Pesticides and public health: an analysis of the regulatory approach to assessing the carcinogenicity of glyphosate in the European Union

Peter Clausing, Claire Robinson, Helmut Burtscher-Schaden

ABSTRACT

The present paper scrutinises the European authorities’ assessment of the carcinogenic hazard posed by glyphosate based on Regulation (EC) 1272/2008. We use the authorities’ own criteria as a benchmark to analyse their weight of evidence (WoE) approach. Therefore, our analysis goes beyond the comparison of the assessments made by the European Food Safety Authority and the International Agency for Research on Cancer published by others. We show that not classifying glyphosate as a carcinogen by the European authorities, including the European Chemicals Agency, appears to be not consistent with, and in some instances, a direct violation of the applicable guidance and guideline documents. In particular, we criticise an arbitrary attenuation by the authorities of the power of statistical analyses; their disregard of existing dose–response relationships; their unjustified claim that the doses used in the mouse carcinogenicity studies were too high and their contention that the carcinogenic effects were not reproducible by focusing on quantitative and neglecting qualitative reproducibility. Further aspects incorrectly used were historical control data, multisite responses and progression of lesions to malignancy. Contrary to the authorities’ evaluations, proper application of statistical methods and WoE criteria inevitably leads to the conclusion that glyphosate is ‘probably carcinogenic’ (corresponding to category 1B in the European Union).

INTRODUCTION

On 15 March 2017, the Risk Assessment Committee of the European Chemicals Agency (ECHA) adopted its opinion that the scientific evidence did not meet the criteria specified in the Classification, Labelling and Packaging (CLP) Regulation to classify glyphosate as a carcinogen, thereby confirming the conclusion of the European Food Safety Authority (EFSA) published in November 2015.1–3 This assessment was contrary to that of the International Agency for Research on Cancer (IARC), which categorised glyphosate in its monograph as a probable human carcinogen (Group 2A, according to IARC’s classification).3

This paper scrutinises the European authorities’ assessment of the carcinogenic potential of glyphosate using their own criteria and their weight of evidence (WoE) approach as the benchmark, concentrating on the hazard assessment according to the CLP Regulation.2,5–7 In contrast, a paper that compared the divergent assessments made by EFSA versus IARC concluded that the difference came partially from the use of different data sets and partially from methodological differences in the evaluation.5 However, a number of flaws in EFSA’s analysis were identified.9 Another paper, again comparing the assessments by the two institutions, identified serious flaws in the scientific evaluation of the Renewal Assessment Report (RAR) as the reason for EFSA’s failure to classify glyphosate as a carcinogen.10 11 Both papers focused on a comparison between the ECHA and IARC evaluations. For the present analysis, it is important to note that ECHA recently published an updated Guidance (version 5.0), which, however, was not yet in place when its opinion on glyphosate was developed.12 Therefore here, we make reference to version 4.1 of this Guidance, which was applicable during glyphosate’s evaluation.8 We show that based on the scrutiny of the available documents,1 3 11 13 14 EFSA’s and ECHA’s main reason for not classifying glyphosate as a carcinogen appears to be inconsistent with, and in some instances a direct violation of, the relevant guidance and guideline documents from the Organisation for Economic Cooperation and Development (OECD) and ECHA itself.5–7

LEGISLATIVE BACKGROUND

While the European pesticide regulation 1107/2009 prohibits in principle the marketing of a pesticide classified as a presumed human carcinogen (category 1B),15 its classification is governed by the CLP Regulation 1272/2008.2 According to this regulation, a chemical is classified as a category 1B carcinogenic hazard (presumed human carcinogen) if there is ‘sufficient evidence of carcinogenicity’ in experimental animals. In its Article 3.6.2.2.3.h, ‘sufficient evidence’ is defined as a causal relationship that has been established between the agent and an increased incidence of malignant neoplasms or an appropriate combination of benign and malignant neoplasms in at least two independently conducted valid animal studies. In cases in which some evidence of carcinogenicity exists but is ‘not sufficiently convincing’, the CLP Regulation provides for classification in category 2, as a ‘suspected human carcinogen’.2

As well as determining whether a sufficient number of studies exist with positive findings (‘sufficient evidence’), the ‘strength’ of evidence is evaluated. According to Article 3.6.2.2.3 of the CLP Regulation, strength of evidence involves the enumeration of tumours in human and animal studies and the determination of their level of statistical significance.
In other words, strength of evidence normally requires statistical significance in the observed increase in tumour incidences. If this strong (statistically significant) evidence is seen in at least two independent studies, it may be considered ‘sufficient’ for category 1B. Because of the variability of biological systems, a number of additional factors need to be taken into account. Therefore, according to ECHA, expert judgement is necessary to determine the most appropriate category for carcinogenicity. This expert judgement, combined with the interpretative methods of the WoE approach, is supposed to be guided by several documents. In our opinion, both WoE approach and expert judgement need to be transparent and exercised within the limits set by guidance documents in order to prevent them from shifting away from science-based decisions, possibly to the advantage of certain interest groups.

STATISTICAL ASSESSMENT
Concerning the classification of glyphosate’s carcinogenicity, the controversy begins with the statistical assessment, the determination of the strength of evidence. This relates to two different problems: (1) whether pairwise comparisons or trend tests are more appropriate and (2) whether one-sided or two-sided test statistics should be used.

Pairwise comparisons or trend tests
A pairwise comparison analyses whether the incidence in just one dose group is increased over the control group. In contrast, a trend test asks whether the tumour incidence in all dose groups increases with rising dose as compared with the control. According to OECD, testing a trend generally has a more specific hypothesis and has greater power than pairwise comparison. This obviously applies to study designs with more than one dose group. OECD’s flowchart depicting a trend test for analysing tumour incidences implies that this method is preferred. In our opinion, trend test results should not be played off against those from pairwise comparisons, as was done by EFSA and ECHA. OECD considers significance in either kind of test as sufficient, and ECHA itself acknowledges that any statistically significant increase in tumour incidences is generally taken as positive evidence of carcinogenicity.

One-tailed or two-tailed test
A two-tailed test considers significance for an increase or a decrease in an observation. In contrast, one-tailed tests consider significance only in one direction, thereby doubling the statistical power. For the assessment of a carcinogenic hazard, the protection of public health is the primary concern, so only one direction of change is relevant—an increase in tumour incidence. In our opinion, the logical conclusion is that the one-tailed test is the relevant type of test. Nevertheless, OECD remains vague. It states that a one-sided test may be considered more appropriate for tumour incidences, but cautions that this can be controversial if the treatment could also be protective. Therefore, it considers a two-sided comparison more appropriate in such situations. The reason for OECD’s ambiguity is not clear, but as described above, the nature of the evaluation (hazard assessment) clearly prioritises one-tailed tests. In contrast, EFSA and ECHA exclusively relied on two-tailed tests.

Overall evaluation of the regulatory approach to statistical significance
The European authorities applied a double attenuation of the power of statistical analysis regarding glyphosate’s carcinogenicity in animal studies. They preferred pairwise comparisons and exclusively used two-tailed tests. This weakened the strength of evidence, even before biological relevance was considered.

CARCINOGENICITY OF GLYPHOSATE: THE DATABASE
The authorities recognised significantly increased tumour incidences in 7 long-term studies out of 12 (2 in rats and 5 in mice), with a total of 11 significant increases when the trend test was applied. The data derived from an official document, the CLH Report, are summarised in table 1. Following EFSA’s and ECHA’s approach, the results of two-tailed tests are depicted. Based on this, there was, for instance, a statistically significant increase in 3/5 mouse studies for renal tumours and malignant lymphomas and in 2/5 mouse studies for haemangiosarcoma. Using one-tailed tests, four more tumour incidences would become significant. Not shown in table 1, after a reanalysis of the original study data, eight further significant increases became evident when using one-tailed trend tests. These were not considered by the authorities.

Epidemiological evidence
IARC evaluated all published evidence and identified five case-control studies from Canada, Sweden and USA with a statistically significant association between non-Hodgkin’s lymphoma (NHL) and glyphosate use, while a large cohort study, though only with 6.7 years follow-up time for NHL, was negative. IARC concluded that this represents ‘limited evidence’ for the carcinogenicity of glyphosate. In the RAR, three of these five publications were discussed and initially dismissed as non-reliable, because information about confounding factors was allegedly missing. In an expert opinion, it was demonstrated that these allegations had no basis, and the German Federal Institute for Risk Assessment (BfR) finally agreed that ‘limited evidence’ existed for the association between glyphosate and NHL. Moreover, two meta-analyses confirmed a statistically significant increase in NHL after occupational exposure to glyphosate. In EFSA’s conclusion, this was modified to ‘very limited evidence’ (a term without legal definition) that was considered ‘overall inconclusive’. ECHA shared this view by concluding that the criteria for assigning a CLP category 2 are not fulfilled.

Remarkably, both authorities discussed the epidemiological evidence for NHL in isolation and not as supportive evidence for the observed increased incidence of malignant lymphoma in three different mouse studies.

WEIGHT OF EVIDENCE
While there are different definitions for the WoE approach, its ultimate goal in hazard and risk assessment is to balance statistical significance (in this case the observed increases in tumour incidences) against biological relevance. One problem is that WoE is often used ‘to refer to a body of scientific evidence that has been examined for some purported risk without any interpretative methodology’. It is noteworthy that this is how EFSA and ECHA use the WoE: they refer to WoE elements as defined in the CLP Regulation, but are not transparent as to how the weighting was performed. Furthermore, the way they used important WoE elements is problematic, as described below.

According to OECD Guidance 116, even non-significant increases could in principle be considered relevant based on WoE. However, in the case of glyphosate the authorities used their WoE approach to dismiss the 11 statistically
### Table 1  Basic study information and summary of tumour incidences in males (unless otherwise indicated) in the studies taken into consideration by EU authorities

| Study (year) duration (months) | Species/strain | Tumour type                          | Doses (mg/kg body wt.) | Number of tumours observed | P values (two-tailed) for trend test/ pairwise comparison (significant values in bold) |
|-------------------------------|----------------|--------------------------------------|------------------------|---------------------------|-----------------------------------------------------------------------------------|
| Lankas (1981) (26 months)     | Rat/SD         | Pancreatic carcinoma                  | 0—3—10.3—31.5          | 0—0—0—1                  | 0.0496/1.000 (Fisher’s)                                                            |
|                              |                |                                      |                        | 1—8—5—7                  | 1.000 (Fisher’s)                                                                  |
| Stout & Ruecker (1990) (24 months) | Rat/SD      | Pancreatic islet cell adenoma         | 0—89—362—940           | 2—2—3—7                  | 0.0171/0.162 (Fisher’s)                                                            |
|                              |                | Liver cell adenoma                    | 0—89—362—940           | 2—2—3—7                  | 0.0752/0.392 (Fisher’s)                                                            |
|                              |                | Liver cell adenoma and carcinoma      | 0—89—362—940           | 5—4—4—9                  |                                                                                   |
| Knezevich and Hogan (1983) (24 months) | Mouse/Crl:CD-1, Charles River Wilmington | Lymphoreticular neoplasms† | 0—157—814—4841 | 2—2—6—6                  | 0.0435/0.168*                                                                     |
|                              |                | Renal carcinoma                       | 0—157—814—4841         | 0—0—1—2                  | 0.0370/0.495 (Fisher’s)                                                            |
|                              |                | Renal adenoma and carcinoma           | 0—157—814—4841         | 1—0—1—3                  | 0.0339/0.617 (Fisher’s)                                                            |
| Atkinson et al (1993) (24 months) | Mouse/Crl:CD-1, Charles River Portage | Malignant lymphoma‡          | 0—100—300—1000          | 4—2—1—6                   | No meaningful statistics possible due to incomplete histopathology, but significant in trend test |
|                              |                | Haemangiosarcoma                      | 0—100—300—1000         | 0—0—0—4                  | 0.0004/0.059 (Fisher’s)                                                            |
| Sugimoto (1997) (18 months)   | Mouse/Crl:CD-1 | Malignant lymphoma                    | 0—165—838—4348         | 2—2—0—6                  | 0.0085/0.269 (Fisher’s)                                                            |
|                              |                | Haemangiosarcoma                      | 0—165—838—4348         | 0—0—0—2                  | 0.0078/0.495 (Fisher’s)                                                            |
|                              |                | Renal tubular adenoma                 | 0—165—838—4348         | 0—0—0—2                  | 0.0078/0.495 (Fisher’s)                                                            |
| Kumar (2001) (18 months)      | Mouse/Swiss Albino | Malignant lymphoma                   | 0—15—151—11460         | 10—15—16—19               | 0.0655/0.077§(Fisher’s)                                                            |
|                              |                | Malignant lymphoma in females          | 0—15—151—11460         | 18—20—19—25              | 0.0680/0.225§ (Fisher’s)                                                           |
|                              |                | Renal tubular adenoma                 | 0—15—151—11460         | 1—0—1—2                  | 0.0390/0.495 (Fisher’s)                                                            |
| Wood et al (2009) (18 months) | Mouse/Crl:CD-1 | Malignant lymphoma                    | 0—71—234—810           | 0—1—2—5                  | 0.0037/0.067 (χ²)                                                                |

* A p value of <0.05 is considered statistically significant; p values for pairwise comparisons are for high-dose vs control group. Two-tailed tests were used. Number of animals per group and sex, in most cases 50, ranged from 43 to 60 in the different studies. Only the seven studies with significant findings are depicted in the table.

* Own calculation, not given in [13].

† No specification of malignant lymphoma.

‡ The incidences shown are only from lymph nodes with macroscopic changes.13

§ Statistically significant in pairwise comparison in the Z-test (used in the original study report).
significant increases they had previously recognised. Focusing on mouse studies, the most important WoE elements used by EFSA and ECHA are discussed below. Notably, one of the most important elements—dose-dependence—was completely neglected by both agencies. The ECHA guidance states: ‘Any statistically significant increase in tumour incidence, especially where there is a dose-response relationship, is generally taken as positive evidence of carcinogenic activity’.5 Table 2 shows that such a dose-dependent relationship existed in at least two studies each for renal tumours and malignant lymphomas.

It should be noted that in the Kumar (2001) study only renal adenomas were seen, while in the Knezevich and Hogan (1983) study the dose-dependent increase was for renal carcinoma. One striking difference between the two studies was the study duration (18 vs 24 months). Thus, the possibility exists that carcinoma could have developed in the subsequent 6 months if the Kumar (2001) study had been of 24 months duration.

**Table 2** Tumour incidences in male mice with dose-dependent increases (for further details, see table 1)

| Tumour type      | Study                  | Study duration | Dose group (mg/kg) | Control | Low    | Mid    | High    |
|------------------|------------------------|----------------|-------------------|---------|--------|--------|---------|
| Renal carcinoma  | Knezevich and Hogan (1983) | 24 months     | Control           | 0       | 0      | 1      | 2       |
| Renal tubular adenoma | Kumar (2001)           | 18 months     | Control           | 0       | 0      | 1      | 2       |
| Malignant lymphoma | Kumar (2001)           | 18 months     | Control           | 0       | 10     | 15     | 16      |
| Malignant lymphoma | Wood et al             | 18 months     | Control           | 0       | 1      | 2      | 5       |

**Historical control data**

HCD (tumour incidences of control animals in earlier studies) can help to interpret the data of the study under consideration. According to applicable OECD guidance, they should only be used if the concurrent control data are appreciably ‘out of line’ with recent previous studies.6 The same guidance emphasises ‘that the concurrent control group is always the most important consideration in the testing for increased tumour rates’ and defines strict requirements for the appropriateness of HCD. According to this guidance, HCD should be from the same laboratory and the same strain of animals and should be collected within a maximum of 5 years prior to the actual study. Furthermore, the median and IQR should be used instead of the arithmetic mean and range.6 BfR and EFSA made extensive use of HCD to dismiss the significant tumour findings, but used HCD from time periods up to 17 years beyond the 5-year limit, from seven different laboratories, and sometimes different substrains, and described the HCD by arithmetic mean and simple range instead of median and IQR. ECHA also relied on these inappropriate HCD. ECHA mentioned the existence of guidance-compliant HCD but did not point out that these HCD actually supported the observation of increased tumour incidences for haemangiosarcoma (one study), malignant lymphoma (two studies) and renal tumours (one study).4 No valid HCD were presented by the authorities to support the opposite conclusion.

**Multisite responses**

Multisite responses (tumours in different organs/tissues in the same study) increase the level of concern for a carcinogenic effect.5 These were observed in 3/5 mouse studies. While EFSA did not discuss multisite responses at all,3 ECHA acknowledged two of the three studies with multisite responses,1 but failed to incorporate this element into its WoE assessment.

**Progression of lesions to malignancy**

This WoE element was not discussed by EFSA.3 ECHA acknowledged such a progression for renal tumours, but considered the evidence equivocal.1 According to ECHA, such a progression from renal adenoma to carcinoma, though equivocal, was seen in one study, but not in two other studies, where only adenomas were observed. However, ECHA did not pay attention to the shorter study duration. Therefore, instead of claiming that the evidence is equivocal, ECHA should have acknowledged that the studies are not comparable. The study with renal adenoma and carcinoma was of 24
months duration, while the two adenoma-only studies lasted 18 months (see table 2).

Effects seen in only one sex

According to applicable guidance, effects seen only in one sex may be less convincing. However, the same guidance states that there is no requirement for a mechanistic understanding of tumour induction in one sex only in order to use such findings to support classification in a given hazard category. With one exception—malignant lymphoma in the Kumar (2001) study—significant increases for haemangiosarcoma, malignant lymphoma and renal tumours were seen in male mice only. While EFSA's conclusion that no hazard classification is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/ © Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1 ECHA. Committee for Risk Assessment RAC Opinion proposing harmonised classification and labelling at EU level of glyphosate (ISO): N-(phosphonomethyl) glycine EC Number: 213-997-4 CAS Number: 1071-83-6 2000001412-86-149/F. Adopted 15 March 2017. 2017 https://echa.europa.ec/documents/10162/263a87cfca/1-31d6-8987-9f124f1ab7ae.

2 Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Official Journal of the European Union L 353.

3 European Food Safety Authority. Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate. Efsa J 2015;13:4302.

4 IARC Working Group. Glyphosate. In: Some organophosphate insecticides and herbicides: diazinon, glyphosate, malathion, parathion, and tetrachlorvinphos. IARC Monogr Prog, 2015:1–92.

5 ECHA Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures. Version 4.1. June 2015. 2015. ISBN: 978-92-9427-413-3.

6 OECD. Guidance document 116 on the conduct and design of chronic toxicity and Carcinogenicity studies, supporting test guidelines 451, 452 and 453. 2nd edition. France, Paris: OECD Publishing, 2012.

7 OECD. Test No. 451: Carcinogenicity Studies. Paris, France: OECD Publishing, 2009.

8 Tarazona IV, Court-Marqués D, Tramani M, et al. Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC. Arch Toxicol 2017;91:2723–43.

9 Porter CI, Clasung F, Tarazona Re, et al. Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC. Arch Toxicol 2017;91:3195–7.

10 Porter CI, Armstrong BK, Baguley BC, et al. Differences in the carcinogenic evaluation of glyphosate between the International Agency for Research on Cancer (IARC) and the European Food Safety Authority (EFSA). J Epidemiol Community Health 2016;70:741–5.

11 ECHA. Renewal Assessment Report. Risk assessment provided by the rapporteur Member State Germany and co-rapporteur Member State Slovakia for the active substance glyphosate. European Food Safety Authority. 2015 http://echa.europa.eu/documents/10162/13626/renewal_assessment_report_addenda_en.pdf.

12 ECHA. Guidance to Regulation (EC) No 1272/2008 on classification, labelling and packaging (CLP) of substances and mixtures Version 5.0 July 2017. 2017.

13 BfA Federal Institute for Occupational Safety and Health. Proposal for harmonized classification and labelling. Substance name: N- phosphonomethylglycine; Glyphosate (ISO). In: ECHA, 2016. https://echa.europa.eu/documents/10162/13626/clh_report_glyphosate_en.pdf.

14 ECHA. Renewal Assessment Report. Glyphosate Addendum 1 to RAR. Assessment of IARC Monographs (2015): Glyphosate. European Food Safety Authority, 31 August 2015 http://echa.europa.eu/documents/10162/13626/renewal_assessment_report_addenda_en.pdf.

15 Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. 24 November 2009. Official Journal of the European Union L 309.

16 Weed DL. Weight of Evidence: A Review of Concept and Methods. Risk Analysis 2005;25:1545–57.

17 Greiser E. Gutachten zu epidemiologischen Studien zum möglichen Zusammenhang zwischen der Exposition mit Glyphosat Non-Hodgkin-Lymphomen bzw. Störungen der menschlichen Fortpflanzung im Zusammenhang mit Bewertungen des Bundesinstituts für Risikobewertung der Bundesrepublik Deutschland und der European Food Safety Authority, im Auftrag von GLOBAL 2000. Wien und Bremen, 2016.https://www.global2000.at/sites/global/files/Gutachten%20Prof%20Greiser_Glyphosate-Studien.pdf.

18 Schinasi L, Leon M. Non-Hodgkin Lymphoma and Occupational Exposure to Agricultural Pesticide Chemical Groups and Active Ingredients: A Systematic Review and Meta-Analysis. Int J Environ Res Public Health 2014;11:4849–527.

19 Chang ET, Delzell E. Systematic review and meta-analysis of glyphosate exposure and risk of lymphohematopoietic cancers. Journal of Environmental Science and Health, Part B 2016;51:402–34.

20 OECD. Combined Chronic Toxicity/Carcinogenicity Studies 453. Paris, France: OECD Publishing, 2009.