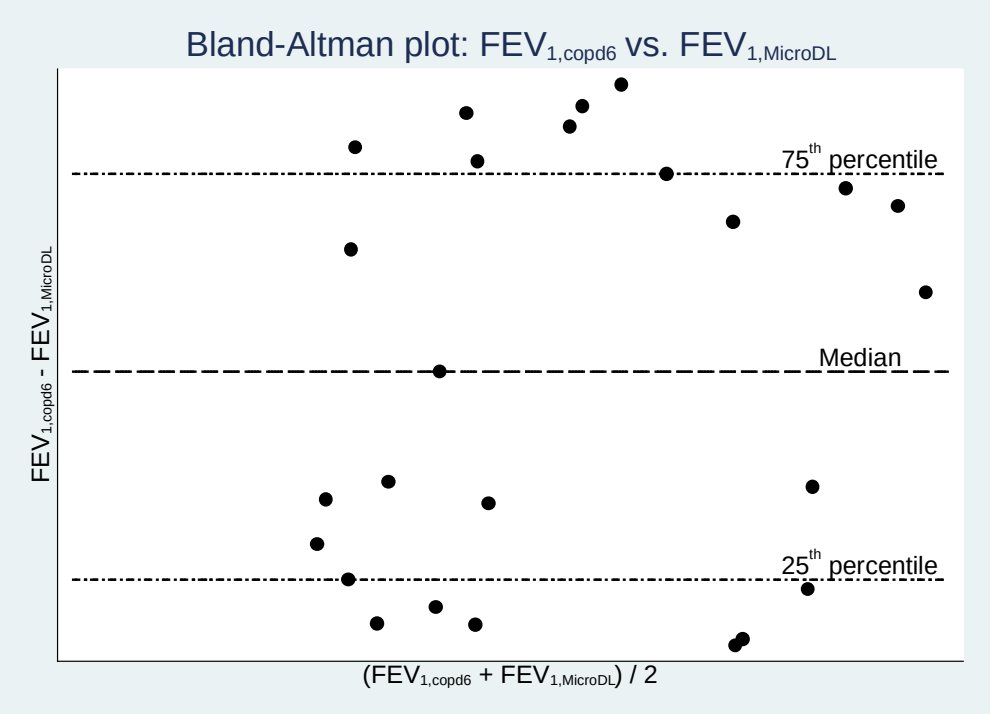
$$\Delta FEV_1 = FEV_{1,copd6} - FEV_{1,MicroDL}$$

We will report the following summary measures for ΔFEV_1 :

- Median
- Range
- 95% confidence interval for median
- Interquartile range
- 95% prediction interval for median

The median with 95% confidence interval will serve as an overall test of statistically significant differences in FEV_1 between copd-6 and MicroDL.

Trends in ΔFEV_1 as a function of FEV_1 will be depicted using a Bland-Altman plot as shown below (the data in the example figure are pseudo-random and do not come from the actual dataset):



To assess whether any bias in measurements from the copd-6 depends of the size of FEV_1, we will perform a linear regression with ΔFEV_1 as dependent variable and $\overline{FEV_1} = FEV_{1,copd6} - FEV_{1,MicroDL}$ as independent variable. We will take non-linearity into account by modelling $\overline{FEV_1}$ using restricted cubic splines with four knots. The location on the knots will depend on the distribution of the data, as implemented in the Stata command `mkspline`.

13.2.3 Secondary analyses

13.2.3.1 Analysis stratified by order of testing

Because of learning effects, fatigue, etc., results may depend on the order of testing (MicroDL first or copd-6 first). Therefore, the primary analysis will be repeated, stratified by order of testing.

13.2.3.2 Analysis of calibration check data

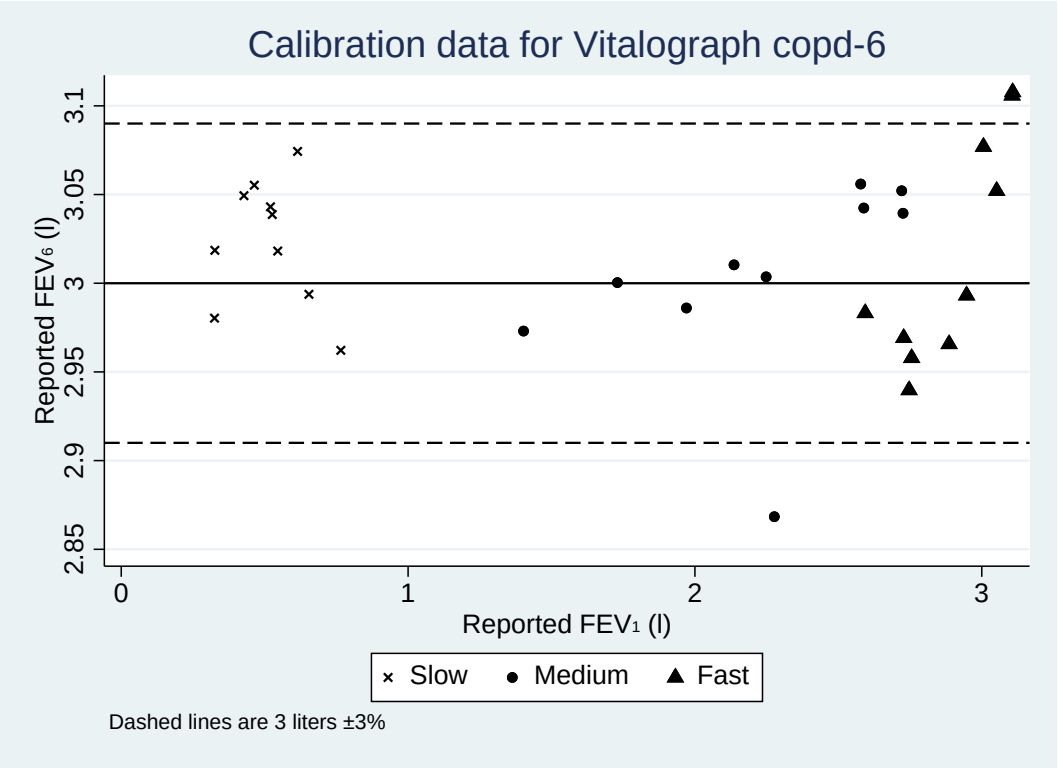
On each day of spirometric testing in phase 1, the calibration of the copd-6 devices was checked using a 3-liter calibration syringe. The syringe was emptied under three conditions:

1) Fast: As fast as possible without banging the piston against the wall of the syringe.
2) Slow: As slow as possible while finishing within 6 seconds (the copd-6 gave an auditory signal after 6 seconds).
3) Medium: Piston pushed at speed in-between 1 and 2

We emptied the syringe three times under each condition and recorded both the FEV_1, FEV_6 and FEV_1/FEV_6 reported by the device.

According to ATS criteria, spirometers must measure volumes with an accuracy of ±3 percent[15]. We will check the validity of the copd-6 by creating a scatterplot of the reported FEV_6 (which should theoretically be 3.00 liters, if there was zero imprecision) as a function of the reported FEV_1 (which is a measure of the speed at which the piston was pushed). The data points will have different symbols to show whether the test

condition was fast/medium/slow. An example with pseudo-random data is shown below:



In addition to the graphical representation, precision and validity will also be assessed by drawing the following table:

	Subjective flow speed			Total
	Slow	Medium	Fast	
Measurements	n	n	n	n
Reported FEV_1 (l)	Median (95% CI)	Median (95% CI)	Median (95% CI)	Median (95% CI)
Reported FEV_6 (l)	Median (95% CI)	Median (95% CI)	Median (95% CI)	Median (95% CI)
Measurements with $FEV_6 < 2.91$ l	n (%)	n (%)	n (%)	n (%)
Measurements with $FEV_6 > 3.09$ l	n (%)	n (%)	n (%)	n (%)

Any trend in accuracy as a function of subjective flow speed will be assessed by Spearman's rank correlation of reported FEV_6 vs. flow speed (1 = slow, 2 = medium, 3 = fast).

13.2.3.3 Analysis based on FEV_6 and FVC instead of FEV_1

We will repeat the primary analysis, based on FEV_6 and FVC instead of FEV_1. The Vitalograph copd-6 device reports FEV_6, while the MicroMedical MicroDL reports FVC. Those two lung function indices are not directly comparable, unless a blow has taken ≤ 6 seconds from start to finish. A blow with duration ≤ 6 seconds, where

the volume-time curve has reached a plateau at the end of the blow, will have FEV_6 = FVC. Therefore, this analysis will be limited to those persons where all acceptable blows with the MicroDL took ≤ 6 seconds (without adjusting for slow starts).

13.2.4 Sensitivity analyses

13.2.4.1.1 REPEATED PRIMARY ANALYSIS, STRATIFIED BY DEVICES USED

Two different copd-6 devices and a number of different MicroDL devices were used in the project. To check whether any imprecision or inaccuracy in the copd-6 results was due to differences in calibration or any faulty devices (copd-6 or MicroDL), we will repeat the primary analysis stratified by the combination of devices used. E.g., copd-6 device number 1 + MicroDL device number 3. Only stratums with at least 15 observations will be analyzed. Out of these, linear regression analysis of ΔFEV_1 as a function of $\overline{FEV_1}$ will only be carried out if there is at least 4 × 15 = 60 observations in the stratum.

13.2.4.1.2 ANALYSIS OF CALIBRATION CHECK DATA, STRATIFIED BY DAY AND DEVICE

Two different copd-6 devices were used in the project. To check whether any imprecision or inaccuracy in the calibration check data was due to temperature differences between days, spirometer turbines wearing down, differences in device calibration etc., the analyses of calibration check data will be repeated with simultaneous stratification by day of testing and device used. This analysis has already been carried out at the time of writing (June 11, 2019).

13.2.4.1.3 REPEATED PRIMARY ANALYSIS, WITH ALTERNATIVE SPIROMETRY QUALITY CRITERIA

When participants performed spirometry with the MicroMedical MicroDL device, they always got five attempts to start with. If, despite coaching from the nurse, their test did not fulfill ATS quality criteria[15] after these five attempts, they were given an additional four attempts. The nurse coached participants based both on her/his observations of the participant during the test, and on the spirograms displayed by the spirometry software. When participants were tested with the Vitalograph copd-6 device, they were always given three attempts, no matter the quality of these. Coaching for the use of the copd-6 was based exclusively on the nurse's observations of the participant, as the copd-6 cannot display spirograms. While the copd-6 manual does suggest testing until three good blows have been performed[30], for pragmatic reasons we only asked participant to blow three times. This approach was chosen because we did not want to tire the participants (they already had to blow 8-12 times). The types of problems that the copd-6 can detect are slow start and cough.[30]

In the primary analysis described above, $FEV_{1,MicroDL}$ will be based on all 5 or 9 blows from the MicroDL, with standard ATS quality criteria.[15] The $FEV_{1,copd6}$ will be the best FEV_1 recorded with the copd-6, and will only be calculated if all three blows performed were OK according to the automatic classification by the device.

Hence, in the primary analysis we are comparing both across different devices and across different ways of testing.

In this sensitivity analysis, we will apply similar quality criteria to both MicroDL and copd-6, to investigate if any discrepancies between the two devices in the primary analysis are due to different testing protocols or due to differences between the devices *per se*. For the MicroDL, we will only use the first three blows performed, and exclude blows where spirograms show slow start or cough. For the copd-6, we will use all three blows and exclude blows where the device showed a warning (because of either slow start or cough). We will base our analyses on only those participants who have three blows with the MicroDL and three blows with the copd-6 fulfilling these criteria. The values that we will compare are the best FEV_1 out of the three measurements from each device. We will present overall results (corresponding to the analyses in section 13.2.2) and stratified by the order of testing (corresponding to the analyses in section 13.2.3.1).

14 Ethical considerations

This study has been approved by the "Higher Degrees, Research and Ethics Committee" at Makerere University School of Public Health, Kampala, Uganda (protocol number 577). Is has also been approved by the "Uganda National Council for Science and Technology", Kampala, Uganda (HS234ES). Since the proposed project is not carried out in Denmark, we were informed by The National Committee on Health Research Ethics in Denmark that approving the project fell outside their jurisdiction. The project has been registered with the Danish Data Protection Agency (Datatilsynet, www.datatilsynet.dk/english).

The project was carried out in accordance with the Declaration of Helsinki. Participation was voluntary, and all participants signed an informed consent form before inclusion. Participant information was given in English or local language (Luganda) as appropriate. Participants were compensated for lost earnings, as well as travel and lunch expenses on the day of examination.

At the time of writing, the biological materials collected during the PEXADU project were still stored in Uganda. We are in the process of getting permission from the Uganda National Council for Science and Technology (UNCST) to export the samples to Denmark for storage and analysis.

References

1 Hansen MRH, Jørs E, Sandbæk A et al. O1A.3 Occupational insecticide exposure, glycemic regulation and bronchoconstriction: preliminary results from a short-term cohort study among small-scale farmers in Uganda. *Occupational and Environmental Medicine* 2019;76:A3-A3.

2 Evangelou E, Ntritsos G, Chondrogiorgi M et al. Exposure to pesticides and diabetes: A systematic review and meta-analysis. *Environ Int* 2016;91:60-8.

3 WHO. Diabetes country profiles - Uganda 2016.

4 Malekirad AA, Faghih M, Mirabdollahi M, Kiani M, Fathi A, Abdollahi M. Neurocognitive, Mental Health, and Glucose Disorders in Farmers Exposed to Organophosphorus Pesticides. *Archives of Industrial Hygiene and Toxicology* 2013;64:1-8.

5 Velmurugan G, Ramprasath T, Swaminathan K et al. Gut microbial degradation of organophosphate insecticides-induces glucose intolerance via gluconeogenesis. *Genome Biol* 2017;18:8.

6 van Gemert F, Kirenga B, Chavannes N et al. Prevalence of chronic obstructive pulmonary disease and associated risk factors in Uganda (FRESH AIR Uganda): a prospective cross-sectional observational study. *The Lancet Global Health* 2015;3:e44-e51.

7 Doust E, Ayres JG, Devereux G et al. Is pesticide exposure a cause of obstructive airways disease? *Eur Respir Rev* 2014;23:180-92.

8 Negatu B, Kromhout H, Mekonnen Y, Vermeulen R. Occupational pesticide exposure and respiratory health: a large-scale cross-sectional study in three commercial farming systems in Ethiopia. *Thorax* 2017;72:498-499.

9 Casida JE. Pest toxicology: the primary mechanisms of pesticide action. *Chem Res Toxicol* 2009;22:609-19.

10 Peter JV, Sudarsan TI, Moran JL. Clinical features of organophosphate poisoning: A review of different classification systems and approaches. *Indian Journal of Critical Care Medicine : Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine* 2014;18:735-745.

11 World Health Organization. STEPwise approach to surveillance (STEPS).

12 Colagiuri S. Glycated haemoglobin (HbA1c) for the diagnosis of diabetes mellitus--practical implications. *Diabetes Res.Clin.Pract.* 2011;93:312-313.

13 O'Connell SG, Kincl LD, Anderson KA. Silicone wristbands as personal passive samplers. *Environ Sci Technol* 2014;48:3327-35.

14 Donald CE, Scott RP, Blaustein KL et al. Silicone wristbands detect individuals' pesticide exposures in West Africa. *R Soc Open Sci* 2016;3:160433.

15 Miller MR, Hankinson J, Brusasco V et al. Standardisation of spirometry. *European respiratory journal* 2005;26:319-338.

16 Armstrong RA. When to use the Bonferroni correction. *Ophthalmic Physiol Opt* 2014;34:502-8.

17 Fukuto TR. Mechanism of action of organophosphorus and carbamate insecticides. *Environmental health perspectives* 1990;87:245-254.

18 Mason H. The recovery of plasma cholinesterase and erythrocyte acetylcholinesterase activity in workers after over-exposure to dichlorvos. *Occupational Medicine* 2000;50:343-347.

19 U.S. Environmental Protection Agency, Office of Pesticide Programs. Organophosphorus Cumulative Risk Assessment - 2006 Update 2006.

20 Office of Pesticide Programs, United States Environmental Protection Agency. Preliminary Cumulative Hazard and Dose-response Assessment for Organophorus Pesticides: Determination of Relative Potency and Points of Departure for Cholinesterase Inhibition, Washington, D.C., United States 2001.

21 Geldenhuys WJ, Mohammad AS, Adkins CE, Lockman PR. Molecular determinants of blood-brain barrier permeation. *Ther Deliv* 2015;6:961-71.

22 Samuel Fuhrimann PS, Christian H Lindh, Berna van Wendel de Joode, Ana M Mora, Mirko S Winkler, Hans Kromhout. Variability and predictors of weekly pesticide exposure in applicators from organic, sustainable, and conventional smallholder farms in Costa Rica (article under review) 2019.

23 Adams P. Preserving pyrethroids. *World Health Organization. Bulletin of the World Health Organization* 2014;92:158.

24 Organization WH. WHO recommended long-lasting insecticidal mosquito nets. *Geneva: World Health Organization* 2011.

25 Thomas KW, Dosemeci M, Coble JB et al. Assessment of a pesticide exposure intensity algorithm in the agricultural health study. *J Expo Sci Environ Epidemiol* 2010;20:559-69.

26 Dosemeci M, Alavanja MC, Rowland AS et al. A quantitative approach for estimating exposure to pesticides in the Agricultural Health Study. *Ann Occup Hyg* 2002;46:245-60.

27 Whittaker M. 9. Cholinesterase in Clinical Medicine *Cholinesterase*: Karger Publishers 1986;65-85.

28 Suttorp MM, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Graphical presentation of confounding in directed acyclic graphs. *Nephrol Dial Transplant* 2015;30:1418-23.

29 Vitalograph. Vitalograph 4000 series - Respiratory monitors & screening devices 2018.

30 -. Respiratory monitor copd-6 model 4000 user training manual, Ennis, Ireland 2013.

31 Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol* 2016;45:1887-1894.

15 Appendix A: Deriviation of relative potencies for red blood cell acetylcholine esterase

```
-------------------------------------------------------------------------------------------------------
      name:  <unnamed>
       log:  C:\Users\au231481\Desktop\Potency EPA\predict_rbc_ache.log
  log type:  text
 opened on:   2 Aug 2019, 16:30:52

.
. //      Open table of brain ChE potency from USEPA 2006. List the data.
.         import excel using "brain_che_potency_USEPA_2006.xlsx", firstrow clear

.         rename Chemical chemical

.         rename Oralrelativepotencybrain potency_brain

.         label variable potency_brain "Relative potency, brain"

.         sort chemical

.         list, ab(32)

     +-----------------------------------+
     |            chemical   potency_brain |
     |-----------------------------------|
  1. |            Acephate             .08 |
  2. |     Azinphos-methyl              .1 |
  3. |           Bensulide            .003 |
  4. |      Chlorethoxyfos             .13 |
  5. |        Chlorpyrifos             .06 |
     |-----------------------------------|
  6. | Chlorpyrifos-methyl            .005 |
  7. |            Diazinon             .01 |
  8. |          Dichlorvos             .03 |
  9. |         Dicrotophos            1.91 |
 10. |          Dimethoate              .32 |
     |-----------------------------------|
 11. |          Disulfoton            1.26 |
 12. |            Ethoprop             .06 |
 13. |          Fenamiphos             .04 |
 14. |            Fenthion             .33 |
 15. |         Fosthiazate             .07 |
     |-----------------------------------|
 16. |           Malathion           .0003 |
 17. |       Methamidophos                1 |
 18. |        Methidathion             .32 |
 19. |    Methyl-parathion             .12 |
 20. |           Mevinphos             .76 |
     |-----------------------------------|
 21. |               Naled             .08 |
 22. |           Omethoate             .93 |
 23. |   Oxydemeton-methyl             .86 |
 24. |             Phorate             .39 |
 25. |           Phosalone             .01 |
     |-----------------------------------|
 26. |            Phosmet             .02 |
 27. |       Phostebupirim             .22 |
 28. |   Pirimiphos-methyl             .04 |
 29. |          Profenofos            .004 |
 30. |            Terbufos             .85 |
     |-----------------------------------|
 31. |   Tetrachlorvinphos            .001 |
 32. |            Tribufos              .02 |
 33. |         Trichlorfon            .003 |
     +-----------------------------------+

.
. //      Save temporary file.
.         tempfile myFile

.         save `myFile', replace
```

```
(note: file C:\Users\au231481\AppData\Local\Temp\ST_3ea0_000001.tmp not found)
file C:\Users\au231481\AppData\Local\Temp\ST_3ea0_000001.tmp saved

.
. //      Open table of RBC ChE potency from USEPA 2002. List the data.
.         import excel using "rbc_che_potency_USEPA_2002.xlsx", firstrow clear

.         de,f

Contains data
  obs:            49
 vars:             6
 size:         2,107
-------------------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------------------
chemical        str17   %17s                  chemical
sex             str1    %9s                   sex
n               byte    %10.0g                n
potency_rbc     double  %10.0g                potency_rbc
potency_rbc_ll  double  %10.0g                potency_rbc_ll
potency_rbc_ul  double  %10.0g                potency_rbc_ul
-------------------------------------------------------------------------------------------------------------
Sorted by:
     Note: Dataset has changed since last saved.

.
.         label variable potency_rbc "Relative potency, RBC"

.         label variable potency_rbc_ll "Relative potency, RBC (lower limit of 95% CI)"

.         label variable potency_rbc_ul "Relative potency, RBC (upper limit of 95% CI)"

.         label variable n "Number of data points"

.
.         sort chemical sex

.         list, ab(32)

     +-------------------------------------------------------------------------------+
     |          chemical   sex    n   potency_rbc   potency_rbc_ll   potency_rbc_ul |
     |-------------------------------------------------------------------------------|
  1. |          Acephate     F   15         .0216           .00906            .0517 |
  2. |          Acephate     M   15         .0207            .0094            .0455 |
  3. |   Azinphos-methyl     F    8          .349             .148             .821 |
  4. |   Azinphos-methyl     M    8          .351             .199             .619 |
  5. |         Bensulide     F    5         .0113           .00974            .0132 |
     |-------------------------------------------------------------------------------|
  6. |      Chlorpyrifos     F    9         .0894            .0153              .52 |
  7. |      Chlorpyrifos     M    9          .102            .0511             .206 |
  8. |          Diazinon     F   12          .269             .103             .703 |
  9. |          Diazinon     M   12          .145             .036             .585 |
 10. |        Dichlorvos     F    7           .23              .07              .78 |
     |-------------------------------------------------------------------------------|
 11. |        Dichlorvos     M    6           .14               .1              .21 |
 12. |        Dimethoate     F    9          .392             .203             .757 |
 13. |        Dimethoate     M    9          .431             .278             .666 |
 14. |        Disulfoton     F   10          4.87             4.43             5.36 |
 15. |        Disulfoton     M   10          3.55             2.94             4.28 |
     |-------------------------------------------------------------------------------|
 16. |        Fenamiphos     F    9          .753               .6              .95 |
 17. |        Fenamiphos     M    8           .56              .44              .72 |
 18. |       Fosthiazate     F   12          .432             .311               .6 |
 19. |       Fosthiazate     M   10          .265             .122             .579 |
 20. |         Malathion     F    7         .0041           .00312           .00539 |
     |-------------------------------------------------------------------------------|
 21. |         Malathion     M    7        .00424           .00293           .00613 |
 22. |     Methamidophos     F   10             1               .7             1.44 |
 23. |     Methamidophos     M   10          1.23              .82             1.83 |
 24. |      Methidathion     F    8           .29              .05             1.63 |
 25. |      Methidathion     M    7           .25              .05             1.14 |
     |-------------------------------------------------------------------------------|
```

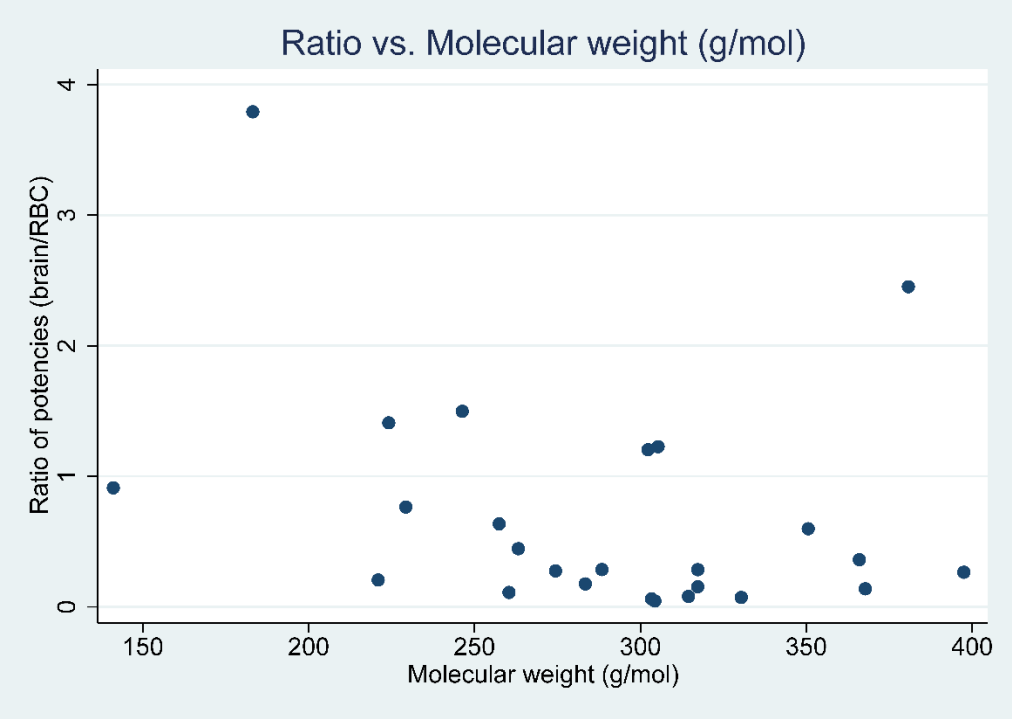
```
 26. |  Methyl-parathion     F   10          .249            .0607            1.02 |
 27. |  Methyl-parathion     M   10          .303            .0525            1.75 |
 28. |         Mevinphos     F    5          .602            .0988            3.67 |
 29. |         Mevinphos     M    5           .46            .0527            4.01 |
 30. |             Naled     F    5           .05            .0263             .113 |
     |-------------------------------------------------------------------------------|
 31. |             Naled     M    5           .03            .0235            .0449 |
 32. | Oxydemeton-methyl     F    9          .448             .251              .8 |
 33. | Oxydemeton-methyl     M   15          .994              .51             1.94 |
 34. |           Phorate     F    5          4.49             3.88              5.2 |
 35. |           Phorate     M    5          3.02             2.69             3.39 |
     |-------------------------------------------------------------------------------|
 36. |         Phosalone     F    8           .05              .02              .14 |
 37. |         Phosalone     M    7           .09              .04               .2 |
 38. |           Phosmet     F    7           .14              .12              .18 |
 39. |           Phosmet     M    7           .12               .1              .15 |
 40. | Pirimiphos-methyl     F   16          .034            .0244            .0475 |
     |-------------------------------------------------------------------------------|
 41. | Pirimiphos-methyl     M   16         .0319             .025            .0407 |
 42. |          Terbufos     F   17          2.01              .45             8.94 |
 43. |          Terbufos     M   17          3.79             1.17             12.3 |
 44. | Tetrachlorvinphos     F    6        .00534           .00152            .0188 |
 45. | Tetrachlorvinphos     M    2        .00246          .000979           .00616 |
     |-------------------------------------------------------------------------------|
 46. |          Tribufos     F    6           .18              .09              .34 |
 47. |          Tribufos     M    6           .28              .19              .42 |
 48. |       Trichlorfon     F    7        .00457           .00216           .00967 |
 49. |       Trichlorfon     M    5        .00479           .00259           .00884 |
     +-------------------------------------------------------------------------------+

.
. /*      We note that for Bensulide we only have data points for female rats. The USEPA 2002 publication
>         does not list numerical results for Bensulide for male rats (I don't want to try and read value
> s
>         from a graph with a logarithmic axis).  */
.
. /*      Perform fixed-effect meta-analysis using inverse variance weights to get common (rather than
>         gender-specific) estimates.     */
.         gen log_est = log(potency_rbc)

.         gen log_ll = log(potency_rbc_ll)

.         gen log_ul = log(potency_rbc_ul)

.         gen log_se = (log_ul - log_ll) / (2 * 1.96)

.         gen log_sd = log_se * sqrt(n)

.         gen log_variance = log_sd^2

.         gen w = 1/log_variance

.         gen weighted_log_est = w * log_est

.         collapse (sum) weighted_log_est w, by(chemical)

.         gen log_est = weighted_log_est / w

.         gen potency_rbc = exp(log_est)

.         keep chemical potency_rbc

.         label variable potency_rbc "Relative potency, RBC"

.         list, ab(32)

     +--------------------------------+
     |          chemical   potency_rbc |
     |--------------------------------|
  1. |          Acephate      .0211007 |
  2. |   Azinphos-methyl       .350389 |
  3. |         Bensulide         .0113 |
  4. |      Chlorpyrifos      .1001979 |
  5. |          Diazinon      .2204856 |
```

```
     |--------------------------------|
  6. |        Dichlorvos      .1453171 |
  7. |        Dimethoate      .4186764 |
  8. |        Disulfoton      4.564723 |
  9. |        Fenamiphos      .6503644 |
 10. |       Fosthiazate      .3963854 |
     |--------------------------------|
 11. |         Malathion      .0041491 |
 12. |     Methamidophos      1.096886 |
 13. |      Methidathion      .2658181 |
 14. |  Methyl-parathion      .2689685 |
 15. |         Mevinphos      .5390574 |
     |--------------------------------|
 16. |             Naled      .0326339 |
 17. | Oxydemeton-methyl      .5740668 |
 18. |            Phorate      3.517057 |
 19. |         Phosalone      .0722202 |
 20. |           Phosmet      .1296148 |
     |--------------------------------|
 21. | Pirimiphos-methyl       .032617 |
 22. |          Terbufos      2.973561 |
 23. | Tetrachlorvinphos      .0027661 |
 24. |          Tribufos      .2493234 |
 25. |       Trichlorfon      .0047176 |
     +--------------------------------+

.
. //      Merge the two files
.         merge 1:1 chemical using `myFile'
(note: variable chemical was str17, now str19 to accommodate using data's values)

    Result                           # of obs.
    -----------------------------------------
    not matched                             8
        from master                         0  (_merge==1)
        from using                          8  (_merge==2)

    matched                                25  (_merge==3)
    -----------------------------------------

.         sort _merge

.         drop _merge

.
. //      There are eight compounds for which we have potency data for brain, but not RBC
.         list, ab(32)

     +-------------------------------------------------+
     |            chemical   potency_rbc   potency_brain |
     |-------------------------------------------------|
  1. |          Profenofos             .            .004 |
  2. |       Phostebupirim             .             .22 |
  3. |           Omethoate             .             .93 |
  4. |            Fenthion             .             .33 |
  5. |            Ethoprop             .             .06 |
     |-------------------------------------------------|
  6. |         Dicrotophos             .            1.91 |
  7. | Chlorpyrifos-methyl             .            .005 |
  8. |      Chlorethoxyfos             .             .13 |
  9. |         Trichlorfon      .0047176            .003 |
 10. |            Tribufos      .2493234             .02 |
     |-------------------------------------------------|
 11. |   Tetrachlorvinphos      .0027661            .001 |
 12. |           Terbufos      2.973561             .85 |
 13. |   Pirimiphos-methyl       .032617             .04 |
 14. |             Phosmet      .1296148             .02 |
 15. |          Phosalone      .0722202             .01 |
     |-------------------------------------------------|
 16. |            Phorate      3.517057             .39 |
 17. |   Oxydemeton-methyl      .5740668             .86 |
 18. |              Naled      .0326339             .08 |
 19. |          Mevinphos      .5390574             .76 |
 20. |    Methyl-parathion      .2689685             .12 |
```

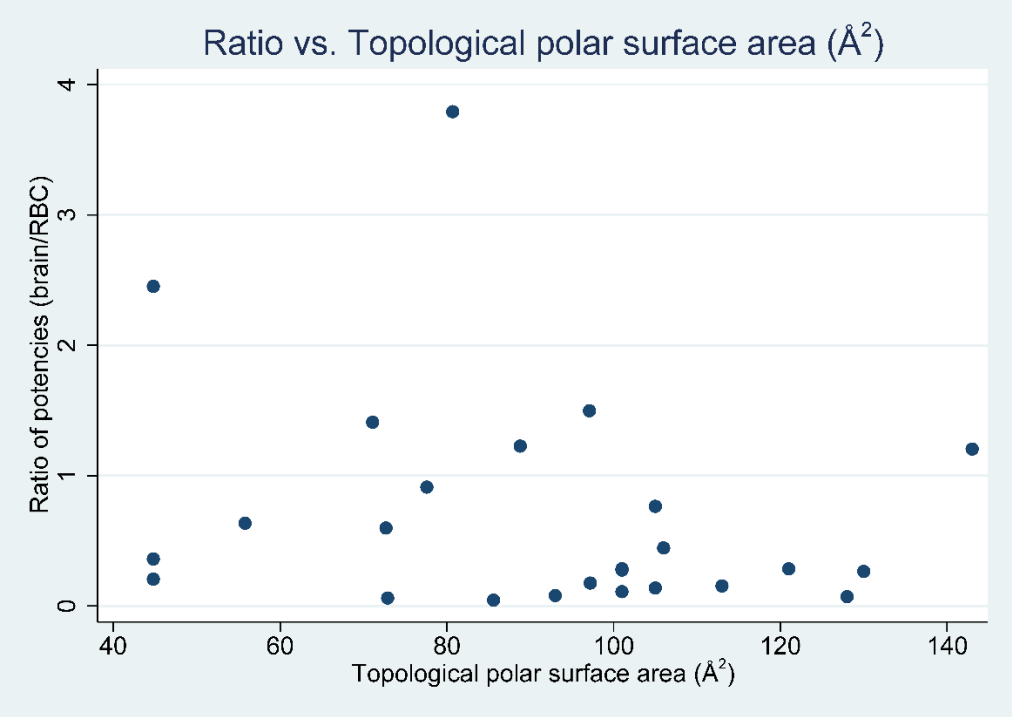
```
     |-------------------------------------------------|
 21. |        Methidathion      .2658181             .32 |
 22. |       Methamidophos      1.096886               1 |
 23. |           Malathion      .0041491           .0003 |
 24. |         Fosthiazate      .3963854             .07 |
 25. |          Fenamiphos      .6503644             .04 |
     |-------------------------------------------------|
 26. |          Disulfoton      4.564723            1.26 |
 27. |          Dimethoate      .4186764             .32 |
 28. |          Dichlorvos      .1453171             .03 |
 29. |            Diazinon      .2204856             .01 |
 30. |        Chlorpyrifos      .1001979             .06 |
     |-------------------------------------------------|
 31. |           Bensulide         .0113            .003 |
 32. |     Azinphos-methyl       .350389              .1 |
 33. |            Acephate      .0211007             .08 |
     +-------------------------------------------------+

.
. //      Generate a variable equal to the ratio between the brain and the RBC potencies
.         gen ratio = potency_brain / potency_rbc
(8 missing values generated)

.         label variable ratio "Ratio of potencies (brain/RBC)"

.
. //      Save temporary file
.         save `myFile', replace
file C:\Users\au231481\AppData\Local\Temp\ST_3ea0_000001.tmp saved

.
. //      Add data on physicochemical properties of organophosphates from PubChem.
.         import delimited using "pubchem_data.csv", clear encoding("utf-8")
(6 vars, 33 obs)

.         label variable molecularweight "Molecular weight (g/mol)"

.         label variable tpsa "Topological polar surface area (Å{sup:2})"

.         label variable xlogp "XLogP"

.         label variable hbonddonorcount "Hydrogen bond donor count"

.         de,f

Contains data
  obs:            33
 vars:             6
 size:         1,188
-------------------------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------------------------------------
chemical        str19   %19s
pubchem_id      long    %12.0g
molecularweight float   %9.0g                 Molecular weight (g/mol)
tpsa            float   %9.0g                 Topological polar surface area (Å{sup:2})
xlogp           float   %9.0g                 XLogP
hbonddonorcount byte    %8.0g                 Hydrogen bond donor count
-------------------------------------------------------------------------------------------------------------
Sorted by:
     Note: Dataset has changed since last saved.

.
.         list, ab(32)

     +-------------------------------------------------------------------------------+
     |            chemical   pubchem_id   molecularweight   tpsa   xlogp   hbonddonorcount |
     |-------------------------------------------------------------------------------|
  1. |            Acephate         1982            183.17   80.7     -.8                 1 |
  2. |     Azinphos-methyl         2268             317.3    121     2.8                 0 |
  3. |           Bensulide        12932             397.5    130     4.2                 1 |
  4. |      Chlorethoxyfos        91655               336   59.8     4.6                 0 |
  5. |        Chlorpyrifos         2730             350.6   72.7     5.3                 0 |
```

```
     |-------------------------------------------------------------------------------|
  6. | Chlorpyrifos-methyl        21803             322.5   72.7     4.3           0 |
  7. |            Diazinon         3017            304.35   85.6     3.8           0 |
  8. |          Dichlorvos         3039            220.97   44.8     1.4           0 |
  9. |         Dicrotophos      5371560            237.19   65.1       0           0 |
 10. |          Dimethoate         3082             229.3    105      .8           1 |
     |-------------------------------------------------------------------------------|
 11. |          Disulfoton         3118             274.4    101       4           0 |
 12. |            Ethoprop         3289             242.3   76.9     3.6           0 |
 13. |          Fenamiphos        31070            303.36   72.9     3.2           1 |
 14. |            Fenthion         3346             278.3   85.1     4.1           0 |
 15. |         Fosthiazate        91758             283.4   97.2       2           0 |
     |-------------------------------------------------------------------------------|
 16. |           Malathion         4004             330.4    128     2.4           0 |
 17. |       Methamidophos         4096            141.13   77.6     -.9           1 |
 18. |        Methidathion        13709             302.3    143     2.4           0 |
 19. |    Methyl-parathion         4130            263.21    106     2.9           0 |
 20. |           Mevinphos      5355863            224.15   71.1     1.2           0 |
     |-------------------------------------------------------------------------------|
 21. |               Naled         4420            380.78   44.8     2.5           0 |
 22. |           Omethoate        14210            213.19   89.9     -.9           1 |
 23. |   Oxydemeton-methyl         4618             246.3   97.1     -.7           0 |
 24. |             Phorate         4790             260.4    101     3.6           0 |
 25. |           Phosalone         4793             367.8    105     4.4           0 |
     |-------------------------------------------------------------------------------|
 26. |             Phosmet        12901             317.3    113     2.8           0 |
 27. |       Phostebupirim        93516            318.37   85.6     4.2           0 |
 28. |   Pirimiphos-methyl        34526            305.34   88.8     4.2           0 |
 29. |          Profenofos        38779            373.63   60.8     4.7           0 |
 30. |            Terbufos        25670             288.4    101     4.5           0 |
     |-------------------------------------------------------------------------------|
 31. |   Tetrachlorvinphos      5284462               366   44.8     3.5           0 |
 32. |            Tribufos         5125             314.5     93     3.2           0 |
 33. |         Trichlorfon         5853            257.43   55.8      .5           1 |
     +-------------------------------------------------------------------------------+

.
. /*      These properties were selected as a priori candidates for the relationship between brain and
>         RBC potency for OP insecticides. The choice was made based on the following article that
>         describes determinants of whether a compound can cross the blood-brain barrier:
>
>         Geldenhuys, Werner J., et al. "Molecular determinants of blood-brain barrier permeation"
>         Therapeutic delivery 6.8 (2015): 961-971. DOI: 10.4155/tde.15.32
>
>         The article also mentions pKa as an important property, but that was deemed irrelevant based
>         on the structure of the compounds.
>         */
.
. //      Merge with the main dataset
.         merge 1:1 chemical using `myFile', nogen

    Result                           # of obs.
    -----------------------------------------
    not matched                             0
    matched                                33
    -----------------------------------------

.
. //      Summarize the physicochemical properties
.         su molecularweight tpsa xlogp

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
molecularw~t |         33    289.4324     59.4423     141.13      397.5
        tpsa |         33    87.17576    24.79146       44.8        143
       xlogp |         33    2.660606    1.833361        -.9        5.3

.         tab hbonddonorcount, missing

   Hydrogen |
 bond donor |
      count |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         26       78.79       78.79
```

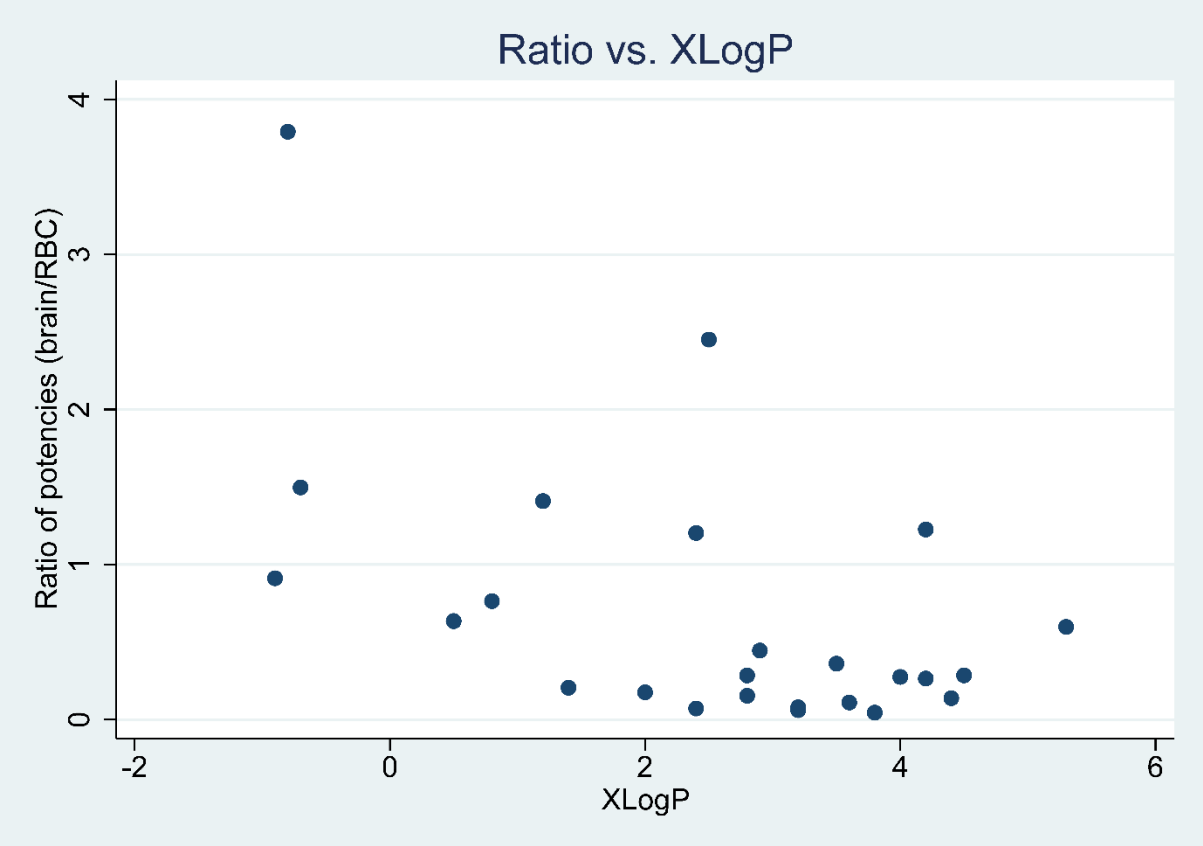
```
          1 |          7       21.21      100.00
------------+-----------------------------------
      Total |         33      100.00

.
. /*      We notice how we have too little data to meaningfully include the number of hydrogen bond donor
> s
>         in any models. We will therefore ignore the variable hbonddonorcount from now on.       */
.
. /*      Draw scatterplots of each of the three continuous properties vs. the ratio of potencies. These
>         graphs will help us determine the best modelling strategy.      */
.         graph drop _all

.         foreach v of var molecularweight tpsa xlogp {
  2.                 local title: variable label `v'
  3.                 twoway scatter ratio `v', title("Ratio vs. `title'") xsize(11.7) ysize(8.3)
  4.                 graph export Graphs/ratio_vs_`v'.pdf, replace
  5.                 }
(file Graphs/ratio_vs_molecularweight.pdf written in PDF format)
```



```
(file Graphs/ratio_vs_tpsa.pdf written in PDF format)
```




```
(file Graphs/ratio_vs_xlogp.pdf written in PDF format)
```




```
.         graph drop _all

.
. /*      For each of the three potential explanatory variables, do a regression with 'ratio' as the
>         dependent variable.     */
.         foreach v of var molecularweight tpsa xlogp {
  2.                local title: variable label `v'
  3.                disp _n(2) "Ratio vs. `title'"
```

```
  4.                 regress ratio `v'
  5.                 }
```

Ratio vs. Molecular weight (g/mol)

```
      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      2.57
       Model |  1.84619638         1  1.84619638   Prob > F        =    0.1223
    Residual |  16.4957038        23  .717204511   R-squared       =    0.1007
-------------+----------------------------------   Adj R-squared   =    0.0616
       Total |  18.3419001        24  .764245839   Root MSE        =    .84688

------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
molecularweight |  -.0045044   .0028075    -1.60   0.122    -.0103122    .0013034
       _cons |   2.000971   .8293902     2.41   0.024     .2852463    3.716695
------------------------------------------------------------------------------
```

Ratio vs. Topological polar surface area (Å{sup:2})

```
      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      1.46
       Model |  1.09322796         1  1.09322796   Prob > F        =    0.2396
    Residual |  17.2486722        23  .749942269   R-squared       =    0.0596
-------------+----------------------------------   Adj R-squared   =    0.0187
       Total |  18.3419001        24  .764245839   Root MSE        =    .86599

------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        tpsa |  -.0080207   .0066431    -1.21   0.240    -.0217629    .0057216
       _cons |   1.430099   .6303487     2.27   0.033     .1261236    2.734075
------------------------------------------------------------------------------
```

Ratio vs. XLogP

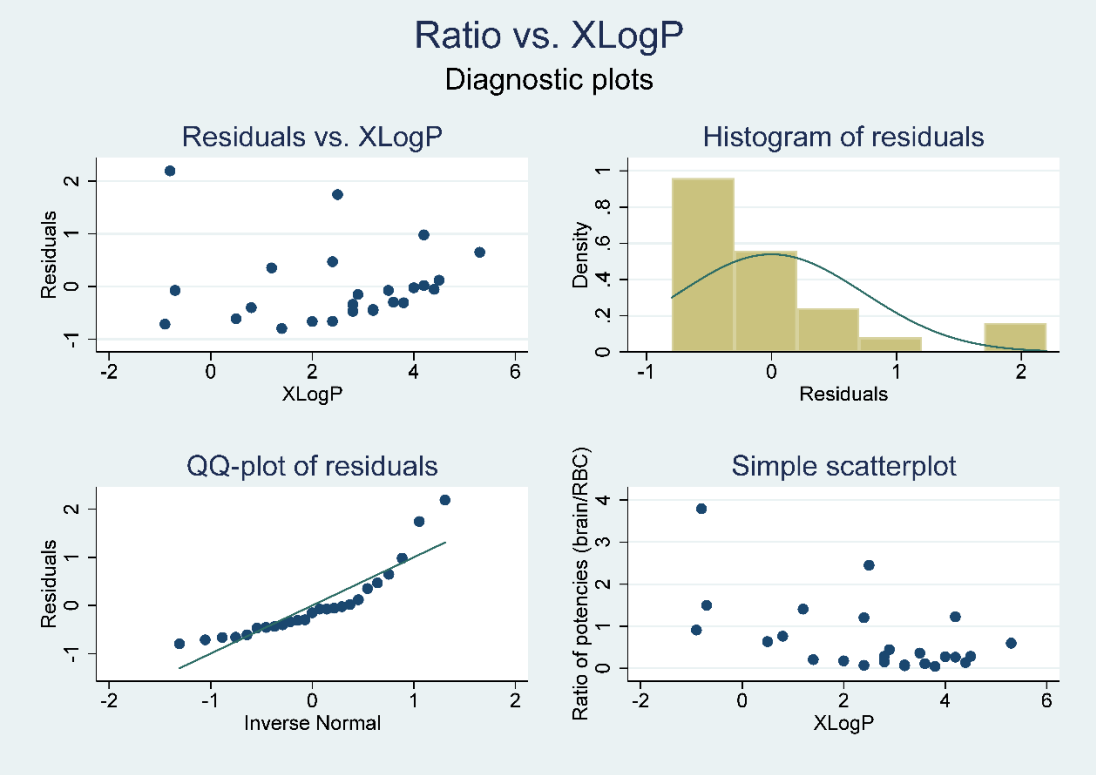
```
      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      9.17
       Model |  5.22938199         1  5.22938199   Prob > F        =    0.0060
    Residual |  13.1125182        23  .570109485   R-squared       =    0.2851
-------------+----------------------------------   Adj R-squared   =    0.2540
       Total |  18.3419001        24  .764245839   Root MSE        =    .75506

------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.2705725   .0893383    -3.03   0.006    -.4553828   -.0857622
       _cons |   1.382333   .2716824     5.09   0.000     .8203153    1.944351
------------------------------------------------------------------------------
```

```
.
. /*      Based on the adjusted R^2, XLogP is clearly the best predictor of the ratio between brain and
>         RBC potencies. We try to fit two new models of the ratio, each using XLogP and another
>         variable as predictors. */
.         foreach v of var molecularweight tpsa {
  2.                 local title: variable label `v'
  3.                 disp _n(2) "Ratio vs. `title' (in addition to xlogp)"
  4.                 regress ratio xlogp `v'
  5.                 }
```

Ratio vs. Molecular weight (g/mol) (in addition to xlogp)

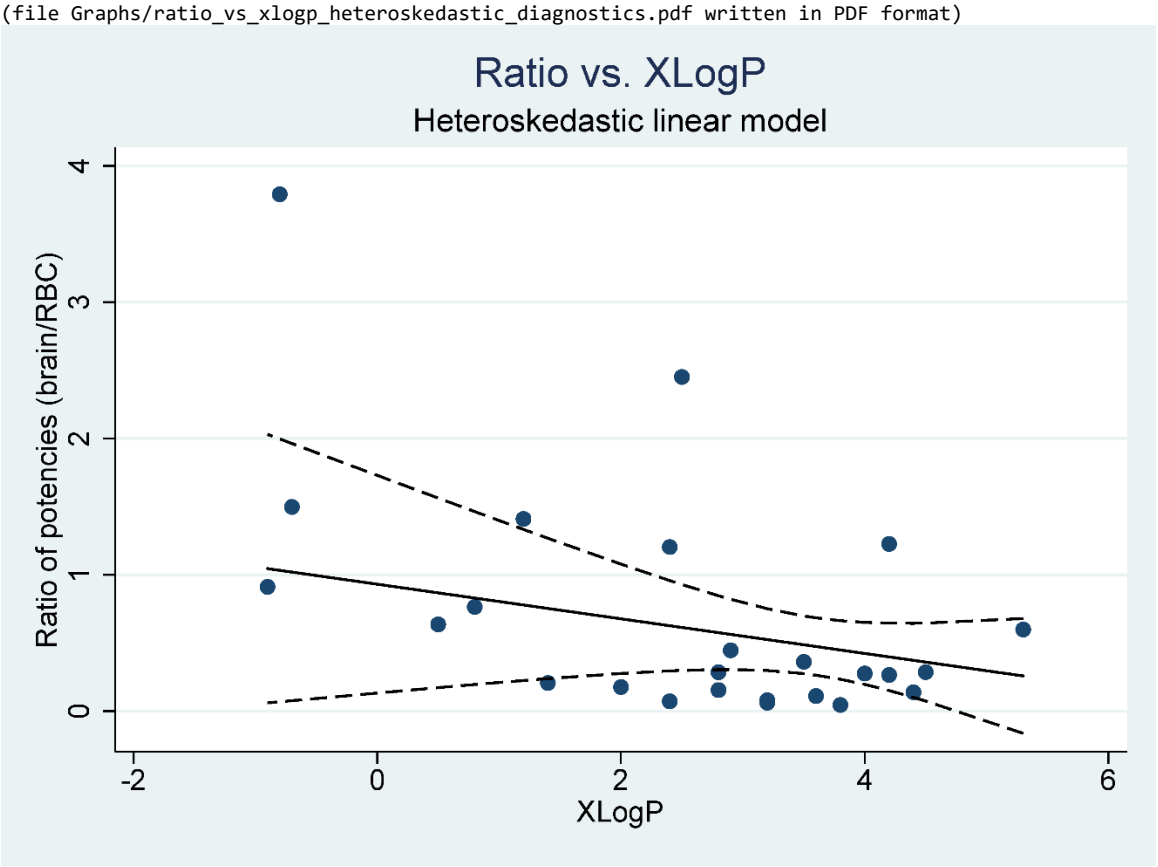
```
      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(2, 22)        =      4.76
       Model |  5.53651052         2  2.76825526   Prob > F        =    0.0192
    Residual |  12.8053896        22  .582063164   R-squared       =    0.3019
-------------+----------------------------------   Adj R-squared   =    0.2384
       Total |  18.3419001        24  .764245839   Root MSE        =    .76293
```

```
------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.3457251   .1373044    -2.52   0.020     -.630477   -.0609733
molecularweight |   .0027945    .003847     0.73   0.475    -.0051838    .0107728
       _cons |   .7641765   .8941699     0.85   0.402    -1.090218    2.618571
------------------------------------------------------------------------------


Ratio vs. Topological polar surface area (Å{sup:2}) (in addition to xlogp)

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(2, 22)        =      4.85
       Model |  5.60971898         2  2.80485949   Prob > F        =    0.0180
    Residual |  12.7321812        22  .578735508   R-squared       =    0.3058
-------------+----------------------------------   Adj R-squared   =    0.2427
       Total |  18.3419001        24  .764245839   Root MSE        =    .76075

------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.2562514   .0917288    -2.79   0.011    -.4464852   -.0660175
        tpsa |  -.0048211   .0059471    -0.81   0.426    -.0171545    .0075123
       _cons |   1.785987   .5682069     3.14   0.005     .6075984    2.964376
------------------------------------------------------------------------------

.
. /*      Based on the adjusted R^2 in these analyses, it is clear that we might as well stick with a
>         model where the only independent variable is xlogp. We fit this model again, and we draw
>         diagnostic plots to make sure that model assumptions are fulfilled.     */
.         local title: variable label xlogp

.         disp _n(2) "Ratio vs. `title'"


Ratio vs. XLogP

.         regress ratio xlogp

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      9.17
       Model |  5.22938199         1  5.22938199   Prob > F        =    0.0060
    Residual |  13.1125182        23  .570109485   R-squared       =    0.2851
-------------+----------------------------------   Adj R-squared   =    0.2540
       Total |  18.3419001        24  .764245839   Root MSE        =    .75506

------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.2705725   .0893383    -3.03   0.006    -.4553828   -.0857622
       _cons |   1.382333   .2716824     5.09   0.000     .8203153    1.944351
------------------------------------------------------------------------------

.         predict residual if e(sample), res
(8 missing values generated)

.         graph drop _all

.         scatter residual xlogp, title("Residuals vs. `title'") name(g1)

.         histogram residual, normal width(0.5) title("Histogram of residuals") name(g2)
(bin=6, start=-.7970866, width=.5)

.         qnorm residual, title("QQ-plot of residuals") name(g3)

.         scatter ratio xlogp, title("Simple scatterplot") name(g4)

.         graph combine g1 g2 g3 g4, cols(2) title("Ratio vs. `title'") subtitle("Diagnostic plots")
>  ///
>                 xsize(11.7) ysize(8.3)

.         graph export Graphs/ratio_vs_xlogp_diagnostics.pdf, replace
(file Graphs/ratio_vs_xlogp_diagnostics.pdf written in PDF format)
```



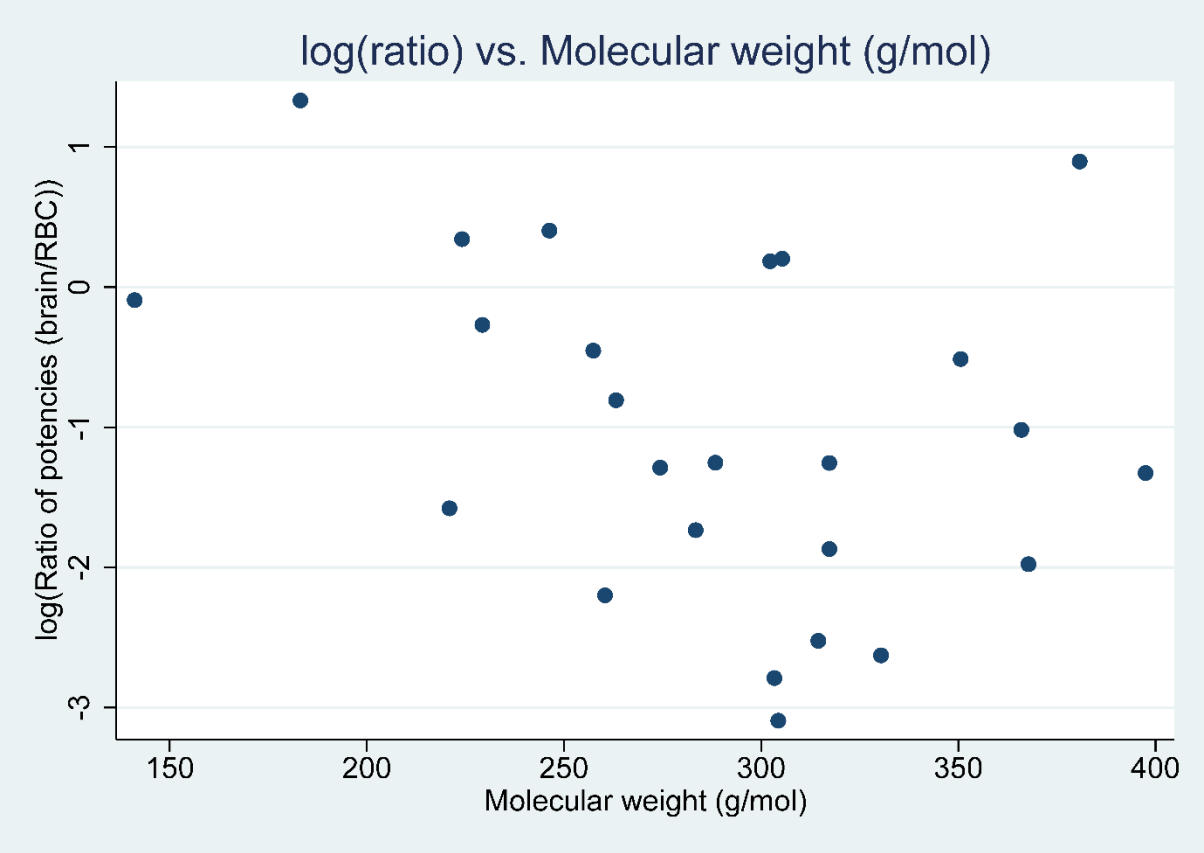

```
.         graph drop _all

.         drop residual

.
. /*      Unfortunately, it is clear from the diagnostic plots that the model does not hold. Residuals ar
> e
>         right-skewed (they should be normally distributed), and when comparing high and low values
>         of xlogp, the residuals show a funnel shape.    */
.
. /*      The problem can be solved by performing "a multiplicative heteroskedastic linear regression by
>         modeling the variance as an exponential function of" the explanatory variable (wording between
>         quotes from Stata help file).   */
.         hetregress ratio xlogp, het(xlogp)

Fitting full model:

Iteration 0:   log likelihood =  -23.82729
Iteration 1:   log likelihood = -23.700717
Iteration 2:   log likelihood = -23.700448
Iteration 3:   log likelihood = -23.700448

Heteroskedastic linear regression               Number of obs     =         25
ML estimation

                                                Wald chi2(1)      =       1.35
Log likelihood = -23.70045                      Prob > chi2       =     0.2449

------------------------------------------------------------------------------
       ratio |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ratio        |
       xlogp |  -.1269779   .1091968    -1.16   0.245    -.3409996    .0870439
       _cons |   .9311111   .4070543     2.29   0.022     .1332993    1.728923
-------------+----------------------------------------------------------------
lnsigma2     |
       xlogp |  -.5089955   .2096058    -2.43   0.015    -.9198153   -.0981757
       _cons |   .3448994   .6006467     0.57   0.566    -.8323465    1.522145
------------------------------------------------------------------------------
LR test of lnsigma2=0: chi2(1) = 7.41                     Prob > chi2 = 0.0065

.
. /*      Make predictions based on the heteroskedastic linear regression. Then plot the predictions,
```

```
>         along with the actual data behind the model.    */
.         gen sampled = 1 if e(sample)
(8 missing values generated)

.         predict ratio_predicted if sampled == 1, xb
(8 missing values generated)

.         predict se_ratio_predicted if sampled == 1, stdp

.         gen ul_ci_ratio_predicted = ratio_predicted + 1.96 * se_ratio_predicted
(8 missing values generated)

.         gen ll_ci_ratio_predicted = ratio_predicted - 1.96 * se_ratio_predicted
(8 missing values generated)

.
.         sort ratio_predicted

.         local ytitle: variable label ratio

.         twoway
>                                                                                 ///
>                 (scatter ratio xlogp)
>                                                                 ///
>                 (line ratio_predicted xlogp, lpattern(solid) lcolor(black))
>                                 ///
>                 (line ul_ci_ratio_predicted xlogp, lpattern(dash) lcolor(black))
>                         ///
>                 (line ll_ci_ratio_predicted xlogp, lpattern(dash) lcolor(black))
>                         ///
>                 , legend(off) title("Ratio vs. `title'") subtitle("Heteroskedastic linear model")
>          ///
>                 ytitle("`ytitle'")

.         graph export Graphs/ratio_vs_xlogp_heteroskedastic_diagnostics.pdf, replace
(file Graphs/ratio_vs_xlogp_heteroskedastic_diagnostics.pdf written in PDF format)
```

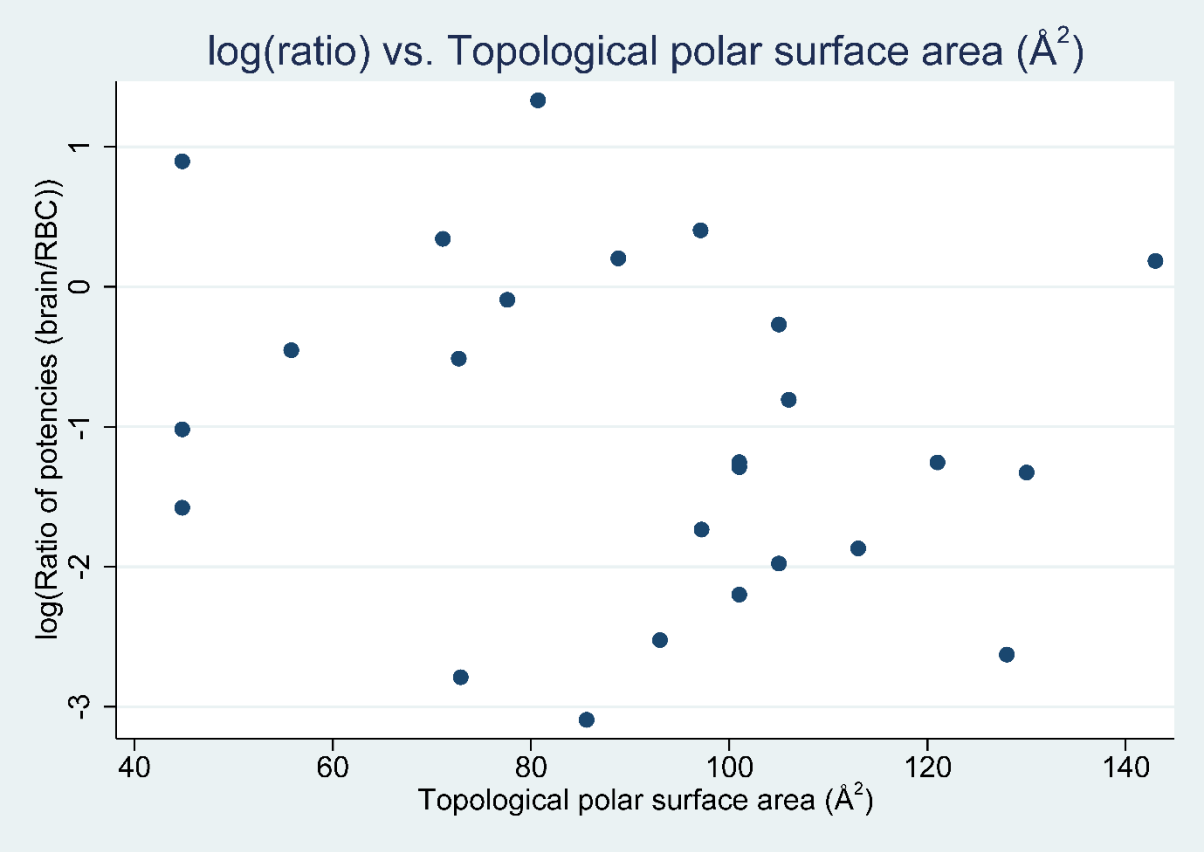



```
.         graph drop _all

.         drop *predicted* sampled
```

```
.
. /*      The heteroskedastic model is a better description of the data, but we still have a problem:
>         when XLogP become large, the predicted ratio can become negative, which is meaningless, as the
>         potency of a compound in both brain and RBC must be positive. We therefore have to try another
>         modelling strategy. We log-transfor the ratio of the potencies as use the transformed ratio as
>         the dependent variable in a linear regression model. Modelling the relationship in this way
>         insures that the ratio will always be positive (on the original scale), as as we will see, it
>         also fits the data.     */
.
.         local ratioLabel: variable label ratio

.         gen log_ratio = log(ratio)
(8 missing values generated)

.         label variable log_ratio "log(`ratioLabel')"

.
. /*      Draw scatterplots of each of the three continuous properties vs. the log(ratio of potencies).
>         These graphs will help us determine the best modelling strategy.        */
.         graph drop _all

.         foreach v of var molecularweight tpsa xlogp {
  2.                 local title: variable label `v'
  3.                 twoway scatter log_ratio `v', title("log(ratio) vs. `title'") xsize(11.7) ysize(8.3)
  4.                 graph export Graphs/log_ratio_vs_`v'.pdf, replace
  5.                 }
(file Graphs/log_ratio_vs_molecularweight.pdf written in PDF format)
```



```
(file Graphs/log_ratio_vs_tpsa.pdf written in PDF format)
```



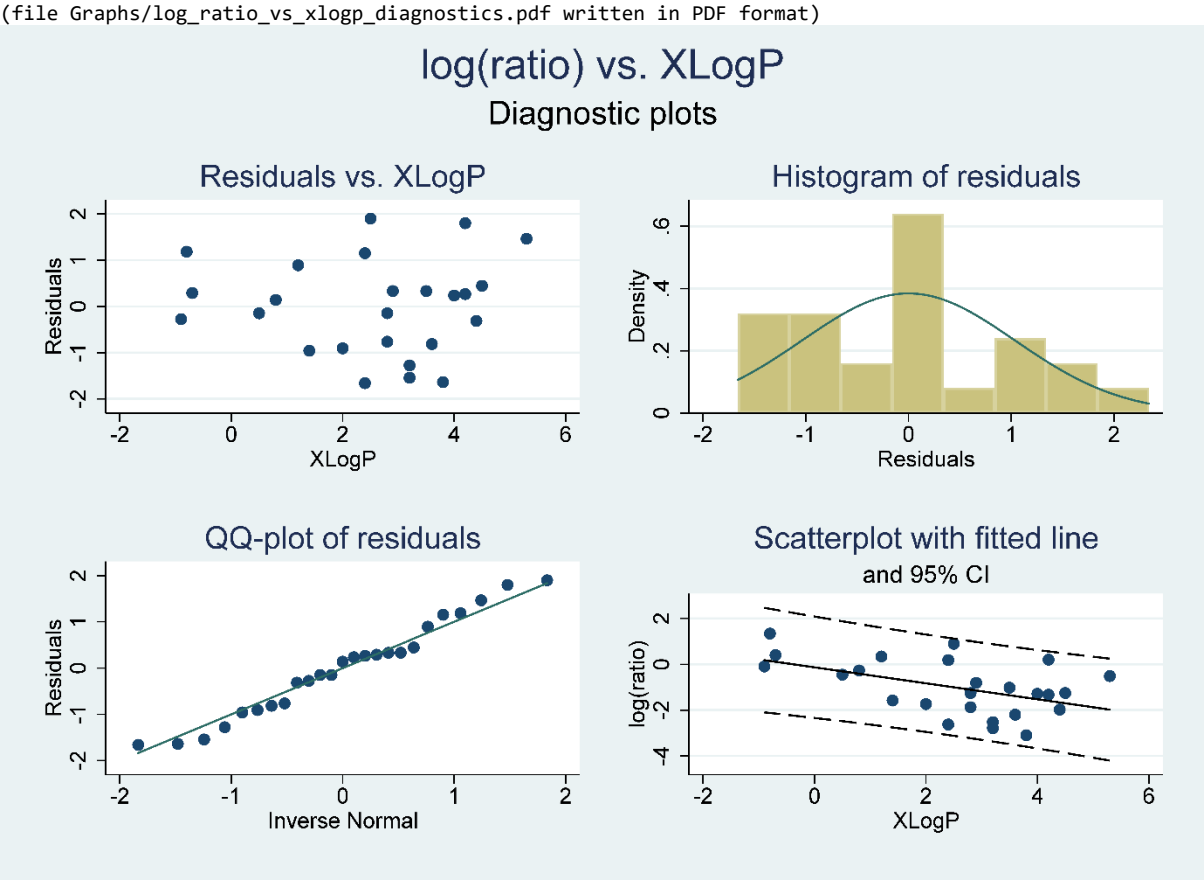

```
(file Graphs/log_ratio_vs_xlogp.pdf written in PDF format)
```


```
. graph drop _all
```

```
.
. /*      Showing that also the log-transformed ratio is best predicted based on a model with xlogp as th
> e
>         only explanatory variable.      */
.         foreach v of var molecularweight tpsa xlogp {
  2.                 local title: variable label `v'
  3.                 disp _n(2) "log(ratio) vs. `title'"
  4.                 regress log_ratio `v'
  5.                 }


log(ratio) vs. Molecular weight (g/mol)

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      2.78
       Model |  3.72169112         1  3.72169112   Prob > F        =    0.1089
    Residual |  30.7628127        23  1.33751359   R-squared       =    0.1079
-------------+----------------------------------   Adj R-squared   =    0.0691
       Total |  34.4845038        24  1.43685432   Root MSE        =    1.1565

------------------------------------------------------------------------------
      log_ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
molecularweight |  -.0063955    .003834    -1.67   0.109    -.0143267    .0015357
          _cons |   .8377997   1.132626     0.74   0.467    -1.505215    3.180815
------------------------------------------------------------------------------


log(ratio) vs. Topological polar surface area (Å{sup:2})

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      1.12
       Model |  1.60406701         1  1.60406701   Prob > F        =    0.3005
    Residual |  32.8804368        23  1.42958421   R-squared       =    0.0465
-------------+----------------------------------   Adj R-squared   =    0.0051
       Total |  34.4845038        24  1.43685432   Root MSE        =    1.1957

------------------------------------------------------------------------------
   log_ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        tpsa |  -.0097155   .0091719    -1.06   0.300    -.0286891     .009258
       _cons |  -.1253069   .8703057    -0.14   0.887    -1.925671    1.675058
------------------------------------------------------------------------------


log(ratio) vs. XLogP

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      7.72
       Model |  8.66425769         1  8.66425769   Prob > F        =    0.0107
    Residual |  25.8202461        23   1.1226194   R-squared       =    0.2513
-------------+----------------------------------   Adj R-squared   =    0.2187
       Total |  34.4845038        24  1.43685432   Root MSE        =    1.0595

------------------------------------------------------------------------------
   log_ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.3482765   .1253645    -2.78   0.011    -.6076128   -.0889402
       _cons |  -.1312697   .3812402    -0.34   0.734     -.919925    .6573857
------------------------------------------------------------------------------

.         foreach v of var molecularweight tpsa {
  2.                 local title: variable label `v'
  3.                 disp _n(2) "log(ratio) vs. `title' (in addition to xlogp)"
  4.                 regress log_ratio xlogp `v'
  5.                 }


log(ratio) vs. Molecular weight (g/mol) (in addition to xlogp)

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(2, 22)        =      3.80
       Model |  8.85718899         2  4.42859449   Prob > F        =    0.0382
    Residual |  25.6273148        22  1.16487795   R-squared       =    0.2568
```

```
-------------+----------------------------------   Adj R-squared   =    0.1893
       Total |  34.4845038        24  1.43685432   Root MSE        =    1.0793

------------------------------------------------------------------------------
      log_ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
          xlogp |  -.4078407   .1942404    -2.10   0.047    -.8106706   -.0050108
molecularweight |   .0022148   .0054423     0.41   0.688    -.0090718    .0135015
          _cons |  -.6212062   1.264955    -0.49   0.628    -3.244563    2.002151
------------------------------------------------------------------------------


log(ratio) vs. Topological polar surface area (Å{sup:2}) (in addition to xlogp)

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(2, 22)        =      3.99
       Model |  9.17260067         2  4.58630033   Prob > F        =    0.0333
    Residual |  25.3119031        22  1.15054105   R-squared       =    0.2660
-------------+----------------------------------   Adj R-squared   =    0.1993
       Total |  34.4845038        24  1.43685432   Root MSE        =    1.0726

------------------------------------------------------------------------------
   log_ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.3317199   .1293352    -2.56   0.018    -.5999446   -.0634951
        tpsa |  -.0055737   .0083852    -0.66   0.513    -.0229635    .0118162
       _cons |   .3353938   .8011568     0.42   0.680    -1.326104    1.996891
------------------------------------------------------------------------------

.
. /*      Based on the adjusted R^2 in these analyses, it is clear that we might as well stick with a
>         model where the only independent variable is xlogp. We fit this model again, and we draw
>         diagnostic plots to make sure that model assumptions are fulfilled.     */
.         local title: variable label xlogp

.         disp _n(2) "log(ratio) vs. `title'"


log(ratio) vs. XLogP

.         regress log_ratio xlogp

      Source |       SS           df       MS      Number of obs   =        25
-------------+----------------------------------   F(1, 23)        =      7.72
       Model |  8.66425769         1  8.66425769   Prob > F        =    0.0107
    Residual |  25.8202461        23   1.1226194   R-squared       =    0.2513
-------------+----------------------------------   Adj R-squared   =    0.2187
       Total |  34.4845038        24  1.43685432   Root MSE        =    1.0595

------------------------------------------------------------------------------
   log_ratio |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       xlogp |  -.3482765   .1253645    -2.78   0.011    -.6076128   -.0889402
       _cons |  -.1312697   .3812402    -0.34   0.734     -.919925    .6573857
------------------------------------------------------------------------------

.         gen sampled = 1 if e(sample)
(8 missing values generated)

.         predict residual if sampled == 1, res
(8 missing values generated)

.         predict log_ratio_predicted, xb

.         predict se_log_ratio_predicted if sampled == 1, stdf
(8 missing values generated)

.         gen ul_ci_predicted = log_ratio_predicted + 1.96 * se_log_ratio_predicted
(8 missing values generated)

.         gen ll_ci_predicted = log_ratio_predicted - 1.96 * se_log_ratio_predicted
(8 missing values generated)

.         graph drop _all
```

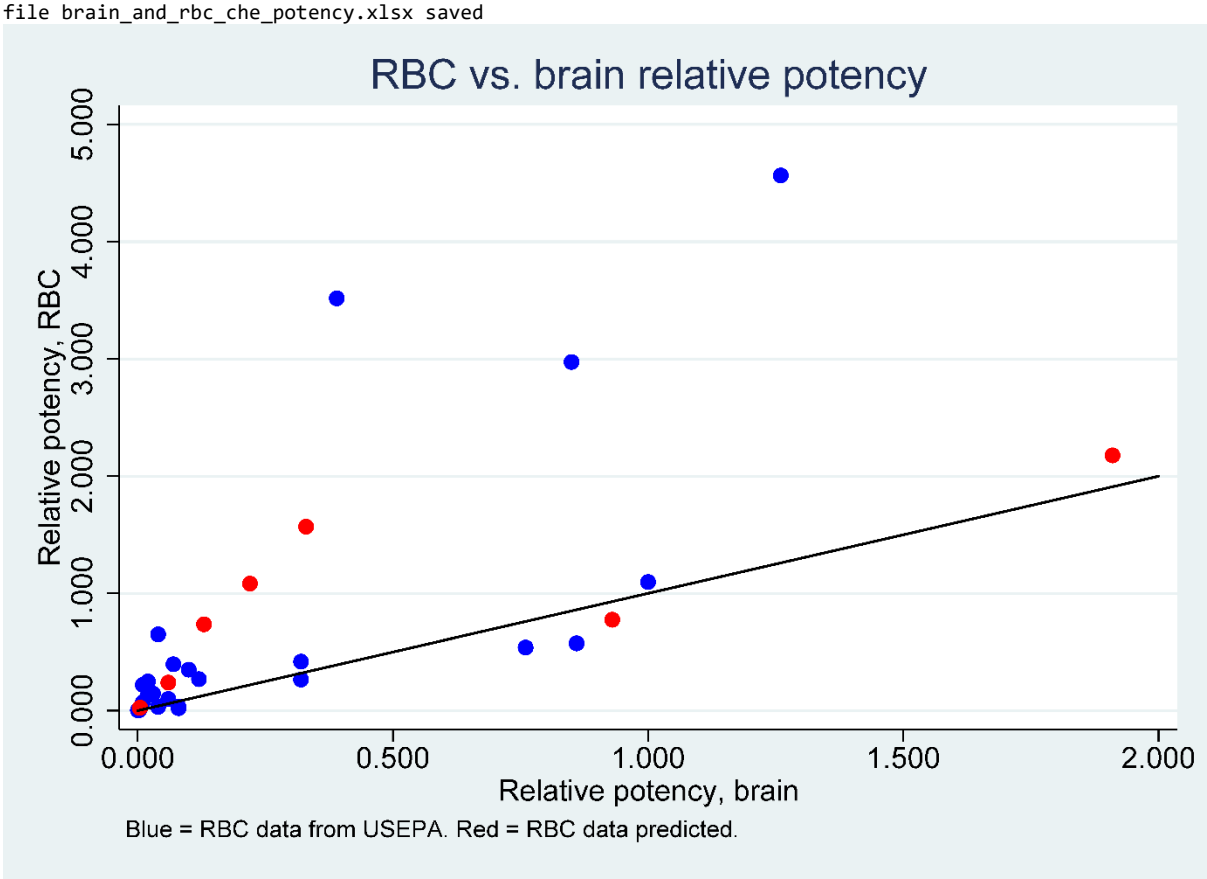
```
.         scatter residual xlogp, title("Residuals vs. `title'") name(g1)

.         histogram residual, normal width(0.5) title("Histogram of residuals") name(g2)
(bin=8, start=-1.6597197, width=.5)

.         qnorm residual, title("QQ-plot of residuals") name(g3)

.         twoway
>                                                                                 ///
>                 (scatter log_ratio xlogp)
>                                                                 ///
>                 (line log_ratio_predicted xlogp, lpattern(solid) lcolor(black))
>                         ///
>                 (line ll_ci_predicted xlogp, lpattern(dash) lcolor(black))
>                                 ///
>                 (line ul_ci_predicted xlogp, lpattern(dash) lcolor(black))
>                                 ///
>                 , title("Scatterplot with fitted line") subtitle("and 95% CI")
>                         ///
>                 legend(off) name(g4) ytitle("log(ratio)")

.         graph combine g1 g2 g3 g4, cols(2) title("log(ratio) vs. `title'")
>                         ///
>                 subtitle("Diagnostic plots") xsize(11.7) ysize(8.3)

.         graph export Graphs/log_ratio_vs_xlogp_diagnostics.pdf, replace

(file Graphs/log_ratio_vs_xlogp_diagnostics.pdf written in PDF format)
```




```
.         graph drop _all

.         drop residual

.
. //      Prettify dataset.
.         replace log_ratio = log_ratio_predicted if sampled != 1
(8 real changes made)

.         replace ratio = exp(log_ratio) if sampled != 1
```

```
(8 real changes made)

.         replace potency_rbc = potency_brain/ratio if sampled != 1
(8 real changes made)

.         recode sampled (1 = 1 "Original") (. = 0 "Predicted"), gen(rbc_group)
(8 differences between sampled and rbc_group)

.         keep chemical potency* rbc_group

.         format %15.3f potency*

.
. //      List dataset
.         sort chemical

.         list, ab(32)

     +-------------------------------------------------------------+
     |            chemical   potency_rbc   potency_brain   rbc_group |
     |-------------------------------------------------------------|
  1. |            Acephate         0.021           0.080    Original |
  2. |     Azinphos-methyl         0.350           0.100    Original |
  3. |           Bensulide         0.011           0.003    Original |
  4. |      Chlorethoxyfos         0.736           0.130   Predicted |
  5. |        Chlorpyrifos         0.100           0.060    Original |
     |-------------------------------------------------------------|
  6. | Chlorpyrifos-methyl         0.025           0.005   Predicted |
  7. |            Diazinon         0.220           0.010    Original |
  8. |          Dichlorvos         0.145           0.030    Original |
  9. |         Dicrotophos         2.178           1.910   Predicted |
 10. |          Dimethoate         0.419           0.320    Original |
     |-------------------------------------------------------------|
 11. |          Disulfoton         4.565           1.260    Original |
 12. |            Ethoprop         0.240           0.060   Predicted |
 13. |          Fenamiphos         0.650           0.040    Original |
 14. |            Fenthion         1.569           0.330   Predicted |
 15. |         Fosthiazate         0.396           0.070    Original |
     |-------------------------------------------------------------|
 16. |           Malathion         0.004           0.000    Original |
 17. |       Methamidophos         1.097           1.000    Original |
 18. |        Methidathion         0.266           0.320    Original |
 19. |    Methyl-parathion         0.269           0.120    Original |
 20. |           Mevinphos         0.539           0.760    Original |
     |-------------------------------------------------------------|
 21. |              Naled         0.033           0.080    Original |
 22. |           Omethoate         0.775           0.930   Predicted |
 23. |   Oxydemeton-methyl         0.574           0.860    Original |
 24. |             Phorate         3.517           0.390    Original |
 25. |           Phosalone         0.072           0.010    Original |
     |-------------------------------------------------------------|
 26. |             Phosmet         0.130           0.020    Original |
 27. |       Phostebupirim         1.083           0.220   Predicted |
 28. |   Pirimiphos-methyl         0.033           0.040    Original |
 29. |          Profenofos         0.023           0.004   Predicted |
 30. |            Terbufos         2.974           0.850    Original |
     |-------------------------------------------------------------|
 31. |   Tetrachlorvinphos         0.003           0.001    Original |
 32. |            Tribufos         0.249           0.020    Original |
 33. |         Trichlorfon         0.005           0.003    Original |
     +-------------------------------------------------------------+

.
. /*      Graphically show the result.    */
.         local ytitle: variable label potency_rbc

.         local xtitle: variable label potency_brain

.         graph drop _all

.         twoway
>                                                                              ///
>                 (scatter potency_rbc potency_brain if rbc_group == 1, mcolor(blue))
>                  ///
```

```
>                 (scatter potency_rbc potency_brain if rbc_group == 0, mcolor(red))
>                         ///
>                 (function y = x, range(0 2) lcolor(black))
>                                         ///
>                 , legend(off) ytitle("`ytitle'") xtitle("`xtitle'")
>                                 ///
>                 note("Blue = RBC data from USEPA. Red = RBC data predicted.")
>                         ///
>                 xsize(11.7) ysize(8.3) title("RBC vs. brain relative potency")

.         graph export Graphs/rbc_vs_brain_relative_potency.pdf, replace
(file Graphs/rbc_vs_brain_relative_potency.pdf written in PDF format)

.         graph drop _all

.
. //      Export data.
.         export excel chemical potency* rbc_group using brain_and_rbc_che_potency.xlsx,
>          ///
>                 firstrow(varlabels) replace keepcellfmt

file brain_and_rbc_che_potency.xlsx saved
```




```
.
. //      Clear data
.         clear

.
. //      Stop logging
.         log close
      name:  <unnamed>
       log:  C:\Users\au231481\Desktop\Potency EPA\predict_rbc_ache.log
  log type:  text
 closed on:   2 Aug 2019, 16:31:15
-------------------------------------------------------------------------------------------------------------
```

16 Appendix B: Table of assumed causal relationships between study variables

The table on the following pages lists all the causal relationships between that we believe (a priori) to be causally related to each other. The table is provided to provide a better overview of the same relationships on the DAGs (directed acyclic graphs) in Appendix C.

		Variable name in DAG	Variable content	Measured in PEXADU project?	Causally influences these DAG variables
Participant factors	**Personal characteristics**	AChE	Acetylcholine esterase	Measured	FPG
		Age	Age in years	Measured	AChE
					Basal_Metabolism
					Alcohol
					Biofuel_Burning
					Diet
					Physical_Activity
					Tobacco
					Organophosphate_Farming
					Other_Pesticides_Farming
					FPG
					Lung_Function
		BMI	Body mass index	Measured	FPG
					Lung_Function
		Basal_Metabolism	Basal metabolism	Unmeasured	BMI
		Height	Height in centimeters	Measured	Lung_Function
		SES	Socioeconomic status	Measured	Alcohol
					Diet
					Height
					Physical_Activity
					Tobacco
					Organophosphate_Farming
					Other_Pesticides_Farming
					Biofuel_Burning
		Sex	Sex	Measured	AChE
					BMI
					Basal_Metabolism
					SES
					Alcohol
					Biofuel_Burning
					Diet
					Height

					Tobacco
					Organophosphate_Farming
					Other_Pesticides_Farming
					FPG
					Lung_Function
	Behavior	Alcohol	Alcohol consumption	Measured	AChE
					BMI
					FPG
		Biofuel_Burning	Burning of biofuels	Measured	Biofuel_Smoke
		Diet	Diet	Measured	BMI
					Organophosphate_Diet
					Other_Pesticide_Diet
					FPG
		Physical_Activity	Physical activity level	Measured	Diet
					BMI
					FPG
		Tobacco	Tobacco smoking	Measured	Diet
					FPG
					Lung_Function
	Exposure	Biofuel_Smoke	Exposure to biofuel smoke	Unmeasured	Lung_Function
		Organophosphate_Diet	Exposure to organophosphates through diet	Unmeasured	Organophosphate_Total
		Organophosphate_Farming	Exposure to organophosphates through farming	Measured	Organophosphate_Total
		Organophosphate_Total	Total organophosphate exposure	Unmeasured	AChE
					FPG
					Lung_Function
		Other_Pesticide_Diet	Exposure to other pesticides through diet	Unmeasured	Other_Pesticide_Total
		Other_Pesticide_Total	Total exposure to other pesticides	Unmeasured	FPG
					Lung_Function
		Other_Pesticides_Farming	Exposure to other pesticides through farming	Measured	Other_Pesticide_Total
	Genes	AChE_Genes	Genes for AChE activity	Unmeasured	AChE
		Diabetes_Genes	Genes for diabetes susceptibility	Unmeasured	FPG
		Height_Genes	Genes for height	Unmeasured	Height

		Lung_Function_Genes	Genes for lung function	Unmeasured	Lung_Function
		Obesity_Genes	Genes for obesity	Unmeasured	BMI
					Basal_Metabolism
Parental factors that can causally affect participant	**Personal characteristics**	Diabetes_(Parent)	Diabetes in parent	Measured	SES
					Alcohol
					Diet
					Physical_Activity
					Tobacco
		Lung_Function_(Parent)	Parent's lung function	Measured	Biofuel_Burning
					Tobacco
		SES_(Parent)	Parent's socioeconomic status	Unmeasured	SES
	Behavior	Alcohol_(Parent)	Parent's alcohol consumption	Unmeasured	Alcohol
		Biofuel_Burning_(Parent)	Parent's burning of biofuels	Unmeasured	Biofuel_Burning
		Diet_(Parent)	Parent's diet	Unmeasured	Diet
		Physical_Activity_(Parent)	Parent's physical activity level	Unmeasured	Physical_Activity
		Tobacco_(Parent)	Parent's tobacco smoking	Unmeasured	Tobacco
	Genes	AChE_Genes_(Parent)	Parent's genes for AChE activity	Unmeasured	AChE_Genes
		Diabetes_Genes_(Parent)	Parent's genes for diabetes susceptibility	Unmeasured	Diabetes_Genes
		Height_Genes_(Parent)	Parent's genes for height	Unmeasured	Height_Genes
		Lung_Function_Genes_(Parent)	Parent's genes for lung function	Unmeasured	Lung_Function_Genes
		Obesity_Genes_(Parent)	Parent's genes for obesity	Unmeasured	Obesity_Genes

FPG = fasting plasma glucose

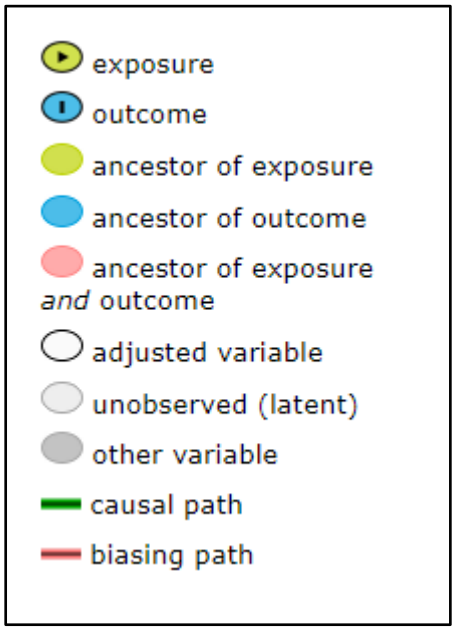
All causal relationship between variables listed as “Participant factors” are assumed to also exist for participants’ parents. E.g., a participant’s smoking is assumed to causally influence his/her lung function (as listed in the table), and his/her mother’s smoking is also assumed to influence the mother’s lung function (even though this is not listed in the table).

The table indicates that we have data on the participants' parents' lung function. A previous diagnosis of asthma or COPD in the parents, reported by the participant, is used as a proxy for the parents' lung functions.

17 Appendix C: DAGs and causal effect reports from DAGitty

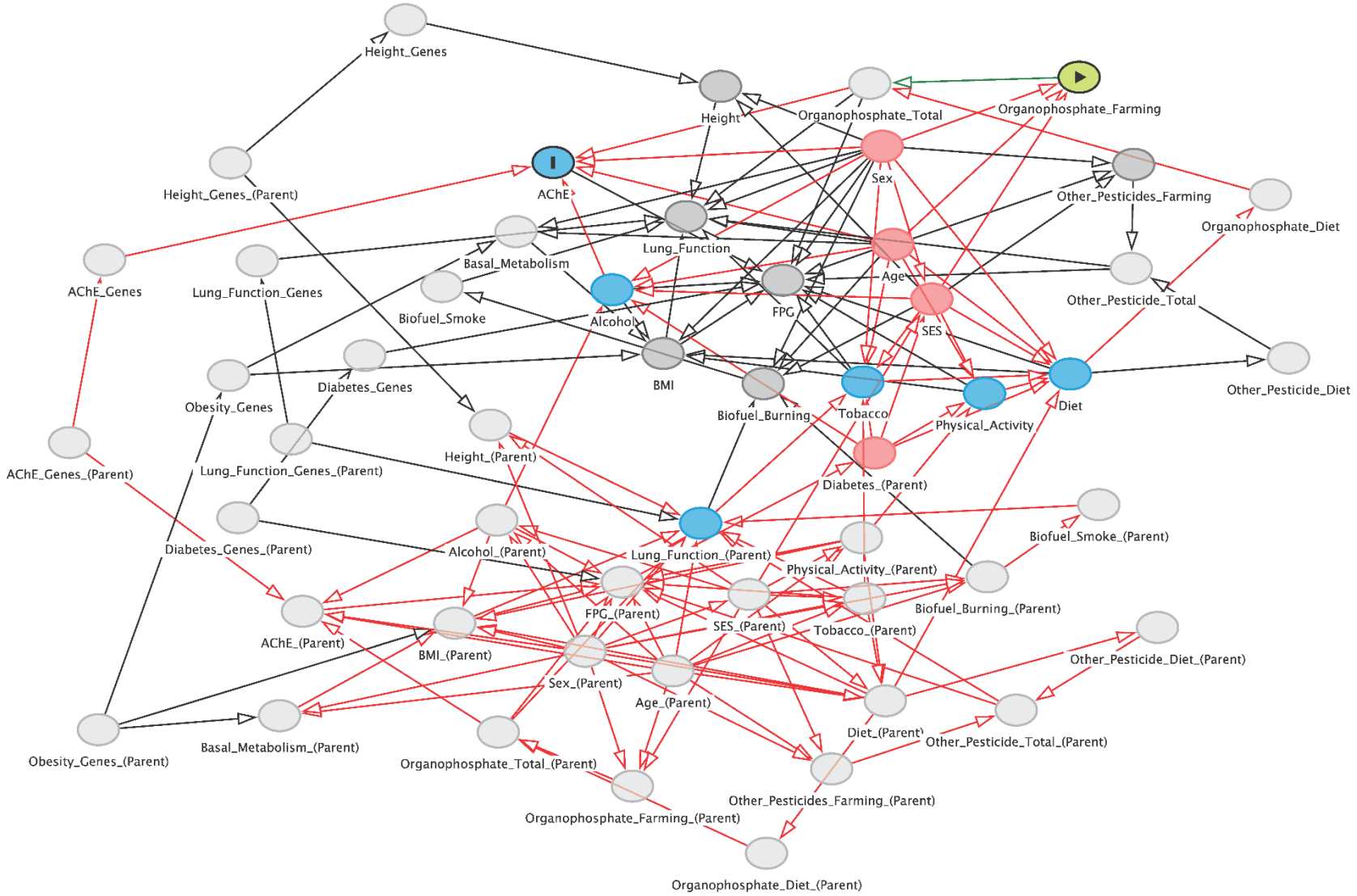
The DAGs (Directed Acyclic graphs = causal diagrams) on the following pages were drawn using the DAGitty[31] software, freely available from dagitty.net. Because of the complexity of the DAGs, they were analyzed automatically by DAGitty. Under each DAG, we have listed the output from the analysis.

Legend for all DAGs:



17.1 Outcome = red blood cell acetylcholine esterase

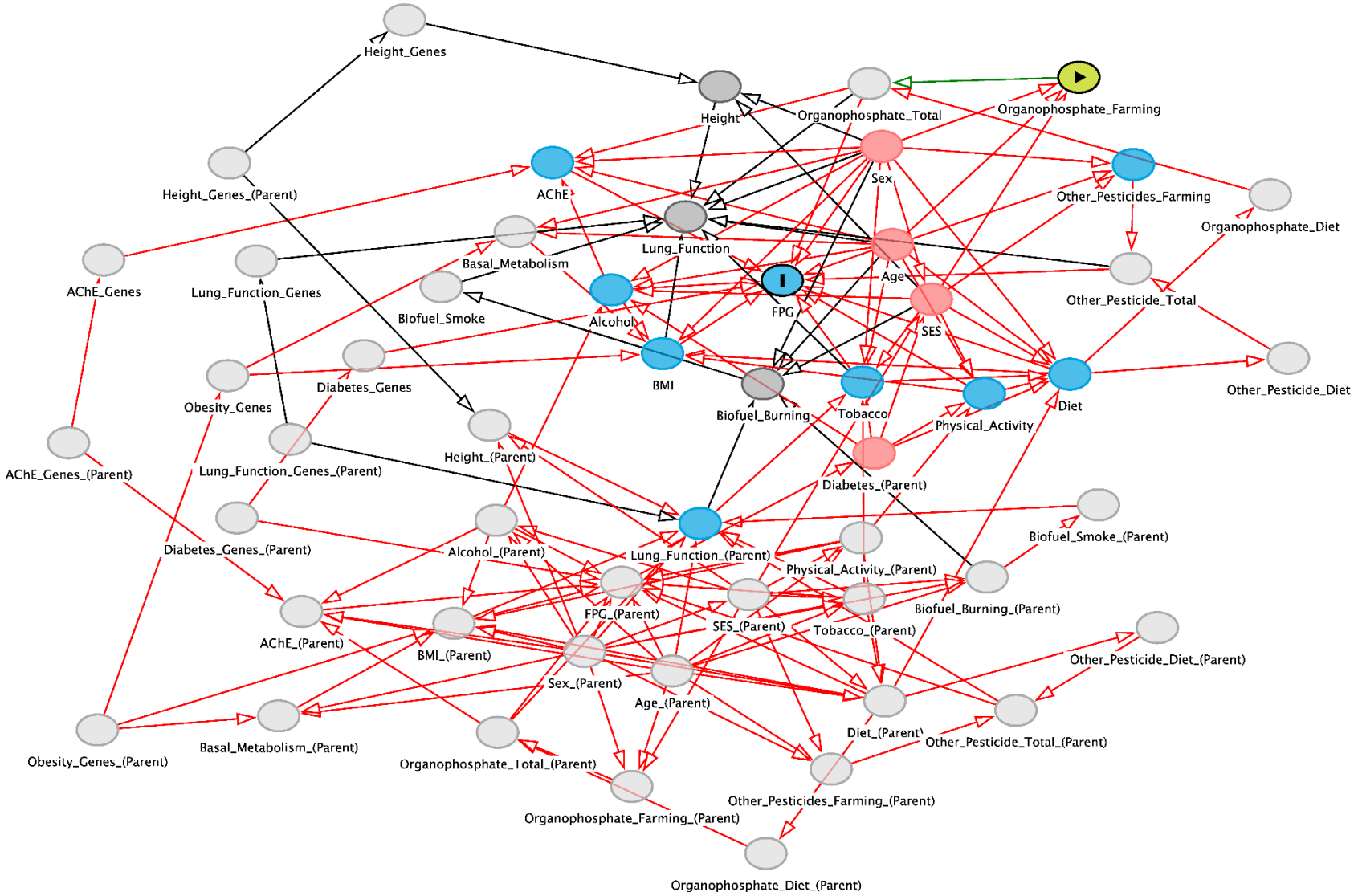
17.1.1 Exposure metric = self-reported use of organophosphate insecticides in farming



Causal effect identification
Minimal sufficient adjustment sets for estimating the total effect of Organophosphate_Farming on AChE: Age, SES, Sex
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex

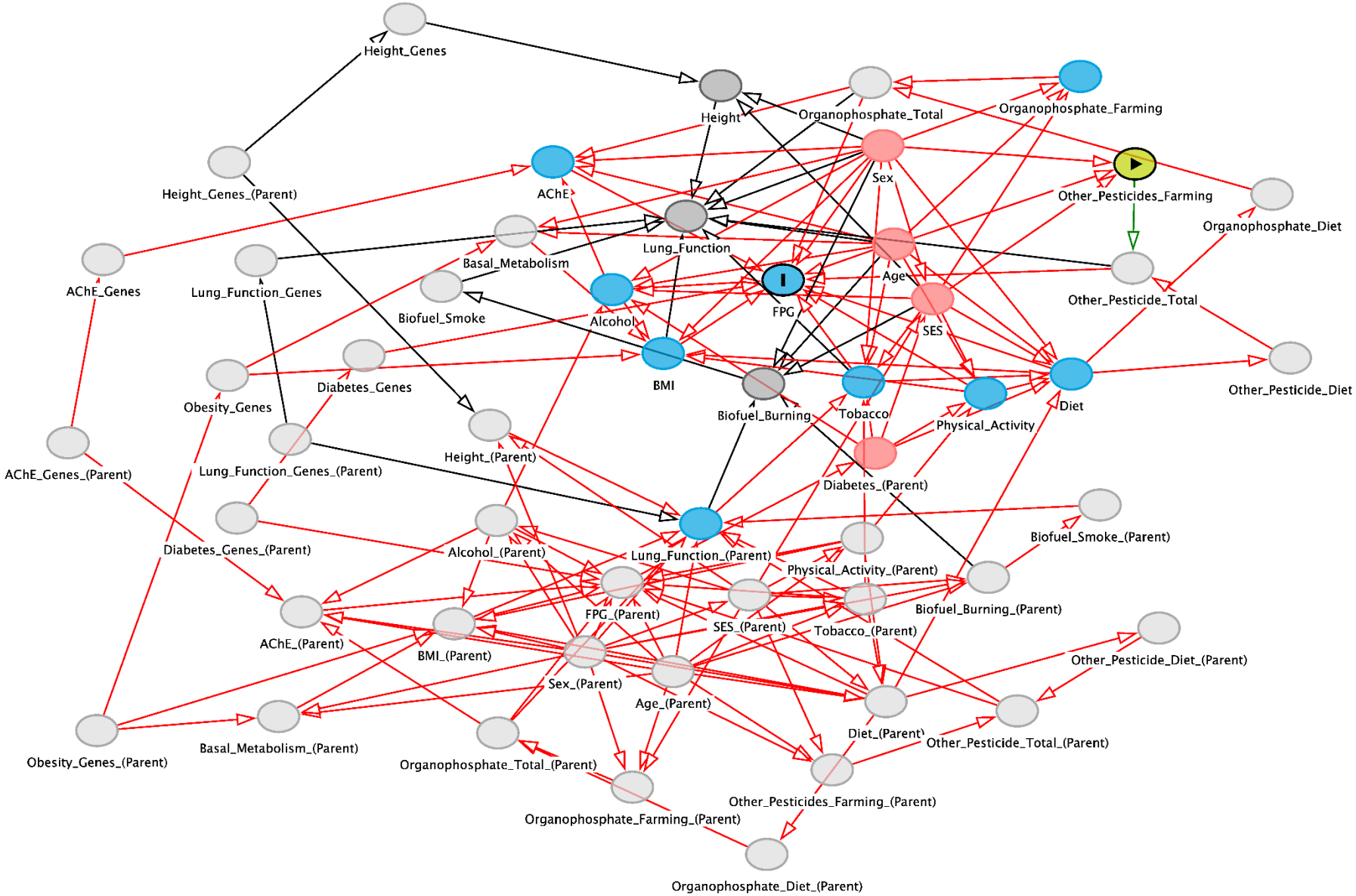
17.2 Outcome = glycemic regulation, exemplified by fasting plasma glucose

17.2.1 Exposure metric = self-reported use of organophosphate insecticides in farming



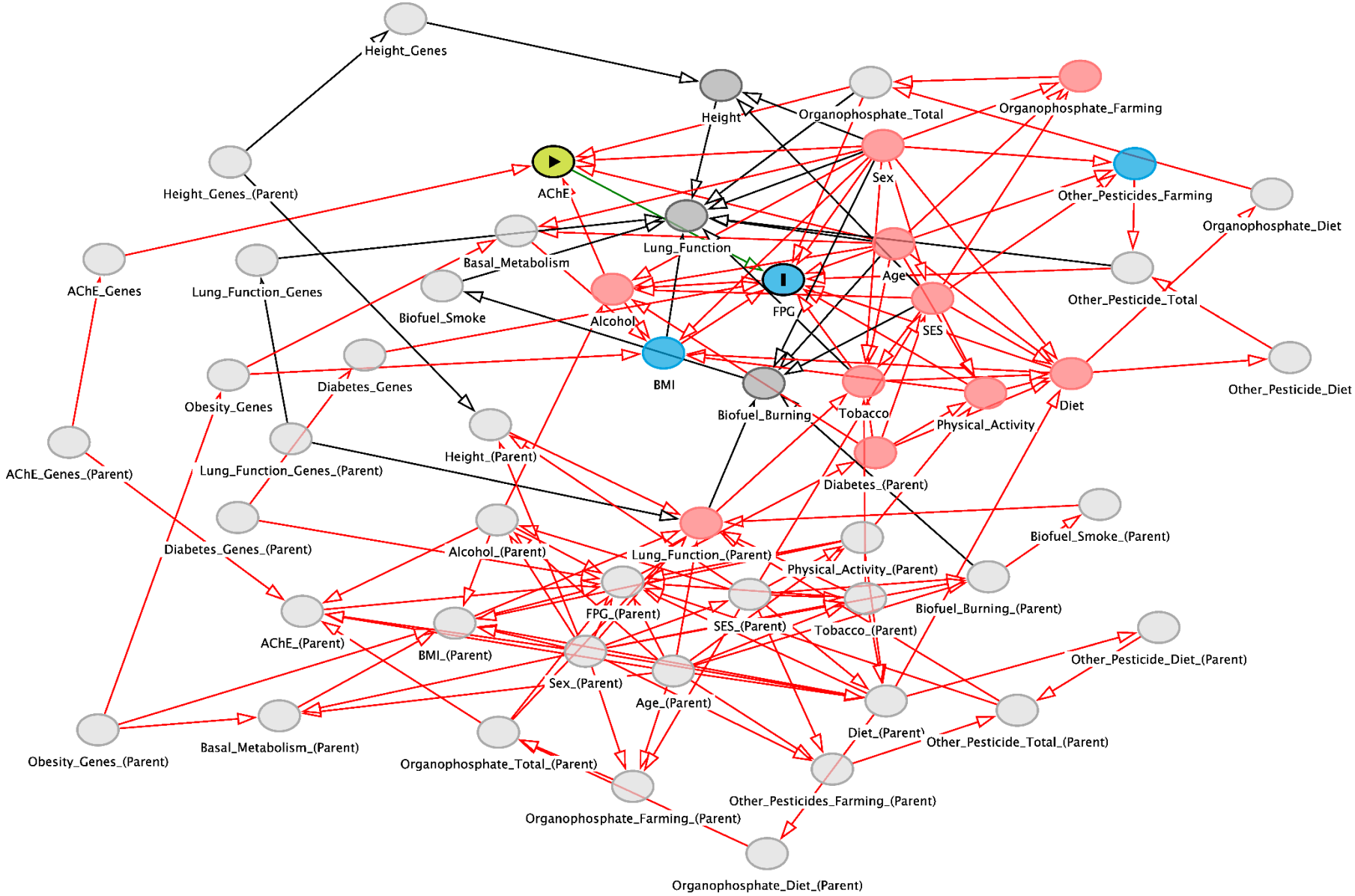
Causal effect identification
Minimal sufficient adjustment sets for estimating the total effect of Organophosphate_Farming on FPG: Age, SES, Sex
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex

17.2.2 Exposure metric = self-reported use of other classes of pesticides in farming



Causal effect identification
Minimal sufficient adjustment sets for estimating the total effect of Other_Pesticides_Farming on FPG: Age, SES, Sex
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex

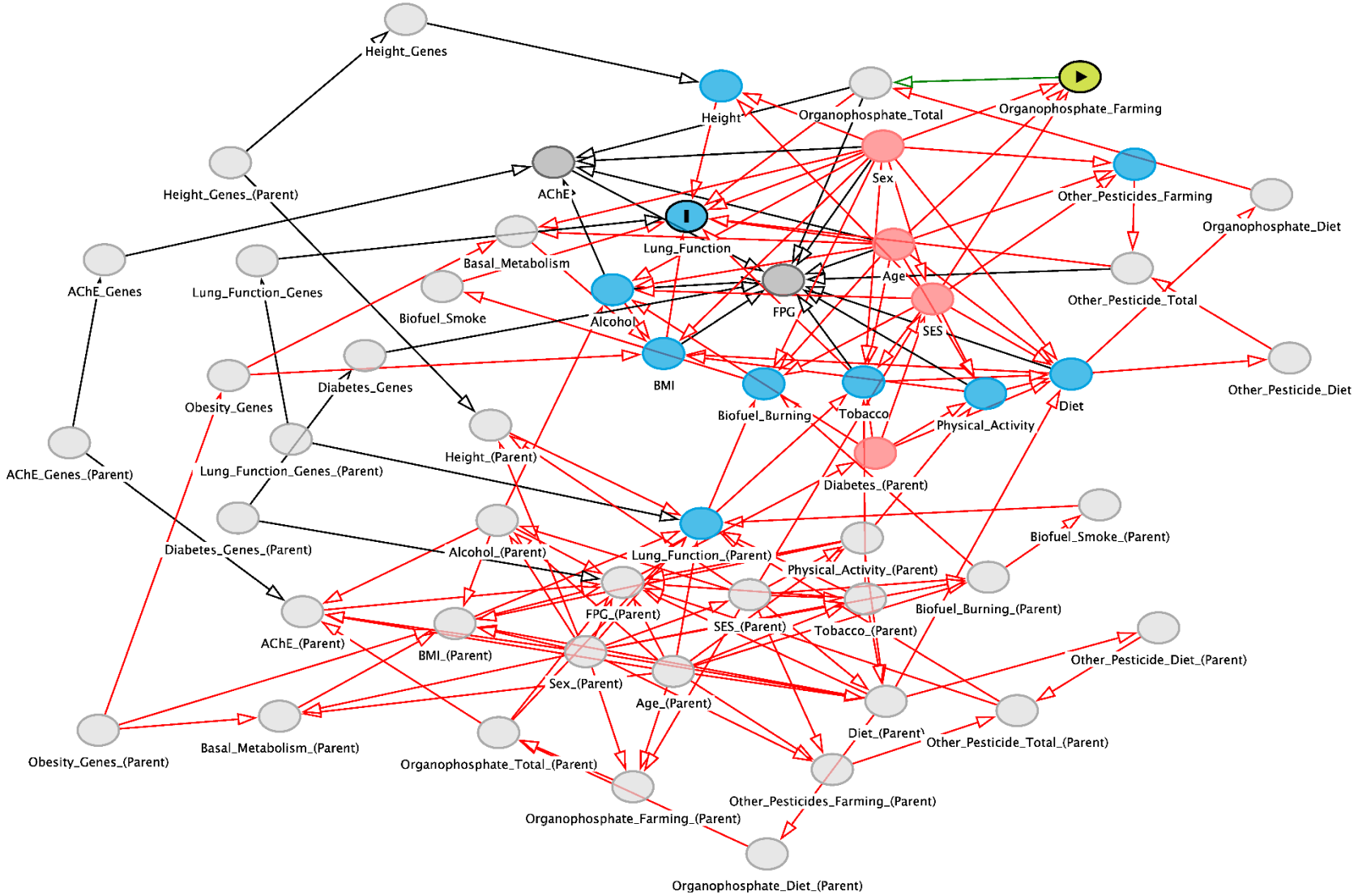
17.2.3 Exposure metric = red blood cell acetylcholine esterase activity



Causal effect identification
The total effect cannot be estimated by covariate adjustment.
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex

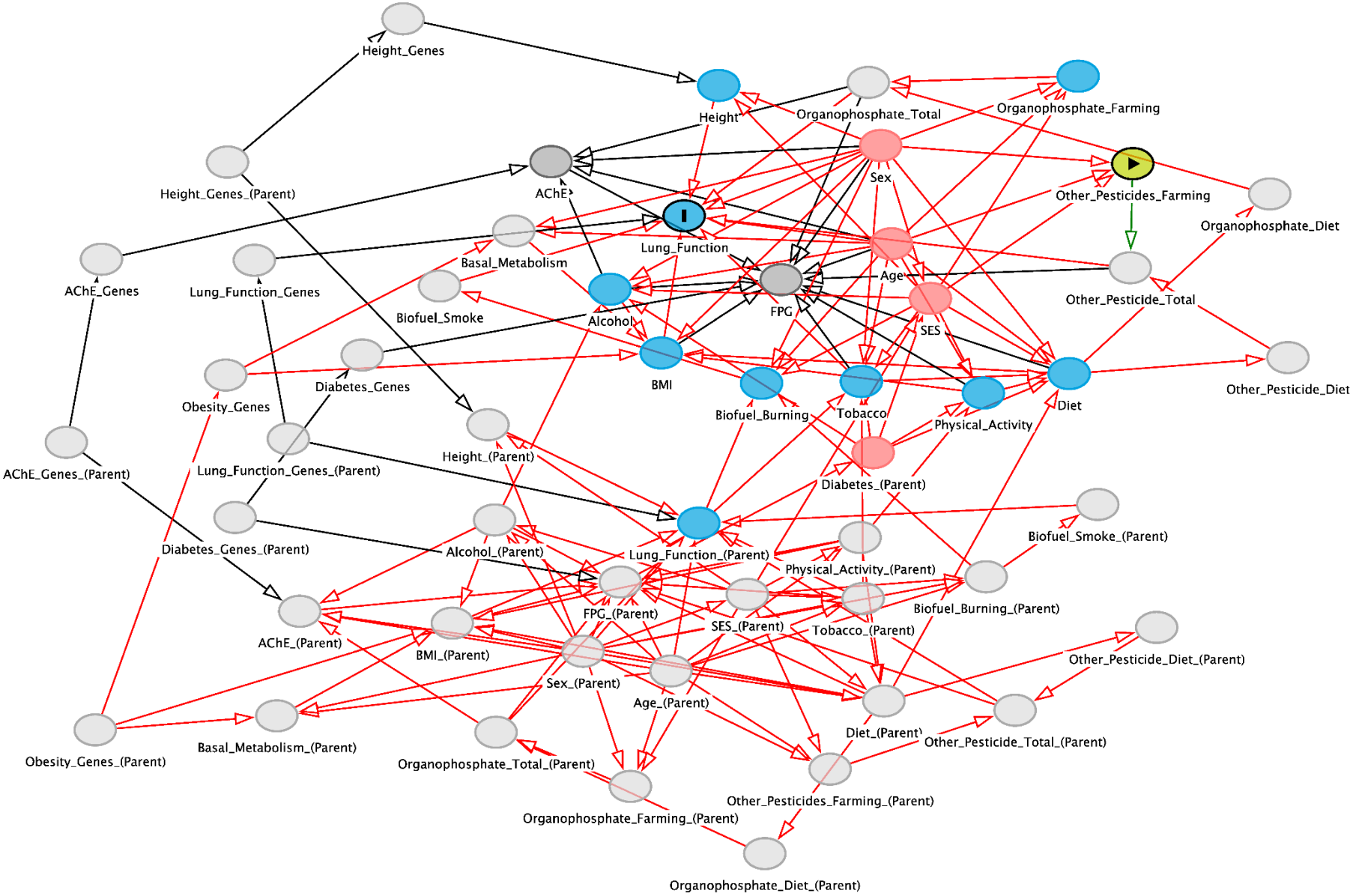
17.3 Outcome = lung function

17.3.1 Exposure metric = self-reported use of organophosphate insecticides in farming



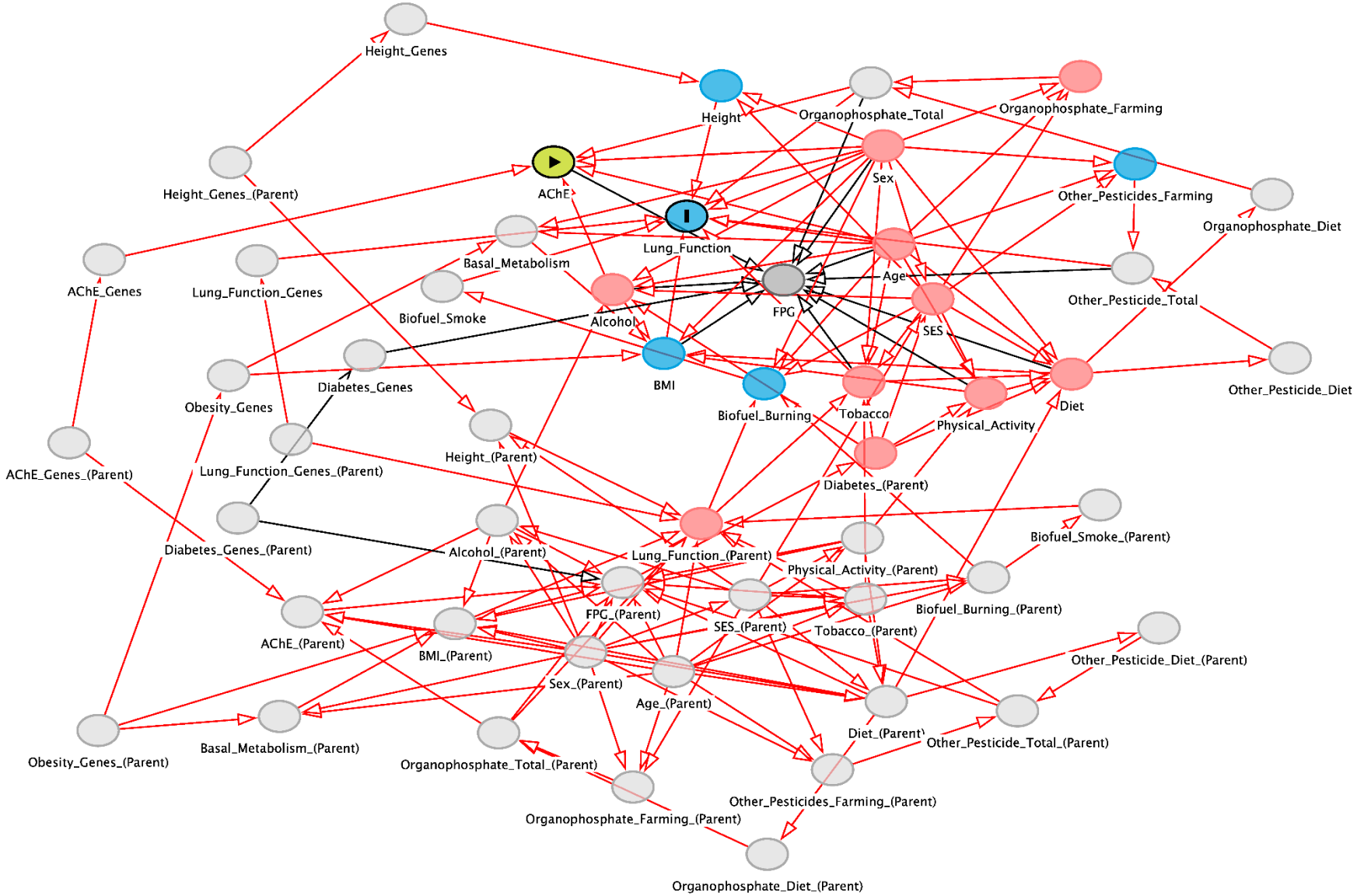
Causal effect identification
Minimal sufficient adjustment sets for estimating the total effect of Organophosphate_Farming on Lung_Function: Age, SES, Sex
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex

17.3.2 Exposure metric = self-reported use of other classes of pesticides in farming



Causal effect identification
Minimal sufficient adjustment sets for estimating the total effect of Other_Pesticides_Farming on Lung_Function: Age, SES, Sex
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex

17.4 Exposure metric = red blood cell acetylcholine esterase activity



Causal effect identification
The total effect cannot be estimated by covariate adjustment.
Testable implications
The model implies the following conditional independences: Age ⊥ Sex Age ⊥ Diabetes_(Parent) Age ⊥ Height Age ⊥ Lung_Function_(Parent) Age ⊥ SES Sex ⊥ Diabetes_(Parent) Sex ⊥ Lung_Function_(Parent) AChE ⊥ Other_Pesticides_Farming | Age, SES, Sex Alcohol ⊥ Height | SES, Sex Alcohol ⊥ Organophosphate_Farming | Age, SES, Sex Alcohol ⊥ Other_Pesticides_Farming | Age, SES, Sex BMI ⊥ Organophosphate_Farming | Age, SES, Sex BMI ⊥ Other_Pesticides_Farming | Age, SES, Sex Biofuel_Burning ⊥ Organophosphate_Farming | Age, SES, Sex Biofuel_Burning ⊥ Other_Pesticides_Farming | Age, SES, Sex Diabetes_(Parent) ⊥ Height | SES, Sex Diabetes_(Parent) ⊥ Organophosphate_Farming | SES, Sex Diabetes_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Diet ⊥ Organophosphate_Farming | Age, SES, Sex Diet ⊥ Other_Pesticides_Farming | Age, SES, Sex Height ⊥ Organophosphate_Farming | SES, Sex Height ⊥ Other_Pesticides_Farming | SES, Sex Height ⊥ Physical_Activity | SES, Sex Lung_Function_(Parent) ⊥ Organophosphate_Farming | SES, Sex Lung_Function_(Parent) ⊥ Other_Pesticides_Farming | SES, Sex Organophosphate_Farming ⊥ Other_Pesticides_Farming | Age, SES, Sex Organophosphate_Farming ⊥ Physical_Activity | Age, SES, Sex Organophosphate_Farming ⊥ Tobacco | Age, SES, Sex Other_Pesticides_Farming ⊥ Physical_Activity | Age, SES, Sex Other_Pesticides_Farming ⊥ Tobacco | Age, SES, Sex