



HAL
open science

Avis en réponse à la saisine du 24 septembre 2012 relative à l'article de Séralini et al. (Food and Chemical Toxicology, 2012)

Jean-Christophe Pagès, Jean-Jacques Leguay, Yves Bertheau, Pascal Boireau,
Denis Bourguet, Florence Coignard, François Coléno, Jean-Luc Darlix, Elie
Dassa, Maryse Deguergue, et al.

► To cite this version:

Jean-Christophe Pagès, Jean-Jacques Leguay, Yves Bertheau, Pascal Boireau, Denis Bourguet, et al..
Avis en réponse à la saisine du 24 septembre 2012 relative à l'article de Séralini et al. (Food and
Chemical Toxicology, 2012). [0] Haut Conseil des Biotechnologies. 2012. hal-02916042

HAL Id: hal-02916042

<https://hal.inrae.fr/hal-02916042>

Submitted on 17 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

HAUT CONSEIL DES BIOTECHNOLOGIES

COMITE SCIENTIFIQUE

Paris, le 19 octobre 2012

AVIS

en réponse à la saisine¹ du 24 septembre 2012 relative à l'article de Séralini *et al.* (*Food and Chemical Toxicology*, 2012).

Le Haut Conseil des biotechnologies (HCB) a été saisi le 24 septembre 2012 par les Autorités compétentes françaises (le Ministère des Affaires Sociales et de la Santé, le Ministère délégué auprès du Ministère de l'Economie et des Finances, chargé de l'économie sociale et solidaire et de la consommation, le Ministère de l'Ecologie, du Développement Durable et de l'Energie, le Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt) d'une demande d'avis relative à l'article de l'équipe du Professeur Séralini publié dans la revue *Food and Chemical Toxicology*, rapportant des effets délétères chez le rat de la consommation sur le long terme de maïs génétiquement modifié NK603 et du Roundup[®], un herbicide à base de glyphosate.

Suite au travail préliminaire d'un groupe d'experts spécialement constitué pour répondre à cette saisine, le Comité scientifique (CS)² du HCB a procédé à l'examen de l'article le 2 octobre 2012 sous la présidence de Jean-Christophe Pagès.

¹ La saisine du 24 septembre 2012 est reproduite dans l'Annexe 1.

² La composition du CS ainsi que les experts externes du groupe de travail *ad hoc* sont indiqués dans l'Annexe 2.

RESUME DE L'AVIS³

En réponse à la saisine du 24 septembre 2012, le Comité scientifique (CS) du Haut Conseil des biotechnologies (HCB) a analysé l'article de l'équipe du Professeur Séralini intitulé « *Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize* » à paraître dans la revue *Food and Chemical Toxicology* (Séralini et al., 2012). Les auteurs de l'article revendiquent la première démonstration expérimentale documentée d'une toxicité sur le long terme de la consommation d'une plante génétiquement modifiée, le maïs NK603, et d'un herbicide à base de glyphosate, le Roundup®.

L'objet de ce premier avis du CS du HCB vise à déterminer si l'article présente des résultats concluants quant à une éventuelle toxicité du maïs NK603. Le second volet de la saisine demande au HCB de réfléchir à la pertinence des modalités d'évaluation des risques sanitaires des PGM et de proposer toute adaptation qui lui semblerait nécessaire. Ce sujet sera examiné dans les prochains mois sous ses aspects scientifiques et contextuels.

Suite à une expertise pluridisciplinaire, le CS du HCB conclut que l'article, essentiellement descriptif, ne permet d'établir aucune relation de causalité entre des événements observés durant l'étude et la consommation de maïs NK603, traité ou non avec du Roundup®. Plus précisément, le CS du HCB note que :

- le dispositif expérimental mis en œuvre est inadapté aux objectifs de l'étude : le nombre de rats par groupe est insuffisant et les groupes témoin ont un effectif inadapté pour pouvoir conclure à des effets statistiquement significatifs de la consommation de maïs NK603 sur deux ans en termes de toxicité chronique et de tumorigénicité chez le rat ;
- la présentation des résultats est parcellaire et imprécise. Seuls certains résultats sont sélectionnés, présentés ou commentés ; la présentation de ces résultats sélectionnés est imprécise, manque de pertinence biologique, et utilise des « nomenclatures » non conventionnelles. De cette description parcellaire et imprécise sont tirées des conclusions non justifiées, utilisées ensuite pour échafauder des hypothèses physiopathologiques qui ne peuvent être fondées ;
- les conclusions d'effets délétères de la consommation du maïs NK603 ne sont pas soutenues par l'analyse des résultats présentée dans l'article. Les données ne font l'objet d'aucune analyse statistique appropriée. Le CS du HCB a appliqué des méthodologies statistiques standard pour analyser les observations de mortalité et tumorigénicité rapportées par les auteurs dans l'article. Il en ressort qu'aucune différence statistiquement significative de mortalité ou de tumorigénicité des rats ne peut être mise en évidence entre les groupes ayant consommé du maïs NK603 et les groupes témoin. De plus, l'utilisation de données de référence du fournisseur des animaux de l'étude indique que les taux de survie et de tumorigénicité des rats ayant consommé du maïs NK603 se situent globalement dans les intervalles de prévision calculés pour les rats de cet élevage. Il est intéressant de noter que le taux de survie du groupe témoin de femelles utilisé dans cette étude sort de l'intervalle de prévision des rats de l'élevage. Ceci confirme la fragilité de l'interprétation des résultats donnée par les auteurs à partir d'aussi faibles effectifs. Enfin, le CS du HCB montre que la méthodologie statistique employée par les auteurs pour l'analyse des paramètres biochimiques est inadéquate et ne permet pas de conclure à l'existence de différences statistiquement significatives entre les groupes ayant consommé du maïs NK603 et les groupes témoin.
- les auteurs de l'article donnent des interprétations spéculatives de leurs résultats. Le CS du HCB ayant conclu que ces résultats ne mettent en évidence aucune différence statistiquement significative de mortalité, tumorigénicité ou de paramètres biochimiques entre groupes expérimentaux et groupes témoin, il n'a pas jugé utile de commenter l'ensemble des points de la discussion. Le CS du HCB a néanmoins pointé les lacunes réhivitoires du raisonnement des auteurs sur l'hormono-sensibilité des tumeurs, les faiblesses du raisonnement sous-tendant l'hypothèse de perturbations endocriniennes

³ Ce résumé ne se substitue pas à l'analyse de l'article développée dans cet avis.

causées par une modification des teneurs en acides phénoliques du maïs NK603, et les lacunes et incohérences empêchant de valider l'hypothèse des auteurs concernant d'éventuelles atteintes rénales chez les rats ayant consommé du maïs NK603.

En réponse à la saisine, le CS du HCB conclut donc que l'article de Séralini *et al.* (2012) ne présente pas de résultats concluants quant à une éventuelle toxicité du maïs NK603, traité ou non avec du Roundup[®].

TABLE DES MATIERES

1. SAISINE MINISTERIELLE SUITE AUX REVENDEICATIONS D'UNE PUBLICATION SCIENTIFIQUE SUR L'IMPACT SANITAIRE D'UN OGM ET D'UN HERBICIDE.....	5
1.1. L'ARTICLE SERALINI ET AL. (2012).....	5
1.2. SAISINE CONJOINTE DE L'ANSES ET DU HCB PAR QUATRE MINISTERES.....	5
1.3. ORGANISATION DE LA REPOSE DU HCB A LA SAISINE.....	5
1.4. ELEMENTS DE CONTEXTE	6
2. ANALYSE DE L'ETUDE RAPPORTEE PAR L'ARTICLE SERALINI ET AL. (2012)	7
2.1. REVENDEICATIONS DE L'ARTICLE	7
2.2. DONNEES ANALYSEES.....	8
2.3. UN DISPOSITIF EXPERIMENTAL INADAPTE AUX OBJECTIFS	8
2.4. UNE PRESENTATION DES RESULTATS PARCELLAIRE ET IMPRECISE.....	10
2.5. DES CONCLUSIONS NON JUSTIFIEES.....	12
2.6. DES INTERPRETATIONS SPECULATIVES	16
3. CONCLUSIONS ET REPOSE DU HCB A LA PREMIERE PARTIE DE LA SAISINE.....	17
4. BIBLIOGRAPHIE	18
ANNEXE 1 : SAISINE	20
ANNEXE 2 : ELABORATION DE L'AVIS.....	22
ANNEXE 3 : ASPECTS STATISTIQUES DU PROTOCOLE EXPERIMENTAL	23
ANNEXE 4 : ASPECTS STATISTIQUES DE L'ANALYSE DES RESULTATS.....	25

1. Saisine ministérielle suite aux revendications d'une publication scientifique sur l'impact sanitaire d'un OGM et d'un herbicide

1.1. L'article Séralini et al. (2012)

Au terme d'une étude de deux ans chez le rat, l'équipe du Professeur Séralini revendique la première démonstration expérimentale documentée d'une toxicité sur le long terme de la consommation d'une plante génétiquement modifiée (GM) et d'un herbicide Roundup[®], dans un article publié en ligne le 19 septembre 2012, sous presse dans la revue scientifique *Food and Chemical Toxicology* (Séralini et al., 2012).

1.2. Saisine conjointe de l'Anses et du HCB par quatre ministères

Considérant la portée des revendications de cette publication, les ministres en charge de la santé, de l'agriculture, de l'écologie, ainsi que le ministre délégué auprès du ministre de l'économie chargé de la consommation, ont conjointement saisi pour avis l'Anses (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail) et le HCB (Haut Conseil des biotechnologies).

Plus précisément, il est demandé à l'Anses et au HCB de « *procéder à une analyse de l'étude rapportée par cet article afin de déterminer si elle est de nature à remettre en cause ou non les conclusions des évaluations précédentes sur cet OGM et notamment si elle peut être considérée comme conclusive quant au risque sanitaire que pourraient présenter les aliments issues de plantes OGM comportant l'événement NK603.* »

De par sa responsabilité dans l'évaluation des formulations phytopharmaceutiques en vue de l'autorisation de leur mise sur le marché, il est d'autre part demandé à l'Anses de « *déterminer si cette étude est de nature à remettre en cause ou non les conclusions des évaluations précédentes de l'Anses sur l'herbicide Roundup* ».

Enfin, il est demandé aux deux organisations d'« *évaluer si le protocole mis en œuvre et les conclusions de cette étude remettent en cause les lignes directrices actuelles ou à venir en matière d'évaluation des risques sanitaires* ».

Les autorités publiques prennent le soin de distinguer l'analyse de l'article lui-même, de la réflexion sur les lignes directrices d'évaluation des risques sanitaires, en définissant un calendrier de travail en deux temps : « *Nous vous saurions gré de bien vouloir rendre un avis sur cet article avant le 20 octobre 2012 et sur la pertinence des modalités d'évaluation des risques sanitaires et des propositions d'aménagements des lignes directrices, si nécessaire, avant le 20 novembre 2012.* »

La saisine est reproduite en Annexe 1.

1.3. Organisation de la réponse du HCB à la saisine

En réponse à cette saisine conjointe, l'Anses et le HCB se sont concertés et ont établi des plans de travail parallèles, avec des groupes de travail distincts permettant de faciliter l'organisation de l'expertise tout en maintenant des ponts de communication sur la partie commune de la saisine. Un expert commun aux deux groupes de travail a facilité les échanges entre les deux organismes. Une réunion finale de présentation des travaux respectifs s'est tenue le 17 octobre 2012.

Conformément à la saisine, le HCB a établi un plan de travail en deux temps, prévoyant de rendre un premier avis ciblé sur l'analyse de l'article de Séralini et al. (2012) et ses implications pour le 20 octobre 2012, et un second avis sur une réflexion relative aux lignes directrices d'évaluation sanitaire des OGM dans un deuxième temps.

Plus précisément, cet avis du Comité scientifique (CS) du HCB vise à déterminer si l'article présente des résultats concluants quant à une éventuelle toxicité du maïs NK603. Les conclusions spécifiques à la toxicité des herbicides Roundup[®], qui ne relèvent pas des

missions et domaines de compétences du HCB, sont examinées par l'Anses, conformément à la saisine.

Le CS du HCB a mis en place un groupe de travail pluridisciplinaire constitué de quatre experts externes, trois experts internes, et du Président et Vice-président du CS (Voir Annexe 2). Les experts ont été choisis pour leurs compétences dans les disciplines pertinentes pour l'analyse de l'article (toxicologie, cancérologie, santé humaine et animale, statistiques, physiologie végétale), leur appartenance au secteur public et leur indépendance, vérifiée par déclaration d'intérêts. Un expert supplémentaire du secteur privé (Centre International de Toxicologie) a été consulté sur des questions techniques par le groupe de travail. Les experts externes sélectionnés n'ont jamais travaillé sur des OGM ou évalué des OGM dans le passé, et ne se sont pas exprimés publiquement sur cette étude.

Le groupe de travail s'est réuni le 27 septembre, les 2 et 5 octobre 2012, et a poursuivi ses échanges par voie électronique. L'article de Séralini *et al.* (2012) et l'analyse du groupe de travail ont été présentés le 2 octobre 2012 au CS du HCB.

Le Professeur Séralini et trois co-auteurs étudiants en thèse (Robin Mesnage, Steeve Gress et Nicolas Defarge) ont été auditionnés par le HCB le 10 octobre 2012.

Un avis a été élaboré à partir des rapports des experts du groupe de travail et des réflexions supplémentaires des membres du CS. L'avis a été révisé par le groupe de travail, et adopté par les membres du CS par voie électronique le 19 octobre 2012.

1.4. Eléments de contexte

Les études de long terme sur la toxicité des plantes GM (PGM) sont peu nombreuses. Les revues de littérature de Domingo et Bordonaba (2011) et de Snell *et al.* (2012) n'identifient notamment que deux études de deux ans sur rongeurs (Domingo and Bordonaba, 2011; Snell *et al.*, 2012). Les auteurs de l'une de ces études concluent à l'absence d'effet détectable de la consommation d'un régime de 30 % d'un soja GM sur deux ans chez le rat (étude réalisée sur des groupes de 50 rats) (Sakamoto *et al.*, 2008) ; les auteurs de la seconde étude concluent que la consommation d'un régime de 14 % d'un soja GM sur deux ans chez la souris pourrait avoir un effet sur le vieillissement hépatique (étude réalisée sur des groupes de 10 souris) (Malatesta *et al.*, 2008).

Les études de toxicité réalisées dans le cadre de l'évaluation des effets sanitaires des PGM pour leur mise sur le marché présentent des limites. Ceci a été rappelé à plusieurs reprises par le CS du HCB⁴. Le CS du HCB rappelle que l'absence d'un risque ne peut être formulée sans y associer une probabilité d'erreur, qui nécessite une étude de puissance (HCB, 2012).

Concernant les PGM tolérantes à un herbicide (PGM TH), les études de toxicité portent souvent sur des PGM qui n'ont pas été traitées lors de leur culture avec l'herbicide correspondant. L'EFSA⁵ a publié en 2011 de nouvelles lignes directrices qui recommandent que l'évaluation de la toxicité des PGM TH inclue l'évaluation de PGM TH traitées avec l'herbicide auquel elles sont tolérantes (EFSA, 2011). Ces lignes directrices sont en cours de transcription en document contraignant par la Commission européenne.

Les études de toxicité réglementaires sont actuellement réalisées, ou sous-traitées, par les pétitionnaires dans le cadre de la Directive 2001/18/CE⁶ (EC, 2001) et du Règlement (CE)

⁴ Par exemple, dans son dernier avis du 31 juillet 2012, le CS du HCB souligne que le pétitionnaire conclut à l'absence d'effet toxique majeur du maïs GA21 sur la santé à partir d'une étude de toxicité orale de 90 jours chez le rat, sans avoir réalisé d'étude de puissance statistique. Cette conclusion va au-delà de ce qu'il est possible d'interpréter à partir de leurs résultats (HCB, 2012).

⁵ EFSA : *European Food Safety Authority* (Autorité européenne de sécurité des aliments).

⁶ La Directive 2001/18/CE est une directive du Parlement européen et du Conseil du 12 mars 2001 qui fixe les règles communautaires relatives à la dissémination volontaire d'OGM dans l'environnement. Elle abroge la directive 90/220/CEE du Conseil. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0018:FR:HTML>.

1829/2003⁷ (EC, 2003), qui ne prévoient pas de règles particulières d'indépendance dans la réalisation de ces études.

Plusieurs autres aspects de l'évaluation du potentiel toxique ne sont souvent pas pris en compte, et limitent donc l'interprétation que l'on peut faire de ces évaluations. Par exemple, la sensibilité de détection d'effets toxiques est variable selon les espèces, et au sein d'une même espèce. Cette sensibilité est également affectée par les conditions de stress dans lesquelles se trouve l'animal. Il n'existe pas de critère réglementaire normalisé sur ces aspects.

Enfin, comme le souligne une note de 2009⁸ de la Commission européenne sur le protocole 408 de l'OCDE relatif aux études de toxicité sur 90 jours chez les rongeurs (OECD, 1998), il n'existe pas de consensus quant aux tests toxicologiques en matière de nouveaux aliments / nouveaux ingrédients, en particulier pour des aliments complets.

2. Analyse de l'étude rapportée par l'article Séralini *et al.* (2012)

2.1. Revendications de l'article

L'article publié dans la revue scientifique *Food and Chemical Toxicology* (Séralini *et al.*, 2012), rapporte les résultats des effets sanitaires à long terme chez le rat, de la consommation d'un régime enrichi en maïs génétiquement modifié (GM) – le maïs NK603⁹, génétiquement modifié pour être tolérant au glyphosate –, et d'une eau de boisson contenant une formulation herbicide à base de glyphosate, commercialisée sous le nom de Roundup[®].

Durant deux ans, les effets sanitaires de trois types de régimes ont été suivis en parallèle, chez des groupes de 10 rats Sprague-Dawley (SD) de chaque sexe :

1. des régimes contenant du maïs GM dans trois proportions différentes : 11, 22 et 33 % ;
2. des régimes contenant du maïs GM, traité par l'herbicide Roundup[®] (WeatherMax)¹⁰ lors de sa culture, dans les mêmes trois proportions de 11, 22 et 33 % ;
3. des régimes contenant 33 % de maïs non GM quasi isogénique au maïs NK603 et une eau de boisson supplémentée d'herbicide Roundup[®] (GT Plus)¹¹, dans trois proportions différentes : 1,1 x 10⁻⁸, 0,09 et 0,5 %.

⁷ Le Règlement (CE) 1829/2003 est un règlement du Parlement européen et du Conseil du 22 septembre 2003 concernant les denrées alimentaires et les aliments pour animaux génétiquement modifiés. (Plus précisément, pour clarifier une confusion inhérente à la traduction française de ce titre, ce règlement concerne les denrées alimentaires et les aliments pour animaux, ces denrées alimentaires ou aliments pouvant consister en des OGM, contenir des OGM, ou être issus d'OGM.) : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003R1829:FR:HTML>.

⁸ EC position paper on the document (ENV/JM(2009)4): proposal to adapt OECD test guidelines No. 408 "repeated dose 90-day toxicity study in rodents" for whole-food testing. 44th Joint meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology for its meeting on 10-11 June 2009.

⁹ Le maïs génétiquement modifié NK603 exprime l'enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) de la souche CP4 d'*Agrobacterium tumefaciens* (CP4 EPSPS), qui lui confère la tolérance au glyphosate, ingrédient actif d'herbicides non sélectifs comme le Roundup[®]. La toxicité non sélective du glyphosate pour les plantes s'explique par le fait qu'il inhibe la fonction de l'enzyme EPSPS de la majorité des plantes. EPSPS est une enzyme essentielle à la production des acides aminés et autres composés aromatiques chez les plantes, les bactéries et les champignons. Elle n'est pas présente chez les animaux, qui ne synthétisent pas leurs propres composés aromatiques. Plusieurs stratégies ont été déployées pour développer des plantes résistantes au glyphosate. La stratégie la plus utilisée actuellement est l'utilisation du gène *cp4 epsps* de la souche CP4 d'Agrobactérie. Ce gène possède une mutation qui rend l'enzyme produite, CP4 EPSPS, insensible à l'inhibition par le glyphosate. Il s'agit donc de rajouter une ou plusieurs copies de ce gène dans les plantes pour maintenir l'activité de la voie métabolique des composés aromatiques tandis que l'enzyme végétale EPSPS endogène est inhibée par le glyphosate (Duke and Powles, 2008; Funke *et al.*, 2006).

¹⁰ La formulation herbicide Roundup[®] WeatherMax contient 540 g/l de glyphosate. Cet herbicide est utilisé à la dose de 3 litres par hectare.

¹¹ La formulation herbicide Roundup[®] GT Plus contient 450g/l de glyphosate.

Pour chaque sexe, ces 9 groupes de rats (groupes dits « expérimentaux » dans cet avis) ont été suivis en comparaison à un même groupe témoin, nourri avec un régime contenant 33 % de maïs non GM quasi isogénique au maïs NK603 et une eau non additionnée de Roundup®.

Selon les auteurs, les résultats montreraient :

- une mortalité plus élevée et plus précoce pour tous les groupes de femelles expérimentaux, et pour trois des six groupes mâles nourris avec du maïs GM ;
- des tumeurs mammaires « globalement » (cf. texte de l'article « *almost always more often than* ») plus fréquentes et plus précoces que chez les groupes témoin, et une atteinte de l'hypophyse chez les femelles ; l'équilibre hormonal de chaque sexe serait modifié par le maïs GM et le Roundup®. Des tumeurs palpables quatre fois plus nombreuses apparaîtraient 600 jours plus tôt chez les mâles des groupes expérimentaux ;
- des lésions hépatiques (congestion, nécrose) en microscopie optique et électronique et des néphropathies plus fréquentes chez les mâles des groupes expérimentaux. L'analyse biochimique corroborerait ces observations.

Dans leurs conclusions, les auteurs indiquent que ces résultats s'expliqueraient par des effets perturbateurs endocriniens non linéairement liés aux doses de Roundup®, et par l'expression du transgène dans le maïs NK603 et ses conséquences métaboliques.

2.2. Données analysées

Tel que publié, l'article présente un ensemble de faiblesses dont certaines – imprécisions et lacunes dans la description du dispositif expérimental, présentation partielle des résultats, utilisation de nomenclatures non conventionnelles sans justification – auraient pu être corrigées par l'apport de données complémentaires par les auteurs de l'article.

En contradiction avec les usages qui prévalent en matière scientifique et malgré leur engagement à suivre les instructions de la revue¹², les auteurs de l'article ont refusé de transmettre les compléments d'information demandés par le HCB à cet effet¹³.

Le CS du HCB note toutefois que même si ces données complémentaires auraient pu être utiles pour éclairer certains résultats, elles n'étaient pas nécessaires pour répondre aux questions posées par la saisine.

En soutien à son analyse des données rapportées dans l'article, le CS du HCB a pu se procurer des données de référence de la société Harlan, fournisseur des rats SD de l'étude¹⁴.

2.3. Un dispositif expérimental inadapté aux objectifs

Des objectifs confus :

Le dispositif expérimental et le plan d'analyse d'une étude sont normalement conçus pour permettre de répondre à des questions précises. Les objectifs de cette étude sont confus. L'étude semble avoir été initialement conçue pour explorer des perturbations biologiques pouvant résulter de la consommation du maïs GM ou de l'herbicide Roundup® sur le long terme sans cible précise, comme en témoigne la multiplicité des constantes mesurées : 31 paramètres sanguins et 16 paramètres urinaires analysés à 11 dates réparties du début à la

¹² Ethique de publication des journaux Elsevier, dont la revue Food and Chemical Toxicology fait partie : "Data access and retention: Authors may be asked to provide raw data in connection with a paper for editorial review, and should be prepared to provide public access to such data (consistent with the ALPSP-STM Statement on Data and Databases), if practicable, and should in any event be prepared to retain such data for a reasonable time after publication."

Instruction spécifique aux auteurs de la revue Food and Chemical toxicology : "Furthermore, it is understood that with submission of this article, the authors (...) are willing to share the original data and materials if so requested."

¹³ Courrier envoyé au Professeur Séralini le 2 octobre 2012.

¹⁴ Courriers envoyés à la société Harlan le 20 septembre et 12 octobre 2012 ; données de la société réceptionnées le 27 septembre et le 16 octobre 2012.

fin de l'expérience, auxquels s'ajoutent 6 paramètres biologiques spécifiques de la fonction hépatique, mesurés une seule fois en fin d'expérience. Cependant, une grande partie de l'article concerne l'apparition de tumeurs chez les rats, qui semble avoir attiré l'attention des auteurs en cours d'expérience.

Un nombre de rats par groupe insuffisant pour répondre aux questions posées :

L'effectif des groupes de rats se prévoit en fonction de la taille de l'effet biologiquement significatif que l'on souhaite détecter, et selon les caractéristiques de la souche de rat utilisée pour les paramètres mesurés sur la durée de l'étude et dans les conditions expérimentales utilisées. Dans cet article, il n'est fait mention d'aucun calcul préalable du nombre de sujets nécessaires pour mettre en évidence un effet jugé biologiquement significatif pour une étude à deux ans.

Selon ces considérations, les effectifs recommandés dans les lignes directrices de l'OCDE ont été estimés à 20 rats par groupe pour une étude de toxicité chronique sur 12 mois¹⁵ [Test guideline n° 452 (OECD, 2009b)], et 50 rats par groupe pour une étude de cancérogénicité sur 24 mois [Test guideline n° 451 (OECD, 2009a)] ou pour une étude combinée de toxicité chronique et de cancérogénicité [Test guideline n° 453 (OECD, 2009c)]. Avec 10 rats par groupe, cette étude se situe en deçà des recommandations estimées nécessaires pour pouvoir conclure à des effets statistiquement significatifs des traitements à long terme pour les deux types d'analyses effectuées (toxicité chronique, tumorigénicité).

De plus, la souche de rats SD, communément utilisée dans le cadre des études toxicologiques sub-chroniques de 90 jours, est connue pour développer spontanément sur le long terme des tumeurs de la glande mammaire et de l'hypophyse. La fréquence des fibroadénomes des glandes mammaires chez les rats SD du fournisseur Harlan peut naturellement s'élever jusqu'à 70 % (Brix *et al.*, 2005). Les données obtenues de la société Harlan sur des rats SD provenant du même élevage que celui des animaux de l'étude Seralini *et al.* (2012) indiquent une fréquence de 60 % de tumeurs des glandes mammaires dans une étude interne de deux ans, débutée en 2009 (communication de données de référence par la société Harlan). Cette information doit impérativement être prise en compte dans la conception d'un dispositif expérimental au long terme, car plus la fréquence naturelle de développement de tumeurs est élevée, plus il faudra d'animaux dans les groupes expérimentaux pour mettre en évidence une augmentation significative du nombre de tumeurs en lien avec le régime (voir Annexe 3). Ces caractéristiques de la souche de rats SD ne sont ni prises en compte dans le dispositif expérimental de cette étude, ni considérées dans l'interprétation et la discussion des résultats.

Des groupes témoin trop peu nombreux et manquant de pertinence biologique :

Pour chaque sexe, un seul groupe témoin de 10 rats a été constitué. Ce même groupe de 10 rats est systématiquement utilisé pour être comparé aux 9 groupes expérimentaux. Il en résulte un tel manque de puissance statistique qu'il est dès lors très difficile d'établir si les différences observées lors de chacune des 9 comparaisons sont dues à un effet du régime, ou bien simplement à une variabilité naturelle du groupe témoin. Ainsi, si un paramètre est particulièrement élevé dans le groupe témoin du fait du hasard de la constitution des groupes, toutes les différences entre groupes expérimentaux et groupe témoin iront dans le même sens et présenteront toutes une diminution du caractère considéré, sans que l'on puisse pour autant en conclure à un effet du régime.

Enfin, la définition des groupes est imprécise : l'article n'indique pas si les régimes sont constitués de telle sorte que les apports totaux en maïs soient égalisés à 33 % de maïs avec du maïs non GM. Si ce n'est pas le cas, chaque groupe expérimental nourri avec un régime d'une proportion donnée de maïs GM devrait être comparé à un groupe témoin nourri avec un

¹⁵ Le protocole 452 de l'OCDE pour les études de toxicité chronique précise qu'une période de 12 mois est suffisamment longue pour détecter une éventuelle toxicité cumulative, tout en évitant les effets liés au vieillissement des rats, qui pourraient prêter à confusion. p. 3 *Principle of the test*: "This duration is chosen to be sufficiently long to allow any effects of cumulative toxicity to become manifest, without the confounding effects of geriatric changes." (OCDE, 2009a).

régime de même proportion de maïs non GM. Enfin, Il manque un lot témoin nourri avec un régime standard de rat (sans addition de maïs).

Ce dispositif expérimental ne permet donc pas de déduire des observations faites au cours de cette étude des relations de causalité entre les traitements et les effets rapportés, notamment concernant la tumorigénicité.

2.4. Une présentation des résultats parcellaire et imprécise

Bien que les faiblesses du dispositif expérimental limitent la portée des conclusions de cette étude, le CS du HCB a procédé à l'analyse complète des résultats présentés dans l'article pour déterminer les conclusions qui pourraient néanmoins en être tirées.

Les résultats de l'étude sont présentés sur un mode essentiellement descriptif. A part les données biochimiques, qui bénéficient d'un traitement statistique particulier, les données rapportées ne font l'objet d'aucune analyse statistique. Une telle démarche pourrait être acceptable si les auteurs se contentaient d'une description, et si cette description couvrait l'ensemble des données obtenues. Or, seuls certains résultats sont sélectionnés, présentés et commentés, et de cette description parcellaire sont tirées des conclusions non justifiées, utilisées ensuite pour échafauder des hypothèses physiopathologiques non fondées.

Plus précisément, la description des résultats souffre de :

1. Lacunes de présentation de certaines données nécessaires en toxicologie :

Les données requises pour l'interprétation de l'analyse toxicologique, comme les contrôles de composition et la présence de contaminants dans les régimes, les apports nutritionnels et la croissance pondérale des animaux, ne sont pas présentées dans l'article. L'équilibre énergétique des différents régimes et les données de consommation sont d'autant plus importants que les incidences tumorales peuvent varier selon les apports nutritionnels (Keenan *et al.*, 1997).

2. Sélection injustifiée des résultats présentés :

Dans un contexte où les résultats ne sont pas appuyés par l'analyse statistique, choisir d'isoler des résultats dans la présentation d'un travail expérimental multi-factoriel conduit à donner une dimension à ces analyses qui n'a aucune valeur scientifique. Sans justification, les auteurs ont choisi de présenter les résultats de quatre paramètres biochimiques et deux paramètres hormonaux qui présenteraient la plus forte variation par rapport au témoin (Figure 5B). Ce choix a été effectué *a posteriori*. On peut évidemment s'attendre à ce qu'il y ait des différences parmi les 864¹⁶ comparaisons effectuées par les auteurs pour les valeurs des 48 paramètres biochimiques au 15^{ème} mois de l'étude. La présentation de résultats sélectionnés peut alors être trompeuse pour un lecteur non spécialiste des comparaisons multiples, qui conclura à tort que les différences observées sont représentatives de différences entre groupes expérimentaux et groupes témoin.

De même, la présentation de photos n'a de valeur autre qu'illustrative que si la sélection des éléments présentés est justifiée, équilibrée et représentative pour chaque groupe de l'étude. Les photos de la Figure 3 illustrent des différences cytologiques objectives. En l'absence de données chiffrées d'occurrence des anomalies, les différences potentielles entre les groupes ne peuvent être appréciées : aucune information ne peut être tirée de la simple visualisation d'un parenchyme normal et d'un parenchyme pathologique. De la même façon pour la microscopie électronique, le choix et la représentativité des éléments présentés en comparaison sont questionnables (Figure 4). Enfin, au-delà d'un impact empathique pour les animaux, aucune information ne peut être tirée de photos de tumeurs de rats expérimentaux en l'absence de photos de tumeurs de rats des groupes témoin.

¹⁶ 864 est le nombre de comparaisons effectuées entre les données obtenues pour les lots traités au 15^{ème} mois de l'étude (18 lots expérimentaux x 48 paramètres biochimiques = 864) et les données correspondantes des groupes témoin.

3. Sélection des éléments commentés dans le texte :

Les résultats de mortalité sont commentés de manière sélective : “Before this period, 30% control males (three in total) and 20% females (only two) died spontaneously, while up to 50% males and 70% females died in some groups on diets containing the GM maize (Fig. 1).”¹⁷ Le panel supérieur de la Figure 1 de l'article est reproduit dans la Figure A ci-dessous pour le cas des mâles. A la simple analyse de l'histogramme, on note que si le groupe de mâles nourris avec 11 % de maïs GM présente un taux de mortalité de « 50% » (5 rats morts) à 600 jours d'étude, les groupes de mâles ayant été nourris avec des doses supérieures de maïs GM présentent des taux de mortalités de seulement « 10 % » (1 rat mort), soit un taux de mortalité inférieur à celui du groupe témoin. Ces résultats ne sont pas mentionnés. D'autre part, choisir d'extraire un taux de mortalité à 600 jours est arbitraire (voir section 2.5, Analyse de survie) : le taux de mortalité chez les mâles du groupe témoin passe de 30 % à 600 jours, à 50 % quelques jours après (courbe pointillée). Ce simple exemple montre que des effets différents, voire même des tendances opposées, peuvent aisément être mis en évidence parmi les résultats obtenus, et qu'une analyse statistique est indispensable pour valider les conclusions que les observations seraient susceptibles de suggérer.

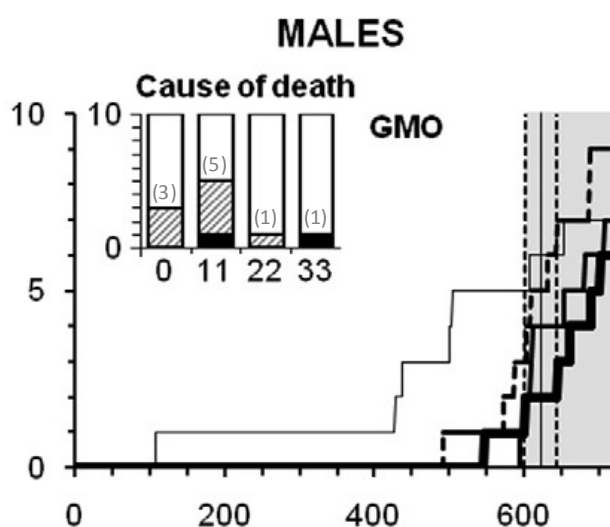


Fig. A. Panel extrait de la Figure 1 de l'article (Séralini et al., 2012), représentant la mortalité des rats mâles nourris avec des régimes contenant du maïs GM à 11, 22 et 33 % (lignes mince, moyenne et grasse) par rapport au témoin (ligne en pointillé). L'histogramme représente la mortalité des rats à ~600 jours d'étude. En noir sont représentés les euthanasies et en hachuré, les morts spontanées. Les chiffres ajoutés à l'histogramme entre parenthèses correspondent aux nombres d'animaux morts par groupe, que l'on peut déduire à partir de l'échelle graduée en ordonnées.

Les résultats de tumorigénicité sont également commentés de manière anecdotique, mettant par exemple l'accent sur deux tumeurs de Wilm (néphroblastomes) chez les mâles sans qu'il ne soit possible de rien en conclure : “It is noteworthy that the first two male rats that died in both GM treated groups had to be euthanized due to kidney Wilm's tumors that were over 25% of body weight. This was at approximately a year before the first control animal died.”¹⁸. Aucun test statistique n'est effectué pour comparer la fréquence des tumeurs entre groupes de rats.

¹⁷ Traduction du texte en français : « Avant cette période, 30% des mâles témoin (trois en tout), et 20% des femelles (seulement deux) sont morts de façon spontanée, alors que jusqu'à 50% de mâles et 70% de femelles sont morts dans certains groupes nourris avec un régime contenant du maïs GM. »

¹⁸ Traduction : « Il est notable que les deux premiers morts des deux groupes nourris avec du maïs GM ont dû être euthanasiés à cause de tumeurs de Wilm qui représentaient plus de 25 % du poids du corps des animaux. Ceci est arrivé approximativement un an avant que le premier animal du groupe témoin ne meurt. »

4. Présentation imprécise, sans pertinence biologique, avec l'utilisation d'une « nomenclature » non conventionnelle :

La présentation des résultats de mortalité et de tumorigénicité en graphes est imprécise et prête à confusion : les graphes de mortalité (Figure 1) ne permettent pas de comparer les espérances de vie selon les régimes ; les graphes décrivant la tumorigénèse (Figure 2) rapportent notamment des nombres de « tumeurs palpables », une entité sans support nosologique et ne permettant de tirer aucune conclusion d'ordre statistique, mécanistique ou étiologique.

Aucune logique biologique n'apparaît dans la description des paramètres biochimiques, un seul tableau (Table 3) classant les résultats sous la forme d'« *increase* »¹⁹ ou de « *decrease* »²⁰ des paramètres tout en y mêlant des valeurs négatives et positives.

L'analyse des pathologies néoplasiques et non néoplasiques souffre d'insuffisances descriptives. Les anomalies histologiques et macroscopiques ne sont pas clairement identifiées ; aucune distinction n'est faite entre tumeurs bénignes, tumeurs malignes et autres lésions des organes et tissus (Table 2). Il serait impossible de conclure à une différence significative entre les groupes expérimentaux et les groupes témoin en se fondant uniquement sur les données composites exposées dans cette publication. Par ailleurs, les auteurs n'indiquent pas les modalités de découverte des anomalies (tumeur létale, tumeur diagnostiquée chez un animal décédé d'une autre cause, tumeur détectée lors de l'examen systématique pratiqué au sacrifice terminal) (Gart *et al.*, 1986).

Le tableau décrivant la fréquence des pathologies relevées dans les différents groupes (Table 2) utilise une « nomenclature » propre aux auteurs qui, en négligeant la nomenclature conventionnelle d'anatomie pathologique et faute d'explication, n'est pas informative pour les spécialistes du domaine (cf. Table 2, Colonne 1, ligne 1 : « *Males in liver* »²¹, ligne 2 : « *In hepatodigestive tract* »²², cette entité est inconnue des spécialistes, le foie en fait-il partie, et si c'est le cas, les anomalies hépatiques seraient-elles comptées deux fois ?).

2.5. Des conclusions non justifiées

L'analyse des résultats devrait permettre de passer de la description des données obtenues dans l'étude à des conclusions portant sur des différences d'effet entre traitements, et donc éventuellement sur des relations de causalité entre un traitement donné et un effet observé. Des conclusions de cette nature sont difficilement justifiables *a priori* dans cet article du fait non seulement de l'inadéquation du dispositif expérimental aux questions posées – notamment concernant la tumorigénicité –, mais également de l'absence d'analyse statistique appropriée des données.

Le CS du HCB a appliqué des méthodologies statistiques standard pour analyser les résultats présentés dans l'article, et ainsi mieux éclairer les conclusions qui peuvent en être légitimement tirées. Cette analyse est détaillée dans l'Annexe 4 ; les points clefs en sont repris ici :

1. Analyse de survie

La durée de vie des animaux n'est pas correctement évaluée par les auteurs de l'article. §3.1 : « *Control male animals survived on average 624 ± 21 days, whilst females lived for 701 ± 20* »²³. Ces valeurs sont inexactes, issues d'un calcul incorrect puisque les auteurs n'ont pas pris en compte le phénomène de censure des données (résultant entre autres du sacrifice des animaux en fin d'étude). En particulier, les précisions annoncées de 20 et 21 jours sous-estiment très largement la précision avec laquelle on peut estimer les espérances de vie des

¹⁹ Traduction : « augmentation ».

²⁰ Traduction : « diminution ».

²¹ Traduction : « mâles dans le foie ».

²² Traduction : « dans l'appareil hépatodigestif ».

²³ Traduction : « Les animaux témoins mâles ont survécu en moyenne 624 ± 21 jours, alors que les femelles ont vécu 701 ± 20 ».

rats à partir de ces données. Il en découle que la position des lignes verticales et zones grisées apparaissant dans la Figure 1 de l'article, et que l'on peut voir dans la Figure A de cet avis, est erronée.

L'examen des courbes de mortalité (Figure 1) ne permet pas de conclure à une différence entre groupes de rats. Il faudrait mettre en œuvre une procédure statistique qui prenne en compte la variabilité des individus et donc les courbes de survie. On peut néanmoins tester si les animaux d'un groupe expérimental donné ont tendance à mourir plus tôt que ceux du groupe témoin. Pour l'étude présentée, il faudrait alors effectuer 18 tests (9 conditions mâles et 9 conditions femelles à comparer à leur unique témoin). Par exemple : H_0 « le régime contenant 11 % de maïs GM n'a pas d'effet sur la durée de vie des rats femelles » vs H_1 « le régime contenant 11 % de maïs GM entraîne une diminution de la durée de vie des rats femelles ».

La statistique de test, pour un test de rang, est définie comme la somme des rangs du groupe témoin. La Figure B ci-dessous montre les statistiques de tests calculées à partir des données de l'article et classées par ordre décroissant ainsi que les intervalles de prévision obtenus sous l'hypothèse nulle, mais qui prend correctement en compte la multiplicité des comparaisons effectuées. Les 18 statistiques de tests sont toutes à l'intérieur des intervalles de prévision de niveau 90 % correspondants : on ne peut donc pas conclure à un effet statistiquement significatif des différents régimes sur la durée de vie des rats.

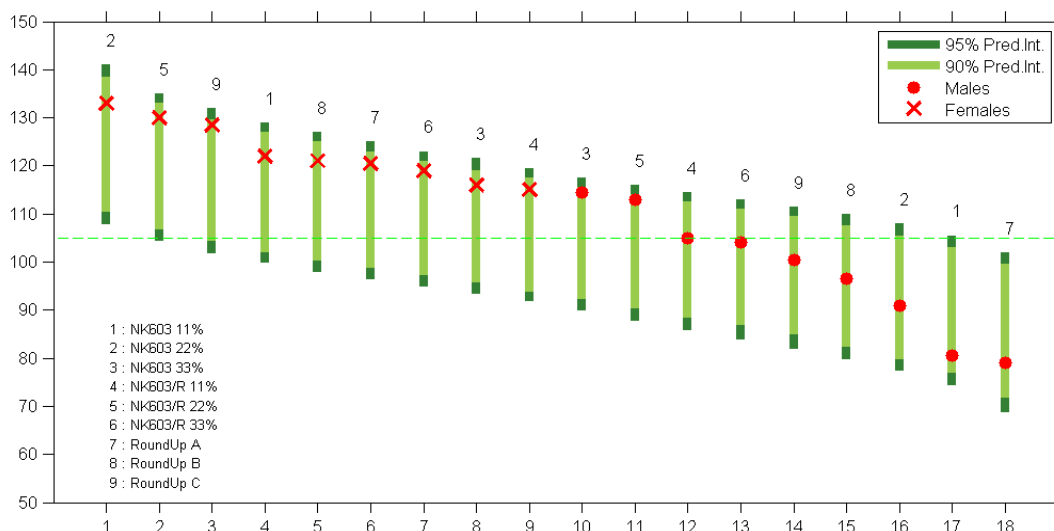


Fig. B. Intervalles de prévision de niveau 90% et 95% des 18 statistiques de tests par ordre décroissant et valeurs observées de ces statistiques de tests pour la durée de vie des rats.

Le manque de puissance statistique, dû à un très faible effectif des groupes témoin, interdit de conclure de façon formelle à la présence ou à l'absence d'un effet du régime sur la mortalité, en particulier chez les femelles. Ce manque de puissance peut être compensé par l'introduction d'informations *a priori* sur le comportement attendu des groupes témoin. Des données de mortalité de la souche SD utilisée dans cette étude ont été obtenues de la société Harlan²⁴, et peuvent compléter l'information apportée par l'étude relatée dans cet article.

L'utilisation des données de référence fournies par l'éleveur confirme que l'on ne peut expliquer les différences observées dans les courbes de survie entre groupes expérimentaux et groupes témoin par un effet du régime : la Figure C montre que les groupes expérimentaux sont globalement distribués à l'intérieur de ces intervalles de prévision. Un groupe expérimental se situe à la marge de l'intervalle : sur 18 groupes, il n'y a rien d'anormal à ce qu'une observation se situe hors d'un intervalle de prévision de niveau 95 %. Les taux de

²⁴ Données provenant d'une étude interne à la société Harlan, d'une durée de 24 mois à partir de 2009, réalisée sur des rats SD de même provenance que les rats utilisés dans l'étude de Séralini *et al.* (2012).

survie observés à deux ans pour les groupes expérimentaux sont donc tout à fait compatibles avec les données de référence fournies par l'éleveur pour les rats SD. En revanche, on peut remarquer que les taux de survie observés dans les groupes témoin sont assez éloignés de ce que les valeurs de référence laissent présager pour ce groupe (la proportion observée de rats femelles en vie après deux ans se situe hors de l'intervalle de prévision de niveau 95 %). Cela confirme la fragilité statistique de résultats obtenus à partir d'aussi faibles effectifs : on ne peut donc tirer des données de l'article aucune conclusion formelle quant à l'effet des différents régimes sur la survie des rats.

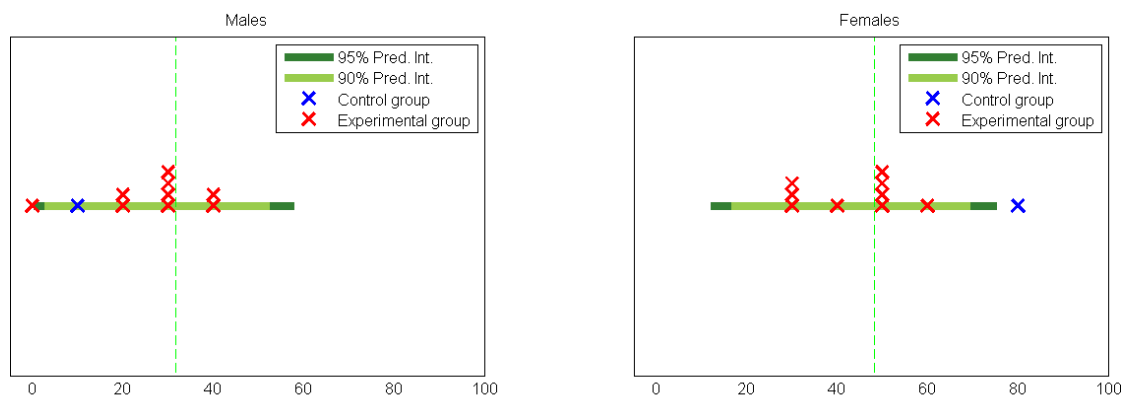


Fig.C. Intervalles de prévision de niveau 90% et 95% des taux de survie à 2 ans obtenus à partir des données de Harlan et taux de survie observés dans les groupes expérimentaux et les groupes témoin.

Enfin, un intervalle de prévision des courbes de survie de deux groupes expérimentaux regroupant l'ensemble des rats mâles et l'ensemble des rats femelles des groupes expérimentaux de l'étude²⁵, peut être construit par simulation. Les courbes de survie des groupes témoin sont à l'intérieur de ces intervalles de prévision (Voir Annexe 4). Il est donc impossible de conclure à une différence statistiquement significative entre la survie des rats témoin et des rats expérimentaux. Là encore, les données de référence fournies par la société Harlan viennent étayer cette conclusion puisque les taux de survie observés dans les groupes expérimentaux se situent dans les intervalles de prévision des taux de survie à deux ans (obtenus avec un échantillon de taille $n = 60$) (Voir Annexe 4).

On ne peut donc conclure à un effet statistiquement significatif d'aucun régime expérimental testé dans cette étude (maïs GM, maïs GM traité au Roundup®, Roundup®) sur la survie des rats.

2. Analyse de la tumorigénicité

Comme pour les courbes de survie, un simple examen des courbes présentées dans l'article (Figure 2) ne permet pas de conclure à une quelconque différence de tumorigénicité entre les groupes de rats. Il aurait été nécessaire de mettre en œuvre un test statistique prenant en compte la variabilité au cours du temps. Comme précédemment, le protocole de l'étude n'est pas adapté pour effectuer les 18 comparaisons proposées. Comme illustré dans l'annexe 4, les courbes de nombres de tumeurs des groupes témoin et expérimentaux se situent à l'intérieur des intervalles de prévision. On ne peut donc conclure à un effet statistiquement significatif du régime sur l'évolution du nombre de tumeurs.

De plus, comme mentionné plus haut, la notion de « tumeurs palpables » est sujette à interprétation et ne tient absolument pas compte de l'origine tissulaire des masses et ne pourrait donc sous-tendre aucun mécanisme physiopathologique explicatif d'un effet des différents régimes.

²⁵ Stratégie de regroupement des groupes permettant de mettre en œuvre des tests statistiques plus robustes et plus puissants, mise en œuvre et discutée dans l'Annexe 4.

3. Analyse des paramètres biochimiques

Quarante-sept paramètres biochimiques ont été mesurés et un paramètre biochimique a été calculé lors de l'étude. Pour chaque sexe, pour chaque condition expérimentale, une méthode de type OPLS-DA²⁶ est mise en œuvre par les auteurs pour discriminer le groupe témoin des groupes expérimentaux. La méthode OPLS-DA est largement utilisée en chimiométrie ou en génomique pour identifier un sous-ensemble de variables qui différencient au mieux différents sous-groupes. Elle est pertinente lorsque le nombre de variables explicatives est grand devant le nombre d'observations. De plus, la méthode OPLS-DA permet de construire un modèle prédictif, qui, pour un jeu de variables explicatives donné, fournit les probabilités d'appartenir à chacun des sous-groupes considérés.

Le choix de cette méthode et son utilisation par les auteurs appellent plusieurs commentaires (voir l'annexe pour l'ensemble des commentaires) :

1. Quelle que soit la méthode utilisée, il est important de valider le modèle obtenu (*i.e.* de s'assurer qu'il possède de bonnes propriétés prédictives) :
 - i) par un échantillon test, afin de s'assurer que le modèle ajusté sur un échantillon d'apprentissage conserve de bonnes propriétés prédictives sur de nouvelles données qui n'ont pas été utilisées pour la construction du modèle ;
 - ii) par des méthodes de validation croisée, en faisant jouer aux différentes données alternativement le rôle d'échantillon d'apprentissage et d'échantillon test.

Les auteurs de l'étude n'ont pas validé les modèles obtenus, ils ne peuvent donc être utilisés en prédiction.

2. L'utilisation de ces méthodes suppose implicitement une distribution symétrique des variables explicatives. Les paramètres biochimiques peuvent avoir une distribution asymétrique ; une transformation préalable est donc nécessaire. Rien n'indique que celle-ci a été réalisée.
3. Calculer des intervalles de confiance pour chaque paramètre n'est pas pertinent lorsque de nombreux paramètres sont utilisés. En effet, les éventuelles corrélations entre paramètres sont totalement ignorées.

Les auteurs ont également sélectionné les données présentées. Parmi les 18 comparaisons entre groupes expérimentaux et groupes témoin, seule celle du groupe des femelles nourries avec un régime de 33 % de maïs NK603, qui présenterait le plus de différences, est présentée. De plus, parmi les 48 paramètres biochimiques considérés, les 6 paramètres présentant, selon les auteurs, le plus de différences entre le groupe femelle NK603 (33 %) et le groupe témoin sont retenus. Il est attendu qu'en sélectionnant à la fois le groupe et les 6 paramètres qui présentent le plus de différences, des différences entre groupe expérimental et groupe témoin seront visibles.

Cette méthode ne permet pas d'expliquer les différences observées par la différence de régime administré. Il est également impossible de rejeter l'hypothèse que ce serait la variabilité naturelle (due aux fluctuations d'échantillonnage) et/ou le critère de sélection des paramètres (les paramètres présentant le plus de différences) qui expliqueraient les différences observées.

Comme mentionné plus haut, aucun mécanisme physiopathologique ne vient expliquer les observations faites. Dans l'exemple des atteintes rénales, le qualificatif de « *severe* »²⁷ qui s'applique à des lésions observées chez les animaux morts, sans précision de la cause ni de la date de la mort, n'est pas corroboré par les données du Tableau 3. Des variations de natrémie (1 à 7 %) ou de kaliémie sont physiologiques. La variation de 10 % de kaliémie observée pour un groupe peut sembler importante, mais elle n'est pas hors norme, sachant que la kaliémie normale oscille entre 4,1 et 4,9 mmol/L. De plus, l'absence d'anomalie de la protéinurie est inhabituelle pour une atteinte rénale. Enfin, les paramètres urinaires sont à utiliser avec précaution, du fait d'une variabilité importante du recueil des urines chez le rat.

²⁶ OPLS-DA: *Orthogonal Partial Least Squares Discriminant Analysis*.

²⁷ Traduction : « sévères ».

Au terme de l'analyse des résultats de l'article Séralini *et al.* (2012), le CS du HCB conclut que le dispositif expérimental et les outils statistiques utilisés souffrent d'importantes lacunes et faiblesses méthodologiques qui ne permettent pas de soutenir les conclusions avancées par les auteurs. Le CS du HCB montre qu'une analyse statistique rigoureuse des données de l'article ne met en évidence :

- aucune différence statistiquement significative de la mortalité des rats entre les groupes expérimentaux et témoin,
- aucune différence statistiquement significative des nombres de tumeurs entre les groupes expérimentaux et témoin.

De plus, il montre que la méthodologie statistique employée par les auteurs pour l'analyse des paramètres biochimiques est inadéquate et ne permet pas de conclure à l'existence de différences statistiquement significatives entre les groupes expérimentaux et témoin.

2.6. Des interprétations spéculatives

La discussion de l'article présente l'interprétation, par les auteurs, des résultats obtenus lors de cette étude. Le CS du HCB ayant conclu que ces résultats ne mettent en évidence aucune différence statistiquement significative de mortalité, tumorigénicité ou de paramètres biochimiques entre groupes expérimentaux et témoin, il n'a pas jugé utile d'en commenter les interprétations spéculatives. Quelques uns des points critiquables peuvent toutefois être soulignés, parmi les sujets qui n'ont pas encore été discutés dans cet avis :

Le raisonnement portant sur l'hormono-sensibilité des tumeurs présente des lacunes réhibitoires. D'une part, il n'est pas rigoureux de discuter de l'hormono-sensibilité des tumeurs sans en donner le type histologique. D'autre part, il n'est pas justifié de soutenir l'occurrence, dans tous les groupes expérimentaux, d'un effet endocrinien variant non linéairement en fonction du degré de l'exposition aux traitements, à partir de paramètres hormonaux (testostérone, oestradiol) présentés pour un seul groupe (femelles nourries avec un régime contenant 33 % de maïs GM), un seul intervalle de mesure (prélèvement à 15 mois), et sous forme de valeurs individuelles exprimées en pourcentage des valeurs des témoins. En l'absence (1) de données chiffrées faisant apparaître des moyennes et les écarts types, (2) d'une analyse statistique des variations et (3) d'une prise en compte de la variabilité des valeurs en fonction de l'état physiologique (cycle œstral), il est difficile de se prononcer sur la signification de ces observations. Les enseignements tirés par les auteurs relèvent essentiellement de spéculations. L'effet potentiel du déséquilibre hormonal illustre bien cette tendance spéculative : le raisonnement s'appuie d'une part sur les expérimentations *in vitro* produites par les auteurs eux-mêmes et dont les résultats ont été précédemment contestés par la communauté scientifique, et d'autre part sur des références (Vandenberg *et al.*, 2012) dont les conclusions sont incorrectement rattachées.

Dans leur discussion, les auteurs émettent l'hypothèse que l'expression du transgène dans le maïs NK603 n'ayant pas reçu de traitement au Roundup® entraînerait une modification du métabolisme secondaire de la voie de biosynthèse des composés phénoliques chez la plante GM, causant ainsi des perturbations endocriniennes chez les rats nourris avec cet OGM. Deux types de métabolites secondaires ont été considérés par les auteurs : les isoflavones, et les acides phénoliques tels l'acide caféique et l'acide férulique. Dans un premier temps, les auteurs notent qu'il n'y a aucune différence de teneurs en isoflavones œstrogéniques entre les différents régimes de l'étude. Ce résultat n'est pas surprenant sachant que ce type de composés n'existe pas chez le maïs (Dixon, 2004; Yu *et al.*, 2000). Dans un second temps, les auteurs indiquent que les régimes contenant du maïs GM présentent des teneurs en acides caféique et férulique moindres par rapport aux régimes témoin. Les auteurs attribuent cette différence au maïs GM sans examiner les autres composantes du régime. Selon eux, ces acides exerceraient des effets protecteurs contre la tumorigénicité chez les mammifères, et interféreraient avec le métabolisme des œstrogènes. Les acides phénoliques existent rarement sous leurs formes libres solubles, mais incorporés au niveau des parois végétales (liaison à des arabinogalactanes et à des polysaccharides) réduisant ainsi fortement leur biodisponibilité (Buanafina, 2009; Manach *et al.*, 2004; Manach *et al.*, 2005). Leur effet antitumorigène est essentiellement documenté par des études réalisées *in vitro* et sur les

formes libres de ces métabolites secondaires (Gani *et al.*, 2012). De ce fait, le raisonnement indiquant que la diminution du taux de ces acides phénoliques dans le maïs GM pourrait moduler les récepteurs œstrogéniques des cellules de mammifères et provoquer des perturbations endocriniennes reste hautement spéculatif.

Concernant les paramètres biologiques, on ne peut interpréter les variations décrites comme étant associées à des atteintes hépatiques ou rénales sévères. Les atteintes hépatiques ne sont décrites que morphologiquement et les variations des paramètres sanguins sont, soit non biologiquement significatives (par exemple, deux paramètres qui devraient varier dans le même sens (l'ALAT et l'ASAT) varient en sens opposés, Figure 5A) soit non rapportées par les auteurs. Les variations des paramètres rénaux sont modestes et ne sont pas toujours corrélées entre elles (Figure 5 et Table 3). L'insuffisance rénale est définie par une diminution du débit de filtration glomérulaire (DFG). La créatinine est le paramètre le plus largement utilisé pour estimer le DFG, qui est calculé à partir de la mesure de la concentration urinaire et sérique de la créatinine. Une atteinte rénale se traduit par une augmentation de la créatinine et de l'urée sériques, une diminution de la créatinine urinaire, une diminution de la clairance de la créatinine, et une protéinurie. Dans cette étude, on ne retrouve pas cette cohérence entre paramètres biologiques. En effet, dans la Figure 5A (femelles nourries au régime de 33 % de maïs GM), la créatinine sérique serait diminuée et l'urée serait normale, la créatinine urinaire serait diminuée et la clairance de la créatinine (notée « U.Clearance »²⁸) serait diminuée. Si la diminution de la clairance est un signe d'insuffisance rénale, une diminution de la créatininémie est anormale dans cette circonstance. En présence de telles incohérences biologiques, l'hypothèse d'insuffisance rénale ou d'atteinte rénale n'est pas vérifiée.

3. Conclusions et réponse du HCB à la première partie de la saisine

Au terme de l'analyse des résultats de l'article de Séralini *et al.* (2012), le CS du HCB conclut que le dispositif expérimental et les outils statistiques utilisés souffrent de lacunes et faiblesses méthodologiques rédhibitoires, qui ne permettent pas de soutenir les conclusions avancées par les auteurs.

En s'appuyant sur une analyse statistique rigoureuse, le CS du HCB montre que les données de l'article ne mettent en évidence :

- aucune différence statistiquement significative de la mortalité des rats, entre les groupes expérimentaux et le groupe témoin,
- aucune différence statistiquement significative du nombre de tumeurs, entre les groupes expérimentaux et le groupe témoin.

De plus, la méthodologie statistique employée par les auteurs pour l'analyse des paramètres biochimiques est inadéquate et ne permet pas de conclure à l'existence de différences statistiquement significatives entre les groupes expérimentaux et témoin.

Les désordres hormonaux revendiqués par les auteurs chez les rats des groupes expérimentaux ne sont pas spécifiques d'une pathologie reconnue et ne peuvent être différenciés des désordres reconnus comme associés à la souche Sprague-Dawley.

Enfin, malgré une sélection importante, par les auteurs, des résultats présentés, la pertinence physiopathologique des valeurs biologiques commentées n'est pas démontrée. Les insuffisances rénales et/ou hépatiques, proposées à partir des observations anatomiques des animaux à l'autopsie, sans comparaison statistique d'occurrence avec le groupe témoin, ne sont pas validées par les données biologiques obtenues au 15^{ème} mois. De plus, la description inadéquate et lacunaire des lésions anatomo-pathologiques ne permet pas de corroborer les résultats. Ces observations ne sont donc pas conclusives d'un effet du maïs GM sur les rats.

²⁸ A noter que la présentation de l'« U.Creatinine » et l'« UEx.Creatinine » dans la Figure 5A prête à confusion, et peut laisser penser qu'il s'agit de paramètres indépendants qui montreraient, chacun, une variation significative. Il s'agit en fait de paramètres liés entre eux par calcul via le volume urinaire. Le volume urinaire est également représenté dans la Figure 5A.

En conséquence, le CS du HCB considère que la publication n'apporte aucune information qui soutienne l'existence d'un risque sanitaire lié à la consommation de maïs NK603, traité ou non traité par une formulation de Roundup®.

4. Bibliographie

Brix, A.E., Nyska, A., Haseman, J.K., Sells, D.M., Jokinen, M.P., and Walker, N.J. (2005). Incidences of selected lesions in control female Harlan Sprague-Dawley rats from two-year studies performed by the National Toxicology Program. *Toxicol Pathol* 33, 477-483.

Buanafina, M.M.D. (2009). Feruloylation in Grasses: Current and Future Perspectives. *Mol Plant* 2, 861-872.

Dixon, R.A. (2004). Phytoestrogens. *Annu Rev Plant Biol* 55, 225-261.

Domingo, J.L., and Bordonaba, J.G. (2011). A literature review on the safety assessment of genetically modified plants. *Environ Int* 37, 734-742.

Duke, S.O., and Powles, S.B. (2008). Glyphosate: a once-in-a-century herbicide. *Pest Manag Sci* 64, 319-325.

EC (2001). Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. *Official Journal of the European Communities L106*, 1-36.

EC (2003). Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. *Official Journal of the European Union L268*, 1-23.

EFSA (2011). Scientific opinion on guidance for risk assessment of food and feed from genetically modified plants. *The EFSA Journal* 9 (5): 2150, 37 pp.

Funke, T., Han, H., Healy-Fried, M.L., Fischer, M., and Schonbrunn, E. (2006). Molecular basis for the herbicide resistance of Roundup Ready crops. *Proc Natl Acad Sci U S A* 103, 13010-13015.

Gani, A., Wani, S.M., Masoodi, F.A., and Hameed, G. (2012). Whole-grain cereal bioactive compounds and their health benefits: a review. *Food Processing and Technology* 3, 1-10.

Gart, J., Krewski, D., Lee, P.N., Tarone, R.E., and Wahrendorf, J. (1986). *Statistical Methods in Cancer Research: Vol.III : The Design and Analysis of Long Term Animal Experiments*. IARC Scientific Publications No.79.

HCB (2012). Avis HCB-2012.07.31 du Haut Conseil des biotechnologies relatif au dossier EFSA-GMO-UK-2008-60 portant sur une demande de mise sur le marché du maïs génétiquement modifié GA21 pour la culture, l'importation, la transformation et l'utilisation en alimentation animale. (Paris, France), pp. 35.

Keenan, K.P., Ballam, G.C., Dixit, R., Soper, K.A., Laroque, P., Mattson, B.A., Adams, S.P., and Coleman, J.B. (1997). The effects of diet, overfeeding and moderate dietary restriction on Sprague-Dawley rat survival, disease and toxicology. *J Nutr* 127, S851-S856.

Malatesta, M., Boraldi, F., Annovi, G., Baldelli, B., Battistelli, S., Biggiogera, M., and Quaglino, D. (2008). A long-term study on female mice fed on a genetically modified soybean: effects on liver ageing. *Histochem Cell Biol* 130, 967-977.

Manach, C., Scalbert, A., Morand, C., Remesy, C., and Jimenez, L. (2004). Polyphenols: food sources and bioavailability. *Am J Clin Nutr* 79, 727-747.

Manach, C., Williamson, G., Morand, C., Scalbert, A., and Remesy, C. (2005). Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *Am J Clin Nutr* 81, 230S-242S.

OECD (1998). OECD guideline for the testing of chemicals - Repeated Dose 90-day Oral Toxicity Study in Rodents. n°408.

OECD (2009a). OECD guideline for the testing of chemicals - Carcinogenicity Studies. n°451.

OECD (2009b). OECD guideline for the testing of chemicals - Chronic Toxicity Studies. n°452.

OECD (2009c). OECD guideline for the testing of chemicals - Combined Chronic Toxicity/Carcinogenicity Studies. n°453.

Sakamoto, Y., Tada, Y., Fukumori, N., Tayama, K., Ando, H., Takahashi, H., Kubo, Y., Nagasawa, A., Yano, N., Yuzawa, K., *et al.* (2008). A 104-week feeding study of genetically modified soybeans in F344 rats. *J Food Hyg Soc Japan* 49, 272-282.

Séralini, G.-E., Clair, E., Mesnage, R., Gress, S., Defarge, N., Malatesta, M., Hennequin, D., and de Vendôme, J.S. (2012). Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize. *Food Chem Toxicol (in press)*.

Snell, C., Bernheim, A., Berge, J.B., Kuntz, M., Pascal, G., Paris, A., and Ricoch, A.E. (2012). Assessment of the health impact of GM plant diets in long-term and multigenerational animal feeding trials: A literature review. *Food Chem Toxicol* 50, 1134-1148.

Vandenberg, L.N., Colborn, T., Hayes, T.B., Heindel, J.J., Jacobs, D.R., Lee, D.H., Shioda, T., Soto, A.M., vom Saal, F.S., Welshons, W.V., *et al.* (2012). Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. *Endocr Rev* 33, 378-455.

Yu, O., Jung, W.S., Shi, J., Croes, R.A., Fader, G.M., McGonigle, B., and Odell, J.T. (2000). Production of the isoflavones genistein and daidzein in non-legume dicot and monocot tissues. *Plant Physiol* 124, 781-793.

Annexe 1 : Saisine



Ministère des Affaires
sociales
et de la Santé

Ministère délégué
auprès du Ministère de
l'économie et des
finances, chargé de
l'économie sociale et
solidaire et de la
consommation

Ministère de
l'Écologie,
du Développement
durable
et de l'Énergie

Ministère de
l'Agriculture,
de
l'Agroalimentaire
et de la Forêt

Monsieur Marc MORTUREUX

Directeur général de l'Agence
nationale de sécurité sanitaire de
l'alimentation, de l'environnement
et du travail

Monsieur Jean-François DHAINAUT

Président du Haut conseil des
biotechnologies

Paris, le 24 septembre 2012

Monsieur le Directeur général,

Monsieur le Président,

Un article intitulé « *Long term toxicity of a Roundup herbicide and a Roundup-tolerant genetically modified maize* » vient d'être publié dans la revue scientifique *Food and Chemical Toxicology* par l'équipe du professeur Gilles-Eric Séralini. Cet article porte sur une étude conduite sur des rats ayant consommé pendant 2 ans du maïs génétiquement modifié NK603 traité ou non avec l'herbicide Roundup, ou de l'herbicide Roundup seul.

Nous vous demandons, par la présente saisine, de bien vouloir vous rapprocher afin de procéder à une analyse de l'étude rapportée par cet article afin de déterminer si elle est de nature à remettre en cause ou non les conclusions des évaluations précédentes sur cet OGM et notamment si elle peut être considérée comme conclusive quant au risque sanitaire que pourraient présenter les aliments issus de plantes OGM comportant l'événement NK603.

Il est d'autre part demandé à l'ANSES de déterminer si cette étude est de nature à remettre en cause ou non les conclusions des évaluations précédentes de l'ANSES sur l'herbicide Roundup.

Sur la base de cette analyse, nous vous demandons d'évaluer si le protocole mis en œuvre et les conclusions de cette étude remettent en cause les lignes directrices actuelles ou à venir en matière d'évaluation des risques sanitaires.

Nous vous saurions gré de bien vouloir rendre un avis sur cet article avant le 20 octobre 2012 et sur la pertinence des modalités d'évaluation des risques sanitaires et des propositions d'aménagements des lignes directrices, si nécessaire, avant le 20 novembre 2012.

Nous vous prions de croire, Monsieur le Président, Monsieur le Directeur général, à l'assurance de notre considération distinguée.

La Ministre des Affaires
sociales
et de la Santé

Marisol TOURAINE

Le Ministre délégué
auprès du Ministre de
l'économie et des
finances, chargé de
l'économie sociale et
solidaire et de la
consommation

Benoît HAMON

La Ministre de
l'Écologie,
du Développement
durable
et de l'Énergie

Delphine BATHO

Le Ministre de
l'Agriculture,
de l'Agroalimentaire
et de la Forêt

Stéphane LE FOLL

Annexe 2 : Elaboration de l'avis

Un groupe de travail *ad hoc* a été constitué en réponse à la saisine du 24 septembre 2012, composé d'experts rapporteurs sélectionnés pour leurs compétences dans les disciplines pertinentes pour l'analyse de l'article de Séralini *et al.* (2012), leur appartenance au secteur public et leur indépendance :

- quatre experts rapporteurs externes au HCB : Dr Avner Bar-Hen, Professeur à l'Université Paris-Descartes, statisticien, Dr Martine Kolf-Clauw, Professeur à l'École Nationale Vétérinaire de Toulouse, toxicologue vétérinaire, Dr Francelyne Marano, Professeur Emérite de l'Université Paris-Diderot, spécialisée en biologie cellulaire, toxicologie et génotoxicité, et Dr Daniel Marzin, Professeur Emérite de l'Université de Lille II, toxicologue ;
- trois experts rapporteurs membres du CS du HCB : Dr Joël Guillemain, pharmacien expert toxicologue, Chargé de cours à l'Université de Tours, Dr Marc Lavielle, Directeur de Recherche à l'INRIA, statisticien, et Dr Rémy Maximilien, Directeur de recherche au CEA, toxicologue ;
- deux experts rapporteurs membres du CS du HCB, Président et Vice-Président du CS du HCB : Dr Jean-Christophe Pagès, Professeur de biochimie et biologie moléculaire, Praticien hospitalier à la Faculté de médecine de Tours, et Dr Jean-Jacques Leguay, Directeur de recherche du CNRS en biologie et physiologie végétale.

Un expert supplémentaire du secteur privé (Dr Roy Forster du Centre International de Toxicologie, toxicologue) a été consulté par le groupe de travail sur des questions techniques.

Le groupe de travail a réalisé une analyse de l'article de Séralini *et al.* (2012) sous le pilotage du Dr Rémy Maximilien, et proposé un rapport pour examen par le CS du HCB le 2 octobre 2012. Chaque expert externe a signé un engagement de confidentialité, et a certifié n'avoir aucun conflit d'intérêts avec le dossier concerné. Les experts rapporteurs du groupe de travail ont analysé le dossier dans leur domaine d'expertise, et ont été auditionnés par le CS. Ils n'ont toutefois pas contribué directement à l'élaboration de cet avis, qui reste de la responsabilité du CS du HCB.

L'avis définitif a été élaboré par le CS du HCB, sous la présidence du Dr Jean-Christophe Pagès, sous la vice-présidence du Dr Jean-Jacques Leguay, et sous la coordination scientifique du Dr Catherine Golstein, Responsable scientifique et Chargée des affaires européennes au HCB.

Le CS du HCB est un comité pluridisciplinaire composé de personnalités scientifiques nommées par décret au titre de leur spécialité en relation avec les missions du HCB. Par ordre alphabétique des noms de famille, le CS du HCB est composé de :

Claude Bagnis, Yves Bertheau, Pascal Boireau, Denis Bourguet, François-Christophe Coléno, Denis Couvet, Jean-Luc Darlix, Elie Dassa, Maryse Deguerque, Marion Desquilbet, Hubert de Verneuil, Robert Drillien, Nathalie Eychenne, Anne Dubart-Kupperschmitt, Claudine Franche, Philippe Guerche, Joël Guillemain, Mireille Jacquemond, André Jestin, Bernard Klonjowski, Marc Lavielle, Jane Lecomte, Olivier Le Gall, Jean-Jacques Leguay, Didier Lereclus, Rémy Maximilien, Antoine Messéan, Nicolas Munier-Jolain, Jacques Pagès, Jean-Christophe Pagès, Daniel Parzy, Catherine Regnault-Roger, Pierre Rougé, Patrick Saindrenan, Annie Sasco, Pascal Simonet, Virginie Tournay, Bernard Vaissière, Jean-Luc Vilotte.

Antoine Messéan n'a contribué ni à l'élaboration ni à la rédaction de cet avis en raison de son appartenance à l'EFSA, qui a également été saisi sur le sujet. Aucun des autres membres du CS n'a déclaré avoir de conflits d'intérêts qui auraient pu interférer avec l'élaboration de cet avis.

La participation à l'élaboration de l'avis n'implique pas que l'avis adopté ait reçu l'assentiment plein et entier de tous les participants mais indique qu'une majorité s'est dégagée en sa faveur, dans la limite des compétences des experts et après exposé de l'ensemble des points de vue.

Annexe 3 : Aspects statistiques du protocole expérimental

Un échantillon de 200 rats, constitué de 100 mâles et 100 femelles, a été randomisé en 20 groupes de 10 rats de même sexe, chaque groupe recevant le même régime alimentaire. Pour chaque sexe, un seul groupe témoin de 10 rats a été constitué. C'est donc uniquement ce même groupe de 10 rats qui est systématiquement utilisé pour être comparé aux 9 groupes expérimentaux de même sexe.

Il en résulte un manque de puissance tel qu'il est dès lors très difficile d'établir si des différences observées lors de chacune de ces 9 comparaisons sont dues à un effet du régime, ou bien simplement à une variabilité naturelle du groupe contrôle. Ainsi, si un paramètre est particulièrement élevé chez le groupe contrôle, toutes les différences entre groupes expérimentaux et contrôles iront dans le même sens et présenteront toutes une diminution du caractère considéré, sans que l'on puisse pour autant en conclure à un effet du régime. A titre d'exemple, les auteurs concluent §3.3 sans aucune précaution que « *Creatinine or clearance decreased in urine for all treatment groups in comparison to female controls (Table 3)* » alors qu'il est très probable que ce sont des valeurs particulièrement élevées de ces paramètres dans le groupe contrôle qui expliquent ces différences.

Aucun calcul préalable du nombre de sujets nécessaires pour mettre en évidence un effet jugé biologiquement significatif n'est mentionné. Un tel calcul aurait été particulièrement pertinent pour évaluer la quantité d'information, par exemple sur les durées de vie ou les nombres de tumeurs, que l'on peut espérer obtenir avec le protocole mis en place. Il aurait ainsi été possible d'évaluer quantitativement le manque d'intérêt d'utiliser une souche de rats qui développe naturellement des tumeurs avec une grande probabilité. En effet, plus le risque de développer naturellement des tumeurs est élevé, et plus il faudra d'animaux dans les groupes expérimentaux pour mettre en évidence une augmentation significative du nombre de tumeurs dû au régime. Considérons à titre d'exemple 2 souches de rats A et B pour qui le risque de développer naturellement des tumeurs au cours d'une période de temps donnée est de 10% pour A et 60% pour B. Le tableau 1 indique le nombre de rats nécessaire dans un groupe expérimental pour mettre en évidence une augmentation du nombre de tumeurs de 10%, 20% ou 30%, lorsque le risque d'erreur de première espèce α et celui de seconde espèce β sont tous les deux égaux à 5% ou 10%.

		Souche de rats			
		A p=10%	B p=60%		
Effet à mettre en évidence	+ 10%	135	248	86	156
	+ 20%	41	60	25	36
	+ 30%	24	24	15	15
		$\alpha=\beta=5\%$		$\alpha=\beta=10\%$	

Tableau 1. Nombres de rats nécessaires dans un groupe expérimental pour mettre en évidence une augmentation donnée du nombre de tumeurs dans deux souches A et B (le risque de développer naturellement des tumeurs au cours d'une période de temps donnée est de 10% pour A et 60% pour B) et pour différents risques d'erreur de première et seconde espèces²⁹.

²⁹ Le risque de première espèce α est le risque de conclure à une augmentation alors qu'il n'y a pas de différence ; le risque de seconde espèce β consiste à ne pas détecter d'augmentation alors qu'elle existe.

On voit ainsi que l'utilisation de la souche B nécessite plus d'animaux que la souche A, quels que soient les risques d'erreurs donnés et lorsque l'augmentation à mettre en évidence est inférieure à 30%.

Plus généralement, le plan d'analyse statistique n'est pas mentionné. Les auteurs de l'étude semblent avoir effectué leur étude statistique en fonction des résultats obtenus, ce qui est en totale contradiction avec les règles élémentaires de bonnes pratiques statistiques. En effet, le degré de signification d'une différence observée pour un paramètre donné n'est pas du tout le même, suivant que ce paramètre a été sélectionné a priori (avant d'obtenir les résultats), ou bien *a posteriori* (parmi les paramètres présentant le plus de différences). Les auteurs présentent ainsi Figure 5-B les 4 paramètres biochimiques et les deux hormones qui présentent le plus de différences, parmi le groupe qui présente le plus de différences. Ce choix a été fait a posteriori, *i.e.* une fois les résultats obtenus : il est donc attendu que certaines comparaisons parmi les $18 \times 48 = 864$ présentent des différences qui peuvent sembler importantes. La présentation brute de ces résultats partiels peut alors être trompeuse pour un lecteur non spécialiste des comparaisons multiples, qui risque de considérer à tort ces différences observées comme représentatives des différences entre groupes expérimentaux et groupes contrôles.

Annexe 4 : Aspects statistiques de l'analyse des résultats

1. Inférence statistique

C'est la statistique inférentielle qui permet d'évaluer les incertitudes et les probabilités de se tromper en concluant à la présence ou à l'absence d'effets. C'est-à-dire si les différences observées peuvent être expliquées par un effet du régime, ou bien simplement par les fluctuations aléatoires d'échantillonnage. En d'autres termes, la question qui se pose naturellement dans ce genre d'étude concerne la reproductibilité des résultats observés : si l'on répète la même expérience dans les mêmes conditions, quelles seront les chances d'obtenir des résultats similaires à ceux observés ici ?

En ce qui concerne les études de durées de vie et de nombres de tumeurs, les auteurs ont totalement négligé cet aspect de la statistique, tout en s'autorisant à des interprétations non justifiées de leurs résultats expérimentaux. On lit ainsi p. 8-9 :

- ***“All treatments in both sexes enhanced large tumor incidence by 2–3-fold in comparison to our controls [...]”***
- ***“Suffering inducing euthanasia and deaths corresponded mostly in females to the development of large mammary tumors. These appeared to be clearly related to the various treatments when compared to the control groups.”³⁰***

On ne trouve pas un seul argument statistique dans cet article qui démontre l'existence de telles relations de cause à effet. On ne trouve pas la moindre analyse statistique qui mette en évidence une différence statistiquement significative des durées de vie et des nombres de tumeurs entre groupes expérimentaux et groupes témoin.

Concernant les paramètres biochimiques, on peut lire dans la conclusion de l'article p. 10 :

- ***“The results of the study presented here clearly demonstrate that lower levels of complete agricultural glyphosate herbicide formulations, at concentrations well below officially set safety limits, induce severe hormone-dependent mammary, hepatic and kidney disturbances.”***
- ***“Altogether, the significant biochemical disturbances and physiological failures documented in this work confirm the pathological effects of these GMO and Roundup® treatments in both sexes, with different amplitudes.”³¹***

De telles affirmations méritent d'être rigoureusement justifiées et validées. Or, il est ici absolument impossible de conclure de façon définitive à la toxicité du NK603 sur la base de données aussi limitées.

³⁰ Traduction : « **Tous les traitements dans les deux sexes ont augmenté l'incidence des larges tumeurs** de 2 à trois fois en comparaison avec nos témoins [...] ». « Les souffrances qui ont conduit à des euthanasies et des morts correspondaient principalement, chez les femelles, au développement de larges tumeurs mammaires. **Celles-ci apparaissent clairement associées aux différents traitements en comparaison aux témoins** »

³¹ Traduction : « Les résultats de l'étude présentée ici **démontrent clairement** que de faibles niveaux de préparations complètes d'herbicide agricole à base de glyphosate, à des concentrations bien inférieures aux limites de sécurité fixées officiellement, **induisent des troubles hormono-dépendant sévères d'ordre mammaire, hépatique et rénal**. ». « Au final, les perturbations biochimiques **significatives** et les troubles physiologiques présentés dans ce travail **confirment les effets pathologiques de ces traitements OGM et Roundup® pour les deux sexes, avec des amplitudes différentes.** »

2. Analyse de survie

2.1. Durée de vie

Les auteurs expliquent §3.1 que “Control male animals survived on average 624 ± 21 days, whilst females lived for 701 ± 20 ”. Ces valeurs sont le résultat d’un calcul incorrect puisque nous sommes en présence de données censurées (on ignore quand les animaux encore vivants en fin d’étude seraient morts s’ils n’avaient pas été euthanasiés). Ce sont les moyennes et écart-types empiriques des valeurs observées non censurées et des temps de censures qui ont été calculées (comme si un rat encore vivant à $T=720j$ était considéré comme mort à $T=720j$). Les résultats présentés sont donc inexacts puisque cette procédure introduit un biais en sous-estimant bien sûr le temps moyen de mort, mais surtout l’erreur standard de l’estimateur. Avoir choisi de griser tout ce qui se passe après $624-21=603$ jours n’est donc pas justifié puisque cette valeur est issue d’un calcul incorrect.

Un calcul correct de la distribution des durées de vie dans les différents groupes nécessite d’utiliser un modèle paramétrique, mais l’intérêt d’une telle approche est limitée avec aussi peu de données par groupe. A titre d’exemple, si l’on ajuste un modèle gaussien pour le temps de survie des mâles, les moyennes et écart-type estimées sont respectivement 626 jours et 68 jours. Pour les femelles, les moyennes et écart-type estimés sont respectivement 892 jours et 206 jours. On ne peut, comme le font les auteurs, calculer une erreur standard pour la moyenne en divisant simplement l’écart-type par $\sqrt{10}$, comme on le ferait pour des variables gaussiennes non censurées. A cause du phénomène de censure, la distribution de l’estimateur de la moyenne est beaucoup plus dispersée et très asymétrique, ce qui rend de plus l’utilisation de l’erreur standard peu pertinente pour calculer un intervalle de confiance.

2.2. Comparaisons entre groupes expérimentaux et groupes témoin

L’analyse de survie réalisée dans cette étude se limite à une représentation graphique des courbes de mortalité dans chaque groupe (nombre de rats morts en fonction du temps).

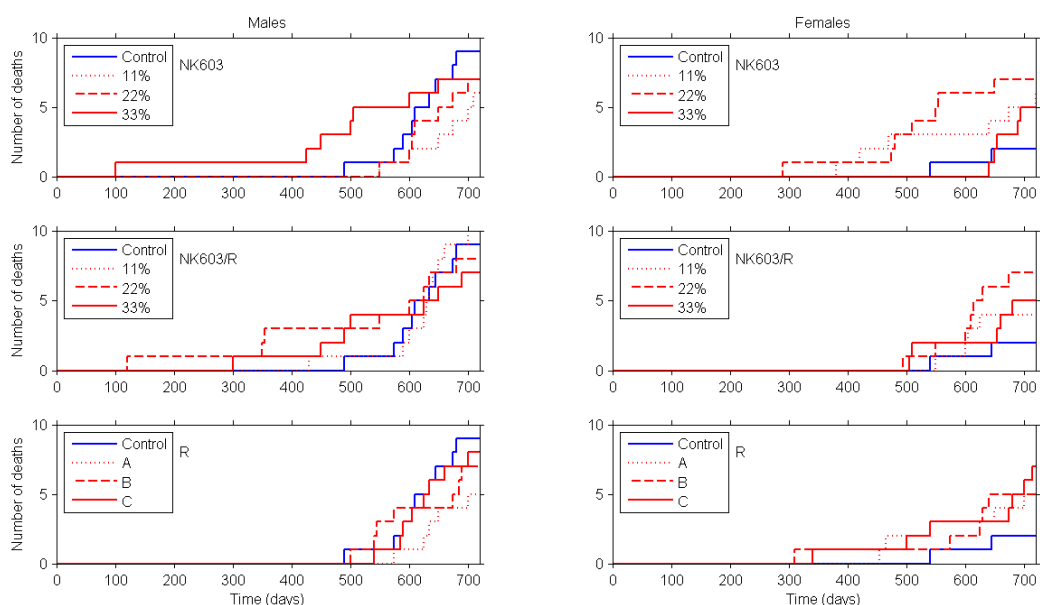


Fig. 1. Courbes de mortalité pour les 18 groupes expérimentaux et les 2 groupes témoin (*control groups*).

Un simple examen des courbes de mortalité observées ne permet pas de conclure à une quelconque différence au niveau de la population. Il faut pour cela mettre en œuvre une procédure statistique rigoureuse qui prend en compte la variabilité des individus et donc des courbes de survie.

De nombreuses techniques statistiques existent pour comparer des courbes de survie. On peut dans un premier temps considérer les 18 comparaisons entre groupes expérimentaux et groupes témoin. Le test de rang (ou test de Wilcoxon) est un test non paramétrique qui permet de comparer les statistiques de rang de 2 échantillons.

On peut ainsi tester si les animaux d'un groupe expérimental donné ont tendance à mourir plus tôt que ceux du groupe témoin. Par exemple :

H_0 « le régime NK603 à 11% n'a pas d'effet sur la durée de vie des rats femelles »

vs

H_1 « le régime NK603 à 11% entraîne une diminution de la durée de vie des rats femelles »

La statistique de test est définie comme la somme des rangs du groupe témoin. Pour chacune des 18 comparaisons, cette statistique de test peut être calculée et comparée à un intervalle de prévision obtenu sous l'hypothèse nulle. Un degré de signification peut être calculé pour chacun des 18 tests effectués.

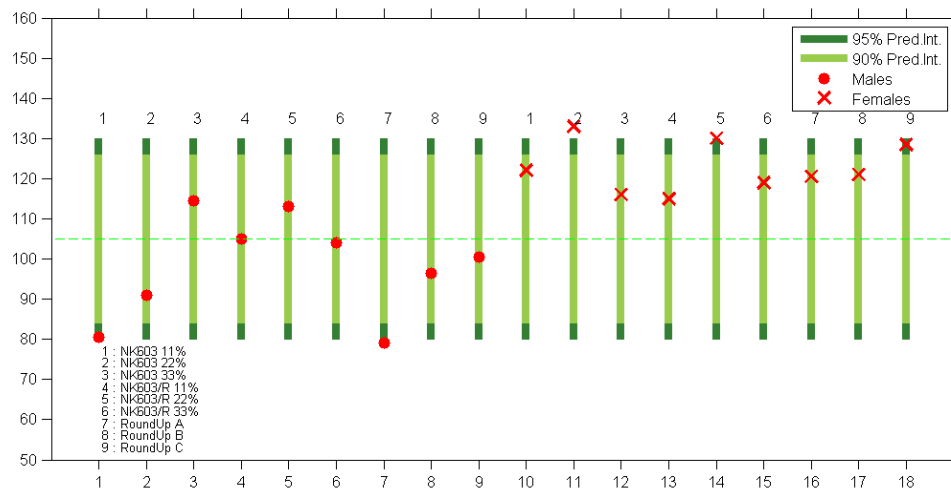


Fig. 2a. Intervalles de prévision de niveau 90 % et 95 % des 18 statistiques de tests et valeurs observées de ces statistiques de tests.

Néanmoins, ces intervalles de prévision ne prennent pas en compte la multiplicité des comparaisons effectuées. Plutôt que mettre en œuvre un test trop conservateur (qui tendrait à conserver trop systématiquement l'hypothèse nulle) on peut estimer par simulation, ou par permutation, la distribution de probabilité des 18 statistiques utilisées pour ce test. La figure ci-dessous montre les intervalles de prévision des 18 statistiques de tests lorsqu'elles sont classées par ordre décroissant (ces intervalles sont estimés par simulation). Les 18 statistiques de tests sont toutes à l'intérieur des intervalles de prévision de niveau 90 % correspondants :

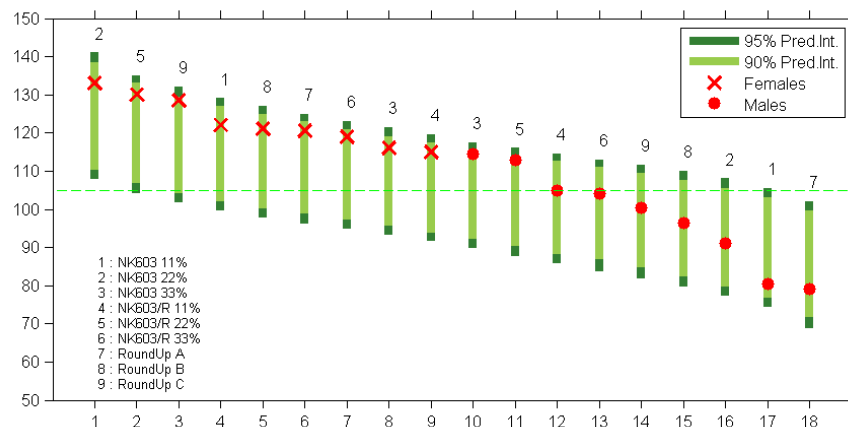


Fig. 2b. Intervalles de prévision de niveau 90 % et 95 % des 18 statistiques de tests ordonnées par ordre décroissant et valeurs observées de ces statistiques de tests.

On peut de plus estimer le degré de signification de chaque comparaison comme un quantile empirique. Le tableau présente pour chacun des 18 tests effectués la statistique de test, *i.e.* la somme des rangs du groupe témoin (sous l'hypothèse nulle, la valeur attendue de cette statistique est $(1+2+\dots+19+20)/2=105$), le degré de signification du test de rang correspondant, le degré de signification corrigé qui prend en compte la multiplicité des tests effectués (par simulation ou par permutation). Les groupes qui ont tendance à mourir avant le groupe témoin (statistique de test supérieure à 105) sont en rouge et les autres en bleu :

Groupe expérimental	Statistique de test	Degré de signification	Degré de signification corrigé (simulation)	Degré de signification corrigé (permutation)
M - NK603 11%	80.5	0.972	0.879	0.895
M - NK603 22%	91	0.865	0.665	0.624
M - NK603 33%	114.5	0.247	0.151	0.073
M - NK603/R 11%	105	0.515	0.385	0.291
M - NK603/R 22%	113	0.285	0.167	0.072
M - NK603/R 33%	104	0.545	0.368	0.263
M - RoundUp A	79	0.978	0.819	0.808
M - RoundUp B	96.5	0.753	0.514	0.452
M - RoundUp C	100.5	0.648	0.422	0.343
F - NK603 11%	122	0.072	0.199	0.174
F - NK603 22%	133	0.011	0.158	0.166
F - NK603 33%	116	0.176	0.195	0.124
F - NK603/R 11%	115	0.188	0.185	0.104
F - NK603/R 22%	130	0.021	0.122	0.104
F - NK603/R 33%	119	0.116	0.152	0.092
F - RoundUp A	120.5	0.092	0.140	0.098
F - RoundUp B	121	0.085	0.178	0.139
F - RoundUp C	128.5	0.029	0.087	0.067

Tableau 1. Les 18 statistiques de tests (test de Wilcoxon) et les degrés de signification associés.

Le groupe qui présente le plus de différences en ce qui concerne la survie est le groupe des femelles nourries avec un régime contenant 22 % de maïs NK603 non traité. Le degré de signification de ce test est de 1,1 % (*i.e.* la probabilité d'obtenir une statistique de test supérieure ou égale à 133 sous l'hypothèse nulle est de 1,1 %). En prenant compte de la multiplicité des tests, cette probabilité est de 15,8 % lorsqu'elle est estimée par simulation et de 16,6 % lorsqu'elle est estimée par permutation : la probabilité que la plus grande statistique de test parmi 18 soit supérieure ou égale à 133 sous l'hypothèse nulle est d'environ 16 %. Ce tableau nous permet donc de conclure que :

Aucune différence observée entre les courbes de survie des groupes expérimentaux et celles des groupes témoin n'est statistiquement significative.

Enfin, une étude de puissance par simulation montre, à titre d'exemple, que si la mort de 5 rats d'un groupe expérimental de 10 rats survient avant celle des rats témoins, le degré de signification du test est de 8 %. Ce degré de signification du test n'est plus que de 2 % (resp. 0,8 %) si ce sont 6 (resp. 7) rats qui meurent avant les témoins.

2.3. Utilisation de données de référence

Le manque de puissance, dû à un très faible effectif des groupes témoin (10 rats), interdit bien entendu de conclure de façon formelle à la présence ou à l'absence d'un effet du régime sur la mortalité, en particulier chez les femelles. Ce manque de puissance peut être compensé par l'introduction d'informations à priori sur le comportement attendu des groupes témoin. Ainsi, des données de mortalité de la souche SD fournies par la société Harlan sont disponibles et peuvent venir compléter l'information apportée par l'expérimentation. Bien sûr, ces études n'ont pas été réalisées exactement dans les mêmes conditions que l'étude qui nous intéresse et un biais peut donc être introduit en utilisant cette information a priori. Au contraire, l'information apportée par les groupes témoin n'est pas biaisée si tous les groupes ont été suivis dans les mêmes conditions, mais comme nous l'avons vu, cette information est entachée d'une très grande variabilité. La combinaison de l'information « a priori » fournie par l'éleveur et de celle fournie par les données de l'expérience permet un bon compromis « biais-variance ».

Ici, les données fournies par la société Harlan indiquent un taux de survie à 2 ans de 32 % pour les mâles et 48 % pour les femelles. Pour chaque sexe, le nombre de rats en vie après 2 ans est une variable aléatoire binomiale dont on peut construire des intervalles de prévision de niveau 90 % ou 95 %³².

³² Pour une variable binomiale, seuls les quantiles d'ordre (i/n , $i=0, 2, \dots, n$) peuvent être calculés directement à partir de la distribution de probabilité. On obtient n'importe quel autre quantile par interpolation linéaire.

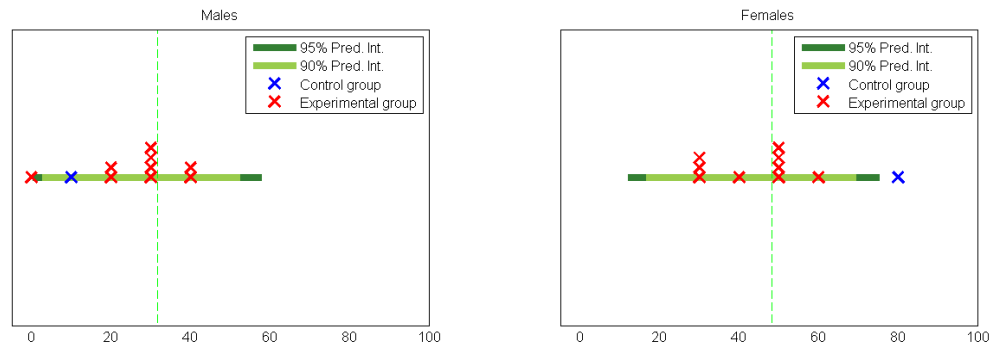


Fig. 3. Intervalles de prévision de niveau 90 % et 95 % des taux de survie à 2 ans obtenus à partir des données Harlan et taux de survie observés dans les groupes expérimentaux et les groupes témoin (*control groups*).

Les groupes expérimentaux sont très globalement distribués à l'intérieur de ces intervalles de prévision. Sur 18 groupes, il n'y a rien d'anormal à ce qu'une observation soit au bord d'un intervalle de prévision de niveau 95 %. Les taux de survie observés à 2 ans sont tout à fait compatibles avec les données de référence fournies par l'éleveur, lorsque les rats sont élevés en condition normale.

L'utilisation de données de référence fournies par l'éleveur confirme que l'on ne peut expliquer les différences observées dans les courbes de survie entre groupes expérimentaux et groupes témoin par un effet régime.

On peut également remarquer que les taux de survie observés dans les groupes témoin sont au contraire assez éloignés de ce que les valeurs de références laissent présager (la proportion observée de rats femelles en vie après 2 ans est hors de l'intervalle de prévision de niveau 95 %). Cela confirme la fragilité statistique de résultats obtenus à partir d'aussi faibles effectifs : on ne peut tirer de ces données aucune conclusion définitive.

2.4. Moins de groupes, mais plus de puissance

Ces résultats montrent qu'il est impossible de considérer avec une puissance suffisante les 18 comparaisons possibles entre groupes expérimentaux et groupes témoin : le protocole n'est pas du tout adapté pour un objectif aussi ambitieux.

Limiter les comparaisons en regroupant certains groupes permet de mettre en œuvre des tests plus robustes et plus puissants. On peut se limiter par exemple à tester si la mortalité chez les rats témoins est plus faible qu'au sein des groupes expérimentaux. On regroupe alors, pour chaque sexe, les groupes expérimentaux et on construit une unique courbe de survie (probabilité d'être en vie au cours du temps). Les 2 groupes expérimentaux (mâles et femelles) sont maintenant chacun formés de 90 animaux : on peut donc raisonnablement approcher les vraies fonctions de survie inconnues par les courbes de survie empiriques, obtenues à partir de ces échantillons de 90 rats. On peut alors comparer ces fonctions de survie aux courbes de survie empiriques des groupes témoin.

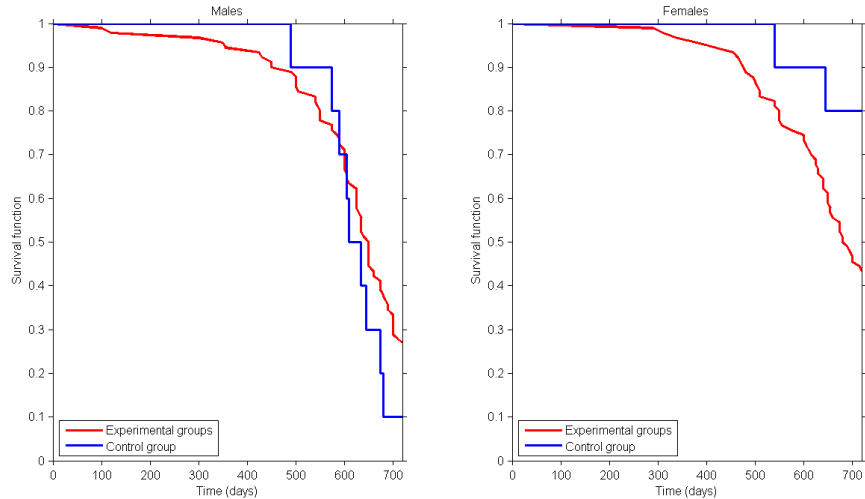


Fig. 4. Courbes de survie dans les groupes expérimentaux et groupes témoin (*control groups*). Pour chaque sexe, le groupe expérimental regroupe les 9 groupes expérimentaux initiaux.

La question qui se pose est de savoir si la courbe de survie bleue (groupe témoin observé) peut avoir été obtenue à partir de 10 rats dont la probabilité de survie est décrite par la courbe rouge.

Un intervalle de prévision des courbes de survie des 2 groupes expérimentaux (mâles et femelles) peut être facilement construit par simulation. Pour chaque sexe, on utilise la courbe de survie du groupe expérimental (rouge) pour simuler un très grand nombre (10 000 dans cet exemple) de groupes de 10 rats et leurs dates de décès. On peut alors construire les 10 000 courbes de survie empiriques obtenues à partir de ces 10 000 groupes simulés. On construit finalement un intervalle de prévision de niveau $1-\alpha$ en calculant à chaque instant les quantiles empiriques d'ordre $\alpha/2$ et $1-\alpha/2$ des 10 000 courbes de survie. Ainsi, un intervalle de prévision de niveau 90% (resp. 95 %) est obtenu en calculant les quantiles empiriques d'ordre 5 % (resp. 2,5 %) et 95 % (resp. 97,5 %).

Le graphique ci-dessous représente pour chaque sexe les courbes de survie des 9 groupes expérimentaux et les intervalles de prévision de niveau 90 % et 95 %.

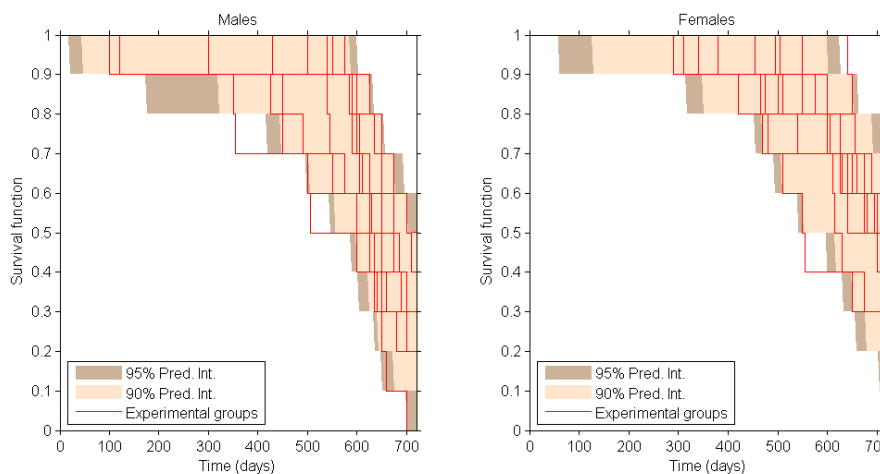


Fig. 5. Intervalles de confiance de niveau 90 % et 95 % de la survie du groupe expérimental et courbes de survie observées des groupes expérimentaux.

Le graphique ci-après représente maintenant pour chaque sexe les mêmes intervalles de prévision, mais avec cette fois-ci la courbe de survie du groupe témoin :

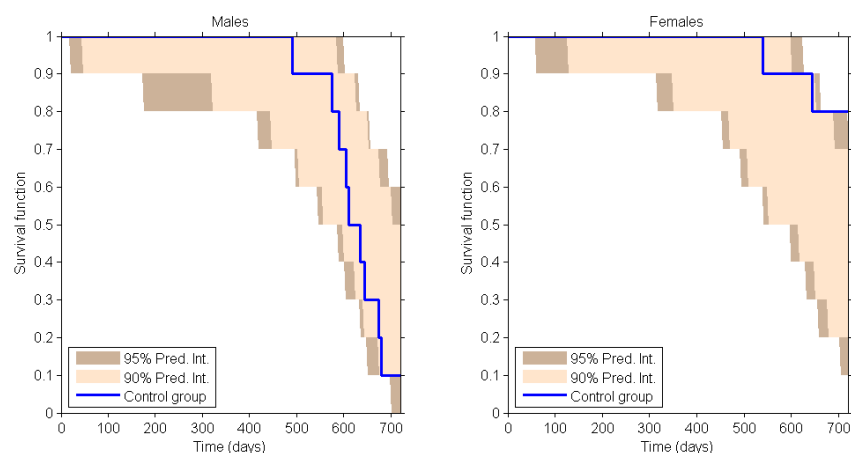


Fig. 6. Intervalles de prévision de niveau 90 % et 95 % de la survie du groupe expérimental et courbes de survie observées des groupes témoin (*control groups*).

Les courbes de survie des groupes témoin sont très globalement à l'intérieur de ces intervalles de prévision :

On ne peut donc conclure à une différence statistiquement significative entre la survie des rats témoins et des rats tests.

Là encore, les données de référence fournies par l'éleveur viennent étayer cette conclusion puisque les intervalles de prévision des taux de mortalité à 2 ans (pour un échantillon de taille $n = 90$) contiennent les taux observés dans les groupes expérimentaux :

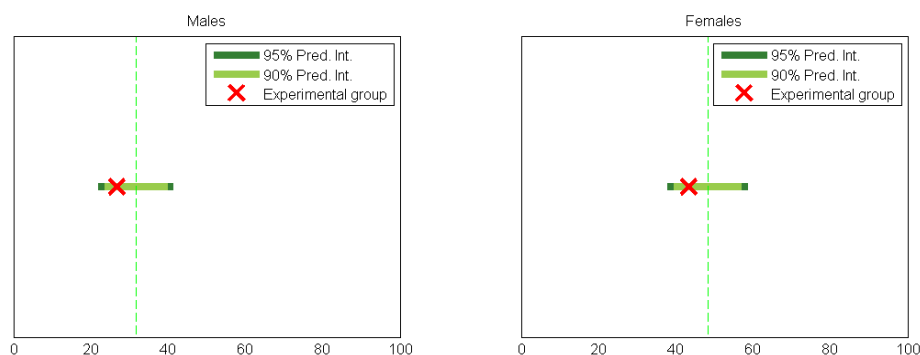


Fig. 7. Intervalles de prévision de niveau 90 % et 95 % des taux de mortalités à 2 ans obtenus à partir des données Harlan et taux mortalités observés dans les groupes expérimentaux (chacun de taille 90).

Le choix de regrouper les 9 groupes expérimentaux et tester si leur distribution est différente de celle du groupe témoin est bien entendu discutable. On peut vouloir mettre en œuvre d'autres tests, et tester par exemple si la consommation d'un maïs OGM a un impact sur la survie, et ce, quelle que soit la dose de maïs OGM dans le régime et quel que soit le traitement associé (Roundup® ou non). Le groupe expérimental est alors formé des 6 premiers groupes expérimentaux (NK603 et NK603 traité au Roundup®, 11 %, 22 % et 33 %) tandis que le groupe témoin est formé du groupe témoin initial (ni OGM, ni OGM traité au Roundup®, ni Roundup®) et des 3 groupes de rats ayant absorbé du Roundup®. Le graphique ci-dessous montre qu'il n'existe pas de différences significatives dans les survies au sein de ces groupes, aussi bien pour les mâles que pour les femelles. On ne peut donc conclure à un effet significatif du maïs NK603 sur la mortalité des rats.

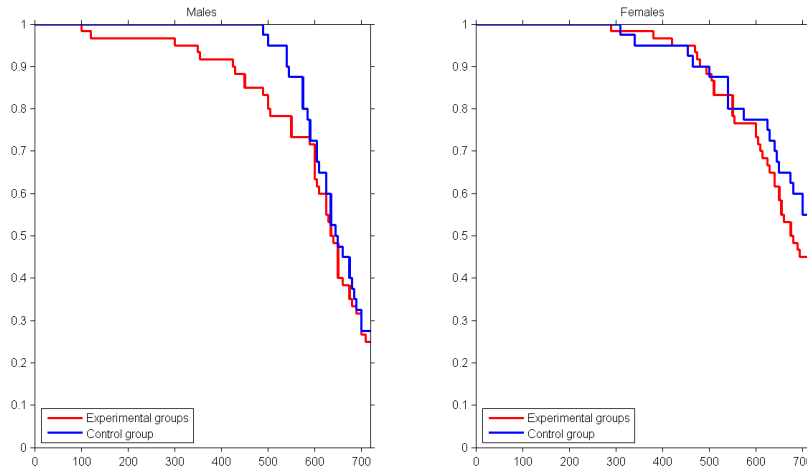


Fig. 8. Courbes de survie dans les groupes expérimentaux et groupes témoin. Pour chaque sexe, le groupe expérimental regroupe les 6 groupes expérimentaux initiaux ayant consommé du maïs NK603 traité au Roundup® ou non traité. Le groupe témoin contient le groupe témoin initial et les 3 groupes ayant absorbé du Roundup®.

On peut également tester si l'absorption de Roundup® (sous forme liquide) a un impact sur la survie, et ce, quelle que soit la quantité de Roundup® absorbée. Le groupe expérimental est alors formé des 3 groupes de rats ayant absorbé du Roundup® tandis que le groupe témoin est formé du groupe témoin initial (ni OGM, ni OGM traité au Roundup®, ni Roundup®) et des 6 premiers groupes expérimentaux (NK603 et NK603 traité au Roundup®, 11 %, 22 % et 33 %). Le graphique ci-dessous montre qu'il n'existe pas de différences significatives dans les survies au sein de ces groupes, aussi bien pour les mâles que pour les femelles. On ne peut donc conclure à un effet significatif du Roundup® sur la mortalité des rats.

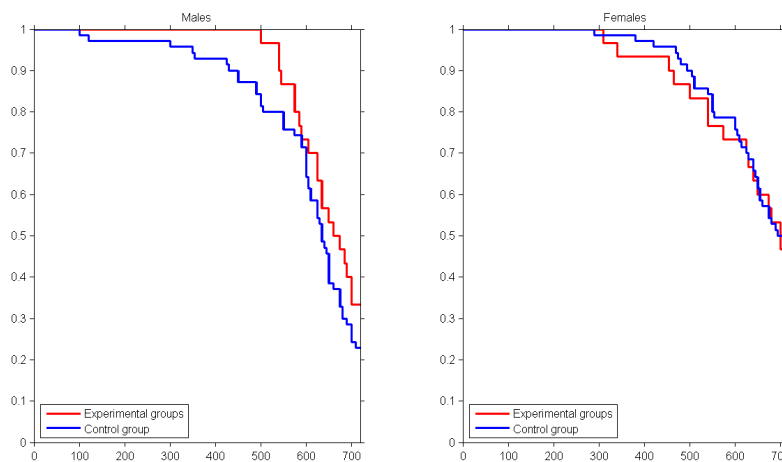


Fig. 9. Courbes de survie dans les groupes expérimentaux et groupes témoin. Pour chaque sexe, le groupe expérimental regroupe les 3 groupes ayant absorbé du Roundup®. Le groupe témoin contient le groupe témoin initial et les 6 groupes expérimentaux initiaux ayant consommé du maïs NK603 traité au Roundup® ou non traité.

En conclusion, on ne peut conclure à un effet statistiquement significatif d'aucun traitement (NK603, NK603 traité au Roundup®, Roundup®) sur la survie des animaux.

3. Nombres de tumeurs

L'analyse des nombres de tumeurs réalisée dans cette étude se limite à une représentation graphique nombres de tumeurs palpables observées dans chaque groupe en fonction du temps :

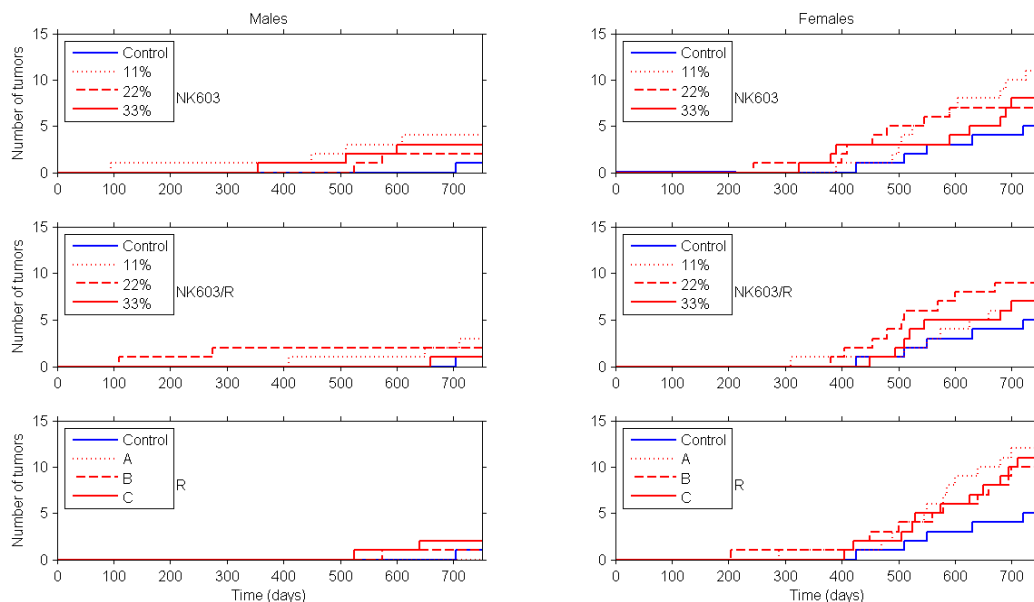


Fig. 10. Evolution des nombres de tumeurs palpables dans les 20 groupes.

Comme pour les courbes de survie, un simple examen de ces courbes observées ne permet pas de conclure à une quelconque différence au niveau de la population. Il faut pour cela mettre en œuvre un test statistique rigoureux qui prend en compte la variabilité des individus et donc des courbes de nombres de tumeurs.

Et comme précédemment, le protocole n'est pas du tout adapté pour effectuer les 18 comparaisons proposées. On peut, pour chaque sexe, regrouper les groupes expérimentaux et construire une unique courbe de nombres de tumeurs que l'on peut comparer à celle obtenue à partir du groupe témoin.

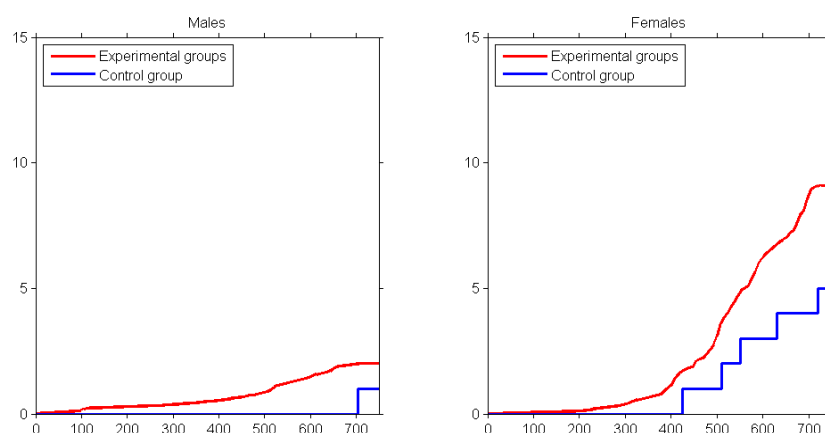


Fig. 11. Nombres de tumeurs palpables dans les groupes expérimentaux et groupes témoin. Pour chaque sexe, le groupe expérimental regroupe les 9 groupes expérimentaux initiaux.

La question qui se pose ici est de savoir si la courbe de nombres de tumeurs bleue (groupe témoin observé) peut avoir été obtenue à partir de 10 rats dont l'évolution du nombre de tumeurs est décrite par la courbe rouge.

Un intervalle de prévision des courbes de nombres de tumeurs des 2 groupes expérimentaux (mâles et femelles) peut être facilement construit en supposant que, pour chaque sexe, le nombre de tumeurs est un processus de Poisson non homogène dont l'intensité est donnée à chaque instant par la courbe rouge. Ainsi, un intervalle de prévision de niveau 90 % (resp. 95 %) est obtenu en calculant à chaque instant les quantiles d'ordre 5 % (resp. 2,5 %) et 95 % (resp. 97,5 %) d'une variable de Poisson.

Le graphique ci-dessous représente pour chaque sexe les nombres de tumeurs des 9 groupes expérimentaux et les intervalles de prévision de niveau 90 % et 95 % :

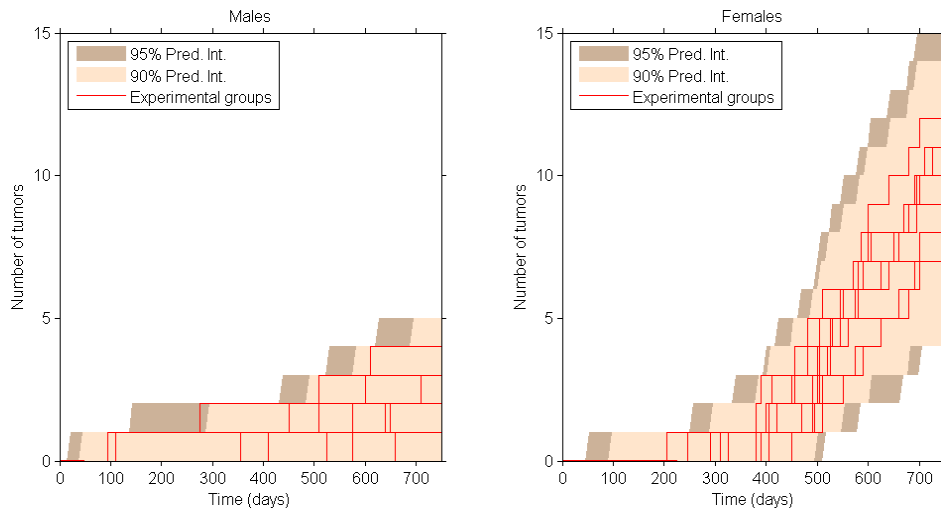


Fig. 12. Intervalles de prévision de niveau 90 % et 95 % des nombres de tumeurs du groupe expérimental et nombres de tumeurs observées des groupes expérimentaux.

Le graphique ci-dessous représente maintenant pour chaque sexe les mêmes intervalles de prévision, mais avec cette fois-ci le nombre de tumeurs du groupe témoin :

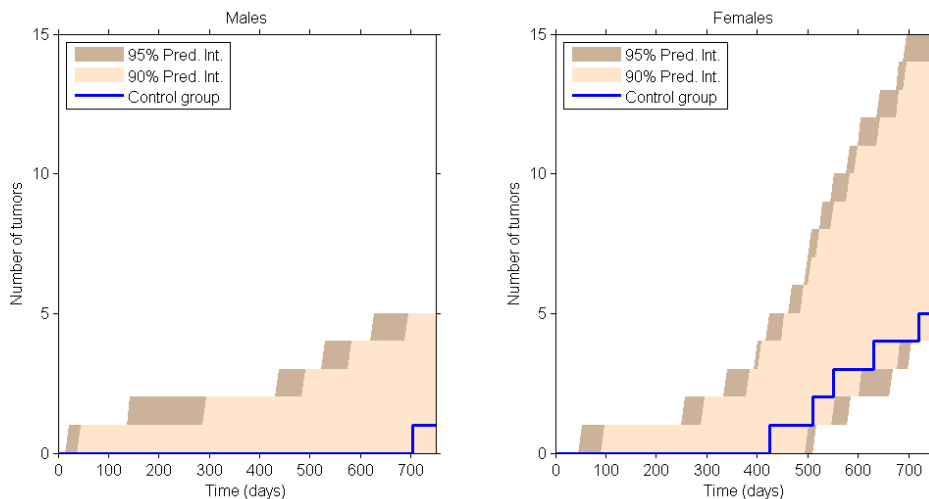


Fig. 13. Intervalles de prévision à 90 % et 95 % des nombres de tumeurs du groupe expérimental et nombres de tumeurs observées des groupes témoin (*control groups*).

Les courbes de nombres de tumeurs des groupes témoin sont à l'intérieur de ces intervalles :

On ne peut donc conclure à un effet statistiquement significatif du régime sur l'évolution du nombre de tumeurs.

4. Paramètres biochimiques

47 paramètres biochimiques ont été mesurés et un paramètre supplémentaire a été calculé lors de l'étude auprès de chacun des 20 groupes. Pour chaque sexe, pour chaque condition expérimentale, une méthode de type « *Orthogonal Partial Least Squares Discriminant Analysis* » (OPLS-DA) est mise en œuvre pour discriminer le groupe témoin et le groupe expérimental.

La méthode OPLS-DA est largement utilisée en chimiométrie ou en génomique pour identifier un sous-ensemble de variables qui différencient au mieux différents sous-groupes. Elle est tout particulièrement pertinente lorsque le nombre de variables explicatives est grand devant le nombre d'observations. De plus, la méthode OPLS-DA permet de construire un modèle prédictif, qui, pour un jeu de variables explicatives donné, fournit les probabilités d'appartenir à chacun des sous-groupes considérés.

Le choix de cette méthode et son utilisation par les auteurs dans ce contexte appellent plusieurs commentaires :

1. Ce type de méthode est réputé pour « sur-ajuster » les données observées lorsque le nombre de variables explicatives est grand devant le nombre d'observations (ce qui est le cas ici). En effet, on pourra toujours trouver un modèle défini par 48 paramètres qui séparera parfaitement 2 groupes de 10 sujets, et ce, quels que soient les groupes ! Pour valider le modèle obtenu (*i.e.* s'assurer qu'il possède bonnes propriétés prédictives), on peut utiliser :
 - un échantillon test afin de s'assurer que le modèle ajusté sur un échantillon d'apprentissage conserve de bonnes propriétés prédictives sur de nouvelles données qui n'ont précisément pas été utilisées pour la construction du modèle,
 - des méthodes de validation croisées (ce qui revient à faire jouer aux différentes données alternativement le rôle d'échantillon d'apprentissage et d'échantillon test).

Les auteurs de l'étude ne présentent aucun critère de validation des modèles obtenus à partir des groupes expérimentaux de 10 rats. Ces modèles ne peuvent donc être utilisés en prédiction.

2. Un modèle donné est défini par 48 paramètres : on construit donc ici 18 modèles définis chacun par 48 paramètres. Autant de modèles construits avec aussi peu de données n'ont que très peu d'intérêt : ces modèles sont peu stables et offrent un pouvoir prédictif très limité. Il aurait été plus pertinent de construire un modèle unique intégrant des effets régimes et sexe ainsi que d'éventuelles interactions régime-sexe (voir d'éventuelles relations dose-effet non linéaires). L'avantage d'un modèle unique est qu'il est construit à partir de l'ensemble des données. On évite ainsi une sur-paramétrisation du modèle qui gagne en stabilité et en pouvoir prédictif.
3. L'utilisation de ces méthodes suppose implicitement une distribution symétrique (proche autant que possible de la distribution normale) des variables explicatives. Il est connu que des paramètres tels que les paramètres biochimiques ont une distribution asymétrique et qu'une transformation préalable est nécessaire pour les rendre le plus « gaussien » possible. Il est par exemple classique d'utiliser certains log-paramètres plutôt que les paramètres d'origine. Un rapport de l'Anses³³ suggère d'utiliser une transformation Box-Cox pour chaque paramètre, le paramètre de puissance étant le même pour les différents groupes, différents paramètres de position caractérisant chaque groupe.
4. Calculer des intervalles de confiance pour chaque paramètre n'est pas pertinent lorsque de nombreux paramètres sont utilisés. En effet, les éventuelles corrélations entre paramètres et l'aspect multidimensionnel sont totalement ignorés. Il faudrait donc :

³³ Recommandations pour la mise en œuvre de l'analyse statistique des données issues des études de toxicité sub-chronique de 90 jours chez le rat dans le cadre des demandes d'autorisation de mise sur le marché d'OGM : <http://www.afssa.fr/Documents/BIOT2009sa0285Ra.pdf>.

- pouvoir calculer des ellipses de confiance, afin de prendre en compte d'éventuelles corrélations entre paramètres,
 - corriger les intervalles de confiance afin de contrôler de façon correcte le risque de première espèce (intervalles multiples).
5. Ces méthodes de classification restent très empiriques et ne sont pas adaptées dans un contexte inférentiel, pour lequel il est nécessaire de calculer des degrés de signification (p-values) et/ou construire des intervalles de confiance. En effet, les lois des statistiques utilisées sont très mal connues et les techniques de type bootstrap ou jack-knife mises en œuvre pour calculer des intervalles de confiance n'ont pas de justification rigoureuse.

Au-delà du choix discutable de la méthode OPLS-DA dans le cadre de cette étude, une erreur méthodologique vient remettre en question les résultats présentés. En effet :

- i) 18 comparaisons entre groupes expérimentaux et groupes témoin sont proposés. Le groupe des femelles nourries avec un régime NK603 33% est celui qui présente le plus de différences : c'est donc ce groupe que les auteurs choisissent de présenter.
- ii) 48 paramètres sont comparés. Les paramètres biochimiques présentant le plus de différences (entre le groupe femelle NK603 33% et le groupe témoin) sont les paramètres *Na*, *Cl*, *U.Cl*, *U.N* tandis que les 2 hormones qui présentent le plus de différences sont *Testosterone* et *Estradiol* : ce sont donc ces 6 paramètres que les auteurs choisissent de présenter.

Il est alors attendu qu'en sélectionnant à la fois le groupe et les 6 paramètres qui présentent le plus de différences, des différences entre groupe expérimental et groupe témoin seront visibles. **Une telle approche ne permet pas :**

- **d'expliquer les différences observées par la différence de régime administré ;**
- **de rejeter l'hypothèse que c'est la variabilité naturelle des données (dues aux fluctuations d'échantillonnage) ainsi que le critère de sélection des paramètres (les paramètres présentant le plus de différences) qui expliquent les différences observées.**

5. Conclusion

Le protocole et les outils statistiques utilisés souffrent de graves lacunes et faiblesses méthodologiques qui ne permettent absolument pas de soutenir les conclusions avancées par les auteurs.

1. **Une analyse statistique rigoureuse des résultats obtenus lors de cette étude ne met en évidence :**
 - aucune différence statistiquement significative de la mortalité des rats dans les groupes témoin et expérimentaux,
 - aucune différence statistiquement significative des nombres de tumeurs dans les groupes témoin et expérimentaux.
2. **La méthodologie statistique employée pour l'analyse des paramètres biochimiques est inadéquate et ne permet pas de conclure à l'existence de différences statistiquement significatives entre les groupes témoin et expérimentaux.**