

AVIS

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

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.

RAPPEL DE LA SAISINE

L'Afssa s'est autosaisie sur l'analyse des pratiques actuelles en matière de traitement statistique des données de l'étude de toxicité sub-chronique de 90 jours chez le rat avec les objectifs suivants :

- 1) Etablir un état des lieux des tests qui sont ou qui pourraient être mis en œuvre pour analyser les données, en insistant sur leurs intérêts et limites.
- 2) Proposer une méthodologie adaptée aux objectifs des études de toxicité sub-chronique de 90 jours chez le rat.
- 3) Analyser la pertinence de ces outils à partir des données fournies par un pétitionnaire pour la demande de mise sur le marché d'un OGM particulier.
- 4) Formuler des recommandations que le pétitionnaire devrait suivre afin de faciliter l'évaluation de la qualité des protocoles mis en œuvre et des résultats obtenus.

CONTEXTE

L'évaluation des risques liés à la consommation de plantes génétiquement modifiées (PGM) ou d'aliments dérivés de la PGM suit une stratégie dite « d'équivalence substantielle¹ » basée sur la comparaison des caractéristiques moléculaires, agronomiques et de la composition chimique en vue de l'évaluation de la sécurité entre la PGM et son équivalent non transgénique².

En 2006, l'EFSA a défini les Lignes Directrices (EFSA 2006; EFSA 2008) relatives aux études à fournir par le pétitionnaire lors d'une demande de mise sur le marché, dont l'étude de toxicité sub-chronique de 90 jours chez le rongeur, qui s'applique au cas par cas.

Plus récemment, elle a précisé le plan expérimental et fait des recommandations pour l'analyse statistique des données issues des essais en champs, visant à comparer la composition des plantes GM à celle des plantes contrôles (EFSA 2009).

¹ **Equivalence substantielle** : la notion d'équivalence substantielle exprime l'idée que les organismes existants qui sont utilisés en tant qu'aliment ou source d'aliment peuvent servir de base à une comparaison lors de l'évaluation de l'innocuité d'un aliment nouveau ou modifié (OCDE 1993; WHO/FAO 2000)

² **Equivalent ou comparateur** : correspond à la plante non génétiquement modifiée ayant le fonds génétique le plus proche possible de la PGM.

Dès 2002, L'Afssa a identifié les points sensibles de l'évaluation des risques sanitaires liés à la consommation d'OGM ; elle a noté en particulier l'importance de la taille des effectifs mis en œuvre dans les essais sur animaux cibles, conditionnant la puissance statistique (AFSSA 2002).

Dans son avis de 2007, l'Agence recommande fortement la mise en œuvre d'une étude de toxicité sub-chronique 90 jours chez le rat, par administration orale de la plante *via* l'aliment, pour les événements primaires de transformation génétique (AFSSA 2007).

L'étude de toxicité sub-chronique, menée selon les recommandations de la norme OCDE 408 (OCDE 1998) a été définie initialement pour l'évaluation des substances chimiques. Elle a ensuite été appliquée à l'étude des « nouveaux aliments » avant leur mise sur le marché, ainsi qu'à celle des OGM.

Récemment, l'Agence a initié un travail pour mieux adapter le protocole de cette étude aux spécificités des OGM. Au cours de ce travail, il est apparu important d'approfondir les aspects liés à l'analyse statistique des données, les documents guides disponibles ne paraissant pas suffisamment explicites sur ce point.

L'objectif de cet avis est de formuler des recommandations pour l'analyse statistique des données du test de toxicité sub-chronique de 90 jours dans le cadre de l'évaluation de la sécurité d'une plante GM. Ces recommandations pourraient s'appliquer à toutes les situations dans lesquelles ce test est mis en œuvre.

METHODE D'EXPERTISE

Un groupe de travail incluant des statisticiens, biostatisticiens et toxicologues a été constitué. La démarche proposée est présentée en détail dans un rapport émis par le groupe de travail intitulé « **Bon usage des outils statistiques dans l'analyse 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** ». Ce rapport a été validé par le comité d'experts spécialisé « Biotechnologie », réuni le 16 décembre 2010.

Le présent avis est basé sur le rapport du groupe de travail.

ARGUMENTAIRE

Seront successivement envisagés 1) les principes du protocole et l'analyse de son contenu à travers les études publiées dans la littérature dans le cadre de l'évaluation de la sécurité des PGM, 2) les méthodes statistiques disponibles, leur intérêt et leurs limites et 3) une application de la méthodologie jugée la plus pertinente à un cas concret de PGM.

1. Principe du protocole OCDE 408 et application actuelle aux plantes GM.

Le protocole expérimental mis en œuvre actuellement est celui de l'OCDE 408, initialement destiné à l'évaluation de la toxicité des produits chimiques. Il recommande des effectifs d'au moins 10 animaux par sexe et par groupe, avec 3 doses du produit étudié et un groupe témoin. L'étude comporte des examens cliniques, hématologiques, biochimiques et histologiques.

Une revue³ des études publiées dans la littérature pour les plantes GM utilisant ce protocole a permis de dégager les points suivants :

- l'espèce utilisée est le rat, le plus souvent « Sprague-Dawley » ;
- les effectifs sont de 10 à 20 animaux de chaque sexe par groupe ;
- le nombre de « doses » est au plus de 2 (par exemple pour le maïs : 11 et 33%) ;

³ 17 articles publiés dans la littérature scientifique ont été analysés.

Les doses maximales incorporées dans les régimes dépendent de l'espèce végétale, la limite de la dose maximale étant celle du respect de l'équilibre alimentaire.

- les groupes témoins incluent toujours des rats nourris aux mêmes « doses » que l'OGM avec la plante quasi-isogénique et parfois avec des variétés commerciales ;
- les examens hématologiques et biochimiques (environ 50 paramètres) sont réalisés à une ou deux dates ;
- en fin d'étude, les organes sont pesés chez tous les animaux avec des examens macroscopique et microscopique (histologie) selon les principes habituels.

Sur le plan statistique, le protocole OCDE 408 n'est pas directif. Dans les publications analysées, les auteurs utilisent des tests statistiques de différence paramétriques ou non paramétriques. Les notions de puissance et d'équivalence ne sont pas évoquées. Les données de croissance et de consommation ne sont pas traitées selon un modèle adapté à la répétition des mesures dans le temps.

2. Intérêts et limites des différentes méthodes statistiques applicables.

2.1 Analyse descriptive des données

• Identification des données atypiques

Avant de procéder à toute analyse statistique, il est essentiel d'examiner les données mesurées pour évaluer leur qualité, leur variabilité ou leur distribution.

Les données dites « atypiques »⁴ peuvent avoir une influence importante sur certains paramètres statistiques, comme la moyenne et la variance, et invalider les résultats des tests statistiques paramétriques. Les données atypiques peuvent être identifiées, à l'aide de méthodes statistiques appropriées (le test de Grubbs ou test de Dixon) et présentées graphiquement pour juger facilement de leur répartition dans les différents groupes (traitement, sexe) ou période de traitement (en cours ou fin). La pertinence de leur élimination ou conservation peut alors être analysée par des toxicologues en tenant compte du phénomène à l'origine des valeurs extrêmes. Lorsqu'elles sont conservées, des méthodes statistiques adaptées doivent être utilisées.

• Transformation des données

Les tests paramétriques usuels de comparaison de moyennes et les calculs d'intervalle de confiance de la moyenne font l'hypothèse que les variables étudiées sont distribuées selon une loi normale. Dans certains cas, cette hypothèse n'est pas réaliste et il est alors possible de procéder à une transformation des données pour se rapprocher d'une distribution normale. La transformation est souvent nécessaire lorsque les tailles d'échantillons sont réduites.

2.2 Analyse des paramètres biologiques hors évolution pondérale

2.2.1 Test de différences

• Les hypothèses testées

Un test statistique est un outil d'aide à la décision permettant de rejeter ou de valider une hypothèse à partir d'observations. Classiquement, les tests statistiques, dits de différence fonctionnent sur le rejet ou l'acceptation de l'hypothèse nulle qui, dans le cas des études de toxicité sub-chronique de 90 jours chez le rat, correspond à l'absence d'effet du régime à base de PGM par rapport au régime à base de plante témoin.

Les conclusions des tests sont assorties de risque d'erreur. Le risque d'erreur appelé « risque de première espèce » (ou risque α) est le risque de conclure à tort à une différence entre les animaux ayant reçu le régime contenant la plante GM et ceux ayant reçu la plante témoin.

Le «risque de seconde espèce» (ou risque β) est le risque de ne pas détecter une différence entre les animaux ayant reçu le régime contenant la plante GM et ceux ayant reçu la plante témoin. Il est intéressant de pouvoir mieux maîtriser le niveau de risque β pour réduire le risque de conclure à tort à un résultat négatif. Les risques α et β évoluent en sens contraire, or si le risque de 1^{ère} espèce est contrôlé (le plus souvent, il est fixé à une valeur faible 5%), le risque de 2^{nde} espèce est par contre

⁴ Une *valeur atypique*, est une valeur qui s'écarte de façon marquée de l'ensemble des autres observations de l'échantillon. Elles peuvent résulter de la variabilité inhérente au critère biologique mesuré, d'une erreur de mesure, d'une erreur d'exécution.

généralement subi car il dépend, notamment, de l'importance de l'effet testé. La puissance, égale à $1 - \beta$, mesure la probabilité de conclure correctement à l'existence d'un effet donné.

Il est donc proposé pour réduire le risque β , d'augmenter le risque α (à 10% au lieu de 5%), ce qui conduira le toxicologue à examiner un nombre plus élevé de différences statistiquement significatives. Le niveau de vigilance est ainsi accru diminuant la probabilité de ne pas détecter une vraie différence (diminution du nombre de « faux négatifs »).

- **Les tests paramétriques et non paramétriques**

Les tests de différence peuvent être réalisés avec des méthodes paramétriques lorsque la structure des données l'autorise (distribution symétrique ou gaussienne, constance de la variance). La transformation des données peut parfois permettre de leur donner une distribution gaussienne. Généralement, les tests paramétriques sont les plus puissants et offrent la possibilité d'adapter l'analyse au plan expérimental de l'étude (par exemple, en prenant en compte des données répétées sur un même animal ou en testant les éventuelles interactions entre les différents facteurs de l'étude). Le recours aux tests non paramétriques se justifie lorsque les conditions d'application des tests paramétriques ne sont pas satisfaites, même après transformation des variables.

Les tests de différence, paramétriques ou non paramétriques, sont conçus pour mettre en évidence des différences entre les animaux ayant subi un traitement expérimental différent, mais ne peuvent pas apporter la preuve de l'absence de différence.

- **Conséquence de la multiplicité des tests sur le risque de première espèce**

La multiplicité des tests statistiques sur un même jeu de données augmente rapidement l'erreur de première espèce, c'est-à-dire la probabilité d'observer des différences faussement significatives. Les corrections de type FDR (false discovery rate) proposés par Benjamini (Benjamini and Hochberg 1995) peuvent être utilisées pour estimer le taux de faux positifs dans les résultats.

- **Conséquence d'un manque de puissance**

Lorsque la puissance est insuffisante, les tests de différence peuvent conduire à conclure de manière erronée à une absence d'effet du traitement. Une puissance faible peut résulter de la variabilité de la mesure ou d'une taille trop faible des échantillons. La puissance peut varier aussi en fonction de la valeur fixée pour le risque α de 1^{ère} espèce.

Il est habituellement recommandé de calculer la puissance du test *a priori* en fonction de la taille d'effet souhaitée, ce qui permet de déterminer la taille de l'échantillon nécessaire pour une puissance jugée suffisante.

Cependant, la puissance peut également être calculée *a posteriori* de façon à quantifier le risque de conclure à tort à l'absence d'effet du traitement. Il est alors nécessaire d'estimer avec une certaine probabilité, la taille d'effet minimum détectable entre les groupes d'animaux ayant reçu un régime contenant une PGM et ceux ayant reçu le régime contenant la plante témoin.

2.2.2 Test de non différence ou d'équivalence

L'utilisation de test d'équivalence a été récemment recommandée par l'EFSA dans le cadre de l'analyse comparée de composition chimique réalisée sur des plantes GM et témoins, cultivées simultanément en champs dans différentes conditions environnementales (EFSA 2009). Elle a aussi été proposée dans le cadre de l'évaluation du risque environnemental et plus spécifiquement pour évaluer l'impact potentiel des OGM sur les organismes non cibles (EFSA 2010).

Dans le cadre des essais toxicologiques sur animal de laboratoire, l'utilisation du test d'équivalence est délicate car il requiert un nombre important d'animaux par groupe et par sexe pour être interprétable.

2.3 Analyse des données de poids

La répétition dans le temps des mesures pondérales réalisées sur les rats nécessite l'utilisation des modèles mixtes adaptés aux données répétées sur le même animal. Divers logiciels sont disponibles pour l'analyse des courbes de croissance.

2.4 Conclusions sur cette analyse

Au regard des différents points examinés, et dans l'objectif d'optimiser l'interprétation des tests, il est proposé d'augmenter le risque de 1^{ère} espèce α , et parallèlement, d'utiliser des effectifs d'animaux suffisants dans les études expérimentales pour atteindre une puissance de test statistique satisfaisante. En augmentant la puissance des tests, le nombre de différences statistiquement significatives, potentiellement indicatrices de risque pour la santé, s'accroît. Le risque de conclure à tort à un résultat négatif est réduit.

La partie 3 ci-après, vise à tester cette méthodologie sur un jeu concret de données, afin d'en vérifier le côté opérationnel et d'estimer la taille minimum des effectifs d'animaux nécessaires pour atteindre une puissance suffisante.

3- Application des méthodes statistiques aux données d'une étude fournie pour la demande de mise sur le marché d'un OGM.

Les tests de différence ont été appliqués aux données de l'étude réalisée avec les maïs génétiquement modifiés MON810, fournies par le pétitionnaire à l'appui de la demande d'autorisation de mise sur le marché de ces maïs dans l'Union Européenne. Les résultats montrent l'importance d'une analyse descriptive approfondie pour rechercher des données potentiellement atypiques et pour vérifier les conditions d'application des tests paramétriques.

Dans cette application, la plupart des données expérimentales ont été transformées de façon à se rapprocher d'une distribution gaussienne.

Afin d'accroître la vigilance du toxicologue, mais aussi la puissance du test, le risque α de première espèce a été augmenté (10% au lieu de 5% habituellement).

Dans ces conditions, les résultats des tests paramétriques mettent en évidence 33 différences significatives sur 432 comparaisons « Régime OGM versus Régime quasi-isogénique ». Lorsque l'on contrôle le FDR, aucune différence n'est jugée significative. Les données de croissance pondérale ont été traitées par un modèle non linéaire mixte de type Mitscherlich qui tient compte de la répétition des données et aucune différence significative n'a été détectée.

Pour chaque paramètre, des calculs de puissance ont été réalisés sur la base d'une taille d'effet détectable, équivalente à la valeur de un écart-type calculé à partir des données des groupes nourris avec les variétés commerciales ou à partir de données historiques issues de la littérature. Cette démarche est nouvelle par rapport aux pratiques actuelles. Sur cette base, la puissance des tests de différence apparaît insuffisante pour certains paramètres, avec un risque de conclure à tort à une absence de différence.

La statistique étant un outil d'aide à la décision, toute différence significative ou toute puissance estimée insuffisante devra être examinée par le toxicologue 1) en considérant les paramètres individuellement, 2) en les analysant selon les critères classiques de faisceau de preuves et 3) en prenant en compte les données histologiques qui sont exclues de ces analyses statistiques. Cette démarche a conduit les toxicologues à estimer qu'il n'y avait pas de différences ayant un sens toxicologique entre les animaux recevant la plante GM étudiée et les animaux témoins recevant la plante quasi-isogénique.

Des calculs complémentaires ont, par ailleurs, montré que les valeurs de puissance atteignent des niveaux acceptables (supérieures à 80%) en faisant passer le nombre d'animaux par groupe et par sexe de 10 à 20. Cette augmentation de la taille des échantillons apparaît d'autant plus logique que certaines de ces études comportent déjà des groupes de 20 animaux par sexe, et que certains paramètres sont déjà mesurés sur ces 20 animaux. Dans ces conditions, la taille d'effet détectable est au moins égale à un écart type des données témoins.

Ce constat, issu de la prise en compte de la notion de puissance, conduit à une recommandation importante relative à l'évaluation de la puissance des tests. L'augmentation de la puissance est de nature à faciliter et à renforcer les conclusions des toxicologues.

Sur la base des seuils de tolérance précédemment définis et pour ce jeu de données, la mise en œuvre du test statistique d'équivalence nécessiterait des effectifs élevés (plusieurs centaines d'animaux par groupe, soit au moins 10 fois plus que ce qui est recommandé dans le protocole OCDE) pour permettre de conclure avec une puissance suffisante (80%).

Cette conclusion, qui soulève notamment des questions d'éthique animale en regard de la pertinence de l'information attendue, conduit à proposer une réflexion complémentaire sur l'applicabilité du test statistique d'équivalence aux études de toxicité sub-chronique de 90 jours chez le rat.

CONCLUSION ET RECOMMANDATIONS

L'Anses rappelle qu'elle recommande, dans le contexte de l'évaluation d'une nouvelle plante génétiquement modifiée (nouvel événement de transformation génétique), la mise en œuvre d'une étude de toxicité sub-chronique par administration de la plante *via* l'aliment, pendant 90 jours chez le rat.

Cette étude de toxicité sub-chronique s'appuie sur le référentiel OCDE 408. Au regard de l'expertise menée par l'Agence, il apparaît pertinent de compléter ce référentiel en particulier sur « l'analyse statistique des données » qui constitue une partie importante de la démarche d'évaluation de risque. L'objectif des recommandations listées ci-après, qui s'appuient sur une augmentation de la puissance des tests statistiques, est de proposer une méthodologie d'analyse statistique des données qui offre un haut niveau d'exigence.

Le référentiel d'évaluation de risque pour les OGM relevant de l'autorité européenne de sécurité des aliments (EFSA), les recommandations listées ci-dessous, constituent une contribution scientifique de l'Anses à l'EFSA. Ces recommandations, visant à améliorer l'analyse statistique des données, sont basées sur les neuf points suivants :

- Identifier et analyser les données atypiques, préalablement aux traitements statistiques, à l'aide de méthodes descriptives. Ces données doivent faire l'objet d'une analyse spécifique de la part des toxicologues.
- Transformer les distributions asymétriques pour se rapprocher de distributions gaussiennes et ceci d'autant plus que la taille des échantillons est faible.
- Utiliser des modèles statistiques adaptés au plan d'expérience et aux mesures réalisées :
 - l'utilisation de modèles mixtes (linéaires ou non linéaires) si les mesures sont répétées sur les mêmes individus à plusieurs dates ;
 - l'utilisation de modèles gaussiens si les distributions des données sont symétriques (éventuellement après transformation) ;
 - l'utilisation de méthodes non paramétriques si les distributions sont asymétriques.
- Accepter un risque de 1ère espèce de 10% (supérieur au niveau usuel de 5 %). La conséquence attendue est une augmentation du nombre de différences statistiquement significatives ; elles sont destinées à accroître la vigilance du toxicologue sur un plus grand nombre de paramètres dont il aura à juger la pertinence.
- Evaluer les conséquences de la multiplicité des tests statistiques de différence et estimer la probabilité d'obtenir des résultats faussement positifs (FDR).
- Evaluer la puissance des tests de différence en calculant pour chaque paramètre biologique, la taille d'effet minimale détectable entre les traitements conduisant à un test de différence significatif avec une probabilité de 0,8 (i.e. une puissance de 80%). Les toxicologues jugeront si les différences détectables pour une puissance de 80% ont une signification toxicologique sur la base de leur expertise et des données historiques disponibles ou internes à l'étude (groupes nourris avec des variétés commerciales). En considérant un seuil de tolérance conservateur égal à un écart-type calculé à partir des données des groupes nourris avec les variétés commerciales, il est recommandé d'utiliser 20 animaux par groupe (un traitement, un sexe, une dose) pour atteindre une puissance de 80% pour la quasi totalité des paramètres. Le choix d'autres seuils de tolérance pourrait conduire à des tailles d'échantillon différentes.

- Les conclusions des études utilisant le terme « d'équivalence entre les deux régimes » devront être justifiées par des tests d'équivalence. Une réflexion sur l'applicabilité de ce test aux études de toxicité sub-chronique de 90 jours chez le rat devrait être menée.
- Faciliter l'interprétation des résultats à l'aide de représentations graphiques notamment sur les tailles d'effet.
- Disposer des données brutes sous forme numérique pour permettre toute vérification ou analyse complémentaire jugée nécessaire par les experts.

L'adoption de ces recommandations devrait conduire à une présentation plus explicite des résultats des études de toxicité sub-chronique de 90 jours chez le rat, notamment en identifiant les incertitudes. Ainsi, l'interprétation des résultats par les toxicologues gagnera en rapidité, objectivité et reproductibilité.

Le directeur général

Marc MORTUREUX

MOTS-CLÉS

Mots clés : OGM, étude de toxicité sub-chronique de 90 jours, analyse statistique, OCDE 408.

BIBLIOGRAPHIE

AFSSA (2002) Evaluation des risques relatifs à la consommation de produits alimentaires composés ou issus d'organismes génétiquement modifiés.

AFSSA (2007) Avis relatif aux études de toxicité réalisées dans le cadre des demandes de mises sur le marché d'OGM.

Benjamini Y, Hochberg D (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289-300.

EFSA (2006) Guidance document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. *The EFSA Journal* **99**, 1-100.

EFSA (2008) **DRAFT** Updated Guidance document for the risk assessment of genetically modified plants and derived food and feed *The EFSA Journal* **727**, 1-135.

EFSA (2009) Scientific opinion on statistical considerations for the safety evaluation of GMOs, on request of EFSA. *The EFSA Journal* **1250**, 1-62.

EFSA (2010) Scientific opinion on the assessment of potential impacts of genetically modified plants on non-target organisms *The EFSA Journal* **8**, 1877.

OCDE (1993) Safety evaluation of food derived by modern biotechnology : concept and principles. *Paris, France*.

OCDE (1998) Guideline for testing of chemicals N°408. Repeated Dose 90-day Oral Toxicity Study in Rodents. *Paris, France*.

WHO/FAO (2000) Safety aspects of genetically modified foods of plant origin. Report of a joint FAO/WHO expert consultation on foods derived from Biotechnology. *WHO/FAO* **2000**(29).