A Commentary on Some Epidemiology Data for Chlorpyrifos

Michael Dourson[a,*], Bernard Gadagbui[a], Chijioke Onyema[a], Patricia McGinnis[a]

[a]Toxicology Excellence for Risk Assessment (TERA), 1250 Ohio Pike, Suite 197

Amelia, OH 45102

* Corresponding author. Tel.: +1 513 542-7475 ext. 101; Fax: 513.488.1990.

E-mail addresses: dourson@tera.org (M. Dourson), gadagbui@tera.org (B. Gadagbui), chijiokeonyema@gmail.com (C. Onyema), mcginnis@tera.org (P. McGinnis).

Abstract

Extensive animal and human studies on chlorpyrifos (CPF) point to changes in a blood enzyme as its first biological effect, and governments and health groups around the world have used this effect in the determination of its safe dose. Preventing this first biological effect, referred to in risk assessment parlance as the critical effect, is part and parcel of chemical regulation in general and of CFP specifically. Rauh et al. (2011), one of the published studies from the Columbia Center for Children's Environmental Health (CCCEH), reported evidence of deficits in Working Memory Index[1] and Full-Scale IQ in children at 7 years old as a function of prenatal CPF exposures that are much lower than levels causing cholinesterase inhibition. Since the raw data on which Rauh et. al. (2011) publicly-funded (in part) findings were based have not been made available despite repeated requests, we show extracted data in Figures 1A and 1E of Rauh et al. (2011), and plotted these extracted data as response versus log dose, a common risk assessment approach. Surprisingly, a significant portion of the data stated to be available in Rauh et al. (2011) were not found in these published figures, perhaps due to data point overlay. However, the reported associations of chlorpyrifos levels with Working Memory and Full Scale IQ were also not replicated in our analysis due perhaps to this missing data. Multiple requests were made to Rauh et al. (2011) for access to data from this, in part, publicly funded study, so that confirmation could be attempted. This general lack of data and inconsistency with cholinergic responses in other researches raises concerns about

[1] Working Memory Index assesses children's ability to memorize new information, hold it in short-term memory, concentrate, and manipulate information.

the lack of data transparency.

Highlights:

- Epidemiology findings suggest that the critical effect of chlorpyrifos is neurological deficits, rather than cholinesterase inhibition.
- Data from these epidemiology findings are needed to ensure replicability.
- Science transparency is needed in any significant set of data so that risk management decisions can be made with reasonable confidence.

Keywords: Chlorpyrifos, Chlorpyrifos-oxon, cholinesterase inhibition, Working Memory Index, Full-Scale IQ scores, science transparency

1. Introduction

The discipline of environmental risk assessment for human health necessitates a careful study of effects caused by chemicals of concern, extrapolation of observations from one population (usually experimental animals) to others (usually sensitive human subgroups) and monitoring or modeling appropriate human exposures. The former two activities are used in part to determine a chemical's safe dose, where human data are preferred over data in experimental animals if available and credible. The latter activity is used in part to determine whether monitored or modeled human exposures are at or below this safe dose.

Pesticides are useful chemicals for assuring adequate food production, but since these chemicals are designed to kill organisms, they need to be carefully studied so that they do not otherwise harm non-target species, such as humans. Since environmental exposures to pesticides are anticipated, it is not surprising that agencies, such as the U.S. Environmental Protection Agency (EPA) spend a lot of resource in reviewing the plethora of studies submitted by manufacturers. In the case of the pesticide chlorpyrifos, (CPF) scores of studies evaluated by EPA (2016) suggest that its sentinel[2] effect, that is, the first biological effect (or marker of exposure) or its known precursor, is cholinesterase

[2] This is also referred to as the chemical's critical effect.

inhibition.[3] This finding is so well accepted that health agencies across the world have focused on cholinesterase inhibition as the basis for determining chlorpyrifos' safe dose (e.g., Health Canada, 2003; United Kingdom, 2003; World Health Organization (WHO), 2004).

A publication by Rauh et al. (2011), one of several studies from the Columbia Center for Children's Environmental Health (CCCEH), showed neurological effects in children associated with prenatal levels of CPF that were much lower than those that showed cholinesterase inhibition. The effects reported by Rauh et al. (2011) are important to consider, since they are in contrast to the weigh-of-evidence from other researchers suggesting cholinesterase inhibition as the critical effect. These effects suggest a different critical effect, specifically, that the critical effect for CPF should be based on human neurological changes rather than cholinesterase inhibition.

Specifically, Rauh et al. (2011) show statistically significant, negative associations of Working Memory and Full Scale IQ scores after adjustment by the natural logarithm with dose shown in normal units. This adjustment of the IQ scores compresses the top of the y-axis in relationship to the bottom. This compression of top IQ scores and stretching out of the lower IQ scores give the subtle appearance of a downward shift, which lessens when IQ scores on the y-axis are not mathematically so adjusted.

From a risk assessment perspective, a common data display would be to show unadjusted

[3] Furthermore this inhibition occurs at roughly the same dose and time course in experimental animals and humans (Zhao et al., 2006).

IQ scores, which are already expected to be normally distributed,[4] as a function of dose that is in either normal units or adjusted by logarithm (base 10). The dose x-axis when adjusted into log units will also stretch out the lower part of the axis in relationship to the higher part of the axis. In this case the logarithm adjustment is appropriate, however, because most of the exposure data lie in the lower part of the dose x-axis.

The purpose of this brief technical report is to confirm the findings of Rauh et al. (2011) so that human neurological effects can be compared with experimental animal cholinesterase inhibition as the critical effect.

2. Methods

Epidemiologists often study associations among a plethora of effects versus exposures to multiple chemicals. This is a good strategy since associations can lead to further, more definitive, investigations, based on a more clearly defined hypothesis. One hypothesis that could be developed from the Rauh et al. (2011) study, in particular, would be that neurological effects of CPF occur at doses lower than cholinesterase inhibition. In this regard, Figures 1A (Working Memory) and 1E (Full Scale IQ) of Rauh et al. (2011) were reviewed and the data from the plots were extracted both manually and using DigitizeIt Software v2.3 (https://www.digitizeit.de). The results for chlorpyrifos levels and test scores were entered into an excel spreadsheet for further analysis.

[4] Each of the measurements in the Rauh et al. (2011) study is on a "standardized scale has [with] a mean of 100 and SD of 15." (See Rauh et al. page 1197, column 3, line 16-17.). This means that the y-axis should be expected to have a normal bell-shaped distribution, although in this population it may not.

For Figure 1A of Rauh et al. (2011), 33 data points were shown by Rauh et al. as zero or non-detectable. For Figure 1E, 60 points were shown as zero or non-detectable. Consistent with the approach of Raul et al. (2011, page 1198), ~80% of these values were assigned a chlorpyrifos level of 0.5 picograms per gram (pg/g) and ~20% of them were assigned a level of 1.0 pg/g. The results were then plotted as natural logarithm-adjusted response versus reported dose (as per Rauh et al., 2011), and as un-adjusted response versus $\log_{10}$ dose. Linear regressions were developed using excel spreadsheet software.

During this reanalysis, approximately 35% of the data, as stated to be available in the publication in Rauh et al. (2011, page 1197), in Figure 1A and approximately 15% of the data in Figure 1E appeared to be missing. These missing data could be due in part to data point overlay. However, at least two high dose data points were not found in either graph, which would not be a data-overlay problem. EPA staff also noticed these missing high dose data points. The CCCEH response to EPA staff stated that 4 high dose data points were actually not plotted, and further stated that the inclusion of at least one of these four un-plotted data points would have attenuated or perhaps even eliminated the statistical significance of their findings.[5] Further examination of Figure 1A in Rauh et al. (2011)

[5] Memo to: Carrol Christensen, Ph.D; From: Robin M. Whyatt, DrPH; Date: April 9, 2015
Re: July 2011 letter to Deborah Smegal, M.P.H.: [Full memo available as Appendix A]

> EPA comment: In Figure 1 page 29, the upper bound of the x axis (chlorpyrifos) is shown to be 25 pg/gm. However, in the second paragraph of page 11 it was reported that the maximum CPF exposure is 63 pg/g. It was not clear to us why in Figure 1 the range of CPF was truncated.

revealed a tremendous amount of scatter at CPF blood concentrations of 5 pg/g or less, but little scatter at higher blood levels. Consequently, it would seem reasonable to include one or more of these 4 data points in the higher blood concentration range in any calculation, or explain more fully why these data were excluded. EPA has asked for these raw data on at least two occasions. Staff of our nonprofit organization has also asked for these data, suitably anonymized, on several occasions.

3. Results

Figure 1 attempts to replicate Figure 1A from the Rauh et al. (2011) publication, specifically Working Memory. This replication seems reasonable from a comparison of where the regression lines lie in relationship to the high dose points in either figure, despite the fact that we are missing approximately 35% of the data stated to be available in Rauh et al. (2011) (see Appendix B for our raw data and Appendix C for a comparison

---

CCCEH response: The maximum CPF exposure in the sample was indeed 63 pg/g. The number of children with CPF levels above 25 pg/g were 4. The x-axis was truncated at 25 pg/gm for the following reasons:

1) One of the subjects did not have the outcomes measured.

2) The subject with 63 pg/g was a highly influential observation (outlier) and drastically impacts inference. This was confirmed based on residual analysis in most analyses. Where appropriate, this observation was removed from the analysis. This influence was observed in the spline plots as well and this lone outlier at the extreme end of the exposure made the plot unstable and uninformative.

3) With just two observation left in this range, the data were too sparse and the splines too unstable in this region.

Moreover, being exploratory in nature, the spline plots were constructed to assess the adequacy of a linear relationship between log-transformed CPF and WISC scores. We therefore restricted the splines to the range of CPF values where the data were not sparse, and the curves were stable.

between data sets). However, and importantly, Rauh et al. also do not include high dose data on their charts (e.g., see reference to 63 pg/g on Rauh et al. page 1198, column 2, which is not found on Figure 1A). Apparently also missing is one child with a value of 32 pg/g (see stated CPF range in Rauh et al., Table 1).

Figure 2 reflects the data from Figure 1 plotted with the response unadjusted and the dose in logarithmic units (log_{10}). Figure 2 shows a reduced effect on Working Memory Index when compared with Figure 1, found by comparing where the regression lines lie in relationship to high dose data points. This indicates that the way Rauh et al. (2011) presented the data may have had an effect on interpretation. Moreover, the response y-axis is not compressed in Figure 2, eliminating the subtle visual effect of downward trend due to this compression of the y-axis found in Figure 1. The R^2 for both regression lines are very small, which indicates that chlorpyrifos does not well explain the data variability (i.e., the scatter in the data). Confirmation of this small R^2 value, which was not otherwise reported by Rauh et al. (2011), would depend on access to the full data set, however. This would also enabled us to do a comprehensive data analysis similar to Rauh et al. (2011).

Figure 3 attempts to replicate the findings of Rauh et al. (2011), specifically their Figure 1E (Full Scale IQ). This replication is not as close as Figures 1 and 2. Again, compare where the regression lines lie in relationship to high dose points in either Figure 3 or Rauh et al. (2011) Figure 1E. As in the previous comparison of Figures 1-2, some of the

data stated to be available in Rauh et al. (2011) Figure 1E are missing (in this case approximately 15%; see Appendix B for our raw data).

Figure 4 reflects the data from Figure 3 plotted with the response unadjusted and the dose in logarithmic units ($\log_{10}$). Figure 4 shows no effect on Full Scale Composite score when compared with Figure 3. As before, the y-axis is not compressed in Figure 4, eliminating the subtle visual effect of downward trend due to this compression in the y-axis of Figure 3. The R^2 is for Full Scale IQ is even smaller than for Working Memory, suggesting that chlorpyrifos is a poor predictor of the outcome. As with Figure 2, confirmation of this small R^2 value would depend on access to the full data set.

The bottom line of this simple reanalysis is that evidence of effect for Full-Scale IQ does not exist when the study data are presented in another manner (Figure 4). Working Memory shows a lessened evidence of a negative statistical association with dose (Figure 2), but this evidence is problematic due to missing data, including data for the highest exposed individual that Rauh et al. (2011) state "was a highly influential observation (outlier) and drastically impacts inference."

Overall, the lack of raw data from this study makes statistical analysis and confirmation of the authors' data and results, a hallmark of scientific inquiry, impossible. Because we are unable to confirm the findings of Rauh et al. (2011), we cannot make a judgment of human neurological effects as a possible replacement of cholinesterase inhibition as the critical effect for CFP.

4. Discussion

The most significant challenge, by far, in any reanalysis of the Rauh et al. (2011) study is the absence of data to conduct a credible replication to confirm their data analysis. For example, Rauh et al. (2011) state that:

> "Of 725 consenting women, 535 were active participants in the ongoing cohort study at the time of this report, and 265 of their children had reached the age of 7 years with complete data on the following: *a*) prenatal maternal interview data, *b*) biomarkers of prenatal CPF exposure level from maternal and/or cord blood samples at delivery, *c*) postnatal covariates, and *d*) neurodevelopmental outcomes."

However, the Results section of Rauh et al. (2011) show a series of 5 graphs, each of which would be expected to offer a more or less complete picture of effects based on 265 children. Yet, our analysis of two of these graphs (Figures 1A and 1E) show ~35% or ~15% missing data points, respectively, and neither of these graphs include the data points from the highest cord blood CPF exposures of 63 pg/g, or another higher dose data point of 32 pg/g, as stated by Rauh et al. (2011, Table 1) (see Appendix B for our raw data).

Despite these missing data, what do our analyses show? Although negative neurological associations are reported in the Rauh et al. (2011) with CPF exposure when a plot of Working Memory Composite Scores are normalized by their natural logarithm, and plotted against dose, this manner of data display is not the only one possible. A more typical risk assessment approach would be to plot the unadjusted scores, which are already expected to be normally distributed in the human population (as per Rauh et al., 2011 Experimental Design), against the logarithm of dose.

When the results of Rauh et al. (2011) are plotted using logarithmic scales in this way, a reduced association is found. For example, Figure 1 is a representation of Rauh et al.'s Figure 1A plotted as the natural logarithm of response versus dose. Figure 2 shows these same data, but where the response, Working Memory, is plotted as unadjusted response versus $\log_{10}$ dose. A comparison of Figures 1 and 2 show that the negative trend of Figure 1 for the Working Memory is less in Figure 2. When a similar analysis is performed for Full Scale IQ (or Composite Score), the slight negative trend of Figure 3, which is a representation of Rauh et al. (2011) Figure 1E, disappears; compare Figures 3 and 4.[6]

Whether one method of plotting these data is superior to another may be important, but a strong true association should not be affected by the method of data plotting. A more appropriate, scientific approach to confirm the findings of Rauh et al. (2011) would be to

[6] What about including the missing high dose data? Adding the two high dose data points described in Rauh et al. (2011) to figures 2 and 4 and supposing only average responses further decrease the negative slopes, but only slightly (data not shown). This indicates even less of an effect, if any, from chlorpyrifos exposure.

have access to the underlying raw data, suitably redacted for confidentiality concerns. For example, access to the raw data would enable us to discuss our results in more statistical terms, by comparing the differences in the slopes of the regressions and the low R^2 values. This would allow a stronger statement on whether a statistically significant association is found (or not). Further, access to the raw data, suitably redacted, would allow us and others to adjust for confounding factors as was performed by Rauh et al. (2011) in their regression. Moreover, we might be able to refine our analysis from a simple linear approach to an alternate approach in a manner similar to that shown by Rauh et al. (2011) who presented a smooth cubic spine curve.

We acknowledge that our analysis from published graphs is a rudimentary way to obtain the raw data of Rauh et al. (2011), because data points may often overlay one another in published figures.[7] Still, such an analysis of the Rauh et al. (2011) data shows that no CPF exposures greater than 25 pg/g are plotted. In fact, correspondence between EPA and Raul et al. suggest that statistical significance is lost when at least one high dose point is included in the regression calls for more data transparency.

Not surprisingly, as co-sponsors of the study, scientists with the EPA have asked for the raw data from Rauh et al. (2011) and earlier publications (EPA, 2016). Such a request would seem reasonable, because as described by Rauh and coworkers:

[7] Are all of the missing data points in the Rauh et al. (2011) Figures 1A and 1E are underneath the other points? One point is ~0.01% of the graph area and all data points combined covers less than ~2% of the graph area. There are 265 children in the study, but only approximately 170 data points observable in Figure 1A. Thus, the chance that all of the missing data points are hidden below other data points appears to be nil. Rather it appears that many of these data points were not added to these figures.

"This study was supported by the National Institute of Environmental Health Sciences (grants 5P01ES09600, P50ES015905, and 5R01ES08977), the U.S. Environmental Protection Agency (grants R827027, 8260901, and RR00645), the Educational Foundation of America, the John and Wendy Neu Family Foundation, the New York Community Trust, and the Trustees of the Blanchette Hooker Rockefeller Fund."

As EPA has noted in its request for the raw data, its scientists are familiar with rules for handling confidential data. Moreover, personal information of the subjects can be redacted while maintaining the ability to replicate findings, or a trusted third party can certify the results as suggested by Perignon et al. (2019). Scientists with our organization have also asked for these raw data on several occasions, and have noted that an anonymized data set would be sufficient.

Obtaining these raw data, suitably redacted, would also allow the exploration of additional risk assessment questions that the Rauh et al. (2011) study suggest. For example, how is it that the full scale composite score graph of Rauh et al. (2011; Figure 1 E) has more data points than Rauh et al. (2011) graph of working memory composite score (Figure 1A), if the former depends on the latter? Is this due to data-point overlay?

Also, according to Rauh et al. (2011), umbilical cord blood samples were not collected at birth in 12% of the study population. Nor were prenatal blood lead levels, a known

neurological risk for children, collected for 66% of the maternal study population. In addition, blood lead samples were only collected in 89 out of 265 children, or 34%. Could these deficiencies be addressed in a reasonable way for risk assessment purposes?

Rauh et al. (2011) is an association study, and as such is very good at suggesting hypotheses. One hypothesis in particular, would be that neurological effects of CPF occur at doses lower than cholinesterase inhibition. However, this hypothesis has been tested in part in a number of experimental animals and found not to be supported. Specifically, neurological effects do occur in experimental animals, but only at doses that are more than 100-fold greater than those showing these associations (EPA, 2016). Although experimental animal studies are not typically able to monitor for the types of neurological effects associated with increasing CPF dose in Rauh et al. (2011), this disparity in dose makes using such human data more difficult in risk assessment, especially when human and experimental animal studies are similar in dose with respect to cholinesterase inhibition, the current critical effect for CPF.

Finally, the metabolite responsible for the toxicity of cholinesterase inhibition, CPF-oxon, is formed in the liver and 99% of this oxon derivative irreversibly binds to cholinesterase in the blood (EPA, 2014). Since it is so bound, it might not be expected to reach the brain to affect neurological development of the fetus at levels much lower than levels that do not otherwise show any effect on the sentinel blood enzyme. In fact, an analysis by Marty et al. (2012) showed no systemic bioavailability, nor any brain cholinesterase inhibition with the CPF-oxon at doses comparable to the established safe doses.

Risk assessments based on human data are generally preferred over those based on experimental animals when such data can be verified. A risk assessment for CPF should be no different. Epidemiology studies showing neurological effects at doses lower than the well-established critical effect of cholinesterase inhibition, such as by Rauh et al. (2011), should carefully address other factors that are known to affect neurodevelopmental effects in infants and children. This is especially important for CFP in light of the fact that the specific results described by Rauh et al. (2011) are not generally reproduced in other epidemiology studies as cited by Burns et al. (2013) and Reiss et al. (2015).

5. Conclusion

One of the papers from the CCCEH, specifically Rauh et al. (2011), shows a statistical association between CPF exposure and measures of intelligence that is at odds with the determination of cholinesterase inhibition as the critical effect by numerous authorities. This study raises a number of challenges due in part to the absence of data to confirm findings and the reluctance by the authors to share these data, suitably anonymized, despite multiple requests by several groups and despite public funding in part. An analysis of the published figures of Rauh et al. (2011) shows that up to 35% of the data appear to be missing, and that the associations are lessened or no longer apparent when different logarithmic assumptions were used in the reanalysis. This lack of transparency makes it difficult to confirm this study's findings and raises serious scientific doubt about

the use of such data in public decisions.

6. Acknowledgements

The origin of this work came about from briefings of MLD as a U.S. Environmental Protection Agency senior advisor in 2017. Afterwards, DowAgro Sciences funded, in part, an independently developed report by Toxicology Excellence for Risk Assessment (TERA), which confirmed the findings of EPA staff. The development of this manuscript and several presentations has been supported by the Internal Development Reserve funds of TERA.

7. References

Burns, Carol J., Laura J. McIntosh , Pamela J. Mink , Anne M. Jurek and Abby A. Li. 2013. Pesticide Exposure and Neurodevelopmental Outcomes: Review of the Epidemiologic and Animal Studies. Journal of Toxicology and Environmental Health, Part B Critical Reviews. Volume 16, 2013 - Issue 3-4.

EPA (United States Environmental Protection Agency), 2014. Chlorpyrifos: Revised Human Health Risk Assessment for Registration Review. Office of Chemical Safety and Pollution Prevention. Available at: http://op.bna.com.s3.amazonaws.com/env.nsf/r%3FOpen%3Ddscz-9shll9

EPA (United States Environmental Protection Agency), 2016. Revised Human Health Risk Assessment for Registration Review. Office of Pesticide Programs, Washington, DC. December 29. Available at: https://www.epa.gov/ingredients-used-pesticide-products/revised-human-health-risk-assessment-chlorpyrifos. https://www.regulations.gov/document?D=EPA-HQ-OPP-2015-0653-0454

Health Canada. (2003). Proposed acceptability for continuing registration: Phase 2 of the re-evaluation of chlorpyrifos. Ottawa, Canada. PACR2003-

Marty, M.S., Andrus, A.K., Bell, M.P., Passage, J.K., Perala, A.W., Brzak, K.A., Bartels, Beck, M.J., Juberg, D.R., 2012. Cholinesterase inhibition and toxicokinetics in immature and adult rats after acute or repeated exposures to chlorpyrifos or chlorpyrifos–oxon. Reg. Toxicol. Pharmacol. 63 (2): 209-224. Available at: http://works.bepress.com/melissa_beck/13/

Perignon, C., K. Gadouche, C. Hurlin, R. Siberman and E. Debonnel. 2019. Certify reproducibility with confidential data. Science. Vol 365 (644():127-128.

Rauh, V., Arunajadai, S., Horton, M., Perera, F., Hoepner, L., Barr, D.B., Whyatt, Rauh, R., 2011. Seven-Year Neurodevelopmental Scores and Prenatal Exposure to Chlorpyrifos, a Common Agricultural Pesticide, Environ Health Perspect. 119(8): 1196-1201. doi:10.1289/ehp.1003160

Reiss, Richard, Ellen T. Chang, Rudy J. Richardson and Michael Goodman. 2015. A review of epidemiologic studies of low-level exposures to organophosphorus insecticides in non-occupational populations. Critical Reviews in Toxicology. Volume 45, 2015 - Issue 7.

United Kingdom. 2003. Chlorpyrifos human health review: Evaluation of further papers requested by the Advisory Committee on Pesticides. York, UK. ACP 6 (299/03).

WHO (World Health Organization). (2004). Chlorpyrifos in drinking-water: Background document for development of WHO Guidelines for Drinking-water Quality. World Health Organization, Geneva. WHO/SDE/WSH/03.04/87.

Zhao, Q., Dourson, M., Gadagbui, B., 2006. A Review of the Reference Dose (RfD) for Chlorpyrifos. Reg. Toxicol. Pharmacol. 44(2):111-124. https://doi.org/10.1016/j.yrtph.2005.10.003

Figures