Accounting for Multiple Comparisons in Statistical Analysis of the Extensive Bioassay Data on Glyphosate

Kenny Crump
[No business affiliation]
2220 South Vienna
Ruston, Louisiana 71270
KennyCrump@email.com

Edmund Crouch
Green Toxicology
106 Sumner Road
Brookline, MA 02445
crouch@greentoxicology.com

Daniel Zelterman
Biostatistics
PO Box 208034, 60 College Street
New Haven, CT, 06520-8034
Daniel.Zelterman@yale.edu

Casey Crump
Departments of Family Medicine and Community Health and of Population Health Science and Policy
Icahn School of Medicine at Mount Sinai
One Gustave L. Levy Place
Box 1077
New York, NY 10029
Casey.Crump@mssm.edu

Joseph Haseman
J.K. Haseman Consulting
1054 Tacketts Pond Drive
Raleigh, NC 27614-7886
hasemanjk@aol.com

Abstract

Glyphosate is a widely used herbicide worldwide. In 2015, the International Agency for Research on Cancer (IARC) reviewed glyphosate cancer bioassays and human studies and declared that the evidence for carcinogenicity of glyphosate is sufficient in experimental animals. We analyzed ten glyphosate rodent bioassays, including those in which IARC found evidence of carcinogenicity, using a multi-response permutation procedure that adjusts for the large number of tumors eligible for statistical testing and provides valid false-positive probabilities. The test statistics for these permutation tests are functions of p-values from a standard test for dose-response trend applied to each specific type of tumor. We evaluated three permutation tests, using as test statistics the smallest p-value from a standard statistical test for dose-response trend and the number of such tests for which the p-value is less than or equal to 0.05 or

0.01. The false-positive probabilities obtained from two implementations of these three permutation tests are: smallest p-value: 0.26, 0.17, p-values ≤ 0.05: 0.08, 0.12, p-values ≤ 0.01: 0.06, 0.08. In addition, we found more evidence for negative dose-response trends than positive. Thus, we found no strong evidence that glyphosate is an animal carcinogen. The main cause for the discrepancy between IARC's finding and ours appears to be that IARC did not account for the large number of tumor responses analyzed and the increased likelihood that several of these would show statistical significance simply by chance. This work provides a more comprehensive analysis of the animal carcinogenicity data for this important herbicide than previously available.

Key Words: glyphosate, carcinogenesis, rodent bioassay, multiple comparisons

Introduction

Glyphosate is a phosphonomethyl amino acid herbicide used extensively throughout the world in weed control. In 2015, after a review of the scientific evidence, the International Agency for Research on Cancer (IARC) concluded that the evidence for the carcinogenicity of glyphosate was limited in humans but sufficient in experimental animals (rats and mice). IARC further concluded that glyphosate was probably carcinogenic in humans. However, other agencies have not concurred with IARC's conclusions (EFSA 2015, WHO 2016, U.S. EPA 2019). Whether or not glyphosate presents a cancer risk remains controversial.

IARC (2015) reviewed ten rodent bioassays and reported on five tumors in three of these bioassays that were interpreted as showing evidence of carcinogenicity: A positive dose-response trend was reported for hepatocellular adenoma in male rats (2/60, 2/60, 3/60, 7/60, p = 0.016) and thyroid C-cell adenoma in female rats (2/60, 2/60, 6/60, 6/60, p = 0.031) (Stout and Ruecker 1990), renal tubule adenoma in male mice (0/49, 0/49, 1/50, 3/50, p = 0.016) (Knezevich and Hogan, 1983), and hemangiosarcoma in male mice (0/50, 0/50, 0/50, 4/50, p < 0.001) (Atkinson 1983b). In addition, there was an increase in pancreatic islet cell adenoma in the low dose group of male rats (1/58, 8/57, 5/60, 7/59, p = 0.016) (Stout and Ruecker 1990), which was also interpreted by IARC as offering evidence of carcinogenicity.

The animal carcinogenicity data on glyphosate are unusually extensive; U.S. EPA (2016) identified 15 long-term rodent oral bioassays of glyphosate, EFSA (2016) identified a further seven, and IARC (2015) another one, together with one skin bioassay. Each bioassay was conducted in both sexes, with each sex potentially having 40-60 unique tumor types, resulting in over 1,000 potential statistical tests, which could easily result in many significant (p≤0.05) tumor increases occurring by chance alone. With such a large number of statistical tests, roughly five percent of them are expected to provide a p-value ≤0.05 simply by chance even if exposure had no effect on carcinogenicity. Thus, in evaluating such a large data base, it is not sufficient to identify sites in individual bioassays in which a statistical test is significant. One must also take into account the large number of statistical tests performed and the attendant possibility that statistically significant findings could be due to chance.

This problem was addressed by conducting a combined analysis of ten glyphosate bioassays which include all three bioassays cited by IARC as showing evidence of carcinogenicity. This article is not a formal systematic review but applies a multi-response permutation approach with the underlying data to quantitatively inform conclusions, consistent with previously recommended guidelines for systematic

reviews (Sena et al 2014 and NAS 2017). The analytic approach provides valid statistical tests of the global hypothesis "glyphosate was carcinogenic in these bioassays." By "valid statistical tests" we mean tests that correct for the multiple comparison problem and consequently have correct false positive rates.

Several methods have been suggested for constructing statistical tests that provide correct false positive rates when evaluating evidence from multiple tumor types possibly in multiple studies. The implementation of each of these methods involves some form of repeated random reassigning of animals to dose groups. Brown and Fears (1981), Heyse and Rom (1987), and Farrar and Crump (1988, 1990) all recommended statistical tests of this type, specifically for analyzing animal carcinogenicity data, that involve repeated permuting of animals among dose groups. Westfall and Young (1989) proposed bootstrap resampling methods for more general types of data. Westfall and Young (1993) concluded that bootstrap and permutation methods yield very similar results. In the present analysis, we employed a slight modification of the permutation approach of Farrar and Crump (1988, 1990) to evaluate the evidence for the carcinogenicity of glyphosate.

The analysis method of Farrar and Crump requires access to individual animal data on histopathological information and tumors, the length of time each animal was on test, and their doses. The ten bioassays used in our analysis represent all the glyphosate animal bioassays for which we had access to this information on individual animals. The ten bioassays include four mouse bioassays (CD-1 strain) and six rat bioassays (three of Sprague Dawley strain and three of Wistar strain) (Table 1).

IARC also considered ten bioassays, two of which were determined to be "inadequate" (George et al 2010 and Séralini et al 2014) and two others for which serious shortcomings were noted (JMPR 2006 had a duration of only one year and Chruscielska et al 2000 had "limited information on dosing regimen, histological examination methods, and tumour incidences"). The remaining six bioassays evaluated by IARC were included in our analysis, including the three bioassays which IARC reported as providing evidence for the carcinogenicity of glyphosate.

Methods

The bioassay data

Individual animal data for the 15 bioassays examined by U.S. EPA through Docket EPA-HQ-OPP-2016-0385 at https://www.regulations.gov (Docket numbers 0018–0047, 0099, 0100, and 0325) were abstracted from the original reports. Of these 15 bioassays, six were eliminated from consideration: two for having incomplete individual data, three for potentially confounding formulation (trimethylsulfonium glyphosate in two, and the sodium salt of n-nitroso glyphosate in the other), and one due to potential confounding by a viral outbreak. The individual animal data for a tenth bioassay (Sugimoto, 1997) were obtained indirectly from EFSA (see Acknowledgements). Of the other eight bioassays identified by U.S. EPA, EFSA, or IARC, five were considered inadequate or unsuitable for evaluation of carcinogenicity by EFSA or IARC, and for none could we locate individual animal data. Thus, our analysis included data from ten glyphosate bioassays (Table 1). The abstracted data from these ten bioassays, along with information on all 23 glyphosate bioassays of which we are aware, are publicly available on Dryad (Crouch et al 2019).

The analysis included four bioassays that were not included in the IARC 2015 review (Suresh 1996, Wood et al 2009a, 2009b and Sugimoto 1997). IARC had access to detailed summary information on the tumor incidences and doses in each of these four bioassays, but stated that they were unable to evaluate them because of the limited experimental data provided in the review article and supplemental information (Greim et al 2015, supplemental online information).

Greim et al (2015) assigned a Klimisch score to each of the ten bioassays included in the analysis that indicates the reliability of a study (Klimisch et al 1997). Eight of the ten studies were assigned a Klimisch score of 1 (indicating that a study is "fully reliable based on compliance with Good Laboratory Practice and adherence to appropriate study guidelines"). Knezevich and Hogan (1983) was assigned a Klimisch score of 2 (signifying that "some guideline requirements are not met, but these deficiencies do not negatively affect the validity of the study for its regulatory purpose"), which apparently was primarily due to the fact that the study was conducted prior to the institution of Good Laboratory Practice, rather than because of any deficiency. Lankas (1981) was assigned a Klimisch score of 3, signifying "a test design that is not fit for the scientific purpose of the study, due to significant scientific flaws, or the objective of the study not covering the regulatory endpoints, or both. Such studies can provide supplemental information but do not allow a stand-alone appraisal of a regulatory endpoint." The apparent reason for the low Klimisch score was low power due to low doses used in Lankas (1981) (see Table 2), rather than any other deficiency. One member of EPA's FIFRA Scientific Advisory Panel (USEPA, 2017) argued that significant carcinogenic effects were seen in this study.

Statistical tests

The statistical tests applied in the analysis were functions of p-values obtained from conventional continuity-corrected poly-3 tests for trend applied to each type of tumor or combination of tumor types in each bioassay. The continuity corrected poly-3 test (Bailer and Portier 1988, Moon et al 2006, Peddada and Kissling 2005) is a survival-adjusted Cochran-Armitage test. This test will have power to detect monotone as well as most non-monotone dose effects, although we know of no evidence or theory that glyphosate causes non-monotone dose-responses. If a non-monotone response is caused by differential mortality, this would be dealt with appropriately by the poly-3 test which adjusts for survival. The poly-3 test was not available when Farrar and Crump (1988, 1990) did their work, but it since has become widely used and is now used by the U.S. National Toxicology Program (NTP) to analyze bioassay data (NTP 2005). In the present analysis, the continuity-corrected version of the poly-3 test used (Peddada and Kissling, 2005) was copied from a key portion of the computer program used by the NTP (provided by Dr. Grace Kissling, NTP), and direct comparisons have shown that our implementation gives the same results as the version used by the NTP. Throughout this paper, all implementations of the poly-3 test, are one-sided (i.e., one-tailed), as are the NTP implementations of the test.

Results from three multi-response permutation tests are presented. In the simplest such test, referred to as the "min test", the test statistic is the smallest p-value obtained from applying the poly-3 test to all tumor types in all of the ten bioassays. In the simplest implementation of this test (a slightly more complex implementation is required in the present situation, due to the rules for conducting pathology examinations used in some of the bioassays, see "Details of the testing procedure" below), animals are randomly reassigned to dose groups (i.e., permuted among dose groups) in a Monte Carlo analysis, keeping the total numbers of animals in each dose group equal to the number in the original data. The tumors in

each such reassignment are analyzed using the poly-3 test in exactly the same way as in the original data. Males and females are permuted separately. The false positive rate is the proportion of random reassignments that result in a smallest poly-3 p-value that is smaller than or equal to the smallest poly-3 p-value obtained from the original data.

For example, suppose the smallest p-value obtained from poly-3 analyses of all the tumor types that occur in the ten bioassays was 0.015 from a statistical analysis of, say, thyroid C-cell adenoma or carcinoma. The animals are redistributed at random among the dose groups, keeping the total number of animals in each dose group equal to the original number. This random redistribution is repeated 5,000 times and each time the permuted data are analyzed in exactly the same way as the original data and the smallest p-value obtained from any tumor is noted. Suppose that in the 5,000 redistributions the smallest p-value obtained is less than or equal to 0.015 in 400 of the 5,000 redistributions. If we had decided a priori that we were interested only in whether the incidence of thyroid C-cell adenoma or carcinoma was increased by exposure, and in no other lesion, then 0.015 would be the correct p-value to consider. But, if, as is more likely, a post-hoc decision was made to focus on thyroid C-cell adenoma or carcinoma because statistical analysis of this lesion gave the smallest p-value out of many lesions statistically analyzed, the appropriate false positive rate would be estimated as 400/5,000 = 0.08. Thus, the true significance of the original p-value of 0.015 is estimated as 0.08 after accounting for the multiplicity of statistical tests applied.

In addition to the min test, two additional such tests were computed. The test statistics for these tests were the number of poly-3 tests of tumors in the original data for which the p-value is less than or equal to the critical value of 0.05 (the "05 test") or 0.01 (the "01 test"). The false positive rates for these tests are the proportion of random permutations of the data for which the number of poly-3 p-values from the permuted data that are less than or equal to the critical value equal or exceed the number from the original data. For example, in conventional poly-3 analyses of all tumor sites in the 10 datasets, if 15 tumor sites provided a p-value <= 0.05, and in 5,000 random permutations of these data across dose groups, 100 of the redistributions resulted in 15 or more sites with p-values <= 0.05, the p-value of the 05 test would be estimated as 100/5,000 = 0.02.

A large number of such tests can be envisioned, each having power for detecting certain departures from the null hypothesis. The min test could have enhanced power in a situation in which a test agent causes cancer at a single site, whereas the 05 test could have enhanced power when a test agent causes detectable cancer of several types.

Each of these tests is a member of the same family of tests as Fisher's exact test, which is often used in testing tumor data for a dose effect. Fisher's exact test, when applied to a particular tumor in a cancer bioassay, is conditional on the total number of animals with this tumor. Under the null hypothesis of no carcinogenic effect, the distribution into dose groups of animals with the tumor is assumed to be random. The permutation tests described above are also conditional, not just on the total numbers of tumors, but on the patterns of tumors occurring in individual animals. No assumption (such as independence) is required regarding the joint distributions of tumors within animals. Repeated permutation of animals is used to compute false positive rates because, unlike the situation with Fisher's exact test, a direct calculation is too difficult. However, false positive rates for Fisher's exact test could also be computed using permutation.

Details of the testing procedures

To test for a dose-response trend using the poly-3 test, estimates of the glyphosate dose in each dose group are needed. Our analysis examines doses in units of mg/kg/day. These come from the individual bioassays with one exception: Knezevich and Hogan (1983) do not provide doses in mg/kg/day by dose group. We carried out our own dose calculations using the information in the Knezevich and Hogan report and obtained doses that varied only trivially from those reported by Greim et al (2015) for this study, so we used the Greim et al doses in our analysis. The doses in each of the ten bioassays used for our analyses are listed in Table 2 and are also publicly available on Dryad (Crouch et al 2019).

In addition to conducting conventional poly-3 tests on specific types of tumors, tests were also conducted on combinations of tumor types thought to have a common origin (e.g., liver adenomas and carcinomas). Tumors to combine were selected by J. H. in a manner patterned after the combinations used by the NTP (2019). Complete listings of the combinations in each of the ten studies used in the analysis are available on Dryad (Crouch et al 2019). Since including these combinations resulted in the same tumors being present in multiple analyses, it was decided to perform two analyses, one (the "primary analysis") that included all of the individual tumors and combinations, and one (the "reduced analysis") in which individual tumors and combinations of tumors were removed from the analysis if they were part of a more inclusive tumor combination. E.g., if a combination consisted of liver adenoma or carcinoma, the tumor categories of liver adenoma and liver carcinoma were removed and only the combination was used in the analysis.

Each of the original reports of the ten bioassays contains a list of the tissues scheduled to be routinely given a histopathological examination. In addition, sometimes other tissues that looked suspicious at necropsy were also examined histologically ad hoc. For five of the glyphosate bioassays (Table 1), if a tissue was listed for histological examination, that tissue was scheduled for a histological examination in all animals in all the dose groups (complete histology [CH] bioassays). In the remaining five bioassays (incomplete histology [ICH] bioassays), control and high dose animals were all given a complete histopathological examination, along with non-surviving animals (animals that died before the final sacrifice) in the intermediate dose groups. In addition, certain tissues ("mandatory tissues") in all animals were scheduled for a histopathological examination (most including kidney, liver and lung; see the lists in Table 1), regardless of when they died.

Systemic tumors, which can appear in multiple tissues, require a special treatment. It was decided that if any tissue in an animal was examined histologically, that animal would be counted as being examined for systemic tumors; hence "system" was considered a mandatory tissue. We believe that this approach will provide the least likelihood of an error, particularly since systemic tumors were often verified in ad hoc examinations. This approach also appears to be consistent with standard practice. The few non-systemic tumors discovered in ad hoc analyses were not included in our analysis.

In the simulations, the object was to randomly permute animals among dose groups. Simple randomization suffers from a potential bias due to dose-related differential survival, and, for the ICH studies, a problem of data comparability — e.g., it would result in high dose and control animals that survived to final sacrifice and had certain tumor sites examined being placed in intermediate dose groups where surviving animals did not have these sites examined and vice-versa. The ICH studies thus required special treatment in the permutation analysis, as explained in the following paragraph and summarized in Figure 2.

In each of the ten bioassays, dose-related effects on survival were tested using a Cochran-Armitage test for negative trend on the proportions of animals surviving to final sacrifice in the various dose groups (Table 2). However, regardless of the outcome of this test, to control for potential dose-related differences in survival

in both the ICH and CH studies, each randomization of the data maintained the same number of survivors and non-survivors in each group as was seen in the actual data. Moreover, in the intermediate dosed groups survivors in the ICH studies, histopathology was carried out only for mandatory tissues. Thus, for the other tissues in the ICH studies, in the intermediate dosed groups only the non-survivors could be used in the trend test analyses. For mandatory tissues (all tissues for CH studies), the survivors and non-survivors were separately permuted, keeping the number of each within each dosed group the same. A similar randomization scheme was carried out for the other tissues in the ICH studies, but these randomizations for survivors included only control and high dose animals. Thus, in ICH studies mandatory tissues and other tissues had to be separately randomized (effectively separating each animal into two sets of tissues) in order to include in the analysis all the pathological information routinely collected in the ICH studies. Figure 2 provides an outline of this permutational analysis.

It is important to note that in all applications of the poly-3 test, the test is applied only to data from one sex in a single study. Thus, data from different sexes or studies are not combined, but rather p-values from poly-3 tests are combined to create the "global" tests (min, 05 test and 01 test) having the correct false positive rates. In addition to the randomization procedures described above for testing for positive dose-response trends in tumor occurrence, the same procedures were repeated after reconfiguring the poly-3 test to assess for negative trends.

Since different permutations are independent, the false positive rates reported from the permutation tests times the number of permutations are binomially (N, p) distributed with N equal to the number of permutations (N = 5,000 in all results reported herein) and p equal to the true false positive rate. This information can be used to calculate exact confidence limits for the true false positive rates. We have chosen not to report these for two reasons: It would make the tables less readable, and we believe the accuracy resulting from 5,000 simulations is sufficient to guide conclusions. However, if a reader is interested in computing confidence intervals for any reported false positive rates, the necessary information is contained in Tables 4 or 5.

Results

Figure 1 shows the frequency of poly-3 p-values for positive trend computed from all tumors in all ten bioassays in which at least two tumors occurred (and in which therefore a p-value ≤ 0.05 was theoretically possible). This figure suggests an excess of large p-values (those close to 1.0) compared to small p-values (those close to 0.0). Since the version of the poly-3 trend test applied is a one-sided test for a positive trend, p-values close to 1.0 would translate into p-values near 0.0 for one-sided trend tests for anti-carcinogenicity. Thus, the overall pattern in Figure 1 is more consistent with an anti-carcinogenic than a carcinogenic effect. However, this is not necessarily evidence that glyphosate is anti-carcinogenic and in the discussion section we mention other plausible reasons for this response.

Results of tests for a dose-related decrease in survival in each study are shown in Table 2. In none of the bioassays was this test statistically significant. Moreover, four of the datasets had p-values in excess of 0.95 which indicate a significant positive trend in survival with increasing dose. Overall, animals exposed to the highest doses of glyphosate tended to have enhanced survival compared to controls (Table 2).

Table 3 lists, for the primary analysis, the 24 tumors in the ten bioassays for which the poly-3 test for a positive dose-related trend was significant at the 0.05 level. This list includes four of the five tumors cited by IARC as providing evidence of carcinogenicity. The missing tumor is pancreatic islet-cell adenoma in male

rats (Stout and Ruecker 1990), which had responses of 1/58, 8/57, 5/60, and 7/59 and did not have a significant dose-related trend. In an identical analysis except that the poly-3 test was configured to test for a negative dose-related trend, there were 26 tumors for which the dose-response trend was significantly negative at the 0.05 level.

Table 4 shows the results for the three permutation tests for positive trend, both for the primary analysis and the reduced analysis. The most significant poly-3 trend in all ten bioassays was 0.0013, which was for hemangiosarcoma in male mice (Atkinson et al 1993b). As noted in Table 3, this response was also the most significant of those reported by IARC (2015). However, the actual significance of this smallest p-value, which is the false positive rate for the min test, was 0.26 based on the primary analysis, rather than the naive value of 0.0013, which means that 26% of the randomizations of the ten datasets gave a smallest p-value less than or equal to the smallest (0.0013) from the original data. Similarly, the 05-test used as test statistic the number of poly-3 p-values ≤ 0.05 from analysis of all the tumors in all ten bioassays. That number was 24, based on the primary analysis. The corresponding false positive rate was 0.08, which means that 8% of randomizations of the ten datasets found at least 24 sites for which the poly-3 p-value was ≤ 0.05. The results from all permutation tests based on the reduced data were similar to those based on the primary data. The false positive rate for the 01 test was 0.06 in the primary analysis (which may be considered borderline significant) and 0.08 in the reduced analysis. Overall, the findings from Table 4 suggest that, after accounting for the number of statistical tests performed, there was no clear evidence of a positive dose-related trend in tumor occurrence.

Table 5 presents the same information as Table 4 except that the poly-3 test was reconfigured to test for negative dose-response trends. Comparing Tables 4 and 5, the evidence for negative trends is greater than that for positive trends in all analyses. The smallest poly-3 p-value for a negative trend is 0.0008 (which was for bronchiolar-alveolar adenoma in female mice in Knezevich and Hogan (1983)), whereas the smallest p-value for a positive trend was 0.0013 (Tables 3 and 4). The 01 test for a negative trend was highly significant in both the primary and reduced analyses (p = 0.002 for each). These findings thus suggest stronger evidence for negative rather than positive dose-response trends in tumor occurrence.

Discussion

The highest doses given to any animal groups in the ten bioassays were 5,873 mg/kg/day and 4,841 mg/kg/day in high-dose female mice and male mice, respectively, in Knezevich and Hogan (1983). U.S. EPA guidance states there is no need to expose animals to daily doses in excess of 1,000 mg/kg/day (U.S. EPA 1998). Despite the extremely high doses, there was no evidence of reduced survival in this study (Table 2). In fact, there was statistically significantly enhanced survival in male mice, as well as in male animals in several other bioassays (Table 2). In no study was there a statistically significant decreasing trend in longevity in either sex. Thus, glyphosate was relatively non-toxic in these bioassays based on survival.

Having access to the individual animal data from all ten bioassays was critical to the analyses conducted. This allowed us to treat an animal as the basic unit of measurement. E.g., use of the individual animal data allowed us to distinguish between an adenoma and a carcinoma occurring in separate animals (which our analysis counts as two animals with tumors), and both tumors occurring in a single animal (which our

analysis counts as a single tumor-bearing animal). It would not be possible to distinguish these occurrences from data on individual tumors summarized by dose groups.

In order to adjust analyses for the specific times different animals were on study, the poly-3 test requires knowledge of the age at death of each animal. Thus, we could not have adjusted our analyses for duration of exposure adequately without individual animal data. Our analysis found a statistically significant trend toward more animals surviving to final sacrifice at higher doses in several of the bioassays (Table 2), suggesting that it may be important to control for age in analyzing these bioassays. In addition to employing the poly-3 test, which is an age-adjusted Cochran-Armitage test, age was also controlled in our analyses by keeping the numbers of animals surviving to final sacrifice in each dose group the same in all permutations as in the original data.

In all ten bioassays combined, our primary analysis conducted 525 poly-3 analyses (Table 4) of individual tumor responses, of which a total of 174 were on combinations of individual tumor types that may have similar etiologies. In the primary analysis, individual tumors can appear in more than one poly-3 analysis. Since this will happen in the original data and the permuted data with equal frequency, it will not bias the analysis. Nevertheless, we also conducted a reduced analysis in which individual tumors and combinations of tumors were removed from the analysis if they were part of a more inclusive tumor combination. This reduced analysis involved 304 poly-3 analyses (Table 4). Results from these two analyses were quite similar (Tables 4 and 5).

Three permutation statistical tests that provide proper control for false positives were applied in both the primary and reduced analysis: The min test – with the most significant poly-3 test result from all tumors in all bioassays as the test statistic; and the 05 test and 01 test having as test statistic the number of poly-3 tests that resulted in a p-value ≤ 0.05 and 0.01, respectively. These two limits, 0.05 and 0.01, were selected because of their traditional importance in evaluating the result of statistical tests.

The smallest poly-3 p-value found in the analysis of the ten datasets was 0.0013 for hemangiosarcoma in male mice in Atkinson et al (1993b). This tumor was also the most significant in IARC's evaluation. However, our analysis showed that the actual false positive rate for this finding after accounting for multiple comparisons was 0.26 in the primary analysis and 0.17 in the reduced analysis (min test, Table 4), demonstrating the importance of accounting for multiple comparisons in the glyphosate data. Similarly, neither the 05 test (p = 0.08, primary analysis and p = 0.12, reduced analysis) nor the 01 test (p = 0.06, primary analysis and p = 0.08, reduced analysis) gave a false positive rate that was clearly less than 0.05 (Table 4), although the false positive rate for the 01 test in the primary analysis was near the boundary of 0.05.

In the primary analysis, three of the seven tumors or tumor combinations that provided poly-3 p-values ≤ 0.01 involved hemangioma and/or hemangiosarcoma in mice (Table 3). However, the two (of the three) in female mice (Table 3) included the same tumors. The increased incidence in males was due to hemangiosarcoma; the increase in females was due primarily to hemangioma.

To further evaluate these findings, we compared them to the incidence of hemangioma and hemangiosarcoma found in the other mouse bioassays. The significant response in CD-1 males in Atkinson et al resulted from four hemangiosarcomas at a dose of 1,000 mg/kg/day (p=0.0013), with no hemangiomas or hemangiosarcomas identified at three lower doses. However, Knezevich and Hogan (1983) exposed CD-1 male mice to a dose nearly five times that of Atkinson et al (4,831 mg/kg/day) and no

hemangiomas or hemangiosarcomas were found, although Knezevich and Hogan did find these tumors at lower doses. Hemangiosarcomas are not rare tumors in CD-1 mice. Giknis and Clifford (2005) report a range of 0-12% in control male CD-1 mice. Thus, the 8% (4/50) incidence of hemangiosarcoma seen in high dose male CD-1 mice in the Atkinson et al study (the highest response seen in all of the glyphosate studies) is within the range seen in male control CD-1 mice. Furthermore, Atkinson et al. provide summary background incidences in six other studies performed under similar conditions as 2/50, 2/50, 4/50, 0/50, 1/50, and 1/50 and consequently considered this finding as not due to administration of glyphosate.

In female CD-1 mice, the maximum hemangioma/hemangiosarcoma response of 5/50 in the Sugimoto (1997) study occurred at a dose of 4,116 mg/kg/day, whereas Knezevich and Hogan (1983) found a response of only 2/50 at a dose of 5,873 mg/kg/day which was 43% higher than the highest dose in Sugimoto. Mice in Knezevich and Hogan also were exposed 31% longer than those in Sugimoto before final sacrifice. The 10% incidence reported in the Sugimoto study is within the control range of 0-12% incidence of hemangiosarcoma reported by Giknis and Clifford (2005) for female control CD-1 mice.

The lack of a consistent dose response in either males or females suggests that finding significant responses in hemangioma and hemangiosarcoma in both sexes of mice may be attributable to chance, especially considering that this represents the “worst case” of more than 100 tumor sites/types in these bioassays that could have shown evidence of carcinogenicity.

The only other tumor in the mouse studies that the IARC regarded as being clearly related to glyphosate exposure was the marginally significant increase (0/49, 0/49, 1/50, 3/50) in kidney adenoma in male mice observed in the Knezevich and Hogan study (see Table 3). However, the data in Table 3 do not reflect the fact that additional step sectioning of kidneys in the dosed and control groups revealed one kidney adenoma in the control group, but no additional kidney tumors in the dosed groups. The data provided to us did not identify this animal, so we could not factor this additional tumor into our analysis. However, inclusion of this tumor-bearing control animal has been reported to eliminate the significant ($p<0.05$) trend for this tumor (Tarone 2018a), adding to the evidence that the tumor increases reported in the glyphosate studies are due to chance.

In addition to testing for positive dose-response trends, both our primary and reduced analyses were repeated using the poly-3 test configured to detect negative dose-related trends in tumor occurrence (Table 5). Comparing the results of these analyses with those testing for a positive trend (Table 4), the evidence for an effect was stronger for negative than for positive trends. This finding agrees with the impression obtained from the histogram of p-values in Figure 1. The smallest p-value for a positive trend was 0.0013 versus 0.0008 for a negative trend, although the corresponding false positive rates after correcting for multiple comparisons were 0.26 and 0.11 (Tables 4 and 5, min test, primary analysis), demonstrating how adjusting for multiple comparisons can change the interpretation of analyses of individual tumors. The only clearly significant results for any of the three permutation tests were highly significant 01 tests for negative trend in both the primary analysis and the reduced analysis ($p = 0.002$ in both cases, Table 5). We caution against assuming this finding is evidence of an anti-carcinogenic effect of glyphosate exposure, as there are other possible explanations. It is known that reduced body weight in rodents can result in fewer tumors (Rao et al 1987, Haseman et al 1997), and perhaps the massive doses of glyphosate fed to the animals made their food less palatable and caused a reduction in body weights at higher doses. An investigation of this possibility is beyond the scope of the present work.

IARC (2015) evaluated ten bioassays, two of which they considered "inadequate" and two others which they also noted had serious shortcomings (one had a duration of only one year and the other IARC claimed had "limited information on dosing regimen, histological examination methods, and tumour incidences"). None of these four studies were included in our analysis. The remaining six bioassays are all included in our analysis. These include two mouse bioassays and four rat bioassays. In their review of these six bioassays, IARC reported four tumors in three bioassays for which the dose-response trend was significant with $p \leq 0.05$ and which were cited as providing evidence of the carcinogenicity of glyphosate. In addition, there was a $p \leq 0.05$ excess of pancreatic islet adenoma over background in male rats in one of these bioassays, which was also cited, although there was no significantly positive dose response trend. These five tumors and the three bioassays are listed in the introduction to this paper. In a fourth bioassay (Lankas 1981), IARC identified a barely significant $p \leq 0.05$ excess over background of pancreatic islet cell tumors at the lowest dose in male rats. IARC did not claim this finding as providing evidence of carcinogenicity, stating that there was no statistically significant positive dose-response trend and there was no apparent progression to carcinoma. With this one exception, every analysis noted as coming from an adequate bioassay that gave a p-value less than 0.05 was cited as evidence for the carcinogenicity of glyphosate. Thus, IARC's method of evaluating the evidence for carcinogenicity of glyphosate seemingly consisted primarily of identifying statistical analyses of individual tumors that exhibited a p-value less than 0.05. No discussion was provided of the extent of data from which these statistical analyses arose, nor of the possibility that these five significant findings could have arisen by chance from the statistical analysis of many tumors in multiple bioassays.

Our analysis of the six bioassays that were also reviewed by IARC identified eight tumors for which the poly-3 test found a $p \leq 0.05$ positive trend. This included the four tumors reported by IARC as showing a $p \leq 0.05$ positive trend, and four additional tumors that were not listed by IARC (Table 3). In all ten bioassays, our analysis identified 24 tumors that exhibited a poly-3 positive trend p-value ≤ 0.05. Nevertheless, after accounting for the multitude of statistical tests our analysis did not find that number statistically significant ($p = 0.08$, Table 4).

Summarizing, our statistical analysis of ten glyphosate bioassays, which included all of the bioassays IARC reported as providing evidence of carcinogenicity, found no strong statistical evidence that glyphosate is carcinogenic, whereas IARC found the evidence for glyphosate carcinogenicity in these bioassays "sufficient." The main cause for this discrepancy appears to be that IARC failed to consider the large number of statistical tests performed in the multiple bioassays they reviewed and the resulting multiple comparison problem. IARC and other organizations involved with interpreting results from large data sets to which a large number of statistical tests have been applied should consider applying analyses of the type used in this paper to make informed and reasonable decisions.

The IARC declared that glyphosate is probably carcinogenic to humans, noting a positive association for non-Hodgkin lymphoma (NHL) ((IARC 2015). The principal human data on glyphosate and NHL come from five case-control studies and two cohort studies. Crump (2020) examined these studies and concluded that the case-control studies are at risk of recall bias resulting from information on exposure to pesticides being collected from cases and controls based on their memories. Two of the case-control studies are additionally at risk of a form of selection bias that can exacerbate the effect of recall bias. Both biases are in the direction of making glyphosate appear carcinogenic. He concluded that the evidence in these studies for the carcinogenicity of glyphosate comply closely with what would be expected if this evidence results from statistical bias in the case-control studies (Crump, 2020).

The IARC's conclusion that glyphosate was probably carcinogenic to humans was influenced by what IARC considered to be "sufficient" evidence of carcinogenicity in animals. However, several reviews by regulatory bodies in the U.S. and Europe disputed that conclusion (EFSA 2015, WHO 2016, USEPA 2019). In EPA's FIFRA Scientific Advisory Panel on glyphosate (USEPA, 2017) report, some panelists noted that the number of significantly positive results in this large database was no greater than would be expected from random assignment of animals to dose groups. These panelists also noted the serious multiple comparison problem resulting from conducting so many statistical analyses. Similarly, Williams et al (2016) reviewed the glyphosate bioassay data and noted that statistical analysis of sites from the large number of bioassays would be expected to generate false positive results. Tarone (2018a, b) likewise noted that IARC finding evidence that glyphosate was carcinogenic based on marginal significance of the most extreme finding from dozens of statistical tests was scientifically unsound. Thus, a number of sources have contradicted IARC's conclusion and several have drawn attention to the multiple comparison problem inherent in the statistical analysis of the many bioassays of glyphosate. The present analysis provides new information on the potential carcinogenicity of glyphosate by being the first to provide results from statistical tests with correct false positive rates. These tests found no strong or convincing evidence that glyphosate is an animal carcinogen.

Conflicts of Interest

Dr. Kenny Crump and Dr. Zelterman served on the EPA Federal Insecticide, Fungicide, and Rodenticide Act Science Advisory Panel (SAP), which met to review an EPA document on glyphosate on December 13–16, 2016. Dr. Haseman testified before this panel on behalf of Monsanto and had a one-year consulting agreement with Monsanto that extended from November 14, 2016 to November 13, 2017. However, this agreement was limited explicitly to his work related to the SAP. He has had no contact with, nor received any compensation, advice, or data from Monsanto related to this paper. The remaining authors have no conflicts of interest to report. This paper was self-funded.

Acknowledgements

We thank Martin Pigeon of Corporate Europe Observatory for providing the individual animal data released by EFSA for the Sugimoto (1997) study. We thank Dr. Grace Kissling for providing the NTP poly-3 computer program.

References

Atkinson C, Strutt AV, Henderson W, Finch J, Hudson P. (1993a) 104-Week Chronic Feeding/Oncogenicity study in rats with 52-week interim kill. Unpublished, MRID[1] 49631701.

Atkinson C, Martin T, Hudson P, Robb D. (1993b) Glyphosate: 104 week dietary carcinogenicity study in mice. Unpublished. Inveresk Research International, Tranent, EH33 2NE, Scotland. IRI Project No. 438618. MRID 49631702.

[1] An MRID is a unique eight-digit number assigned to each study submitted to the U. S. EPA.

Bailer AJ, Portier CJ. (1988) Effects of Treatment-induced Mortality and Tumor-induced Mortality on Tests for Carcinogenicity in Small Samples. Biometrics 44, 417–431.

Brammer A. (2001) Glyphosate Acid: Two Year Dietary Toxicity and Oncogenicity Study in Wistar Rats. Unpublished. Central Toxicology Laboratory, Alderley Park Macclesfield, Cheshire, UK: Syngenta. MRID 49704601.

Brown C, Fears T (1981) Exact significant levels for multiple binomial testing with applications to carcinogenicity screens. Biometrics 37: 763-774.

Chruscielska K, Brzezinski J, Kita K, Kalhorn D, Kita I, Graffstein B, Korizeniowski P. (2000) Glyphosate - evaluation of chronic activity and possible far-reaching effects. Part 1. Studies on chronic toxicity. Pestycydy (Warsaw), 3–4:11–20, ISSN 0208-8703.

Crouch E, Haseman J, Crump K. (2019) Individual animal data from rodent carcinogenicity bioassays for glyphosate, Dryad, Dataset, https://doi.org/10.5061/dryad.wwpzgmsfv

Crump, K. (2019) The potential effects of recall bias and selection bias on the epidemiological evidence for the carcinogenicity of glyphosate. Risk Analysis DOI: 10.1111/risa.13440.

EFSA (European Food Safety Authority). (2015) Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate (EFSA-Q-2014-00546, EFSA-Q2015-00279, approved on 30 October 2015 by European Food Safety Authority). EFSA J. 13:4302. [107 p.]. doi:10.2903/j.efsa. 2015.4302.

EFSA (European Food Safety Authority). (2016) CLH report. Proposal for Harmonised Classification and Labelling. Substance Name: N-(phosphonomethyl)glycine; Glyphosate (ISO). May 2016. Available at http://dar.efsa.europa.eu/dar-web/provision.

Farrar D, Crump KS. (1988) Exact statistical tests for any carcinogenic effect in animal bioassays. Fundamental and Applied Toxicology 11:652-663.

Farrar D, Crump KS. (1990) Exact statistical tests for any carcinogenic effect in animal bioassays. II. Age-adjusted tests. Fundamental and Applied Toxicology 15:710-721.

George J, Prasad S, Mahmood Z, Shukla Y. (2010) Studies on glyphosate-induced carcinogenicity in mouse skin: a proteomic approach. J Proteomics, 73(5):951–64. doi:10.1016/j.jprot.2009.12.008 PMID:20045496

Giknis MLA, Clifford CB (2005) Spontaneous Neoplastic Lesions in the CrI:CD-1(ICR) Mouse in Control Groups from 18 Month to 2 year Studies, Charles River Laboratories http://www.centerforfoodsafety.org/files/charles-river-2005_39907.pdf

Greim H, Saltmiras D, Mostert V, Strupp C. (2015) Evaluation of carcinogenic potential of the herbicide glyphosate, drawing on tumor incidence data from fourteen chronic/carcinogenicity rodent studies, Crit. Rev. Toxicol. 45(3): 185-208.

Haseman JK, Young E, Eustis SL, Hailey JR. (1997) Body weight/tumor incidence correlations in long term rodent carcinogenicity studies. Toxicologic Pathology 25: 256-263.

Heyse J, Rom D. (1987) Adjusting for multiplicity of statistical tests in the analysis of carcinogenicity studies using multiresponse randomization tests. Presented at the Annual Meeting of the American Statistical Association.

IARC (International Agency for Research on Cancer). (2015) Some Organophosphate Insecticides and Herbicides: Diazinon, Glyphosate, Malathion, Parathion, and Tetrachlorvinphos, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Volume 112. Lyon.

JMPR. (2006) Glyphosate. In: Pesticide residues in food—2004 Toxicological evaluations. Sponsored jointly by FAO and WHO. With the support of the International Programme on Chemical Safety (IPCS). Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group. Rome, Italy, 20–29 September 2004.

Klimisch HJ, Andreae M, Tillmann U. (1997) A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regulatory Toxicology and Pharmacology 25: 1 – 5.

Knezevich AL, Hogan GK. (1983) A chronic feeding study of glyphosate in mice. Unpublished report prepared by Bio/Dynamic Inc., dated July 21, 1983. Report No. 77-2011. MRID 00130406.

Lankas, GP. (1981) A Lifetime Study of Glyphosate in Rats. Unpublished Report. No. 77-2062 prepared by Bio Dynamics, Inc. EPA Accession. No. 247617 – 247621. December 23, 1981. MRID 00093879.

Moon H, Ahn H, Kodell RL. (2006) A Computational Tool for Testing Dose-related Trend Using an Age-adjusted Bootstrap-based Poly-k Test. Journal of Statistical Software 16,7.

NTP (National Toxicology Program) (2005) NTP Technical Report On The Toxicology And Carcinogenesis Studies Of Citral (Microencapsulated) (Cas No. 5392-40-5) In F344/N Rats And B6c3f1 Mice (Feed Studies). National Toxicology Program, P.O. Box 12233, Research Triangle Park, NC 27709.

NTP (National Toxicology Program) (2019) NTP Historical Controls Report, All Routes and Vehicles, Harlan Sprague-Dawley RATS. Available at **https://ntp.niehs.nih.gov/ntp/historical_controls/ntp2000_2019/r_hcrpt_allrte20190400.pdf**. Last accessed on 2019/01/10.

Peddada SD, Kissling GE. (2005) A survival-adjusted quantal-response test for analysis of tumor incidence rates in animal carcinogenicity studies. Environmental Health Perspectives 114(4): 537–541.

Rao GN, Piegorsch WW, Haseman JK. (1987) Influence of body weight on the incidence of spontaneous tumors in rats and mice of long-term studies. American Journal of Clinical Nutrition 45: 252-260.

Séralini GE, Clair E, Mesnage R, Gress S, Defarge N, Manuela Malatesta M et al (2014). Republished study: long-term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize. Environmental Sciences Europe, 26(1):1–14. doi:10.1186/s12302-014-0014-5.

Suresh TP. (1996) Combined Chronic Toxicity and Carcinogenicity Study with Glyphosate Technical in Wistar Rats. Unpublished. Toxicology Department Rallis Research Centre, Rallis India Limited, TOXI-1559, 002/1-GPT-CARCI-M. MRID 49987401.

Stout LD, Ruecker PA. (1990). Chronic Study of Glyphosate Administered in Feed to Albino Rats. Unpublished. MRID 41643801; Historical Controls MRID 41728700.

Sugimoto K. (1997), HR-001: 18-Month Oral Oncogenicity Study in Mice, Vol. 1 and 2. Unpublished. The Institute of Environmental Toxicology, 2-772, Suzuki-cho, Kodaira-shi, Tokyo, 187, Japan, Study No.:IET 94-0151. MRID 50017108, 50017109.

Tarone RE. (2018a) On the International Agency for Research on Cancer classification of glyphosate as a probable human carcinogen. European Journal of Cancer Prevention 27: 82-87.

Tarone RE. (2018b) Conflict of Interest, bias and the IARC Monographs Program. Regulatory Toxicology and Pharmacology 98: A1-A4.

U. S. EPA (United States Environmental Protection Agency). (1998) Health effects test guidelines OPPTS 870.4200 carcinogenicity. EPA 712–C–98–211, August 1998.

U. S. EPA (United States Environmental Protection Agency). (2016) Glyphosate Issue Paper: Evaluation of Carcinogenic Potential. EPA's Office of Pesticide Programs, September 12, 2016.

U. S. EPA (United States Environmental Protection Agency). (2017) FIFRA Scientific Advisory Panel Meeting Minutes and Final Report No. 2017-01. A set of scientific issues being considered by the Environmental Protection Agency regarding: EPA's evaluation of the carcinogenic potential of Glyphosate. December 13–16, 2016 FIFRA Scientific Advisory Panel Meeting Held at the EPA Conference Center, One Potomac Yard Arlington, Virginia. Retrieved from https://www.epa.gov/sites/production/files/2017-03/documents/december_13-16_2016_final_report_03162017.pdf

U. S. EPA (United States Environmental Protection Agency). (2019) News Releases from Headquarters Chemical Safety and Pollution Prevention (OCSPP) "EPA Takes Next Step in Review Process for Herbicide Glyphosate, Reaffirms No Risk to Public Health" https://www.epa.gov/newsreleases/epa-takes-next-step-review-process-herbicide-glyphosate-reaffirms-no-risk-public-health Accessed August 13, 2019.

Westfall P, Young S. (1989) P-value adjustment for multiple tests in multivariate binomial models. Journal of the American Statistical Association 84, 780-786.

Westfall P, Young S. (1993) Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment, John Wiley and Sons, Inc. New York.

WHO (World Health Organization). (2016) Pesticide residues in food-2016: toxicological evaluations. Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues, Geneva, Switzerland, 9-13 May 2016, World Health Organization 2017. ISBN 978-92-5-109246-0.

Wood E, Dunster J, Watson P, Brooks P. (2009a) Glyphosate Technical: Dietary Combined Chronic Toxicity/Carcinogenicity Study in the Rat. Unpublished. Harlan Laboratories Limited, Shardlow Business Park, Shardlow, Derbyshire DE72 2GD, UK. Study No. 2060-012. April, 23, 2009. MRID 49957404.

Wood E, Dunster J, Watson P, Brooks P. (2009b) Glyphosate Technical: Dietary Carcinogenicity Study in the Mouse. Unpublished. Harlan Laboratories Limited, Shardlow Business Park, Shardlow, Derbyshire DE72 2GD, UK. Study No. 2060-011. April, 22, 2009. MRID 49957402.

Table 1. Characteristics of Bioassays Used in This Analysis

Bioassay	Species	Strain	# Dose Groups/ Sex	Animals/Dose	Maximum Dose[a] (mg/kg/d)		Maximum Weeks on test	Sites Where Histopathology was Conducted in All Dose Groups[b]
					Males	Females		
Atkinson et al 1993b[c]	Mouse	CD-1	4	50	988	1000	105	Kidney, Liver, Lung, Vascular System
Knezevich and Hogan 1983[c]	Mouse	CD-1	4	50	4841	5873	102	(all)
Wood et al 2009b	Mouse	CD-1	4	51	810	1081	81	Kidney, Liver and Lung
Sugimoto 1997	Mouse	CD-1-ICR	4	50	4348	4116	78	(all)
Atkinson et al 1993a[c]	Rat	SD[d]	5	50	1007	1018	105	Kidney, Liver, Lung and Salivary Glands: Parotid, Mandibular and Sublingual
Lankas 1981[c]	Rat	SD	4	50	31.49	34.02	111	(all)
Stout and Ruecker 1990[c]	Rat	SD	4	60	940	1183	105	(all)
Brammer 2001[c]	Rat	Wistar	4	64	1214	1498	104	(all)
Suresh 1996	Rat	Wistar	4	50	595.2	886	107	(none)
Wood et al 2009a	Rat	Wistar	4	51	1077	1382	105	Kidney, Liver, Lung and Bone Marrow

[a] All doses in each bioassay are listed in Table 2.

[b] Systemic tumors are assumed to have been searched for if at least one tissue in an animal was given a histopathological examination.

[c] These six studies were evaluated by IARC. IARC (2015) also reviewed two additional studies in which they identified shortcomings, but which they did not claim were "inadequate": Chruscielska et al (2000) and JMPR (2006). No glyphosate-related tumor responses were noted in either of these studies.

[d] SD = Sprague Dawley.

Table 2. Results for Test of Dose-Related Decrease in Survival

Bioassay	Species/ Sex	1st Row: Dose (mg/kg/day) 2nd Row: # survivors/total #					p-value[a]
Atkinson et al. 1993b	M/M	0	98	279	988		
		26/50	25/50	29/50	24/50		0.33
	M/F	0	102	298	1000		
		21/50	16/50	26/50	24/50		0.86
Knezevich and Hogan 1983	M/M	0	157	814	4841		
		20/50	16/50	17/50	26/50		0.98
	M/F	0	190	955	5873		
		20/50	12/50	27/50	23/50		0.88
Wood et al. 2009b	M/M	0	71.4	234.2	810		
		39/51	41/51	39/51	35/51		0.10
	M/F	0	97.9	299.5	1081		
		37/51	38/51	38/51	40/51		0.75
Sugimoto 1997	M/M	0	165	838.1	4348		
		26/50	34/50	27/50	29/50		0.47
	M/F	0	153.2	786.8	4116		
		32/50	36/50	40/50	35/50		0.51
Atkinson et al. 1993a	R/M	0	10	101	306	1007	
		28/50	25/50	31/50	28/50	31/50	0.79
	R/F	0	10	103	311	1018	
		21/50	22/50	22/50	20/50	26/50	0.85
Lankas 1981	R/M	0	3.05	10.3	31.49		
		15/50	26/50	16/50	26/50		0.94
	R/F	0	3.37	11.22	34.02		
		18/50	23/50	30/50	15/50		0.09
Stout and Ruecker 1990	R/M	0	89	362	940		
		14/60	19/60	17/60	17/60		0.58
	R/F	0	113	457	1183		
		22/60	22/60	17/60	18/60		0.18
Brammer 2001	R/M	0	121	361	1214		
		16/64	17/64	18/64	26/64		0.98
	R/F	0	145	437	1498		
		32/64	28/64	39/64	30/64		0.41
Suresh 1996	R/M	0	6.3	59.4	595.2		
		20/50	20/50	18/50	29/50		0.99
	R/F	0	8.6	88.5	886		
		24/50	26/50	33/50	21/50		0.07
Wood et al. 2009a	R/M	0	85.5	285.2	1077		
		39/51	37/51	38/51	45/51		0.97
	R/F	0	104.5	348.6	1382		
		37/51	34/51	36/51	39/51		0.80

[a]Test for trend toward progressive fewer surviving animals at higher doses.

Table 3. Tumors Giving a Poly-3 p-value ≤ 0.05 for a Positive Trend in the Ten Glyphosate Bioassays Investigated

Bioassay	Species/ Sex	Tumor	Summary Tumor Incidence					Poly-3 p-value	Cited by IARC[a]
Atkinson et al. 1993b	M/M	Haemangiosarcoma	0/50	0/50	0/50	4/50		0.0013	IARC
Lankas 1981	R/F	Thyroid: C-cell Carcinoma	1/47	0/49	2/50	6/47		0.0015	
Sugimoto 1997	M/F	Hemangioma	0/50	0/50	2/50	5/50		0.0028	
Sugimoto 1997	M/F	Hemangioma, Hemangiosarcoma	0/50	0/50	3/50	5/50		0.0062	
Stout and Ruecker 1990	R/F	Adrenal: Cortical Carcinoma	0/60	0/60	0/60	3/60		0.0072	
Sugimoto 1997	M/F	Osteoma, Osteosarcoma	0/50	0/50	0/50	3/50		0.0074	
Wood et al. 2009b	M/M	Lymphoma	0/51	1/51	2/51	5/51		0.0076	
Brammer 2001	R/M	Liver: Hepatocellular Adenoma	0/64	2/64	0/64	5/64		0.014	
Lankas 1981	R/M	Testis: Interstitial Cell Tumor	0/50	3/50	1/50	6/50		0.021	
Stout and Ruecker 1990	R/M	Liver: Hepatocellular Adenoma	3/60	2/60	3/60	8/60		0.022	IARC[b]
Atkinson et al. 1993a	R/F	Lipoma	0/50	0/50	0/50	0/50	2/50	0.022	
Wood et al. 2009b	M/M	Lung: Adenocarcinoma	5/50	5/51	7/51	11/51		0.025	
Knezevich and Hogan 1983	M/M	Kidneys: Renal Tubal Adenoma	0/49	0/49	1/50	3/50		0.034	IARC
Lankas 1981	R/F	Lipoma	0/50	0/50	0/50	2/50		0.036	
Sugimoto 1997	M/M	Malignant Lymphoma	2/50	3/50	0/50	6/50		0.038	
Knezevich and Hogan 1983	M/F	Lymphoblastic Lymphosarcoma	0/50	1/50	0/50	3/50		0.041	
Sugimoto 1997	M/F	Osteosarcoma	0/50	0/50	0/50	2/50		0.041	
Sugimoto 1997	M/M	Kidney: Adenoma	0/50	0/50	0/50	2/50		0.042	
Sugimoto 1997	M/M	Hemangiosarcoma	0/50	0/50	0/50	2/50		0.043	
Stout and Ruecker 1990	R/M	Neurofibroma, Neurofibrocarcinoma	0/60	0/60	0/60	2/60		0.045	
Sugimoto 1997	M/F	Harderian Gland: Adenoma	1/50	3/50	0/50	5/50		0.046	
Stout and Ruecker 1990	R/F	Thyroid Gland: C-cell Adenoma	2/60	2/60	6/60	6/60		0.047	IARC
Suresh 1996	R/M	Lymphoma	0/50	0/50	0/50	2/50		0.049	
Stout and Ruecker 1990	R/F	Thyroid Gland: C-cell Adenoma or Carcinoma	2/60	2/60	7/60	6/60		0.049	

[a] Indicates tumor responses cited by IARC (2015) as evidence of carcinogenicity. Pancretic islets in male rats in Stout and Ruecker (1990) was also cited by IARC (1/58, 8/57, 5/60 and 7/59) but this response did not give a p-value ≤ 0.05 by the Poly-3 trend test.

[b] IARC (2015) resported tumor responses of 2, 2, 3, 7.

Table 4. Results of Multi-response Permutation Tests for Positive Dose-Related Trends in Tumor Occurrence

	Primary Analysis		Reduced Analysis[a]	
Description of Test	Test Statistic[b]	Statistical Significance of Test Statistic[c]	Test Statistic[b]	Statistical Significance of Test Statistic[c]
Min Test	p = 0.0013	p = 0.26	p = 0.0013	p = 0.17
05 Test	24	p = 0.08	14	p = 0.12
01 Test	7	p = 0.06	4	p = 0.08
Number of trend tests	525		304	

[a] The reduced analysis removed from the analysis tumors and combinations of tumors that were included in larger combinations.

[b] The test statistic of the min test is the smallest poly-3 p-value obtained from any tumor in any study in the original data. The test statistics of the 05 test and the 01 test are the number of tumors for which the poly-3 p-value was ≤ 0.05 or ≤ 0.01, respectively.

[c] Calculated using 5,000 simulations.

Table 5. Results of Multi-response Permutation Tests for Negative Dose-Related Trends in Tumor Occurrence

	Primary Analysis		Reduced Analysis[a]	
Description of Test	Test Statistic[b]	Statistical Significance of Test Statistic[c]	Test Statistic[b]	Statistical Significance of Test Statistic[c]
Min Test	p = 0.0008	p = 0.11	p = 0.0011	p = 0.10
05 Test	26	p = 0.08	15	P = 0.12
01 Test	10	p = 0.002	6	p = 0.002
Number of trend tests	525		304	

[a] The reduced analysis removed from the analysis tumors and combinations of tumors that were included in larger combinations.

[b] The test statistic of the min test is the smallest poly-3 p-value obtained from any tumor in any study in the original data. The test statistics of the 05 test and the 01 test are the number of tumors for which the poly-3 p-value was ≤ 0.05 or ≤ 0.01, respectively.

[c] Calculated using 5,000 simulations.

Figure 1. Histogram of p-values obtained from one-sided poly-3 tests for positive trend applied to ten glyphosate bioassays from tumor groupings that contain at least two tumors

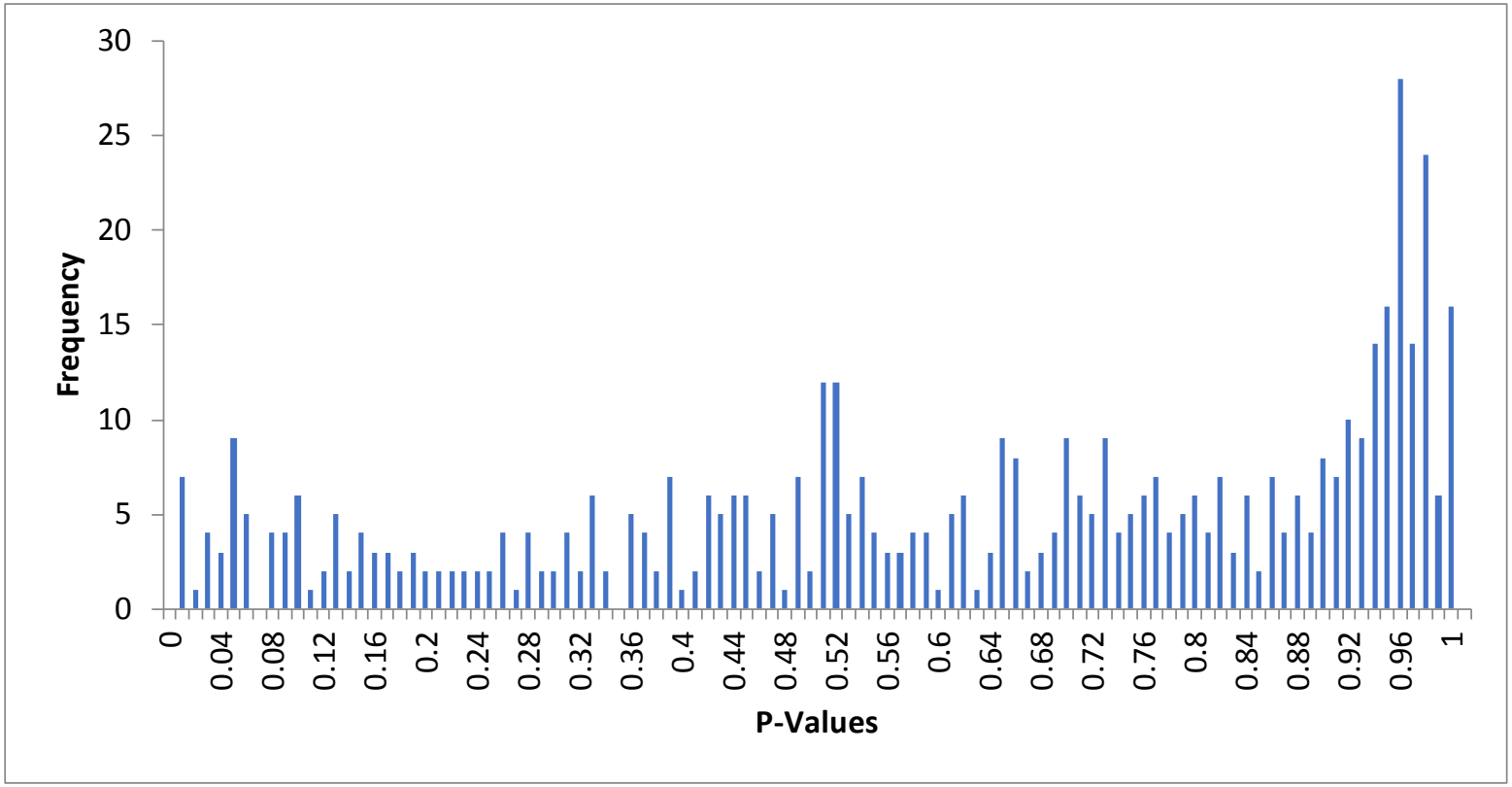


Figure 2: Outline of permutation analysis

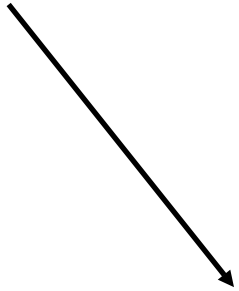
For Each of the Ten Bioassays in Each of 5,000 Iterations:

CH studies and mandatory tissues (including systemic tumors) from ICH studies

Randomly permute animals across dosed and control groups, maintaining the same number of survivors in each group as seen in the original data.

Other tissues from ICH studies

Randomly permute non-survivors across dosed and control groups, maintaining the same number in each group as was seen in the actual data; permute the survivors (high dose and control only), maintaining the same numbers in each group as was seen in the original data; exclude the surviving intermediate dosed animals.



Carry out poly-3 trend tests for all tumors and identified tumor combinations, noting if a trend test produces a p-value ≤ 0.05 or 0.01. Also, note the most significant trend p-value observed in all ten studies.

Repeat this process 5,000 times (i.e., 5,000 randomizations). Compare the results of the trend tests with that of the actual data. Specifically, determine the proportions of randomizations in which the number of trend p-values ≤ 0.05 (for the 0.05 test) or 0.01 (for the 0.01 test) equal or exceed the number in the original data, and determine the proportion of randomizations in which the most significant trend p-value is equal to or more significant than the most significant one from the original data (for the min test).

The above primary

procedure is carried out four times: for both increasing and decreasing trends in tumor incidence, and for both the analysis and the reduced analysis.

Abbreviations: CH = complete histology, ICH = incomplete histology.

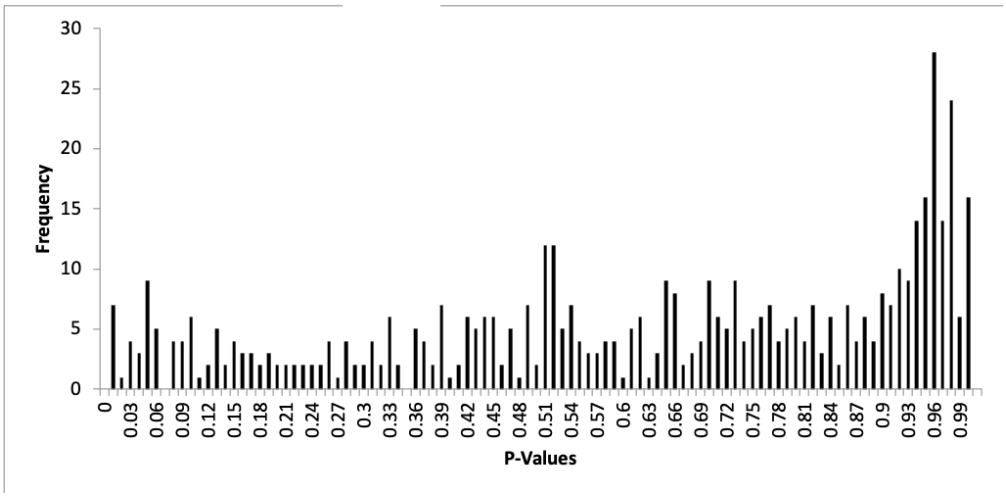


Figure 1. Histogram of p-values obtained from one-sided poly-3 tests for positive trend applied to ten glyphosate bioassays from tumor groupings that contain at least two tumors

173x84mm (150 x 150 DPI)

Figure 2: Outline of permutation analysis

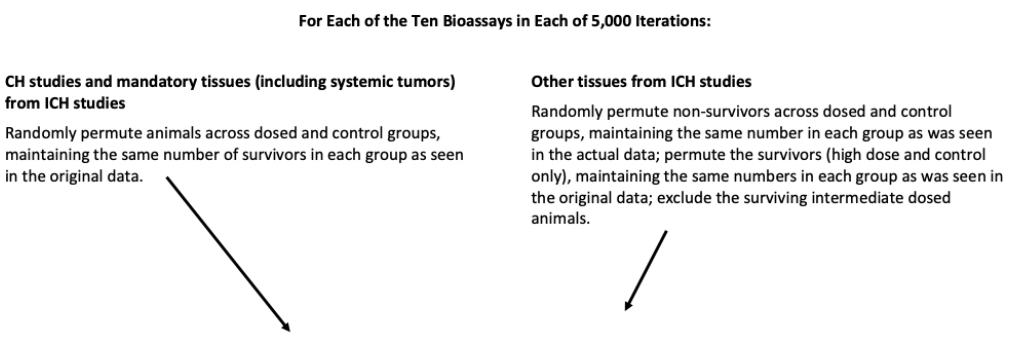


Figure 2

164x107mm (150 x 150 DPI)