

# Considerations on the statistical methods used to assess carcinogenicity studies of pesticides with emphasis on glyphosate

by Peter Clausing, PAN Germany

.....

# Introduction

Much of the controversy concerning glyphosate being a carcinogenic in animal studies or not revolves around the type of statistical analysis that was used to assess the incidence of tumors. In essence it boils down to the question whether trend-analyses or pair-wise comparisons are more appropriate. In the latter case each group of animals treated with glyphosate is compared individually with the control group. In case of trend analyses (e.g. Cochran-Armitage trend test, Mantel's test for trends or Poly-k test) the data, i.e. incidences of a particular tumor type, are analyzed at once across all groups of the study also taking into account dose-dependent changes from group to group.

# Views of the Reporting Member State, the EFSA and the Peer Reviewers

In the final draft of the Renewal Assessment Report (RAR) on glyphosate dated 31 March 2015 no particular discussion was devoted to the type of the statistical analysis. The Reporting Member State (i.e. Germany's Federal Institute for Risk Assessment or BfR for its German acronym) presented the results just as in the study reports. In case of the controversial mouse carcinogenicity studies these were exclusively pair-wise comparisons. Only one tumor type (malignant lymphoma) in one out of five studies (the one from 2001) exhibited a statistically significant increase. However, after the International Agency for Research on Cancer (IARC) published its monograph on glyphosate on 30 July 2015 the BfR revisited the carcinogenicity studies contained in the RAR and presented the results in an Addendum dated 31 August 2015. There the Cochran-Armitage trend test was used for the analyses demonstrating statistically significant increases of tumor incidences in all five mouse studies.

In the present analysis we leave out our earlier critique (1) of the arguments (historical control data, high-dose effects etc.) used by the BfR and the EFSA to dismiss these findings. Instead we focus on the arguments used during the peer review by some of the Member States and by the EFSA itself to claim that trend tests are not acceptable for the analysis of tumor incidences.

Originally the BfR stated that "the weak increase in malignant lymphoma was clearly confined to the single study and strain since it was not reproducible in four other valid long-term studies" (2, p.65). Then, in the Addendum to the RAR it was claimed that "it should be avoided to base any conclusion only on the statistical significance of an increased tumour incidence identified in a single study without consideration of the

biological significance of the finding" (3, p. iii). Finally, the EFSA highlighted the "lack of statistical significance in pair-wise comparison tests" (4, p.11). Although the BfR as well as the EFSA insist that the assessment was made in an open and transparent manner, the way how the EFSA came to the conclusion that glyphosate does not pose a carcinogenic hazard is all but transparent.<sup>1</sup> Very different opinions were expressed by the European Member

<sup>&</sup>lt;sup>1</sup> The EFSA did not demonstrate or disclose whether and, if so, what <u>valid</u> historical controls were used to dismiss the observed increases in tumor incidences in the mouse studies, after the BfR partly presented invalid historical control data and partly used the available historical control data in an erroneous and distorted manner (cf. 3, p. iii). Neither did the EFSA explain how the change came from a <u>suspicion</u> that the Swiss Albino mice used in the 2001 mouse study might be infected by a virus, based on the reference to just one publication (Taddesse-Heath et al. 2001) to the conclusion that the 2001 study was "not acceptable due to viral infections" (4, p.10) even although it was explicitly stated in the RAR that "it is not known to which extent such a latent infection might have contributed to lymphoma incidences reported earlier or even in the studies described in this RAR (9, p.511).

States during the peer review of the RAR (see Table 1 for details). From this discussion neither a "majority view" nor a consensus opinion becomes apparent.

Comments by	Quote from Comments
Belgium	relying on significant trends to declare a finding eligible for
(p. 872)*	classification is a policy decision not supported by the EU.
EFSA	The statistical analysis methods should be discussed considering that,
(p. 876)	according to the OECD 451 guideline for carcinogenicity studies, "the statistical method most appropriate for the analysis of results, given the experimental design and objectives, should be established <b>before commencing the study</b> (emphasis added).
Norway (p. 885)	Why does the RMS consider the statistical evaluation provided with the study reports as more appropriate than the trend test used by IARC? According to the OECD guidance document on the conduct and design of chronic toxicity and carcinogenicity studies, significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result.
Sweden (p. 888)	It is generally believed that trend tests are more sensitive than e.g. pair-wise comparison methods and from a precautionary assessment point of view trend tests would be preferable.
UK	We would question the suitability of relying on a trend test where
(p.986)	the incidence in controls is zero and the only group responding, with no
	pairwise significance, is the top dose group.
EFSA Executive Director	the studies under consideration were designed for pair-wise comparisons.
(5, p.6)	

Table 1: Opinions expressed on statistical analyses (4, 7)

\*page numbers refer to the PDF-page number of the document posted on EFSA's Website (7)

Annotations on the comments listed in Table 1

# Comment by Belgium

It should be noted that contrary to this statement trend tests have been accepted in the (Renewal) Assessment Reports for other pesticides, e.g. for 2,4-D, carbaryl, diazinon, ethylene, isoproturon and trifluralin. Belgium's reference to a "policy decision" actually deviates from the science-based approach repeatedly claimed by BfR and EFSA. In fact isoproturon had been proposed as a CLP class 2 carcinogen based on trend tests in the Final RAR of April 2015 (8).

#### Comment by EFSA

First a minor, but important deviation from the original text in the OECD should be noted:

While the OECD refers to "the statistical methods most appropriate" (implying that more than just one method can and should be defined in the study plan) the EFSA quotes this part of the OECD guideline not in plural but in singular ("the statistical method most appropriate").

More importantly however, the EFSA missed to provide the full quote; that is:

"The statistical methods most appropriate for the analysis of results, given the experimental design and objectives, should be established before commencing the study. **Issues to consider include whether the statistics should include adjustment for survival, analysis of cumulative tumour risks relative to survival duration, analysis of the time to tumour and analysis in the event of premature termination of one or more groups. Guidance on the appropriate statistical analyses and key references to internationally accepted statistical methods are given in Guidance Document No. 116, and also in Guidance Document No. 35 on the analysis and evaluation of chronic toxicity and carcinogenicity studies (20)" (10, p.2-3, emphasis added).** 

The considerations emphasized by the OECD do not refer at all to the question whether pairwise comparisons or trend tests should be used. The description of different circumstances to be considered which cannot be foreseen at the commencement of the study implies that a proper definition of statistical methods in the study plan should take into account the various options that can be encountered during the course of the study. Besides, the Guidance Document No.116 contains a flow-chart that **explicitly recommends trend tests** for the assessment of **histopathological findings** and **tumor** incidences (11, p.123).

#### Comments by Norway and Sweden

While Norway emphasizes that the OECD Guidance states "significance in either kind of test (i.e. pair-wise comparison or trend test, P.Cl.) is sufficient to reject the hypothesis that chance accounts for the result", (7, p.885) Sweden points out that "from a precautionary assessment point of view trend tests would be preferable." In other words, these peer-reviewing member states are in favor of accepting the results of trend tests (7, p.888).

# Comment by the UK

The UK statement quoted in Table 1 has a very limited impact on the overall outcome concerning the applicability of trend tests. The tumor incidences of five mouse studies and two rat studies were listed in the Addendum to the RAR, yielding a total of 23 statistical analyses. Just three of these 23 analyses had a zero incidence in controls with the top dose group as the only group responding. Of the other 20 analyses (including analyses for separated and combined carcinoma/adenoma incidences) 7 yielded a significant increase.<sup>2</sup>

#### Statement by EFSA's Executive Director

To "design" carcinogenicity studies for pair-wise comparisons is simply not possible. The comment of EFSA "the study was designed for pair-wise comparison" implies that the design of a study that uses trend-test would be different; which is not true. The study design for pair-wise comparisons is the same as the study design for trend tests.

<sup>&</sup>lt;sup>2</sup> If a one-sided error probability would have been applied the number of significant increases would be 11. In the context of carcinogenicity studies the use of a one-sided error probability is scientifically justified, although rarely used. The scientific rationale behind using a one-sided error probability is that in terms of hazard assessment one can assume a one-directional change, i.e. an increase of the tumor incidence.

A healthy world for all. Protect humanity and the environment from pesticides. Promote alternatives.

# General considerations concerning the statistical analysis of carcinogenicity tests

The OECD Guidance document 116 states that "the specific statistical analyses are tactical methods used to help answer the questions. Therefore, the statistical methods most appropriate for the analysis of the data collected should be established at the time of designing the experiment and before the study starts." (11, p. 114).

But as described above the type of data generated may require different statistical approaches, which cannot be foreseen when planning the study. That is why a good study plan should consider different options of statistical analysis and define the circumstances when to use which option. This however is rarely done in the day-to-day practice of regulatory toxicology. Furthermore, part of the peer-review process of a study is to revise the statistical analysis. It is a common practice for the experts, during the peer-review to ask for a different statistical test, because of its adequacy for the specific study, even if the study has been completed.

The opinion that "deviations from the statistical analysis used by the study authors should be limited and properly justified" (5, Annex p. 6) should not preclude the BfR or EFSA from a sound assessment of the appropriateness of the statistical method. In case of doubt, a proper post-hoc statistical analysis should be performed. While the OECD Guidance of 2012 discusses both approaches for statistical assessment, i.e. pair-wise comparisons and trend tests, it clearly points to the use of trend tests as the method of choice in its flow chart on page 123. In addition, this guidance states: "In general, testing a trend which is a more specific hypothesis has greater power than a pair-wise comparison" (11, p.118). In contrast, it appears that EFSA insists on the use of the statistical method with less power. In addition, the EFSA states: "It should also be noted that there are no valid studies with statistically significant effects confirmed by both statistical approaches" (5, Annex, p.6). This statement gives the impression that a confirmation by both statistical approaches would be a precondition for considering differences as significant and ignores the clear wording used in the old (12, p. 62) as well as in the new (11, p. 116) OECD Guidance: "Significance in either kind of test is sufficient to reject the hypothesis that chance accounts for the result".

#### Conclusion

The use of pair-wise testing claimed by the BfR and the EFSA as the only appropriate statistical approach to assess tumor incidences of rodent carcinogenicity studies is not supported by OECD guidelines. Likewise, it is a false claim that a post-hoc application of more appropriate statistical analyses by regulatory authorities is forbidden according to OECD guidelines. Instead, the arguments used by BfR and EFSA are either based on invalid contentions (studies were "designed" for pair-wise comparison) or with an out-of-context-reference to OECD Guidance No. 116 and OECD Guideline 451.

#### References

- Clausing, P. (2015): The EFSA Conclusion on the Peer Review of the Glyphosate Risk Assessment - A Reality Check. <u>http://www.pan-</u> germany.org/download/Analysis EFSA-Conclusion 151201.pdf
- (2) Renewal Assessment Report, 31 March 2015 Version, Volume 1; Renewal Assessment Report on Glyphosate, Public version.
- (3) Addendum 1 (2015): Glyphosate Addendum 1 to RAR, 31 August 2015. Renewal Assessment Report on Glyphosate, Public version.
- (4) EFSA (2015a): Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate. EFSA Journal 2015;13 (11):4302
- (5) EFSA (2016): EFSA response to Open letter: Review of the Carcinogenicity of Glyphosate by EFSA and BfR.
  - http://www.efsa.europa.eu/sites/default/files/EFSA\_response\_Prof\_Portier.pdf
- (6) Addendum 1 (2015): Glyphosate Addendum 1 to RAR, 31 August 2015, p. 90.
  Renewal Assessment Report on Glyphosate, Public version.
- (7) EFSA (2015b) Peer Review Report on Glyphosate.
- (8) Renewal Assessment Report, Volume 3, Annex B.6., 28 February 2014, Renewal Assessment Report on Isoproturon
- (9) Renewal Assessment Report, Volume 3, B.6., Renewal Assessment Report on Glyphosate, Public version.
- (10) OECD (2009): OECD Guideline for the Testing of Chemicals No. 451, Carcinogenicity Studies. <u>http://www.oecd-ilibrary.org/environment/test-no-451-carcinogenicity-</u> <u>studies 9789264071186-en;jsessionid=16odl65wcetn2.x-oecd-live-03</u>
- (11) (OECD 2012): Guidance Document 116 on the Conduct and Design of Chronic Toxicity and Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453, 2nd Edition Series on Testing and Assessment No. 116.
- (12) OECD (2002): OECD Environment, Health and Safety Publications. Series on Testing and Assessment No. 35 and Series on Pesticides No. 14. Guidance Notes for Analysis and Evaluation of Chronic Toxicity and Carcinogenicity Studies. <u>http://www.oecd.org/officialdocuments/displaydocument/?cote=env/jm/mono(2002)19</u> <u>&doclanguage=en</u>

# Imprint

# **Pestizid Aktions-Netzwerk (PAN) e.V.** Nernstweg 32

D-22765 Hamburg Phone: +49 (0)40-399 19 10-0 Email: info@pan-germany.org

Hamburg, 25 April 2016

Author: Dr. Peter Clausing with support from Leonie Sontheimer Phone: +49-176 7801 2705, Email: <u>peter.clausing@pan-germany.org</u>,