**DATA SUMMARY**

***Initial dataset:***

**Number of effect sizes**: 2185

**Number of studies**: 92

**Number of species**: 36

**Number of observations with shared control group**: 2034

**Number of studies with shared control group**: 79

***Filtering initial dataset:***

After correcting sample size for shared control groups, 98 observations have sample size lower than 1 and are removed from dataset (sampling variance calculation only for n > 1). This leaves a dataset of 2087 observations, 90 studies and 36 species.

There are 5 observations with standard deviations for control and treatment groups of 0. These should be checked just in case these are typos/error in the dataset (and sample sizes for these observations are large, unlikely that SD is just 0). These observations are ‘Study.ID’: 1549, 1746, 2161, 2162, 2164. For now, I remove them from the data, but any chance you or Colette can double check these data in the original paper?

Up until now, the dataset has 2082 observations, 90 studies and 36 species.

Then, I remove 5 studies that present correlative data (i.e., no experimental), these five studies account for 75 observations. After removing these five studies, the dataset has 2007 observations, 85 studies and 33 species

**Observations (and studies) by moderator levels:**

Biomarker and developmental stage:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Embryo** | **Tadpole** | **Adult** |
| **Enzymatic** | 97 (11) | 708 (45) | 547 (17) |
| **Indicator** | 5 (2) | 172 (25) | 166 (15) |
| **Non-enzymatic** | 22 (4) | 131 (15) | 159 (16) |

Biomarker, developmental stage and pollutant class:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Embryo** | **Tadpole** | **Adult** |
| **Herbicide** | **Enzymatic** | 5 (1) | 170 (12) | 10 (2) |
| **Indicator** | 0 (0) | 60 (10) | 2 (1) |
| **Non-enzymatic** | 0 (0) | 27 (2) | 0 (0) |
| **Metallic elements** | **Enzymatic** | 16 (2) | 196 (6) | 8 (2) |
| **Indicator** | 0 (0) | 3 (1) | 16 (4) |
| **Non-enzymatic** | 0 (0) | 44 (2) | 24 (7) |
| **Other inorganic compounds** | **Enzymatic** | 36 (4) | 51 (7) | 3 (1) |
| **Indicator** | 0 (0) | 14 (2) | 3 (1) |
| **Non-enzymatic** | 3 (1) | 6 (1) | 3 (1) |
| **Other organic compounds** | **Enzymatic** | 0 (0) | 16 (3) | 20 (2) |
| **Indicator** | 0 (0) | 4 (2) | 12 (2) |
| **Non-enzymatic** | 0 (0) | 2 (1) | 11 (2) |
| **Pesticide** | **Enzymatic** | 40 (5) | 275 (17) | 506 (12) |
| **Indicator** | 5 (2) | 91 (10) | 133 (8) |
| **Non-enzymatic** | 19 (3) | 52 (9) | 121 (6) |

**ANALYSIS SUMMARY**

***1 – Overall meta-analysis***

I first ran an overall meta-analysis with no moderators and the full dataset.

I get a warning message: “*Ratio of largest to smallest sampling variance extremely large. May not be able to obtain stable results”*

This message seems to mean that there are observations with very large and / or very small sampling variance and that causes problems. I inspect the data and indeed there are two observations (‘Study.ID’ 71 and 1756) that have very large sampling variance compared to the rest of the data set. It would be good to double check the data for these two effect sizes, just in case there is a type in the SD value of controls or treatments. For now, I have run a second model in which I have removed the top 10 and bottom 10 observations by their sampling variance. These second model does not generate the warning and the results are extremely similar to the first model (so, all good).

*These are the heterogeneity results:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ***I*2 total** | ***I*2 study ID** | ***I*2 species** | ***I*2 phylogeny** | ***I*2 residual** |
| **Full dataset**  **(k = 2007)** | 99.95 | 26.97 | 1.30 | 2.61 | 69.06 |
| **Trimmed data set**  **(k = 1987)** | 99.81 | 27.62 | 1.21 | 2.42 | 68.57 |

*These are the model coefficients:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Low 95%CI** | **High 95%CI** | **p** |
| **Full dataset**  **(k = 2007)** | 0.1489 | 0.2719 | 0.0258 | 0.018 |
| **Trimmed data set**  **(k = 1987)** | 0.1500 | 0.2710 | 0.0290 | 0.015 |

These estimates mean that overall treatments caused approximately a 16% increase in biomarker levels compared to controls.

*Plot of results:*

*Chart

Description automatically generated*

***2 – Overall effect per biomarker***

In the second meta-analysis, I include ‘Biomarker category’ as a moderator. I get a similar warning as before and check results with the trimmed dataset, also as above (both models provide very similar results).

Biomarker category only explains 1.57% of the variation in lnRR in the whole dataset (k = 2007 observations). The estimate for ‘Enzymatic’ and “Indicator’ are positive and significantly different from zero (table and plot below).

*These are the model coefficients:*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Estimate** | **Low 95%CI** | **High 95%CI** | **p** |
| **Full dataset**  **(k = 2007)** | **Enzymatic** | 0.120 | 0.008 | 0.232 | 0.035 |
| **Indicator** | 0.297 | 0.175 | 0.418 | < 0.001 |
| **Non-enzymatic** | 0.121 | -0.002 | 0.244 | 0.054 |
| **Trimmed data set**  **(k = 1987)** | **Enzymatic** | 0.121 | 0.011 | 0.231 | 0.031 |
| **Indicator** | 0.298 | 0.178 | 0.418 | <0.001 |
| **Non-enzymatic** | 0.121 | -0.001 | 0.242 | 0.051 |

*Chart

Description automatically generated*

***3 – Effect per biomarker and developmental stage***

Then, I do the analysis above, including biomarker category as a moderator, but for each developmental stage separately.

*Results for embryos:*

Biomarker category explains 8.2% of the variation in lnRR in the subset for embryos (k = 124 observations), and the estimate for ‘Non-enzymatic’ is positive and significantly different from zero (table and plot below).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Low 95%CI** | **High 95%CI** | **p** |
| **Enzymatic** | 0.054 | -0.213 | 0.320 | 0.694 |
| **Indicator** | 0.221 | -0.205 | 0.646 | 0.309 |
| **Non-enzymatic** | 0.387 | 0.040 | 0.733 | 0.029 |

Chart, scatter chart

Description automatically generated

*Results for tadpoles:*

Biomarker category only explains 0.59% of the variation in lnRR in the subset for tadpoles (k = 1011 observations), and none of the estimates are significantly different from zero (table and plot below). I get a similar warning as before and check results with a trimmed dataset, also as above (both models provide very similar results).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Estimate** | **Low 95%CI** | **High 95%CI** | **p** |
| **Full dataset**  **(k = 1011)** | **Enzymatic** | 0.179 | -0.012 | 0.370 | 0.066 |
| **Indicator** | 0.063 | -0.145 | 0.270 | 0.554 |
| **Non-enzymatic** | 0.203 | -0.010 | 0.416 | 0.061 |
| **Trimmed data set**  **(k = 991)** | **Enzymatic** | 0.181 | -0.001 | 0.370 | 0.060 |
| **Indicator** | 0.068 | -0.139 | 0.274 | 0.520 |
| **Non-enzymatic** | 0.203 | -0.008 | 0.414 | 0.060 |

***Chart

Description automatically generated***

*Results for adults:*

Biomarker category explains 10.02% of the variation in lnRR in the subset for adults (k = 872 observations), and the estimate for ‘Indicator’ is positive and significantly different from zero (table and plot below).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Low 95%CI** | **High 95%CI** | **p** |
| **Enzymatic** | 0.071 | -0.080 | 0.223 | 0.358 |
| **Indicator** | 0.465 | 0.306 | 0.624 | < 0.001 |
| **Non-enzymatic** | 0.076 | -0.083 | 0.234 | 0.348 |

Chart

Description automatically generated

*Results across the three life stages for comparison (same plots as above but limits of the x axis are the same across the three plots):*

**Embryos**

***Chart, scatter chart

Description automatically generated***

**Tadpoles**

***Chart

Description automatically generated***

**Adults**

***Chart, diagram

Description automatically generated***

***4 – Effect of pollutants per biomarker – To Be Continue***

*Results for non-enzymatic:*

*Results for enzymatic:*

*Results for indicator:*