Author’s response to reviews

Title: THE RAMAZZINI INSTITUTE 13-WEEK PILOT STUDY ON GLYPHOSATE-BASED HERBICIDES ADMINISTERED AT HUMAN-EQUIVALENT DOSE TO SPRAGUE DAWLEY RATS: EFFECTS ON DEVELOPMENT AND ENDOCRINE SYSTEM

Authors:

Authors. Manservisi (manservisif@ramazzini.it)
Corina Lesseur (corina.lesseur@mssm.edu)
Simona Panzacchi (panzacchis@ramazzini.it)
Daniele Mandrioli (mandriolid@ramazzini.it)
Laura Falcioni (falcionil@ramazzini.it)
Luciano Bua (bual@ramazzini.it)
Marco Manservigi (manservigim@ramazzini.it)
Marcella Spinaci (marcella.spinaci@unibo.it)
Giovanna Galeati (giovanna.galeati@unibo.it)
Alberto Mantovani (alberto.mantovani@iss.it)
Stefano Lorenzetti (stefano.lorenzetti@iss.it)
Rossella Miglio (rossella.miglio@unibo.it)
Anderson Andrade (martino.andrade@gmail.com)
David Kristensen (david@moebjerg.com)
Melissa Perry (mperry@gwu.edu)
Shanna Swan (shanna.swan@mssm.edu)
Jia Chen (jia.chen@mssm.edu)
Fiorella Belpoggi (belpoggif@ramazzini.it)

Version: 1 Date: 08 Jan 2019
Author’s response to reviews:

Tuesday, January 8, 2019

Reviewer reports:

Reviewer #1: The goal of this study was to determine how human-relevant exposures to glyphosate or glyphosate-based herbicides (GBH) in Sprague-Dawley rats affected developmental and endocrine outcomes at differing life stages. Comparisons of the commercial product and the active ingredient is a strength of the study. Several significant effects were found: for AGD (larger, especially in Roundup groups), hormones in males: DHT (13-week Roundup), TSH (6-week glyphosate, 13-week Roundup), BDNF (6-wk Roundup), fT/TT ratio (6-wk glyphosate, 13-wk Roundup), E2/SHBG (6-wk Roundup); and in females, the DHT/TT ratio and TT/SHBG ratios (13-wk Roundup). The study is sound and the data suggest that endocrine systems are disrupted by glyphosate and/or Roundup, with more effects of the commercial product. The authors are commended for excellent statistical data analysis.

Critique:

1. The Discussion needs to be reframed. Authors focus on several endpoints that were affected (TSH, BDNF, steroid hormones, etc.) and they do a lot of speculation based on single-point hormone measures. For example, increased TSH in the absence of measuring other thyroid hormones is not terribly informative about "subclinical reduced thyroid function" proposed by authors (line 439). The discussion on both thyroid function (beginning line 488) and BDNF (line 500) can be shortened, as the biological meaning is unknown. The fact that animals were killed across a large range of times introduces the possibility that diurnal rhythms contributed to differences; and females were not killed on the same day of the estrous cycle, making hormones in females difficult to interpret. Rather, it would be more interesting for authors to discuss important issues such as: why do the 6- and 13 week endpoints differ? Why do results of glyphosate and Roundup differ? The Discussion needs some rewriting to focus on key findings.
Response: Thank you for the comment. Following the reviewer’s suggestions, we have condensed the discussion on hormonal activity, reducing its extension and introducing sub-headers. We are aware that our hormone measurements present some pitfalls (i.e. single-point hormone measures, time of killing, stage of the estrous cycle at termination) and the significant results on hormonal activity need to be interpreted with caution. However, if data are considered with a weight of evidence approach, the outcome of statistically significant differences in hormonal activity together with changes in apical endpoints (such as AGD and delayed first estrous) should be considered a “positive” result as reported by the “OECD Guidance document n.150 on standardised test guidelines for evaluating chemicals for endocrine disruption” (OECD, 2014).

We added in the discussion the following sentence on the data interpretation of our overall results:

“Considering these outcomes with a weight of evidence approach, statistically significant differences in apical endpoints (AGD and FE) together with changes in hormonal activity detected in both the treatment groups should be should be taken into account suggesting evidence of concern for reproductive toxicity via an endocrine disruption mechanism (OECD, 2014). Indeed, a longer AGD at birth in both sexes and an increased age at FE, together with the increased TT in females offspring, are considered endpoints for androgen-mediated activity by the weight of evidence assessment (OECD, 2014).”

We reported the result on TSH in the discussion with more caution and objectivity:

“Finally, we observed a significant increase in TSH in glyphosate-treated males (6-week cohort) and Roundup-treated males (13-week cohort). Due to funding limitations, we could not investigate T3 and T4 yet (this study only aimed to be a pilot study), we are planning this work in the near future. We did not observe histological changes in the thyroid gland, therefore the altered level of TSH together with a normal histological pattern of the thyroid gland are not indicative of a thyroid-related activity. Nevertheless, our findings prompt further detailed investigation on the effect of GBH on thyroid function and development.”

The discussion on both thyroid function and BDNF was clarified:

“Indeed, BDNF is an explorative and new endpoint for neurodevelopment and the utility of neurotrophins as potential biomarkers of thyroid hormones effects in brain is not completely understood. At the moment, any adverse impact of GBH exposure on neurodevelopment through the thyroid axis can only be pointed out as a topic for further investigation.”
The comparison between the results obtained in the two cohorts and between glyphosate and Roundup formulation was provided at the end of the discussion. We understand the importance to give more depth to these issue but we also believe that no more than what we reported could be stressed, as it was a pilot study with few animals/group. We only add in the discussion the sentence:

“Our results corroborate prior mixture studies (Defarge et al., 2018), indicating that technical glyphosate and components of formulations may have cumulative (e.g., additive or synergistic) effects on endocrine-sensitive endpoints. Therefore, ADI calculations and other regulatory experiments should be performed with the full formulations and all components should be declared and/or measured”.

2. Sample size information is unclear. Specify the number of dams per group, the numbers of offspring, and how litter was considered as a variable in analyses if multiple littermates were included in the same endpoint group.

Response: We thank the Reviewer for pointing out these omissions and have thus added the missing information in the material and methods:

“Each of twenty-four virgin female SD rats (17 weeks old, 270-315 g) was cohabited outbred with one breeder male rat of the same age and strain. Every day, the females were examined for presence of sperm”.

The numbers related to the offspring rats were already reported in the material and methods:

“After weaning, on PND 28, offspring were randomly distributed in two cohorts: 8M+8F/group animals belonging to the 6-week cohort were sacrificed at PND 73 ± 2, i.e. 6 weeks after weaning; 10M+10F/group animals belonging to the 13-week cohort were sacrificed at PND 125 ± 2, i.e. 13 weeks after weaning. After weaning, the offspring (F1) were treated through drinking water until sacrifice. Altogether, 108 SD rats (54 males and 54 females) were enrolled in the post-weaning treatment phase”.

When the observations were performed on the entire litter, a fixed and mixed effect models were estimated (litter as random effect) and both reported, as for the AGD analysis.
3. The weakness that females were not euthanized on the same day of the cycle, and that rats were killed sometime between 9 am and 3 pm, is acknowledged by the authors. They do not state, though, whether they plan to correct these experimental design weaknesses in future work - please do that. The wording describing blood collection [Line 269 ("Blood collection was continuous from 9 a.m. to 3 p.m.")] is confusing and should be changed. Blood collection was not continuous - that implies multiple sampling within animals, which was not the case.

Response: We are aware that time of killing and the stage of the estrous cycle might have influenced the variability and absolute concentration of the hormone determinations. However, the issue of sacrificing animals in the same cycling period (e.g estrous) is still controversial. The updated OECD Test Guidelines on reproductive-developmental toxicity do not require the sacrifice of females in the same stage of estrous, they only recommend: “vaginal smear is examined on the day of necropsy to determine the stage of the estrous cycle and allow correlation with histopathology in reproductive organs”.

We added in the discussion the following sentence:

“However, even if sacrificing animals on a specific day of the cycle might improve the ability to observe changes in the baseline hormone levels, the issue of sacrificing animals in the same cycling period (e.g estrous) is still controversial. The updated OECD Test Guidelines on reproductive-developmental toxicity do not require the sacrifice of females in the same stage of estrous, only the examination of estrous cycle on the day of necropsy is recommend to allow correlation with histopathology in reproductive organs (OECD 2018).”

We are planning, in our future work, to increase the number of animals for hormone measurements (20 M + 20 F/group) in order to allow the statistical analysis of the data by avoiding the possible bias of killing animal the same day of the cycle.

Regarding the time of killing, we agree with the reviewer that collection of blood for hormones should ideally be carried out at a comparable time of day in case of diurnal variations but the technical procedures for clinical chemistry (sampling, blood specimen application to the hematology and chemistry analyzer) require a waiting period between a necropsy and the other. This was the reason of a prolonged scheduled time for necropsies.
We added a sentence in the discussion in order to clarify this point of concern:

“Several hormones were measured in the dams and offspring, but not all hormones were measured in all the animals, due to insufficient material for a complete data set of hormone profiling after the full scale hematology and clinical biochemistry (data not yet published at the time when this work is presenting). Furthermore, the number and timing of blood sample collection was limited to the final sacrifice of animals, considering that this was a pilot study and that in vivo blood sampling could lead to maternal and pups stress”

We also changed in the material and methods the sentence "Blood collection was continuous from 9 a.m. to 3 p.m….", replacing it with:

“The necropsy session was scheduled from 9 a.m. to 3 p.m, time needed to first collect blood and perform hematology and chemistry analyses, then the necropsy of the rat. The time and date of necropsy were recorded”.

4. Hormone assays: Details on each hormone assay coefficient of variation need to be provided.
Response: The data on each hormone assay coefficient of variation has been provided as supplementary material (Fig. S4-5).

5. The Introduction is far too long and is a review of the entire literature. Please condense.
Response: We understand the reviewers’ critic, nevertheless in our introduction we aimed to provide the reader the necessary background information for interpreting our findings, which include an overview of the studies on several outcomes and endpoints related to reproductive and developmental toxicity.

Minor points:

6. Line 85 ("In vivo, sexual development…"), cite the original literature on sexual development, not Dallegrave et al. 2007.
Response: We have deleted the reference of Dallegrave et al. 2001 in the sentence “In vivo, sexual development is controlled by hormones and is therefore highly sensitive to exogenous substances with endocrine-related effects”.
7. Line 292, sperm analysis: What do authors mean by single male animals? N=1 per group? This is not a meaningful sample size and this section should be excluded if this is the case.

Response: We apologize for the misunderstanding. Each animal was evaluated for sperm analysis. We have rephrased the sentence in the material and methods:

“Sperm analyses were performed on each male animals from both cohorts, at scheduled necropsies on PND 73 ± 2 and PND 125 ± 2”.

8. Results: The use of sub-headers would be helpful in delineating separate outcomes.

Response: We appreciate the Reviewer’s suggestion and we introduced sub-headers in the section reporting the results and in the discussion, what gives the readers the opportunity to scan the list of outcomes and to locate the information they want.

9. Table 8a: In the table legend, what do the lettered footnotes mean (e.g., a refers to 7 out of 8, etc.)? 7 out of 8 what? Weren't the hormones measured in all animals?

Response: Yes, as already pointed out we didn’t measure hormones in all animals for two reasons: 1) the total costs allocated for this pilot study couldn’t cover the hormone analysis in all the experimental animals; 2) plasma and serum samples from each animal were subjected to many investigations, among which full-scale haematology and clinical biochemistry. The amount of material was not always sufficient for a complete data set of hormone profiling.

10. Title. Why is it important to refer to the Ramazzini Institute? This is extraneous information.

Response: Yes, we agree that the name of the Test Facility performing the study in the Title is quite unusual. However, this idea came directly after a clear position of the Ramazzini Institute asking for a re-evaluation of Glyphosate and GBHs with solid independent results, obtained by a shared research project on which regulatory agencies and policymakers can confidently base their risk assessments and their evaluations, including the upcoming decision for the reauthorization for glyphosate use in Europe in 2022 [Landrigan P and Belpoggi F, 2018]. The pilot study, started in 2015, represents the starting point of an integrated long-term project that we will begin in 2019, whose results will be available by the time of the next EU decision.

Reviewer #2: This manuscript describes the results of a pilot study investigating endocrine disrupting properties of glyphosate, and of a Roundup formulation. This pilot study is well conducted and well interpreted. It is a very interesting study because it is an unexplored topic. This study provides an answer to a very important gap. Authors have performed a large range of analysis. The main gap of this study is that only one dose has been tested. Most statistically significant differences are scattered and it is difficult to find a clear toxicity pattern. It is thus very difficult to understand if the health effects which are detected in this study are dose dependent and if they would be replicable in other studies.

Addressing the few comments below would improve the clarity of the study and could provide more insights into the consistency of the statistical differences.

L 60 - In the conclusions of the abstract, I would specify that it is a pilot study
Response: Thank you, specification added.

L 77 and further - Which Roundup formulation? It is an important detail since formulations are likely to have different effects. This remark is valid for other citations in the manuscript. In the case where several formulations have been tested, better stay general and mention GBH. The name 'Roundup' is a trade name referring to the formulation 'Roundup' also called MON 2139 (glyphosate 360g/l and ethoxylated tallowamine). The Roundup herbicide tested in this study, Roundup Bioflow, is the representative EU formulation 52276 (https://www.sdslibrary.monsanto.com/Lists/MSDS%20Library/DispForm.aspx?ID=151). This is important to take into account as it is likely that this formulation does not contain an ethoxylated tallowamine surfactant because they were banned from the European market.
Response: The commercial formulation used in this study, Roundup Bioflow, was the representative formulated product recently evaluated for the renewal of the approval of glyphosate in EU and considered in the European Food Safety Authority peer review (MON 52276). These specifications were reported in our previous paper describing the study design [Panzacchi et al. 2018]. We agree that it is equally important to report also in the present manuscript the characteristics of the GBH used in our study, so we added the following paragraph:

“It is noteworthy that the commercial formulation used in this study, Roundup Bioflow, was the representative formulated product recently evaluated for the renewal of the approval of glyphosate in EU and considered in the European Food Safety Authority peer review (MON 52276)”

L 109 - 'No evidence of interaction of glyphosate and GBHs'. Are you sure that the EDSP program included tests on GBHs?

Response: Thank you for the right Reviewer’s observation, EDSP program did not include tests on GBHs. We deleted the word “GBHs”.

L 111 - 'EPA dismissed statistically significant differences consistent with estrogenic activity'. Was there a consistent profile indicating estrogenic effects across different assays? It would be more accurate to indicate which assays were found to be positive.

Response: Plasma vitellogenin (VTG), a reproductive biomarker in the Fish Early Life-Stage Toxicity (Threespine Stickleback) assay, was significantly decreased 55% (p<0.05) in female fish at the mid-high treatment level (6.2 mg a.i./L) compared to the negative control.

We now specified these details in the sentence:

“However, in Fish Early Life-Stage Toxicity (Threespine Stickleback) assay, EPA dismissed statistically significant differences in plasma vitellogenin, consistent with estrogenic activity, because of a non-monotonic dose response (Vandenberg et al. 2017)”

L143 - 'exposure to GBHs' -> exposure to a GBH

Response: revised.
Was the analysis of estrous cycle characterization and sperm analysis, and other aspects of the experiment, performed in a blinded manner?

Response: Yes, it was.

L 355 - Which statistical software have you used?

Response: The statistical analysis was performed using Stata/IC 10.1 (for all regression) and Statistix 10 (for all the other tests); graphs were obtained using Excel.

L 380 - If your statistical software can do it, it is always better to indicate the exact p-values instead of thresholds. A p-value is a probability and it's value is a very important piece of information.

Response: We agree that reporting individual exact p-values is a more elegant and precise way to express statistical evaluations, but in line with common practices at NTP and IARC and in order to simplify for the reader the difference between significant and non-significant results, we preferred to indicate $p < 0.05$ or $p < 0.01$ (when statistically significant) and use the exact p-value only for border-line case (i.e. $p = 0.056$).

L 402 - 'borderline significant ($p = 0.056$)', I would avoid calling this borderline 'significant' since it is not significant

Response: Thank you for this correction. We rephrased the sentence: “Plasma TSH levels showed an increase, even if not statically significant ($p = 0.056$) in the glyphosate-treated males and a marked and significant increase in Roundup-treated males versus control ($P = 0.0004$).”

Table 2, and other tables. It will improve the clarity of the article if the authors could provide dot plots showing the spread of the data for some important statistical differences. I understand that it will not be reasonable to do this for every variable, but the spread of the data is a very important information to evaluate the quality of a dataset and the accuracy of interpretations. It can be provided as a supplementary material.
The addition of these dot plots may provide important information for some statistical differences. For instance, I am wondering if the statistical difference for the age at first oestrous is not due to potential outliers since the variance for this point is larger in comparison to the variance for glyphosate or for the control. In another case, I would like to see to which extent the spread for the male AGD values are overlapping if the effect size is very clear. The most important is perhaps the effect on DHT levels in males for the 13-week cohort. It seems that it is the most pronounced effect in this study. The exposure to Roundup Bioflow decreased the DHT serum levels by a factor of 10.

Response: We added as supplementary material the box and dot plots of AGD (Fig. S1), age at first oestrous (Fig. S2), and DHT (Fig. S3).

Were there toxic effects in other tissues that could explain these endocrine disrupting effects?
Response: The histopathological findings on other tissues are currently under evaluation.

Reviewer #3: This manuscript describes one of the few studies that has compared the effects of glyphosate alone with a glyphosate-based herbicide that is representative of formulations in use within Europe. The authors test a single but regulatory relevant dose of glyphosate/Roundup (US chronic reference dose). An important component of the experimental design is that exposure to the test substances was initiated pre-natally, representing a more realistic scenario with respect to human exposures. The authors have conducted a very thorough analysis to evaluate effects on hormone systems and developmental/reproductive consequences. All analyses are conducted to a very high technical standard and thus the results obtained are compelling as far go for a single dose experiment. Given the controversies surrounding the toxicology of glyphosate based herbicides this is a very timely article that will be found of interest to many in the field and by the general public.

The authors find statistically significant outcomes in a number of measurements suggesting disruption of hormone systems (eg TSH, DHT, BDNF). Associated developmental defects observed were an increase in anogenital distance (males and females) and delay in first estrous. Interestingly, more frequent and pronounced disturbances were found in the Roundup compared to the glyphosate alone treatment groups. This suggest that the co-formulants/adjuvants in the Roundup are a significant contributing factor to the observed physiological/biochemical disturbances. Again this is of importance with respect to human health as populations are exposed to the commercial herbicide formulation and not just glyphosate alone.
Nevertheless, the following points should be addressed before publication should be considered.

1. Anogenital distance (AGD) was measured at post-natal day 4 and found to be increased in Roundup treated males and females and in glyphosate treated males. Was this increase in AGD maintained or did it resolve and return to normal as the animals aged to adulthood?

Response: We did not measure AGD later in life. Recent epidemiological studies are exploring if AGD at birth would correlate to the AGD later in life, representing a phenotypic signature throughout life [Priskorn L. et al, 2018]. The animal literature has focused mainly on the relationship between neonatal AGD and reproductive tissues weights and malformations measured in adulthood. Few rodent studies have explored the postnatal plasticity of AGD [Van den Driesche et al. 2011; Kita DH et al. 2016]. It seems that under normal physiological conditions and environmentally relevant exposure scenarios to endocrine disruptors no significant changes in AGD are expected in postnatal life, although it can also be slightly responsive to changes in the androgen environment following pubertal exposure [Kita DH et al. 2016]. Even if the usefulness of AGD as a long-life marker of prenatal androgen environment in rodent bioassays needs to be validated, we will consider to explore this marker in future experiments.


2. Between lines 85 to 98 the authors summarise findings from other studies of Roundup formulations that found alterations in hormone systems in vivo. However, the authors fail to refer to study by Seralini and colleagues (Seralini R et al. Environ Sci Eur. 2014;26(1):14), which also found hormone (testosterone, estrogen) changes in response to even a very low dose of Roundup over a 2 year period of treatment.
Response: Since we there was no consensus among the co-authors about the relevance of this reference, we preferred not to cite it.

3. Line 144: "Admissible" should be "Acceptable".

Response: Corrected.

4. Lines 400-401: it is stated that "for females only a few samples were available for these further analyses". Why?

Response: The initial focus of our pilot study was to assess techniques and methods for glyphosate detection in different matrices [Panzacchi et al., 2018], then to evaluate endocrine disrupting activities, together with microbiome alterations [Mao et al., 2018], target organ toxicity, genotoxicity and omics (data not yet published). For these reasons, plasma and serum samples from each animal were subjected to many investigations, among which full-scale haematology and clinical biochemistry. In addition, blood from all animals was taken and stored for possible analysis at a later time to help clarify equivocal effects or to generate internal exposure data. If more funding will be available, we will further explore the missing hormones from our back-up stored samples.


As per our response to Reviewer#1 we added the following sentence in the discussion:

“Several hormones were measured in the dams and offspring, but not all hormones were measured in all the animals, due to insufficient material for a complete data set of hormone profiling after the full scale haematology and clinical biochemistry (data not yet published at the time when this work is presenting). Furthermore, the number and timing of blood sample collection was limited to the final sacrifice of animals, considering that this was a pilot study and that in vivo blood sampling could lead to maternal and pups stress”

5. Lines 445-446: it is stated that "Our findings suggest that both Roundup and glyphosate result in an increased in utero exposure to androgens". It is not clear to this reviewer how the authors arrive at this conclusion since androgen measurements were not taken during pregnancy. Androgen measurements are only shown in the offspring. This point needs to be clarified.

Response: We understand the Reviewer’s concern and we agree that the assessment of serum/plasma levels of hormones (particularly testosterone) in dams during gestation and/or in exceeding pups during perinatal life could be extremely useful, although not specifically required by the current reproductive/developmental toxicity guidelines [OECD TG 443]. As this was a pilot study, the number and timing of blood sample collection was limited to the final sacrifice of animals, also considering the possible source of maternal stress induced by in vivo blood sampling during gestation. As a result, the increased serum testosterone in females offspring together with a longer AGD both in males and in females at birth, are considered endpoints for androgen-mediated activity by a weight of evidence approach, as reported in the “OECD guidance document on standardised test guidelines for evaluating chemicals for endocrine disruption” [OECD Guidance document n. 150]. We added in the manuscript a sentence reporting the current OECD guidance on how to interpret changes indicative of endocrine disruption.

“A longer AGD at birth in both sexes and an increased age at FE, together with the increased TT in females offspring, are considered endpoints for androgen-mediated activity by the weight of evidence assessment (OECD, 2014)”
6. There are a number of measurements that show significant differences at 6 weeks but not at 13 weeks; increase in TSH levels in glyphosate-treated males; Roundup increases BDNF levels in males; increase in E2/SHBG ratio in Roundup treated males; decrease in fT/TT ration in glyphosate treated males. Can the authors offer any suggestions as why this is the case? That is, why were the 6 week differences not maintained at 13 weeks?

Response: In our pilot study, we collected a number of measurements as an initial assessment of all these endpoints that will be useful for our future long-term project on GBHs. It is not rare for potential endocrine disruptive chemicals to induce effects only in certain windows of susceptibility and the study design used (a pilot study with few animals/group and one low dose) can not detect dose-response and long-term effects providing a clear interpretation on the mode of action. We stressed in the text that our results only suggest hormone alterations that need to be interpreted with caution.

7. Lines 503-510: it is difficult to follow the arguments of the authors in this section and thus not possible to understand the point they are trying to make. Clarification is required.

Response: Clarification added:

“Indeed, BDNF is an explorative and new endpoint for neurodevelopment and the utility of neurotrophins as potential biomarkers of thyroid hormones effects in brain is not completely understood. At the moment, any adverse impact of GBH exposure on neurodevelopment through the thyroid axis can only be pointed out as a topic for further investigation.”

8. Line 586: "30.000" should be "30,000".

Response: Corrected.

9. Sentence structure and grammar is poor in many places. The manuscript should be reviewed by a native English speaker to raise to the required standard so that the risk of confusion is avoided.

Response: The revised version has been re-edited and reviewed by two native speakers.