Taylor & Francis
Taylor & Francis Group

Critical Reviews in Biotechnology

ISSN: 0738-8551 (Print) 1549-7801 (Online) Journal homepage: http://www.tandfonline.com/loi/ibty20

Published GMO studies find no evidence of harm when corrected for multiple comparisons

Alexander Y. Panchin & Alexander I. Tuzhikov

To cite this article: Alexander Y. Panchin & Alexander I. Tuzhikov (2016): Published GMO studies find no evidence of harm when corrected for multiple comparisons, Critical Reviews in Biotechnology

To link to this article: http://dx.doi.org/10.3109/07388551.2015.1130684

Published online: 14 Jan 2016.

Critical Reviews in Biotechnology

http://informahealthcare.com/bty
ISSN: 0738-8551 (print), 1549-7801 (electronic)

Crit Rev Biotechnol, Early Online: 1–5
 DOI: 10.3109/07388551.2015.1130684

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

Published GMO studies find no evidence of harm when corrected for multiple comparisons

Alexander Y. Panchin[1] and Alexander I. Tuzhikov[1,2]

[1]Institute for Information Transmission Problems RAS, Moscow, Russian Federation and [2]Department of Ophthalmology, School of Medicine, Bascom Palmer Eye Institute, University of Miami Miller Miami, FL, USA

Abstract

A number of widely debated research articles claiming possible technology-related health concerns have influenced the public opinion on genetically modified food safety. We performed a statistical reanalysis and review of experimental data presented in some of these studies and found that quite often in contradiction with the authors' conclusions the data actually provides weak evidence of harm that cannot be differentiated from chance. In our opinion the problem of statistically unaccounted multiple comparisons has led to some of the most cited anti-genetically modified organism health claims in history. We hope this analysis puts the original results of these studies into proper context.

Keywords

Genetic engineering, genetic modification, GMO, health, safety, statistics, transgenic

History

Received 1 March 2015
Revised 8 November 2015
Accepted 25 November 2015
Published online 11 January 2016

Introduction

The overall negative public perception of the use of genetically modified organisms (GMOs) in food production has caused severe difficulties for the development of GM crops and their commercialization both in developed and developing countries.[1,2] The debates have been stirred by a number of publications claiming that GM crops might not be equivalent to conventional crops and could pose health hazards for the consumer due to some yet unknown mechanisms. Here, we show that a number of articles some of which have strongly and negatively influenced the public opinion on GM crops and even provoked political actions, such as GMO embargo, share common flaws in the statistical evaluation of the data. Having accounted for these flaws, we conclude that the data presented in these articles does not provide any substantial evidence of GMO harm.

In statistics, the null hypothesis refers to the default position that there is no relationship between two measured phenomena (such as treatment and outcome). Rejection of the null hypothesis allows one to conclude that a relationship may exist between the two phenomena in question. The null hypothesis is generally assumed true unless evidence indicates otherwise. Before testing the null hypothesis, a researcher chooses a threshold, called the significance level (α) value. If the p values of a statistical test is smaller than α, the observed data is considered inconsistent with the assumption that the null hypothesis is true, and thus the null hypothesis should be rejected.

Address for correspondence: Alexander Y. Panchin, Institute for Information Transmission Problems RAS, Moscow, Russian Federation. E-mail: alexpanchin@yahoo.com

If multiple hypotheses are tested simultaneously, the problem of multiple comparisons arises. If one tests several independent null hypotheses and leaves α at 0.05 for each comparison, the probability of obtaining at least one ''statistically significant'' result is >5% even if all null hypotheses are true. This problem is illustrated in Figure 1. The probability of at least one ''significant'' outcome in a comparison is computed from the number of comparisons (N) using the following equation: $(1.00 - 0.95^N) \times 100\%$.

In 2010, Bennett et al. [3] published an article called ''Neural Correlates of Interspecies Perspective Taking in the Postmortem Atlantic Salmon: An Argument For Proper Multiple Comparisons Correction''. The authors convincingly provide an anecdotal example of unexpected biological findings due to a statistical methodology flaw mocking the conclusion of a ''miraculous ability of a dead fish brain to retain activity long after the death of the fish''. Similar problems have been documented in related fields.[4]

The Bonferroni correction [5] is the simplest method used to counteract the problem of multiple comparisons. The correction states that if an experimenter is testing n hypotheses, he should test each hypothesis at α level of $1/n$ times what it would be if only one hypothesis was tested. The Bonferroni correction is conservative yet justified [6] in such cases as those discussed below because virtually any difference between GMOs and conventional crops is presented as a cause for concern, a large number of tests are carried out without a preplanned hypothesis on what these differences might be and false positive errors have important social consequences.

Previously, Nicolia et al. [7] summarized the existing evidence on GMO safety in a review of over 1500 published articles and found little to no evidence of harm. Snell et al. [8]

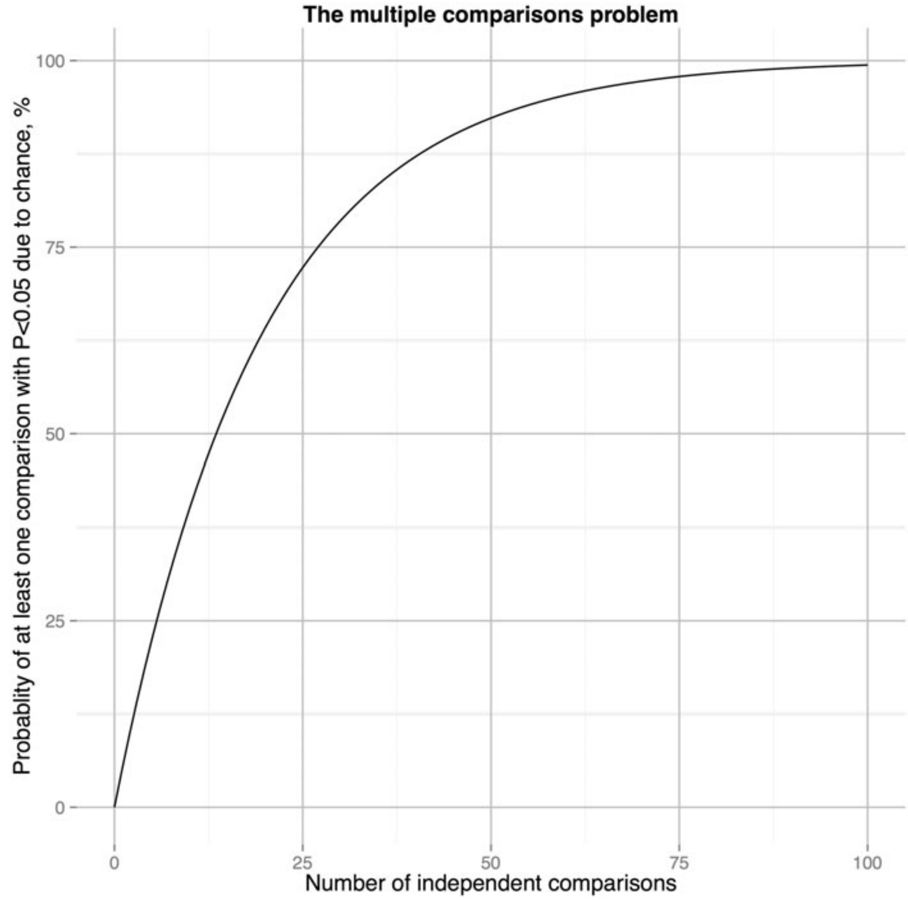

Figure 1. The problem of multiple comparisons. The probability of at least one ‘‘significant’’ comparison is computed from the number of comparisons (N) using the following equation: $(1.00 - 0.95^N) \times 100\%$.

have reviewed articles about the long-term consequences of GMO consumption, for up to five generations of animals and failed to find any evidence of harm. The European Commission in their report ‘‘A decade of EU-funded GMO research’’ that presents the results of 50 projects, involving >400 research groups concludes that ‘‘biotechnology, and in particular GMOs, are not *per se* more risky than, e.g. conventional plant breeding technologies’’.[9] In addition Van Eenennaam and Young [10] have argued that ‘‘field data sets, representing over 100 billion animals following the introduction of GE crops, did not reveal unfavorable or perturbed trends in livestock health and productivity’’.

During the period of 2001–2011, scientists from the Scientific Research Institute of Nutrition in Russia performed over a dozen studies to evaluate the safety of some commercially grown GMOs. One of the largest included a three-generation study of 630 adult rats with 2837 rats of progeny.[11] None of these government funded studies has revealed any evidence of harm from diets containing commercially available GM foods comparing to foods derived from isogenic organisms.

Only a handful of studies claim to have provided evidence that GM crops differ from related conventional crops by more than just the desired effects of the genetic modification, however these studies received much media attention. Because a vast majority of studies supports GMO safety, credible evidence to the contrary must be very strong and accompanied by a proposed mechanism of action. This is an additional argument as to why a conservative correction for multiple comparisons is warranted when the safety of GMOs is brought into question by experimental data.

We reviewed published articles in which undesired and statistically significant differences between GMOs and conventional crops were reported. A summary of the number of multiple comparisons performed for each reviewed claim is presented in Table 1. Since the problem of multiple comparisons can only lead to false positive results, we did not reanalyze the abundant studies in which no significant difference between GMO and conventional crops had been observed. The application of the Bonferroni correction would not alter the conclusions of these articles.

Application of the Bonferroni correction

We begin with the 1999 article by Ewans and Pusztai [12] published in Lancet. The authors incorporated a gene encoding a lectin (lectins are considered toxic) [13,14] into a potato genome and checked whether or not the lectin-GM potato would cause a different effect on rats rather than a wild-type potato with lectin added as a food supplement. Mucosal thickness (stomach) or crypt length (jejunum, ileum, cecum and colon) were measured. A statistically significant effect of the genetic modification itself, as measured using Tukey’s test ($p = 0.041$), was observed for the jejunum, although barely passing the $\alpha = 0.05$ threshold. It was concluded that the genetic modification

Table 1. A summary of GMO health claims, the number of multiple comparisons and how the application of the Bonferroni correction would have altered the study's conclusions.

Main GMO health claim	Study	Before Bonferroni correction	Number of multiple comparisons	After Bonferroni correction
Crypt length	Ewen and Pusztai [12]	Claim supported	5	Claim not supported
Gut microbiota	Xu et al. [17]	Claim supported	42	Claim not supported
Female mortality	Seralini et al. [19]	Claim not supported	See text	Claim not supported
Organ pathology	Seralini et al. [19]	Claim not supported	>60	Claim not supported
Organ pathology	Carman et al. [24]	Claim supported	42	Claim not supported
Liver proteome and histology	Malatesta et al. [25]	Claim supported	>1400	Claim not supported
Nutrient composition	Bohn et al. [27]	Claim supported	35	See text

changed the potato's effect on the jejunum (but not on the other organs).

The statistical evaluation did not account for the five independent outcomes (multiple comparisons): stomach mucosal thickness or crypt length for jejunum, ileum, cecum or colon. Bonferroni corrected α is $0.05/5 = 0.01$, <0.041. The Royal Society and independent scientists criticized this study on different grounds [15,16] as there had been differences in protein content of the potatoes and the experiments had not been ''blind''. However, this does not matter if the observed differences between rats fed lectin-GM potato and wild-type potato with added lectin are not statistically significant.

An article, which was published in the Journal of Food Sciences in 2011 [17] and gained wide publicity, states that *Lactobacillus* bacteria were more abundant in the cecum of male rats fed with 70% non-genetically modified rice comparing to the cecum of rats fed with 70% genetically modified rice (GMR) ($p < 0.05$). However, the study effectively involved 42 multiple comparisons: two sexes of rats were compared independently, for each there were seven types of microbial differences analyzed at three different feeding levels. By chance alone one may expect an average of two between-groups comparisons to yield significant results at $\alpha = 0.05$. In addition, since both an increase and decrease of microbial content could be regarded as an effect of the GM diet a two-tailed test should be used. Having applied the Bonferroni correction, we conclude that there are no statistically significant differences observed between rats fed non-GM or GMR in this study, invalidating the authors' conclusions that their results ''suggested that GMR had a complex effect on cecal microflora that may be related to the health of the host''. It is also worth mentioning that the effect was observed only in male, but not in female rats and that high diversity it the amount of *Lactobacillus* and other bacteria has been observed in other studies even between rats fed the same diet.[18]

Next, we will review the study by Seralini et al.[19] The authors chose to drop the ''statistics part'', related to the main conclusions of the article, all together [20,21] so we had to obtain the p values ourselves using a straight-forward and transparent statistical significance test used in the analysis of contingency tables – Fishers exact test (other tests would yield similar results). In the current reanalysis of the hypothesis presented by Seralini et al., the numerical data was taken from the original article. The number of multiple comparisons is discussed but is not relevant in this case because the study reveals no significant effects of GM food on animal health even without the Bonfferoni correction.

In this study, male and female rats were fed genetically modified maize (GMM) in three different concentrations (11%, 22% and 33%) with and without an added herbicide (six GMM-fed groups for each sex). A modest control set of rats ($N = 10$ for each sex) was used. One of the main conclusions was that ''in females, all the treated groups died 2–3 times more than controls''. As follows from the article: 2 female rats out of 10 died before the mean survival time in the control group, compared to 29 out of 60 in the pooled six GMM-fed groups. However, this difference is not statistically significant ($p = 0.09$, two-tailed Fishers exact test, not adjusted for multiple comparisons).

Seralini et al. state: ''20% females (only two) died spontaneously, while up to 50% males and 70% females died in some groups on diets containing the GMM''. The highest mortality was observed for the group of female rats fed 22% GMO: 7 out of 10 rats died prematurely. Even if we compare the mortality in this particular group of rats with controls, we conclude that the observed difference is not statistically significant ($p = 0.07$, two-tailed Fishers exact test, not adjusted for 12 multiple comparisons). In addition, mortality rates among female rats were lower in the 33% GMO group comparing to the 22% GMO group, thus a dose–response effect was not observed.

Among males, 3 rats out of 10 died prematurely in the control group, compared to 19 out of 60 rats in the six pooled GM-fed groups (not significant $p = 0.615$, two-tailed Fishers exact test). One of the most curious facts is that if we set aside statistical significance, take the data provided by Seralini et al. and use same ''cherry picking'' strategy, we could state that ''in males, groups with 22% and 33% GMM in their diet died 3 times less than controls'' ($p = 0.291$, two-tailed Fishers exact test, not significant). Does a high dose of GMM prolong the lifespan of male rats? The authors did not report this ''observation''.

The article states: ''in treated males, liver congestions and necrosis were 2.5–5.5 times higher'' and that ''females developed large mammary tumors almost always more often than and before controls''. Two male rats out of 10 had liver pathologies in the control group, compared to 30 out of 60 GMM fed male rats. About 5 female rats out of 10 developed mammary tumors in the control group, compared to 44 out of 60 GMM fed female rats. Neither of these differences are statistically significant ($p = 0.076$ and $p = 0.133$, two-tailed Fishers exact test).

Over 30 different organs were analyzed in the study and the two sexes were analyzed independently. Seralini et al. provide the data selectively, however none of the reported differences between the number of rats with specific organ pathologies in control and GMM fed groups are statistically significant even at $\alpha = 0.05$ (two-tailed Fishers exact test without the Bonferroni correction). In his response to the criticism of the absent statistical analysis, Dr. Seralini gave a laconic reply: ‘‘Statistics do not tell the truth, but may help in understanding results’’.[22] Indeed and clearly so! As a side note: the rat tumors presented as scary imagery in the study are frequent in Sprague–Dawley rats and have a reported spontaneous incidence of 45% over a 1.5-year period.[23]

We now review an article titled ‘‘a long-term toxicology study on pigs fed a combined genetically modified (GM) soy and GMM diet’’.[24] In this study, the authors claim that female GM-fed pigs had median uterine weights that were 25% greater than non-GM-fed pigs ($p = 0.025$). The authors measured the weight changes in the pigs' kidneys, heart, liver, spleen, lung, stomach and ovaries. That is eight multiple comparisons, leading to $\alpha = 0.05/8 = 0.00625$. Thus, the observed difference is not statistically significant.

The authors also claim that the pigs fed the GM diet showed 2.6 times the rate of severe stomach inflammation compared to non-GM-fed pigs ($p = 0.004$). And that, GM-fed male pigs showed severe stomach inflammation at a rate of 4.0 times that of the non-GM-fed male pigs ($p = 0.041$); and female pigs showed a rate of severe stomach inflammation that was 2.2 the rate of the non-GM-fed female pigs ($p = 0.034$). The considered set of pathologies includes kidney, heart, liver, spleen, intestine and ovary abnormalities, lung pneumonia, pericarditis and abnormal lymph nodes. Even separate fields for mild, moderate and severe inflammation, erosion, pin-point ulcers, frank ulcers and bleeding ulcers were chosen. The moderate and severe stomach inflammations combined demonstrate no significant effects at $\alpha = 0.05$. In any case, there are at least 17 multiple comparisons. The corrected α is at most $0.05/17 = 0.0029$. Even $p = 0.004$ for severe stomach inflammation turns out not statistically significant. The authors also looked at 17 blood biochemistry parameters but reported nothing of significance. Strictly speaking, the total number of tested hypothesis in this study was 42, thus our estimates of α could be even more conservative.

In 2008, Malatesta et al. [25] published a study on the effects of GM soy on liver aging in female mice. For this analysis, the authors used a modest sample size of 10 rats in the experimental and control groups. No macroscopic alterations or pathologic lesions were observed in any organ of any animals. However, the authors compared the quantity of $\sim$1400 proteins in the liver between groups. The total number of tested hypothesis was even greater since the authors also examined a number of histological parameters and performed a number of immunoelectron microscopy tests. Unsurprisingly, some significant differences were found (including difference for the quantity of 49 proteins and some other parameters, with some p values as low as 0.001), since multiple comparisons had not been accounted for. The Bonferroni corrected α is $\sim 0.05/1400 = 0.000035$, thus no conclusions can be drawn from the analysis. The problem with similar studies with small sample sizes and a large flexibility in designs, definitions and outcomes is discussed elsewhere.[26]

We close with a 2014 article where the authors compared the composition of nutrients and elements in GM, conventional and organic soybeans.[27] The data showed that organic soybeans had more total protein ($p = 0.003$) than the GM soy, more ash ($p = 0.005$), more lysine ($p = 0.002$), more arginine ($p = 0.04$), more indispensable amino acids ($p = 0.037$), less palmitic acid ($p = 0.046$), less total saturated fatty acids ($p = 0.001$) more linoleic acid ($p = 0.01$), more barium ($p = 0.00005$), less selenium ($p = 0.0003$) and more zinc ($p = 0.0002$). Granting that the total number of hypothesis tested in this article was 35, application of Bonferroni correction is due. The corrected significance level is $0.05/35 = 0.00142$, and thus only four of these comparisons are significant. In all of the truly significant cases (and in general), GM soy appeared very close to conventional soy, so it is mostly the difference between organic and non-organic soy we are dealing with. A comparison of organic and conventional soy based on this data falls outside the scope of this review.

Conclusions

The presented articles suggesting possible harm of GMOs received high public attention. However, despite their claims, they actually weaken the evidence for the harm and lack of substantial equivalency of studied GMOs. We emphasize that with over 1783 published articles on GMOs over the last 10 years [7] it is expected that some of them should have reported undesired differences between GMOs and conventional crops even if no such differences exist in reality.

It has been argued that we might be underestimating the number of false-positive results in science in general due to bias, improper use of statistics, analysis of highly improbably hypothesis and other factors.[26] The suggested solution was to take preference for large studies or low-bias meta-analyses and to take into account the pre-study probabilities of a finding being true.

We argue that the totality of the evidence should be taken into account when drawing conclusion on GMO safety, instead of far-fetched evidence from single studies with a high risk of bias due to a large number of multiple comparisons. Perhaps more focus should be drawn to clear and relevant outcomes such as mortality rates, life expectancy or reproductive success.

Unfortunately, it takes just a single article claiming a mild difference between GM and non-GM products to stir the public debate and cause a long-lasting hysteria. Though the article by Seralini was retracted,[28] it still keeps being cited by the media. We are convinced that policy makers, media representatives and the public should pay less attention to individual articles about GMO health concerns until the results are confirmed by independent studies. Such studies should undergo rigorous review, including statistical evaluation. Quoting Carl Sagan: ‘‘extraordinary claims require extraordinary evidence’’.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article. This article was not financed by any non-government organization or lobbied by any entity, except for the common sense. This study was supported by the Russian Science Foundation (project 14-50-00150). The funding body did not in any way interfere with or participate in the preparation of the article. A.I.T. was supported by NIH grant EY02238. We are grateful to anonymous reviewers who suggested a number of improvements for our article.

References

Herring RJ. On risk and regulation: Bt crops in India. GM Crops Food. 2014;5:204–209.

Kuntz M. The GMO case in France: politics, lawlessness and postmodernism. GM Crops Food. 2014;5:163–169.

Bennett C, Baird A, Miller M, et al. Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: an argument for proper multiple comparisons correction. J Serendipitous Unexpected Results. 2010;1:1–5.

Ford I. Can statistics cause brain damage? J Cereb Blood Flow Metab. 1983;3:259–262.

Abdi H. Bonferroni test. In N. Salkind (Ed.), Encyclopedia of measurement and statistics. Thousand Oaks, CA: SAGE Publications, 104–108.

Armstrong R. When to use the Bonferroni correction. Ophthal Physiol Opt. 2014;34:502–508.

Nicolia A, Manzo A, Veronesi F, et al. An overview of the last 10 years of genetically engineered crop safety research. Crit Rev Biotechnol. 2014;34:77–88.

Snell C, Bernheim A, Berge JB, et al Assessment of the health impact of GM plant diets in long-term and multigenerational animal feeding trials: a literature review. Food Chem Toxicol. 2012;50:1134–1148.

European_Comission. A decade of EU-funded GMO research. 2010. Available from: http://ec.europa.eu/research/biosociety/pdf/a_decade_of_eu funded_gmo_research.pdf.

Van Eenennaam AL, Young AE. Prevalence and impacts of genetically engineered feedstuffs on livestock populations. J Anim Sci. 2014;92: 4255–4278.

Tyshko NV, Zhminchenko VM, Pashorina VA, et al Assessment of the impact of GMO of plant origin on rat progeny development in 3 generations. Vopr Pitan. 2011;80:14–28.

Ewen SW, Pusztai A. Effect of diets containing genetically modified potatoes expressing Galanthus nivalis lectin on rat small intestine. Lancet. 1999;354:1353–1354.

Miyake K, Tanaka T, McNeil PL. Lectin-based food poisoning: a new mechanism of protein toxicity. PLoS One. 2007;2:e687.

Cordain L, Toohey L, Smith MJ, et al. Modulation of immune function by dietary lectins in rheumatoid arthritis. Br J Nutr. 2000;83:207–217.

Royal Society UK. Review of data on possible toxicity of GM potatoes. Annual Review of Plant Biology. 1999;59:771–812.

Kuiper HA, Noteborn HP, Peijnenburg AA. Adequacy of methods for testing the safety of genetically modified foods. Lancet. 1999;354: 1315–1316.

Xu W, Li L, Lu J, et al Analysis of caecal microbiota in rats fed with genetically modified rice by real-time quantitative PCR. J Food Sci. 2011;76:M88–M93.

Hashimoto N, Nakamura Y, Noda T, et al Effects of feeding potato pulp on cholesterol metabolism and its association with cecal conditions in rats. Plant Foods Hum Nutr. 2011;66:401–407.

Seralini GE, Clair E, Mesnage R, et al Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize. Food Chem Toxicol. 2012;50:4221–4231.

Panchin AY. Toxicity of Roundup-tolerant genetically modified maize is not supported by statistical tests. Food Chem Toxicol. 2013;53:475.

Trewavas A. Science requires the dispassionate presentation of information. Food Chem Toxicol. 2013;53:449.

Seralini GE, Mesnage R, Defarge N, et al Answers to critics: why there is a long term toxicity due to a Roundup-tolerant genetically modified maize and to a Roundup herbicide. Food Chem Toxicol. 2013;53: 476–483.

Prejean JD, Peckham JC, Casey AE, et al Spontaneous tumors in Sprague-Dawley rats and Swiss mice. Cancer Res. 1973;33: 2768–2773.

Carman J, Vlieger H, Steeg L, et al A long-term toxicology study on pigs fed a combined genetically modified (GM) soy and GM maize diet. J Org Syst. 2013;8:38–54.

Malatesta M, Boraldi F, Annovi G, et al A long-term study on female mice fed on a genetically modified soybean: effects on liver ageing. Histochem Cell Biol. 2008;130:967–977.

Ioannidis JP. Why most published research findings are false. PLoS Med. 2005;2:e124.

Bohn T, Cuhra M, Traavik T, et al Compositional differences in soybeans on the market: glyphosate accumulates in Roundup Ready GM soybeans. Food Chem. 2014;153:207–215.

Retraction notice to ''Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize'' [Food Chem Toxicol. 2012;50:4221–4231]. Food Chem Toxicol. 2014;63:244.